Robust estimates in generalized partially linear single-index models

Graciela Boente & Daniela Rodriguez

TEST
An Official Journal of the Spanish Society of Statistics and Operations Research

ISSN 1133-0686
Volume 21
Number 2

TEST (2012) 21:386-411
DOI 10.1007/s11749-011-0249-z

Test (2012) 21:386–411
DOI 10.1007/s11749-011-0249-z

ORIGINAL PAPER

Robust estimates in generalized partially linear single-index models

Graciela Boente · Daniela Rodriguez

Received: 6 July 2010 / Accepted: 29 April 2011 / Published online: 1 June 2011

Abstract A natural generalization of the well known generalized linear models is to allow only for some of the predictors to be modeled linearly while others are modeled nonparametrically. However, this model can face the so called "curse of dimensionality" problem that can be solved by imposing a nonparametric dependence on some unknown projection of the carriers. More precisely, we assume that the observations $(y_i, \mathbf{x}_i, \mathbf{t}_i)$, $1 \leq i \leq n$, are such that $\mathbf{t}_i \in \mathbb{R}^q$, $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i|(\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$ with $\mu_i = H(\eta(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i) + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$, for some known distribution function F and link function H. The function $\eta : \mathbb{R} \to \mathbb{R}$ and the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown and to be estimated. This model is known as the *generalized partly linear single-index* model.

In this paper, we introduce a family of robust estimates for the parametric and nonparametric components under a generalized partially linear single-index model. It is shown that the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are root-n consistent and asymptotically normally distributed. Through a Monte Carlo study, we compare the performance of the proposed estimators with that of the classical ones.

Keywords Asymptotic properties · Generalized partly linear single-index models · Rate of convergence · Robust estimation · Smoothing techniques

Mathematics Subject Classification (2000) 62F35 · 62H25

Communicated by Domingo Morales.

G. Boente · D. Rodriguez (✉)
Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Pabellón 2, Buenos Aires, 1428, Argentina
e-mail: drodrig@dm.uba.ar

G. Boente
e-mail: gboente@dm.uba.ar

1 Introduction

Generalized linear models (GLM, McCullagh and Nelder 1989) extend linear models to allow for parametrically modeling the relation between a transformation of the mean response and some covariates. They assume that the observations $(y_i, \mathbf{x}_i, t_i)$, $1 \leq i \leq n$, $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{t}_i \in \mathbb{R}^q$, are independent such that the conditional distribution of $y_i|(\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$ belongs to the canonical exponential family. In this situation, the mean $\mu_i = \mu(\mathbf{x}_i, \mathbf{t}_i) = \mathbb{E}(y_i|(\mathbf{x}_i, \mathbf{t}_i))$ is modeled linearly through a known link function, H, i.e., $\mu_i = H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i)$.

However, in some situations, the relationship between the response and covariates may be so complex that the linear relation stated by the GLM is not enough to capture it. Different approaches were given to face this lack of linearity and to solve problems arising when a misspecified model is fitted. A fully nonparametric model can be considered, but, as is well known, this model faces the so called "curse of dimensionality". In an attempt to solve this problem and to preserve, in some sense, the easy interpretation of the generalized linear models, Hastie and Tibshirani (1990) introduced the generalized additive model (GAM) by assuming that a transformation of the mean response can be written as a sum of nonparametric components of the predictors. Hence, under a GAM we have $\mu_i = H(\beta_0 + \sum_{j=1}^{p} \nu_j(x_{ij}) + \sum_{\ell=1}^{q} \eta_\ell(t_{i\ell}))$ where $\nu_j : \mathbb{R} \to \mathbb{R}$ and $\eta_\ell : \mathbb{R} \to \mathbb{R}$ are smooth functions with $\mathbb{E}(\nu_j(x_{ij})) = 0$ and $\mathbb{E}(\eta_\ell(t_{i\ell})) = 0$. This model provides a helpful generalization of the usual generalized linear model and a way of modeling data that does not conform the linear assumption present in the GLM.

Even if the generality of the GAM is attractive, one should keep in mind that the practitioner may lose some precision and power if a nonparametric component is adopted when a linear or other parametric term is appropriate. In this case, semi-parametric models provide a solution. To improve the efficiency of the generalized additive models but still keeping some flexibility, Severini and Staniswalis (1994) and Härdle et al. (1998) studied generalized partially linear models which allow modeling, through the link function H, the mean of the response linearly on some of the carriers and nonparametrically on the remaining ones. To be more precise, generalized partially linear models assume that $\mathbf{t}_i \in \mathbb{R}$, i.e., $q = 1$ and that $\mu_i = H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + \eta(\mathbf{t}_i))$. The function $\eta : \mathbb{R} \to \mathbb{R}$ and the parameters $\boldsymbol{\beta}$ are unknown and to be estimated. These models offer a flexible alternative to generalized linear models but they may be insufficient to explain the relationship between the response variable and its associated covariables if more than one covariate enters into the model nonlinearly.

A natural generalization is to study estimators under a generalized partially linear model when the carriers $\mathbf{t}$ take values in $\mathbb{R}^q$, in which case, one may consider similar estimators to those defined when $\mathbf{t} \in \mathbb{R}$. However, as in a fully nonparametric model, this approach involves a smoothing procedure in a q-dimensional space and so, the nonparametric components are subject to the "curse of dimensionality" and can only accommodate low dimensional covariates $\mathbf{t}$. Hence, to reduce the dimensionality of the nonparametric part and to mimic the generalized additive model, one may consider a model in which the dependence on $\mathbf{t} = (t_1, \ldots, t_q)^{\mathrm{T}} \in \mathbb{R}^q$ is explained nonparametrically by each of the components of the vector. In this situation, $\mu_i = H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + \eta(\mathbf{t}_i))$ but the nonparametric dependence on the carriers $\mathbf{t}$ satisfies the additive model $\eta(\mathbf{t}) = \sum_{\ell=1}^{q} \eta_\ell(t_\ell)$ with $\eta_\ell : \mathbb{R} \to \mathbb{R}$ smooth functions. These models usually called generalized additive partially linear models are an extension of the generalized additive models because some covariates can be modeled linearly and we refer to Härdle et al. (2006) for a discussion on estimation procedures.

As with generalized additive models, the flexibility of the generalized additive partially linear models has the disadvantage of introducing a loss of efficiency when the dependence on the covariates $\mathbf{t}$ can be done through a smaller number of projections. To remedy this, the generalized partially linear single-index model (GPLSIM), introduced by Carroll et al. (1997), provides a dimension reduction model since it assumes that the influence of the covariate $\mathbf{t}$ can be collapsed to a single index, $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}$. These models follow similar ideas to those considered by Friedman and Stuetzle (1981) in projection pursuit regression that involves a small number of nonparametric functions of linear combinations $\boldsymbol{\alpha}_\ell^{\mathrm{T}}\mathbf{t}$ of the covariates $\mathbf{t}$. In this way, the GPLSIM allows some reduction in the dimensionality of the space in which the nonparametric estimation is carried out. To be more precise, generalized partially linear single-index models assume that $y|(\mathbf{x}, \mathbf{t}) \sim F(\cdot, \mu(\mathbf{x}, \mathbf{t}))$ with $\mu(\mathbf{x}, \mathbf{t}) = H(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} + \eta(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}))$ and $\mathrm{VAR}(y|(\mathbf{x}, \mathbf{t})) = V(\mu(\mathbf{x}, \mathbf{t}))$ where H is a given link function, V is a known function while $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\alpha} \in \mathbb{R}^q$ are unknown parameters and η is an unknown continuous function. These models are useful to make inference on the effects $\mathbf{x}$, by making minimal assumptions on $\mathbf{t}$, when the covariates $\mathbf{t}$ have large dimension and are of little interest. In this situation, the component $\eta(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})$ may be seen as a nuisance parameter. On the other hand, since a smooth function η is applied to the index $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}$, interactions between theses covariates can be modeled. Thus, generalized partially linear single-index models are a useful alternative to generalized additive partially linear models, which also reduce dimensionality but do not incorporate interactions.

Note that when $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ is assumed to be known, the GPLSIM becomes the GPLM with covariates $(\mathbf{x}, \boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t})^{\mathrm{T}}$. On the other hand, when η is the identity function, the GPLSIM is the GLM. Clearly, to identify the parameter $\boldsymbol{\alpha}$ and the function η, some restrictions need to be introduced. As in Carroll et al. (1997) where estimators for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and η under a GPLSIM are defined, we will assume that $\|\boldsymbol{\alpha}\| = 1$ where $\| \cdot \|$ stands for the Euclidean norm. Moreover, as is usual in partially linear models, we will assume that the vector $\mathbf{1}_n$ is not in the space spanned by the column vectors of $\mathbf{x}$, that is, we do not allow $\boldsymbol{\beta}$ to include an intercept so that the model is identifiable. Due to the generality of the GPLSIM, identifiability implies that only "slope" coefficients can be estimated. Moreover, we do not allow any linear combination of $\mathbf{x}$ to be predicted by $\mathbf{t}$, otherwise, the model will be purely nonparametric and $\boldsymbol{\beta}$ will not be identifiable (see Robinson 1988). For further discussion on models related to the generalized partially linear single-index model, we also refer to Xia et al. (1999), Delecroix et al. (2003), Xia and Härdle (2006), Yi et al. (2009) and Wang et al. (2010), among others.

As is well known, in generalized linear models, large deviations of the response from its mean, as measured by the Pearson residuals, or outlying points in the covariate space can have a large influence on the classical estimators based on the quasi-likelihood. Robust estimators in GLM were considered by Stefanski et al. (1986), Künsch et al. (1989), Bianco and Yohai (1995), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco et al. (2005).

As in generalized linear models, in a semiparametric setting, outliers can also have a devastating effect. In this setting, extreme points on the carriers can easily affect the scale and the shape of the function estimate of η, leading to possible wrong conclusions on $\boldsymbol{\beta}$. Robust proposals for generalized partial linear models were introduced by Boente et al. (2006) and Boente and Rodriguez (2010). On the other hand, a robust generalized estimating equations approach, for generalized partially linear models with clustered data, using regression splines and Pearson residuals was given in He et al. (2002) while a robust proposal for generalized additive models was recently considered in Azadeh and Salibian-Barrera (2011). Since they are based on a quasi-likelihood approach, the estimators introduced in Carroll et al. (1997) may be sensitive to outliers, in particular, when high leverage points in the covariates $\mathbf{x}$ are present. This sensitivity to outliers in the covariates is illustrated in Sect. 5 under a logistic model. The goal of this paper is to introduce robust procedures under a generalized partially linear single-index model to provide reliable estimators of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and the function η, when outliers are present in the sample.

This paper is organized as follows. The robust proposal is described in Sect. 2. In Sect. 3, we state results related to the consistency of the estimators while the asymptotic distribution of the estimators of the regression parameter $\boldsymbol{\beta}$ and the index parameter $\boldsymbol{\alpha}$ are derived in Sect. 4. The results of a Monte Carlo study are summarized in Sect. 5, while proofs are relegated to the Appendix.

2 The proposal

Let $(y_i, \mathbf{x}_i, \mathbf{t}_i) \in \mathbb{R}^{p+q+1}$ be independent observations such that $y_i|(\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$ with $\mu_i = H(\eta(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i) + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ and $\mathrm{VAR}(y_i|(\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$ for some known function V. Let $\eta_0(t)$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ denote the true parameter values and $\mathbb{E}_0$ the expectation under the true model, thus $\mathbb{E}_0(y|(\mathbf{x}, t)) = H(\eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$.

Let $w_i : \mathbb{R}^p \to \mathbb{R}$ for $i = 1, 2$ be weight functions to control leverage points on the carriers $\mathbf{x}$ and $\rho : \mathbb{R}^2 \to \mathbb{R}$ a loss function. We define for each $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and any continuous function $v : \mathbb{R} \to \mathbb{R}$ the functions

$$R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \mathbb{E}_0\left[\rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)w_1(\mathbf{x})\middle|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t} = u\right], \tag{1}$$

$$G(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \mathbb{E}_0\left[\rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + v\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}\right)\right)w_2(\mathbf{x})\right]. \tag{2}$$

Denote by $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u) = \operatorname{argmin}_{a\in\mathbb{R}} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$. Throughout the paper, we will assume Fisher consistency, i.e., that $w_2(\cdot)$ and $\rho(\cdot)$ are such that $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) = \operatorname{argmin}_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in\mathbb{R}^{p+q}} G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_0)$ being the unique minimum. Conditions ensuring Fisher consistency under the generalized linear model have been studied by several authors such as Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco et al. (2005). Under a generalized partially linear model, when $\mathbf{t}_1 \in \mathbb{R}$, i.e., $q = 1$, Boente et al. (2006) and Boente and Rodriguez (2010) showed that if

$$P\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta} = \gamma\middle|\mathbf{t}_1 = \tau\right) < 1\,, \tag{3}$$

for any $(\boldsymbol{\beta}, \gamma) \neq 0$ and τ in the support of $\mathbf{t}_1$, Fisher consistency holds under the same regularity conditions on the loss function stated for generalized linear models by the above mentioned authors. Standard arguments allow one to show that under (3) Fisher consistency still holds under a GPLSIM for the families of loss functions studied in Bianco and Yohai (1995) and Croux and Haesbroeck (2002) for the logistic model and by Bianco et al. (2005) for the gamma model. Note that (3) does not allow $\boldsymbol{\beta}_0$ to include an intercept, so that the model will be identifiable, as mentioned in the Introduction. On the other hand, as is well known, when $H(u) = u$, i.e., under the partially linear single index model $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i) + \epsilon_i$, the standard choice for the loss function is $\rho(y, u) = \rho_0(y - u)$ with $\rho_0 : \mathbb{R} \to \mathbb{R}$ a loss function. In this situation, Fisher consistency holds if, for instance, the errors ϵ_i have a symmetric distribution and the score function $\psi_0(u) = (\partial\rho_0)/(\partial u)$ is odd.

