Robust inference in generalized partially linear models

Graciela Boente [*], Daniela Rodriguez

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
CONICET, Argentina

ABSTRACT

In many situations, data follow a generalized partly linear model in which the mean of the responses is modeled, through a link function, linearly on some covariates and nonparametrically on the remaining ones. A new class of robust estimates for the smooth function η, associated to the nonparametric component, and for the parameter β, related to the linear one, is defined. The robust estimators are based on a three-step procedure, where large values of the deviance or Pearson residuals are bounded through a score function. These estimators allow us to make easier inferences on the regression parameter β and also improve computationally those based on a robust profile likelihood approach. The resulting estimates of β turn out to be root-n consistent and asymptotically normally distributed. Besides, the empirical influence function allows us to study the sensitivity of the estimators to anomalous observations. A robust Wald test for the regression parameter is also provided. Through a Monte Carlo study, the performance of the robust estimators and the robust Wald test is compared with that of the classical ones.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The generalized linear model (McCullagh and Nelder, 1989) is a popular technique for modeling a wide variety of data and assumes that the observations $(y_i, \mathbf{x}_i, t_i), 1 \leq i \leq n, \mathbf{x}_i \in \mathbb{R}^p, t_i \in \mathbb{R}$, are independent with the same distribution as $(y, \mathbf{x}, t) \in \mathbb{R}^{p+2}$ such that the conditional distribution of $y|(\mathbf{x}, t)$ belongs to the canonical exponential family $\exp[y\theta(\mathbf{x}, t) - B(\theta(\mathbf{x}, t)) + C(y)]$, for known functions B and C. In this situation, the mean $\mu(\mathbf{x}, t) = \mathrm{E}(y|(\mathbf{x}, t))$ is modeled linearly through a known link function, g, i.e., $g(\mu(\mathbf{x}, t)) = \theta(\mathbf{x}, t) = \beta_0 + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \alpha t$. Robust procedures for generalized linear models have been considered among others by Stefanski et al. (1986), Künsch et al. (1989), Bianco and Yohai (1995), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco et al. (2005). Recently, robust tests for the regression parameter under a logistic model were considered by Bianco and Martínez (2009).

As is well known, semiparametric models may be introduced when the linear model is insufficient to explain the relationship between the response variable and its associated covariates. This approach has been used to extend generalized linear models to allow most predictors to be modeled linearly while one or a small number of them enter the model nonparametrically. In this paper, we consider the semiparametric generalized partially linear model, denoted GPLM, that is, $y_i|(\mathbf{x}_i, t_i) \sim F(., \mu_i)$ where $\mathrm{VAR}(y_i|(\mathbf{x}_i, t_i)) = V(\mu_i)$, with V a known function and $\mu_i = \mu(\mathbf{x}_i, t_i)$ such that

$$\mu(\mathbf{x}, t) = \mathrm{E}(y|(\mathbf{x}, t)) = H(\eta(t) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}), \qquad (1)$$

where $H = g^{-1}$ is a known link function, $\boldsymbol{\beta} \in \mathbb{R}^p$ is an unknown parameter and η is an unknown continuous function. As is usual in partially linear models, we will assume that the vector $\mathbf{1}_n$ is not in the space spanned by the column vectors of $\mathbf{x}$, that is, we do not allow $\boldsymbol{\beta}$ to include an intercept so that the model is identifiable. Due to the generality of the semiparametric

* Corresponding address: Instituto de Cálculo, FCEyN, UBA, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina. Tel.: +54 11 45763375; fax: +54 11 45763375.

E-mail addresses: gboente@dm.uba.ar (G. Boente), drodrig@dm.uba.ar (D. Rodriguez).

model (1), identifiability implies that only "slope" coefficients can be estimated. Moreover, we do not allow any linear combination of **x** to be predicted by t, otherwise, the model will be purely nonparametric and $\boldsymbol{\beta}$ will not be identifiable (see Robinson (1988)).

When $H(t) = t$, model GPLM is simply the well known partly linear regression model that has been studied in great depth. We refer, for instance, to Härdle et al. (2000) for a review, to Chang and Qu (2004) for an approach based on wavelets and to Liang (2006) for a comparison of different procedures and a discussion regarding the rate of the smoothing parameter. In a semiparametric setting, outliers can have a devastating effect, since the extreme points can easily affect the scale and the shape of the function estimate of η, leading to possibly wrong conclusions on $\boldsymbol{\beta}$. Robust proposals for this model were introduced among others by Gao and Shi (1997), He et al. (2002) and Bianco and Boente (2004).

For generalized partly linear models, estimators based on the concept of generalized profile likelihood were considered by Severini and Wong (1992) and Severini and Staniswalis (1994); see also Härdle et al. (2006) for a review. The sensitivity to outliers of the classical estimates for these models was described in Boente et al. (2006) where a robust procedure was introduced. On the other hand, a robust generalized estimating equations approach, for GPLM models with clustered data, using regression splines and Pearson residuals was given in He et al. (2005). A related approach for generalized semiparametric mixed models for longitudinal data was considered in Qin and Zhu (2007). The main disadvantage of the estimators proposed by Boente et al. (2006) is that, since they are based on a generalized profile likelihood approach, their asymptotic covariance matrices depend on the derivatives of the robust profile regression function $\eta_{\boldsymbol{\beta}}$ with respect to $\boldsymbol{\beta}$ making difficult their estimation.

In this paper, we introduce a three-step robust procedure to estimate the parameter $\boldsymbol{\beta}$ and the function η, under a GPLM model, which is easier to compute than the one introduced by Boente et al. (2006) and that will allow to make inference on the regression parameter. The proposal is a robustified version of the estimators considered in Carroll et al. (1997). It is shown that the robust estimates of $\boldsymbol{\beta}$ are root-n consistent and asymptotically normal. Through a Monte Carlo study, we compare the performance of these estimators with that of the classical ones. Besides, through their empirical influence function we study the sensitivity of the estimators to anomalous observations. A robust procedure to test the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is also discussed. The paper is organized as follows. The robust proposal is given in Section 2, its consistency is derived in Section 3. The asymptotic distribution of the regression estimators and a robust Wald test for the regression parameter are provided in Section 4, while an expression for the empirical influence function is obtained in Section 5. The results of a Monte Carlo study are summarized in Section 6.

2. The proposal

Let $(y_i, \mathbf{x}_i, t_i) \in \mathbb{R}^{p+2}$ be independent observations such that $y_i|(\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$ with $\mu_i = H(\eta(t_i) + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ and $\mathrm{VAR}\,(y_i|(\mathbf{x}_i, t_i)) = V(\mu_i)$. Let $\eta_0(\tau)$ and $\boldsymbol{\beta}_0$ denote the true parameter values and E_0 the expectation under the true model, thus $\mathrm{E}_0(y|(\mathbf{x}, t)) = H(\eta_0(t) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0)$. Let $w_1 : \mathbb{R}^p \to \mathbb{R}$ be a weight function to control leverage points on the carriers $\mathbf{x}$ and $\rho : \mathbb{R}^2 \to \mathbb{R}$ a loss function. Define

$$S_n(a, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{n} W_i(\tau)\rho\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x}_i), \tag{2}$$

$$S(a, \boldsymbol{\beta}, \tau) = \mathrm{E}_0\left[\rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x})|t = \tau\right], \tag{3}$$

where $W_i(\tau)$ are weights depending on the closeness of t_i to τ and which will be taken as the kernel weights, for the sake of simplicity, i.e.,

$$W_i(\tau) = K\left(\frac{\tau - t_i}{h}\right)\left\{\sum_{j=1}^{n} K\left(\frac{\tau - t_j}{h}\right)\right\}^{-1}.$$

Let us assume that $w_1(\cdot)$ and $\rho(\cdot)$ are such that $S(\eta_0(\tau), \boldsymbol{\beta}_0, \tau) = \min_{a,\boldsymbol{\beta}} S(a, \boldsymbol{\beta}, \tau)$, then in order to estimate $\eta_0(\tau)$ and $\boldsymbol{\beta}_0$ one can minimize $S_n(a, \boldsymbol{\beta}, \tau)$ that provides, under mild conditions, a consistent estimator of $S(a, \boldsymbol{\beta}, \tau)$. It is worth noting that, with such a choice, the estimators of $\boldsymbol{\beta}_0$ will not have a root-n rate of convergence. In order to provide $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}_0$, let us define for each $\boldsymbol{\beta} \in \mathbb{R}^p$ and any continuous function $v : \mathbb{R} \to \mathbb{R}$,

$$L_n(\boldsymbol{\beta}, v) = \frac{1}{n} \sum_{i=1}^{n} \rho\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + v(t_i)\right) w_2(\mathbf{x}_i), \tag{4}$$

$$L(\boldsymbol{\beta}, v) = \mathrm{E}_0\left[\rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + v(t)\right) w_2(\mathbf{x})\right], \tag{5}$$

where $w_2(\cdot)$ is again a weight function decreasing the effect of high leverage points.

