Annals of the Institute of Statistical Mathematics (2020) 72:855–893
https://doi.org/10.1007/s10463-019-00712-8

Robust estimation in single-index models when the errors have a unimodal density with unknown nuisance parameter

Claudio Agostinelli[1] · Ana M. Bianco[2] · Graciela Boente[3]

Received: 26 January 2018 / Revised: 15 February 2019 / Published online: 21 March 2019

Abstract
This paper develops a robust profile estimation method for the parametric and nonparametric components of a single-index model when the errors have a strongly unimodal density with unknown nuisance parameter. We derive consistency results for the link function estimators as well as consistency and asymptotic distribution results for the single-index parameter estimators. Under a log-Gamma model, the sensitivity to anomalous observations is studied using the empirical influence curve. We also discuss a robust K-fold cross-validation procedure to select the smoothing parameters. A numerical study carried on with errors following a log-Gamma model and for contaminated schemes shows the good robustness properties of the proposed estimators and the advantages of considering a robust approach instead of the classical one. A real data set illustrates the use of our proposal.

Keywords Kernel weights · Fisher consistency · Local polynomials · Single-index models · Robustness

✉ Graciela Boente
gboente@dm.uba.ar

Claudio Agostinelli
claudio.agostinelli@unitn.it

Ana M. Bianco
abianco@dm.uba.ar

1 Dipartimento di Matematica, Università di Trento, Via Sommarive, 14, 38123 Trento, Italy

2 Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Pabellón 2, 1428 Buenos Aires, Argentina

3 Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IMAS, CONICET, Ciudad Universitaria, Pabellón 1, 1428 Buenos Aires, Argentina

1 Introduction

Semiparametric models are an appealing compromise between parametric and non-parametric paradigms. These models represent an intermediate point between a fully parametric model, which is usually of easy interpretation but vulnerable to poor specification, and a fully nonparametric model, which is more flexible but suffers from the well-known curse of dimensionality. Semiparametric modelling combines parametric components with nonparametric ones, retaining the advantages of both types of approaches and avoiding their drawbacks.

Single-index models are a relevant topic within the broad class of semiparametric methods with a great potential to model data in different scientific disciplines. These models have raised a lot of interest in part due to the fact that they reduce the dimensionality of the covariates through a suitable projection linked to the parametric component, while at the same time they capture a possible nonlinear relationship through an unknown smooth function.

Under a single-index model, the response variable y is related to the covariates $\mathbf{x}$ through the equation

$$y = \eta(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \epsilon, \tag{1}$$

where the single-index parameter $\boldsymbol{\beta} \in \mathbb{R}^q$ and the link univariate real-valued function $\eta : \mathbb{R} \to \mathbb{R}$ are both unknown. For the sake of identifiability, it is assumed with no loss of generality that $\|\boldsymbol{\beta}\| = 1$ and the last component of $\boldsymbol{\beta}$ is positive, where $\|\cdot\|$ denotes the Euclidean norm. Furthermore, in the classical setting, it is usually assumed that $\mathbb{E}(\epsilon|\mathbf{x}) = 0$ and $\mathbb{E}(\epsilon^2|\mathbf{x}) < \infty$.

As noted above, in our framework $\|\boldsymbol{\beta}\| = 1$ and we may assume that $\beta_q \neq 0$, without loss of generality. However, some authors consider a different parametrization given by

$$y = \eta^\star(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{x}) + \epsilon, \tag{2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}^\star, \theta_q)$ with $\theta_q = 1$ and $\boldsymbol{\theta}^\star = (\theta_1, \ldots, \theta_{q-1}) \in \mathbb{R}^{q-1}$, which also leads to an identifiable model. One of the advantages of parametrization (1) over that given in (2) is that the finite-dimensional parameter $\boldsymbol{\beta}$ naturally belongs to a compact set. The relation between both parametrizations is given by $\boldsymbol{\beta} = \boldsymbol{\theta}/\|\boldsymbol{\theta}\|$ and $\eta(u) = \eta^\star(u\,\|\boldsymbol{\theta}\|)$, while $\boldsymbol{\theta} = \boldsymbol{\beta}/\beta_q$ and $\eta^\star(u) = \eta(u\,\beta_q)$. So, estimators in any of these two parametrizations lead to estimators in the other one.

Single-index models have received an increasing amount of attention in the last years, probably because they have an appealing feature: they cope with the *curse of dimensionality* combining nonparametric and parametric-driven approaches. Beneath single-index models underlies the idea that the contribution of the vector of covariates $\mathbf{x}$ to the response y can be expressed in terms of a one-dimensional projection. In this sense, these models can be seen as a dimension reduction technique since, once $\boldsymbol{\beta}$ has been estimated, the unidimensional variable $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ can be used as a univariate carrier to estimate nonparametrically the function η.

There is an extensive literature in this area. Among the first works, we can mention Powell et al. (1989), Härdle and Stoker (1989), Härdle et al. (1993), Xia et al. (2002) and Carroll et al. (1997). More recently, Xia and Härdle (2006) studies the asymptotic distribution of a class of estimators, Chang et al. (2010) consider the heteroscedastic case and Xia et al. (2012) propose a family of estimators of the nonparametric component for which it is not necessary to undersmooth in order to obtain a $\sqrt{n}$-rate estimator of the parametric component. On the other hand, Wu et al. (2010) consider the estimation of the single-index quantile regression, while Liu et al. (2013) propose robust estimators by means of the mode, without taking into account the estimation of a possible scale factor. Xue and Zhu (2006) focus on the problem of looking for confidence regions and intervals, and Zhang et al. (2010) study the problem of testing hypotheses that involve $\boldsymbol{\beta}$. Recently, Li and Patilea (2017) considered a quadratic form criterion involving kernel smoothing and propose a resampling method to build confidence intervals for the index parameter. Wang et al. (2014) also consider the extension of these models to the situation in which there are missing responses. All the aforementioned procedures are based on classical methods, and hence, they are very sensitive to the presence of outliers.

Indeed, even when different approaches have been proposed for fitting model (1), such as kernel smoothing or sliced inverse regression methods, in most cases it is assumed that the error distribution has finite first moment. In the robust framework, this assumption is generally replaced by the symmetry of the error term distribution, in order to achieve Fisher-consistent estimators. However, in practice, situations arise in which the errors are asymmetric, as it is the case when the error term distribution belongs to some class of exponential families, such as the log-Gamma distribution. In this paper, we focus on the problem of robust estimating the parametric and nonparametric components of model (1) when the density of the error ϵ is of the form

$$g(s,\gamma) = Q(\gamma)\exp^{\gamma\, t(s)}, \tag{3}$$

where $\gamma > 0$ is an unknown parameter and t is a continuous function with unique mode at e_0. Under a linear regression model, this family of exponential distributions has been previously considered by Bianco et al. (2005) in their attempt to extend MM-estimators to the case of asymmetric errors. An attractive feature of this family of distributions is that it enables to model either symmetric or asymmetric errors, as well. A prominent member of this family is the log-Gamma distribution that is frequently used to fit asymmetric data. Furthermore, in linear regression models, this family of errors distribution leads to the log-Gamma regression model which corresponds to a generalized linear model with log link function.

A first approach to deal with outliers in the responses was given in Delecroix et al. (2006) who considered M-type estimators for single-index models with known nuisance parameter. In contrast, Boente and Rodriguez (2012) proposed robust estimators of the parametric and nonparametric components under a generalized partially linear single-index model by assuming that the conditional model of the responses given the covariates belongs to a canonical exponential family. In this sense, a first contribution of our approach is that when the errors in Eq. (1) have symmetric distribution with unknown scale σ, our proposal is distributional free. This means that, under symmetry, by taking the nuisance parameter γ as σ, we do not need to assume a known density for the errors, as it is the case in generalized linear models. Last but not least, it should be emphasized that in both Delecroix et al. (2006) and Boente and Rodriguez (2012) it is assumed that the nuisance parameters are known, which may be restrictive for practical uses. Moreover, a linear regression model with asymmetric errors is typically fitted using a log-Gamma distribution where γ represents the unknown shape parameter. In most cases, the estimation of γ is crucial to down-weight large residuals. In fact, as in linear regression, it is necessary to determine the size of the residuals to decide if an observation is an outlier or not and this task strongly depends on a good preliminary nuisance parameter estimator. The symmetric and asymmetric errors situations show how important is to estimate γ in order to calibrate the robust estimators.

Consequently, in this paper, we go beyond and we contemplate a more realistic situation for model (1) with error distribution in (3), in which additional parameters of shape or scale have to be estimated. For this purpose, we introduce a stepwise procedure based on robust profile estimators. We make special emphasis in the case of errors with log-Gamma distribution, which is often employed in applications, and then we extend the proposal to the general setting. Under mild conditions, the estimators of η are consistent and the parametric component estimators are consistent and asymptotically normal with $\sqrt{n}$-rate. We also provide a class of initial estimators and a robust K-fold cross-validation procedure to select the bandwidth parameters involved in our proposal.

The outline of the paper is as follows. In Sect. 2, the three-step procedure for robust estimation under a single-index model is introduced first for log-Gamma errors, and then, it is extended to more general situations. In Sect. 3, we give some asymptotic properties of the proposal, while in Sect. 4, we compute the empirical influence function which may be helpful to study the sensitivity of the estimators to atypical observations. Section 5 presents a robust K-fold cross-validation method to select the smoothing parameters for the proposed robust estimators. The robustness and performance for finite samples of the proposed method are analysed by means of a numerical study in Sect. 6. Finally, in Sect. 7 we present an application to a real data set that illustrates the use of our proposal. Proofs are relegated to the "Appendix".

2 The estimators

Let $(y_i, \mathbf{x}_i) \in \mathbb{R}^{q+1}$ be independent observations that follow model (1) for $\eta = \eta_0$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and assume that the errors ϵ_i are independent, independent of $\mathbf{x}_i$ and have density (3) with $\gamma = \gamma_0$. Denote $\mathbb{E}_0$ the expectation under the true model and α_0 the true nuisance parameter which as mentioned above is a function of γ_0.

2.1 The log-Gamma setting

In order to introduce the proposed estimators, let us first revisit the particular case of the purely parametric regression model with log-Gamma errors, that is, with density

$$g(s,\gamma) = \frac{\gamma^\gamma}{\Gamma(\gamma)} \exp^{\gamma\,(s-\exp(s))}\,. \tag{4}$$

Assume that the variable $z \in \mathbb{R}_{\geq 0}$ and the covariates $\mathbf{x} \in \mathbb{R}^q$ are such that $z|\mathbf{x} \sim \Gamma(\gamma_0, \mu(\mathbf{x}))$, where the parametrization is such that $\mathbb{E}(z|\mathbf{x}) = \mu(\mathbf{x})$ and $\log \mu(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0$. Hence, defining $u = z/\mu(\mathbf{x})$, we have that $u \sim \Gamma(\gamma_0, 1)$ and therefore, if $y = \log(z)$ and $\epsilon = \log(u)$, we get that

$$y = \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x} + \epsilon, \tag{5}$$

where $\epsilon \sim \log(\Gamma(\gamma_0, 1))$ has a density given by (4) with $\gamma = \gamma_0$, i.e. it belongs to the family given in (3).

In the log-Gamma model, the classical estimators are based on the maximum likelihood method and are defined through the minimization of the deviance, whose components are given by $d(y, a) = \exp(y-a)-(y-a)-1$. A natural way to robustify these estimators is by means of an M-estimation procedure. Thus, if $(y_i, \mathbf{x}_i) \in \mathbb{R}^{q+1}$, $1 \leq i \leq n$, are independent observations following model (5), an M-estimator is defined as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \phi(y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \widehat{c}) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \rho\left(\frac{\sqrt{d(y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)}}{\widehat{c}}\right), \tag{6}$$

where $\widehat{c}$ is a preliminary estimate of a tuning constant c_0 and ρ is a bounded and continuous loss function such as the Tukey's biweight function given by $\rho(s) = \rho_{\mathrm{T}}(s) = \min\left(1, 3s^2 - 3s^4 + s^6\right)$. For this family of distributions, the nuisance parameter can be taken as the tuning constant c_0 that is related to the unknown shape parameter γ_0. Fisher consistency for this family of estimators has been studied in Bianco et al. (2005), under general conditions.

With this background in mind, let us now consider the case of a single-index model with log-Gamma errors, that is, $(y_i, \mathbf{x}_i) \in \mathbb{R}^{q+1}$, $1 \leq i \leq n$, is a random sample where

$$y_i = \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}) + \epsilon_i \quad \text{and} \quad \epsilon_i \sim \log(\Gamma(\gamma_0, 1))\,. \tag{7}$$

We will borrow some of the previous ideas to introduce a robust profile method that involves smoothing and parametric techniques. 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). In order to introduce the smoothers, we will consider local weights. For the sake of simplicity, given $\boldsymbol{\beta}$ we define the kernel weights $W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$ as

$$W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i) = K_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u\right)\left\{\sum_{j=1}^n K_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j - u\right)\right\}^{-1},$$

where $K_h(u) = (1/h)\, K(u/h)$ with K a kernel function, i.e. a nonnegative integrable function on $\mathbb{R}$ and h is the bandwidth parameter. The weights $W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$ depend on the closeness between the point u and the projection of $\mathbf{x}_i$ on the direction $\boldsymbol{\beta}$, i.e. between u and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$. To assume that a consistent estimator of the tuning constant, $\widehat{c}$, is available, let $\widehat{\gamma}_{\mathrm{R}}$ stand for a preliminary robust consistent estimator of γ_0 allowing to define $\widehat{c} = \widehat{c}(\widehat{\gamma}_{\mathrm{R}})$. The latter estimators must be properly computed according to the underlying errors distribution whose density we assume in the family given in (3). In Sect. 2.3, we introduce a robust consistent estimator of the nuisance parameter for the usual regression model with symmetric errors and for the log-Gamma regression model, as well.

Then, for the particular situation of model (7) we propose the following stepwise procedure

Step LG1 For each fixed $\boldsymbol{\beta}$, with $\|\boldsymbol{\beta}\| = 1$, let

$$\widehat{\eta}_{\boldsymbol{\beta}}(u) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d(y_i, a)}}{\widehat{c}}\right) W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i).$$

Step LG2 Define the estimators $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ as the minimum of $\Delta_n(\boldsymbol{\beta})$ among $\|\boldsymbol{\beta}\| = 1$, where

$$\Delta_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{\sqrt{d\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)\right)}}{\widehat{c}}\right)\tau(\mathbf{x}_i)$$

and τ is a weight function.

Step LG3 Define the final estimator $\widehat{\eta}$ of η_0 as $\widehat{\eta}(u) = \widehat{a}(u)$ with

$$(\widehat{a}(u), \widehat{b}(u)) = \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{i=1}^{n} W_h(u, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\rho\left(\frac{\sqrt{d(y_i, a + b\ (\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_j - u))}}{\widehat{c}}\right).$$

The robust estimators are obtained by controlling large values of the deviance with a bounded loss function ρ. A popular choice is the Tukey's bisquare loss function $\rho = \rho_{\mathrm{T}}$, while $\widehat{c}$ estimates the tuning constant c_0 selected to attain a given efficiency. As mentioned above, c_0 depends on the shape parameter γ_0 (see Bianco et al. 2005). Note that the three steps involve the function

$$\phi(y, a, c) = \rho\left(\frac{\sqrt{d(y, a)}}{c}\right),$$

where, as above, $d(y, a) = \exp(y-a)-(y-a)-1$. As mentioned in the Introduction, the tuning constant c plays the role of the nuisance parameter.

2.2 The proposal for the general setting (3)

Let us now consider the general case in which the errors have a density g in family (3). In order to extend the proposal given in Sect. 2.1 to this situation, one may consider a loss function ϕ bounding the deviances. To be more precise, let us denote as

$$\phi(y,a,\alpha)=\rho\left(\frac{\sqrt{d(y,a)}}{\alpha}\right)\,,$$

where $d(y,a)=t(e_0)-t(y-a)$, with e_0 the unique mode of the density g and α is the tuning constant related to the nuisance parameter. As in Maronna et al. (2006), $\rho:\mathbb{R}\to\mathbb{R}_+$ is a ρ-function, that is, an even function, non-decreasing on $|s|$, increasing for $s>0$ when $\rho(s)<\lim_{x\to+\infty}\rho(x)$ and such that $\rho(0)=0$.

