Computational Statistics and Data Analysis 106 (2017) 46–64

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust estimation in partially linear errors-in-variables models

Ana M. Bianco [a,*], Paula M. Spano [b]

[a] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Ciudad Universitaria, Buenos Aires, Argentina

[b] Departamento de Ciencias Exactas, Ciclo Básico Común, Universidad de Buenos Aires, Ciudad Universitaria, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form 31 August 2016
Accepted 1 September 2016
Available online 10 September 2016

Keywords:
Fisher-consistency
Kernel weights
M-location functionals
Nonparametric regression
Robust estimation

ABSTRACT

In many applications of regression analysis, there are covariates that are measured with errors. A robust family of estimators of the parametric and nonparametric components of a structural partially linear errors-in-variables model is introduced. The proposed estimators are based on a three-step procedure where robust orthogonal regression estimators are combined with robust smoothing techniques. Under regularity conditions, it is proved that the resulting estimators are consistent. The robustness of the proposal is studied by means of the empirical influence function when the linear parameter is estimated using the orthogonal M-estimator. A simulation study allows to compare the behaviour of the robust estimators with their classical relatives and a real example data is analysed to illustrate the performance of the proposal.

1. Introduction

Two important branches of regression analysis arise from parametric and nonparametric models. The fully parametric models are readily interpretable, but they can be severely affected by misspecification. On the other hand, nonparametric models are very flexible to assess the relationship among variables, but they suffer from the well known *curse of dimensionality*. In the last decades semiparametric models, that amalgamate these two branches, have deserved a lot of attention. They take the best and avoid the worst of the parametric and nonparametric models. Among them, partially linear models have been extensively studied in the last years. Let $(y, \mathbf{x}^{\mathrm{T}}, t)$ be the observation in a subject or experimental unit, where y is the response that is related to the covariates $(\mathbf{x}^{\mathrm{T}}, t) \in \mathbb{R}^p \times \mathbb{R}$. The partially linear model assumes that

$$y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t) + e,$$

where the error e is independent of the covariates $(\mathbf{x}^{\mathrm{T}}, t)$. By means of a nonparametric component, partially linear models are flexible enough to cover many situations; indeed, they can be a suitable choice when one suspects that the response y linearly depends on $\mathbf{x}$, but that it is nonlinearly related to t. An extensive description of the different results obtained in partially linear regression models can be found in Härdle et al. (2000). Among the robust literature, we find He et al. (2002) that consider M-type estimates for repeated measurements using B-splines and Bianco and Boente (2004) who introduce a kernel-based stepwise procedure to define robust estimates under a partially linear model.

* Correspondence to: Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Piso 2, 1428, Buenos Aires, Argentina. Fax: +54 11 45763375.
E-mail addresses: abianco@dm.uba.ar (A.M. Bianco), paumerspano@yahoo.com.ar (P.M. Spano).

http://dx.doi.org/10.1016/j.csda.2016.09.002
0167-9473/

In practice, however, there often exist covariate measurement errors. This is a common situation in economics, medicine and social sciences. Errors-in-variables (EV) models have drawn a lot of attention and generated a wide literature, surveyed in Fuller (1987) and Carroll et al. (1995). The effect of measurement errors is well-known, indeed they can cause biased and inconsistent parameter estimators. Two approaches are adopted in order to overcome these difficulties according to the nature of the problem: the *functional* and *structural* modelling. In the functional model it is assumed that the covariates are deterministic, while in the *structural* model, which is treated in this paper, the covariates are considered as random variables. In our setting, we assume that we cannot observe $\mathbf{x}$ directly, but instead we observe a surrogate variable $\mathbf{v}$ which is related to $\mathbf{x}$ through the equation $\mathbf{v} = \mathbf{x} + \mathbf{e}_x$. In other words, the response and the vector of covariates $\mathbf{x}$ are observed with errors, while the scalar variable t is observable, that is, we assume the partially linear errors-in-variables (PLEV) model given by

$$\begin{aligned} y &= \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} + g(t) + e, \\ \mathbf{v} &= \mathbf{x} + \mathbf{e}_x, \end{aligned} \tag{1}$$

where the vector of measurement errors

$$\boldsymbol{\epsilon} = \begin{pmatrix} e \\ \mathbf{e}_x \end{pmatrix} \tag{2}$$

is independent of $(\mathbf{x}^{\mathrm{T}}, t)$.

In order to correct for measurement error, some additional information or data is usually required. In the classical approach, at this point, there are two variants. In the first one, it is assumed that the covariance matrix of the measurement errors, $\boldsymbol{\Sigma}_{\mathbf{e}_x}$, is known and the approach is a correction for attenuation. Following these ideas, Liang et al. (1999) adapt the estimators of Severini and Staniswalis (1994), which combine local smoothers and linear parametric techniques, by including an attenuation term based on $\boldsymbol{\Sigma}_{\mathbf{e}_x}$ that enables to adjust the regression coefficients for the effects of measurement error. If $\boldsymbol{\Sigma}_{\mathbf{e}_x}$ were unknown, the estimation of the covariance matrix could be possible when replicates are available. In the second variant, it is assumed that the ratio between the variance of the error model e and the measurement errors $\mathbf{e}_x$ is known. This assumption allows for identification of the model. In this case, Liang et al. (1999) propose to estimate $\boldsymbol{\beta}$ by total least squares method.

Even when in practice the feasibility of any of these conditions depends on the problem, in general, in the robust framework assumptions involving the existence of first or second moments of the errors are avoided and replaced by weaker conditions on the errors distribution, such as symmetry. So, in this paper, we will extend the second variant by assuming that the vector of errors $\boldsymbol{\epsilon}$ follows a spherically symmetric distribution, which is a standard assumption in errors-in-variables models. In this case, if $\boldsymbol{\epsilon}$ has a density, it is of the form $\phi(\|\mathbf{u}\|)$ for some non-negative function ϕ. Spherical symmetry implies that e and each component of $\mathbf{e}_x$ have the same distribution. Cui and Kong (2006) justify this assumption by noticing that in some situations the response y and the covariate x are measured in the same way or, even more, the response and the non-observable covariate are two methods that measure the same quantity. As motivating example, we can consider the problem of predicting cholesterol serum level (CS) from a previous register of CS and age, which corresponds to the case of the real dataset we analyse below. First, it is sensible to assume that both cholesterol serum variables (the response and the covariate) are affected by an error, justifying to fit an EV model. Second, since both measures are of the same nature, it seems natural to assume that the errors of the response and the covariate follow the same distribution, making reasonable the sphericity assumption.

Among the literature in partially linear EV models, we can highlight the contribution of several authors. As mentioned, Liang et al. (1999) introduce a semiparametric version of the parametric correction for attenuation, while He and Liang (2000) consider consistent regression quantile estimates of $\boldsymbol{\beta}$. Partially linear models with measurement errors have been also studied by Ma and Carroll (2006), who propose locally efficient estimators in semiparametric models, Liang et al. (2007) that consider missing not at random responses, Pan et al. (2008) who deal with longitudinal data and by Liang and Li (2009) who focus on variable selection. As mentioned, we deal with the case in which variable t is observable. Measurement errors in both the parametric and the non-parametric part represent a much more complicated problem and would deserve a different approach, that is beyond the scope of this paper. In the classical setting, Liang (2000) and Zhu and Cui (2003), who deal with an unobservable variable t in the context of a partially linear model, consider deconvolution techniques to handle this type of situations.

However, if the smoothers involved in the estimation process are not resistant to outliers, then the resulting estimators can be severely affected by a relatively small fraction of atypical observations. The same can be asserted with respect to the estimation of the regression parameter when it is estimated by total least squares or least squares corrected for attenuation. For this reason, in this paper we consider an intuitively appealing way to obtain robust estimators for model (1) with spherically symmetric errors, which combines robust univariate smoothers with robust parametric estimators for a linear EV model. It is expected that the good robustness properties of estimates for linear EV models, such as M-orthogonal estimators or weighted orthogonal estimators introduced by Zamar (1989) and Fekri and Ruiz-Gazen (2004), respectively, combined with local smoothers, such as local medians or local M-type estimators, would result in estimators with good robustness properties as well. In what follows, we introduce a three-step procedure that yields robust and consistent estimators. We also derive the empirical influence function of the proposal when M-orthogonal estimators are used to estimate the regression parameter. The simulation results show that, regardless of the presence of outliers in the sample, the proposed estimators of the parametric and nonparametric components are very stable, making clear the advantage of using this kind of procedures.

The outline of the paper is as follows. In Section 2 we remind the classical estimators and the three-step procedure for robust estimation in the partially linear EV model is outlined. In Section 3 we prove the consistency of the proposal. In Section 4 we derive the empirical influence function in order to study the sensitivity of the parametric component of the model to outlying observations in the case in which the linear parameter is estimated using the orthogonal M-estimator. The robustness and performance for finite samples of the proposal are studied by means of a numerical study in Section 5 and a real data set is analysed in Section 6. Proofs are relegated to Appendix A.

2. Estimators

In this section we consider the estimation of $\boldsymbol{\beta}$ and g in a partially linear EV model, where for $1 \le i \le n$

$$\begin{aligned} y_i &= \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + g(t_i) + e_i, \\ \mathbf{v}_i &= \mathbf{x}_i + \mathbf{e}_{xi}. \end{aligned} \tag{3}$$

2.1. Classical approach

Assume that $(y, \mathbf{x}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}, t)$ is a random vector with the same distribution as $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i)$, i.e., that satisfies (1) and (2). In the classical approach, it is assumed that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and the existence of higher order moments of $\boldsymbol{\epsilon}$ and $\mathbf{x}$. In this case, taking conditional expectation in (1), we have that

$$\begin{aligned} E(y|t=\tau) &= \boldsymbol{\beta}^{\mathrm{T}}E(\mathbf{x}|t=\tau) + g(\tau), \\ E(\mathbf{v}|t=\tau) &= E(\mathbf{x}|t=\tau), \end{aligned} \tag{4}$$

resulting that $g(\tau) = \nu_o(\tau) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\nu}(\tau)$, where $\nu_o(\tau) = E(y|t=\tau)$ and $\boldsymbol{\nu}(\tau) = \left(\nu_1(\tau), \dots, \nu_p(\tau)\right)^{\mathrm{T}}$ with $\nu_j(\tau) = E(v_j|t=\tau)$ for $1 \le j \le p$. From (1) and (4), we obtain that

$$\begin{aligned} y - \nu_o(\tau) &= \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\nu}(\tau)) + e, \\ \mathbf{v} - \boldsymbol{\nu}(\tau) &= \mathbf{x} - \boldsymbol{\nu}(\tau) + \mathbf{e}_x, \end{aligned} \tag{5}$$

which reduces to a linear EV model. The regression parameter $\boldsymbol{\beta}$ can be estimated through the total least squares estimator that corresponds to perform orthogonal regression.

Let us remind the definition of the total least squares estimator in the simplest case of a linear EV model, where we observe $\mathbf{z}_i^{\mathrm{T}} = (y_i, \mathbf{v}_i^{\mathrm{T}})$, $1 \le i \le n$, such that

$$\begin{aligned} y_i &= \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + e_i, \\ \mathbf{v}_i &= \mathbf{x}_i + \mathbf{e}_{xi}. \end{aligned} \tag{6}$$

In this case, $\mathbf{z}_i = z_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i^{\mathrm{T}} = (e_i, \mathbf{e}_{xi}^{\mathrm{T}})$ is as in (2) and the vectors $z_i^{\mathrm{T}} = (\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \mathbf{x}_i^{\mathrm{T}})$ belong to a hyperplane $\mathcal{H}_{\boldsymbol{\beta}}$, with $\mathcal{H}_{\mathbf{b}} = \{z = (r, \mathbf{x}^{\mathrm{T}}) : r = \mathbf{x}^{\mathrm{T}}\mathbf{b}\}$. The orthogonal regression method looks for the p-dimensional hyperplane that provides the best fit to $\mathbf{z}_1, \dots, \mathbf{z}_n$, i.e., the total least squares estimator minimizes

$$\sum_{i=1}^{n} \|\mathbf{z}_i - \Pi_{\mathcal{H}_{\mathbf{b}}}(z_i)\|^2, \tag{7}$$

where $\Pi_{\mathcal{H}_{\mathbf{b}}}$ is the orthogonal projector on $\mathcal{H}_{\mathbf{b}}$.