Throughout the paper, we will assume Fisher-consistency, i.e., that $L(\boldsymbol{\beta}_0, \eta_0) = \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \eta_0)\boldsymbol{\beta}_0$ being the unique minimum (see Remark 2.1 below). The estimators can thus be defined as

• *Step* 1: For each fixed τ, let

$$\left(\widetilde{\eta}(\tau), \widetilde{\boldsymbol{\beta}}(\tau)\right) = \underset{a,\boldsymbol{\beta}}{\mathrm{argmin}}\, S_n(a, \boldsymbol{\beta}, \tau). \tag{6}$$

- *Step* 2: Define the estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ as

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, L_n(\boldsymbol{\beta}, \widetilde{\eta}). \tag{7}$$

- *Step* 3: Define the final estimator $\widehat{\eta}(\tau)$ of $\eta(\tau)$ as

$$\widehat{\eta}(\tau) = \underset{a}{\operatorname{argmin}}\, S_n(a, \widehat{\boldsymbol{\beta}}, \tau). \tag{8}$$

Step 3 is introduced in order to improve the performance of the regression function estimator. In those cases in which the regression function plays the role of a nuisance parameter, this last step can be avoided. This approach improves computationally the proposal given by Boente et al. (2006). Effectively, these authors considered a robust profile likelihood approach defining for each $\boldsymbol{\beta} \in \mathbb{R}^p$, $\widehat{\eta}_{\boldsymbol{\beta}}(\tau) = \operatorname{argmin}_a S_n(a, \boldsymbol{\beta}, \tau)$. The final estimator of $\boldsymbol{\beta}$ satisfies $\widehat{\boldsymbol{\beta}}_{\text{BHZ}} = \operatorname{argmin}_{\boldsymbol{\beta}} L_n(\boldsymbol{\beta}, \widehat{\eta}_{\boldsymbol{\beta}})$, while that of η turns out to be $\widehat{\eta}_{\text{BHZ}} = \widehat{\eta}_{\widehat{\boldsymbol{\beta}}_{\text{BHZ}}}$. Therefore, to compute the estimators defined in Boente et al. (2006) the functions $\widehat{\eta}_{\boldsymbol{\beta}}$ need to be computed at each data point of the sample t_i over a set of candidates $\boldsymbol{\beta}_j$ to compute the value of the objective function $L_n(\boldsymbol{\beta}_j, \widehat{\eta}_{\boldsymbol{\beta}_j})$ which increases the computing time. Note that the estimator computed in Step 3, is simply $\widehat{\eta} = \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}$. However, the estimator $\widehat{\boldsymbol{\beta}}$ differs of $\widehat{\boldsymbol{\beta}}_{\text{BHZ}}$ since the first one is computed using an initial estimator of the regression function η while the latter uses an estimator of the least favorable function. For instance, if robustified Pearson residuals are considered, our proposal computes pseudo-residuals using a preliminary estimator of η and then, fits the pseudo-residuals versus **x** while the robust profile approach tries to fit simultaneously both components by fitting first the nuisance parameter for each fixed $\boldsymbol{\beta}$. Indeed, in our proposal the estimator $\widetilde{\eta}$ is computed once in the first step and thus, a complete search over a grid of values for $\boldsymbol{\beta}$ involving a minimization of S_n to compute $\widehat{\eta}_{\boldsymbol{\beta}}$ for each of them is avoided. In this sense, our method is based on the fact that conditionally on t, the model can be parametrized by a finite-dimensional parameter and so, we have a *conditionally parametric* model as defined in Severini and Wong (1992). Instead of providing a consistent estimator of the *robustified least favorable curve* $\eta_{\boldsymbol{\beta}}$ to obtain a robust *nearly efficient* estimator of $\boldsymbol{\beta}$ maximizing a robust version of the generalized profile log-likelihood, our procedure takes advantage of the finite-dimensional structure conditional on t and obtains a consistent estimator of the nuisance parameter that simplifies the numerical complexity. Our procedure can be thought as a robust backfitting procedure related to the local linear backfitting method introduced by Opsomer and Ruppert (1999) in partially linear additive models.

As in Boente et al. (2006), a robust cross-validation procedure to select the smoothing parameter h can also be considered.

When ρ is continuously differentiable, if we denote by $\Psi(y, u) = \partial \rho(y, u)/\partial u$, $(\boldsymbol{\beta}_0, \eta_0(\tau))$ and $(\widetilde{\boldsymbol{\beta}}(\tau), \widetilde{\eta}(\tau))$ satisfy the differentiated equations $S^1(a, \boldsymbol{\beta}, \tau) = 0$ and $S_n^1(a, \boldsymbol{\beta}, \tau) = 0$, respectively, where

$$S^1(a, \boldsymbol{\beta}, \tau) = \mathrm{E}_0\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x})\mathbf{z} | t = \tau\right), \tag{9}$$

$$S_n^1(a, \boldsymbol{\beta}, \tau) = \sum_{i=1}^n W_i(\tau)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x}_i)\mathbf{z}_i, \tag{10}$$

and $\mathbf{z} = (1, \mathbf{x}^{\mathrm{T}})^{\mathrm{T}}$. On the other hand, the regression estimator $\widehat{\boldsymbol{\beta}}$, is a solution of $L_n^1(\boldsymbol{\beta}, \widetilde{\eta}) = 0$ while $\boldsymbol{\beta}_0$ solves $L^1(\boldsymbol{\beta}_0, \eta_0) = 0$ where

$$L^1(\boldsymbol{\beta}, v) = \mathrm{E}_0\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + v(t)\right) w_2(\mathbf{x})\mathbf{x}\right), \tag{11}$$

$$L_n^1(\boldsymbol{\beta}, v) = \frac{1}{n}\sum_{i=1}^n \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + v(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i. \tag{12}$$

When $L_n(\boldsymbol{\beta}, \widetilde{\eta})$ has only one critical point, i.e., when the equation $L_n^1(\boldsymbol{\beta}, \widetilde{\eta}) = 0$ has only one root, corresponding to the minimum of $L_n(\boldsymbol{\beta}, \widetilde{\eta})$, the estimator $\widehat{\boldsymbol{\beta}}$ can be computed using a Newton–Raphson approach and initiating the iterative procedure with $\widehat{\boldsymbol{\beta}}_{\text{INI}} = \sum_{i=1}^n \widetilde{\boldsymbol{\beta}}(t_i)/n$.

Boente et al. (2006) proposed two classes of loss functions ρ. The first one aims to bound the deviances, while the second one introduced by Cantoni and Ronchetti (2001) bounds the Pearson residuals. For the sake of completeness, we recall their definition.

The first class of loss function takes the form of

$$\rho(y, u) = \phi[-\ln f(y, H(u)) + A(y)] + G(H(u)), \tag{13}$$

where ϕ is a bounded nondecreasing function with continuous derivative φ, and $f(\cdot, s)$ is the density of the distribution function $F(\cdot, s)$ with $y|(\mathbf{x}, t) \sim F\left(\cdot, H\left(\eta_0(t) + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0\right)\right)$. To avoid triviality, it is assumed that ϕ is non-constant in a positive probability set. Typically, ϕ is a function performing like the identity function in a neighborhood of 0. The function $A(y)$ is typically used to remove a term from the log-likelihood that is independent of the parameter, and can be defined as

$A(y) = \ln(f(y, y))$ in order to get the deviance. The correction term G is used to guarantee the Fisher-consistency, and satisfies

$$G'(s) = \int \varphi[-\ln f(y, s) + A(y)] f'(y, s) d\mu(y)$$
$$= E_s \left(\varphi[-\ln f(y, s) + A(y)] f'(y, s) / f(y, s) \right),$$

where E_s indicates expectation taken under $y \sim F(\cdot, s)$ and $f'(y, s)$ is shorthand for $\partial f(y, s)/\partial s$. With this class of ρ functions, we call the resulting estimator a *modified likelihood estimator* denoted MOD. Note that, when considering generalized linear models, the maximum likelihood estimator corresponds to the choice $\phi(t) = t, A(y) = \ln(f(y, y)), G(u) = 0$ and $w_1 = w_2 \equiv 1$. The estimators based on this choice for the loss function ρ will be denoted as DEV and may be considered as the classical counterpart of the MOD.

In a logistic regression setting, in order to guarantee existence of solution, Croux and Haesbroeck (2002) proposed using the score function

$$\phi(t) = \begin{cases} t \exp(-\sqrt{c}) & \text{if } t \leq c \\ -2(1 + \sqrt{t}) \exp(-\sqrt{t}) + \left(2(1 + \sqrt{c}) + c\right) \exp(-\sqrt{c}) & \text{otherwise.} \end{cases}$$

It is worth noting that, when considering the deviance and a continuous family of distributions with strongly unimodal density function, the correction term G can be avoided, as discussed in Bianco et al. (2005).

The second class of loss functions is based on the proposal given by Cantoni and Ronchetti (2001) for generalized linear models, where they consider a general class of M-estimators of Mallows type, by bounding separately the influence of deviations on y and $(\mathbf{x}, t)$. Their approach is based on robustifying the quasi-likelihood, which is an alternative to the generalizations given for generalized linear regression models by Stefanski et al. (1986) and Künsch et al. (1989). Let $r(y, \mu) = (y - \mu) V^{-1/2}(\mu)$ be the Pearson residuals with VAR $(y_i|(\mathbf{x}_i, t_i)) = V(\mu_i)$. Denote $v(y, \mu) = V^{-1/2}(\mu)\psi_c(r(y, \mu))$, with ψ_c an odd nondecreasing score function with tuning constant c, such as the Huber function, and

$$\rho(y, u) = -\left[\int_{s_0}^{H(u)} v(y, s) ds + G(H(u)) \right], \tag{14}$$

where s_0 is such that $v(y, s_0) = 0$. To ensure Fisher-consistency, the correction term $G(s)$ satisfies $G'(s) = -E_s(v(y, s))$. With such a ρ function, we call the resulting estimator a *robust quasi-likelihood estimator*. For the Binomial and Poisson families, explicit forms of the correction term $G(s)$ are given in Cantoni and Ronchetti (2001). The classical counterpart of this approach corresponds to the choice $\psi_c(u) = u, w_1 = w_2 \equiv 1$, and the estimators will be denoted as QAL since they are based on the quasi-likelihood.

Remark 2.1. Let $\mathcal{T}$ be the support of the random variable t. Under a logistic partially linear regression model, Fisher-consistency can easily be derived for the loss function given by (13), when ϕ satisfies the regularity conditions stated in Bianco and Yohai (1995) and

$$P\left(\mathbf{x}^T \boldsymbol{\beta} = \alpha | t = \tau\right) < 1, \quad \forall (\boldsymbol{\beta}, \alpha) \neq 0 \text{ and } \tau \in \mathcal{T}. \tag{15}$$

Moreover, it is easy to verify that $\boldsymbol{\beta}_0$ is the unique minimizer of $L(\boldsymbol{\beta}, \eta_0)$ in this case. The same assertion can be verified for the robust quasi-likelihood proposal if ψ_c is bounded and increasing.

Condition (15) does not allow $\boldsymbol{\beta}_0$ to include an intercept, so that the model will be identifiable, as mentioned in the Introduction.