We define for each $\boldsymbol{\beta}$ and any continuous function $v:\mathbb{R}\to\mathbb{R}$ the functions

$$\Upsilon(\boldsymbol{\beta},a,u,\alpha)=\mathbb{E}_0\left[\phi\left(y,a,\alpha\right)|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}=u\right]\,, \tag{8}$$

$$\Delta(\boldsymbol{\beta},v,\alpha)=\mathbb{E}_0\left[\phi\left(y,v(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\alpha\right)\tau(\mathbf{x})\right]\,, \tag{9}$$

where τ is a weight function as above. Denote as $\eta_{\boldsymbol{\beta}}(u)=\operatorname{argmin}_{a\in\mathbb{R}}\Upsilon(\boldsymbol{\beta},a,u,\alpha_0)$. Note that, since we are considering the deviance and a continuous family of distributions with strongly unimodal density, there is no need to introduce a correction term to attain Fisher consistency (see Bianco et al. 2005). More precisely, we have that $\boldsymbol{\beta}_0=\operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^q}\Delta(\boldsymbol{\beta},\eta_0,\alpha_0)$ and $\eta_{\boldsymbol{\beta}_0}=\eta_0$. Furthermore, $\boldsymbol{\beta}_0$ is the unique minimum of $\Delta(\boldsymbol{\beta},\eta_0,\alpha_0)$.

In order to define consistent estimators of the parametric and nonparametric components, let us consider the empirical versions of the objective functions (8) and (9), respectively, as

$$\Upsilon_n(\boldsymbol{\beta},a,u,\alpha)=\sum_{i=1}^{n}W_h(u,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\phi\left(y_i,a,\alpha\right),$$

$$\Delta_n(\boldsymbol{\beta},v,\alpha)=\frac{1}{n}\sum_{i=1}^{n}\phi\left(y_i,v(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i),\alpha\right)\tau(\mathbf{x}_i),$$

where v is any continuous function $v:\mathbb{R}\to\mathbb{R}$.

Assume that an initial robust estimator of α, $\widehat{\alpha}_{\mathrm{R}}$, is available. For a general single-index model, the *robustified profile method* can thus be defined as

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

$$\widehat{\eta}_{\boldsymbol{\beta}}(u)=\operatorname*{argmin}_{a\in\mathbb{R}}\Upsilon_n(\boldsymbol{\beta},a,u,\widehat{\alpha}_{\mathrm{R}}).$$

Step 2 Define the estimators $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ as

$$\widehat{\boldsymbol{\beta}} = \underset{\|\boldsymbol{\beta}\|=1}{\operatorname{argmin}}\, \Delta_n(\boldsymbol{\beta}, \widehat{\eta}_{\boldsymbol{\beta}}, \widehat{\alpha}_{\mathrm{R}}).$$

Step 3 Define the final estimator $\widehat{\eta}$ of η_0 as $\widehat{\eta}(u) = \widehat{a}(u)$ with

$$(\widehat{a}(u), \widehat{b}(u)) = \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{i=1}^{n} W_h(u, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\phi\left(y_i, a + b\,(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i - u), \widehat{\alpha}_{\mathrm{R}}\right).$$

Note that the stepwise procedure defined by **Step LG.1–Step LG.3** corresponds to **Step 1–Step 3** for a particular choice of the function ϕ.

It is worth noticing that this stepwise procedure only involves unidimensional nonparametric smoothers, circumventing the sparsity of the data induced by the dimensionality of the covariates. In the third step, a local polynomial of first degree is computed in order to improve the estimation of the link function η_0. Taking into account that most kernels K attain their maximum at 0, the contribution of an observation $(y_i, \mathbf{x}_i)$ to the smoothers computed in **Steps 1** and **3** is determined by the closeness between u and $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i$. When nuisance parameters are present, they may be estimated using a preliminary S-estimator which will allow to define also the tuning constant as motivated in the next section.

2.3 Initial estimators

The calibration of the robust estimators will need the computation of a preliminary estimator of the nuisance parameter γ_0. As described in the Introduction, as for many robust estimators, this is a crucial issue for the three-step procedure and it can be accomplished in different ways according to the underlying error distribution. We will illustrate the computation of an initial estimator of the nuisance parameter for the log-Gamma model, which can be extended to the case of errors with density in the family given in (3). In Sect. 2.4, we consider the situation in which the errors have a symmetric distribution.

The preliminary estimator of the shape parameter γ_0 under model (7) allows to compute the tuning constant by means of an S-estimator. S-estimators were introduced by Rousseeuw and Yohai (1984) for ordinary regression and studied in the framework of linear regression with asymmetric errors in Bianco et al. (2005). Let ρ_{T} be the bisquare ρ-function and consider the following S-estimator.

Step ILG.1 For each value of a, u and $\boldsymbol{\beta}$, compute $s_{n,\boldsymbol{\beta},u}(a)$ as the solution of

$$\sum_{i=1}^{n} \rho_{\mathrm{T}}\left(\frac{\sqrt{d(y_i, a)}}{s_{n,\boldsymbol{\beta},u}(a)}\right) W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i) = b,$$

where, for instance, $b = 1/2$ and $d(y, a) = \exp(y - a) - (y - a) - 1$. Define $\widetilde{\eta}_{\boldsymbol{\beta}}(u)$ as the value $\widetilde{\eta}_{\boldsymbol{\beta}}(u) = \operatorname{argmin}_a s_{n,\boldsymbol{\beta},u}(a)$.

Step ILG.2 For each $\boldsymbol{\beta}$, let $\widetilde{\sigma}(\boldsymbol{\beta})$ be the solution of

$$\frac{1}{\sum_{i=1}^n \tau(\mathbf{x}_i)} \sum_{i=1}^n \rho_{\mathrm{T}} \left(\frac{\sqrt{d\left(y_i, \widetilde{\eta}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)\right)}}{\widetilde{\sigma}(\boldsymbol{\beta})} \right) \tau(\mathbf{x}_i) = b.$$

Now, the estimator of $\boldsymbol{\beta}_0$ is given by $\widetilde{\boldsymbol{\beta}} = \operatorname{argmin}_{\|\boldsymbol{\beta}\|=1} \widetilde{\sigma}(\boldsymbol{\beta})$ and $\widehat{s}_n = \widetilde{\sigma}(\widetilde{\boldsymbol{\beta}})$.

Step ILG.3 Define the estimator of γ_0 as $\widehat{\gamma} = S^{\star\,-1}(\widehat{s}_n)$ where $S^{\star}(\gamma)$ is the solution of

$$\mathbb{E}_{\gamma}\rho_{\mathrm{T}}\left(\frac{\sqrt{d(\epsilon, 0)}}{S^{\star}(\gamma)}\right) = \mathbb{E}_{\gamma}\rho_{\mathrm{T}}\left(\frac{\sqrt{\exp(\epsilon) - 1 - \epsilon}}{S^{\star}(\gamma)}\right) = b$$

where ϵ has density $g(s, \gamma)$ given in (4).

This method provides an estimator of γ_0 as well as an initial estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$, which is robust, but may be inefficient. It also provides an estimator of the function η_0 as $\widehat{\eta} = \widetilde{\eta}_{\widetilde{\boldsymbol{\beta}}}$. These estimators may be used to start the stepwise estimation procedure in **Steps LG1–LG3** given above. In Bianco et al. (2005) it is shown that $S^{\star}(\gamma)$ is a one-to-one function and thus invertible. For this reason, they recommend to take the data-driven tuning constant in (6) as $\widehat{c}_n \geq \widehat{s}_n = S^{\star}(\widehat{\gamma})$.

It is worth noting that if we replace $d(y, a) = \exp(y - a) - (y - a) - 1$ by $d(y, a) = t(e_0) - t(y - a)$ in the initial Steps **ILG.1–ILG.3**, the described procedure provides preliminary estimators when the errors have density given by (3).

2.4 The model with symmetric errors

As it is noted above, the family of densities given in (3) also includes symmetric distributions. In this case, a suitable initial method that exploits this feature of the errors distribution can be introduced. Thus, as a second example, we consider the symmetric setting. We set $\alpha = \sigma$ and $\rho_0(u) = \rho_{\mathrm{T}}(u/c_0)$, where c_0 is the tuning constant needed to obtain a scale Fisher-consistent estimator. For instance, when dealing with Tukey's bisquare function ρ_{T}, the choice $c_0 = 1.54764$ and $b = 1/2$ leads to a scale Fisher-consistent estimator at the normal distribution with breakdown point 50%. Then, to provide a preliminary estimator of the true scale parameter $\alpha_0 = \sigma_0$, let us consider an S-estimator that can easily be computed as follows.

Step IS.1 For each value of u and $\boldsymbol{\beta}$, compute $\widetilde{\eta}_{\boldsymbol{\beta}}(u)$ as the median of the empirical local distribution

$$F_{n,\boldsymbol{\beta},h}(s) = \sum_{i=1}^n \mathbb{I}_{(-\infty,s]}(y_i) W_h(u, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i).$$

Step IS.2 For each $\boldsymbol{\beta}$, let $\widetilde{\sigma}(\boldsymbol{\beta})$ be the solution of

$$\frac{1}{\sum_{i=1}^{n}\tau(\mathbf{x}_i)}\sum_{i=1}^{n}\rho_0\left(\frac{y_i-\widetilde{\eta}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)}{\widetilde{\sigma}(\boldsymbol{\beta})}\right)\tau(\mathbf{x}_i)=b,$$

where, for instance, $b=1/2$. Now, the estimators of $\boldsymbol{\beta}_0$ and σ_0 are given as $\widetilde{\boldsymbol{\beta}}=\operatorname{argmin}_{\|\boldsymbol{\beta}\|=1}\widetilde{\sigma}(\boldsymbol{\beta})$ and $\widehat{\alpha}_{\mathrm{R}}=\widehat{\sigma}=\widetilde{\sigma}(\widetilde{\boldsymbol{\beta}})$.

To improve the efficiency of the estimators of $\boldsymbol{\beta}_0$, consider $\rho_1(u)=\rho_{\mathrm{T}}(u/c_1)$, with $c_1>c_0$, and define an MM-procedure as follows.

Step S.1 For each value of u and $\boldsymbol{\beta}$, compute $\widetilde{\eta}_{\boldsymbol{\beta}}(u)$ as

$$\widehat{\eta}_{\boldsymbol{\beta}}(u)=\underset{a\in\mathbb{R}}{\operatorname{argmin}}\sum_{i=1}^{n}W_h(u,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\rho_1\left(\frac{y_i-a}{\widehat{\sigma}}\right).$$

Step S.2 Define the estimator $\widehat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\beta}}=\underset{\|\boldsymbol{\beta}\|=1}{\operatorname{argmin}}\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\frac{y_i-\widehat{\eta}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)}{\widehat{\sigma}}\right)\tau(\mathbf{x}_i).$$

Step S.3 For each value of u, define the final estimator $\widehat{\eta}$ of η_0 as $\widehat{\eta}(u)=\widehat{a}(u)$ with

$$(\widehat{a}(u),\widehat{b}(u))=\underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}}\sum_{i=1}^{n}W_h(u,\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\rho_1\left(\frac{y_i-a-b(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i-u)}{\widehat{\sigma}}\right).$$

Note that the stepwise procedure defined by **Steps S.1** to **S.3** corresponds to **Step 1–Step 3** for a particular choice of the function ϕ, that is, $\phi(y,a,\alpha)=\rho_1((y-a)/\alpha)$.

3 Asymptotic results

In this section, we derive, under some regularity conditions, the consistency of the estimators defined in Sect. 2.2 through Steps 1 to 3. We will assume that $\mathbf{x}\in\mathcal{X}\subset\mathbb{R}^p$. Let $\mathcal{X}_0\subset\mathcal{X}$ be a compact set and define the set $\mathcal{U}(\mathcal{X}_0)=\{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}:\mathbf{x}\in\mathcal{X}_0,\ \boldsymbol{\beta}\in\mathcal{S}_1\}$, where $\mathcal{S}_1$ is the unit ball in $\mathbb{R}^p$, i.e. $\mathcal{S}_1=\{\boldsymbol{\beta}\in\mathbb{R}^p:\|\boldsymbol{\beta}\|=1\}$. For any continuous function $v:\mathcal{U}(\mathcal{X}_0)\to\mathbb{R}$ denote $\|v\|_{0,\infty}=\sup_{u\in\mathcal{U}(\mathcal{X}_0)}|v(u)|$. We consider the following set of assumptions:

A1 The loss function ρ and the function t defined in (3) are continuous. Moreover, ρ and τ are bounded.

A2 The kernel $K:\mathbb{R}\to\mathbb{R}$ is an even, nonnegative, continuous and bounded function, with bounded variation, satisfying $\int K(u)\mathrm{d}u=1$, $\int u^2K(u)\mathrm{d}u<\infty$ and $|u|K(u)\to 0$ as $|u|\to\infty$.

A3 The bandwidth sequence $h = h_n$ is such that $h \to 0$, $nh/\log(n) \to \infty$ when $n \to \infty$.

A4 i) The marginal density $f_{\mathbf{X}}$ of $\mathbf{x}$ is bounded in $\mathcal{X}$.
ii) Given any compact set $\mathcal{X}_0 \subset \mathcal{X}$, there exists a positive constant $A_1(\mathcal{U}(\mathcal{X}_0))$ such that $A_1(\mathcal{U}(\mathcal{X}_0)) < f_{\boldsymbol{\beta}}(u)$ for all $u \in \mathcal{U}(\mathcal{X}_0)$ and $\|\boldsymbol{\beta}\| = 1$, where $f_{\boldsymbol{\beta}}$ is the marginal density of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$.

A5 The function $\Upsilon(\boldsymbol{\beta}, a, u, \alpha)$ satisfies the following equicontinuity condition: given $\mathcal{X}_0 \subset \mathcal{X}$ and $\mathcal{K} \subset \mathbb{R}_{>0}$ compact sets, for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $u_1, u_2 \in \mathcal{U}(\mathcal{X}_0)$; $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{S}_1$ and $\alpha_1, \alpha_2 \in \mathcal{K}$,

$$\begin{aligned} &|u_1 - u_2| < \delta\ ,\ |\alpha_1 - \alpha_2| < \delta \ \text{ and } \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta \\ &\Rightarrow \sup_{a\in\mathbb{R}} |\Upsilon(\boldsymbol{\beta}_1, a, u_1, \alpha_1) - \Upsilon(\boldsymbol{\beta}_2, a, u_2, \alpha_2)| < \epsilon\ . \end{aligned}$$

A6 The function $\Upsilon(\boldsymbol{\beta}, a, u, \alpha)$ is continuous and $\eta_{\boldsymbol{\beta},\alpha}(u) = \operatorname{argmin}_{a\in\mathbb{R}} \Upsilon(\boldsymbol{\beta}, a, u, \alpha)$ is a continuous function on $(\boldsymbol{\beta}, u, \alpha)$.

A7 The initial estimator of α, $\widehat{\alpha}_{\mathrm{R}}$, is a consistent estimator.

A8 The functions ρ and t are differentiable functions.

Remark 1 Condition **A1** is fulfilled by the loss functions commonly used in the framework of robustness such as Tukey's bisquare function and guarantees that $\phi(y, a, \alpha)$ is a continuous and bounded function. Assumptions **A2** and **A3** are standard in nonparametric regression. Moreover, **A2** is verified for the Epanechnikov and Gaussian kernels, while **A3** is satisfied choosing $h_n = n^{-q}$ for $q > 0$. **A4** is a standard condition in semiparametric models; in particular ii) is achieved if $f_{\mathbf{X}}(\mathbf{x}) > B_1(\mathcal{X}_0)$ for any $\mathbf{x} \in \mathcal{U}(\mathcal{X}_0)$. Note that **A8** entails that $\phi(y, a, \alpha)$ is a continuously differentiable function with respect to a. We will denote as $\phi'(y, a, \alpha)$ its partial derivative with respect to a.

The following Lemma gives the uniform convergence of $\widehat{\eta}_{\boldsymbol{\beta},\alpha}$ to $\eta_{\boldsymbol{\beta},\alpha}$. Its proof is omitted since it follows using analogous arguments to those considered in the proof of Lemma 3.1 in Boente and Rodriguez (2012).