Estimators of $\nu_o(\tau)$ and $\boldsymbol{\nu}(\tau)$ can be plugged in Eq. (5) prior to the estimation of the regression parameter. For this purpose, consider the weights

$$\omega_i(\tau) = \frac{K\left(\frac{t_i-\tau}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{t_j-\tau}{h}\right)}, \tag{8}$$

then, according to the classical approach, the conditional expectations can be estimated through

$$\widehat{\nu}_{o,\mathrm{LS}}(\tau) = \sum_{i=1}^{n} \omega_i(\tau) y_i \quad \text{and} \quad \widehat{\nu}_{j,\mathrm{LS}}(\tau) = \sum_{i=1}^{n} \omega_i(\tau) v_{ij}, \tag{9}$$

where K is a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h is the bandwidth parameter. The classical estimator of $\boldsymbol{\beta}$ can be based on the total least squares method replacing in (7) $\mathbf{z}_i$ by the residual vectors

$$\widetilde{\mathbf{z}}_i = \begin{pmatrix} y_i - \widehat{\nu}_{o,\mathrm{LS}}(t_i) \\ \mathbf{v}_i - \widehat{\boldsymbol{\nu}}_{\mathrm{LS}}(t_i) \end{pmatrix},$$

where the vector $\widehat{\boldsymbol{\nu}}_{\mathrm{LS}}(\tau)$ has components $\widehat{\nu}_{j,\mathrm{LS}}(\tau)$.

2.2. Proposal

Unlike the classical model, in the robust setting no moment conditions are assumed, thus, resembling Bianco and Boente (2004), we prefer to consider robust conditional location functionals. More precisely, consider local robust scale functionals corresponding to $y|t = \tau$ and $v_j|t = \tau$, $s_o(\tau)$ and $s_j(\tau)$, respectively. Define the related functionals $\nu_j(\tau),\ 0 \le j \le p$ as the solution of

$$E_{F_o}\left[\psi_1\left(\frac{y-\nu_o(\tau)}{s_o(\tau)}\right)\Big| t=\tau\right]=0 \quad \text{and} \quad E_{F_j}\left[\psi_1\left(\frac{v_j-\nu_j(\tau)}{s_j(\tau)}\right)\Big| t=\tau\right]=0, \quad 1\le j\le p, \tag{10}$$

where ψ_1 is an odd, increasing, bounded and continuous function and E_{F_o} and $E_{F_j}, j = 1, \ldots, p$, denote the expectations respect to $F_o(\mathrm{y}|t = \tau)$ and $F_j(v|t = \tau)$, the (cumulative) distribution functions of $y|t = \tau$ and $v_j|t = \tau$, respectively. We propose to estimate these conditional location functionals by using a robust smoothing and the regression parameter by means of a robust regression estimator. This prompts the following three-step procedure.

- **Step 1**: Estimate $\nu_o(\tau)$ and $\nu_j(\tau)$ through a robust conditional location estimator. Denote $\widehat{\nu}_o(\tau)$ and $\widehat{\nu}_j(\tau)$ the obtained estimates and $\widehat{\boldsymbol{\nu}}(\tau) = \left(\widehat{\nu}_1(\tau), \ldots, \widehat{\nu}_p(\tau)\right)^{\mathrm{T}}$.
- **Step 2**: Consider the residual variables $\widetilde{y}_i = y_i - \widehat{\nu}_o(t_i)$, $\widetilde{\mathbf{x}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\nu}}(t_i)$ and $\widetilde{v}_i = \mathbf{v} - \widehat{\boldsymbol{\nu}}(t_i)$. Estimate the regression parameter of the linear regression EV model in the residuals $\widetilde{y}_i$, $\widetilde{\mathbf{x}}_i$ and $\widetilde{v}_i$ through a robust and consistent estimator. Denote $\widehat{\boldsymbol{\beta}}$ the obtained estimator.
- **Step 3**: Define the estimate of the regression function g as $\widehat{g}(\tau) = \widehat{\nu}_o(\tau) - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\nu}}(\tau)$.

In Step 1 local medians or local M-type estimators can be considered. Take weights $\omega_i(\tau)$ as in (8). Then, local M-type estimators, $\widehat{\nu}_o(\tau)$ and $\widehat{\nu}_j(\tau)$, are defined as the location M-estimators related to the empirical conditional distribution functions $\widehat{F}_o(\mathrm{y}|t = \tau)$ and $\widehat{F}_j(v|t = \tau)$ defined as

$$\widehat{F}_o(\mathrm{y}|t=\tau)=\sum_{i=1}^{n}\omega_i(\tau)I_{(-\infty,\mathrm{y}]}(y_i), \tag{11}$$

$$\widehat{F}_j(v|t=\tau)=\sum_{i=1}^{n}\omega_i(\tau)I_{(-\infty,v]}(v_{ij}). \tag{12}$$

It is worth noticing that $\widehat{F}_o(\mathrm{y}|t = \tau)$ and $\widehat{F}_j(v|t = \tau)$ estimate the distribution of $y|t = \tau$ and $v_j|t = \tau$, that were denoted above as $F_o(\mathrm{y}|t = \tau)$ and $F_j(v|t = \tau)$, respectively. Thus, local M-type estimators are the solutions of

$$\sum_{i=1}^{n}\omega_i(\tau)\psi_1\left(\frac{y_i-\widehat{\nu}_o(\tau)}{\widehat{s}_o(\tau)}\right)=0 \quad \text{and} \quad \sum_{i=1}^{n}\omega_i(\tau)\psi_1\left(\frac{v_{ij}-\widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)=0,$$

and $\widehat{s}_o$ and $\widehat{s}_j$, $1 \le j \le p$, are local robust scale estimates. The score function ψ_1 can be chosen as the Huber or the bisquare ψ-functions; the scales $\widehat{s}_o$ and $\widehat{s}_j$ can be taken as the local MAD, that is, the MAD with respect to the distributions $\widehat{F}_o(\mathrm{y}|t = \tau)$ and $\widehat{F}_j(v|t = \tau)$, $1 \le j \le p$, defined in (11) and (12).

Once the residual variables $\widetilde{y}_i = y_i - \widehat{\nu}_o(t_i)$ and $\widetilde{v}_i = \mathbf{v}_i - \widehat{\boldsymbol{\nu}}(t_i)$ are computed, the regression parameter can be estimated by means of any robust method that yields consistent estimators in the framework of linear regression with measurement errors. Possible choices are the orthogonal regression M-estimates due to Zamar (1989) and the robust weighted orthogonal regression estimators proposed by Fekri and Ruiz-Gazen (2004). We briefly remind the definition of both families of estimators in the case of a linear EV model. In order to robustify the total least squares estimator, Zamar (1989) proposes to use a bounded loss function in the least squares criterion instead of the square loss function. To be more precise, assume a linear model with measurement errors where we observe $\mathbf{z}_i = z_i + \boldsymbol{\epsilon}_i$, with $z_i \in \mathcal{H}_{\boldsymbol{\beta}}$ as described above. Since $z_i \in \mathcal{H}_{\boldsymbol{\beta}}$, we have that

$$\frac{r_i}{\sqrt{1+\|\boldsymbol{\beta}\|^2}}-\frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i}{\sqrt{1+\|\boldsymbol{\beta}\|^2}}=0,$$

so the model can be reparametrized as

$$\mathbf{a}^{\mathrm{T}}z_i=0,$$

where $\mathbf{a} = (a_0, a_1, \ldots, a_p)^{\mathrm{T}}$, with $a_0 = 1/\sqrt{1+\|\boldsymbol{\beta}\|^2}$, $-a_j/a_0 = \beta_j, j = 1, \ldots, p$. For this new parametrization of the model, Zamar (1989) considers the orthogonal regression M-estimator which solves the minimization problem

$$\min_{\|\mathbf{a}\|=1}\sum_{i=1}^{n}\rho\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{z}_i}{s_n}\right), \tag{13}$$

where ρ is a bounded loss function and s_n is some robust estimate of the residuals scale σ.

On the other hand, the robust weighted orthogonal regression estimator in a linear EV model introduced by Fekri and Ruiz-Gazen (2004) minimizes

$$\sum_{i=1}^{n} \eta(\|\mathbf{z}_i - \boldsymbol{\mu}_n\|^2_{\boldsymbol{\Sigma}_n^{-1}})\,\|\mathbf{z}_i - \Pi_{\mathcal{H}_{\mathbf{b}}}(z_i)\|^2, \tag{14}$$

where η is a weight function, $\|\mathbf{z}_i - \boldsymbol{\mu}_n\|^2_{\boldsymbol{\Sigma}_n^{-1}} = (\mathbf{z}_i - \boldsymbol{\mu}_n)^{\mathrm{T}}\boldsymbol{\Sigma}_n^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_n)$ and $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$ are multivariate robust estimators of location and scatter, respectively. Note that the total least squares estimator obtained in (7) corresponds to $\eta \equiv 1$. The robust estimators of multivariate position and dispersion can be chosen among different robust families of estimates; *M*-, *S*-and *MCD* estimators are possible choices.

Therefore, robust estimators of $\boldsymbol{\beta}$ can be obtained in Step 2 replacing in (13) or (14) $\mathbf{z}_i$ by $\widetilde{\mathbf{z}}_i = (y_i - \widehat{\nu}_o(t_i), \mathbf{v}_i^{\mathrm{T}} - \widehat{\boldsymbol{\nu}}(t_i)^{\mathrm{T}})$.

Note that if the model were a purely linear regression model with errors-in-variables, Step 1 and 3 would be not necessary and the proposal would reduce just to Step 2, going down to the orthogonal robust method chosen to estimate the linear component. Besides, if there were no measurement errors, in Step 2 an ordinary regression robust estimator, such as *GM*- or *MM*-estimators could be used and hence, the proposal would result in the procedure introduced by Bianco and Boente (2004).

3. Consistency

We will assume a set of conditions in order to derive the consistency of the proposed estimators for the regression parameter and the nonparametric function g.

H1. ψ_1 is an odd function, strictly increasing, bounded and continuous differentiable, such that $z\psi_1'(z) \le \psi_1(z)$.

H2. $F_o(\mathrm{y}|t = \tau)$ and $F_j(v|t = \tau)$, $1 \le j \le p$, are symmetric around $\nu_o(\tau)$ and $\nu_j(\tau)$, respectively.

H3. For any compact set $\mathbf{C} \subset \mathbb{R}$, the density f_t of t is bounded on $\mathbf{C}$ and $\inf_{t\in\mathbf{C}} f_t(\tau) > 0$.

H4. $F_o(\mathrm{y}|t = \tau)$ and $F_j(v|t = \tau)$, $1 \le j \le p$, are continuous functions in t and satisfy the following equicontinuous condition

$$\forall\varepsilon > 0 \quad \exists\,\delta > 0/|a-b| < \delta \Rightarrow \sup_{\tau\in\mathbf{C}}\max_{0\le j\le p}\left(|F_j(a|t=\tau) - F_j(b|t=\tau)|\right) < \varepsilon \text{ for any compact set } \mathbf{C}.$$

H5. The kernel $K : \mathbb{R} \to \mathbb{R}$ is nonnegative, bounded and Lipschitz function such that

$$\int K(z)\,dz = 1, \qquad \int |z|K(z)\,dz < \infty \quad \text{and} \quad |z|K(z) \to 0 \quad \text{if } |z| \to \infty.$$

H6. The sequence $h = h_n$ verifies $h_n \to 0,\ nh_n \to \infty$ and $nh_n/\log n \to \infty$.

Under this set of conditions, Theorem 3.3 of Boente and Fraiman (1991) implies the strong uniform convergence of the local *M*-estimators. More precisely, assumption **H1** is on the score function, which is chosen by the user, and is standard in the context of robustness. The fact that the score function ψ_1 is odd, strictly increasing, bounded and continuous ensures the uniqueness of $\nu_i(\tau)$, introduced in (10), when the conditional distribution $F_i(\cdot|t = \tau)$ is symmetric. As noted in Boente and Fraiman (1990) and Härdle (1984), among other possibilities, the score function $\psi(u) = \arctan(u)$ verifies these four conditions and satisfies **H1**. When $s_o(\tau)$ and $s_j(\tau)$ are the *MAD* of the conditional distribution of $y|t = \tau$ and $v_j|t = \tau$, respectively, **H2**–**H4** state regularity conditions of the marginal density of t and the conditional distribution functions, that imply that for any compact $\mathbf{C}$ $\inf_{\tau\in\mathbf{C}} s_j(\tau) > 0$ and $\sup_{\tau\in\mathbf{C}} s_j(\tau) < \infty$, for $0 \le j \le p$. **H5** and **H6** refer to the rate of convergence of the smoothing parameter and restrictions over the class of kernel functions to be chosen.