Under a generalized partially linear model with the response having a gamma distribution with a fixed shape parameter, Theorem 1 of Bianco et al. (2005) allows us to derive Fisher-consistency for the nonparametric and parametric components, if the score function ϕ is bounded and strictly increasing on the set where it is not constant and if (15) holds.

Remark 2.2. As for generalized linear models, the correction factor, denoted $G(s)$, is included to guarantee Fisher-consistency under the true model. Otherwise, one can only ensure that the estimators will be consistent to the solution $(\boldsymbol{\beta}(F), \eta(F, \tau))$ of the related functional equations. To be more precise, let $(\boldsymbol{\beta}^\star(F, \tau), \eta^\star(F, \tau)) = \text{argmin}_{a, \boldsymbol{\beta}} S(a, \boldsymbol{\beta}, \tau)$ and denote by $\boldsymbol{\beta}(F) = \text{argmin}_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \eta^\star(F, \cdot))$ and $\eta(F, \tau) = \text{argmin}_a S(a, \boldsymbol{\beta}(F), \tau)$ where $S(a, \boldsymbol{\beta}, \tau)$ and $L(\boldsymbol{\beta}, v)$ are defined in (3) and (5), respectively. The results stated in Section 3 can be easily adapted to show that $\widehat{\boldsymbol{\beta}} \xrightarrow{\text{a.s.}} \boldsymbol{\beta}(F)$ and $\|\widehat{\eta} - \eta(F, \cdot)\|_{0,\infty} \xrightarrow{\text{a.s.}} 0$ while the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ can also be derived by centering with $\boldsymbol{\beta}(F)$. On the other hand, as it is well known, when $H(u) = u$, i.e., under the partially linear model $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \eta(t_i) + \epsilon_i$, Fisher-consistency holds if, for instance, the errors ϵ_i have a symmetric distribution and the score function ψ is odd.

3. Consistency

In this section, we will derive, under some regularity conditions, the consistency of the estimators defined through (6) and (7). Note that consistency of $\widehat{\boldsymbol{\beta}}$ implies the consistency of $\widehat{\eta}$ using Theorem 3.1 of Boente et al. (2006). We will assume

that $t \in \mathcal{T} \subset \mathbb{R}$. For any continuous function $v : \mathcal{T}_0 \to \mathbb{R}$ denote $\|v\|_{0,\infty} = \sup_{\tau \in \mathcal{T}_0} |v(\tau)|$ and $\|v\|_{\infty} = \sup_{\tau \in \mathcal{T}} |v(\tau)|$ where $\mathcal{T}_0 \subset \mathcal{T}$ is a compact set.

C1. The function $\rho(y, a)$ is continuous and bounded, and the functions $w_1(\cdot)$ and $w_2(\cdot)$ are non-negative bounded functions.
C2. The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even, non-negative, continuous and bounded function, with bounded variation, satisfying $\int K(u)\mathrm{d}u = 1$, $\int u^2 K(u)\mathrm{d}u < \infty$ and $|u|K(u) \to 0$ as $|u| \to \infty$.
C3. The bandwidth sequence h_n is such that $h_n \to 0$, $nh_n / \log(n) \to \infty$.
C4. The marginal density f_T of t is a bounded function in $\mathcal{T}$. Moreover, given any compact set $\mathcal{T}_0 \subset \mathcal{T}$, there exists a positive constant $A_1(\mathcal{T}_0)$ such that $A_1(\mathcal{T}_0) < f_T(\tau)$ for all $\tau \in \mathcal{T}_0$.
C5. The function $S(a, \boldsymbol{\beta}, \tau)$ satisfies the following equicontinuity condition: given $\mathcal{T}_0 \subset \mathcal{T}$, $\mathcal{K} \subset \mathbb{R}^p$ compact sets, for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $\tau_1, \tau_2 \in \mathcal{T}_0$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{K}$,

$$|\tau_1 - \tau_2| < \delta \quad \text{and} \quad \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta \Rightarrow \sup_{a \in \mathbb{R}} |S(a, \boldsymbol{\beta}_1, \tau_1) - S(a, \boldsymbol{\beta}_2, \tau_2)| < \epsilon.$$

C6. The functions $S(a, \boldsymbol{\beta}, \tau)$ and $\eta_0(\tau)$ are continuous.

Lemma 3.1. *Let $\mathcal{K} \subset \mathbb{R}^p$ and $\mathcal{T}_0 \subset \mathcal{T}$ compact sets such that $\mathcal{T}_\delta \subset \mathcal{T}$ with $\mathcal{T}_\delta$ the closure of a δ-neighborhood of $\mathcal{T}_0$. Assume that C1 to C6 hold and that the family of functions $\mathcal{F} = \{f(y, \mathbf{x}) = \rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, a \in \mathbb{R}\}$ has covering number $\sup_{\mathbb{Q}} N\left(\epsilon, \mathcal{F}, L^1(\mathbb{Q})\right) \leq A\epsilon^{-W}$ for any $0 < \epsilon < 1$. Then, we have that $\|\widetilde{\eta} - \eta_0\|_{0,\infty} \xrightarrow{\text{a.s.}} 0$ and $\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_{0,\infty} \xrightarrow{\text{a.s.}} 0$.*

Theorem 3.1. *Let $\widehat{\boldsymbol{\beta}}$ be defined in (7). Assume that C1 holds, $\Psi(y, u) = \partial\rho(y, u)/\partial u$ is bounded and that $\widetilde{\eta}$ verifies*

$$\|\widetilde{\eta} - \eta_0\|_{0,\infty} \xrightarrow{\text{a.s.}} 0. \tag{16}$$

Moreover, assume that the family of functions $\mathcal{H} = \{f_{\boldsymbol{\beta}}(y, \mathbf{x}, t) = \rho(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + \eta_0(t))w_2(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p\}$ has finite bracketing number, $N_{[\,]}\left(\epsilon, \mathcal{H}, L^1(P)\right) < \infty$, for any $0 < \epsilon < 1$, where P is the distribution of $(y_1, \mathbf{x}_1, t_1)$ or that $\log N\left(\epsilon, \mathcal{H}, L^1(P_n)\right) = o_P(n)$ with P_n the empirical distribution. Then,

(a) $\sup_{\boldsymbol{\beta} \in \mathbb{R}^p} |L_n(\boldsymbol{\beta}, \widetilde{\eta}) - L(\boldsymbol{\beta}, \eta_0)| \xrightarrow{\text{a.s.}} 0$.

(b) *If $L(\boldsymbol{\beta}, \eta_0)$ has a unique minimum in $\boldsymbol{\beta}_0$ and $\liminf_{\|\boldsymbol{\beta}\| \to \infty} L(\boldsymbol{\beta}, \eta_0) > L(\boldsymbol{\beta}_0, \eta_0)$, then $\widehat{\boldsymbol{\beta}} \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0$.*

We omit the proofs of Lemma 3.1 and Theorem 3.1 since they follow arguing as in Boente et al. (2006) using Lemma A.1 in Carroll et al. (1997). Details can be found in Boente and Rodriguez (2008).

Remark 3.1. The condition that the family of functions $\mathcal{F} = \{f(y, \mathbf{x}) = \rho\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x}), \boldsymbol{\beta} \in \mathbb{R}^p, a \in \mathbb{R}\}$ has covering number $N\left(\epsilon, \mathcal{F}, L^1(\mathbb{Q})\right) \leq A\epsilon^{-W}$ is fulfilled for the estimators defined through (13) if the function $\phi(s)$ and $G(H(s))$ are of bounded variation and if the densities are such that the covering number of the class $\mathcal{F}_0 = \{g(y, \mathbf{x}) = \ln f\left(y, H\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)\right), \boldsymbol{\beta} \in \mathbb{R}^p, a \in \mathbb{R}\}$ grows at a polynomial rate, i.e., it is bounded by $A\epsilon^{-W}$. For the score functions ϕ usually considered in robustness, such as Tukey's biweight function or the score function introduced in Croux and Haesbroeck (2002) for the logistic model, ϕ and $G(H(s))$ have bounded variation. On the other hand, for the logistic and Gamma model to be considered below, it is easy to see that for $r \geq 1$, $N(\epsilon, \mathcal{F}_0, L^r(P))$ grows at polynomial rate, since the family of functions $\{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a, \boldsymbol{\beta} \in \mathbb{R}^p, a \in \mathbb{R}\}$ has finite dimension. Using that $N(\epsilon, \mathcal{H}_1 + \mathcal{H}_2, L^r(P)) \leq N(\epsilon, \mathcal{H}_1, L^r(P)) N(\epsilon, \mathcal{H}_2, L^r(P))$, the desired result is easily derived. On the other hand, since $N_{[\,]}\left(\epsilon, \mathcal{H}, L^1(P)\right) \leq N(\epsilon, \mathcal{H}, L^\infty(P))$, using similar arguments one can derive that the conditions required in Theorem 3.1 are usually fulfilled.

4. Asymptotic distribution and tests on the regression parameter

In this section, we derive under mild conditions the asymptotic distribution of the regression parameter estimator defined in Section 2. The obtained asymptotic distribution can be used to construct a Wald-type statistic to make inferences on the regression parameter, that is, when we want to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$.