Boente et al. (2006) proposed two classes of loss functions ρ. The first one aims to bound the deviances, while the second one introduced by Cantoni and Ronchetti (2001) bounds the Pearson residuals. We refer to Boente et al. (2006) or Boente and Rodriguez (2010) for their definition.

In order to define consistent estimators of the parametric and nonparametric components, let us consider the empirical versions of the objective functions (1) and (2), respectively, as

$$R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \sum_{i=1}^{n} W_{\boldsymbol{\alpha}, i}(u)\rho\big(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + a\big)w_1(\mathbf{x}_i), \tag{4}$$

$$G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \frac{1}{n}\sum_{i=1}^{n} \rho\big(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + v\big(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i), \tag{5}$$

where v is any continuous function $v : \mathbb{R} \to \mathbb{R}$, the functions w_1, w_2 and ρ are chosen to guarantee Fisher consistency and $W_{\boldsymbol{\alpha}, i}(u)$ are weights depending on the closeness between u and the projection of $\mathbf{t}$ on the direction $\boldsymbol{\alpha}$, i.e., between u and $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}$. For the sake of simplicity, throughout this paper, $W_{\boldsymbol{\alpha}, i}(u)$ will be taken as the kernel weights, i.e.,

$$W_{\boldsymbol{\alpha}, i}(u) = K\left(\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i - u}{h}\right)\left\{\sum_{j=1}^{n} K\left(\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_j - u}{h}\right)\right\}^{-1}.$$

The estimation procedure to estimate $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ and η_0 can thus be defined as follows.

Step 1: Compute an initial robust consistent estimator of $\boldsymbol{\beta}_0$ and an initial robust consistent and equivariant by orthogonal transformations estimator of $\boldsymbol{\alpha}_0$, denoted, respectively, $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\hat{\boldsymbol{\alpha}}_{\mathrm{R1}}$. Let $\hat{\boldsymbol{\alpha}}_{\mathrm{R}} = \hat{\boldsymbol{\alpha}}_{\mathrm{R1}}/\|\hat{\boldsymbol{\alpha}}_{\mathrm{R1}}\|$.

Step 2: Define the estimator $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}(u)$ of η_0 as

$$\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}(u) = \operatorname*{argmin}_{a \in \mathbb{R}} R_n(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}, a, u). \tag{6}$$

Step 3: Define estimators $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ as

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \operatorname*{argmin}_{\boldsymbol{\alpha}, \boldsymbol{\beta}} G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}). \tag{7}$$

The estimation procedure involves initial estimators of the parametric components that might be inefficient. However, they need to be robust, consistent and, in the case of the projection index, equivariant by orthogonal transformations. To guarantee their existence we will introduce an estimation procedure based on a robustified profile method. Profile likelihood procedures were studied by van der Vaart (1988) and applied to generalized partially linear models by Severini and Wong (1992) and Severini and Staniswalis (1994). The estimators defined in Boente et al. (2006) for the GPLM correspond to a robustified profile method. As noted by Boente and Rodriguez (2010), the main disadvantage of the estimators proposed by Boente et al. (2006) is that their asymptotic covariance matrix depend on the derivatives of the robust profile regression function with respect to $\boldsymbol{\beta}$ making difficult its estimation. For the GPLSIM, the *robustified profile method* can thus be defined as

Step P1: For each fixed $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, with $\|\boldsymbol{\alpha}\| = 1$, let

$$\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}}\, R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u). \tag{8}$$

Step P2: Define the estimators $(\hat{\boldsymbol{\alpha}}_{\text{PR}}, \hat{\boldsymbol{\beta}}_{\text{PR}})$ of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ as

$$(\hat{\boldsymbol{\alpha}}_{\text{PR}}, \hat{\boldsymbol{\beta}}_{\text{PR}}) = \underset{\|\boldsymbol{\alpha}\|=1,\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}}\; G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}). \tag{9}$$

Rodriguez (2008) studied some properties of these estimators, such as consistency and equivariance by orthogonal transformations, leading to robust initial estimators as required in Step 1. Throughout the paper, we will only focus on the asymptotic properties of the estimators defined through Steps 1 to 3.

When ρ is continuously differentiable, let us denote by $\Psi(y,u) = \partial\rho(y,u)/\partial u$ and by

$$R^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \mathbb{E}_0\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)w_1(\mathbf{x})\middle|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t} = u\right),$$

$$R_n^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \sum_{i=1}^{n} W_{\boldsymbol{\alpha},i}(u)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + a\right)w_1(\mathbf{x}_i).$$

Then, $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ and $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\text{R}},\hat{\boldsymbol{\beta}}_{\text{R}}}(u)$ satisfy the differentiated equations $R^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u), u) = 0$ and $R_n^{(1)}(\hat{\boldsymbol{\alpha}}_{\text{R}}, \hat{\boldsymbol{\beta}}_{\text{R}}, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\text{R}},\hat{\boldsymbol{\beta}}_{\text{R}}}(u), u) = 0$, respectively. Besides, using the fact that $\|\boldsymbol{\alpha}_0\| = 1$ and $\|\hat{\boldsymbol{\alpha}}\| = 1$, we see that $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ and $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ satisfy $G^{(1)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}) = \mathbf{0}$ and $G_n^{(1)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \eta_{\hat{\boldsymbol{\alpha}}_{\text{R}},\hat{\boldsymbol{\beta}}_{\text{R}}}) = \mathbf{0}$, respectively, with

$$G^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \mathbb{E}_0\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + v\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}\right)\right)w_2(\mathbf{x})\begin{pmatrix}\mathbf{x}\\ v'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t})\mathbf{t}\end{pmatrix}\right) + \theta\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\alpha}\end{pmatrix}, \tag{10}$$

$$G_n^{(1)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, v) = \frac{1}{n}\sum_{i=1}^{n}\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + v\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i\right)\right)w_2(\mathbf{x}_i)\begin{pmatrix}\mathbf{x}_i\\ v'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i)\mathbf{t}_i\end{pmatrix} + \theta_n\begin{pmatrix}\mathbf{0}\\ \boldsymbol{\alpha}\end{pmatrix}, \tag{11}$$

where θ and θ_n are the Lagrange multipliers.

Remark 2.1 It is easy to see that the estimators of the projection index $\boldsymbol{\alpha}$ are equivariant by orthogonal transformations. More precisely, we see that, if $\tilde{\mathbf{t}}_i = \boldsymbol{\Gamma}\mathbf{t}_i$ is an orthogonal transformation of the predictors $\mathbf{t}_i$, then the projection index for the transformed variables $(y_1, \mathbf{x}_1, \tilde{\mathbf{t}}_1)$ is $\tilde{\boldsymbol{\alpha}}_0 = \boldsymbol{\Gamma}\boldsymbol{\alpha}_0$ while the linear regression parameter remains unchanged. It is easy to see that if we compute the estimators $(\hat{\tilde{\boldsymbol{\alpha}}}, \hat{\tilde{\boldsymbol{\beta}}})$ of $(\tilde{\boldsymbol{\alpha}}_0, \boldsymbol{\beta}_0)$ based on the new sample $(y_i, \mathbf{x}_i, \tilde{\mathbf{t}}_i)$, then $\hat{\tilde{\boldsymbol{\alpha}}} = \boldsymbol{\Gamma}\hat{\boldsymbol{\alpha}}$ and $\hat{\tilde{\boldsymbol{\beta}}} = \hat{\boldsymbol{\beta}}$. The proof of this assertion can be found in Rodriguez (2008).

3 Consistency

In this section, we will derive, under some regularity conditions, the consistency of the estimators defined in Sect. 2 through Steps 1 to 3. We will assume that $\mathbf{t} \in \mathcal{T} \subset \mathbb{R}^q$. Let $\mathcal{T}_0 \subset \mathcal{T}$ be a compact set and define the set $\mathcal{U}(\mathcal{T}_0) = \{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t} : \mathbf{t} \in \mathcal{T}_0, \|\boldsymbol{\alpha}\| = 1\}$. For any continuous function $v : \mathcal{U}(\mathcal{T}_0) \to \mathbb{R}$ denote $\|v\|_{0,\infty} = \sup_{u \in \mathcal{U}(\mathcal{T}_0)} |v(u)|$. From now on, $\mathcal{S}_1$ will stand for the unit ball in $\mathbb{R}^q$, i.e., $\mathcal{S}_1 = \{\boldsymbol{\alpha} \in \mathbb{R}^q \ \|\boldsymbol{\alpha}\| = 1\}$.

D1. $\rho(y, a)$ is a continuous and bounded function. Moreover, w_1 and w_2 are non-negative bounded functions.

D2. The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even, nonnegative, continuous and bounded function, with bounded variation, satisfying $\int K(u)\,du = 1$, $\int u^2 K(u)\,du < \infty$ and $|u|K(u) \to 0$ as $|u| \to \infty$.

D3. The bandwidth sequence h_n is such that $h_n \to 0$, $nh_n/\log(n) \to \infty$ when $n \to \infty$.

D4. (i) The marginal density $f_{\mathbf{T}}$ of $\mathbf{t}_1$ is a bounded function in $\mathcal{T}$.
(ii) Given any compact set $\mathcal{T}_0 \subset \mathcal{T}$, there exists a positive constant $A_1(\mathcal{U}(\mathcal{T}_0))$ such that $A_1(\mathcal{U}(\mathcal{T}_0)) < f_{\boldsymbol{\alpha}}(u)$ for all $u \in \mathcal{U}(\mathcal{T}_0)$ and $\|\boldsymbol{\alpha}\| = 1$, where $f_{\boldsymbol{\alpha}}$ is the marginal density of $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_1$.

D5. The function $R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$ satisfies the following equicontinuity condition: given $\mathcal{T}_0 \subset \mathcal{T}$ and $\mathcal{K} \subset \mathbb{R}^p$ compact sets, for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $u_1, u_2 \in \mathcal{U}(\mathcal{T}_0)$; $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathcal{S}_1$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{K}$,

$$|u_1 - u_2| < \delta,\ \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta \ \text{ and } \|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\| < \delta$$
$$\Rightarrow \quad \sup_{a \in \mathbb{R}} \big| R(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1, a, u_1) - R(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2, a, u_2) \big| < \epsilon.$$

D6. The function $R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$ is continuous and $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ is a continuous function on $(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$.

D7. The initial estimators $(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}})$ are consistent estimators of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$.

Remark 3.1

– Assumption D1 is fulfilled for the score functions usually considered to obtain robust estimators in generalized linear models, such as the function introduced by Croux and Haesbroeck (2002) for the logistic model and the Tukey's bisquare function used in Bianco et al. (2005) for the Gamma model.