For the sake of completeness we remind the following two lemmas, which are proved in Bianco and Boente (2004). They are needed to derive the consistency of the proposed estimators that is stated in Theorem 3.1 and Corollary 3.1.

Lemma 3.1. *Assume that* ***H1***–***H6*** *hold. Then, for any compact set* $\mathbf{C}$ *we have that* $\sup_{\tau\in\mathbf{C}} |\widehat{\nu}_j(\tau) - \nu_j(\tau)| \xrightarrow{\text{c.s.}} 0,\ 0 \le j \le p$.

Lemma 3.2. *Let* $(r_i, \mathbf{u}_i^{\mathrm{T}}, t_i)^{\mathrm{T}} \in \mathbb{R}^{p+2},\ 1 \le i \le n$ *be independent random vectors distributed over* $(\Omega, \mathcal{A}, \mathcal{P})$ *such that* $(r_i, \mathbf{u}_i^{\mathrm{T}})$ *have common distribution P. Let* $\widehat{\eta}_o(\tau)$ *and* $\widehat{\boldsymbol{\eta}}(\tau) = (\widehat{\eta}_1(\tau), \ldots, \widehat{\eta}_p(\tau))^{\mathrm{T}}$ *be random functions such that for any compact set* $\mathbf{C} \subset \mathbb{R}$

$$\sup_{\tau\in\mathbf{C}} |\widehat{\eta}_j(\tau)| \xrightarrow{\text{c.s.}} 0, \quad 0 \le j \le p.$$

Denote P_n *and* Q_n *the following empirical measures over* $\mathbb{R}^{p+1}$

$$P_n(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(r_i, \mathbf{u}_i) \quad \textit{and} \quad Q_n(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(r_i + \widehat{\eta}_o(t_i), \mathbf{u}_i + \widehat{\boldsymbol{\eta}}(t_i)),$$

where $A \subset \mathbb{R}^{p+1}$ is a Borel set. Therefore,

(a) *for any bounded and continuous function $f : \mathbb{R}^{p+1} \to \mathbb{R}$ we have that*

$$|E_{Q_n}(f) - E_{P_n}(f)| \xrightarrow{\text{c.s.}} 0,$$

(b) *$\Pi(Q_n, P) \xrightarrow{\text{c.s.}} 0$, where Π stands for the Prohorov distance.*

The following theorem states that, under regularity conditions, the uniform convergence over compact sets of the local estimators of Step 1 and the continuity of the regression functional related to the estimation procedure in Step 2, implies the consistency of the resulting regression estimator $\widehat{\boldsymbol{\beta}}$.

Theorem 3.1. *Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i)$, $1 \le i \le n$ be a random sample of vectors that satisfy model (3) and P the distribution of $(y_i - \nu_o(t_i), \mathbf{v}_i^{\mathrm{T}} - \boldsymbol{\nu}(t_i)^{\mathrm{T}})$, where $\nu_o(\tau)$ and $\boldsymbol{\nu}(\tau)$ are defined in* **H2**, *with $\nu_o(\tau) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\nu}(\tau) + g(\tau)$. Assume that $\widehat{\nu}_j(\tau)$, $0 \le j \le p$, are estimators of $\nu_j(\tau)$ such that for any compact set $\mathbf{C} \subset \mathbb{R}$*

$$\sup_{\tau \in \mathbf{C}} |\widehat{\nu}_j(\tau) - \nu_j(\tau)| \xrightarrow{\text{c.s.}} 0, \quad 0 \le j \le p. \tag{15}$$

Let $\boldsymbol{\beta}(G)$ be a regression functional for the linear EV model

$$\mathbf{z} = \begin{pmatrix} \boldsymbol{\beta}^{\mathrm{T}}\mathbf{u} \\ \mathbf{u} \end{pmatrix} + \begin{pmatrix} e \\ \mathbf{e}_u \end{pmatrix}, \tag{16}$$

where $\mathbf{z} \sim G$, $(e, \mathbf{e}_u^{\mathrm{T}})$ is independent of $\mathbf{u}$. Assume that $\boldsymbol{\beta}(G)$ is continuous in P and provides Fisher-consistent estimators. If $\widehat{P}_n(A) = 1/n \sum_{i=1}^n I_A(\widetilde{y}_i, \widetilde{v}_i)$, where $\widetilde{y}_i = y_i - \widehat{\nu}_o(t_i)$, $\widetilde{v}_i^{\mathrm{T}} = \mathbf{v}_i^{\mathrm{T}} - \widehat{\boldsymbol{\nu}}(t_i)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\widehat{P}_n)$, then

$$\widehat{\boldsymbol{\beta}} \xrightarrow{\text{c.s.}} \boldsymbol{\beta}.$$

As in the case of observable covariates $\mathbf{x}$'s, in Theorem 3.1 it is required that **H2** holds with $\nu_o(\tau) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\nu}(\tau) + g(\tau)$. In fact, if $\nu_o(\tau) - \boldsymbol{\nu}(\tau) = h(\tau)$, then $y - \nu_o(\tau) - \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\nu}(\tau)) = g(\tau) - h(\tau) + \epsilon$, so $g(\tau) - h(\tau)$ should be 0 in order to obtain Fisher-consistent estimators.

As a direct consequence of the previous results, we get the consistency of the nonparametric component g, as stated in the following corollary.

Corollary 3.1. *Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i)$, $1 \le i \le n$, be independent random vectors that satisfy model (3). Let $\widehat{\nu}_j(\tau)$, $0 \le j \le p$, be estimators of $\nu_j(\tau)$ such that for any compact set $\mathbf{C} \subset \mathbb{R}$*

$$\sup_{\tau \in \mathbf{C}} |\widehat{\nu}_j(\tau) - \nu_j(\tau)| \xrightarrow{\text{c.s.}} 0, \quad 0 \le j \le p,$$

and suppose that the assumptions of Theorem 3.1 hold. Then, the estimator of the regression function g given by $\widehat{g}(\tau) = \widehat{\nu}_o(\tau) - \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\widehat{\boldsymbol{\nu}}(\tau)$ is uniformly convergent over compact sets.

4. Empirical influence curve

In this section we derive the empirical influence function of the regression parameter estimator when in Step 2 an orthogonal M-estimator is used. The empirical influence function (EIF), introduced by Tukey (1977), is a useful measure of the robustness of an estimator with respect to a single outlier. In fact, it reflects the effect on a given estimator of adding an arbitrary datum to the sample, that may not follow the central model. Mallows (1974) considers a finite version of the influence function defined by Hampel (1974) that is computed at the sample empirical distribution. Diagnostics for identifying outliers can be developed from the empirical influence functions. Even when in parametric models the EIF has been extensively used, it has received less attention in nonparametrics. Manchester (1996) introduces a graphical method to display sensitivity of a scatter plot smoother. Aït Sahalia (1995) presents a smoothed functional approach to nonparametric kernel estimators, that Tamine (2002) uses to define a smoothed influence function in nonparametric regression assuming a fixed smoothing parameter.

Following Boente and Rodriguez (2010), we compute an empirical influence function that addresses Manchester (1996) approach and is close to the version of the EIF introduced by Mallows (1974). It corresponds to the influence function of the functional under study computed for the empirical distribution and evaluated at each datum in the sample.

More precisely, let $\widehat{\boldsymbol{\beta}}$ be the regression parameter estimator based on the original data set $(y_i, \mathbf{v}_i^{\mathrm{T}}, t_i)$, $1 \le i \le n$. Denote P_n the empirical measure that gives weight $1/n$ to each point in the sample, hence $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(P_n)$. Let $P_{n,\varepsilon}$ be the empirical measure that gives mass $(1-\varepsilon)/n$ to each $(y_i, \mathbf{v}_i^{\mathrm{T}}, t_i)$, $1 \le i \le n$ and ε to the observation $(y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)$. That is, we have a new sample with the original data set representing an $1-\varepsilon$ proportion and the new observation an ε proportion. Define $\widehat{\boldsymbol{\beta}}_\varepsilon$

the regression parameter estimator for this new sample. We compute the empirical influence function of $\widehat{\boldsymbol{\beta}}$ at a given point $(y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)$ as

$$\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = \lim_{\varepsilon \to 0} \frac{\widehat{\boldsymbol{\beta}}_\varepsilon - \widehat{\boldsymbol{\beta}}}{\varepsilon}.$$

For simplicity, we will assume that the scale is known and equal to σ. In this case the orthogonal M-estimator in Step 2 is related to the minimization problem

$$\min_{\|\mathbf{a}\|=1} \sum_{i=1}^n \rho\left(\frac{\mathbf{a}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right), \tag{17}$$

with $\widetilde{\mathbf{z}}_i = (\widetilde{z}_{i0}, \ldots, \widetilde{z}_{ip})^{\mathrm{T}}$, $\widetilde{z}_{i0} = y_i - \widehat{\upsilon}_0(t_i)$ and $\widetilde{z}_{ij} = \upsilon_{ij} - \widehat{\upsilon}_j(t_i)$, $1 \le j \le p$ and $\widehat{\upsilon}_j(\tau)$ obtained in Step 1. Let $\widehat{\mathbf{a}}$ be the solution of (17).

Since $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = \left(\partial \widehat{\boldsymbol{\beta}}_\varepsilon / \partial\varepsilon\right)|_{\varepsilon=0}$, from the relationship between $\boldsymbol{\beta}$ and $\mathbf{a}$, we get that

$$\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = \frac{-1}{\widehat{a}_0}\left[\frac{-\widehat{\mathbf{a}}^*}{\widehat{a}_0}\mathbf{I}\right] \qquad \mathrm{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)),$$

where $\widehat{\mathbf{a}}^*$ contains the p last components of $\widehat{\mathbf{a}}$, i.e., $\widehat{\mathbf{a}}^* = (\widehat{a}_1, \ldots, \widehat{a}_p)^{\mathrm{T}}$ and $\mathrm{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))$ is the empirical influence function of the regression parameter after the reparametrization. Denote $\psi = \rho'$, then if

$$\mathbf{A} = \frac{1}{\sigma}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\frac{1}{n}\sum_{i=1}^n \psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\,\mathbf{I} + \widehat{\mathbf{a}}\widetilde{\mathbf{z}}_i^{\mathrm{T}}\right),$$

and $\widehat{\boldsymbol{\upsilon}}^*(\tau) = (\widehat{\upsilon}_0(\tau), \widehat{\boldsymbol{\upsilon}}(\tau)^{\mathrm{T}})$, we have that

$$\begin{aligned}\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = \frac{1}{\widehat{a}_0}\left[\frac{-\widehat{\mathbf{a}}^*}{\widehat{a}_0}\mathbf{I}\right]\mathbf{A}^{-1}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\Bigg[&\psi\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_0\right)\widetilde{\mathbf{z}}_0 - \frac{1}{n}\sum_{i=1}^n \psi'\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\right)\widetilde{\mathbf{z}}_i\widehat{\mathbf{a}}^{\mathrm{T}}\mathrm{EIF}(\widehat{\boldsymbol{\upsilon}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\\ &- \frac{1}{n}\sum_{i=1}^n \psi\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\right)\mathrm{EIF}(\widehat{\boldsymbol{\upsilon}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\Bigg],\end{aligned} \tag{18}$$

where $\mathrm{EIF}(\widehat{\boldsymbol{\upsilon}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau) = \left(\partial\widehat{\boldsymbol{\upsilon}}_\varepsilon^*(\tau)/\partial\varepsilon\right)|_{\varepsilon=0}$ and $\widehat{\boldsymbol{\upsilon}}_\varepsilon^*(\tau)$ is the vector of local M-estimators of Step 1 obtained in the new sample.