4.1. Asymptotic normality

Throughout this section we will assume that $\mathcal{T}$ is a compact set. We begin by fixing some notation. For any symmetric matrix $\mathbf{B} \in \mathbb{R}^{q \times q}$, we denote by $\lambda_1(\mathbf{B}) \geq \lambda_2(\mathbf{B}) \geq \cdots \geq \lambda_q(\mathbf{B})$ the eigenvalues in decreasing order. Let $(y, \mathbf{x}, t)$ be a random vector with the same distribution as $(y_1, \mathbf{x}_1, t_1)$ and denote

$$\chi(y, a) = \frac{\partial}{\partial u}\Psi(y, u) \quad \text{and} \quad \chi_1(y, a) = \frac{\partial^2}{\partial u^2}\Psi(y, u).$$

Let $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ be defined as

$$\boldsymbol{\Sigma} = \mathrm{E}_0\left\{\Psi^2\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right)\left[w_2(\mathbf{x})\mathbf{x} + D(\mathbf{x}, t)\boldsymbol{\gamma}(t)f_T(t)\right]\left[w_2(\mathbf{x})\mathbf{x} + D(\mathbf{x}, t)\boldsymbol{\gamma}(t)f_T(t)\right]^{\mathrm{T}}\right\},$$

where $\boldsymbol{\gamma}(\tau) = \mathrm{E}_0(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_2(\mathbf{x})\mathbf{x}f_T^{-1}(t)|t = \tau)$, $D(\mathbf{x}, \tau)$ is the first element of $\mathbf{A}(\tau)^{-1}w_1(\mathbf{x})\mathbf{z}$, with $\mathbf{A}(\tau) = \mathrm{E}_0\left(\chi(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t))w_1(\mathbf{x})\mathbf{z}\mathbf{z}^{\mathrm{T}}|t = \tau\right)$ and $\mathbf{z} = (1, \mathbf{x}^{\mathrm{T}})^{\mathrm{T}}$. Denote by $G_i(\tau) = G(y_i, \mathbf{x}_i, \tau) = \Psi(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\tau))D(\mathbf{x}_i, \tau)$, i.e., the first element of $\mathbf{A}(\tau)^{-1}\Psi(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\tau))w_1(\mathbf{x}_i)\mathbf{z}_i$.

N1. η_0 and $\widetilde{\eta}$ are continuously differentiable functions, such that $\|\widetilde{\eta}' - \eta_0'\|_\infty \xrightarrow{p} 0$.

N2. The functions Ψ, χ, χ_1, w_1, w_2, $\psi_1(\mathbf{x}) = \mathbf{x}w_1(\mathbf{x})$ and $\psi_2(\mathbf{x}) = \mathbf{x}w_2(\mathbf{x})$ are bounded and continuous. Besides, $w_1(\cdot)$ and $w_2(\cdot)$ are non-negative functions.

N3. (a) The matrix $\mathbf{A} = \mathrm{E}_0\left[\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right)\mathbf{x}\mathbf{x}^{\mathrm{T}}w_2(\mathbf{x})\right]$ is non-singular.
 (b) The matrix $\mathbf{A}(\tau)$ defined above is non-singular. Moreover, $\inf_{\tau \in \mathcal{T}} \lambda_{p+1}(\mathbf{A}(\tau)) > 0$.

N4. $\boldsymbol{\Sigma}$ is a positive definite matrix.

N5. $\mathrm{E}_0\left\{\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right)|(\mathbf{x}, t)\right\} = 0$.

N6. $\mathrm{E}_0\left(w_2(\mathbf{x})\|\mathbf{x}\|^2\right) < \infty$.

N7. f_T the marginal density of t is bounded continuously differentiable with bounded continuous derivative. Besides, $\inf_{\tau \in \mathcal{T}} f_T(\tau) > 0$.

N8. (a) $m_1(\tau_1, \tau_2) = \mathrm{E}_0(G_1(t_2)|t_1 = \tau_1, t_2 = \tau_2)$, $m_1'(\tau_1, \tau_2) = \partial m_1(\tau_1, \tau_2)/\partial\tau_2$ and $m_1''(\tau_1, \tau_2) = \partial^2 m_1(\tau_1, \tau_2)/\partial\tau_2^2$ are bounded continuous functions.
 (b) $m_{is}(\tau_1, \tau_2, \tau_3, \tau_4) = \mathrm{E}_0([G_i(t_j) - G_i(t_i)][G_s(t_\ell) - G_s(t_s)]|t_i = \tau_1, t_j = \tau_2, t_s = \tau_3, t_\ell = \tau_4)$, $m_{is,\ell}'(\tau_1, \tau_2, \tau_3, \tau_4) = \partial m_{is}(\tau_1, \tau_2, \tau_3, \tau_4)/\partial\tau_\ell$ and $m_{is,\ell r}''(\tau_1, \tau_2, \tau_3, \tau_4) = \partial^2 m_{is}(\tau_1, \tau_2, \tau_3, \tau_4)/\partial\tau_\ell\partial\tau_r$ are bounded continuous functions.
 (c) $\boldsymbol{\gamma}(\tau)$ is continuously differentiable with continuous derivative.

N9. The kernel $K : \mathbb{R} \to \mathbb{R}$ is an even non-negative bounded function with bounded variation. Moreover, it satisfies a Lipschitz condition of order one and $\int K(u)\mathrm{d}u = 1$, $\int uK(u)\mathrm{d}u = 0$ and $\int u^2 K(u)\mathrm{d}u < \infty$.

In order to show that $\widehat{\boldsymbol{\beta}}$ is asymptotically normally distributed, we will need the following lemma whose proof can be found in the Appendix.

Lemma 4.1. *Assume that* N1, N2, N3, N7 *and* N9 *hold. If in addition,* $w_1(\mathbf{x})\|\mathbf{x}\|^3$ *is bounded,* $nh^4 \to 0$, $\lim_{n\to\infty} nh^2/\log^2(1/h) = +\infty$, *we have that*

(a) $n^{\frac{1}{4}}\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_\infty \xrightarrow{p} 0$ *and* $n^{\frac{1}{4}}\|\widetilde{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$.

(b) $\sup_{\tau \in \mathcal{T}} |\widetilde{\eta}(\tau) - \eta_0(\tau) - (nf_T(\tau))^{-1}\sum_{j=1}^{n} K_h(\tau - t_j)G_j(\tau)| = o_p(n^{-1/2})$, *where* $G_j(\tau)$ *is defined above and* $K_h(u) = (1/h)K(u/h)$.

Remark 4.1.

- In Lemma 4.1, the assumption that $w_1(\mathbf{x})\|\mathbf{x}\|^3$ is bounded can be relaxed by requiring that, for some $s > 4$, $Ew_1^s(\mathbf{x})\|\mathbf{x}\|^{3s} < \infty$ and that the kernel K has bounded support, by using similar arguments to those considered in Mack and Silverman (1982).
- It is worth noting that if $\mathbf{A}(\tau)$ is non-singular, thus, the condition $\inf_{\tau \in \mathcal{T}} \lambda_{p+1}(\mathbf{A}(\tau)) > 0$ will be fulfilled if $\mathbf{A}(\tau)$ is a continuous function of τ.
- On the other hand, the continuous differentiability of the kernel K and the implicit function theorem entail that $\widetilde{\eta}(\tau)$ and $\widetilde{\boldsymbol{\beta}}(\tau)$ are continuously differentiable functions of τ. Moreover, using that $S_n^1(\widetilde{\eta}(\tau), \widetilde{\boldsymbol{\beta}}(\tau), \tau) = 0$, we get that

$$\widetilde{\mathbf{A}}_n(\tau)\begin{pmatrix}\widetilde{\eta}'(\tau) \\ \widetilde{\boldsymbol{\beta}}'(\tau)\end{pmatrix} = -\frac{1}{nh_n^2}\sum_{i=1}^{n} K'\left(\frac{\tau - t_i}{h_n}\right)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau)\right)\mathbf{z}_i,$$

where

$$\widetilde{\mathbf{A}}_n(\tau) = -\frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{\tau - t_i}{h_n}\right)\chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau)\right)\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}},$$

and so, the uniform consistency required in N1 to $\widetilde{\eta}'(\tau)$ can be derived through analogous arguments to those considered in Theorem 3.1, if the following requirements hold
(a) K is continuously differentiable with derivative K' bounded and with bounded variation
(b) for any compact sets $\mathcal{K} \in \mathbb{R}^p$ and $\mathcal{K}_1 \in \mathbb{R}$

$$\sup_{\tau \in \mathcal{T}} E\left(\sup_{\boldsymbol{\beta} \in \mathcal{K}, a \in \mathbb{R}} |\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)\|\mathbf{x}\||t = \tau\right) < \infty,$$

$$\sup_{\tau \in \mathcal{T}} E\left(\sup_{\boldsymbol{\beta} \in \mathcal{K}, a \in \mathbb{R}} |\chi_1\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)\|\mathbf{x}\||t = \tau\right) < \infty,$$

$$\inf_{\substack{\boldsymbol{\beta} \in \mathcal{K}, a \in \mathcal{K}_1 \\ \tau \in \mathcal{T}}} E\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)|t = \tau\right) > 0.$$

- Assumptions N2–N4 are standard conditions on the score function, in particular, N3 is a standard requirement in robust regression in order to get root-n estimators of $\boldsymbol{\beta}$. As noted in Boente et al. (2006), for the score functions considered by

Bianco and Yohai (1995), Croux and Haesbroeck (2002) and Cantoni and Ronchetti (2001), N5 is satisfied. This condition is the conditional Fisher-consistency property as stated in the generalized linear regression model by Künsch et al. (1989).

- It is worth noting that N6 can be replaced by $E\left(\sup_{\boldsymbol{\beta}\in\mathcal{K}, a\in\mathbb{R}} |\chi_1\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)| \|\mathbf{x}\|^2 w_2(\mathbf{x})\right) < \infty$.
- Finally, N7–N9 and the convergence requirement in N1 are standard conditions in nonparametric estimation.