- Assumptions D2 and D3 are standard conditions in the nonparametric setting. The first one is fulfilled, for instance, for the Epanechnikov kernel $K(u) = (3/4)(1 - u^2)\mathrm{I}_{[-1,1]}(u)$ and also for the Gaussian kernel while the latter is satisfied, for instance, if $h_n = n^{-\alpha}$ for $\alpha > 0$. Besides, assumption D3 ensures that for each fixed a, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ we have convergence of the kernel estimates $R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)$ to their mean, while D5 guarantees that the bias term converges to 0.
- Assumption D4 is a standard condition in semiparametric models. D4(ii) holds, for instance, if $f_{\mathbf{T}}(\mathbf{t}) > B_1(\mathcal{T}_0)\ \forall \mathbf{t} \in \mathcal{T}_0$. As mentioned in Boente et al. (2006), in the classical case and for generalized partially linear models, i.e., when $\boldsymbol{\alpha}$ is known, it corresponds to Condition (D) of Severini and Staniswalis (1994, p. 511).
- Assumption D5 is fulfilled under D1 if the conditional distribution of $\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t} = u$ is continuous with respect to $(u, \boldsymbol{\alpha})$ and if the following equicontinuity condition holds: for any $\epsilon > 0$ there exist compact sets $\mathcal{K}_1 \subset \mathbb{R}$ and $\mathcal{K}_p \subset \mathbb{R}^p$ such that for any $u \in \mathcal{U}(\mathcal{T}_0)$, $P((y, \mathbf{x}) \in \mathcal{K}_1 \times \mathcal{K}_p | \boldsymbol{\alpha}^{\mathrm{T}} t = u) > 1 - \epsilon$, for any $\boldsymbol{\alpha} \in \mathcal{S}_1$ which holds for instance if, for $1 \le i \le n$ and $1 \le j \le p$, $x_{ij} = \phi_j(\mathbf{t}_i) + u_{ij}$, where ϕ_j are continuous functions and u_{ij} are i.i.d. and independent of $\mathbf{t}_i$.
- Under D1, assumption D6 is fulfilled if the conditional distribution of $\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t} = u$ is continuous with respect to $(u, \boldsymbol{\alpha})$. Moreover, if besides ρ is a continuously differentiable function, the implicit function theorem implies that $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ is continuous.

If we considered the initial estimators defined through Steps P1 and P2, D7 is satisfied, see Rodriguez (2008).

Lemma 3.1 *Let $\mathcal{K} \subset \mathbb{R}^p$ and $\mathcal{T}_0 \subset \mathcal{T}$ be compact sets and assume that there exists $\delta_0 > 0$ such that $\mathcal{T}_{\delta_0} \subset \mathcal{T}$, where $\mathcal{T}_{\delta_0}$ stands for the closure of a δ_0-neighborhood of $\mathcal{T}_0$. Assume that* D1 *to* D6 *hold and that the family of functions $\mathcal{F} = \{f(y, \mathbf{x}) = \rho(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a)w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathcal{K}, a \in \mathbb{R}\}$ has a covering number satisfying $\sup_{\mathbb{Q}} N(\epsilon, \mathcal{F}, L^1(\mathbb{Q})) \le A\epsilon^{-W}$, for any $0 < \epsilon < 1$ and some positive constants A and W where $\mathbb{Q}$ stands for any probability measure for $(y, \mathbf{x})$. Then, we have*

(a) $$\sup_{a\in\mathbb{R},\boldsymbol{\alpha}\in\mathcal{S}_1,\boldsymbol{\beta}\in\mathcal{K}} \left\| R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) \right\|_{0,\infty} \xrightarrow{\text{a.s.}} 0.$$

(b) *If* $$\inf_{\boldsymbol{\alpha}\in\mathcal{S}_1,\ \boldsymbol{\beta}\in\mathcal{K}, u\in\mathcal{U}(\mathcal{T}_0)} \left[\lim_{|a|\to\infty} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) - R\big(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u), u\big) \right] > 0,$$

then $$\sup_{\boldsymbol{\alpha}\in\mathcal{S}_1,\ \boldsymbol{\beta}\in\mathcal{K}} \|\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}} - \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}\|_{0,\infty} \xrightarrow{\text{a.s.}} 0.$$

Remark 3.2 The requirement

$$\inf_{\boldsymbol{\alpha}\in\mathcal{S}_1,\ \boldsymbol{\beta}\in\mathcal{K},\ u\in\mathcal{U}(\mathcal{T}_0)} \left[\lim_{|a|\to\infty} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) - R\big(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u), u\big) \right] > 0 \tag{12}$$

in Lemma 3.1(b) is a natural extension to generalized partially linear single-index models of the condition stated in Boente et al. (2006) for generalized linear models. It ensures that the infimum of objective function (1) is not attained at infinite. The uniformity on $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and u is needed to attain uniform convergence of $\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ to $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}$. For the score functions usually considered in robustness such as those defined in Croux and Haesbroeck (2002) for the logistic model and by Bianco et al. (2005)

for the Gamma model, it is easy to see that $\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, u) = \lim_{|a|\to\infty} R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u), u) > 0$. Using that $\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$ is a continuous function and the fact $\mathcal{S}_1$, $\mathcal{K}$ and $\mathcal{U}(\mathcal{T}_0)$ are compact sets, it follows that (12) is fulfilled.

Lemma 3.2 *Let $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$ be defined in* (7) *where $\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ satisfies*

$$\sup_{\boldsymbol{\alpha}\in\mathcal{S}_1, \boldsymbol{\beta}\in\mathcal{K}} \|\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}} - \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}\|_{0,\infty} \xrightarrow{\text{a.s.}} 0. \tag{13}$$

Assume that D1 *and* D7 *hold and that $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_0)$ has a unique minimum at* $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$. *Then, we have*

(a) $\sup_{\mathbf{a},\boldsymbol{\alpha}\in\mathcal{S}_1;\mathbf{b},\boldsymbol{\beta}\in\mathcal{K}} |G_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\eta}_{\mathbf{a},\mathbf{b}}) - G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_{\mathbf{a},\mathbf{b}})| \xrightarrow{\text{a.s.}} 0$ *for any compact set $\mathcal{K}$.*

(b) *If, in addition, there exists a compact set $\mathcal{K}_1$ such that* $\lim_{m\to\infty} P(\bigcap_{n\geq m} \hat{\boldsymbol{\beta}} \in \mathcal{K}_1)$, *then $\hat{\boldsymbol{\alpha}} \xrightarrow{\text{a.s.}} \boldsymbol{\alpha}_0$ and $\hat{\boldsymbol{\beta}} \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0$.*

4 Asymptotic distribution

In this section, we derive under mild conditions the asymptotic distribution of estimators defined in Sect. 2 through Steps 1 to 3. In Rodriguez (2008), it is shown that the estimators $(\hat{\boldsymbol{\alpha}}_{\text{PR}}, \hat{\boldsymbol{\beta}}_{\text{PR}})$ defined through Steps P1 and P2 are asymptotically normally distributed. Indeed, these estimators can be written as

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}}_{\text{PR}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}}_{\text{PR}} - \boldsymbol{\alpha}_0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) + o_p(1), \tag{14}$$

where the function $\boldsymbol{\varphi} : \mathbb{R}^{p+q+1} \to \mathbb{R}^{p+q}$ is such that $\mathbb{E}(\boldsymbol{\varphi}(y_1, \mathbf{x}_1, \mathbf{t}_1)) = 0$. Thus, in order to derive the asymptotic distribution of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$, we will assume that the initial estimators $(\hat{\boldsymbol{\beta}}_{\text{R}}, \hat{\boldsymbol{\alpha}}_{\text{R}})$ admit a Bahadur expansion, i.e., $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{R}} - \boldsymbol{\beta}_0, \hat{\boldsymbol{\alpha}}_{\text{R}} - \boldsymbol{\alpha}_0)$ can be written as in (14) for some function $\boldsymbol{\varphi} : \mathbb{R}^{p+q+1} \to \mathbb{R}^{p+q}$.

Throughout this section we will assume that $\mathcal{T}$ is a compact set. We begin by fixing some notation.

Denote by $K_h(u) = (1/h)K(u/h)$, $\chi(y, a) = \partial\Psi(y, u)/\partial u$ and $\chi_1(y, a) = \partial^2\Psi(y, u)/\partial u^2$. Moreover, denote by $\boldsymbol{\tau} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$ and $\boldsymbol{\tau}_{0i} = (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i)$ and define

$$D_i(u) = A(u)^{-1}\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(u)\right)w_1(\mathbf{x}_i), \tag{15}$$

$$A(u) = \mathbb{E}_0\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1\right)\right)w_1(\mathbf{x}_1)|\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1 = u\right). \tag{16}$$

The following two functions will be used in assumption M6

$$\begin{aligned} m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) &= \mathbb{E}_0\left(D_1\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_2\right)\big|(\mathbf{t}_1, \mathbf{t}_2) = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)\right) \\ &= A^{-1}\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{\nu}_2\right)\mathbb{E}_0\left(\Psi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{\nu}_2\right)\right)w_1(\mathbf{x}_1)\big|\mathbf{t}_1 = \boldsymbol{\nu}_1\right), \\ m_{is}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) &= \mathbb{E}_0\left(\left[D_i\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_j\right) - D_i\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\right)\right]\right. \\ &\quad \left.\times\left[D_s\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_\ell\right) - D_s\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_s\right)\right]\big|(\mathbf{t}_i, \mathbf{t}_j, \mathbf{t}_s, \mathbf{t}_\ell) = (\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)\right), \end{aligned}$$

while the matrices $\boldsymbol{\Gamma}$ and $\mathbf{C}$ below are related to the asymptotic covariance of the estimators

$$\boldsymbol{\Gamma} = \mathbb{E}\big[\mathbf{w}\mathbf{w}^{\mathrm{T}}\big], \tag{17}$$

$$\mathbf{C} = \mathbb{E}_0\big[\chi\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})^{\mathrm{T}}\big], \tag{18}$$

where

$$\mathbf{w} = \Psi\big(y_1, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0\big(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) + \tilde{\boldsymbol{\varphi}}(y_1, \mathbf{x}_1, \mathbf{t}_1),$$

$$\begin{aligned}\tilde{\boldsymbol{\varphi}}(y_i, \mathbf{x}_i, \mathbf{t}_i) &= \mathbb{E}_0\Bigg(\chi\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) \\ &\quad\times \begin{pmatrix} \frac{\partial}{\partial\boldsymbol{\beta}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{01}} \\ \frac{\partial}{\partial\boldsymbol{\alpha}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{01}} \end{pmatrix}^{\mathrm{T}}\Bigg)\boldsymbol{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) \\ &\quad + D_i\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)f_{\boldsymbol{\alpha}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\tilde{\boldsymbol{\gamma}}(\mathbf{t}_i) \\ &= \mathbb{E}\Bigg(\tilde{\boldsymbol{\gamma}}(\mathbf{t}_i)\begin{pmatrix} \frac{\partial}{\partial\boldsymbol{\beta}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}} \\ \frac{\partial}{\partial\boldsymbol{\alpha}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)|_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}} \end{pmatrix}^{\mathrm{T}}\Bigg)\boldsymbol{\varphi}(y_i, \mathbf{x}_i, \mathbf{t}_i) \\ &\quad + D_i\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)f_{\boldsymbol{\alpha}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\tilde{\boldsymbol{\gamma}}(\mathbf{t}_i),\end{aligned} \tag{19}$$

$$\boldsymbol{\lambda}_i(\boldsymbol{\tau}) = \begin{pmatrix} \mathbf{x}_i \\ \frac{\partial}{\partial u}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)|_{\boldsymbol{\tau}}\ \mathbf{t}_i \end{pmatrix}, \tag{20}$$

$$\tilde{\boldsymbol{\gamma}}(\mathbf{t}) = \mathbb{E}_0\big(\chi\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01})\big|\mathbf{t}_1 = \mathbf{t}\big). \tag{21}$$

For the sake of simplicity, denote

$$\hat{\nu}(\boldsymbol{\alpha}, \boldsymbol{\beta}, u) = \hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u) - \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u), \qquad \hat{\nu}_0(u) = \hat{\nu}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, u), \tag{22}$$

$$\hat{\upsilon}_j(\boldsymbol{\alpha}, \boldsymbol{\beta}, u) = \frac{\partial\hat{\nu}(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)}{\partial\beta_j}, \qquad \hat{\upsilon}_{j,0}(u) = \hat{\upsilon}_j(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, u), \tag{23}$$

$$\hat{w}_\ell(\boldsymbol{\alpha}, \boldsymbol{\beta}, u) = \frac{\partial\hat{\nu}(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)}{\partial\alpha_\ell}, \qquad \hat{w}_{\ell,0}(u) = \hat{w}_\ell(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, u). \tag{24}$$

We will consider the following set of assumptions:

M1. (a) The functions $\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ and $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ are continuously differentiable with respect to $(\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$. Moreover, $\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ and $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ are three times continuously differentiable with respect to u.

(b) For any consistent estimate $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ of $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$, $\|\hat{\eta}_{\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}} - \eta_0\|_\infty \xrightarrow{p} 0$.

(c) For each fixed $\mathbf{t} \in \mathcal{T}$, $\boldsymbol{\alpha} \in \mathbb{R}^q$ and $\boldsymbol{\beta} \in \mathbb{R}^p$, $\hat{\nu}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}) \xrightarrow{p} 0$. Moreover,
 (i) $n^{\frac{1}{4}}\|\hat{\nu}_0\|_\infty \xrightarrow{p} 0$
 (ii) $n^{\frac{1}{4}}\|\hat{\upsilon}_{j,0}\|_\infty \xrightarrow{p} 0$ for $1 \le j \le p$
 (iii) $n^{\frac{1}{4}}\|\hat{w}_{\ell,0}\|_\infty \xrightarrow{p} 0$, for $1 \le \ell \le q$ and
 (iv) $\|\partial^k\hat{\nu}_0(u)/\partial u^k\|_\infty \xrightarrow{p} 0$, for $k = 1, 2, 3$.