To illustrate the behaviour of the estimators and to compare it with the classical ones, we consider the following model. We generate a sample

$$\begin{aligned} y_i &= \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i + \sin\left(\frac{1}{2}\pi t_i\right) + e_i,\\ \mathbf{v}_i &= \mathbf{x}_i + \mathbf{e}_{xi},\end{aligned} \tag{19}$$

for $1 \le i \le 100$, where $\boldsymbol{\beta}^{\mathrm{T}} = (1, 1)$, $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}^{\mathrm{T}} = (0, 0)$ and $\boldsymbol{\Sigma} = \mathbf{I}_2$, $t \sim N(0, 2)$ and then truncated t so that $t \in [-6, 6]$. Besides, we take $\boldsymbol{\epsilon}^{\mathrm{T}} = (e, \mathbf{e}_x^{\mathrm{T}}) \sim N(0, \mathbf{I}_3)$. We consider two situations. In the first one, we take a grid of values $(y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)$ for $t_0 = -1, 0, 1$, $\mathbf{v}_0^{\mathrm{T}} = (\upsilon_0, \upsilon_0)$ and (y_0, υ_0) taking values on an equidistant grid on $y_0 \times \upsilon_0$ of size 41×21 on the rectangle $[-10, 10] \times [-5, 5]$. In the second one, the only difference is that we take $\mathbf{v}_0^{\mathrm{T}} = (\upsilon_0, 0)$. For each fixed t_0, we have computed the empirical influence function given by (18) at each point of this net of 861 points. We compute the influence function of the classical and the robust estimators. The former corresponds to the choice $\psi(u) = \psi_1(u) = u$, while for the robust estimator we choose a score function in the family $\psi_k(u) = \arctan(u/k)/k$ with $k = 0.55$ for the local M-estimators and for the orthogonal M-estimator we select a score function in the Tukey biweight family with tuning constant $c = 4.7$. With $k = 0.55$ the local M-estimator achieves a 90% efficiency under normality, while, as noted in Zamar (1989) with $c = 4.7$ the orthogonal M-estimator achieves an efficiency equal to 95% at the Gaussian linear regression and errors-in-variables models. In the local estimates, we use Gaussian kernel weights with the optimal bandwidth $h_{opt} = 1.30$ in the case of the classical estimator and $h_{opt} = 1.475$ in the robust one. These values of the smoothing parameter are based on the minimization of $CV(h)$ and $D(h)$, the classical and robust cross-validation criteria described in the next section.

Figs. 1 and 2 show that the classical estimator has unbounded influence function. In Fig. 1 we observe that when $\mathbf{v}_0^{\mathrm{T}} = (\upsilon_0, \upsilon_0)$, the norm of the EIF has a moderate growth according to the values of y_0 and υ_0 shrink to 0 and along the first bisection axis of the rectangular area under study, while it increases very quickly in at least one of the corners corresponding to the opposite diagonal. It is worth noticing that points with coordinates $y_0 \to \infty$ and $\upsilon_0 \to -\infty$ or $y_0 \to -\infty$ and $\upsilon_0 \to \infty$, that is in the second and fourth quadrants of the considered rectangle, tend to lie far from the expected value under the regression function, yielding high residuals. Therefore, they may have a potential bad influence

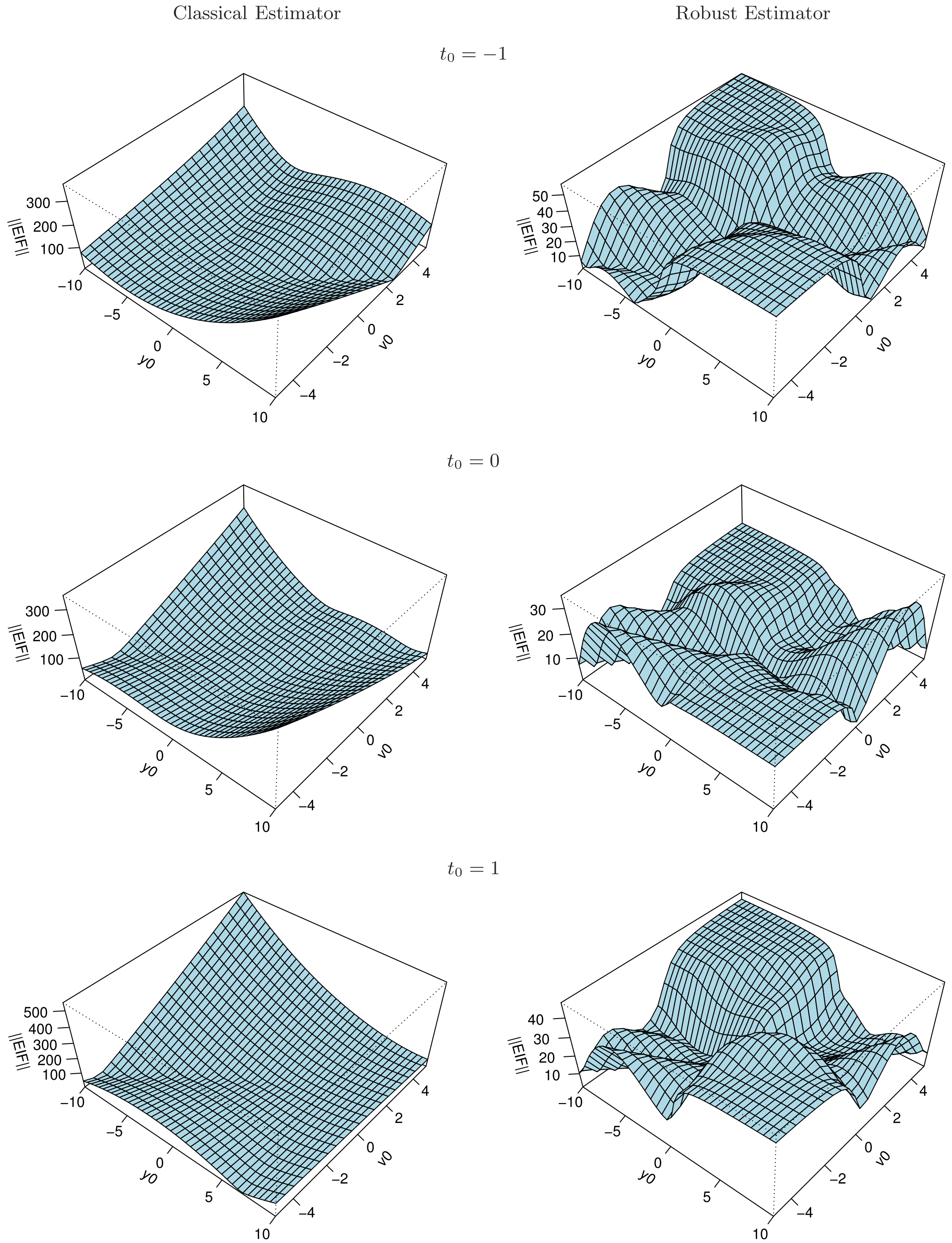

Fig. 1. Norm of the empirical influence function, $\|\text{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))\|$, for $t_0 = -1$, 0 and 1 when $\mathbf{v}_0^{\mathrm{T}} = (v_0, v_0)$.

on any estimator. Figs. 1 and 2 exhibit that the norm of the EIF of the classical estimator is moderate just at the centre of the rectangle, but it grows in the second and fourth quadrants. Fig. 2 also shows that the $\|\text{EIF}\|$ tends to increase at the edges of the rectangle under study when $\mathbf{v}_0^{\mathrm{T}} = (v_0, 0)$. Therefore, the plots given indicate that the $\|\text{EIF}\|$ of the classical estimator increases when $y_0 \to \infty$ and $v_0 \to -\infty$ or $y_0 \to -\infty$ and $v_0 \to \infty$, making these points extremely influential. Besides, the conclusions about the behaviour of the empirical influence function of the classical estimator are similar for the three values of t_0 considered.

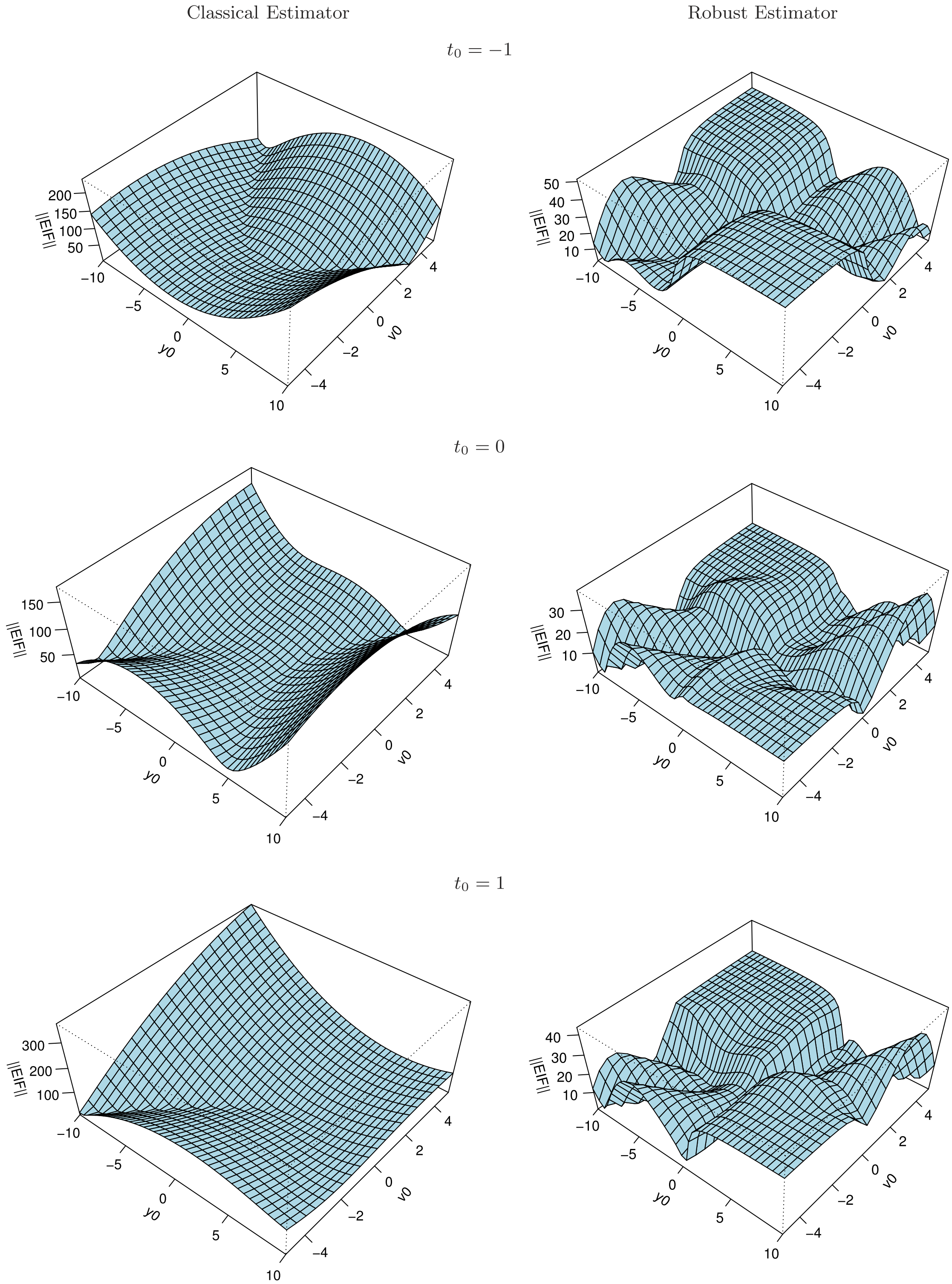

Fig. 2. Norm of the empirical influence function, $\|\mathrm{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))\|$, for $t_0 = -1$, 0 and 1 when $\mathbf{v}_0^{\mathrm{T}} = (v_0, 0)$.

On the other hand, the orthogonal M-estimator is expected to downweight observations with high residuals when the score function ψ is bounded. The norm of the expected influence function of the robust estimator resembles the behaviour of $\|\mathrm{EIF}\|$ of the classical estimator at the centre of the studied region and along the first bisection axis of this area, that correspond to points with small or moderate residuals. However, even when the $\|\mathrm{EIF}\|$ increases rapidly when $y_0 v_0 < 0$, as

$y_0 \to \infty$ and $v_0 \to -\infty$ or $y_0 \to -\infty$ and $v_0 \to \infty$, the $\|\text{EIF}\|$ stabilizes to a plateau, showing that bad influential points on the second and fourth quadrants are downweighted.