Lemma 1 *Let $\mathcal{K} \subset \mathbb{R}_{>0}$ and $\mathcal{X}_0 \subset \mathcal{X}$ be compact sets and assume that there exists $\delta_0 > 0$ such that $\mathcal{X}_{\delta_0,0} \subset \mathcal{X}$, where $\mathcal{X}_{\delta_0,0}$ stands for the closure of a δ_0-neighbourhood of $\mathcal{X}_0$. Assume that **A1** to **A6** hold and that the family of functions $\mathcal{F} = \{f(y) = \phi(y, a, \alpha)\ , a \in \mathbb{R}, \alpha \in \mathcal{K}\}$ has a covering number satisfying $\sup_{\mathbb{Q}} N\left(\epsilon, \mathcal{F}, L^1(\mathbb{Q})\right) \le A\epsilon^{-W}$, 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 that*

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

b) *If* $\displaystyle\inf_{\substack{\boldsymbol{\beta}\in\mathcal{S}_1,\ \alpha\in\mathcal{K} \\ u\in\mathcal{U}(\mathcal{X}_0)}} \left[\lim_{|a|\to\infty} \Upsilon(\boldsymbol{\beta}, a, u, \alpha) - \Upsilon(\boldsymbol{\beta}, \eta_{\boldsymbol{\beta},\alpha}(u), u, \alpha)\right] > 0$, *then* $\displaystyle\sup_{\boldsymbol{\beta}\in\mathcal{S}_1,\ \alpha\in\mathcal{K}} \|\widehat{\eta}_{\boldsymbol{\beta},\alpha} - \eta_{\boldsymbol{\beta},\alpha}\|_{0,\infty} \xrightarrow{a.s.} 0$, *where* $\widehat{\eta}_{\boldsymbol{\beta},\alpha}(u) = \operatorname{argmin}_{a\in\mathbb{R}} \Upsilon_n(\boldsymbol{\beta}, a, u, \alpha)$.

Remark 2 The condition on the infimum assumed in Lemma 1b) warranties that the infimum of function Υ in (8) is not attained at infinity. Recall that finite-dimensional families of functions are VC-classes of functions as defined in Pollard (1984). Hence, using that

$$\mathcal{F} = \left\{ f(y) = \phi\,(y, a, \alpha) = \rho\left(\frac{\sqrt{t(e_0) - t(y-a)}}{\alpha}\right), a \in \mathbb{R}, \alpha \in \mathcal{K} \right\},$$

we obtain that the required condition on the covering number depends on the behaviour of the function $t(s)$. In particular, for the log-Gamma regression model, this condition is satisfied for any ρ-function.

From Lemma 1, the continuity with respect to $\boldsymbol{\beta}$ of the function $\eta_{\boldsymbol{\beta},\alpha}(u)$ defined in **A6** and condition **A7**, we obtain the following result recalling that $\Delta(\boldsymbol{\beta}, \eta_0, \alpha_0)$ has a unique minimum at $\boldsymbol{\beta}_0$.

Theorem 1 *Let* $\widehat{\boldsymbol{\beta}}$ *be defined* $\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}}\, \Delta_n(\boldsymbol{\beta}, \widehat{\eta}_{\boldsymbol{\beta},\widehat{\alpha}_{\mathrm{R}}}, \widehat{\alpha}_{\mathrm{R}})$, *where* $\widehat{\eta}_{\boldsymbol{\beta},\alpha} = \text{argmin}_{a\in\mathbb{R}}\, \Upsilon_n(\boldsymbol{\beta}, a, u, \alpha)$ *satisfies*

$$\sup_{\boldsymbol{\beta}\in\mathcal{S}_1,\ \alpha\in\mathcal{K}} \|\widehat{\eta}_{\boldsymbol{\beta},\alpha} - \eta_{\boldsymbol{\beta},\alpha}\|_{0,\infty} \xrightarrow{a.s.} 0\,. \tag{10}$$

*Assume that **A1** and **A8** hold and that* $\widehat{\alpha}_{\mathrm{R}} \xrightarrow{a.s.} \alpha_0$. *Then, we have that*

a) $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{S}_1; a\in\mathcal{K}} \left|\Delta_n(\boldsymbol{\beta}, \widehat{\eta}_{\mathbf{b},a}, a) - \Delta(\boldsymbol{\beta}, \eta_{\mathbf{b},a}, a)\right| \xrightarrow{a.s.} 0$ *for any compact set* $\mathcal{K} \subset \mathbb{R}_{>0}$.

b) $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$.

The asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ can be derived using the consistency of $\widehat{\alpha}_{\mathrm{R}}$ assuming that the covariates $\mathbf{x}$ lie in a compact set with probability one. In fact, similar arguments to those considered in the proof of Theorem 3.5.3 in Rodriguez (2007) can be used, but taking into account the fact that the estimator of the nuisance parameter is consistent. In particular, we consider below the case of a log-Gamma model.

From now on, we assume that ρ is twice continuously differentiable with first and second derivatives $\Psi(y)$ and $\Psi'(y)$, respectively, and that $\eta_{\mathbf{b},a}$ is continuously differentiable in $(\mathbf{b}, a)$.

Recall that under a log-Gamma model, $y_i = \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) + \epsilon_i$ with $\epsilon_i \sim \log\,(\Gamma(\gamma_0, 1))$ independent of $\mathbf{x}_i$, so $d(y, a) = \exp(y-a) - (y-a) - 1$ and $\phi(y, a, c) = \rho\left(\sqrt{d(y,a)}/c\right)$. If we define $d^*(u) = \exp(u) - u - 1$, we have that $d(y, \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x})) = d^*(y - \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}))$. Let

$$\psi(y, a, c) = \frac{\partial}{\partial a}\phi(y, a, c) \tag{11}$$

$$\chi(y, a, c) = \frac{\partial}{\partial a}\psi(y, a, c)\,. \tag{12}$$

Hence, $\psi(y, \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}), c) = \psi^*(y - \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}), c)$ and $\chi(y, \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}), c) = \chi^*(y - \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}), c)$, where

$$\psi^*(u, c) = \frac{1}{2c}\,\Psi\left(\frac{\sqrt{d^*(u)}}{c}\right)\frac{1-\exp(u)}{\sqrt{d^*(u)}} \tag{13}$$

$$\begin{aligned}\chi^*(u, c) &= \frac{1}{4c^2}\,\Psi'\left(\frac{\sqrt{d^*(u)}}{c}\right)\frac{(1-\exp(u))^2}{d^*(u)}\\ &\quad + \frac{1}{4c}\,\Psi\left(\frac{\sqrt{d^*(u)}}{c}\right)\left[\frac{2\exp(u)}{\sqrt{d^*(u)}} - \frac{(1-\exp(u))^2}{d^*(u)^{3/2}}\right].\end{aligned} \tag{14}$$

Define $\mathbf{B} = \mathbb{E}_0\left[\chi\left(y_1, \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1), c\right)\tau(\mathbf{x})\ \boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)^{\mathrm{T}}\right]$, where

$$\boldsymbol{\nu}_i(\mathbf{b}, t) = \frac{\partial}{\partial\boldsymbol{\beta}}\eta_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)} + \frac{\partial}{\partial s}\eta_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)}\ \mathbf{x}_i\ . \tag{15}$$

Due to the independence between the errors and the covariates, $\mathbf{B}$ can be written as

$$\mathbf{B} = \mathbb{E}\left(\chi^*(\epsilon_1, c)\right)\widetilde{\mathbf{B}}\ , \tag{16}$$

where $\widetilde{\mathbf{B}} = \mathbb{E}\left[\tau(\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)^{\mathrm{T}}\right]$. Furthermore, consider the matrix

$$\begin{aligned}\boldsymbol{\Sigma} &= 4\,\mathbb{E}_0\left\{\psi^2\left(y_1, \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1), c\right)\tau^2(\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)^{\mathrm{T}}\right\}\\ &= \mathbb{E}\left\{\psi^{*\,2}(\epsilon_1, c)\right\}\widetilde{\boldsymbol{\Sigma}}\,,\end{aligned} \tag{17}$$

with $\widetilde{\boldsymbol{\Sigma}} = 4\mathbb{E}\left\{\tau^2(\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)\boldsymbol{\nu}_1(\boldsymbol{\beta}_0, \boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_1)^{\mathrm{T}}\right\}$. Let $\mathbf{B}_1$, $\widetilde{\mathbf{B}}_1$, $\boldsymbol{\Sigma}_1$ and $\widetilde{\boldsymbol{\Sigma}}_1$ be the left superior matrices of dimension $(q-1)\times(q-1)$ of $\mathbf{B}$, $\widetilde{\mathbf{B}}$, $\boldsymbol{\Sigma}$ and $\widetilde{\boldsymbol{\Sigma}}$, respectively. Assume that $\mathbf{B}$ is non-singular, $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, $\mathbf{x}_i$ are random vectors with distribution with compact support $\mathcal{X}$ and the bandwidth $h = h_n$ satisfies $nh^4 \to 0$ and $nh^2/\log(1/h) \to \infty$. Then, using analogous arguments to those considered in Rodriguez (2007) for the case of fixed nuisance parameter and taking into account that $\widehat{c}_n \xrightarrow{p} c$, we obtain that

$$\sqrt{n}(\widehat{\beta}_q - \beta_{0q}) \xrightarrow{p} 0 \tag{18}$$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^{(q-1)} - \boldsymbol{\beta}_0^{(q-1)}) \xrightarrow{D} N(0, \mathbf{B}_1^{-1}\boldsymbol{\Sigma}_1(\mathbf{B}_1^{-1})^{\mathrm{T}})\ , \tag{19}$$

where for any $\mathbf{b} \in \mathbb{R}^q$, $\mathbf{b}^{(q-1)} = (b_1, \ldots, b_{q-1})^{\mathrm{T}}$.

Hence, using (16) and (17), we get that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^{(q-1)} - \boldsymbol{\beta}_0^{(q-1)}) \xrightarrow{D} N\left(0, \frac{\mathbb{E}\psi^{*\,2}(\epsilon_1, c)}{(\mathbb{E}\chi^*(\epsilon_1, c))^2}\,\widetilde{\mathbf{B}}_1^{-1}\widetilde{\boldsymbol{\Sigma}}_1(\widetilde{\mathbf{B}}_1^{-1})^{\mathrm{T}}\right).$$

Since the classical estimator of the single-index parameter corresponds to the choice $\rho(s) = s^2$, its asymptotic covariance matrix is of the form $(1/\gamma_0)\ \widetilde{\mathbf{B}}_1^{-1}\widetilde{\boldsymbol{\Sigma}}_1(\widetilde{\mathbf{B}}_1^{-1})^{\mathrm{T}}$; therefore, the asymptotic efficiency with respect to the classical estimator is given by

$$e = \frac{1}{\gamma_0}\ \frac{(\mathbb{E}\chi^*(\epsilon_1, c))^2}{\mathbb{E}_0\psi^{*2}(\epsilon_1, c)},$$

which equals the efficiency of the MM-regression estimator described in Bianco et al. (2005). Hence, the tuning constant parameter in **Steps LG1** to **3** can be chosen to attain a given efficiency.

Let us consider the parametrization given in (2), and let $\boldsymbol{\theta} = (\boldsymbol{\theta}^\star, \theta_q)$ with $\boldsymbol{\theta}^\star = (\theta_1, \ldots, \theta_{q-1}) \in \mathbb{R}^{q-1}$ and $\theta_q = 1$. Using that the parameter $\boldsymbol{\theta}$ equals $\boldsymbol{\beta}/\beta_q$ ($\beta_q > 0$), we have that $\boldsymbol{\theta}^\star = \boldsymbol{\beta}^{(q-1)}/\beta_q$. Hence, the relation between both parameters suggests to estimate $\boldsymbol{\theta}^\star$ by means of $\widehat{\boldsymbol{\theta}}^\star = \widehat{\boldsymbol{\beta}}^{(q-1)}/\widehat{\beta}_q$. Thus, from

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}_0^\star) = \sqrt{n}\left(\frac{\widehat{\boldsymbol{\beta}}^{(q-1)}}{\widehat{\beta}_q} - \frac{\boldsymbol{\beta}_0^{(q-1)}}{\beta_{0q}}\right) = \frac{1}{\widehat{\beta}_q}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{(q-1)} - \boldsymbol{\beta}_0^{(q-1)}\right)$$
$$+\sqrt{n}\frac{(\beta_{0q} - \widehat{\beta}_q)}{\widehat{\beta}_q\beta_{0q}}\ \boldsymbol{\beta}_0^{(q-1)}\ ,$$

it is easy to see that $\sqrt{n}(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}_0^\star) \xrightarrow{D} N\left(0, (1/\beta_{0q}^2)\,\mathbf{B}_1^{-1}\boldsymbol{\Sigma}_1(\mathbf{B}_1^{-1})^{\mathrm{T}}\right)$, since (18) and (19) entail that $\sqrt{n}(\widehat{\boldsymbol{\theta}}^\star - \boldsymbol{\theta}_0^\star) = \sqrt{n}(\widehat{\boldsymbol{\beta}}^{(q-1)} - \boldsymbol{\beta}_0^{(q-1)})/\beta_{0q} + o_{\mathbb{P}}(1)$.

4 Empirical influence curve

In this section, we derive the empirical influence function of the single-index parameter estimator under a log-Gamma model. The empirical influence function (EIF), introduced by Tukey (1977), measures the robustness of an estimator with respect to a single outlier. Essentially, it assesses the impact on an estimator of adding an arbitrary observation to the sample. Diagnostic measures with the purpose of outlier identification can be defined from the empirical influence function. Mallows (1974) defines a finite version of the influence function, introduced by Hampel (1974), that is computed at the sample empirical distribution. The EIF has been widely used in parametric statistics, but has retrieved less attention in nonparametric literature. Foremost, Manchester (1996) introduces a simple graphical procedure to display the sensitivity of a scatter plot smoother to perturbations in the data. Tamine (2002) defines a smoothed influence function in the context of nonparametric regression with a fixed bandwidth that is based on Aït Sahalia (1995) smoothed functional approach to nonparametric kernel estimators.

Following Boente and Rodriguez (2010), we consider an empirical influence function that is close to Manchester (1996) approach and at the same time, retains the spirit of the EIF definition introduced by Mallows (1974).

To be more precise, denote $\widehat{\boldsymbol{\beta}}$ the single-index parameter estimator based on the original data set $(y_i, \mathbf{x}_i)$, $1 \leq i \leq n$. If P_n is the empirical measure that gives weight $1/n$ to each datum in the sample, we have that $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(P_n)$. Let $P_{n,\varepsilon}$ be the empirical measure that gives mass $(1-\varepsilon)/n$ to each $(y_i, \mathbf{x}_i)$, $1 \leq i \leq n$ and mass ε to the arbitrary observation $(y_0, \mathbf{x}_0)$. In other words, we have a new sample with the original data set accounting an $1-\varepsilon$ proportion and the new observation an ε proportion. Now, denote $\widehat{\boldsymbol{\beta}}_\varepsilon = \widehat{\boldsymbol{\beta}}(P_{n,\varepsilon})$ the single-index parameter estimator for the new sample. We compute the empirical influence function of $\widehat{\boldsymbol{\beta}}$ at a given point $(y_0, \mathbf{x}_0)$ as

$$\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{x}_0)) = \lim_{\varepsilon \to 0} \frac{\widehat{\boldsymbol{\beta}}_\varepsilon - \widehat{\boldsymbol{\beta}}}{\varepsilon}.$$

It is easy to see that the single-index estimator is equivariant under orthogonal transformations. Hence, without loss of generality, we can assume that $\widehat{\boldsymbol{\beta}} = \mathbf{e}_q$, the q-th canonical vector of $\mathbb{R}^q$. To obtain the empirical influence function, we will assume that the matrix $\mathbf{B}_1$ given in (16) is non-singular, as required when deriving the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ in Sect. 3. Furthermore, for simplicity, we will assume that the tuning parameter c and the bandwidth parameter h are fixed.

To avoid heavy notation, denote $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}) = \mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{x}_0))$, $\mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u)) = \mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u), (y_0, \mathbf{x}_0))$, $\mathrm{EIF}(\partial\, \widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial \boldsymbol{\beta}) = \mathrm{EIF}(\partial\, \widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial \boldsymbol{\beta}, (y_0, \mathbf{x}_0))$ and $\mathrm{EIF}(\partial\, \widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u) = \mathrm{EIF}(\partial\, \widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u, (y_0, \mathbf{x}_0))$. Moreover, from now on, given $\mathbf{b} \in \mathbb{R}^q$, $\mathbf{b}^{(q-1)} = (b_1, \ldots, b_{q-1})^{\mathrm{T}}$ stands for the vector of its first $q-1$ elements. Besides, given the kernel K and its first derivative K', we define $K'_h(u) = (1/h)\, K'(u/h)$.