M2. Ψ, χ, χ_1, w_2 and $\psi_2(\mathbf{x}) = \mathbf{x}w_2(\mathbf{x})$ are bounded and continuous functions. Besides, w_1 and w_2 are nonnegative functions.
M3. $\mathbb{E}_0\{\Psi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\mathbf{t}_1))|(\mathbf{x}_1, \mathbf{t}_1)\} = 0$.
M4. $\mathbb{E}_0(w_2(\mathbf{x}_1)\|\mathbf{x}_1\|^2) < \infty$.
M5. $f_{\boldsymbol{\alpha}_0}$, the marginal density of $\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_1$, is a bounded continuously differentiable function with bounded continuous derivative and such that $\inf_{\mathbf{t}\in\mathcal{T}} f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}) > 0$.
M6. (a) $m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)$, $m_1'(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \partial m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)/\partial \boldsymbol{\nu}_2$ and $m_1''(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \partial^2 m_1(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2)/\partial \boldsymbol{\nu}_2 \partial \boldsymbol{\nu}_2^{\mathrm{T}}$ are bounded and continuous functions.
(b) $m_{is}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)$, $m_{is,\ell}'(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = \partial m_{is}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)/\partial \boldsymbol{\nu}_\ell$ and $m_{is,\ell r}''(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4) = \partial^2 m_{is}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \boldsymbol{\nu}_3, \boldsymbol{\nu}_4)/\partial \boldsymbol{\nu}_\ell \partial \boldsymbol{\nu}_r^{\mathrm{T}}$ are bounded continuous functions.
M7. The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even, nonnegative bounded function with bounded variation. Moreover, it satisfies a Lipschitz condition of order one and $\int K(u)\,du = 1$, $\int uK(u)\,du = 0$ and $\int u^2K(u)\,du < \infty$.
M8. The matrix $\mathbf{C}_1 \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$ in non singular, where $\mathbf{C}_1$ is the left superior square submatrix of $\mathbf{C}$ defined in (18).
M9. $\boldsymbol{\Gamma}_1 \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$, the left superior square submatrix of $\boldsymbol{\Gamma}$, is non singular.

Remark 4.1

- The continuous differentiability of the kernel K and the implicit function theorem entail that $\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)$ is a continuously differentiable function of u. Moreover, the uniform consistency required in M1 can be derived through analogous arguments as those considered in Boente et al. (2006), under mild conditions such as

  (a) K is continuously differentiable with derivative K' bounded and with bounded variation.
  (b) For any compact sets $\mathcal{K} \in \mathbb{R}^p$ and $\mathcal{K}_1 \in \mathbb{R}$

$$\sup_{\mathbf{t}\in\mathcal{T}} \mathbb{E}_0\Big(\sup_{\boldsymbol{\beta}\in\mathcal{K}, a\in\mathbb{R}} \big|\chi\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta} + a\big)\big|\|\mathbf{x}_1\|\big|\mathbf{t}_1 = \mathbf{t}\Big) < \infty,$$

$$\sup_{\mathbf{t}\in\mathcal{T}} \mathbb{E}_0\Big(\sup_{\boldsymbol{\beta}\in\mathcal{K}, a\in\mathbb{R}} \big|\chi_1\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta} + a\big)\big|\|\mathbf{x}_1\|\big|\mathbf{t}_1 = \mathbf{t}\Big) < \infty,$$

$$\inf_{\substack{\boldsymbol{\beta}\in\mathcal{K}, a\in\mathcal{K}_1 \\ \mathbf{t}\in\mathcal{T}}} \mathbb{E}_0\big(\chi\big(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta} + a\big)\big|\mathbf{t}_1 = \mathbf{t}\big) > 0.$$

- Assumptions M2, M8 and M9 are standard conditions on the score function, in particular, M8 and M9 are a common requirement in robust regression in order to get root-n estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. As noted in Boente et al. (2006), for the score functions considered by Bianco and Yohai (1995), Croux and Haesbroeck (2002) and Cantoni and Ronchetti (2001), M3 is satisfied. This condition is the conditional Fisher-consistency property as stated in the generalized linear regression model by Künsch et al. (1989).

- It is worth noting that M4 can be replaced by $\mathbb{E}_0(\sup_{\boldsymbol{\beta}\in\mathcal{K},a\in\mathbb{R}}|\chi_1(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}+a)|\|\mathbf{x}_1\|^2 w_2(\mathbf{x}_1))<\infty$.
- Finally, M5, M6 and M7 and the convergence requirement in M1(b) are standard conditions in nonparametric estimation.

In order to show that $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are asymptotically normally distributed, we will need the following lemma whose proof follows using similar arguments to those considered in Lemma 4.1 of Boente and Rodriguez (2010).

Lemma 4.1 *Assume that* M1, M2, M5 *and* M7 *hold. If, in addition,* $w_1(\mathbf{x})\|\mathbf{x}\|^3$ *is bounded,* $A(u)\neq 0$ *for any* u *and* $\inf_{\mathbf{t}\in\mathcal{T}} A(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t})>0$ *where* $A(u)$ *is defined in* (16), $\lim_{n\to\infty} nh^4=0$ *and* $\lim_{n\to\infty} nh^2/\log^2(1/h)=+\infty$, *we have*

$$\sup_{t\in\mathcal{T}}\left|\hat{\eta}_{\boldsymbol{\alpha}_0\boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}\right)-\eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}\right)-\frac{1}{nf_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t})}\sum_{j=1}^{n}K_h\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_j-\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}\right)D_j\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}\right)\right|=o_p\left(n^{-1/2}\right),$$

where $D_j(u)$ *is defined in* (15).

Theorem 4.1 *Let us assume that the* $\mathbf{t}_i$ *have compact support* $\mathcal{T}$ *and that* M1 *to* M9 *hold. Let* $(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})$ *be a solution of* (11) *providing a consistent estimator of* $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$. *If* $\lim_{n\to\infty} nh^4=0$, *the conclusion of Lemma* 4.1 *holds and the initial estimators,* $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$ *and* $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$, *satisfy* (14), *we have*

(a) $\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\\ \hat{\boldsymbol{\alpha}}^{(q-1)}-\boldsymbol{\alpha}_0^{(q-1)}\end{pmatrix}\xrightarrow{D} N(0,\mathbf{C}_1^{-1}\boldsymbol{\Gamma}_1\mathbf{C}_1^{-1})$ *where* $\mathbf{C}_1$ *and* $\boldsymbol{\Gamma}_1$ *are given in* M8 *and* M9, *respectively, and* $\hat{\boldsymbol{\alpha}}^{(q-1)}=(\hat{\alpha}_1,\ldots,\hat{\alpha}_{q-1})$, $\boldsymbol{\alpha}_0^{(q-1)}=(\alpha_{01},\ldots,\alpha_{0q-1})$.

(b) $\sqrt{n}(\hat{\alpha}_q-\alpha_{0q})\xrightarrow{p} 0$.

5 Monte Carlo study

This section contains the results of a simulation study conducted with the aim of comparing the performance of the proposed estimators with the classical ones under a logistic and a Gamma partially linear single index model. Under a logistic model, the responses are bounded and so, the effect of the outliers on the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ is mainly observed when introducing high leverage points. Under this setting, the performance of our proposal is compared with the estimators defined in Carroll et al. (1997) which are based on the quasi-likelihood and also with those defined using as loss function the deviance. The Gamma model allows us to introduce also large values in the responses which will lead to a large effect on the classical estimators of the nonparametric function η. Under a Gamma model, due to the effect of outliers on the quasi-likelihood estimators already observed in the logistic model, we only compare our proposal with the estimators obtained minimizing the deviance that correspond to the classical counterpart of the robust estimators considered. In the reported Tables, the robust estimators introduced in this paper are indicated as ROB_c where c denotes the tuning constant of the loss function used while the classical estimators are indicated as DEV and QAL. The estimators DEV correspond to the estimators based on the deviance while those denoted by QAL are based on the quasi-likelihood, as defined in Carroll et al. (1997). On the other hand, the robust estimators ROB_c correspond to those controlling large values of the deviance together with large values on the $\mathbf{x}$ covariate space. The weight functions w_1 and w_2 used to control high leverage points were taken as the Tukey's biweight function with tuning constant $c_w = 4.685$ while for the classical estimators, $w_1 = w_2 \equiv 1$. In the smoothing procedure, the Epanechnikov kernel $K(t) = (3/4)(1 - t^2)\mathrm{I}_{[-1,1]}(t)$ was selected.

5.1 Logistic model

As mentioned above, this section summarizes the results of a simulation study under a logistic partly linear model. The robust estimators ROB_c control large values of the deviance using as ρ function the score function defined in Croux and Haesbroeck (2002) with tuning constant $c = 0.5$. To estimate the nonparametric component, both the estimators ROB_c and DEV used as bandwidths $h = 0.15$ and $h = 0.30$ in Steps 1 and 2, respectively, while for the QAL estimators the bandwidth $h = 0.30$ was considered. We have performed $NR = 1000$ replications with samples of size $n = 200$.

Under the central model, denoted C_0, we generate samples $(y_i, x_i, \mathbf{t}_i)$, $1 \le i \le n$, where $y_i|(x_i, \mathbf{t}_i) \sim Bi(1, p(x_i, \mathbf{t}_i))$ with $\log(p(x, \mathbf{t})/(1 - p(x, \mathbf{t}))) = x/2 + t_1 - 0.5 + \sin(4\pi t_1)$, i.e., $\beta_0 = 0.5$, $\boldsymbol{\alpha}_0 = (1, 0)$ $\eta_0(u) = (u - 1/2) + \sin(4\pi u)$. The covariates are such that $(x_i, \mathbf{t}_i) \sim N((0, 1/2, 1/2), \Sigma)$, $1 \le i \le n$, with

$$\Sigma = \begin{pmatrix} 1 & 1/(6\sqrt{3}) & 0 \\ 1/(6\sqrt{3}) & 1/36 & 0 \\ 0 & 0 & 1/36 \end{pmatrix}$$

and the variable $\mathbf{t}$ was truncated so that $\mathbf{t} \in [1/4, 3/4] \times [1/4, 3/4]$.

For each sample generated, we have considered the following contamination labeled C_1 in Table 1 and Figs. 1 and 2. We have first generated a sample $u_i \sim \mathcal{U}(0, 1)$, $1 \le i \le n$, and then, the contaminated sample, denoted $(y_{i,c}, x_{i,c}, t_i)$, is defined by $(y_{i,c}, x_{i,c}) = (y_i, x_i)$ if $u_i \le 0.90$ and $(y_{i,c}, x_{i,c}) = (y_{i,\text{new}}, x_{i,\text{new}})$ if $u_i > 0.90$, where $x_{i,\text{new}}$ is a new observation from a N(10, 1) and $y_{i,\text{new}}$ is a new observation from a $Bi(1, 0.05)$.

Table 1 gives summary measures for the different estimators. For the estimators of β_0, we have considered the following summary measures: bias, standard deviation (SD) and mean square error (MSE) computed over replications. Besides, to assess the performance of the estimators of $\boldsymbol{\alpha}_0$, we have considered estimated angles $\hat{\theta}$ for which we report bias, standard deviation (SD) and mean square error (MSE) computed over replications. To study the behavior of the estimators, $\hat{\eta}$, of the regression function η_0 we have considered the average over replications of the mean square error $\text{MSE}(\hat{\eta}) = (1/n)\sum_{i=1}^{n}[\hat{\eta}(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i) - \eta(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i)]^2$. Figures 1 and 2 give the boxplots of the estimators of the regression parameter and of $\hat{\theta}$, respectively. The horizontal lines indicate the true value for the parameters.