5. Monte Carlo study

A Monte Carlo study was carried out to illustrate the behaviour of the proposed estimators and to compare them with the classical ones under different models and contamination schemes. For the numerical experiment we revisit model (19) which was inspired in the simulation study given in Zhu and Cui (2003) and was adapted to our framework, where variable t is assumed to be observable. The sine function in the nonparametric component challenges the ability of the estimators of g to capture the characteristic oscillation of a trigonometric function. Regarding the covariates, we explore different designs in order to compare the performance of the estimators either when covariates $\mathbf{x}$ and t are independent or dependent. Besides, we consider two models for the regressor $\mathbf{x}$. We present below the results concerning the case in which the coordinates of $\mathbf{x}$ are independent. We also experiment with a model where the linear covariate $\mathbf{x}$ has a different covariance structure, i.e., $\mathbf{x}$ has correlated components with different variance. The corresponding results are summarized in the Supplementary material. The supplementary file also contains the R code implementing the proposal and an example script illustrating its use (see Appendix B).

Henceforth, we consider the PLEV model given by (19), where $\boldsymbol{\beta} = (1, 1)^{\mathrm{T}}$, $\mathbf{x} = (x_1, x_2)^{\mathrm{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu} = (0, 0)^{\mathrm{T}}$ and $\boldsymbol{\Sigma} = \mathbf{I}_2$. We analyse two situations, in the first one, we take $t = (x_1 + x_2 + d)(\sqrt{6}/3)$, where $d \sim N(0, 1)$ and then we truncate t so that $t \in [-6, 6]$, leading to a design where the covariates are dependent, while in the second one, we take t with the same marginal distribution as before, but independent of $\mathbf{x}$. In these central models, denoted as C_0, we take $\boldsymbol{\epsilon}^{\mathrm{T}} = (e, \mathbf{e}_x^{\mathrm{T}}) \sim N(0, \sigma^2 \mathbf{I}_3)$, $\sigma = 0.3$. We also consider four contaminated schemes

- C_1 $e \sim 0.9N(0, \sigma^2) + 0.1N(0, (10\sigma)^2)$, that corresponds to enlarge the response errors.
- C_2 $e \sim 0.9N(0, \sigma^2) + 0.1N(0, (10\sigma)^2)$ and $\mathbf{e}_x \sim 0.9N(0, \sigma^2\mathbf{I}_2) + 0.1N(0, (10\sigma)^2\mathbf{I}_2)$, that corresponds to independently enlarge both kinds of errors.
- C_3 $\boldsymbol{\epsilon}^{\mathrm{T}} = (e, \mathbf{e}_x^{\mathrm{T}}) \sim 0.9N(0, \sigma^2\mathbf{I}_3) + 0.1N(0, (10\sigma)^2\mathbf{I}_3)$, that jointly inflates both types of errors.
- C_4 $e \sim 0.9N(0, \sigma^2) + 0.1N(8, 0.01^2)$, that introduces asymmetric response errors.

Note that C_1 and C_4 have the extra effect of breaking the spherical symmetry of the errors distribution.

The classical estimator is computed using the Nadaraya–Watson estimators $\widehat{\nu}_{j,\mathrm{LS}}(\tau)$, given in (9), and combining them with the total least squares estimator of the regression parameter $\boldsymbol{\beta}$. We denote *TLS* the resulting estimator of $\boldsymbol{\beta}$. The robust estimators are based on local M-estimators combined with robust estimators of the parameter $\boldsymbol{\beta}$ under the linear EV model. More precisely, we use the Gaussian kernel weights with standard deviation $0.25/0.675 = 0.37$ such that the interquartile range is 0.5 for the classical and robust smoothing procedures. We compute the robust local M-estimates using the score function based on arctan, that is $\psi_k(u) = 1/k \arctan(u/k)$ with tuning constant $k = 0.55$, which gives a 90% efficiency with respect to its linear relative under normality. Local medians are selected as initial estimates in the iterative procedure for the computation of the local M-estimators.

In both cases, for the smoothing procedures, we choose the bandwidth by means of suitable cross-validation criteria. To describe these procedures, assume that $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i)$, $1 \le i \le n$, follow model (3). The classical cross-validation criterion is given by the minimization of

$$CV(h) = \frac{1}{n}\sum_{t=1}^{n}\left(y_i - \widehat{\boldsymbol{\beta}}_{h,i}^{\mathrm{T}}\mathbf{v}_i - \widehat{g}_{h,i}(t_i)\right)^2, \tag{20}$$

where $\widehat{g}_{h,i}(\tau) = \widehat{\nu}_{o,h,i}(\tau) - \widehat{\boldsymbol{\beta}}_{h,i}^{\mathrm{T}}\widehat{\boldsymbol{\nu}}_{h,i}(\tau)$, $\widehat{\nu}_{o,h,i}(\tau)$ and $\widehat{\boldsymbol{\nu}}_{h,i}(\tau)$ are the smoothers obtained with all the data except $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i)$ using bandwidth h and $\widehat{\boldsymbol{\beta}}_{h,i}$ is obtained from the sample without the ith observation. However, the sensitivity to outliers of the classical methods for the selection of the smoothing parameter is well known in nonparametric regression. For this reason, as in Bianco and Boente (2007) for the robust estimators we consider a robust cross-validation criterion, that makes a trade-off between robust measures of bias and variance. Following this approach, we consider the minimization of

$$D(h) = \widehat{\mu}_n^2\left(y_i - \widehat{\boldsymbol{\beta}}_{h,i}^{\mathrm{T}}\mathbf{v}_i - \widehat{g}_{h,i}(t_i)\right) + \widehat{\sigma}_n^2\left(y_i - \widehat{\boldsymbol{\beta}}_{h,i}^{\mathrm{T}}\mathbf{v}_i - \widehat{g}_{h,i}(t_i)\right), \tag{21}$$

where we take $\widehat{\mu}_n$ as the median and $\widehat{\sigma}_n$ as the Huber τ-scale estimate. We consider the classical cross-validation criterion for the classical estimators and the robust $D(h)$ for our robust proposal.

The robust regression estimators are computed using both, the robust weighted orthogonal regression introduced in Fekri and Ruiz-Gazen (2004) and the orthogonal regression M-estimates as proposed in Zamar (1989). In the case of the weighted orthogonal estimators, we consider three robust estimators of multivariate location and scatter: M-, S-and *MCD* estimators with breakdown point equal to 0.45 for the first one and 0.5 for the two last estimators. We denote the resulting weighted estimators of the regression parameter as WOR_M, WOR_S and WOR_{MCD}. As proposed by Fekri and Ruiz-Gazen (2004), we take hard rejection weights which correspond to the indicator function $\eta \equiv \mathcal{I}_{[0,c]}$, where c is the 97.5% quantile of a distribution χ^2_{p+1}. With respect to the orthogonal regression M-estimator, it was computed by means of the algorithm introduced in

Table 1
Simulation summary: 10% trimmed mean squared error for the estimators of the regression parameter $\boldsymbol{\beta}$, independent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0053	0.0097	0.0104	0.0098	0.0062
C_1	0.2737	0.0135	0.0136	0.0126	0.0203
C_2	0.2674	0.0151	0.0150	0.0145	0.0268
C_3	0.3372	0.0104	0.0108	0.0102	0.0155
C_4	18.0127	0.0106	0.0112	0.0102	0.0139

Table 2
Simulation summary: median of MedSE($\widehat{g}$), independent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0265	0.0399	0.0402	0.0398	0.0389
C_1	0.0589	0.0505	0.0514	0.0509	0.0516
C_2	0.0793	0.0615	0.0615	0.0615	0.0621
C_3	0.0781	0.0582	0.0583	0.0583	0.0582
C_4	0.7746	0.0946	0.0951	0.0940	0.0954

Table 3
Simulation summary: 10% trimmed mean squared error for the estimators of g, independent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0491	0.0814	0.0816	0.0815	0.0809
C_1	0.1172	0.1090	0.1088	0.1087	0.1105
C_2	0.1897	0.1423	0.1424	0.1423	0.1447
C_3	0.1876	0.1312	0.1312	0.1310	0.1315
C_4	1.5745	0.2434	0.2435	0.2433	0.2442

Zamar (1989) using the Tukey biweight loss function with tuning constant $c = 4.7$, yielding the *MOR* estimator for the regression component of the simulated models.

We report on the simulation results based on 1000 replications of samples of size $n = 100$ of each of the described schemes. Fig. 3 gives the boxplots of the estimates of the regression parameters β_1 and β_2 for the two designs. Due to the difference in location and scale between the classical and the robust estimators under contamination, in order to enable a suitable visualization the vertical axis of these boxplots is restricted. Thus, the chosen vertical range is $[-1, 3]$ for C_1 and C_2 and $[-10, 10]$ for C_3 and C_4, even though the classical method yields estimates beyond these bounds. In uncontaminated samples the classical estimator shows its advantage with respect to the robust estimators, while the robust ones do not exhibit an important loss of efficiency. From this Figure the sensitivity of the classical estimator to the selected contaminations becomes evident. Under contaminations C_1 and C_4 the classical estimator has a poor performance, mainly with respect to bias, while under contaminations C_2 and C_3 it becomes very unstable. On the contrary, the robust estimators are very stable in most situations. Robust estimates may have heavy tailed distribution for finite samples and from Fig. 3 this seems to be the case of the orthogonal M-estimators, at least under C_1 and C_2 in the dependent design. This behaviour of robust estimators has been already described, for example in Maronna and Yohai (2010) and Croux et al. (2010). For this reason, as all these authors pointed out, an α-trimmed mean squared error is more representative since it is less influenced by a few extreme values. Hence, we compute the 10% trimmed mean squared error (MSE) of the regression estimates as summary measure, which are reported in Tables 1 and 4 for the independent and dependent designs, respectively. On the other hand, the performance of an estimate $\widehat{g}$ of the nonparametric component g is measured using the 10% trimmed mean squared error, that are summarized in Tables 3 and 6 and also by means of the median squared error computed as

$$\text{MedSE}(\widehat{g}) = \text{median}\left([\widehat{g}(t_i) - g(t_i)]^2\right),$$

shown in Tables 2 and 5.

Tables 1–6 show that, as expected, at the uncontaminated schemes the classical estimators achieve the lowest square errors for both components, the parametric and nonparametric one. However, in general terms, even when there is some loss of efficiency, the behaviour of the robust estimators is satisfactory. Focusing on the parametric component, under any of the considered contaminations, the performance of the classical estimator is very poor. Indeed, Tables 1 and 4 show that the MSE of the regression parameter increases more than fifty times under contamination and things seem to be worse under the dependent structure between the $\mathbf{x}$ and t than under the independent design. The robust estimators are much more stable in all considered circumstances, indeed, they are still being reliable even under the more severe contamination schemes. For instance, the asymmetric contamination on the response errors C_4, that seems to have a devastating effect on the classical estimator of the linear and nonparametric components, is much more harmless for all the robust estimators.