Let us assume that ρ is three times continuously differentiable. Moreover, if in addition, the kernel K is continuously differentiable, we get that $\widehat{\eta}_{\boldsymbol{\beta}}(u)$ defined in **Step LG1** is twice continuously differentiable with respect to $\boldsymbol{\beta}$ and u. Moreover, its first partial derivatives can be computed as

$$\frac{\partial \widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial \boldsymbol{\beta}} = -\frac{\sum_{i=1}^n K'_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u\right) \psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right) \mathbf{x}_i}{\sum_{i=1}^n K_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u\right) \chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)} \quad \text{and}$$

$$\frac{\partial \widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u} = \frac{\sum_{i=1}^n K'_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u\right) \psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)}{\sum_{i=1}^n K_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u\right) \chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)},$$

where, as in (11) and (12), $\psi(y, a, c)$ and $\chi(y, a, c)$ stand for the derivatives with respect to a of $\phi(y, a, c)$ and $\psi(y, a, c)$, respectively.

Proposition 1 *Assume that ρ is three times continuously differentiable, the kernel K is continuously differentiable and that $\mathbf{B}_1$, the left superior matrix of dimension $(q-1) \times (q-1)$ of the matrix $\mathbf{B}$ given in (16) is non-singular. Denote as*

$$
\begin{aligned}
\boldsymbol{\ell}_n(y_0, \mathbf{x}_0) &= \psi\left(y_0, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0), c\right) \widehat{\boldsymbol{\nu}}_0(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i)\ \mathrm{EIF}\left(\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i) \left\{ \mathrm{EIF}\left(\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \right. \\
&\quad \left. + \mathrm{EIF}\left(\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i \right\}
\end{aligned} \tag{20}
$$

$$
\begin{aligned}
\mathbf{M}_n &= \left[\frac{1}{n}\sum_{i=1}^{n} \chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i)\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}} \right. \\
&\quad \left. + \ \frac{1}{n}\sum_{i=1}^{n} \psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right) \mathbf{V}(\widehat{\mathbf{s}}_i)\tau(\mathbf{x}_i)\right],
\end{aligned} \tag{21}
$$

where $\widehat{\mathbf{s}}_i = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)$ *and* $\widehat{\boldsymbol{\nu}}_i(\mathbf{b}, t)$ *are estimates of the quantities given in* (15)*, that is,*

$$
\begin{aligned}
\widehat{\boldsymbol{\nu}}_i(\mathbf{b}, t) &= \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)} + \frac{\partial}{\partial s}\widehat{\eta}_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)}\ \mathbf{x}_i \\
\widehat{\boldsymbol{\nu}}_0(\mathbf{b}, t) &= \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)} + \frac{\partial}{\partial s}\widehat{\eta}_{\boldsymbol{\beta}}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)}\ \mathbf{x}_0
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{V}(\widehat{\mathbf{s}}_i) &= \left[\ \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2\boldsymbol{\beta}}\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2 u}\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \right. \\
&\quad \left. + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u\partial\boldsymbol{\beta}}\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i^{\mathrm{T}} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}\partial u}\bigg|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i^{\mathrm{T}}\ \right].
\end{aligned}
$$

Then, if $\mathbf{M}_{n,1}$*, the left upper* $(q-1)\times(q-1)$ *submatrix of* $\mathbf{M}_n$*, is invertible, we have that*

a) $\mathrm{EIF}(\widehat{\boldsymbol{\beta}})_q = 0$ *and* $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}^{(q-1)}) = -\mathbf{M}_{n,1}^{-1}\boldsymbol{\ell}_n^{(q-1)}$*, where* $\boldsymbol{\ell}_n = \boldsymbol{\ell}_n(y_0, \mathbf{x}_0)$.
b) *the empirical influence functions at* $(y_0, \mathbf{x}_0)$*,* $\mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))$*,* $\mathrm{EIF}(\partial\,\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial\boldsymbol{\beta})$ *and* $\mathrm{EIF}(\partial\,\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u)$ *are given by*

$$
\begin{aligned}
\mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u)) &= -\frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)}{D_n} \\
\mathrm{EIF}\left(\frac{\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}}\right) &= -\frac{\frac{1}{h}K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\mathbf{x}_0 + K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\chi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)}{D_n}
\end{aligned}
$$

$$+\frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0-u)\psi\left(y_0,\widehat{\eta}_{\boldsymbol{\beta}}(u),c\right)}{D_n^2}\left[\frac{1}{h}\mathbf{g}_n+F_n\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right]$$

$$\begin{aligned}\mathrm{EIF}\left(\frac{\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u}\right)=\;&\frac{1}{D_n}\left\{\frac{1}{h}K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0-u)\psi\left(y_0,\widehat{\eta}_{\boldsymbol{\beta}}(u),c\right)\right.\\&\left.-K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0-u)\chi\left(y_0,\widehat{\eta}_{\boldsymbol{\beta}}(u),c\right)\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right\}\\&-\frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0-u)\psi\left(y_0,\widehat{\eta}_{\boldsymbol{\beta}}(u),c\right)}{D_n^2}\left(F_n\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)-\frac{1}{h}E_n\right),\end{aligned}$$

where

$$\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)=\frac{1}{h}F_n^{-1}E_n\,,\qquad \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)=-\frac{1}{h}F_n^{-1}\mathbf{g}_n$$

and

$$\begin{aligned}D_n&=\frac{1}{n}\sum_{i=1}^n K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i-u)\psi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}(u),c\right)\,,\\E_n&=\frac{1}{n}\sum_{i=1}^n K_h'\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i-u\right)\psi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}(u),\alpha\right)\,,\\F_n&=\frac{1}{n}\sum_{i=1}^n K_h\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i-u\right)\chi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}(u),\alpha\right)\,,\\\mathbf{g}_n&=\frac{1}{n}\sum_{i=1}^n K_h'\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i-u}{h}\right)\psi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}(u),\alpha\right)\mathbf{x}_i\,.\end{aligned}$$

Remark 3 In Proposition 1, the left upper submatrix $\mathbf{M}_{n,1}$ is assumed to be non-singular. Using the conditional Fisher consistency, that is, $\mathbb{E}_0\left(\psi\left(y,\eta_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}),c_0\right)|\mathbf{x}\right)=0$, we have that

$$\frac{1}{n}\sum_{i=1}^n\psi\left(y_i,\widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i),c\right)\mathbf{V}(\widehat{\mathbf{s}}_i)\tau(\mathbf{x}_i)\xrightarrow{p}0\,,$$

under mild conditions, while

$$\frac{1}{n}\sum_{i=1}^n\chi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i),c\right)\quad\tau(\mathbf{x}_i)\,\widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\xrightarrow{p}\mathbf{B}\,.$$

Therefore, $\mathbf{M}_n\xrightarrow{p}\mathbf{B}$ which implies that $\mathbf{M}_{n,1}\xrightarrow{p}\mathbf{B}_1$. Hence, taking into account that we have assumed that $\mathbf{B}_1$ is invertible, we get that with probability converging to 1 $\mathbf{M}_{n,1}$ is non-singular.

As mentioned above, the EIF may be used as a tool to assess the robustness of the method under consideration. To the greatest extent, we can say that if an estimator has an unbounded empirical influence function, we may expect the estimator to have a high sensitivity to the presence of outliers. On the contrary, a bounded EIF reflects that the effect of an anomalous point on the estimator is somehow down-weighted. In this sense, a robust method should give stable results when adding any possible point, even an extremely unlikely one. It should be taken into account that, in the present context of semiparametric methods, we are mainly concerned about the influence of large residuals due to an anomalous response (vertical outlier), since, as in nonparametric regression or partly linear models, to derive asymptotic distribution results it is assumed that the covariates related to the nonparametric component lie in a compact set.

It is important to recall that, as in nonparametric regression, when using a kernel with compact support to compute the smoother $\widehat{\eta}_{\boldsymbol{\beta}}$ only atypical responses near the value at which the link function estimator is evaluated may impact the nonparametric estimator. This well-known effect is clear from the dependence of $\text{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))$ on $K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)$. Consequently, the effect of an atypical response y_0 on $\widehat{\eta}_{\boldsymbol{\beta}}(u)$ takes into account not only the size of $\psi_{y_0,\boldsymbol{\beta},u} = \psi(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c)$, but also the distance between $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0$ and u. More precisely, a large value of $\psi_{y_0,\boldsymbol{\beta},u}$ will be more harmful when combined with values of $\mathbf{x}_0$ whose distance between $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0$ and u is small with respect to the considered bandwidth.

In the case of the estimator of the single-index parameter, we observe that the $\text{EIF}(\widehat{\boldsymbol{\beta}})$ depends on the arbitrary point $(y_0, \mathbf{x}_0)$ just through the factor $\boldsymbol{\ell}_n = \boldsymbol{\ell}_n(y_0, \mathbf{x}_0)$, since the matrix $\mathbf{M}_n$ only involves the original sample $(y_i, \mathbf{x}_i)$, $1 \leq i \leq n$. Hence, focusing on $\boldsymbol{\ell}_n$, we realize that its dependency on the discretionary response y_0 is carried out in the first term through $\psi_{y_0,\widehat{\boldsymbol{\beta}},u}$. In the remaining terms of $\boldsymbol{\ell}_n$, the effect of y_0 is related to the behaviour of the empirical influence functions of $\widehat{\eta}_{\boldsymbol{\beta}}(u)$ and its derivatives at $(\boldsymbol{\beta}, u) = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)$, which are larger when the projection $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0$ is close to that of one of the sample points $\mathbf{x}_i$, as described above. It is worth mentioning that the empirical influence functions $\text{EIF}\left(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial\boldsymbol{\beta}\right)$ and $\text{EIF}\left(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u\right)$ depend on y_0 not only through the score functions ψ but also through its partial derivative χ.

For the classical estimator, we have that $\Psi(t) = t$; therefore, the values of $\boldsymbol{\ell}_n(y_0, \mathbf{x}_0)$ and $\text{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))$ may go beyond any limit when y_0 goes to infinity. Furthermore, as mentioned above, the effect on the nonparametric component and its derivatives may be larger when the outlier y_0 corresponds to projection values $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0$ close to u. On the opposite, for our robust proposal, we choose Ψ as a bounded redescending score function leading to bounded ψ and χ functions, which enable to control the local effect of anomalous responses on the single-index estimator $\widehat{\boldsymbol{\beta}}$ and the link function one $\widehat{\eta}$.

It is worth noticing that even when considering a bounded loss function, such as the Tukey's bisquare function, the term in $\boldsymbol{\ell}_n(y_0, \mathbf{x}_0)$ involving $\mathbf{x}_0$ in $\widehat{\boldsymbol{\nu}}_0(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0)$ may not be bounded in directions orthogonal to $\widehat{\boldsymbol{\beta}}$, unless the function $\tau(\mathbf{x}_0)$ controls large values of the covariates. This behaviour is similar to that arising with projection-pursuit estimators when estimating the principal directions (see Croux and Ruiz-Gazen 2005), but unlike this case, in the present context of single-index models, our main concern is on the estimators sensitivity to anomalous responses and the values of $\mathbf{x}_0$ are assumed to be within a compact set.

5 Selection of the smoothing parameters

The estimation of the nonparametric component of the model involves a smoothing parameter both in the first and third steps. Each step may require a different degree of smoothness, and for this reason, the bandwidths may be different. The effect of the bandwidth is crucial on the performance of the nonparametric estimator; the smoothing parameter must warranty a balance between bias and variance. The problem of bandwidth selection has been widely studied in nonparametric and semiparametric models, and leave-one-out cross-validation procedures have been extensively used for this purpose. K-fold cross-validation criteria are also a reasonable choice with a computationally cheaper cost.

However, it is well known that classical cross-validation criteria are very sensitive to outliers. It is worth noticing that robust criteria for the selection of the smoothing parameter are needed even when robust estimators are considered. Leung et al. (1993), Wang and Scott (1994), Boente et al. (1997), Cantoni and Ronchetti (2001) and Leung (2005) discuss these ideas in the fully nonparametric framework, while Bianco and Boente (2007) and Boente and Rodriguez (2008) consider robust cross-validation in semiparametric models.

For the initial and final smoothing steps performed in **Steps 1** and **3** of the proposed method, we consider a robust version of the classical K-fold cross-validation criterion based on the deviance to select the bandwidths. More precisely, let us first randomly split the data set into K subsets of similar size, disjoint and exhaustive, with indexes $\mathcal{I}_j$, $1 \leq j \leq K$, such that $\cup_{j=1}^{K}\mathcal{I}_j = \{1, \ldots, n\}$. Let $\mathcal{H}_n^1 \subset \mathbb{R}$ be the set of bandwidths to be considered in the first step of the proposed procedure. Denote $\widetilde{\boldsymbol{\beta}}_h^{(-j)}$ the robust regression estimator computed in **Step 2** without the observations with indexes in the set $\mathcal{I}_j$ and using as smoothing parameter $h \in \mathcal{H}_n^1$ in the previous step and let $\widehat{\eta}^{(-j)}_{\widetilde{\boldsymbol{\beta}},h}(u)$ be the corresponding nonparametric robust estimator computed in **Step 1**.

Taking into account that for each i, $1 \leq i \leq n$, there exists j, $1 \leq j \leq K$, such that $i \in \mathcal{I}_j$, we define the prediction of observation y_i as $\widehat{y}_i = \widehat{\eta}^{(-j)}_{\widetilde{\boldsymbol{\beta}},h}(\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_h^{(-j)})$. Noticing that in the actual setting, the deviance residuals are a suitable measure of the discrepancy between an observation and its predictor; the robust K-th fold cross-validation smoothing parameter is defined as $\widehat{h}_1 = \operatorname{argmin}_{h \in \mathcal{H}_n^1} RCV(h)$, where

$$RCV(h) = \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d\,(y_i, \widehat{y}_i)}}{c}\right), \tag{22}$$

for a given tuning constant c. Clearly, the choice of the optimal bandwidth depends on the subset $\mathcal{H}_n^1$. It is worth noticing that for unit vectors $\boldsymbol{\beta}$, $|\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}| \leq \|\mathbf{x}_i\|$, so the projected data take values in the interval $[-\sup_i \|\mathbf{x}_i\|, \sup_i \|\mathbf{x}_i\|]$. In practice, to determine meaningful boundaries for the bandwidth set $\mathcal{H}_n^1$, one needs to have a more precise idea of the interspacing. For that purpose, the practitioner could generate M random unit directions $\boldsymbol{\beta}$ and explore the spacing between the values $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$ for each direction.

Denote $\widehat{\boldsymbol{\beta}}$ the robust estimator based on the whole sample when the smoothing parameter is the optimal $h = \widehat{h}_1$. It is worth noticing that the robust K-th fold cross-validation $RCV(h)$ given in (22) is a robustified version of its classical counterpart that seeks for the smoothing parameter minimizing

$$CCV(h) = \sum_{i=1}^{n} d\,(y_i, \widehat{y}_i), \tag{23}$$

where $\widehat{y}_i$ are based on the classical estimators.

In order to select the second bandwidth to be used in the local linear nonparametric estimator described in **Step 3**, we consider a similar procedure. That is, we take $\mathcal{H}_n^2 \subset \mathbb{R}$ the set of bandwidths to be considered in the third step. Denote $\widehat{\eta}^{(-j)}_{\widehat{\boldsymbol{\beta}},h}(u)$ the robust nonparametric estimator without the observations with indexes in the set $\mathcal{I}_j$ and using as smoothing parameter $h \in \mathcal{H}_n^2$ and $\widehat{\boldsymbol{\beta}}$. Again, reasoning as above, for each $1 \le i \le n$, we define the predictor of observation y_i as $\widehat{y}_i = \widehat{\eta}^{(-j)}_{\widehat{\boldsymbol{\beta}},h}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)$ and so, the robust K-th fold cross-validation linear smoothing parameter is defined as $\widehat{h}_2 = \mathrm{argmin}_{h\in\mathcal{H}_n^2}\, RCV(h)$. Once the data-driven bandwidth $\widehat{h}_2$ is obtained, the final nonparametric estimator denoted $\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{h}_2}$ can be computed as in **Step 3** from the whole sample using this bandwidth.

We want to highlight that the set of possible bandwidths in $\mathcal{H}_n^2$ depends now on the spacing between the projections $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i$. Since $\boldsymbol{\beta}_0$ has already been estimated through $\widehat{\boldsymbol{\beta}}$ at this stage, this determination results in a simpler process than for $\mathcal{H}_n^1$.