Theorem 4.1. *Let us assume that t_i have compact support $\mathcal{T}$ and that N1–N9 hold. Let $\widehat{\boldsymbol{\beta}}$ be a solution of (12) providing a consistent estimator of $\boldsymbol{\beta}$. If $nh^4 \to 0$, and the conclusion of Lemma 4.1 holds, we have that*

$$n^{\frac{1}{2}} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1}\right).$$

4.2. Tests on the regression parameter

In many situations we are interested in finding out the impact of the covariates $\mathbf{x}$ on the response variable y. That is, we need to make inference on the slope parameter $\boldsymbol{\beta}$ or on some of its components. In this section, we focus on the problem of testing, under model (1), the parametric hypothesis $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. It seems natural to test $H_{0\boldsymbol{\beta}}$ through the Wald-type statistic

$$D(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}, H_{0\boldsymbol{\beta}}) = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{17}$$

where $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}$ is an estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0} = \mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1}$. To define estimators of $\boldsymbol{\Sigma}$ and $\mathbf{A}$, let us denote by $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}, \eta)$ and $\widehat{\mathbf{A}}(\boldsymbol{\beta}, \eta)$ the matrices

$$\widehat{\mathbf{A}}(\boldsymbol{\beta}, \eta) = \frac{1}{n} \sum_{i=1}^{n} \chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}, \tag{18}$$

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}, \eta) = \frac{1}{n} \sum_{i=1}^{n} \left\{\Psi^2\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta(t_i)\right) \widehat{\mathbf{u}}_i(\boldsymbol{\beta}, \eta)\widehat{\mathbf{u}}_i(\boldsymbol{\beta}, \eta)^{\mathrm{T}}\right\}, \tag{19}$$

$$\widehat{\mathbf{u}}_i(\boldsymbol{\beta}, \eta) = w_2(\mathbf{x}_i)\mathbf{x}_i + \widehat{D}_i(t_i, \boldsymbol{\beta}, \eta)\widehat{\boldsymbol{\gamma}}(t_i, \boldsymbol{\beta}, \eta)\widehat{f}(t_i),$$

$$\widehat{\boldsymbol{\gamma}}(\tau, \boldsymbol{\beta}, \eta) = \sum_{i=1}^{n} W_i(\tau)\chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\widehat{f}^{-1}(t_i), \tag{20}$$

$$\widehat{\mathbf{A}}(\tau, \boldsymbol{\beta}, \eta) = \sum_{i=1}^{n} W_i(\tau)\chi(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \eta(t_i))w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}, \tag{21}$$

where $\widehat{f}(\tau)$ is a kernel estimate of the density $f_T(\tau)$, $\widehat{D}_i(\tau, \boldsymbol{\beta}, \eta)$ denotes the first element of $\widehat{\mathbf{A}}(\tau, \boldsymbol{\beta}, \eta)^{-1}w_1(\mathbf{x}_i)\mathbf{z}_i$. The following two estimates of $\boldsymbol{\Sigma}$ and $\mathbf{A}$ may be considered, $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ and $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widehat{\eta})$ or $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \widetilde{\eta})$ and $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widetilde{\eta})$, where $\widetilde{\eta}$ and $\widehat{\eta}$ are the estimators defined in Step 1 and Step 3, respectively.

Likelihood ratio-type tests based on the robust quasi-likelihood as described in Boente et al. (2006) are another possibility. The main disadvantage of robust quasi-likelihood tests is that they behave asymptotically as a linear combination of independent chi-square random variables with one degree of freedom and thus, p-values are not easily derived.

In regression, one of the most frequent hypothesis testing problems involves only a subset of the regression parameter. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^{\mathrm{T}}, \boldsymbol{\beta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{x} = (\mathbf{x}_{(1)}^{\mathrm{T}}, \mathbf{x}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\beta}_{(1)} \in \mathbb{R}^q$. In order to test $H_{0\boldsymbol{\beta}_{(1)}} : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}, \boldsymbol{\beta}_{(2)}$ unspecified, one may use the statistic

$$D_1(\widehat{\boldsymbol{\beta}}_{(1)}, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}, H_{0\boldsymbol{\beta}_{(1)}}) = (\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0}), \tag{22}$$

where $\widehat{\boldsymbol{\Sigma}}_{11}$ denotes the $q \times q$ submatrix of $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}} = \widehat{\mathbf{A}}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{A}}^{-1}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$.

In the next theorem, we derive the asymptotic distribution of the Wald statistic (17), under the null hypothesis $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ and under a sequence of contiguous alternatives.

Theorem 4.2. *Let $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}, 1 \le i \le n$ be independent random vectors satisfying (1). Let us assume that t_i have compact support $\mathcal{T}$ and that N1–N9 hold. Let $\widehat{\boldsymbol{\beta}}$ be a solution of (12) providing a consistent estimator of $\boldsymbol{\beta}$. If $nh^4 \to 0$, and the conclusion of Lemma 4.1 holds, we have that*

(i) *under $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $W_n = nD(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}, H_{0\boldsymbol{\beta}}) \xrightarrow{\mathcal{D}} \chi_p^2$*

(ii) *under $H_{1\boldsymbol{\beta}} : \boldsymbol{\beta} \ne \boldsymbol{\beta}_0$, $W_n \xrightarrow{p} \infty$, for any fixed $\boldsymbol{\beta}$*

(iii) *under $H_{1\boldsymbol{\beta}(\mathbf{c})} : \boldsymbol{\beta} = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}$, $W_n \xrightarrow{\mathcal{D}} \chi_p^2(\theta)$, where $\theta = \mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}^{-1}\mathbf{c}$, if, in addition, for any $\mathbf{c} \in \mathbb{R}^p$, $\widetilde{\eta}(\tau)$ is such that*

$$\|\widetilde{\eta} - \eta_0\|_{\infty} \xrightarrow{p} 0 \text{ when model (1) holds with } \boldsymbol{\beta} = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}.$$

From Theorem 4.2, to test $H_{0\boldsymbol{\beta}}$ at a given significance level α, the robust Wald test rejects $H_{0\boldsymbol{\beta}}$ if $W_n > \chi_{p,\alpha}^2$.

A similar result to that given in the previous theorem can be derived for the robust statistic (22) used to test $H_{0\beta_{(1)}}$, i.e., when the null hypothesis involves only a subset of q parameters. In Theorem 4.3, we state the asymptotic distribution of the Wald-type statistic. Its proof is similar to that of Theorem 4.2.

Theorem 4.3. Let $(y_i, \mathbf{x}_i^T, t_i)^T$, $1 \le i \le n$ be independent random vectors satisfying (1). Let us assume that t_i have compact support $\mathcal{T}$ and that N1–N9 hold. Let $\widehat{\boldsymbol{\beta}}$ be a solution of (12) providing a consistent estimator of $\boldsymbol{\beta}$. If $nh^4 \to 0$, and the conclusion of Lemma 4.1 holds, we have that N1–N6 hold, we have that

(i) under $H_{0\beta_{(1)}} : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\mathcal{W}_{1,n} = nD_1(\widehat{\boldsymbol{\beta}}_{(1)}, \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}, H_{0\beta_{(1)}}) \xrightarrow{\mathcal{D}} \chi_q^2$

(ii) under $H_{1\beta_{(1)}} : \boldsymbol{\beta}_{(1)} \neq \boldsymbol{\beta}_{(1),0}$, $\mathcal{W}_{1,n} \xrightarrow{p} \infty$

(iii) under $H_{1\beta_{(1)}(\mathbf{c}_{(1)})} : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0} + \mathbf{c}_{(1)}n^{-1/2}$, $\mathcal{W}_{1,n} \xrightarrow{\mathcal{D}} \chi_q^2(\theta_1)$, where $\theta_1 = \mathbf{c}_{(1)}^T \boldsymbol{\Sigma}_{\beta_0,11}^{-1} \mathbf{c}_{(1)}$, if, in addition, for any $\mathbf{c}_{(1)} \in \mathbb{R}^q$, $\widetilde{\eta}(\tau)$ is such that $\|\widetilde{\eta} - \eta_0\|_\infty \xrightarrow{p} 0$ when model (1) holds with $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^T, \boldsymbol{\beta}_{(2)}^T)^T$ where $\boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0} + \mathbf{c}_{(1)}n^{-1/2}$.

5. Empirical influence function

Robust procedures seek for estimates less sensitive to outliers than the classical ones. To measure robustness with respect to single outliers, the empirical influence function has shown to be useful, see Tukey (1977), since it reflects the behavior of the estimator when we change one element of the sample by a new observation that does not follow the original model. Statistical diagnostics and graphical displays for detecting outliers can be built based on empirical influence functions. Mallows (1974) considered an influence function for small samples related to the influence function defined by Hampel (1974) (see Hampel et al., 1986) computed at the sample empirical distribution. In parametric models this topic is widely developed, however, less attention has been given in the nonparametric literature. A smoothed functional approach to nonparametric kernel estimators was introduced by Aït Sahalia (1995) and used by Tamine (2002) to define a smoothed influence function in nonparametric regression assuming that the smoothing parameter is fixed. On the other hand, Manchester (1996) introduced a graphical method to display sensitivity of a scatter plot smoother.

To measure the influence of outlying observations on the proposed estimators, we will follow an approach similar to that given by Manchester (1996). However, instead of considering the finite-sample version of the influence function introduced by Tukey (1977), we will give an approach related to the *empirical influence function* defined by Mallows (1974), which is the influence function of the functional under study computed for the empirical distribution. Given a data set $\{(y_i, \mathbf{x}_i, t_i)\}_{1 \le i \le n}$, let $\widehat{\boldsymbol{\beta}}$ be the regression parameter estimator based on this data set. Denote P_n the empirical measure that gives weight $1/n$ to each sample point, thus, $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}(P_n)$. On the other hand, let $P_{n,0}$ the empirical measure that gives mass $(1 - \varepsilon)/n$ to each $(y_i, \mathbf{x}_i, t_i)$, $1 \le i \le n$, and ε to the observation $(y_0, \mathbf{x}_0, t_0)$. Denote $\widehat{\boldsymbol{\beta}}_{0,\varepsilon}$ the regression parameter estimator for this new sample. We can thus define the empirical influence function (EIF) of $\widehat{\boldsymbol{\beta}}$ at $(y_0, \mathbf{x}_0, t_0)$ as

$$\text{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) = \lim_{\varepsilon \to 0} \frac{\widehat{\boldsymbol{\beta}}_{0,\varepsilon} - \widehat{\boldsymbol{\beta}}}{\varepsilon}.$$

Assume that $\chi(y, u) = \partial \Psi(y, u)/\partial u$ exists, in the Appendix it is shown that,

$$\text{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) = -\left\{ \frac{1}{n} \sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T \right\}^{-1} \left\{ \Psi\left(y_0, \mathbf{x}_0^T\widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_0)\right) w_2(\mathbf{x}_0)\mathbf{x}_0 \right.$$

$$\left. + \frac{1}{n} \sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\text{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(t_i) \right\}, \tag{23}$$

where $\text{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(\tau) = \left(\partial \widetilde{\eta}_{0,\varepsilon}(\tau)/\partial \varepsilon\right)|_{\varepsilon=0}$ and $\widetilde{\eta}_{0,\varepsilon}(\tau)$ is the regression function estimator obtained in Step 1 with the new sample. Moreover, $\text{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(\tau)$ is the first element of the vector

$$v_0(\tau) = -\widehat{\mathbf{A}}_0(\tau)^{-1}K\left(\frac{\tau - t_0}{h}\right) \Psi\left(y_0, \mathbf{x}_0^T\widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau)\right) w_1(\mathbf{x}_0)\mathbf{z}_0, \tag{24}$$

with

$$\widehat{\mathbf{A}}_0(\tau) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{\tau - t_i}{h}\right) \chi\left(y_i, \mathbf{x}_i^T\widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau)\right) w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^T. \tag{25}$$

A similar expression for $\text{EIF}(\widehat{\eta}; (y_0, \mathbf{x}_0, t_0))(\tau)$ can be obtained by replacing in (24) and (25), $\widetilde{\boldsymbol{\beta}}(\tau)$ and $\widetilde{\eta}$ by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}(\tau)$, respectively.