Regarding the estimation of the nonparametric component, in order to give a full picture of the performance of both classical and robust estimators, Figs. 4 and 5 display the functional boxplots of the estimators of function g. Since variable

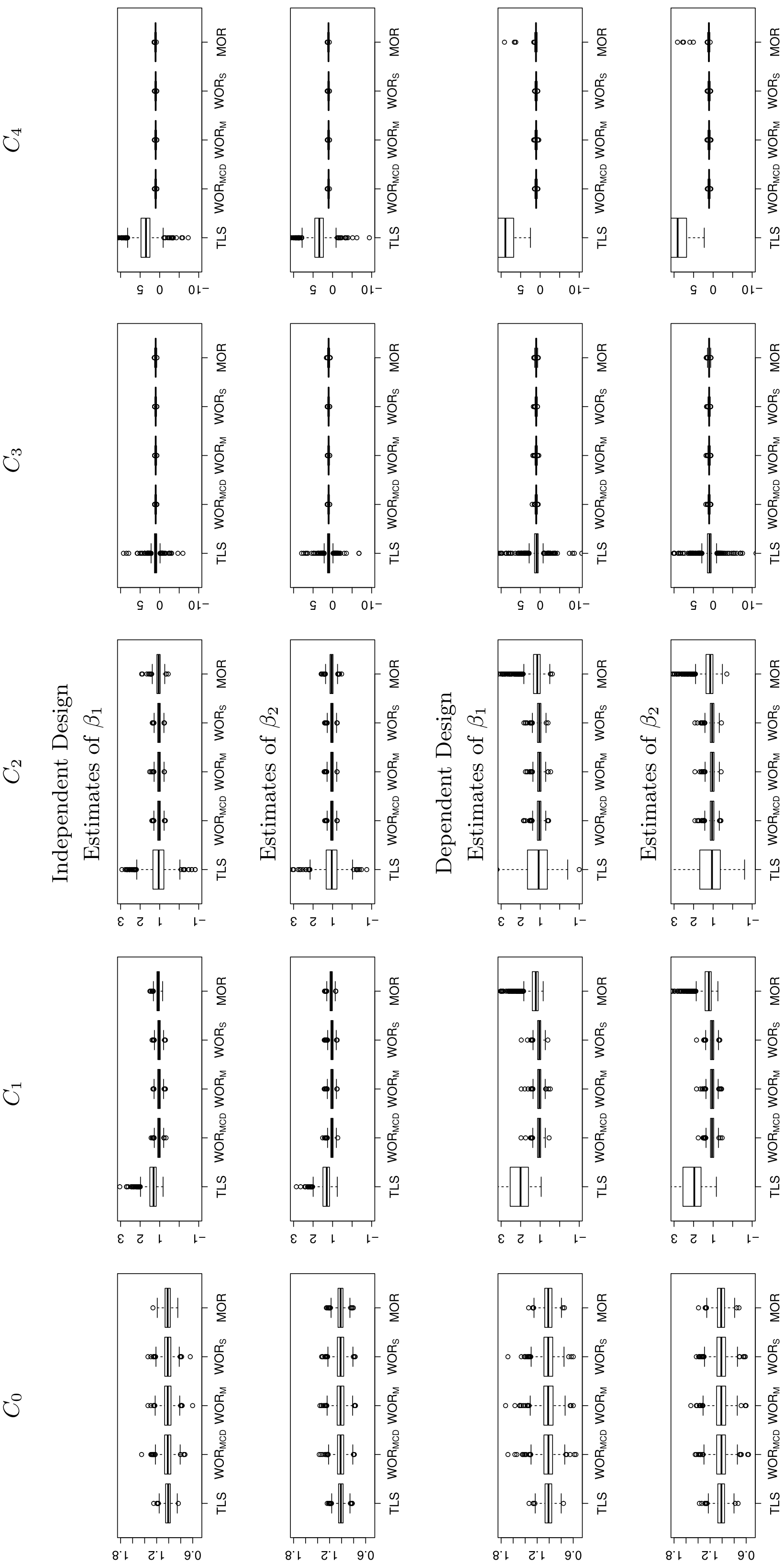

Fig. 3. Boxplots of the estimators of β_1 and β_2.

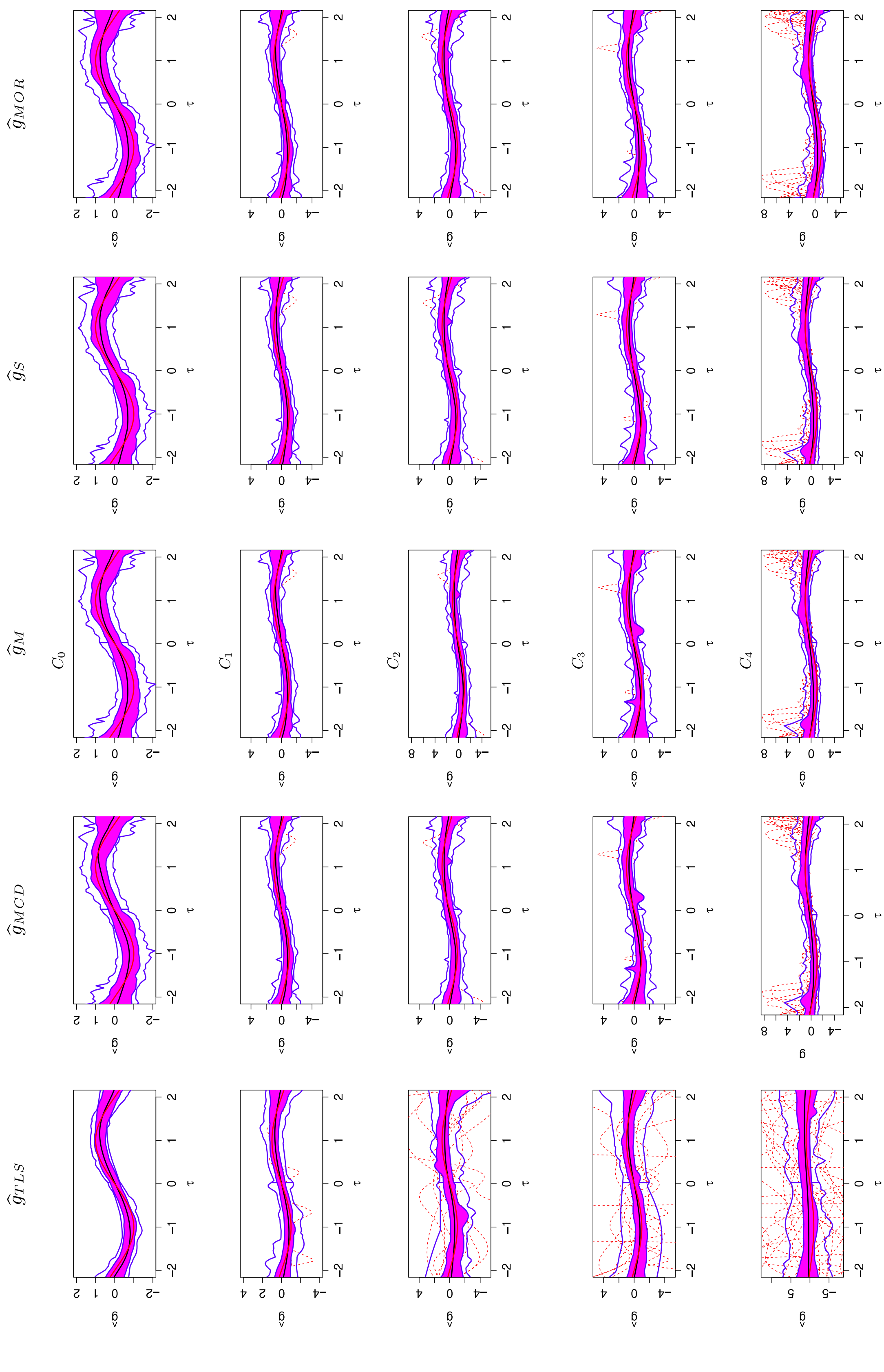

Fig. 4. Functional boxplots of $\widehat{g}(\tau_j)$, $j = 1, \ldots, 100$ with τ_j equally spaced points over $\mathcal{I} = [-2.5, 2.5]$ for the independent design.

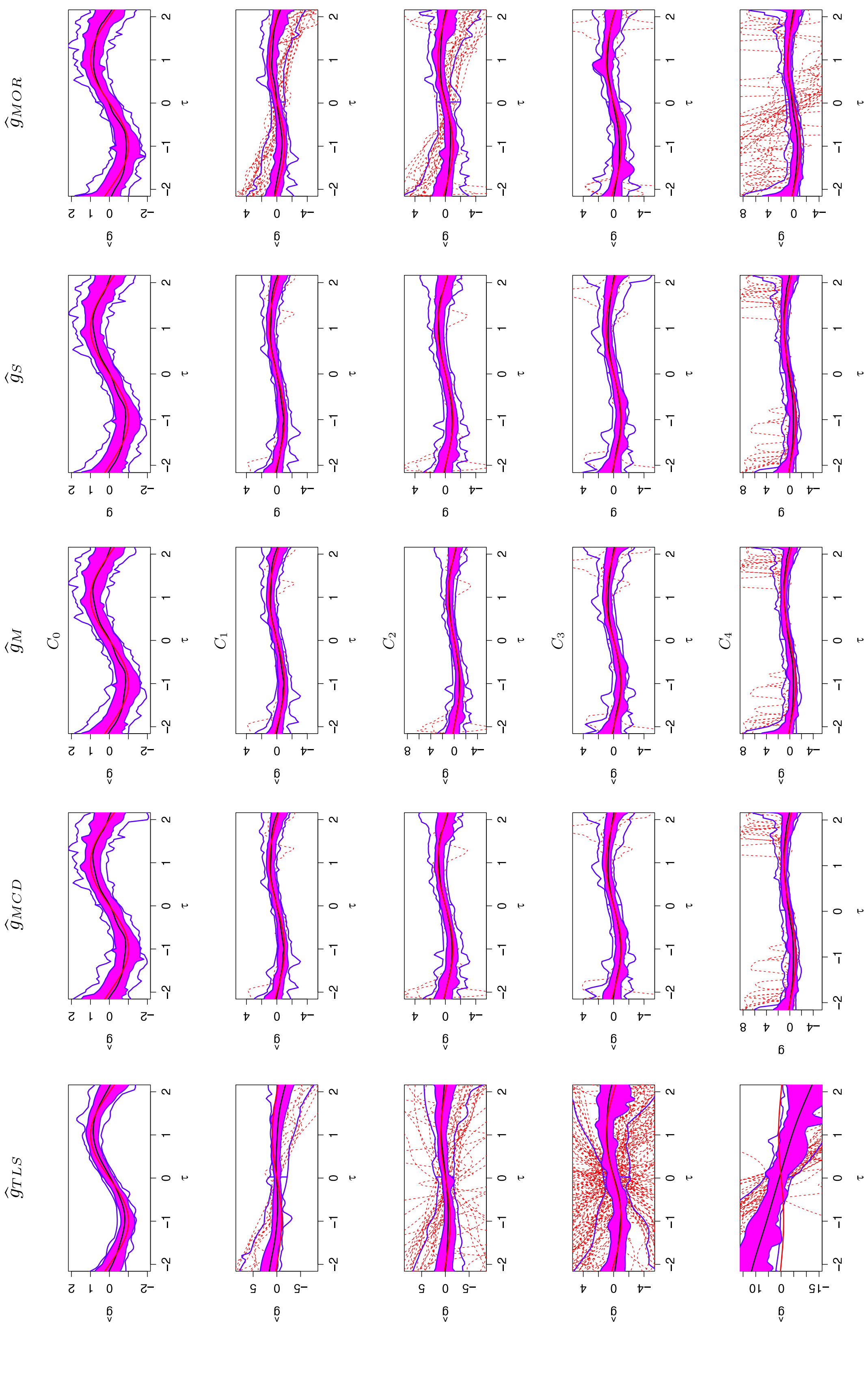

Fig. 5. Functional boxplots of $\widehat{g}(\tau_j), j = 1, \ldots, 100$ with τ_j equally spaced points over $\mathcal{I} = [-2.5, 2.5]$ for the dependent design.

Table 4
Simulation summary: 10% Trimmed mean squared error for the estimators of the regression parameter $\boldsymbol{\beta}$, dependent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0097	0.0169	0.0177	0.0166	0.0112
C_1	2.5142	0.0245	0.0243	0.0233	0.1764
C_2	0.8932	0.0292	0.0280	0.0281	0.1434
C_3	1.3543	0.0177	0.0194	0.0178	0.0341
C_4	151.4738	0.0183	0.0200	0.0181	0.0279

Table 5
Simulation summary: median of MedSE($\widehat{g}$), dependent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0155	0.0273	0.0273	0.0271	0.0267
C_1	0.6399	0.0370	0.0374	0.0364	0.0720
C_2	0.1457	0.0484	0.0474	0.0476	0.0654
C_3	0.1197	0.0408	0.0421	0.0425	0.0432
C_4	33.3647	0.0591	0.0594	0.0580	0.0626

Table 6
Simulation summary: 10% trimmed mean squared error for the estimators of g, dependent case.