6 Numerical results

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 log-Gamma model.

We have performed $N = 1000$ replications with samples of size $n = 100$. For the clean samples, the covariates $\mathbf{x}_i$ are generated as $\mathbf{x}_i \sim \mathcal{U}((0, 1) \times (0, 1))$, while the response variables follow the log-Gamma single-index model $y_i = \eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) + \epsilon$, with $\eta_0(u) = \sin(2\pi u)$, $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$ and $\epsilon \sim \log(\Gamma(3, 1))$.

In all Tables, the results for the uncontaminated samples are denoted as C_0. Furthermore, the robust estimators introduced in this paper are subindicated with R, while their classical counterparts based on the deviance are subindicated with CL. To be more precise, the robust estimators correspond to those controlling large values of the deviance. In this case, the robust estimators were computed using the Tukey's bisquare loss function with adaptive tuning constants computed as in Bianco et al. (2005). It should be stressed that the tuning constant, which plays the role of the nuisance parameter in this setting, allows to measure the size of deviance residuals in order to down-weight the effect of possible influential observations. For that reason, as mentioned in the Introduction, its estimation is a key point to detect possible atypical data. More precisely, at each replication the calibration of the robust estimators was implemented through an S-estimator. For this purpose, we computed the initial procedure described in Steps ILG.1 to ILG.3 in Sect. 2.3. Since the resulting outcome of Step ILG.3 provides an estimator $\widehat{\gamma}$ of the shape parameter $\gamma_o = 3$, we calibrated the robust estimator through the function $S^\star$ defined in Step ILG.3 by taking as tuning constant $\widehat{c} = S^\star(\widehat{\gamma})$.

On the other hand, the classical estimators correspond to choosing the loss function equal to the deviance. With respect to the weight or trimming function, in order to make a fair comparison between the classical and robust estimators, we choose $\tau(\mathbf{x}) = \|\mathbf{x} - \mathbf{c}\|\mathbb{I}_{[0,b_n]}$, with $\mathbf{c} = (0.5, 0.5)$ and $b_n = 0.4\sqrt{\log(\log(n))}$ for both estimators. The value b_n is selected as in Sherman (1994) to avoid the density of $\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}$ to be too small.

The smoothing parameters were selected as described in Sect. 5 using a 5-fold cross-validation procedure. For the classical estimators, we use criterion (23) in each step, while for the robust estimates, we used the robust 5-fold method (22) with $c = 1.6394$ that under the central model corresponds to an asymptotic efficiency of 0.90. In all these cases, the set $\mathcal{H}_n^1$ of candidates for the initial bandwidth h was taken as an equidistant grid of length 13 between 0.05 and 0.35, while for the local linear smoothing parameter, we choose $\mathcal{H}_n^2$ as an equidistant grid of length 25 between 0.05 and 0.35. To simplify the notation, henceforth we denote $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\widehat{\eta}_{\mathrm{R}}$ the robust estimators computed with the two robust cross-validation bandwidths, while $\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}$ and $\widehat{\eta}_{\mathrm{CL}}$ stand for the classical estimators computed with the bandwidths obtained minimizing (23).

To evaluate the performance of each estimator, we compute different measures. For the parametric component, given an estimator $\widehat{\boldsymbol{\beta}}$ of the true single-index parameter $\boldsymbol{\beta}_0$, we consider $\mathrm{MSE}_{\widehat{\boldsymbol{\beta}}}$ as the mean values over replications of $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2$. For the nonparametric component, we compute $\mathrm{MSE}_{\widehat{\eta}}$ as the mean over replications of $(1/n)\sum_{i=1}^n(\eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) - \widehat{\eta}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i))^2$ and also $\mathrm{MedSE}_{\widehat{\eta}}$ as the median over replications of $\mathrm{median}_{i=1:n}(\eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) - \widehat{\eta}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i))^2$, where $\widehat{\eta}$ is a given estimator of the function η_0.

In order to assess the behaviour of the estimators under contamination, we have considered two types of contaminations and samples $(y_{i,c}, \mathbf{x}_{i,c})$ generated from them. The first set of contaminations introduces moderate outlying points, while with the second one we expect a more dramatic effect on the classical estimators.

Three different models, labelled M_1, M_2 and M_3 in all Tables and Figures, are considered in the moderate contamination scheme. To obtain the contaminated samples, we have first generated a sample $u_i \sim \mathcal{U}(0, 1)$ for $1 \leq i \leq n$, and then, we introduce large values on the responses as

$$y_{i,c} = \begin{cases} y_i & \text{if } u_i \leq 0.90 \\ y_i^\star & \text{if } u_i > 0.90, \end{cases} \tag{24}$$

where $y_i^\star = \log(k) + \widetilde{\eta}(\mathbf{x}_i) + \epsilon_i$, with $\epsilon_i \sim \log(\Gamma(3, 1))$, $\widetilde{\eta}(\mathbf{x}_i) = \eta(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0^{\perp})$ where $\boldsymbol{\beta}_0^{\perp}$ is the unit vector orthogonal to the true single-index parameter $\boldsymbol{\beta}_0$ and $k = 3, 4$ and 5 under M_1, M_2 and M_3, respectively.

The second scheme accounts for more severe contaminations, labelled S_1, S_2 and S_3 in all Tables and Figures, and we expect that its effect on the classical estimators would be more dramatic. To obtain the contaminated samples, the observations are

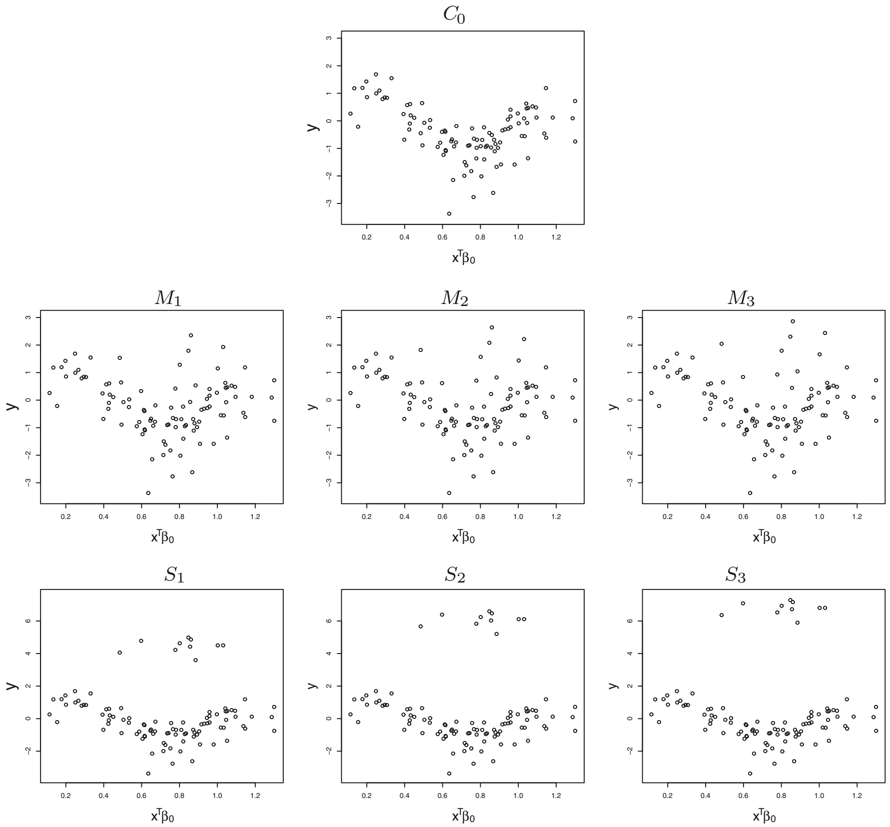

Fig. 1 Generated sample when $\eta_0(u) = \sin(2\pi u)$ and $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$ under the central and contaminated models

generated as in (24) where now $y_i^\star = \log(k) + \epsilon_i$ where as above $\epsilon_i \sim \log(\Gamma(3, 1))$ but $k = 100$, 500 or 1000, respectively. Figure 1 illustrates the considered contaminations in a generated sample.

Table 1 summarizes the results along the $N = 1000$ replications. The reported results show the great stability of the robust procedure against moderate and severe contaminations. As expected, when there is no contamination the classical estimators achieve the lowest square errors for both the parametric and nonparametric components. Nevertheless, the performance of the robust estimators is very satisfactory under C_0 since the loss of efficiency is very small. Focusing on the parametric component, under any of the contaminated schemes, the performance of the classical estimator is very poor. Table 1 exhibits that the mean square error of the single-index parameter increases more than forty times under the moderate contaminations and more than 200 times under the severe ones, while the robust estimator remains very stable in all considered scenarios.

Since an important goal in this framework is to capture the direction of the single-index parameter $\boldsymbol{\beta}_0$, instead of presenting the traditional boxplots of the estimates, in

Table 1 Mean square errors for the estimators of $\boldsymbol{\beta}_0$, Mean over replications of $(1/n)\sum_{i=1}^{n}(\eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) - \widehat{\eta}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i))^2$ and Median over replications of $\mathrm{median}_{i=1:n}(\eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{x}_i) - \widehat{\eta}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i))^2$ denoted $MSE(\widehat{\eta})$ and $\mathrm{MedSE}_{\widehat{\eta}}$, respectively

	$\mathrm{MSE}_{\widehat{\boldsymbol{\beta}}}$		$\mathrm{MSE}_{\widehat{\eta}}$		$\mathrm{MedSE}_{\widehat{\eta}}$	
	$\widehat{\boldsymbol{\beta}}_{\mathrm{CL}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$	$\widehat{\eta}_{\mathrm{CL}}$	$\widehat{\eta}_{\mathrm{R}}$	$\widehat{\eta}_{\mathrm{CL}}$	$\widehat{\eta}_{\mathrm{R}}$
C_0	0.005	0.005	0.041	0.046	0.020	0.021
M_1	0.209	0.008	0.294	0.061	0.164	0.031
M_2	0.357	0.007	0.408	0.060	0.226	0.030
M_3	0.534	0.007	0.521	0.058	0.287	0.028
S_1	1.064	0.013	5.393	0.059	4.510	0.024
S_2	1.098	0.008	13.282	0.057	13.282	0.023
S_3	1.106	0.006	18.057	0.053	17.436	0.022

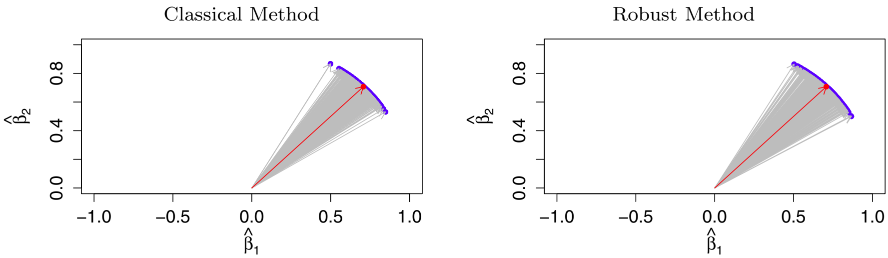

Fig. 2 Classical and robust estimators of the single-index parameter under C_0. The red arrow represents the true direction $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$, while the grey arrows are the estimates

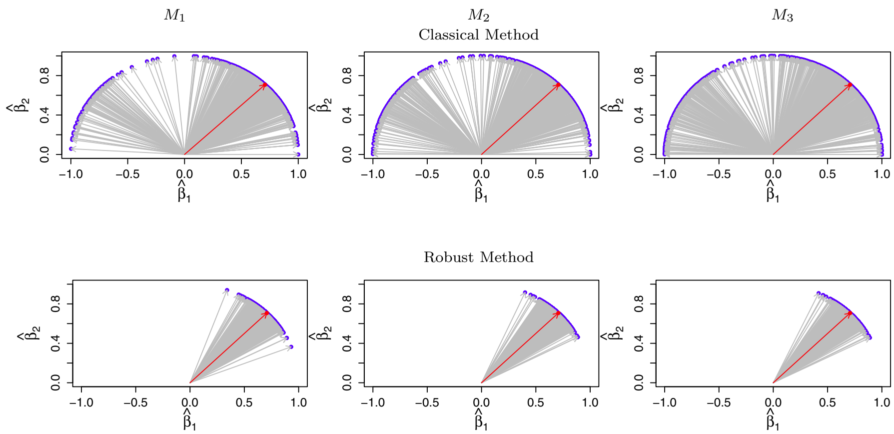

Fig. 3 Classical and robust estimators of the single-index parameter under M_1, M_2 and M_3. The red arrow represents the true direction $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$, while the grey arrows are the estimates

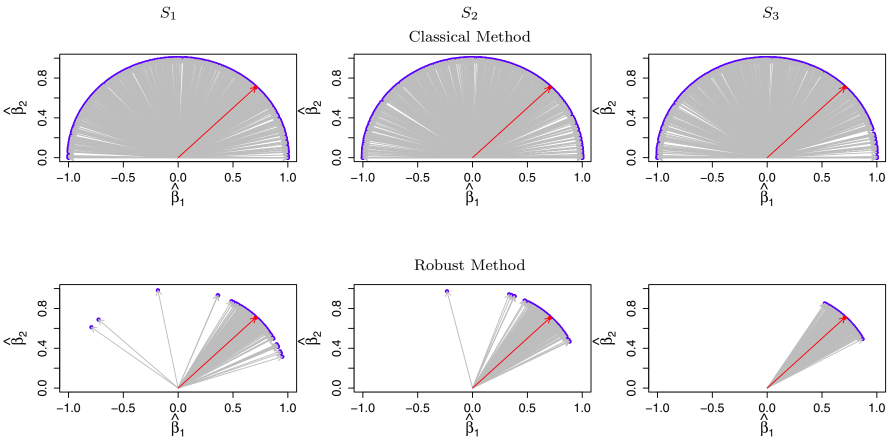

Fig. 4 Classical and robust estimators of the single-index parameter under S_1, S_2 and S_3. The red arrow represents the true direction $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$, while the grey arrows are the estimates

Figs. 2, 3 and 4 we present a two-dimensional graph that reflects the skill of the classical and robust estimators to get the true direction $\boldsymbol{\beta}_0$, for the clean and contaminated samples. In these plots, the red arrow represents the true direction $\boldsymbol{\beta}_0 = (1/\sqrt{2}, 1/\sqrt{2})^{\mathrm{T}}$, that corresponds to an angle $\theta_0 = \pi/4$ and the grey ones to the estimated directions. These figures show that under C_0 the performance of the robust estimator of the parametric component is similar to that of the classical estimator since the robust estimates are more or less spread as the classical estimator around the target direction. It also becomes evident that in contaminated samples the robust estimator of the parametric component is very stable under all the contaminated scenarios, while the classical estimator is completely spoiled. Indeed, under M_1 to M_3, the classical estimates of the single-index parameter tend to be concentrated not only on directions close to the true value $\boldsymbol{\beta}_0$ but also to its orthogonal direction $\boldsymbol{\beta}_0^{\perp}$, showing the impact of the contaminated points. On the other hand, under the severe contaminations S_1 to S_3 the classical estimates cover almost all possible directions in the first and second quadrants, becoming completely unreliable.

Regarding the estimation of the nonparametric component, Table 1 shows the large effect of the considered contaminations on the classical estimator of the nonparametric component, where the mean square error increases at least seven times under the moderate contaminations. Under the severe contaminations S_1 to S_3, the effect of the outliers on the classical estimator is devastating, while it is quite harmless for the robust estimator. It is worth noticing that under all the contamination schemes, the reported values of MedSE_η for the classical estimator, which is a more resistant measure based on the median, are very close to the corresponding values of MSE_η, making evident that in most replications the classical estimator of the nonparametric component is completely spoiled.