Table 1 Summary results for the estimators of β_0, α_0 and η_0 for the logistic model. The performance of $\hat{\alpha}$ was measured through its angle $\hat{\theta}$

	Estimator	Bias($\hat{\beta}$)	SD($\hat{\beta}$)	MSE($\hat{\beta}$)	Bias($\hat{\theta}$)	SD($\hat{\theta}$)	MSE($\hat{\theta}$)	MSE($\hat{\eta}$)
C_0	DEV	0.0532	0.1939	0.0404	0.0921	0.1289	0.0250	0.1564
	QAL	0.0039	0.2019	0.0408	0.0081	0.2120	0.0450	0.7308
	$\text{ROB}_{0.5}$	0.0393	0.1944	0.0393	0.0915	0.1259	0.0242	0.1635
C_1	DEV	−0.6939	0.0526	0.4842	0.0629	0.0896	0.0120	0.1562
	QAL	−0.7136	0.0604	0.5129	−0.5253	1.7454	3.3224	1.2620
	$\text{ROB}_{0.5}$	0.0248	0.208	0.0439	0.0924	0.1268	0.0247	0.1708

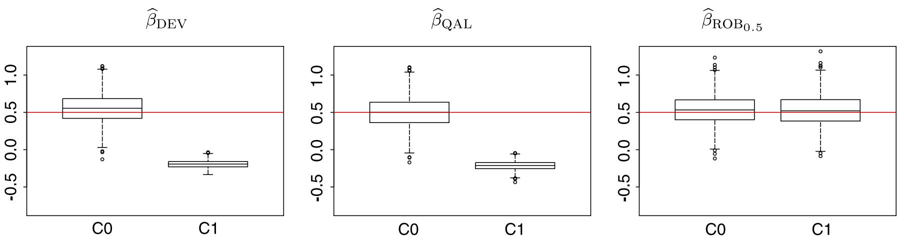

Fig. 1 Boxplots of the estimators of the regression parameter β for the logistic model

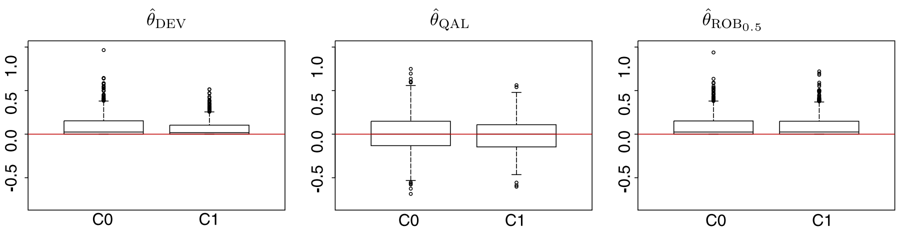

Fig. 2 Boxplots of the estimators of the single index angle θ for the logistic model

Table 1 shows that, under C_0, the mean square error of the robust estimator of the regression parameter $\boldsymbol{\beta}$ is similar to that of the classical estimators. The situation changes under C_1, where the mean square error of the classical procedures for estimating $\boldsymbol{\beta}$ is more than ten times larger than that of the robust ones. With respect to the estimation of $\boldsymbol{\alpha}$, the classical estimators based on the deviance and the robust ones show a similar behavior even under contamination. On the other hand, for the uncontaminated samples, the mean square error of the quasi-likelihood procedure is slightly larger than that of the procedures using the deviance while, under C_1 it increases significantly due to an increase of both the bias and variance. With respect to the estimation of η, again, the quasi-likelihood estimator shows a poor behavior compared to those based on the deviance since its mean square error is more than four times larger, under C_0 and more than nine times larger, under contamination. The fact that both the classical estimators of $\boldsymbol{\alpha}$ and η based on the deviance and the robust ones show similar behavior under C_1 may be due to the fact that under the logistic model the responses are bounded. In the next section, this situation changes since large outliers in the responses may occur.

5.2 Gamma model

In this section we summarize the results of a simulation study designed to compare the performance of the proposed estimators with the classical ones under a model with unbounded responses such as the Gamma model. As mentioned above, in the reported tables, the robust estimators introduced in this paper are indicated as ROB_c where c denotes the tuning constant used while their classical counterparts are indicated as DEV, since they correspond to the estimators based on the deviance. To be more precise, the robust estimators correspond to those controlling large values of the deviance as described in Bianco et al. (2005). They were computed using the Tukey's biweight score function with tuning constants $c = 1.5, 2, 2.3, 2.5, 2.8$ and $c = 3$. The weight functions w_1 and w_2 used to control high leverage points were taken as the Tukey's biweight function with tuning constant $c_w = 4.685$. On the other hand, the classical estimators correspond to the choice $w_1 = w_2 \equiv 1$ and loss function equal to the deviance. To compute the initial estimators, a bandwidth $h = 0.15$ was selected while for the final estimator $h = 0.3$ was chosen. Other bandwidth values were tested and they give quite similar results.

We have performed $NR = 1000$ replications with samples of size $n = 100$. The central model denoted C_0 in tables and figures corresponds to select $(x_i, \mathbf{t}_i)$ independent of each other such that $x_i \sim \text{N}(0, 1)$, $\mathbf{t}_i \sim \mathcal{U}((0, 1) \times (0, 1))$. The response variable was generated following a log-gamma single-index model, i.e., $y_i|(x_i, \mathbf{t}_i) \sim \Gamma(3, 3/\mu(x_i, \mathbf{t}_i))$, where $\mathbb{E}(y_i|(x_i, \mathbf{t}_i)) = \mu(x_i, \mathbf{t}_i)$ with $\log(\mu(x_i, \mathbf{t}_i)) = \beta_0 x_i + \eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i)$, with $\beta_0 = 2$, $\eta_0(u) = \sin(2\pi u)$ and $\boldsymbol{\alpha}_0 = (1, 1)/\sqrt(2)$ corresponding to an angle $\theta_0 = \pi/4$.

For each sample generated we have considered two contaminations schemes labeled C_1 and C_2 in tables and figures that lead to contaminated samples $(y_{i,c}, x_{i,c}, t_i)$. To obtain the contaminated samples, we have first generated a sample $u_i \sim \mathcal{U}(0, 1)$ for $1 \le i \le n$ and then, we have considered the following contamination schemes

- C_1 introduces *bad* high leverage points in the carriers x, without changing the responses already generated, i.e., $y_{i,c} = y_i$, $1 \le i \le n$, while

$$x_{i,c} = \begin{cases} x_i & \text{if } u_i \le 0.90, \\ \text{a new observation } x_i^\star \text{ from a N}(0, 25) & \text{if } u_i > 0.90. \end{cases}$$

- C_2 corresponds to increasing the variance of the carriers x and also to introduce large values on the responses

$$x_{i,c} = \begin{cases} x_i & \text{if } u_i \le 0.90, \\ \text{a new observation from a N}(0, 25) & \text{if } u_i > 0.90, \end{cases}$$

and

$$y_{i,c} = \begin{cases} y_i & \text{if } u_i \le 0.90, \\ y_i^\star & \text{if } u_i > 0.90, \end{cases}$$

where $y_i^\star$ is a new observation from a $\Gamma(3, 3/1000)$.

Table 2 Summary results for the estimators of β_0, α_0 and η_0. The performance of $\hat{\alpha}$ was measured through its angle $\hat{\theta}$

	Estimator	Bias($\hat{\beta}$)	SD($\hat{\beta}$)	MSE($\hat{\beta}$)	Bias($\hat{\theta}$)	SD($\hat{\theta}$)	MSE($\hat{\theta}$)	MSE($\hat{\eta}$)
C_0	DEV	0.0032	0.0621	0.0039	0.0119	0.1503	0.0227	0.0319
	$\text{ROB}_{1.5}$	−0.0967	0.9064	0.8309	−0.0905	0.5392	0.2989	0.6915
	$\text{ROB}_{2.0}$	0.0204	0.8924	0.7968	0.0134	0.5318	0.2830	0.7239
	$\text{ROB}_{2.3}$	0.0575	0.7233	0.5265	0.0437	0.5042	0.2561	0.6782
	$\text{ROB}_{2.5}$	0.0755	0.7972	0.6412	0.0239	0.4662	0.2179	0.7079
	$\text{ROB}_{2.8}$	0.0339	0.5790	0.3364	0.0342	0.4389	0.1938	0.6800
	$\text{ROB}_{3.0}$	0.0531	0.6133	0.3790	0.0388	0.4233	0.1807	0.6615
C_1	DEV	−1.4544	0.3740	2.2551	0.2249	0.4419	0.2459	1.9798
	$\text{ROB}_{2.8}$	0.0029	0.5829	0.3397	0.0303	0.4392	0.1938	0.6977
C_2	DEV	−1.9037	0.3355	3.7366	0.6108	0.5457	0.6709	18.7991
	$\text{ROB}_{2.8}$	0.0520	0.6772	0.4613	0.0279	0.4475	0.2010	0.7046

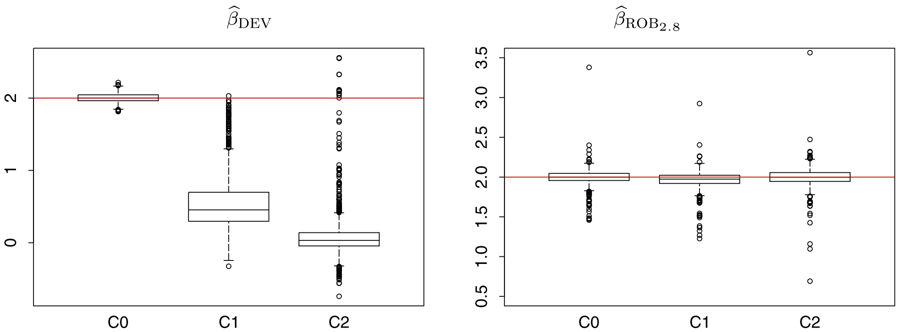

Fig. 3 Boxplots of the estimators of the regression parameter β

Table 2 summarizes the results obtained using the same summary measures as in Sect. 5.1. Figures 3 and 4 give the boxplots of the estimators of the regression parameter and of the angle of the index parameter, respectively. The horizontal lines indicate the true value for the parameters.

As expected, under C_0 the classical estimators show their advantage. The robust estimators show a high loss of efficiency in this case both for the regression parameter and for the angle of the index parameter. Among the several choices for the tuning constant, the best performance for the estimators of β was attained by $c = 2.8$ both in bias and standard deviation. For the estimators of θ the best performance in mean square error was obtained with $c = 3$. Taking into account that, when estimating $\boldsymbol{\alpha}$, the performance of the estimators $\widehat{\theta}$ when using $c = 2.8$ is quite similar to that obtained with $c = 3$, we only report the results for the contaminated samples when $c = 2.8$. Under the selected contamination, the classical estimators of β and θ show a poor behavior, in particular, with respect to bias. It is worth noticing, see Fig. 4, that for some samples the robust procedure gives very low angles while the

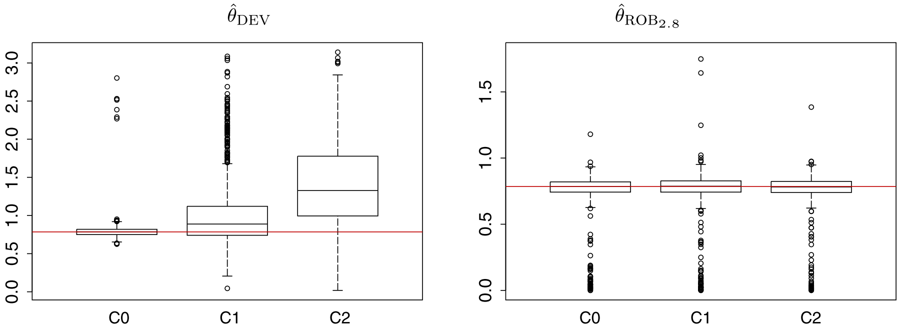

Fig. 4 Boxplots of the estimators of the single index angle θ

classical procedure tends to choose large values for $\hat{\theta}$. On the other hand, for the estimators of β the mean square error of the classical procedure is more than seven times larger than that of the robust ones for both contaminations. Besides, the standard deviation of the classical estimator under both contaminations is such that a test for the regression parameter would reject the null hypothesis $\beta_0 = 2$. For the two contaminations considered, the mean square errors of the classical estimators of η are more than two and twenty times those of the robust procedure which are quite close to the corresponding ones under C_0. On the other hand, contaminating only on the carriers multiplies by fifty the mean square error of the classical estimators $\hat{\eta}$ while under C_2 the MSE is five hundred times that under C_0. Therefore, as expected large responses affect the estimators of the nonparametric component even more than leverage points.

Acknowledgements This research was partially supported by Grants X018 from Universidad of Buenos Aires, PID 112-200801-00216 from CONICET and PICT m 821 from ANPCYT, Argentina. The authors thank Matt Wand for providing the code to compute the estimators defined in Carroll et al. (1997). We also wish to thank two anonymous referees and the Associate Editor for valuable comments which led to an improved version of the original paper.