To study the behavior of the estimators, we have considered a logistic and a Gamma model. For the logistic model, we generate a sample (y_i, x_i, t_i), $1 \le i \le n$, of size $n = 200$ where the response variable y_i is such that $y_i|(x_i, t_i) \sim Bi(1, p(x_i, t_i))$ where $\log(p(x, t)/(1 - p(x, t))) = x/2 + t - 0.5 + \sin(4\pi t)$, i.e., $\beta_0 = 0.5$, $\eta_0(t) = (t - 1/2) + \sin(4\pi t)$. The covariates
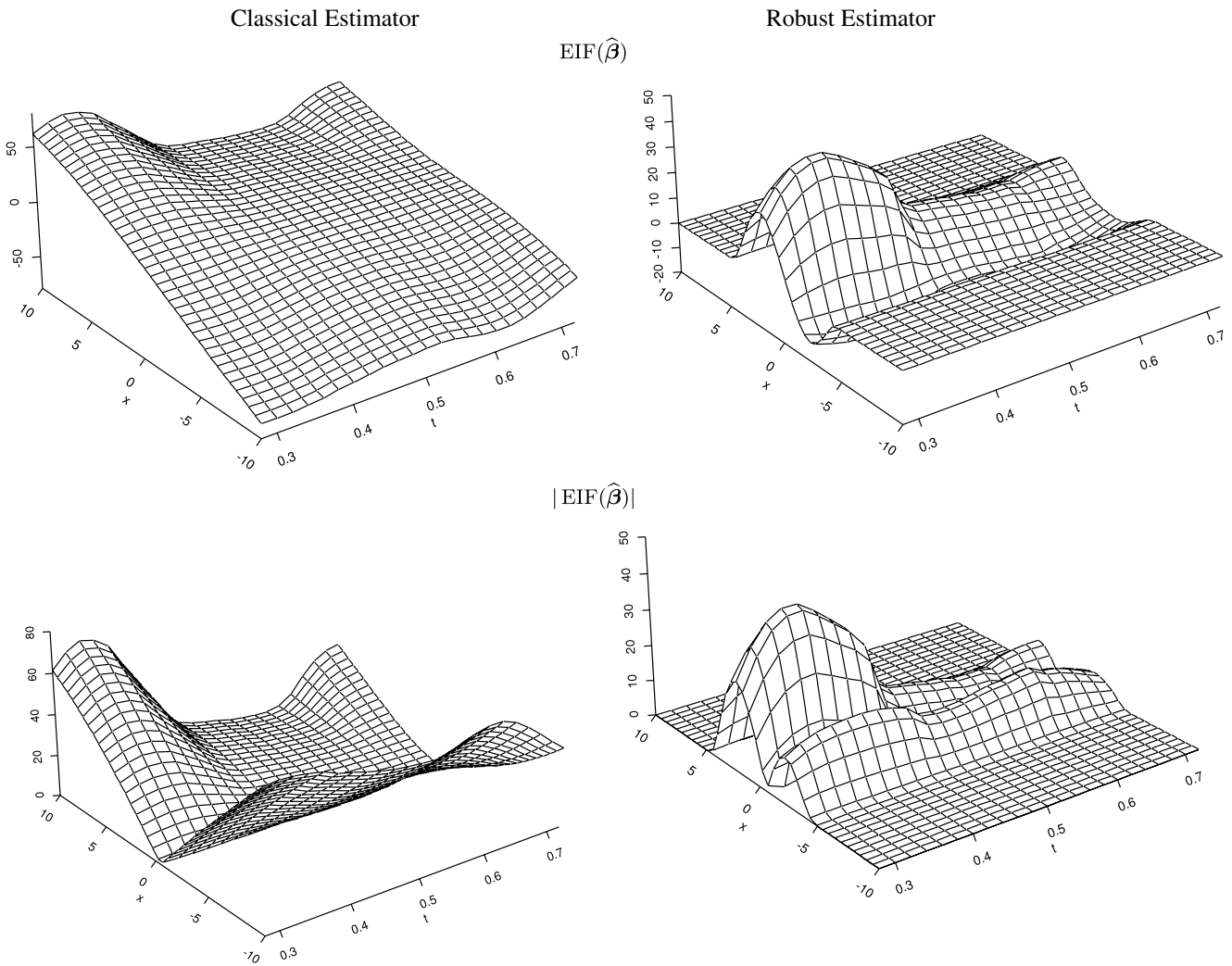
Classical Estimator   Robust Estimator

$\mathrm{EIF}(\widehat{\boldsymbol{\beta}})$

$|\mathrm{EIF}(\widehat{\boldsymbol{\beta}})|$

Fig. 1. Empirical influence function of $\widehat{\boldsymbol{\beta}}$ and its absolute value, when $y_0 = 1$.

are such that $(x_i, t_i) \sim N((0, 1/2), \Sigma)$, $1 \le i \le n$, with $\Sigma = \begin{pmatrix} 1 & 1/(6\sqrt{3}) \\ 1/(6\sqrt{3}) & 1/36 \end{pmatrix}$ and the variable t was truncated so that $t \in [1/4, 3/4]$. We select a grid of values (y_0, x_0, t_0) defined by $y_0 = 0$ or $y_0 = 1$, (t_0, x_0) taking values on a equidistant grid on each axis of size 23×40 on $[0.28, 0.72] \times [-10, 10]$. Thus, for $y_0 = 1$ and $y_0 = 0$, we have a grid of 920 points (x_0, t_0) and for each of them we have computed the empirical influence function, $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}, (y_0, \mathbf{x}_0, t_0))$, given by (23). The influence function was computed for both the classical and robust estimators. The classical procedure corresponds to select the quasi-likelihood in Steps 1 to 3, i.e., with $\Psi(y, a) = -(y - H(a))H'(a)/V(H(a))$ and $w_1 = w_2 \equiv 1$. Under the logistic model, the robust proposal was computed by bounding the deviance as defined in (13) using the score function ϕ proposed by Croux and Haesbroeck (2002) with tuning constant $c = 0.5$. Besides, the weight functions w_1 and w_2 were chosen as Tukey's biweight function with tuning constant $c = 4.685$

$$w_1(x_i) = w_2(x_i) = \begin{cases} \left(1 - \left[\dfrac{x_i - M_n}{4.685}\right]^2\right)^2 & |x_i - M_n| \le 4.685 \\ 0 & |x_i - M_n| \ge 4.685, \end{cases}$$

with M_n the median of x_i since we have considered $x_i \in \mathbb{R}$. The kernel was the Epanechnikov kernel, $K(t) = (3/4)\left(1 - t^2\right) I_{[-1, 1]}(t)$ and the bandwidth was $h = 0.1$. The bandwidth choice was based in the fact that this bandwidth was selected by Boente et al. (2006) in their simulation study. Moreover, in Section 6, we report the results when the estimators are computed using different bandwidths. The best performance was obtained for $h = 0.1$, when no outliers are present, both for the classical and robust estimators of $\boldsymbol{\beta}$.

Figs. 1 and 2 give the plots for $y_0 = 1$. The corresponding ones for $y_0 = 0$ are quite similar and can be found in Boente and Rodriguez (2008). Note that for the classical estimators under a logistic model $\Psi(y, a) = H(a) - y$ and so $\chi(y, a) = H(a)(1 - H(a))$ which implies that EIF is unbounded for large values of the covariates $\mathbf{x}$ since we are assuming that t has compact support. In fact, the plots given show that the absolute value of the influence function of the classical