In order to give a full picture of the performance of both classical and robust estimators of η_0, Figs. 5, 6 and 7 display their functional boxplots which are a very useful graphical tool introduced by Sun and Genton (2011). Since the covariate $\mathbf{x}$ is

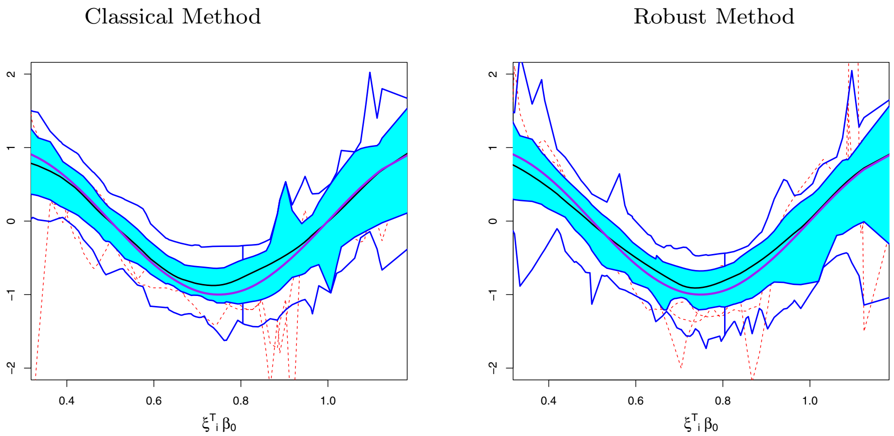

Fig. 5 Classical and robust estimators of η_0 under C_0

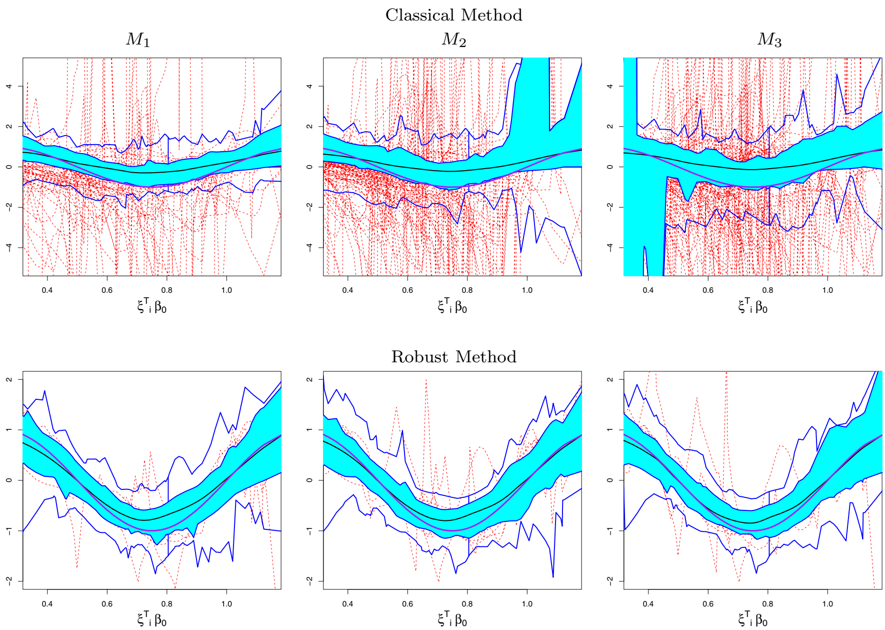

Fig. 6 Classical and robust estimators of η_0 under M_1, M_2 and M_3

random, in order to obtain comparable estimations for η_0, we consider a fixed grid of points $\boldsymbol{\xi}_j$, $j = 1, \ldots, 100$ in $[0, 1] \times [0, 1]$. Thus, for each replication, we estimate $\eta_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\xi}_j)$ using the classical and robust procedures. In the functional boxplots, the area in light blue represents the central region, the dotted red lines correspond to outlying curves, the black line indicates the deepest curve, while the purple line is the true nonparametric function η_0. It is worth noticing that, for the contaminated settings, due to the effect of the outliers, some curves are out of range when using the

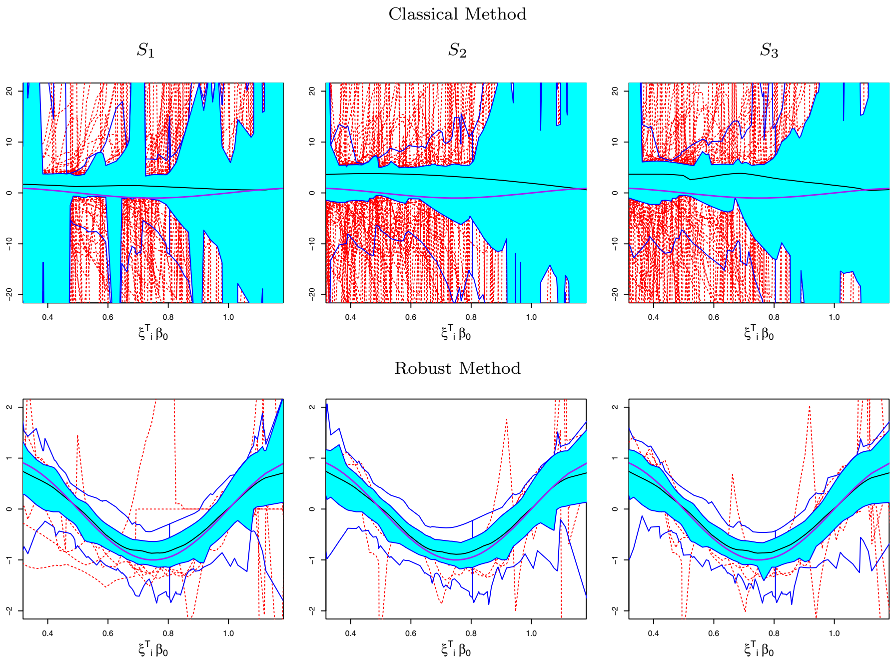

Fig. 7 Classical and robust estimators of η_0 under S_1, S_2 and S_3

classical procedure. For that reason, the functional boxplots of the classical estimators are plotted in a reduced range to allow a clear visualization of the central area. Figure 5 shows that the classical and robust nonparametric estimators of η_0 are quite similar under C_0, while Figs. 6, 7 exhibit the devastating effect of the contaminating points, even the moderate ones, on the classical estimator. The impact of the contaminations on the classical estimates is reflected either in the presence of a great number of outlying curves and also in the enlargement of the width of the bars of the functional boxplots. With respect to the robust estimates, despite the fact that a few outlying curves appear, the range of variation of the curves is almost the same than under C_0, the central region in light blue of all the boxplots always contains the true function η_0 and most curves follow the pattern introduced by the sine function. In general terms, the functional boxplots show the stability of the robust estimates of η_0 which are reliable under the contaminated scenarios as well as the strong effect of the considered contaminations on the classical estimators of the nonparametric component.

A careful study of the bandwidth behaviour is beyond the scope of the paper; however in order to have a deeper insight of the performance of the selectors under C_0 and the considered contaminations, we give a brief analysis of the data-driven parameters obtained in this numerical experiment. Table 2 reports, for both estimators, the median over replications of the cross-validation data-driven bandwidths to be used in Steps 1 and 3 denoted $\widehat{h}_1$ and $\widehat{h}_2$, respectively. On the other hand, in Figs. 8, 9 and 10 we present the bagplots corresponding to $\widehat{h}_1$ and $\widehat{h}_2$ selected through the classical and robust cross-validation criteria. Under C_0, both criteria lead to similar data-driven

Table 2 Median over replications of cross-validation data-driven bandwidths to be used in Steps 1 and 3 denoted $\widehat{h}_1$ and $\widehat{h}_2$, respectively

Method		C_0	M_1	M_2	M_3	S_1	S_2	S_3
Classical	$\widehat{h}_1$	0.175	0.225	0.250	0.250	0.325	0.325	0.325
	$\widehat{h}_2$	0.175	0.238	0.250	0.250	0.350	0.300	0.338
Robust	$\widehat{h}_1$	0.175	0.175	0.175	0.175	0.175	0.200	0.200
	$\widehat{h}_2$	0.188	0.188	0.188	0.188	0.188	0.188	0.188

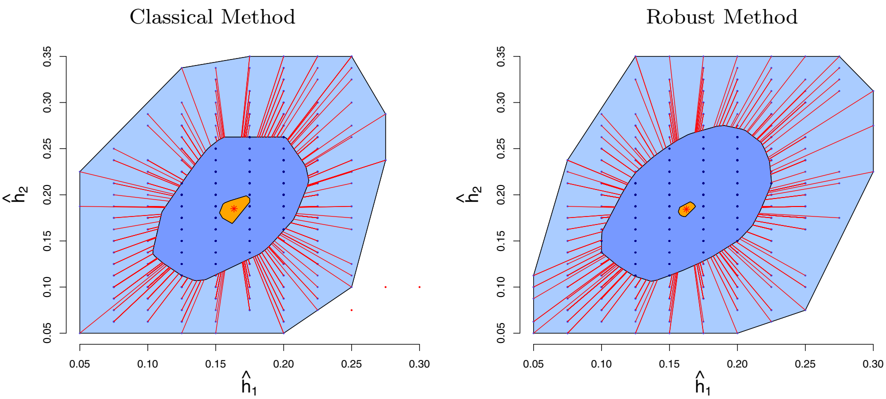

Fig. 8 Bagplots for $(\widehat{h}_1, \widehat{h}_2)$ chosen according the classical and robust cross-validation criteria under C_0

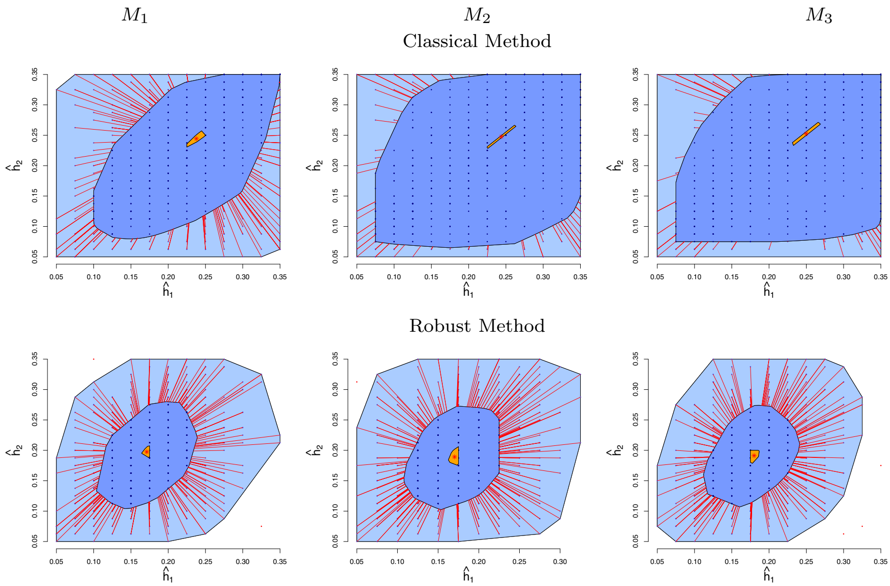

Fig. 9 Bagplots for $(\widehat{h}_1, \widehat{h}_2)$ chosen according the classical and robust cross-validation criteria under M_1, M_2 and M_3

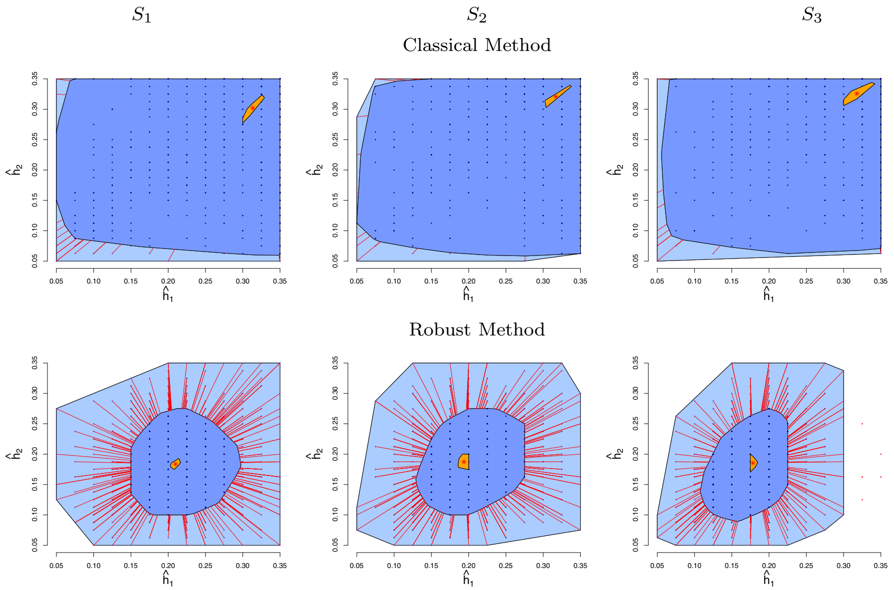

Fig. 10 Bagplots for $(\widehat{h}_1, \widehat{h}_2)$ chosen according the classical and robust cross-validation criteria under S_1, S_2 and S_3

smoothing parameters. However, the lack of robustness of the classical cross-validation criterion under contaminations becomes evident from these plots. The classical cross-validation criterion under contaminations tends to choose greater bandwidths, and this becomes evident, for instance, from the behaviour of their medians reported in Table 2. The poor behaviour of the classical data-driven bandwidths leads towards over-smoothing which may explain the results reported in Table 1. On the other hand, except for a few cases, the selected bandwidths obtained with the robust criterion remain stable in all circumstances.

7 Hospital Costs Data: an exploratory analysis

Marazzi and Yohai (2004) consider a data set related to the costs of patients that suffer from back problems (APDRG 243) in the Centre Hospitalier Universitaire Vaudois in Lausanne, Switzerland, in 1999. The data correspond to 100 patients, and among other binary variables, they contain information about the cost of stay, z, (cost, in Swiss francs) and the following two numerical variables: length of stay in days and age. Our goal is to perform an exploratory analysis to study the nature of the relationship between the cost of stay and these two numerical variables. Cantoni and Ronchetti (2006) fitted the complete data set under a conditional Gamma model for the responses by assuming a log link function. Taking into account that they encountered several outliers and the sensitivity of the classical approach that became evident from our simulations, it is sensible to analyse the data by means of a robust procedure. In our analysis, we choose as covariates x_1 and x_2 the standardized variables length of stay

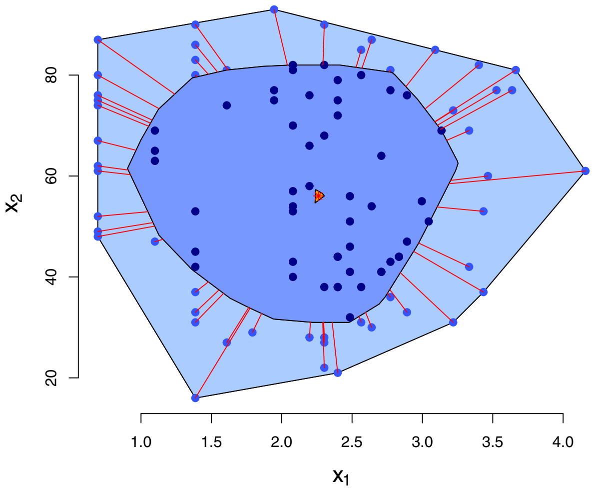

Fig. 11 Hospital Cost Data: bagplot for (x_1, x_2)

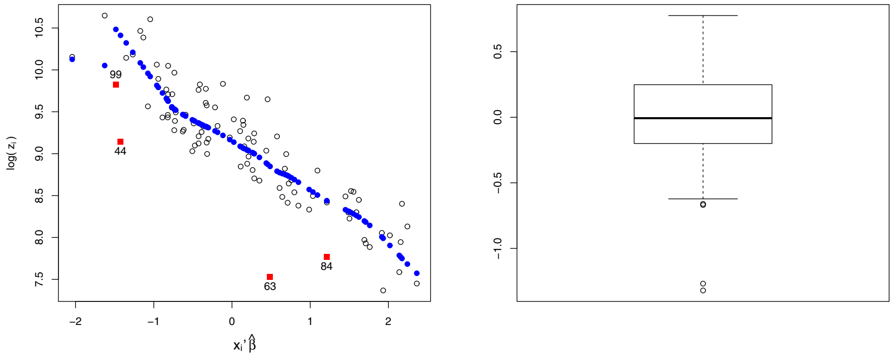

Fig. 12 Hospital Cost Data: On the left panel, the scatter plot of $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i$ vs. $y_i = log(z_i)$ in black and in blue filled circles the fitted values $\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{h}_2}(\mathbf{x}_i)$. Red squares correspond to identified outliers. On the right panel, adjusted boxplot of the residuals

in days and age, since this standardization does not affect the shape of the function g. Both variables are centred with their median and scaled using their MAD.

In a first step, we examine the data graphically. Figure 11 gives the bagplot corresponding to the two covariates and shows that in the two-dimensional plane (x_1 vs. x_2) there are no isolated points.