	TLS	WOR_{MCD}	WOR_M	WOR_S	MOR
C_0	0.0368	0.0660	0.0664	0.0660	0.0628
C_1	1.5359	0.0908	0.0901	0.0898	0.1918
C_2	0.5775	0.1289	0.1272	0.1277	0.2228
C_3	0.7538	0.1089	0.1099	0.1082	0.1195
C_4	79.9371	0.1751	0.1775	0.1752	0.1849

t is random and it is highly concentrated in the interval $\mathcal{I} = [-2.5, 2.5]$, in order to obtain comparable estimations of g, we consider a grid of equally spaced points $\tau_j, j = 1, \ldots, 100$ in $\mathcal{I}$. Thus, in each replication we estimate $g(\tau_j)$ with each procedure using the optimal bandwidth obtained with the corresponding cross-validation criterion. In the functional boxplots, the area in magenta represents the central region, the dotted red lines correspond to outlying curves and the red line to the true nonparametric function g. In general terms, the functional boxplots show the stability of the robust estimates of g and the strong effect of the considered contaminations on the classical estimators of the nonparametric function. The impact of the contaminations seems to be less pronounced in the independent design than in the dependent one. The effect of the contaminations on the classical estimates is reflected either in the presence of a great number of outlying curves or in the enlargement of the width of the bars of the boxplots or in the trend of these bars. As already mentioned, the asymmetric contamination C_4 breaks down the classical estimator under dependency, indeed a spurious trend appears in the classical estimators of g and most of the estimated curves $\widehat{g}_{TLS}$ do not follow the same direction than the true function g, which is not completely included in the magenta central area. This is also evident from Tables 3 and 6 that show that the trimmed MSE of the classical estimate of g increases more than thirty times when the covariates are independent and more two thousand times when $\mathbf{x}$ and t are related under C_4. With respect to the robust estimates, despite the fact that a few outlying curves appear, the central region in magenta of all the boxplots does always contain the true function g and most curves follow the pattern introduced by the sine function.

The results in the Supplementary material go in the same direction to those described herein (see Appendix B). The dependent structure among the linear covariates, introduced through the covariance matrix, strengthens the effect of the contaminations on the classical estimators, even when $\mathbf{x}$ and t are independent, while the robust ones remain being reliable.

6. Example: *LA* data

Afifi and Azen (1979) consider an epidemiological heart disease study on *LA* County based on 200 employees. Among other variables, age and serum cholesterol levels in 1950 and 1962 were recorded. Buonaccorsi (2010) considers the regression of serum cholesterol level in 1962 ($CS62$) on age (Age) and serum cholesterol in 1950 ($CS50$), assuming that ($CS50$) and ($CS62$) are measured with error, while Age is measured without error.

The left panel of Fig. 6 presents a kernel fit of the response variable $CS62$ on Age and the right one, the scatter plot of $CS62$ versus $CS50$. They show a linear pattern between the two serum cholesterol variables and a nonlinear relationship between $CS62$ and Age, so taking into account all these considerations, we fit to the *LA* data a partially linear EV model. Initially, we compute three estimation procedures: a naive classical method, as if there were no errors-in-variables, based on least squares, the classical estimates based on total least squares described above and the proposed estimator using the weighted orthogonal regression estimator based on S-multivariate estimates of location and scatter as described in Section 5. For this

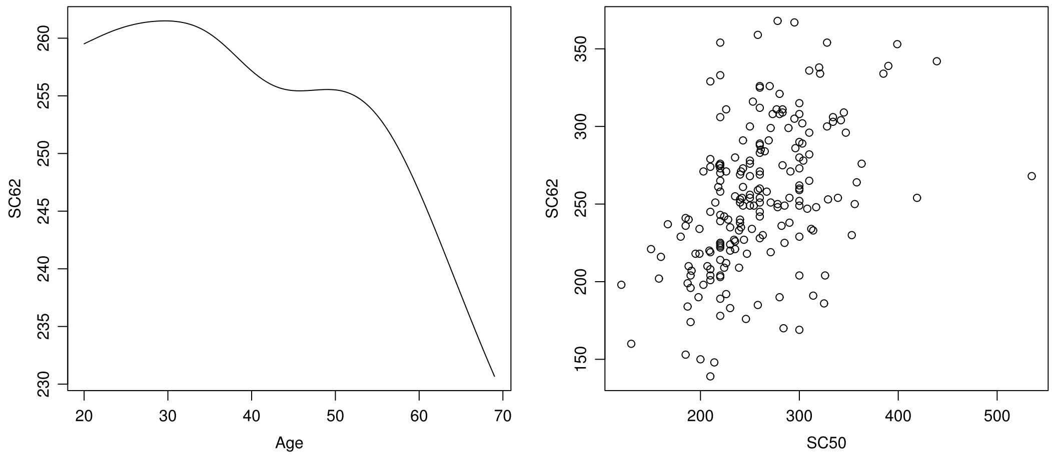

Fig. 6. *LA* data: On the left panel kernel fit of *CS*62 on *Age* and on the right panel scatter of *CS*50 versus *CS*62.

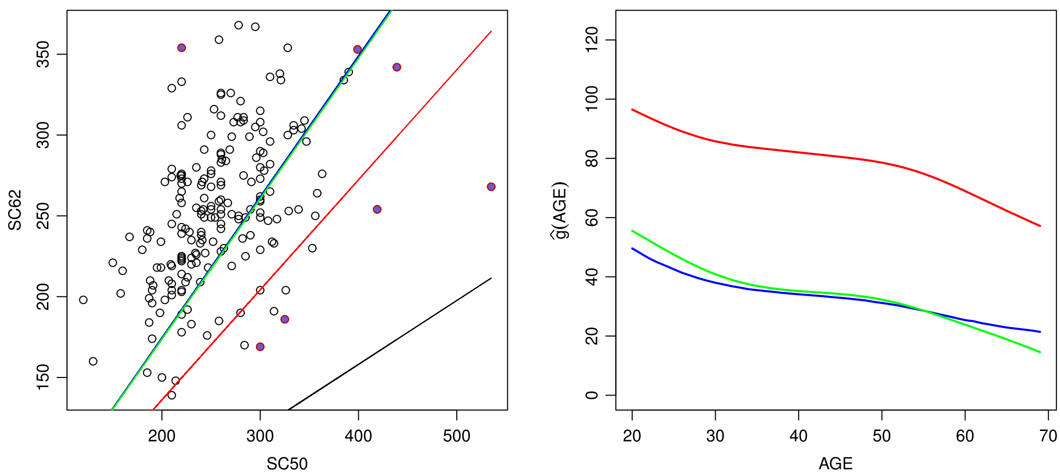

Fig. 7. *LA* data: In the left panel, the line in red corresponds to $\widehat{\beta}_{TLS}$, the line in black to the naive $\widehat{\beta}_{LS}$, the line in blue to $\widehat{\beta}_{ROB}$, while the line in green to $\widehat{\beta}_{TLS}^{-7}$. In the right panel, the curve in red plots the classical $\widehat{g}_{TLS}$, while the curve in blue plots $\widehat{g}_{ROB}$ and the curve in green represents $\widehat{g}_{TLS}^{-7}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Estimators of the regression parameter for the *LA* data.

$\widehat{\beta}_{LS}$	$\widehat{\beta}_{TLS}$	$\widehat{\beta}_{ROB}$	$\widehat{\beta}_{TLS}^{-7}$
0.3953	0.6808	0.8727	0.8689

last estimator, we consider hard rejection weights, which correspond to the indicator function $\eta \equiv \mathcal{I}_{[0,c]}$, where c is the 99% quantile of a distribution χ^2_2.

In a first stage, we select the smoothing parameter for the classical and robust procedures. We consider the classical cross-validation criterion $CV(h)$ given in (20) for the classical estimators and the robust $D(h)$ defined in (21) for our robust proposal. The classical $CV(h)$ is minimized at $h = 17$ and $D(h)$ at $h = 20$. The classical and robust estimators of β obtained using these bandwidths are reported in Table 7 as *TLS* and *ROB*, respectively. The naive least squares estimator computed as if there were no error measurements is also reported and it is labelled as *LS*.

In a second step, we look for those observations downweighted by the weighted orthogonal estimator, they correspond to observations 23, 39, 41, 77, 88, 115 and 137 that are identified as possible outliers. They are plotted in blue in the left panel of Fig. 7, where we show the obtained fits that look remarkably different from each other.

Finally, we remove from the sample these seven identified outliers and we repeat the classical approach based on total least squares with the remaining data. The corresponding estimators are $\widehat{\beta}_{TLS}^{-7}$ and $\widehat{g}_{TLS}^{-7}$, whose fits are represented in green in Fig. 7 together with the classical and robust fits computed from the whole sample. Table 7 shows that the detected atypical points influence the classical estimation procedure. The classical estimators obtained after removing the outliers and the proposed robust estimators computed with the whole sample lead to similar fits as it is shown in Fig. 7.

Acknowledgements

The authors thank the anonymous referee and Associated Editor for their comments and suggestions, that truly contributed to improve the paper. This research was partially supported by Grants 20120130100241BA from Universidad de Buenos Aires, PIP 112-2011-01-00339 from CONICET and PICT 2014-0351 from ANPCYT, Argentina.

Appendix A

Proof of (18). Let $\widehat{\mathbf{a}}_\varepsilon$ be the estimator in the reparametrized form for the new sample. Since $\text{EIF}(\widehat{\boldsymbol\beta}, (y_0, \mathbf{v}_0^{\mathrm T}, t_0)) = \left(\partial\widehat{\boldsymbol\beta}_\varepsilon/\partial\varepsilon\right)|_{\varepsilon=0} = \left(\partial\widehat{\boldsymbol\beta}/\partial\mathbf{a}\right)(\partial\widehat{\mathbf{a}}_\varepsilon/\partial\varepsilon)|_{\varepsilon=0}$, we have that

$$\text{EIF}(\widehat{\boldsymbol\beta}, (y_0, \mathbf{v}_0^{\mathrm T}, t_0)) = \frac{-1}{\widehat{a}_0}\left[\frac{-\widehat{\mathbf{a}}^*}{\widehat{a}_0}\mathbf{I}\right] \qquad \text{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm T}, t_0)),$$

where $\widehat{\mathbf{a}}^* = (\widehat{a}_1, \ldots, \widehat{a}_p)^{\mathrm T}$ and $\text{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm T}, t_0))$ is the empirical influence function of the regression parameter after the reparametrization.

Note that $\widehat{\mathbf{a}}_\varepsilon$ is the solution of

$$\min_{\mathbf{a}} \frac{1-\varepsilon}{n}\sum_{i=1}^n \rho\left(\frac{\mathbf{a}^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma\,\|\mathbf{a}\|}\right) + \varepsilon\rho\left(\frac{\mathbf{a}^{\mathrm T}\widetilde{\mathbf{z}}_{0,\varepsilon}}{\sigma\,\|\mathbf{a}\|}\right), \tag{A.1}$$

with $\widetilde{\mathbf{z}}_{i,\varepsilon} = (y_i, \mathbf{v}_i^{\mathrm T}) - \widehat{\boldsymbol\nu}_\varepsilon^*(t_i)$, $0 \le i \le n$, where $\widehat{\boldsymbol\nu}_\varepsilon^*(\tau)$ are obtained from the new sample. Then, $\widehat{\mathbf{a}}_\varepsilon$ satisfies

$$\left(\mathbf{I} - \frac{\widehat{\mathbf{a}}_\varepsilon}{\|\widehat{\mathbf{a}}_\varepsilon\|}\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}}{\|\widehat{\mathbf{a}}_\varepsilon\|}\right)\left[\frac{(1-\varepsilon)}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma\,\|\widehat{\mathbf{a}}_\varepsilon\|}\right)\widetilde{\mathbf{z}}_{i,\varepsilon} + \varepsilon\,\psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{0,\varepsilon}}{\sigma\,\|\widehat{\mathbf{a}}_\varepsilon\|}\right)\widetilde{\mathbf{z}}_{0,\varepsilon}\right] = \mathbf{0}.$$

We may choose the estimate to have length equal to one, so it verifies that

$$\left(\mathbf{I} - \widehat{\mathbf{a}}_\varepsilon\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\right)\left[\frac{(1-\varepsilon)}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{i,\varepsilon} + \varepsilon\,\psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{0,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{0,\varepsilon}\right] = \mathbf{0}.$$