Appendix

A.1 Proof of the consistency results

Proof of Lemma 3.1 (a) Let $R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = \sum_{i=1}^{n} K_h(\boldsymbol{\alpha}^{\rm T}\mathbf{t}_i - u)\rho(y_i, \mathbf{x}_i^{\rm T}\boldsymbol{\beta} + a)w_1(\mathbf{x}_i)/n$, $R_{0n}(\boldsymbol{\alpha}, u) = \sum_{i=1}^{n} K_h(u - \boldsymbol{\alpha}^{\rm T}\mathbf{t}_i)/n$ with $K_h(u) = K(u/h)/h$ and $R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u) = R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, u)/R_{0n}(\boldsymbol{\alpha}, u)$. Then,

$$\sup_{\substack{\boldsymbol{\alpha}\in\mathcal{S}_1,\boldsymbol{\beta}\in\mathcal{K}\\ a\in\mathbb{R}}} \left\| R_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - R(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot)\right\|_{0,\infty}$$

$$\leq \left[\sup_{\substack{\boldsymbol{\alpha}\in\mathcal{S}_1,\boldsymbol{\beta}\in\mathcal{K}\\ a\in\mathbb{R}}} \left\| R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot) - \mathbb{E}\big(R_{1n}(\boldsymbol{\alpha}, \boldsymbol{\beta}, a, \cdot)\big)\right\|_{0,\infty} \right.$$

$$+ \sup_{\substack{\boldsymbol{\alpha}\in\mathcal{S}_1,,\boldsymbol{\beta}\in\mathcal{K}\\ a\in\mathbb{R}}} \left\|\mathbb{E}\big(R_{1n}(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot)\big) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot)\mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},\cdot)\big)\right\|_{0,\infty}$$

$$+ \|\rho\|_\infty \sup_{\boldsymbol{\alpha}\in\mathcal{S}_1} \left\| R_{0n}(\boldsymbol{\alpha},\cdot) - \mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},\cdot)\big)\right\|_{0,\infty} \Bigg] \Big/ \inf_{\boldsymbol{\alpha}\in\mathcal{S}_1, u\in\mathcal{U}(\mathcal{T}_0)} R_{0n}(\boldsymbol{\alpha},u),$$

where $\|\rho\|_\infty = \sup_{y,a}|\rho(y,a)|$. For n large enough, we have

$$\inf_{\boldsymbol{\alpha}\in\mathcal{S}_1, u\in\mathcal{U}(\mathcal{T}_0)} R_{0n}(\boldsymbol{\alpha},u) \geq \inf_{\boldsymbol{\alpha}\in\mathcal{S}_1, u\in\mathcal{U}(\mathcal{T}_0)} \mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},u)\big) - \sup_{\boldsymbol{\alpha}\in\mathcal{S}_1} \left\| R_{0n}(\boldsymbol{\alpha},\cdot) - \mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},\cdot)\big)\right\|_{0,\infty}.$$

Let $\delta < \delta_0$ and define $\mathcal{U}_\delta = \{u+s : u \in \mathcal{U}(\mathcal{T}_0), |s| \leq \delta\}$. Let R be such that $\int_{|v|\leq R} K(v)\,dv > 1/2$. Then, for $h \leq \delta/R$ we have $hv + u \in \mathcal{U}_\delta$ and so, using the fact that $\mathcal{U}_\delta$ is a compact set and D4, we get $\mathbb{E}(R_{0n}(\boldsymbol{\alpha},u)) = \int K(v) f_{\boldsymbol{\alpha}}(hv+u)\,dv > \frac{1}{2}A_1(\mathcal{U}_\delta)$. Therefore, it is enough to show that

$$\sup_{\substack{\boldsymbol{\alpha}\in\mathcal{S}_1,\boldsymbol{\beta}\in\mathcal{K}\\ a\in\mathbb{R}}} \left\| R_{1n}(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot) - \mathbb{E}\big(R_{1n}(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot)\big)\right\|_{0,\infty} \xrightarrow{\text{a.s.}} 0, \tag{25}$$

$$\sup_{\boldsymbol{\alpha}\in\mathcal{S}_1} \left\| R_{0n}(\boldsymbol{\alpha},\cdot) - \mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},\cdot)\big)\right\|_{0,\infty} \xrightarrow{\text{a.s.}} 0, \tag{26}$$

$$\sup_{\substack{\boldsymbol{\alpha}\in\mathcal{S}_1,\boldsymbol{\beta}\in\mathcal{K}\\ a\in\mathbb{R}}} \left\|\mathbb{E}\big(R_{1n}(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot)\big) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,\cdot)\mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},\cdot)\big)\right\|_{0,\infty} \to 0. \tag{27}$$

Using Theorem 37 in Pollard (1984) and D1, we see that (26) hold. On the other hand, if $u_1 = \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_1$ we obtain that

$$\begin{aligned}
&\left|\mathbb{E}\big(R_{1n}(\boldsymbol{\alpha},\boldsymbol{\beta},a,u)\big) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u)\mathbb{E}\big(R_{0n}(\boldsymbol{\alpha},u)\big)\right| \\
&\quad = \left|\mathbb{E}\big(K_h(u_1-u)\big[R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u_1) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u)\big]\big)\right| \\
&\quad = \left|\int K_h(\tau-u)\big[R(\boldsymbol{\alpha},\boldsymbol{\beta},a,\tau) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u)\big] f_{\boldsymbol{\alpha}}(\tau)\,d\tau\right| \\
&\quad \leq \|f_{\boldsymbol{\alpha}}\|_\infty \int K(v)\big|R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u-vh) - R(\boldsymbol{\alpha},\boldsymbol{\beta},a,u)\big|\,dv.
\end{aligned}$$

Note that the boundedness of $f_{\mathbf{T}}$ lead to $\sup_{\boldsymbol{\alpha}\in\mathcal{S}_1}\|f_{\boldsymbol{\alpha}}\|_\infty < \infty$. Then, (27) follows easily from the boundedness of ρ, the integrability of the kernel, the equicontinuity condition D5 and the fact that $h_n \to 0$. Finally, in order to prove (25), consider the class of functions

$$\begin{aligned}
\mathcal{F}_n = \Bigg\{ & g_{\mathbf{t},a,\boldsymbol{\alpha},\boldsymbol{\beta},h}(y,\mathbf{x},v) = B^{-1}\rho\big(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}+a\big)w_1(\mathbf{x})K\left(\frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}-v}{h}\right) \\
& = B^{-1}\rho\big(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}+a\big)w_1(\mathbf{x})\tilde{K}_{\boldsymbol{\alpha},h,\mathbf{t}}(v)\Bigg\},
\end{aligned}$$

with $B = \|\rho\|_\infty \|w_1\|_\infty \|K\|_\infty$. The proof of (25) follows now as that of Theorem 3.1 in Boente et al. (2006).

(b) Follows using analogous arguments to those considered in Theorem 3.1 in Boente et al. (2006). □

Proof of Lemma 3.2 (a) For any $\varepsilon > 0$, let $\mathcal{T}_0$ be a compact set such that $P(\mathbf{t} \notin \mathcal{T}_0) < \varepsilon$. Then, we find that

$$\sup_{\mathbf{a},\boldsymbol{\alpha}\in\mathcal{S}_1,\mathbf{b},\boldsymbol{\beta}\in\mathcal{K}} \left|G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{\eta}_{\mathbf{a},\mathbf{b}}) - G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}})\right|$$
$$\leq \sup_{\mathbf{a}\in\mathcal{S}_1,\mathbf{b}\in\mathcal{K}} \|\hat{\eta}_{\mathbf{a},\mathbf{b}} - \eta_{\mathbf{a},\mathbf{b}}\|_{0,\infty} \|w_2\|_\infty \|\Psi\|_\infty$$
$$+ 2\|w_2\|_\infty \|\rho\|_\infty \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}_{(\mathbf{t}_i\notin\mathcal{T}_0)}$$

and so, using (13), the fact that $P(\mathbf{t} \notin \mathcal{T}_0) < \varepsilon$ and the Strong Law of Large Numbers, we get

$$\sup_{\mathbf{a},\boldsymbol{\alpha}\in\mathcal{S}_1,\mathbf{b},\boldsymbol{\beta}\in\mathcal{K}} \left|G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{\eta}_{\mathbf{a},\mathbf{b}}) - G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}})\right| \xrightarrow{a.s.} 0.$$

Therefore, it remains to show that $\sup_{\mathbf{a},\boldsymbol{\alpha}\in\mathcal{S}_1,\mathbf{b},\boldsymbol{\beta}\in\mathcal{K}} |G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}}) - G(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}})| \xrightarrow{a.s.} 0$. Define the following class of functions $\mathcal{H} = \{f_{\boldsymbol{\alpha},\boldsymbol{\beta}}(y,\mathbf{x},\mathbf{t}) = \rho(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \eta_{\mathbf{a},\mathbf{b}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}))w_2(\mathbf{x}), \mathbf{b},\boldsymbol{\beta} \in \mathcal{K}, \mathbf{a},\boldsymbol{\alpha} \in \mathcal{S}_1\}$. Using Theorem 3 from Chap. 2 in Pollard (1984), the compactness of $\mathcal{K}$, D1, D6 and analogous arguments to those considered in Lemma 1 from Bianco and Boente (2002), we get $\sup_{\mathbf{a},\boldsymbol{\alpha}\in\mathcal{S}_1;\mathbf{b},\boldsymbol{\beta}\in\mathcal{K}} |G_n(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}}) - G(\boldsymbol{\alpha},\boldsymbol{\beta},\eta_{\mathbf{a},\mathbf{b}})| \xrightarrow{a.s.} 0$.

(b) Let $(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k)$ be a subsequence of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ such that $(\hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_k) \to (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Note that $(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ belongs to the compact set $\mathcal{S}_1 \times \mathcal{K}_1$. Let us assume without loss of generality that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \xrightarrow{a.s.} (\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Then, **D7**, the continuity of $\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}$ and Lemma 3.1(a) entail that $G_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}) - G(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*, \eta_0) \xrightarrow{a.s.} 0$ and $G_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}) - G(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \eta_0) \xrightarrow{a.s.} 0$. Finally, since $G_n(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}) \geq G_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}})$ and $G(\boldsymbol{\alpha}, \boldsymbol{\beta}, \eta_0)$ has a unique minimum at $(\boldsymbol{\beta}_0, \eta_0)$, we conclude the proof. □

A.2 Proof of the asymptotic normality results

Lemma A.1 *Let* $(y_i, \mathbf{x}_i, \mathbf{t}_i)$ *be independent observations such that* $y_i|(\mathbf{x}_i, \mathbf{t}_i) \sim F(\cdot, \mu_i)$ *with* $\mu_i = H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i))$ *and* $\mathrm{VAR}(y_i|(\mathbf{x}_i, \mathbf{t}_i)) = V(\mu_i)$. *Assume that* $\mathbf{t}_i$ *are random variables with distribution on a compact set* $\mathcal{T}$ *and that* M1 *to* M4 *hold. Let* $\hat{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$, $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ *consistent estimators of* $\boldsymbol{\alpha}_0$ *and* $\boldsymbol{\beta}_0$, *respectively. Then,*

$\mathbf{C}_n \xrightarrow{p} \mathbf{C}$ *where* $\mathbf{C}$ *is defined in* (18) *and*

$$\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},i})\hat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\tau}}_{\mathrm{R},i})^{\mathrm{T}}$$
$$+ \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)\hat{\boldsymbol{\lambda}}_i^{(1)}(\tilde{\boldsymbol{\tau}}_{\mathrm{R},i})$$

with $\tilde{\boldsymbol{\tau}}_{\mathrm{R},i} = (\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}, \tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i)$, $\hat{\boldsymbol{\tau}}_{\mathrm{R},i} = (\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}, \hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i)$,

$$\hat{\boldsymbol{\lambda}}_i(\boldsymbol{\tau}) = \begin{pmatrix} \mathbf{x}_i \\ \frac{\partial}{\partial s}\hat{\eta}_{\mathbf{a},\mathbf{b}}(s)|_{(\mathbf{a},\mathbf{b},s)=\boldsymbol{\tau}}\ \mathbf{t}_i \end{pmatrix} \quad \textit{and} \quad \hat{\boldsymbol{\lambda}}_i^{(1)}(\boldsymbol{\tau}) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial^2}{\partial s^2}\hat{\eta}_{\mathbf{a},\mathbf{b}}(s)\mathbf{t}_i\mathbf{t}_i^{\mathrm{T}} \end{pmatrix}\Bigg|_{(\mathbf{a},\mathbf{b},s)=\boldsymbol{\tau}},$$

where $\boldsymbol{\tau} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, u)$.