In order to inquire about the relationship between the response variable, *cost of stay*, and the two explanatory variables of interest, *length of stay* and *age*, we propose the model

$$y_i = \log(z_i) = \eta(\beta_1 x_{1i} + \beta_2 x_{2i}) + \epsilon_i$$

where $\epsilon_i \sim \log(\Gamma(\gamma_0, 1))$ and γ_0 is unknown. We consider the same robust stepwise estimators as in our numerical example, and the smoothing parameters were chosen as described in Sect. 5 using a 5-fold cross-validation procedure. The resulting bandwidths parameters are $h_1 = 0.6$ and $h_2 = 0.975$. The left panel of Fig. 12 corresponds to the scatter plot of $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i$ versus $y_i = \log(z_i)$ (in black). The fitted values according to $\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{h}_2}$, in the notation of Sect. 5, are drawn as blue filled circles. Besides, in the left panel, we show the skewed-adjusted boxplot (see Hubert and Vandervieren 2008) of the residuals $\log(z_i) - \widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{h}_2}(\mathbf{x}_i)$, where four outliers are identified. These atypical residuals correspond to observations 44 , 63, 84 and 99, which are highlighted as red filled squares in the scatter plot of the left panel. The points labelled 44 and 63, suspicious from being outliers in our analysis, were also detected by Cantoni and Ronchetti (2006). Moreover, the fitted values exhibit a curvature which suggests that a nonlinear trend is still present between the responses and the explanatory variables.

Acknowledgements The authors wish to thank an anonymous referee for valuable comments which led to an improved version of the original paper. This research was partially supported by Grants PICT 2014-0351 from ANPCYT, Grants 20120130100279BA and 20020170100022BA from the Universidad de Buenos Aires at Buenos Aires, Argentina and also by the Spanish Project MTM2016-76969P from the Ministry of Science and Innovation, Spain. It was also supported by the Italian–Argentinian project *Metodi robusti per la previsione del costo e della durata della degenza ospedaliera* funded by the joint collaboration program MINCYT-MAE AR14MO6 (IT1306) between MINCYT from Argentina and MAE from Italy.

A Appendix

A.1 Proof of Theorem 1

a) For any $\varepsilon > 0$, let $\mathcal{X}_0$ be a compact set such that $P(\mathbf{x} \notin \mathcal{X}_0) < \varepsilon$. Then, we have that

$$\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{S}_1;a\in\mathcal{K}} \left|\Delta_n(\boldsymbol{\beta}, \widehat{\eta}_{\mathbf{b},a}, a) - \Delta_n(\boldsymbol{\beta}, \eta_{\mathbf{b},a}, a)\right|$$

$$\leq \sup_{\mathbf{b}\in\mathcal{S}_1,a\in\mathcal{K}} \|\widehat{\eta}_{\mathbf{b},a} - \eta_{\mathbf{b},a}\|_{0,\infty}\|\tau\|_\infty\|\phi'\|_\infty + 2\|\phi\|_\infty\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{(\mathbf{x}_i\notin\mathcal{X}_0)}\tau(\mathbf{x}_i)$$

and so, using (10), the fact that $P(\mathbf{x} \notin \mathcal{X}_0) < \varepsilon$ and the strong law of large numbers, we get that

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

Therefore, it remains to show that $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{S}_1;a\in\mathcal{K}} \left|\Delta_n(\boldsymbol{\beta}, \eta_{\mathbf{b},a}, a) - \Delta(\boldsymbol{\beta}, \eta_{\mathbf{b},a}, a)\right| \xrightarrow{a.s.} 0.$

Define the following class of functions $\mathcal{H} = \{f_{\boldsymbol{\beta}}(y, \mathbf{x}) = \phi(y, \eta_{\mathbf{b},a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), a)\tau(\mathbf{x})\,,\ \boldsymbol{\beta}, \mathbf{b} \in \mathcal{S}_1, a \in \mathcal{K}\}$. Using Theorem 3 from Chapter 2 in Pollard (1984), the compactness of $\mathcal{K}$, **A1**, the continuity of $\eta_{\boldsymbol{\beta},\alpha}(u)$ given in **A6** and analogous arguments to those considered in Lemma 1 from Bianco and Boente (2002), we get that $\sup_{\boldsymbol{\beta},\mathbf{b}\in\mathcal{S}_1;a\in\mathcal{K}} \left|\Delta_n(\boldsymbol{\beta},\widehat{\eta}_{\mathbf{b},a},a)-\Delta(\boldsymbol{\beta},\eta_{\mathbf{b},a},a)\right| \xrightarrow{a.s.} 0$ and a) follows.

b) Let $\widehat{\boldsymbol{\beta}}_k$ be a subsequence of $\widehat{\boldsymbol{\beta}}$ such that $\widehat{\boldsymbol{\beta}}_k \to \boldsymbol{\beta}^*$, where $\boldsymbol{\beta}^*$ lies in the compact set $\mathcal{S}_1$. Let us assume, without loss of generality, that $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}^*$. Then, **A7**, the continuity of $\eta_{\boldsymbol{\beta},\alpha}$, the consistency of $\widehat{\alpha}_{\mathrm{R}}$ and a) entail that $\Delta_n(\widehat{\boldsymbol{\beta}},\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}_{\mathrm{R}}},\widehat{\alpha}_{\mathrm{R}})-\Delta(\boldsymbol{\beta}^*,\eta_0,\alpha_0)\xrightarrow{a.s.} 0$ and $\Delta_n(\boldsymbol{\beta}_0,\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}_{\mathrm{R}}},\widehat{\alpha}_{\mathrm{R}})-\Delta(\boldsymbol{\beta}_0,\eta_0,\alpha_0)\xrightarrow{a.s.} 0$, since $\eta_{\boldsymbol{\beta}_0,\alpha_0}=\eta_0$. Now, using that $\Delta_n(\boldsymbol{\beta}_0,\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}_{\mathrm{R}}},\widehat{\alpha}_{\mathrm{R}})\geq\Delta_n(\widehat{\boldsymbol{\beta}},\widehat{\eta}_{\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\alpha}}_{\mathrm{R}}},\widehat{\alpha}_{\mathrm{R}})$ and $\Delta(\boldsymbol{\beta},\eta_0,\alpha_0)$ has a unique minimum at $\boldsymbol{\beta}_0$, we conclude the proof. □

A.2 Proof of Proposition 1

a) The single-index parameter estimation related to **Step LG2** is obtained by means of the minimization with respect to $\boldsymbol{\beta}$ of

$$\sum_{i=1}^{n}\rho\left(\frac{\sqrt{d\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)\right)}}{c}\right)\tau(\mathbf{x}_i),$$

among the vectors of length one, where, at the same time, $\widehat{\eta}_{\boldsymbol{\beta}}(u)$ is defined as

$$\widehat{\eta}_{\boldsymbol{\beta}}(u)=\underset{a\in\mathbb{R}}{\operatorname{argmin}}\sum_{i=1}^{n}\rho\left(\frac{\sqrt{d(y_i,a)}}{c}\right)W_h(u,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i).$$

Hence, if we denote $\mathcal{B}(\boldsymbol{\theta})=\boldsymbol{\theta}/\|\boldsymbol{\theta}\|$, we have that $\widehat{\boldsymbol{\beta}}_\varepsilon=\widehat{\boldsymbol{\theta}}_\varepsilon/\|\widehat{\boldsymbol{\theta}}_\varepsilon\|=\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)$ where $\widehat{\boldsymbol{\theta}}_\varepsilon$ is the solution of

$$\begin{aligned}\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ &\frac{1-\varepsilon}{n}\sum_{i=1}^{n}\rho\left(\frac{\sqrt{d\left(y_i,\widehat{\eta}^{\varepsilon}_{\mathcal{B}(\boldsymbol{\theta})}\left(\mathcal{B}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{x}_i\right)\right)}}{c}\right)\tau(\mathbf{x}_i)\\ &+\varepsilon\,\rho\left(\frac{\sqrt{d\left(y_0,\widehat{\eta}^{\varepsilon}_{\mathcal{B}(\boldsymbol{\theta})}\left(\mathcal{B}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{x}_0\right)\right)}}{c}\right)\tau(\mathbf{x}_0).\end{aligned}$$

Then, $\widehat{\boldsymbol{\theta}}_\varepsilon$ satisfies

$$\begin{aligned}\mathbf{0}=\left(\mathbf{I}-\mathcal{B}\left(\widehat{\boldsymbol{\theta}}_\varepsilon\right)\mathcal{B}\left(\widehat{\boldsymbol{\theta}}_\varepsilon\right)^{\mathrm{T}}\right)&\left[\frac{(1-\varepsilon)}{n}\sum_{i=1}^{n}\psi\left(y_i,\widehat{\eta}^{\varepsilon}_{\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)}(\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)^{\mathrm{T}}\mathbf{x}_i),c\right)\widehat{\boldsymbol{\nu}}_i^{\epsilon}\left(\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon),\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)\mathbf{x}_i\right)\tau(\mathbf{x}_i)\right.\\ &\left.+\,\varepsilon\,\psi\left(y_0,\widehat{\eta}^{\varepsilon}_{\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)}(\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)^{\mathrm{T}}\mathbf{x}_0),c\right)\widehat{\boldsymbol{\nu}}_0^{\epsilon}\left(\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon),\mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)\mathbf{x}_0\right)\tau(\mathbf{x}_0)\right],\end{aligned}$$

where

$$\psi(y,a,c) = \frac{\partial}{\partial a}\phi(y,a,c) = \frac{1}{2c}\Psi\left(\frac{\sqrt{d(y,a)}}{c}\right)\frac{1-\exp(y-a)}{\sqrt{d(y,a)}}$$

as defined in (11), Ψ stands for the derivative of ρ and $\widehat{\boldsymbol{\nu}}_i^{\epsilon}(\mathbf{b},t)$ are given by

$$\widehat{\boldsymbol{\nu}}_i^{\epsilon}(\mathbf{b},t) = \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}^{\epsilon}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)} + \frac{\partial}{\partial s}\widehat{\eta}_{\boldsymbol{\beta}}^{\epsilon}(s)|_{(\boldsymbol{\beta},s)=(\mathbf{b},t)}\ \mathbf{x}_i\ .$$

Using that $\widehat{\boldsymbol{\beta}}_\varepsilon = \mathcal{B}(\widehat{\boldsymbol{\theta}}_\varepsilon)$, we get that the estimator $\widehat{\boldsymbol{\beta}}_\varepsilon$ verifies

$$\begin{aligned}\mathbf{0} = \left(\mathbf{I}-\widehat{\boldsymbol{\beta}}_\varepsilon\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\right)\Bigg[\frac{(1-\varepsilon)}{n}\sum_{i=1}^{n}\psi\left(y_i,\widehat{\eta}_{\widehat{\boldsymbol{\beta}}_\varepsilon}^{\varepsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i),c\right)\widehat{\boldsymbol{\nu}}_i^{\epsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon,\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i)\tau(\mathbf{x}_i)\\ +\varepsilon\ \psi\left(y_0,\widehat{\eta}_{\widehat{\boldsymbol{\beta}}_\varepsilon}^{\varepsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_0),c\right)\widehat{\boldsymbol{\nu}}_0^{\epsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon,\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0)\Bigg]\end{aligned}$$

and $\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u)$ is the solution of

$$\frac{(1-\varepsilon)}{n}\sum_{i=1}^{n}\psi\left(y_i,\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u),c\right)W_h(u,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)+\varepsilon\ \psi\left(y_0,\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u),c\right)W_h(u,\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0)=0\ . \tag{A.1}$$

Then, if we call

$$\begin{aligned}\boldsymbol{\lambda}(\varepsilon) = \frac{(1-\varepsilon)}{n}\sum_{i=1}^{n}\psi\left(y_i,\widehat{\eta}_{\widehat{\boldsymbol{\beta}}_\varepsilon}^{\varepsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i),c\right)\widehat{\boldsymbol{\nu}}_i^{\epsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon,\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i)\tau(\mathbf{x}_i)\\ +\varepsilon\ \psi\left(y_0,\widehat{\eta}_{\widehat{\boldsymbol{\beta}}_\varepsilon}^{\varepsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_0),c\right)\widehat{\boldsymbol{\nu}}_0^{\epsilon}(\widehat{\boldsymbol{\beta}}_\varepsilon,\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0)\end{aligned}$$

we get that, for any $0 \le \epsilon < 1$, $\widehat{\boldsymbol{\beta}}_\varepsilon$ satisfies $\mathbf{0} = \left(\mathbf{I}-\widehat{\boldsymbol{\beta}}_\varepsilon\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\right)\ \boldsymbol{\lambda}(\varepsilon)$. Therefore, differentiating with respect to ε and evaluating at $\varepsilon = 0$ and using that $\boldsymbol{\lambda}(0) = \mathbf{0}$, we obtain that

$$\begin{aligned}\mathbf{0} &= \frac{\partial}{\partial\varepsilon}\left[\left(\mathbf{I}-\widehat{\boldsymbol{\beta}}_\varepsilon\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\right)\boldsymbol{\lambda}(\varepsilon)\right]\Big|_{\varepsilon=0} = \frac{\partial}{\partial\varepsilon}\left[\left(\mathbf{I}-\widehat{\boldsymbol{\beta}}_\varepsilon\widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\right)\right]\Big|_{\varepsilon=0}\boldsymbol{\lambda}(0)+\left(\mathbf{I}-\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right)\frac{\partial}{\partial\varepsilon}\ \boldsymbol{\lambda}(\varepsilon)|_{\varepsilon=0}\\ &= \left(\mathbf{I}-\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right)\frac{\partial}{\partial\varepsilon}\ \boldsymbol{\lambda}(\varepsilon)|_{\varepsilon=0}\ .\end{aligned} \tag{A.2}$$

Henceforth, in order to compute $(\partial\boldsymbol{\lambda}(\varepsilon)/\partial\varepsilon)|_{\varepsilon=0}$ and to simplify the presentation, we consider the following functions:

$$h(\varepsilon,\boldsymbol{\beta},u) = \widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u)\ ,\qquad h_{\boldsymbol{\beta}}(\varepsilon,\boldsymbol{\beta},u) = \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u)\ ,\qquad h_u(\varepsilon,\boldsymbol{\beta},u) = \frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u)$$

and their corresponding derivatives with respect to ε

$$H_i = \frac{\partial}{\partial \varepsilon} h(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}} \mathbf{x}_i)\Big|_{\varepsilon=0}, \qquad H_{\boldsymbol{\beta},i} = \frac{\partial}{\partial \varepsilon} h_{\boldsymbol{\beta}}(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}} \mathbf{x}_i)\Big|_{\varepsilon=0},$$
$$H_{u,i} = \frac{\partial}{\partial \varepsilon} h_u(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}} \mathbf{x}_i)\Big|_{\varepsilon=0}.$$

Thus, we have that

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \boldsymbol{\lambda}(\varepsilon)\Big|_{\varepsilon=0} &= -\frac{1}{n}\sum_{i=1}^{n} \psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right) \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\tau(\mathbf{x}_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n} \Big\{\chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ H_i\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i) \\
&+ \psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ (H_{\boldsymbol{\beta},i} + \mathbf{x}_i H_{u,i})\Big\}\ \tau(\mathbf{x}_i) \\
&+ \psi\left(y_0, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0), c\right) \widehat{\boldsymbol{\nu}}_0(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0).
\end{aligned}$$

Since $\boldsymbol{\lambda}(0) = \mathbf{0}$, we obtain that

$$\begin{aligned}
\frac{\partial}{\partial \varepsilon} \boldsymbol{\lambda}(\varepsilon)\Big|_{\varepsilon=0} &= \frac{1}{n}\sum_{i=1}^{n} \Big\{\chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ H_i\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i) \\
&+ \psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ (H_{\boldsymbol{\beta},i} + \mathbf{x}_i H_{u,i})\Big\}\ \tau(\mathbf{x}_i) \\
&+ \psi\left(y_0, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0), c\right) \widehat{\boldsymbol{\nu}}_0(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0).
\end{aligned} \tag{A.3}$$

It remains to compute the functions H_i, $H_{\boldsymbol{\beta},i}$ and $H_{u,i}$. Straightforward arguments lead to

$$\begin{aligned}
H_i &= \frac{\partial}{\partial \varepsilon} h(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i)\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial \varepsilon} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} + \frac{\partial}{\partial \boldsymbol{\beta}} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial \varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \\
&\quad + \frac{\partial}{\partial u} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial \varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \mathbf{x}_i\ ,
\end{aligned}$$

where $\widehat{\mathbf{s}}_i = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)$. Then, we get that

$$\begin{aligned}
H_i &= \mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \frac{\partial \widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial \boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathrm{EIF}(\widehat{\boldsymbol{\beta}}) + \frac{\partial \widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathrm{EIF}(\widehat{\boldsymbol{\beta}})\mathbf{x}_i \\
&= \mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \widehat{\boldsymbol{\nu}}_i(\widehat{\mathbf{s}}_i)\ .
\end{aligned}$$