Classical Estimator   Robust Estimator

$\mathrm{EIF}(\widehat{\eta})$

$|\mathrm{EIF}(\widehat{\eta})|$
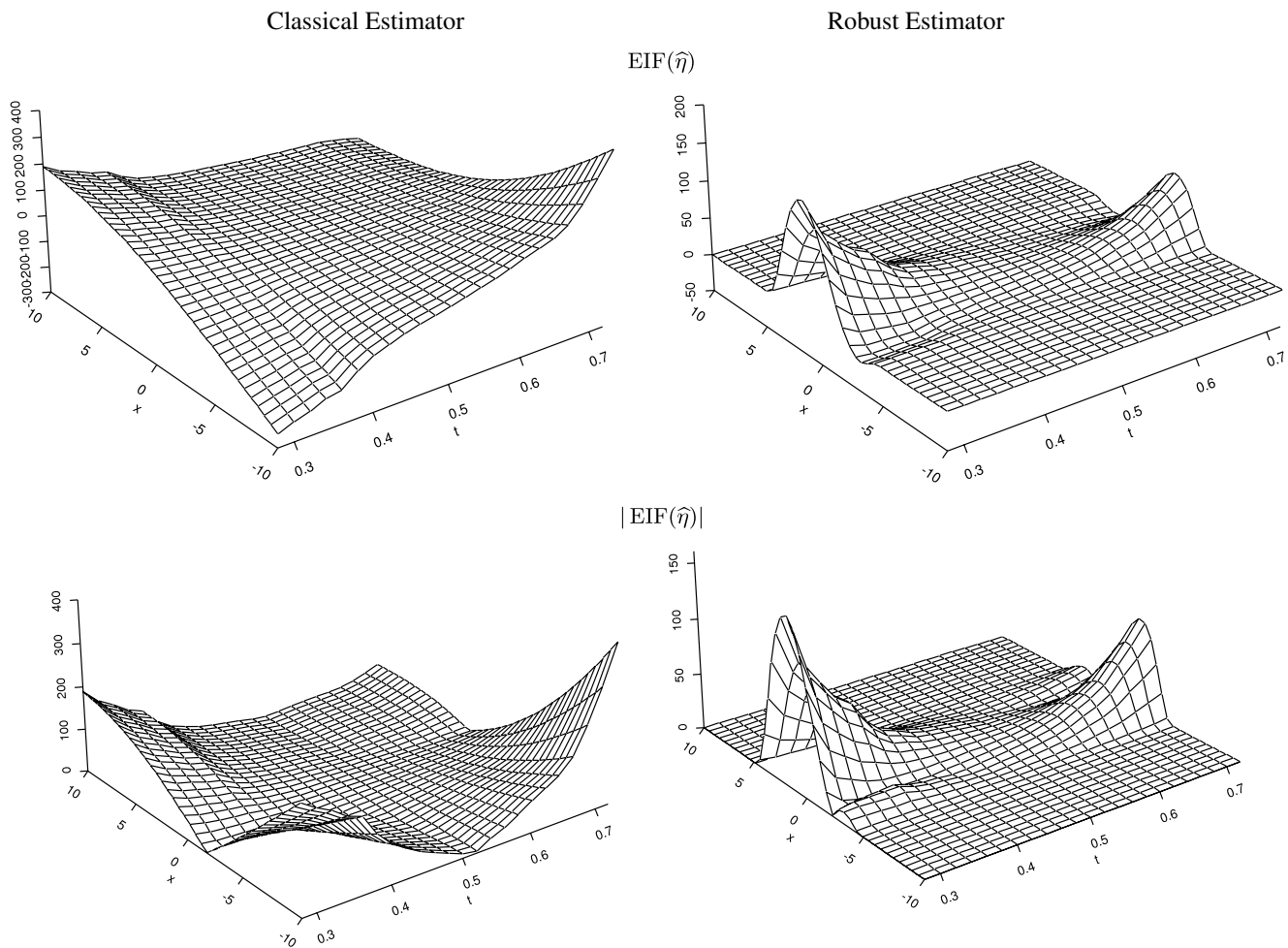
Fig. 2. Empirical influence function of $\widehat{\eta}$ and its absolute value, when $y_0 = 1$.

estimators increases as **x** increase. Negative values with large absolute value are extremely influential for the estimation of β when $y_0 = 1$. On the other hand, the robust estimator has a bounded empirical influence function. For both models, the minimum value for the EIF is attained for values of x close to -10 when $y_0 = 1$. This influence is negative, i.e., it produces a negative bias in the estimators. On the other hand, points in a neighborhood of $t = 0.3$ and $x = 10$, when $y_0 = 1$, correspond to the maximum of the empirical influence function, this maximum being also the maximum of the absolute value of |EIF|. Therefore, for small contaminations the more dangerous contaminations correspond to values close to $x = 10$ and $t = 0.3$ when $y_0 = 1$ producing the larger bias in the estimation, this bias being positive. Similar conclusions can be obtained for the estimation of η.

As mentioned above, we have also studied the behavior of the estimators under a Gamma model. For $1 \le i \le n$, we generated covariates (x_i, t_i) independent of each other such that $x_i \sim N(0, 1)$, $t_i \sim \mathcal{U}(0, 1)$. The response variable was generated as $y_i|(x_i, t_i) \sim \Gamma(3, \lambda_i)$, where $E(y_i|(x_i, t_i)) = 3/\lambda_i = \exp\{\beta_0 x_i + \eta_0(t_i)\}$, with $\beta_0 = 2$, $\eta_0(t) = \sin(2\pi t)$, i.e., $H(a) = \exp(a)$. The sample size equals $n = 100$ and the empirical influence function EIF, given by (23) was computed over a grid of values (y_0, x_0, t_0) defined by $t_0 \in \{0.2, 0.4, 0.6, 0.8\}$ and (y_0, x_0) taking values on a equidistant grid on each axis of size 40×40 on $[0.2, 10] \times [-10, 10]$. As for the logistic case, we have considered the Epanechnikov kernel in the smoothing procedure and the bandwidth equals $h = 0.15$. The robust estimators corresponds to those controlling large values of the deviance, $d(y, a) = y/H(a) - \ln(y/H(a)) - 1$, and were computed using Tukey's biweight score function with tuning constant $c = 2$. The weight functions w_1 and w_2 used to control high leverage points were taken as in the previous example. We only show the obtained surfaces for $t_0 = 6$, for other values of t_0 see Boente and Rodriguez (2008). For the selected Gamma model, the classical estimators considered are not based on the quasi-likelihood but on the deviance, i.e., they correspond to the choice $\phi(t) = t$ in (13) and $w_1 = w_2 \equiv 1$. Thus, $\Psi(y, a) = 1 - y/H(a)$, $\chi(y, a) = y/H(a)$ and so, the empirical influence function will be unbounded for each fixed t_0, when $y_0 \to \infty$ for fixed x_0 and when $x_0 \to -\infty$ for fixed y_0. Note that since $\Psi(y_0, x_0\beta + \upsilon)x_0 = (1 - y_0 \exp(-x_0\beta - \upsilon))x_0$, the empirical influence function of the classical estimators will still be unbounded when $x_0 \to +\infty$, but at a smaller rate than when $x_0 \to -\infty$. Figs. 3 and 4 show that, for large negative values of x and large values of y, the absolute values of the empirical influence function of the classical estimators takes very large values. The worst effect being observed for high leverage points. On the other hand, the robust procedure lead to more stable estimators.
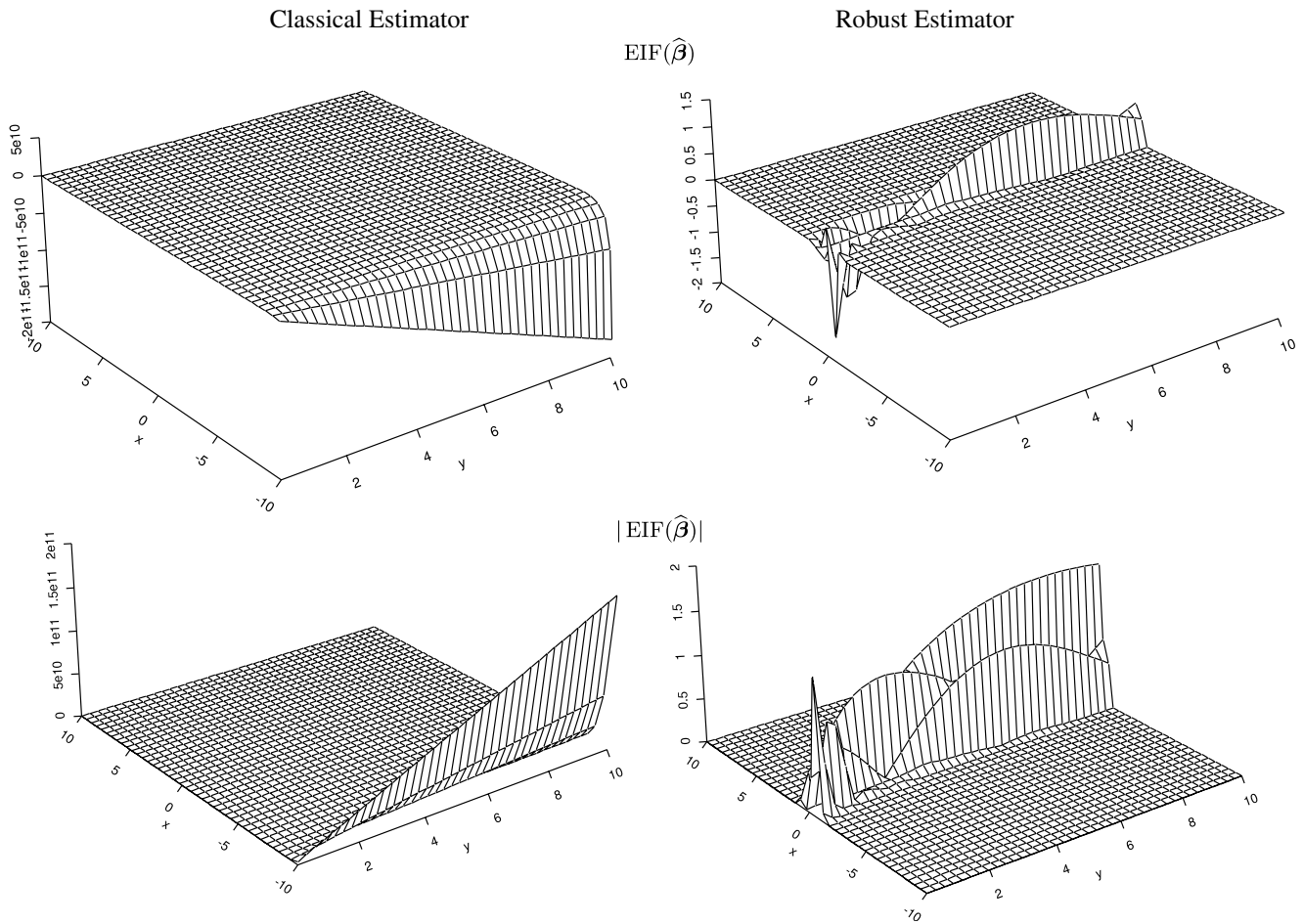
Classical Estimator   Robust Estimator

$\text{EIF}(\widehat{\boldsymbol{\beta}})$

$|\text{EIF}(\widehat{\boldsymbol{\beta}})|$

Fig. 3. Empirical influence function of $\widehat{\boldsymbol{\beta}}$ and its absolute value, under a Gamma model when $t = 0.6$.