Proof Let $\boldsymbol{\lambda}_i^{(1)}(\boldsymbol{\tau}) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{\partial^2}{\partial u^2}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)\mathbf{t}_i\mathbf{t}_i^{\mathrm{T}} \end{pmatrix}|_{\boldsymbol{\tau}=\boldsymbol{\tau}}$ and note that $\mathbf{C}_n$ can be written as $\mathbf{C}_n = \sum_{i=1}^{6}\mathbf{C}_n^{(i)}$ where

$$\mathbf{C}_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})^{\mathrm{T}},$$

$$\mathbf{C}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)$$
$$\times\left[\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},i})\hat{\boldsymbol{\lambda}}_i(\tilde{\boldsymbol{\tau}}_{\mathrm{R},i})^{\mathrm{T}} - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})^{\mathrm{T}}\right],$$

$$\mathbf{C}_n^{(3)} = \frac{1}{n}\sum_{i=1}^{n} \chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \xi_{1in}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\left[\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right) - \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\right)\right],$$

$$\mathbf{C}_n^{(4)} = \frac{1}{n}\sum_{i=1}^{n} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i^{(1)}(\boldsymbol{\tau}_{0i}),$$

$$\mathbf{C}_n^{(5)} = \frac{1}{n}\sum_{i=1}^{n} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right)\right) w_2(\mathbf{x}_i)\left[\hat{\boldsymbol{\lambda}}_i^{(1)}(\tilde{\boldsymbol{\tau}}_{\mathrm{R},i}) - \boldsymbol{\lambda}_i^{(1)}(\boldsymbol{\tau}_{0i})\right],$$

$$\mathbf{C}_n^{(6)} = \frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\tilde{\boldsymbol{\beta}} + \xi_{2\mathrm{in}}\right) w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i^{(1)}(\boldsymbol{\tau}_{0i})\left[\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i\right) - \eta_0\left(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\right)\right],$$

where $\xi_{1\mathrm{in}}$ and $\xi_{2\mathrm{in}}$ are intermediate points between $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}(\tilde{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i)$ and $\eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i)$ and between $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}(\hat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{t}_i)$ and $\eta_0(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i)$, respectively. Using M1, M2, M4 and the fact that $\hat{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\tilde{\boldsymbol{\beta}}$ are consistent estimators of the parameters, we have easily that $\mathbf{C}_n^{(j)} \xrightarrow{p} 0$ for $j = 2, 3, 5, 6$.

The Dominated Convergence Theorem and the consistency of $\tilde{\boldsymbol\beta}$ lead to

$$\mathbb{E}\big(\chi\big(y_1, \mathbf{x}_1^{\mathrm T}\tilde{\boldsymbol\beta} + \eta_0\big(\boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_1)\boldsymbol\lambda_1(\boldsymbol\tau_{01})\big)$$
$$\to \mathbb{E}\big(\chi\big(y_1, \mathbf{x}_1^{\mathrm T}\boldsymbol\beta_0 + \eta_0\big(\boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\boldsymbol\lambda_1(\boldsymbol\tau_{01})\big).$$

On the other hand, Theorem 3 in Chap. 2 of Pollard (1984) implies that, for any compact set $\mathcal{K} \subset \mathbb{R}^p$,

$$\sup_{\boldsymbol\beta\in\mathcal{K}} \Bigg| n^{-1}\sum_{i=1}^n \big[\chi\big(y_i, \mathbf{x}_i^{\mathrm T}\boldsymbol\beta + \eta_0\big(\boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\boldsymbol\lambda_i(\boldsymbol\tau_{0i})$$
$$- \mathbb{E}\big(\chi\big(y_1, \mathbf{x}_1^{\mathrm T}\boldsymbol\beta + \eta_0\big(\boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_1\big)\big)w_2(\mathbf{x}_1)\big)\boldsymbol\lambda_1(\boldsymbol\tau_{01})\big]\Bigg| \xrightarrow{\text{a.s.}} 0.$$

Therefore, we have $\mathbf{C}_n^{(2)} \xrightarrow{p} \mathbf{C}$ with $\mathbf{C}$ defined in (18). Using analogous arguments and M3, we obtain that $\mathbf{C}_n^{(4)} \xrightarrow{p} 0$. □

Proof of Theorem 4.1 Note that by Remark 2.1, we can assume that $\boldsymbol\alpha_0 = \mathbf{e}_q$. Let $\hat{\boldsymbol\tau}_{\mathrm R,0i} = (\hat{\boldsymbol\alpha}_{\mathrm R}, \hat{\boldsymbol\beta}_{\mathrm R}, \boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_i)$ while $\tilde{\boldsymbol\tau}_{\mathrm R,i}$, $\hat{\boldsymbol\tau}_{\mathrm R,i}$, $\hat{\boldsymbol\lambda}_i(\boldsymbol\tau)$, $\hat{\boldsymbol\lambda}_i^{(1)}(\boldsymbol\tau)$, $\boldsymbol\lambda_i^{(1)}(\boldsymbol\tau)$ and $\boldsymbol\tau = (\boldsymbol\alpha, \boldsymbol\beta, u)$ are defined as in Lemma A.1.

Let $(\hat{\boldsymbol\alpha}, \hat{\boldsymbol\beta})$ be a solution of $G_n^{(1)}(\hat{\boldsymbol\alpha}, \hat{\boldsymbol\beta}, \hat\eta_{\hat{\boldsymbol\alpha}_{\mathrm R},\hat{\boldsymbol\beta}_{\mathrm R}}) = 0$ with $G_n^{(1)}$ defined in (11). Using a Taylor expansion of order one, we get

$$0 = \theta_n\begin{pmatrix}0\\ \hat{\boldsymbol\alpha}\end{pmatrix} + \frac{1}{n}\sum_{i=1}^n \Psi\big(y_i, \mathbf{x}_i^{\mathrm T}\hat{\boldsymbol\beta} + \hat\eta_{\hat{\boldsymbol\alpha}_{\mathrm R},\hat{\boldsymbol\beta}_{\mathrm R}}\big(\hat{\boldsymbol\alpha}^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\hat{\boldsymbol\lambda}_i(\hat\tau_{\mathrm R,i})$$
$$= \theta_n\begin{pmatrix}0\\ \hat{\boldsymbol\alpha}\end{pmatrix} + \hat{\mathbf{V}}_n + \mathbf{C}_n\sqrt{n}\begin{pmatrix}\hat{\boldsymbol\beta} - \boldsymbol\beta_0\\ \hat{\boldsymbol\alpha} - \boldsymbol\alpha_0\end{pmatrix}, \tag{28}$$

where

$$\hat{\mathbf{V}}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi\big(y_i, \mathbf{x}_i^{\mathrm T}\boldsymbol\beta_0 + \hat\eta_{\hat{\boldsymbol\alpha}_{\mathrm R},\hat{\boldsymbol\beta}_{\mathrm R}}\big(\boldsymbol\alpha_0^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\hat{\boldsymbol\lambda}_i(\hat{\boldsymbol\tau}_{\mathrm R,0i}),$$
$$\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^n \chi\big(y_i, \mathbf{x}_i^{\mathrm T}\tilde{\boldsymbol\beta} + \hat\eta_{\hat{\boldsymbol\alpha}_{\mathrm R},\hat{\boldsymbol\beta}_{\mathrm R}}\big(\tilde{\boldsymbol\alpha}^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\hat{\boldsymbol\lambda}_i(\hat{\boldsymbol\tau}_{\mathrm R,i})\hat{\boldsymbol\lambda}_i(\tilde{\boldsymbol\tau}_{\mathrm R,i})^{\mathrm T}$$
$$+ \frac{1}{n}\sum_{i=1}^n \Psi\big(y_i, \mathbf{x}_i^{\mathrm T}\tilde{\boldsymbol\beta} + \hat\eta_{\hat{\boldsymbol\alpha}_{\mathrm R},\hat{\boldsymbol\beta}_{\mathrm R}}\big(\hat{\boldsymbol\alpha}^{\mathrm T}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\hat{\boldsymbol\lambda}_i^{(1)}(\tilde{\boldsymbol\tau}_{\mathrm R,i})$$

with $\tilde{\boldsymbol\alpha}$ and $\tilde{\boldsymbol\beta}$ intermediate points between $\boldsymbol\alpha_0$ and $\hat{\boldsymbol\alpha}$ and between $\boldsymbol\beta_0$ and $\hat{\boldsymbol\beta}$, respectively.

Lemma A.1 entails that $\mathbf{C}_n \xrightarrow{p} \mathbf{C}$ where $\mathbf{C}$ is defined in (18). Then, it will enough to study the behavior of $\hat{\mathbf{V}}_n$. Let

$$\mathbf{V}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i}).$$

Straightforward calculations lead to $\hat{\mathbf{V}}_n - \mathbf{V}_n = \sum_{i=1}^{5} \mathbf{V}_{in}$ where

$$\mathbf{V}_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \left[\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right],$$

$$\mathbf{V}_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \left[\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) - \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right]$$
$$\times \left[\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right],$$

$$\mathbf{V}_{3n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})$$
$$\times \left[\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) - \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right],$$

$$\mathbf{V}_{4n} = \frac{1}{2n} \sum_{i=1}^{n} \chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \xi_{\mathrm{in}}\right) w_2(\mathbf{x}_i) \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})$$
$$\times \left[n^{1/4}\left(\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) - \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right)\right]^2,$$

$$\mathbf{V}_{5n} = \frac{1}{2n} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \zeta_{\mathrm{in}}\right) w_2(\mathbf{x}_i) \left[n^{1/4}\left(\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) - \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right)\right]^2$$
$$\times \left[\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right]$$

with ξ_{in} and ζ_{in} intermediate points between $\eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i)$ and $\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}}, \hat{\boldsymbol{\beta}}_{\mathrm{R}}}(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i)$.

Analogous arguments to those considered in Theorem 3.5.3 in Rodriguez (2008), allow one to show that $\mathbf{V}_{1n} \xrightarrow{p} 0$. Furthermore, M1, M2 and the fact that $n^{1/4}\|\hat{\boldsymbol{\alpha}}_{\mathrm{R}} - \boldsymbol{\alpha}_0\| \xrightarrow{\text{a.s.}} 0$ and $n^{1/4}\|\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0\| \xrightarrow{\text{a.s.}} 0$ entail that $\mathbf{V}_{in} \xrightarrow{p} 0$ for $i = 4, 5$.

Therefore, if we show that

(i) $\mathbf{V}_{2n} \xrightarrow{p} 0$
(ii) $\mathbf{V}_{3n} = \sum_{i=1}^{n} \tilde{\boldsymbol{\varphi}}(y_i, \mathbf{x}_i, \mathbf{t}_i)/\sqrt{n} + o_p(1)$

where $\tilde{\boldsymbol{\varphi}}$ is defined in (19), we can conclude that

$$\hat{\mathbf{V}}_n - \mathbf{V}_n - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\boldsymbol{\varphi}}(y_i, \mathbf{x}_i, \mathbf{t}_i) \xrightarrow{p} 0.$$

Hence, M3 and the fact that $\mathbb{E}_0 \boldsymbol{\varphi}(y_1, \mathbf{x}_1, \mathbf{t}_1) = 0$ entail that $\hat{\mathbf{V}}_n$ is asymptotically normally distributed with covariance $\boldsymbol{\Gamma}$.

Let us begin by showing (ii). First, we note that by Lemma 4.1 we have

$$\begin{aligned}\hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}(u)-\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}(u) &= \hat{\eta}_{\hat{\boldsymbol{\alpha}}_{\mathrm{R}},\hat{\boldsymbol{\beta}}_{\mathrm{R}}}(u)-\hat{\eta}_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}(u)+\hat{\eta}_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}(u)-\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}(u)\\ &= \frac{\partial}{\partial\boldsymbol{\alpha}}\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)\bigg|^{\mathrm{T}}_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,u)}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}-\boldsymbol{\alpha}_0)\\ &\quad+\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)\bigg|^{\mathrm{T}}_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,u)}(\hat{\boldsymbol{\beta}}_{\mathrm{R}}-\boldsymbol{\beta}_0)\\ &\quad+\frac{1}{nf_{\boldsymbol{\alpha}_0}(u)}\sum_{j=1}^{n}K_h\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_j-u\big)D_j(u)+o_p\big(n^{-1/2}\big), \qquad (29)\end{aligned}$$

and consider the following expansion:

$$\mathbf{V}_{3n}=\mathbf{V}_{31n}+\mathbf{V}_{32n}+\mathbf{V}_{33n}+\mathbf{V}_{34n}+\mathbf{V}_{35n}+o_p\big(n^{-1/2}\big),$$

where

$$\begin{aligned}\mathbf{V}_{31n} &= \frac{1}{n}\sum_{i=1}^{n}\chi\big(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{\partial}{\partial\boldsymbol{\alpha}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)\bigg|^{\mathrm{T}}_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}}\\ &\quad\times\sqrt{n}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}-\boldsymbol{\alpha}_0),\\ \mathbf{V}_{32n} &= \frac{1}{n}\sum_{i=1}^{n}\chi\big(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{\partial}{\partial\boldsymbol{\beta}}\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)\bigg|^{\mathrm{T}}_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}}\\ &\quad\times\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{R}}-\boldsymbol{\beta}_0),\\ \mathbf{V}_{33n} &= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{f_{\boldsymbol{\alpha}_0}(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_j)}K_h\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_j-\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\chi\big(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\big)\\ &\quad\times w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})D_j\big(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{t}_i\big),\\ \mathbf{V}_{34n} &= \frac{1}{n}\sum_{i=1}^{n}\chi\big(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{\partial}{\partial\boldsymbol{\alpha}}\big[\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)-\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)\big]\bigg|^{\mathrm{T}}_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}}\\ &\quad\times\sqrt{n}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}-\boldsymbol{\alpha}_0),\\ \mathbf{V}_{35n} &= \frac{1}{n}\sum_{i=1}^{n}\chi\big(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\big(\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{t}_i\big)\big)w_2(\mathbf{x}_i)\boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\frac{\partial}{\partial\boldsymbol{\beta}}\big[\hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)-\eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u)\big]\bigg|^{\mathrm{T}}_{\boldsymbol{\tau}=\boldsymbol{\tau}_{0i}}\\ &\quad\times\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{R}}-\boldsymbol{\beta}_0).\end{aligned}$$

Using M1, M2 and the fact that $\sqrt{n}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}}-\boldsymbol{\alpha}_0)=O_p(1)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{R}}-\boldsymbol{\beta}_0)=O_p(1)$, it is easy to see that $\mathbf{V}_{3in}\xrightarrow{p}0$ for $i=4,5$. Using analogous arguments to those considered in Theorem 3.5.3 in Rodriguez (2008) we obtain that

$$\mathbf{V}_{33n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D_i\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) f_{\boldsymbol{\alpha}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) \tilde{\boldsymbol{\gamma}}(\mathbf{t}_i) + o_p(1)$$

with $\tilde{\boldsymbol{\gamma}}(\mathbf{t})$ defined in (21).