Analogously, we have that

$$\begin{aligned}
H_{\boldsymbol{\beta},i} &= \frac{\partial}{\partial\varepsilon} h\boldsymbol{\beta}(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i)\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial\varepsilon}\frac{\partial}{\partial\boldsymbol{\beta}} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} + \frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial\boldsymbol{\beta}} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial\varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \\
&\quad + \frac{\partial}{\partial u}\frac{\partial}{\partial\boldsymbol{\beta}} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial\varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \mathbf{x}_i\ ,
\end{aligned}$$

so

$$\begin{aligned}
H_{\boldsymbol{\beta},i} = \mathrm{EIF}\left(\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2\boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathrm{EIF}(\widehat{\boldsymbol{\beta}}) + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u\partial\boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \\
\mathrm{EIF}(\widehat{\boldsymbol{\beta}})\mathbf{x}_i.
\end{aligned}$$

Finally, in a similar way, we obtain that

$$\begin{aligned}
H_{u,i} &= \frac{\partial}{\partial\varepsilon} h_u(\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon, \widehat{\boldsymbol{\beta}}_\varepsilon^{\mathrm{T}}\mathbf{x}_i)\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial\varepsilon}\frac{\partial}{\partial u} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} + \frac{\partial}{\partial\boldsymbol{\beta}}\frac{\partial}{\partial u} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial\varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \\
&\quad + \frac{\partial}{\partial u}\frac{\partial}{\partial u} h(\varepsilon, \boldsymbol{\beta}, u)\Big|_{(\varepsilon,\mathbf{s})=(0,\widehat{\mathbf{s}}_i)} \frac{\partial}{\partial\varepsilon}\widehat{\boldsymbol{\beta}}_\varepsilon\Big|_{\varepsilon=0} \mathbf{x}_i\ ,
\end{aligned}$$

which implies that

$$\begin{aligned}
H_{u,i} &= \mathrm{EIF}\left(\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}\partial u}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathrm{EIF}(\widehat{\boldsymbol{\beta}}) \\
&\quad + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2 u}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathrm{EIF}(\widehat{\boldsymbol{\beta}})\mathbf{x}_i\ .
\end{aligned}$$

Using the previous expressions, we deduce that

$$\begin{aligned}
H_{\boldsymbol{\beta},i} + \mathbf{x}_i H_{u,i} &= \mathrm{EIF}\left(\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \mathrm{EIF}\left(\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i \\
&\quad + \left[\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2\boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2 u}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\right. \\
&\quad \left. + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u\partial\boldsymbol{\beta}}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i^{\mathrm{T}} + \frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}\partial u}\Big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i} \mathbf{x}_i^{\mathrm{T}}\right] \mathrm{EIF}(\widehat{\boldsymbol{\beta}})\ .
\end{aligned}$$

Now, replacing in (A.3) H_i, $H_{\boldsymbol{\beta},i}$ and $H_{u,i}$ with the obtained expression, we have that

$$
\begin{aligned}
\left.\frac{\partial}{\partial \varepsilon}\boldsymbol{\lambda}(\varepsilon)\right|_{\varepsilon=0} &= \frac{1}{n}\sum_{i=1}^{n}\chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i)\ \mathrm{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\\
&+\frac{1}{n}\sum_{i=1}^{n}\chi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i)\ \widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\widehat{\boldsymbol{\nu}}_i(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\,\mathrm{EIF}(\widehat{\boldsymbol{\beta}})\\
&+\frac{1}{n}\sum_{i=1}^{n}\psi\left(y_i, \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i), c\right)\ \tau(\mathbf{x}_i)\left\{\mathrm{EIF}\left.\left(\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\right.\\
&+\mathrm{EIF}\left.\left(\frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right)\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i\\
&+\left[\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2\boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2 u}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u\partial\boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i^{\mathrm{T}}\right.\\
&\left.\left.+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}\partial u}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i^{\mathrm{T}}\right]\mathrm{EIF}(\widehat{\boldsymbol{\beta}})\right\}\\
&+\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0), c\right)\widehat{\boldsymbol{\nu}}_0(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_0)\tau(\mathbf{x}_0).
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\mathbf{V}(\widehat{\mathbf{s}}_i) = &\left[\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2\boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial^2 u}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial u\partial\boldsymbol{\beta}}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i^{\mathrm{T}}\right.\\
&\left.+\left.\frac{\partial^2\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}\partial u}\right|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}\mathbf{x}_i^{\mathrm{T}}\right].
\end{aligned}
$$

Then, we get that

$$
\left.\frac{\partial}{\partial\varepsilon}\boldsymbol{\lambda}(\varepsilon)\right|_{\varepsilon=0} = \boldsymbol{\ell}_n + \mathbf{M}_n\,\mathrm{EIF}(\widehat{\boldsymbol{\beta}}),
$$

where $\boldsymbol{\ell}_n \in \mathbb{R}^q$ and $\mathbf{M}_n \in \mathbb{R}^{q\times q}$ are defined in (20) and (21). Replacing in (A.2), we have that

$$
\mathbf{0} = \left(\mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right)(\boldsymbol{\ell}_n + \mathbf{M}_n\,\mathrm{EIF}(\widehat{\boldsymbol{\beta}})).
$$

It is worth noticing that since $\|\widehat{\boldsymbol{\beta}}_\varepsilon\|^2 = 1$, differentiating with respect to ε and evaluating at $\varepsilon = 0$, we have that

$$0 = \frac{\partial}{\partial \varepsilon} \widehat{\boldsymbol{\beta}}_{\varepsilon}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{\varepsilon} \Big|_{\varepsilon=0} = 2\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \operatorname{EIF}(\widehat{\boldsymbol{\beta}})$$

which, taking into account that $\widehat{\boldsymbol{\beta}} = \mathbf{e}_q$, implies that $\operatorname{EIF}(\widehat{\boldsymbol{\beta}})_q = 0$. Therefore, we only have to compute $\operatorname{EIF}(\widehat{\boldsymbol{\beta}})_j$ for $j = 1, \ldots, q-1$.

Using again that $\widehat{\boldsymbol{\beta}} = \mathbf{e}_q$, we obtain that

$$\left(\mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right) = \begin{pmatrix} \mathbf{I}_{q-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}.$$

Hence, we have that the left superior matrix of $\left(\mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right) \mathbf{M}_n$ equals the matrix $\mathbf{M}_{n,1} \in \mathbb{R}^{(q-1)\times(q-1)}$, so that $\mathbf{0} = \left(\mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\right) (\boldsymbol{\ell}_n + \mathbf{M}_n \operatorname{EIF}(\widehat{\boldsymbol{\beta}}))$ implies

$$\mathbf{0} = \boldsymbol{\ell}_n^{(q-1)} + \mathbf{M}_{n,1} \operatorname{EIF}(\widehat{\boldsymbol{\beta}}^{(q-1)}). \tag{A.4}$$

Therefore, from (A.4) we get that $\operatorname{EIF}(\widehat{\boldsymbol{\beta}}^{(q-1)}) = -\mathbf{M}_{n,1}^{-1} \boldsymbol{\ell}_n^{(q-1)}$.

It is worth noticing that $\boldsymbol{\ell}_n$ and $\mathbf{M}_n$ involve $\operatorname{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}$, $\operatorname{EIF}(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial\boldsymbol{\beta})\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}$ and $\operatorname{EIF}(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u)\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}$.

b) Let us derive $\operatorname{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u))$. Since $\widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u)$ is the solution of (A.1), we have that

$$\frac{(1-\varepsilon)}{n} \sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u), c\right) + \varepsilon\, K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}^{\varepsilon}(u), c\right) = 0.$$

Differentiating with respect to ε and evaluating at $\varepsilon = 0$, we obtain that

$$\operatorname{EIF}(\widehat{\eta}_{\boldsymbol{\beta}}(u)) = -\frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)}{\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)}. \tag{A.5}$$

Analogously, differentiating first with respect to $\boldsymbol{\beta}$ on both sides of Eq. (A.1) and then, with respect to ε and evaluating at $\varepsilon = 0$, we can obtain an expression for $\operatorname{EIF}(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial\boldsymbol{\beta})\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}$. Alternatively, we may differentiate (A.5) with respect to $\boldsymbol{\beta}$ to obtain

$$\begin{aligned}
\operatorname{EIF}\left(\frac{\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\partial\boldsymbol{\beta}}\right) &= -\frac{\frac{1}{h}K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\mathbf{x}_0 + K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\chi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)}{\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)} \\
&+ \frac{K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_0 - u)\psi\left(y_0, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)}{\left\{\frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\right\}^2} \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_h'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\mathbf{x}_i\right. \\
&\left. + \frac{1}{n}\sum_{i=1}^{n} K_h(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u)\chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), c\right)\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u)\right].
\end{aligned}$$

Similar arguments lead to the expression for $\text{EIF}(\partial\widehat{\eta}_{\boldsymbol{\beta}}(u)/\partial u)\big|_{(\boldsymbol{\beta},u)=\widehat{\mathbf{s}}_i}$.

Finally, note that $\widehat{\eta}_{\boldsymbol{\beta}}(u)$, satisfies

$$\sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right) = 0. \tag{A.6}$$

Hence, differentiating with respect to $\boldsymbol{\beta}$ equation (A.6), we get that

$$\begin{aligned} 0 &= \frac{1}{h}\sum_{i=1}^{n} K'\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right)\mathbf{x}_i \\ &\quad + \sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right) \times \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u), \end{aligned}$$

which implies that

$$\begin{aligned} \frac{\partial}{\partial\boldsymbol{\beta}}\widehat{\eta}_{\boldsymbol{\beta}}(u) = -\frac{1}{h}&\left[\sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right)\right]^{-1} \\ &\sum_{i=1}^{n} K'\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right)\mathbf{x}_i\ . \end{aligned}$$

On the other hand, differentiating (A.6) with respect to u, we obtain that

$$\begin{aligned} 0 &= -\frac{1}{h}\sum_{i=1}^{n} K'\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right) \\ &\quad + \sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right) \times \frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u) \end{aligned}$$

which entails that

$$\begin{aligned} \frac{\partial}{\partial u}\widehat{\eta}_{\boldsymbol{\beta}}(u) = \frac{1}{h}&\left[\sum_{i=1}^{n} K\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\chi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right)\right]^{-1} \\ &\sum_{i=1}^{n} K'\left(\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - u}{h}\right)\psi\left(y_i, \widehat{\eta}_{\boldsymbol{\beta}}(u), \alpha\right)\ . \end{aligned}$$

□

References

Aït Sahalia, Y. (1995). *The delta method for nonaparmetric kernel functionals*. Ph.D. dissertation, University of Chicago.

Bianco, A., Boente, G. (2002). On the asymptotic behavior of one-step estimation. *Statistics and Probability Letters*, *60*, 33–47.

Bianco, A., Boente, G. (2007). Robust estimators under a semiparametric partly linear autoregression model: asymptotic behavior and bandwidth selection. *Journal of Time Series Analysis*, *28*, 274–306.

Bianco, A., García Ben, M., Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canadian Journal Statistics*, *33*, 511–528.

Boente, G., Fraiman, R., Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference*, *57*, 109–142.

Boente, G., Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. *Computational Statistics and Data Analysis*, *52*, 2808–2828.

Boente, G., Rodriguez, D. (2010). Robust inference in generalized partially linear models. *Computational Statistics and Data Analysis*, *54*, 2942–2966.

Boente, G., Rodriguez, D. (2012). Robust estimates in generalized partially linear single-index models. *TEST*, *21*, 386–411.

Cantoni, E., Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, *11*, 141–146.

Cantoni, E., Ronchetti, E. (2006). A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. *Journal of Health Economics*, *25*, 198–213.

Carroll, R., Fan, J., Gijbels, I., Wand, M. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

Chang, Z. Q., Xue, L. G., Zhu, L. X. (2010). On an asymptotically more efficient estimation of the single-index model. *Journal of Multivariate Analysis*, *101*, 1898–1901.

Croux, C., Ruiz-Gazen, A. (2005). High breakdown estimators for principal components: the projection-pursuit approach revisited. *Journal of Multivariate Analysis*, *95*, 206–226.

Delecroix, M., Hristache, M., Patilea, V. (2006). On semiparametric M-estimation in single-index regression. *Journal of Statistical Planning and Inference*, *136*, 730–769.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*, 383–394.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, *21*, 157–178.

Härdle, W., Stoker, T. M. (1989). Investigating smooth multiple regression by method of average derivatives. *Journal of the American Statistical Association*, *84*, 986–95.

Hubert, M., Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, *52*, 5186–5201.

Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics*, *33*, 2291–2310.

Leung, D., Marriott, F., Wu, E. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, *4*, 333–339.

Li, W., Patilea, W. (2017). A new inference approach for single-index models. *Journal of Multivariate Analysis*, *158*, 47–59.

Liu, J., Zhang, R., Zhao, W., Lv, Y. (2013). A robust and efficient estimation method for single index models. *Journal of Multivariate Analysis*, *122*, 226–238.

Mallows, C. (1974). *On some topics in robustness*. Memorandum, Bell Laboratories, Murray Hill, N.J.

Manchester, L. (1996). Empirical influence for robust smoothing. *Australian Journal of Statistics*, *38*, 275–296.

Marazzi, A., Yohai, V. (2004). Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference*, *122*, 271–291.

Maronna, R., Martin, D., Yohai, V. (2006). *Robust statistics: Theory and methods*. New York: Wiley.

Pollard, D. (1984). *Convergence of stochastic processes. Springer series in statistics*. New York: Springer.

Powell, J. L., Stock, J. H., Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, *57*, 1403–30.

Rodriguez, D. (2007). *Estimación robusta en modelos parcialmente lineales generalizados*. Ph.D. Thesis (in spanish), Universidad de Buenos Aires. http://cms.dm.uba.ar/academico/carreras/doctorado/tesisdanielarodriguez.pdf. Accessed 20 Feb 2019.

Rousseeuw, P. J., Yohai, V. J. (1984). Robust regression by means of S-estimators. In J. Franke, W. Hardle, D. Martin (Eds.), *Robust and nonlinear time series, Lecture notes in statistics* (Vol. 26, pp. 256–272). New York: Springer.

Severini, T., Staniswalis, J. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, *89*, 501–511.

Severini, T., Wong, W. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, *20*(4), 1768–1802.

Sherman, R. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics*, *22*, 439–459.

Sun, Y., Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, *20*, 316–334.

Tamine, J. (2002). *Smoothed influence function: Another view at robust nonparametric regression*. Discussion paper 62, Sonderforschungsbereich 373, Humboldt-Universiät zu Berlin.

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

van der Vaart, A. (1988). Estimating a real parameter in a class of semiparametric models. *Annals of Statistics*, *16*(4), 1450–1474.

Wang, F., Scott, D. (1994). The L1 method for robust nonparametric regression. *Journal of the American Statistical Association*, *89*, 65–76.

Wang, Q., Zhang, T., Hädle, W. (2014). *An extended single index model with missing response at random*. SFB 649 Discussion Paper 2014-003.

Wu, T. Z., Yu, K., Yu, Y. (2010). Single index quantile regression. *Journal of Multivariate Analysis*, *101*, 1607–1621.

Xia, Y., Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis*, *97*, 1162–1184.

Xia, Y., Härdle, W., Linton, O. (2012). Optimal smoothing for a computationally and efficient single index estimator. In *Exploring research frontiers in contemporary statistics and econometrics: A Festschrift for Léopold Simar* (pp. 229–261).

Xia, Y., Tong, H., Li, W. K., Zhu, L. (2002). An adaptive estimation of dimension reduction space (with discussion). *Journal of the Royal Statistical Society, Series B*, *64*, 363–410.

Xue, L. G., Zhu, L. X. (2006). Empirical likelihood for single-index model. *Journal of Multivariate Analysis*, *97*, 1295–1312.

Zhang, R., Huang, R., Lv, Z. (2010). Statistical inference for the index parameter in single-index models. *Journal of Multivariate Analysis*, *101*, 1026–1041.