Now, differentiating with respect to ε we have that

$$\begin{aligned}\mathbf{0} = {} & \left(\mathbf{I} - \widehat{\mathbf{a}}_\varepsilon\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\right)\left[-\frac{1}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{i,\varepsilon} + \psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{0,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{0,\varepsilon}\right] + \frac{(1-\varepsilon)}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}_\varepsilon\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\right)\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{i,\varepsilon}\widetilde{\mathbf{z}}_{i,\varepsilon}^{\mathrm T}\frac{\partial\widehat{\mathbf{a}}_\varepsilon}{\partial\varepsilon}\\ & - \frac{(1-\varepsilon)}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\left(\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}\mathbf{I} + \widehat{\mathbf{a}}_\varepsilon\widetilde{\mathbf{z}}_{i,\varepsilon}^{\mathrm T}\right)\frac{\partial\widehat{\mathbf{a}}_\varepsilon}{\partial\varepsilon} + \frac{(1-\varepsilon)}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}_\varepsilon\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\right)\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{i,\varepsilon}\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\frac{\partial\widetilde{\mathbf{z}}_{i,\varepsilon}}{\partial\varepsilon}\\ & + \frac{(1-\varepsilon)}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}_\varepsilon\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\right)\psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{i,\varepsilon}}{\sigma}\right)\frac{\partial\widetilde{\mathbf{z}}_{i,\varepsilon}}{\partial\varepsilon} + \varepsilon\frac{\partial}{\partial\varepsilon}\left[\psi\left(\frac{\widehat{\mathbf{a}}_\varepsilon^{\mathrm T}\widetilde{\mathbf{z}}_{0,\varepsilon}}{\sigma}\right)\widetilde{\mathbf{z}}_{0,\varepsilon}\right].\end{aligned}$$

Thus, evaluating at $\varepsilon = 0$ we get that

$$\begin{aligned}\mathbf{0} = {} & \left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm T}\right)\left[-\frac{1}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i + \psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_0}{\sigma}\right)\widetilde{\mathbf{z}}_0\right]\\ & + \left[\frac{1}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm T}\right)\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^{\mathrm T} - \frac{1}{n}\sum_{i=1}^n \psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i}{\sigma}\right)\left(\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i\mathbf{I} + \widehat{\mathbf{a}}\widetilde{\mathbf{z}}_i^{\mathrm T}\right)\right]\text{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm T}, t_0))\\ & - \frac{1}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm T}\right)\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widehat{\mathbf{a}}^{\mathrm T}\text{EIF}(\widehat{\boldsymbol\nu}^*, (y_0, \mathbf{v}_0^{\mathrm T}, t_0))(t_i)\\ & - \frac{1}{n}\sum_{i=1}^n\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm T}\right)\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm T}\widetilde{\mathbf{z}}_i}{\sigma}\right)\text{EIF}(\widehat{\boldsymbol\nu}^*, (y_0, \mathbf{v}_0^{\mathrm T}, t_0))(t_i).\end{aligned}$$

Therefore, we have that

$$\mathbf{0} = \left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right) \psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_0}{\sigma}\right)\widetilde{\mathbf{z}}_0$$
$$+ \left[\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\,\mathbf{I} + \widehat{\mathbf{a}}\widetilde{\mathbf{z}}_i^{\mathrm{T}}\right)\right]\mathrm{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))$$
$$- \left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sigma}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widehat{\mathbf{a}}^{\mathrm{T}}\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i) + \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\right].$$

From this expression, if we denote $\mathbf{A} = \frac{1}{\sigma}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\frac{1}{n}\sum_{i=1}^{n}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widetilde{\mathbf{z}}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\,\mathbf{I} + \widehat{\mathbf{a}}\widetilde{\mathbf{z}}_i^{\mathrm{T}}\right)$, we get that

$$\mathrm{EIF}(\widehat{\mathbf{a}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = -\mathbf{A}^{-1}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\left[\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_0}{\sigma}\right)\widetilde{\mathbf{z}}_0 - \frac{1}{\sigma}\frac{1}{n}\sum_{i=1}^{n}\psi'\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\widetilde{\mathbf{z}}_i\widehat{\mathbf{a}}^{\mathrm{T}}\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\right.$$
$$\left. - \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i}{\sigma}\right)\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\right].$$

Hence, we need an expression for $\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau)$ in order to compute the empirical influence function. For each component j, $0 \le j \le p$ we have that

$$0 = \frac{(1-\varepsilon)}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1\left(\frac{z_{ji} - \widehat{\nu}_{j,\varepsilon}(\tau)}{\widehat{s}_{j,\varepsilon}(\tau)}\right) + \varepsilon K\left(\frac{t_0 - \tau}{h}\right)\psi_1\left(\frac{z_{j0} - \widehat{\nu}_{j,\varepsilon}(\tau)}{\widehat{s}_{j,\varepsilon}(\tau)}\right). \tag{A.2}$$

Differentiating (A.2) with respect to ε and evaluating at $\varepsilon = 0$, we obtain

$$0 = -\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right) + K\left(\frac{t_0 - \tau}{h}\right)\psi_1\left(\frac{z_{j0} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)$$
$$- \frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1'\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\frac{1}{\widehat{s}_j(\tau)}\mathrm{EIF}(\widehat{\nu}_j, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau)$$
$$- \frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1'\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j^2(\tau)}\mathrm{EIF}(\widehat{s}_j, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau),$$

and so

$$\mathrm{EIF}(\widehat{\nu}_j, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau) = \left[\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1'\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\frac{1}{\widehat{s}_j(\tau)}\right]^{-1}\left[K\left(\frac{t_0 - \tau}{h}\right)\psi_1\left(\frac{z_{j0} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\right.$$
$$\left. - \mathrm{EIF}(\widehat{s}_j, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(\tau)\;\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1'\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j^2(\tau)}\right]$$
$$= \left[\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t_i - \tau}{h}\right)\psi_1'\left(\frac{z_{ji} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\frac{1}{\widehat{s}_j(\tau)}\right]^{-1}\left[K\left(\frac{t_0 - \tau}{h}\right)\psi_1\left(\frac{z_{j0} - \widehat{\nu}_j(\tau)}{\widehat{s}_j(\tau)}\right)\right].$$

Finally, if we denote $\widehat{\mathbf{a}}^* = (\widehat{a}_1, \ldots, \widehat{a}_p)^{\mathrm{T}}$, we have that

$$\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0)) = \frac{1}{\widehat{a}_0}\left[\frac{-\widehat{\mathbf{a}}^*}{\widehat{a}_0}\;\mathbf{I}\right]\mathbf{A}^{-1}\left(\mathbf{I} - \widehat{\mathbf{a}}\widehat{\mathbf{a}}^{\mathrm{T}}\right)\left[\psi\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_0\right)\widetilde{\mathbf{z}}_0 - \frac{1}{n}\sum_{i=1}^{n}\psi'\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\right)\widetilde{\mathbf{z}}_i\widehat{\mathbf{a}}^{\mathrm{T}}\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\right.$$
$$\left. - \frac{1}{n}\sum_{i=1}^{n}\psi\left(\widehat{\mathbf{a}}^{\mathrm{T}}\widetilde{\mathbf{z}}_i\right)\mathrm{EIF}(\widehat{\boldsymbol{\nu}}^*, (y_0, \mathbf{v}_0^{\mathrm{T}}, t_0))(t_i)\right].$$

Proof of Theorem 3.1. We have that

$$\widehat{P}_n(A) = \frac{1}{n}\sum_{i=1}^{n}I_A(y_i - \widehat{\nu}_o(t_i), \mathbf{v}_i - \widehat{\boldsymbol{\nu}}(t_i)).$$

Hence, since (15) holds, from Lemma 3.2 we get that

$$|E_{\widehat{P}_n}(f) - E_{P_n}(f)| \xrightarrow{\text{c.s.}} 0,$$

where

$$P_n(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(y_i - \nu_o(t_i), \mathbf{v}_i - \boldsymbol{\nu}(t_i)),$$

and thus,

$$\Pi(\widehat{P}_n, P) \xrightarrow{\text{c.s.}} 0.$$

Since $\boldsymbol{\beta}(G)$ is continuous in P,

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}(\widehat{P}_n) \xrightarrow{\text{c.s.}} \boldsymbol{\beta}(P).$$

It only remains to show that $\boldsymbol{\beta}(P) = \boldsymbol{\beta}$. Since $(y_i, \mathbf{x}_i^{\mathrm{T}}, \mathbf{v}_i^{\mathrm{T}}, t_i),\ 1 \le i \le n$, satisfy (1) and $g(\tau) = \nu_o(\tau) - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\nu}(\tau)$, we get that

$$\mathbf{z}_i = \begin{pmatrix} y_i - \nu_o(t_i) \\ \mathbf{v}_i - \boldsymbol{\nu}(t_i) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\nu}(t_i)) \\ \mathbf{x}_i - \boldsymbol{\nu}(t_i) \end{pmatrix} + \begin{pmatrix} e_i \\ \mathbf{e}_{xi} \end{pmatrix}.$$

Therefore, $\mathbf{z}_i$, $1 \le i \le n$, follow model (16) with $\boldsymbol{\beta} = \boldsymbol{\beta}$ and hence, $\beta(P) = \boldsymbol{\beta}$.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2016.09.002.

References

Afifi, A.A., Azen, S.P., 1979. Statistical Analysis: A Computer Oriented Approach. Academic Press, New York and London.
Aït Sahalia, Y., 1995. The delta method for nonparametric kernel functionals (Ph.D. dissertation), University of Chicago.
Bianco, A., Boente, G., 2004. Robust estimators in semiparametric partly linear regression models. J. Statist. Plann. Inference 122, 229–252.
Bianco, A., Boente, G., 2007. Robust estimators under semi-parametric partly linear autoregression: Asymptotic behaviour and bandwidth selection. J. Time Series Anal. 28, 274–306.
Boente, G., Fraiman, R., 1990. Robust nonparametric regression estimation for dependent observations. Ann. Statist. 17, 1242–1256.
Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. Internat. Statist. Rev. 59, 355–372.
Boente, G., Rodriguez, D., 2010. Robust inference in generalized partially linear models. Comput. Statist. Data Anal. 54, 2942–2966.
Buonaccorsi, J.P., 2010. Measurement Error: Models, Methods and Applications. Chapman and Hall /CRC, USA.
Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement Error in Nonlinear Models. Chapman and Hall, London.
Croux, C., Fekri, M., Ruiz-Gazen, A., 2010. Fast and robust estimation of the multivariate errors in variables model. TEST 19, 286–303.
Cui, H., Kong, E., 2006. Empirical likelihood confidence region for parameters in semi-linear errors-in-variables models. Scand. J. Stat. 33, 153–168.
Fekri, M., Ruiz-Gazen, A., 2004. Robust weighted orthogonal regression in the errors-in-variables model. J. Multivariate Anal. 88, 89–108.
Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.
Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383–394.
Härdle, W., 1984. Robust regression function estimation. J. Multivariate Anal. 14, 169–180.
Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Physica–Verlag.
He, X., Liang, H., 2000. Quantile regression estimates for a class of linear and partially linear errorsin-variables models. Statist. Sinica 10, 129–140.
He, X., Zhu, Z., Fung, W., 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika 89, 579–590.
Liang, H., 2000. Asymptotic normality of parametric part in partially linear model with measurement error in the non-parametric part. J. Statist. Plann. Inference 86, 51–62.
Liang, H., Härdle, W., Carroll, R.J., 1999. Estimation in a semiparametric partially linear errors–in–variables model. Ann. Statist. 27, 1519–1535.
Liang, H., Li, R., 2009. Variable selection for partially linear models with measurement errors. J. Amer. Statist. Assoc. 104, 234–248.
Liang, H., Wang, S., Carroll, R.J., 2007. Partially linear models with missing response variables and error-prone covariates. Biometrika 94, 185–198.
Ma, Y., Carroll, R.J., 2006. Locally efficient estimators for semiparametric models with measurement error. J. Amer. Statist. Assoc. 101, 1465–1474.
Mallows, C., 1974. On some topics in robustness. Memorandum, Bell Laboratories, Murray Hill, NJ.
Manchester, L., 1996. Empirical influence for robust smoothing. Austral. J. Statist. 38, 275–296.
Maronna, R., Yohai, V.J., 2010. Correcting *MM*-estimates for fat data sets. J. Comput. Statist. Data Anal. 54, 3168–3173.
Pan, W., Zeng, D., Lin, X., 2008. Estimation in semiparametric transition measurement error models for longitudinal data. Biometrics 65, 728–736.
Severini, T., Staniswalis, J., 1994. Quasi-likelihood estimation in semiparametric models. J. Amer. Statist. Assoc. 89, 501–511.
Tamine, J., 2002. Smoothed influence function: another view at robust nonparametric regression. Discussion paper 62, Sonderforschungsbereich 373, Humboldt-Universitat zu Berlin.
Tukey, J., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.
Zamar, R., 1989. Robust estimation in the errors–in–variables model. Biometrika 76, 149–160.
Zhu, L.X., Cui, H.J., 2003. A semi-parametric regression model with errors in the variables. Scand. J. Stat. 30, 429–442.