On the other hand, using the fact that the initial estimators $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\hat{\boldsymbol{\alpha}}_{\mathrm{R}}$ satisfy (14) and the strong law of large umbers, we get

$$\begin{aligned}
&\mathbf{V}_{31n} + \mathbf{V}_{32n} \\
&\quad = \mathbb{E}\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_1\right)\right) w_2(\mathbf{x}_1) \boldsymbol{\lambda}_1(\boldsymbol{\tau}_{01}) \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{01}} \\ \frac{\partial}{\partial \boldsymbol{\alpha}} \eta_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{01}} \end{pmatrix}^{\mathrm{T}} \right) \\
&\qquad \times \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\varphi}(y_i, \mathbf{x}_i \mathbf{t}_i) + o_p(1)
\end{aligned}$$

concluding the proof of (ii).

Let us show (i). Using the expansion (29), we can write $\mathbf{V}_{2n} = \mathbf{V}_{21n} + \mathbf{V}_{22n} + \mathbf{V}_{23n} + o_p(n^{-1/2})$ where

$$\begin{aligned}
\mathbf{V}_{21n} &= \frac{1}{n} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \frac{\partial}{\partial \boldsymbol{\alpha}} \hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s) \Big|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{0i}}^{\mathrm{T}} \\
&\quad \times \sqrt{n}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}} - \boldsymbol{\alpha}_0)\left(\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right), \\
\mathbf{V}_{22n} &= \frac{1}{n} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) w_2(\mathbf{x}_i) \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\eta}_{\boldsymbol{\alpha},\boldsymbol{\beta}}(s)(s) \Big|_{(\boldsymbol{\alpha},\boldsymbol{\beta},s)=\boldsymbol{\tau}_{0i}}^{\mathrm{T}} \\
&\quad \times \sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0)\left(\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right), \\
\mathbf{V}_{23n} &= \frac{1}{n^{3/2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{f_{\boldsymbol{\alpha}_0}(\alpha_0^{\mathrm{T}} \mathbf{t}_j)} K_h\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_j - \boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right) \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_{\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0}\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\right) \\
&\quad \times w_2(\mathbf{x}_i) D_j\left(\boldsymbol{\alpha}_0^{\mathrm{T}} \mathbf{t}_i\right)\left[\hat{\boldsymbol{\lambda}}_i(\hat{\boldsymbol{\tau}}_{\mathrm{R},0i}) - \boldsymbol{\lambda}_i(\boldsymbol{\tau}_{0i})\right].
\end{aligned}$$

Using again that $\sqrt{n}(\hat{\boldsymbol{\alpha}}_{\mathrm{R}} - \boldsymbol{\alpha}_0) = O_p(1)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{R}} - \boldsymbol{\beta}_0) = O_p(1)$, together with assumptions M1 and M2, we obtain that $\mathbf{V}_{2in} \xrightarrow{p} 0$ for $i = 1, 2$. The proof of $\mathbf{V}_{23n} \xrightarrow{p} 0$ can be found in Lemma 3.5.4 in Rodriguez (2008).

To conclude the proof, let $\mathbf{P}_{\hat{\boldsymbol{\alpha}}}$ be the projection matrix $\mathbf{P}_{\hat{\boldsymbol{\alpha}}} = \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_q - \hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\alpha}}^{\mathrm{T}} \end{pmatrix}$. Using (28), we get

$$\mathbf{0} = \mathbf{P}_{\hat{\boldsymbol{\alpha}}} \hat{\mathbf{V}}_n + \mathbf{P}_{\hat{\boldsymbol{\alpha}}} \mathbf{C}_n \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \end{pmatrix}. \tag{30}$$

The consistency of $\hat{\boldsymbol{\alpha}}$ implies that $\mathbf{P}_{\hat{\boldsymbol{\alpha}}} \xrightarrow{p} \mathbf{P}_{\boldsymbol{\alpha}_0}$ and since $\boldsymbol{\alpha}_0 = \mathbf{e}_q$, we get $\mathbf{P}_{\boldsymbol{\alpha}_0} = \begin{pmatrix} \mathbf{I}_{p+q-1} & 0 \\ 0 & 0 \end{pmatrix}$ Then, write $\mathbf{C}_n = \begin{pmatrix} \mathbf{C}_{n1} & \mathbf{C}_{n2} \\ \mathbf{C}_{n3} & C_{n4} \end{pmatrix}$ where $\mathbf{C}_{n1} \in \mathbb{R}^{(p+q-1)\times(p+q-1)}$, $\mathbf{C}_{n2}, \mathbf{C}_{n3}^{\mathrm{T}} \in \mathbb{R}^{(p+q-1)\times 1}$ and $C_{n4} \in \mathbb{R}$ and also, $\mathbf{C}_{n1} = \begin{pmatrix} \mathbf{C}_{n1}^{(1)} & \mathbf{C}_{n1}^{(2)} \\ \mathbf{C}_{n1}^{(3)} & \mathbf{C}_{n1}^{(4)} \end{pmatrix}$ with $\mathbf{C}_{n1}^{(1)} \in \mathbb{R}^{p\times p}$. Let $\mathbf{P}_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)} = (p_{\hat{\boldsymbol{\alpha}}_1} \cdots p_{\hat{\boldsymbol{\alpha}}_{p+q-1}})^{\mathrm{T}}$ where $p_{\hat{\boldsymbol{\alpha}}_i}$ correspond to the columns of $\mathbf{P}_{\hat{\boldsymbol{\alpha}}}$. Then, Lemma A.1 implies that $\mathbf{P}_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\mathbf{C}_n \xrightarrow{p} (\mathbf{C}_1\ \mathbf{C}_2)$ where $\mathbf{C}_1$ is defined in M8 and $\mathbf{C}_2 \in \mathbb{R}^{(p+q-1)\times 1}$.

Note that, since $\|\hat{\boldsymbol{\alpha}}\| = 1$ and $\|\boldsymbol{\alpha}_0\| = 1$, we have $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)^{\mathrm{T}}(\hat{\boldsymbol{\alpha}} + \boldsymbol{\alpha}_0) = 0$. By (30), we obtain

$$\begin{pmatrix} -\mathbf{P}_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\hat{\mathbf{V}}_n \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\mathbf{C}_n \\ 0 \end{pmatrix} \sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \end{pmatrix}.$$

The consistency of $\hat{\boldsymbol{\alpha}}$ and the fact that $\mathbf{C}_1$ is nonsingular and $\hat{\mathbf{V}}_n \xrightarrow{D} N(0, \boldsymbol{\Gamma})$, entail that $-\mathbf{P}_{\hat{\boldsymbol{\alpha}}}^{(p+q-1)}\hat{\mathbf{V}}_n \xrightarrow{D} N(0, \mathbf{P}_{\boldsymbol{\alpha}_0}^{(p+q-1)}\boldsymbol{\Gamma}\mathbf{P}_{\boldsymbol{\alpha}_0}^{(p+q-1)\mathrm{T}})$. Then, $\sqrt{n}(\hat{\alpha}_q - \alpha_{0q}) \xrightarrow{p} 0$ and

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}}^{(q-1)} - \boldsymbol{\alpha}_0^{(q-1)} \end{pmatrix} \xrightarrow{D} N\left(0, \mathbf{C}_1^{-1}\boldsymbol{\Gamma}_1\mathbf{C}_1^{-1}\right),$$

where $\mathbf{C}_1$ and $\boldsymbol{\Gamma}_1$ are defined in M8 and M9, respectively. □

References

Azadeh A, Salibian-Barrera M (2011, to appear). An outlier-robust fit for generalised additive models with applications to outbreak detection. J Am Stat Assoc. Available at http://www.stat.ubc.ca/~matias/rgam-rev1-authors.pdf

Bianco A, Boente G (2002) On the asymptotic behavior of one-step estimation. Stat Probab Lett 60:33–47

Bianco A, Yohai V (1995) Robust estimation in the logistic regression model. In: Lecture notes in statistics, vol 109. Springer, New York, pp 17–34

Bianco A, García Ben M, Yohai V (2005) Robust estimation for linear regression with asymmetric errors. Can J Stat 33:511–528

Boente G, Rodriguez D (2010) Robust inference in generalized partially linear models. Comput Stat Data Anal 54:2942–2966

Boente G, He X, Zhou J (2006) Robust estimates in generalized partially linear models. Ann Stat 34:2856–2878

Cantoni E, Ronchetti E (2001) Robust inference for generalized linear models. J Am Stat Assoc 96:1022–1030

Carroll R, Fan J, Gijbels I, Wand M (1997) Generalized partially linear single-index models. J Am Stat Assoc 92:477–489

Croux C, Haesbroeck G (2002) Implementing the Bianco and Yohai estimator for logistic regression. Comput Stat Data Anal 44:273–295

Delecroix M, Härdle W, Hristache M (2003) Efficient estimation in conditional single-index regression. J Multivar Anal 86:213–226

Friedman J, Stuetzle W (1981) Projection pursuit regression. J Am Stat Assoc 76:817–823

Härdle W, Mammen E, Müller M (1998) Testing parametric versus semiparametric modeling in generalized linear models. J Am Stat Assoc 93:1461–1474

Härdle W, Müller M, Sperlich S, Werwatz A (2006) Nonparametric and semiparametric models. Springer, Berlin

Hastie TJ, Tibshirani RJ (1990) Generalized additive models. Chapman & Hall, CRC, New York
He X, Zhu Z, Fung W (2002) Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika 89:579–590
Künsch H, Stefanski L, Carroll R (1989) Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. J Am Stat Assoc 84:460–466
McCullagh P, Nelder J (1989) Generalized linear models, 2nd edn. Champman and Hall, London
Pollard D (1984) Convergence of stochastic processes. Springer series in statistics. Springer, New York
Robinson P (1988) Root-n consistent semiparametric regression. Econometrica 56:931–954
Rodriguez D (2008) Doctoral thesis, Universidad de Buenos Aires. Available at http://cms.dm.uba.ar/academico/carreras/doctorado/tesisdanielarodriguez.pdf
Severini T, Staniswalis J (1994) Quasi-likelihood estimation in semiparametric models. J Am Stat Assoc 89:501–511
Severini T, Wong W (1992) Profile likelihood and conditionally parametric models. Ann Stat 20:4 1768-1802
Stefanski L, Carroll R, Ruppert D (1986) Bounded score functions for generalized linear models. Biometrika 73:413–424
van der Vaart A (1988) Estimating a real parameter in a class of semiparametric models. Ann Stat 16(4):1450–1474
Wang JL, Xue L, Zhu L, Chong Y (2010) Estimation for a partial-linear single-index model. Ann Stat 38:246–274
Xia Y, Härdle W (2006) Semi-parametric estimation of partially linear single-index models. J Multivar Anal 97:1162–1184
Xia Y, Tong H, Li WK (1999) On extended partially linear single-index models. Biometrika 86:831–842
Yi GY, He W, Liang H (2009) Analysis of correlated binary data under partially linear single-index logistic models. J Multivar Anal 100:278–290