In this situation, the classical estimator of the regression function η can be extremely influenced by an anomalous observation. Note that the maximum value of the |EIF| is attained for large negative values of the covariates x ($x_0 = -10$) combined with large values of the responses ($y_0 = 10$). This corresponds also to the maximum of the empirical influence function, showing that in these regions the estimators will have a huge positive bias. The bias both for the classical regression estimators $\widehat{\boldsymbol{\beta}}$ or the classical regression function estimators $\widehat{\eta}$ is of order 10^{10} and so it is 10^8 times larger than in the logistic setting. The huge effect of outliers can be explained by the fact that under a Gamma model, the response variables can attain large values. As mentioned in Boente et al. (2006), for unbounded response variables y, bounded score functions allow us to deal with large residuals, while, for models with a bounded response, such as the logistic one, the score functions introduced in the robust procedure protects against outliers with large Pearson residuals that for binary responses y correspond only when to contaminated points in which the variances are close to 0. This fact explains the large influence observed under a Gamma model for the regression function. To give an example of this behavior we have computed the values of the empirical influence function at large values of **x** or y, for the classical estimators, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (100, 2, 0.6)) = 4.149$, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (1000, 2, 0.6)) = 61.926$, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (2, 100, 0.6)) = -133.562$ and $\text{EIF}(\widehat{\boldsymbol{\beta}}; (2, 1000, 0.6)) = -1339.309$ while for the robust procedure, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (100, 2, 0.6)) = 2.974$, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (1000, 2, 0.6)) = 0$, $\text{EIF}(\widehat{\boldsymbol{\beta}}; (2, 100, 0.6)) = 2.974$ and $\text{EIF}(\widehat{\boldsymbol{\beta}}; (2, 1000, 0.6)) = 0$. The effect of outliers in the covariates **x** seems to be larger when estimating the function η since in this situation for the classical estimators, $\text{EIF}(\widehat{\eta}; (100, 2, 0.6)) = 2.485$, $\text{EIF}(\widehat{\eta}; (1000, 2, 0.6)) = 45.273$, $\text{EIF}(\widehat{\eta}; (2, 100, 0.6)) = 286.125$ and $\text{EIF}(\widehat{\eta}; (2, 1000, 0.6)) = 2934.640$ while for the robust procedure, $\text{EIF}(\widehat{\eta}; (100, 2, 0.6)) = 5.967$, $\text{EIF}(\widehat{\eta}; (1000, 2, 0.6)) = 0$, $\text{EIF}(\widehat{\eta}; (2, 100, 0.6)) = 5.967$ and $\text{EIF}(\widehat{\eta}; (2, 1000, 0.6)) = 0$. This behavior can be related to the fact that locally the regression function η acts like an intercept when estimating it and outliers affect considerably the intercept in generalized linear models.

It is worth noting that, due to the scale of Figs. 3 and 4, the |EIF| of the classical estimators seems to be equal to 0 for values of x larger than -9. To avoid this masking effect, Fig. 5 give the plots of the EIF for the classical and robust procedures for a reduced range of values of x and t, to compare the behavior of both methods. These plots show that the shape of both surfaces is similar in the central part, i.e., for values of x in the range of [1.5, 3] and [1, 3] for $\text{EIF}(\widehat{\boldsymbol{\beta}})$ and $\text{EIF}(\widehat{\eta})$, respectively. Beyond that range the $\text{EIF}(\widehat{\boldsymbol{\beta}})$ of the classical estimators decrease while that of the robust procedure remains bounded. Similar conclusions hold for $\text{EIF}(\widehat{\eta})$. In fact, when $t_0 = 0.6$, the 25% quantile of the computed values of $|\text{EIF}(\widehat{\boldsymbol{\beta}})|$ and $|\text{EIF}(\widehat{\eta})|$

Classical Estimator   Robust Estimator

$\mathrm{EIF}(\widehat{\eta})$
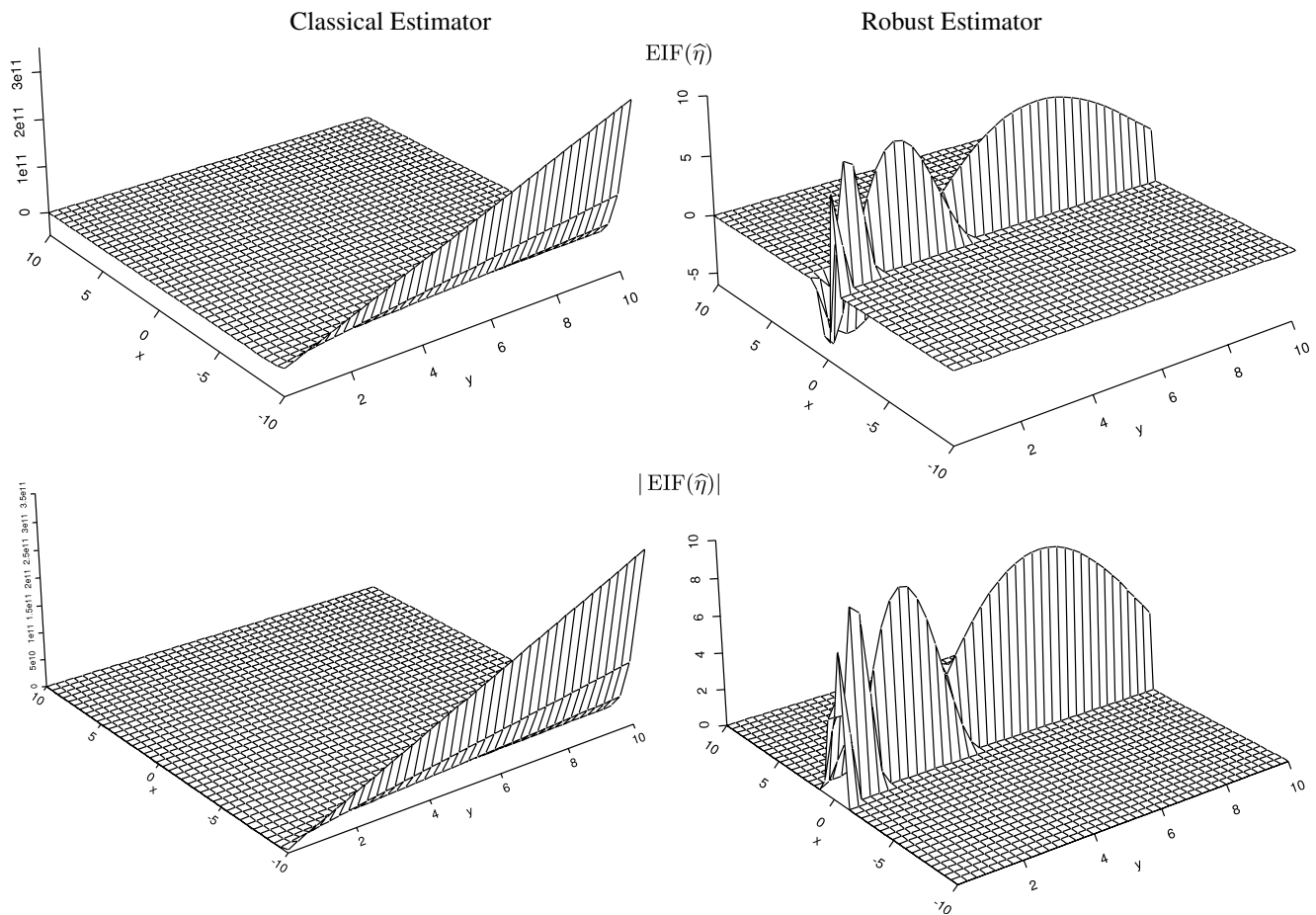
$|\mathrm{EIF}(\widehat{\eta})|$

Fig. 4. Empirical influence function of $\widehat{\eta}$ and its absolute value, under a Gamma model when $t = 0.6$.

are equal to 6.115 and 6.220, respectively for the classical estimators. On the other hand, for the robust estimators, large absolute values of the covariates x combined with any value of the response will lead to a null value of the EIF since leverage points are penalized with a null weight when computing the robust estimators. Moreover, when $t_0 = 0.6$, the maximum values of |EIF| for the robust estimators of β and η equal 1.640 and 9.791, respectively. Fig. 6 plots the values of the empirical influence function of $\widehat{\beta}$ and $\widehat{\eta}$, when $y = 3$ and t_0 take values on a grid of equidistant points in [0, 1] of length 40. The same conclusions described above remain valid in this situation.

6. Monte Carlo study

6.1. Behavior of the estimators

6.1.1. Logistic model

This section contains the results of a simulation study conducted with the aim of comparing the performance of the proposed estimators with the classical one and with those defined in Boente et al. (2006) under a logistic partially linear model. The estimators considered are those defined in Section 2, denoted MOD in Tables and Figures, and the classical estimators, denoted by QAL, as defined in Carroll et al. (1997) and described in Sections 2 and 5, which are an alternative to those, based on profile likelihood, considered in Severini and Staniswalis (1994).

The estimators MOD corresponds to those controlling large values of the deviance and they were computed using the score function defined in Croux and Haesbroeck (2002) with tuning constant $c = 0.5$. The weight functions w_1 and w_2 used to control high leverage points were taken as Tukey's biweight function with tuning constant $c = 4.685$, as in Section 5. The central model denoted C_0 in Tables and Figures corresponds to (y_i, x_i, t_i) such that $(x_i, t_i) \sim N((0, 1/2), \Sigma)$ and $y|(x, t) \sim Bi(1, p(x, t))$ as in Section 5, i.e., $\beta_0 = 0.5$, $\eta_0(t) = (t - 0.5) + \sin(4\pi t)$. We have considered 1000 replications of samples of size $n = 200$ and, as in Section 5, the Epanechnikov kernel $K(t) = (3/4)(1 - t^2)I_{[-1,1]}(t)$ was selected in the smoothing procedure. Several bandwidths were chosen under C_0 to compute the new robust estimators. Fig. 7 gives the boxplots of the QAL and MOD estimators of β for different smoothing parameters. The selection $h = 0.1$ made by Boente et al. (2006) gives also the best results in this case. Thus, we only report the results corresponding to $h = 0.1$.

For each sample generated, we have considered the following contamination labeled C_1 in Tables and Figures. We have first generated a sample $u_i \sim \mathcal{U}(0, 1)$, $1 \le i \le n$, and then, the contaminated sample, denoted $(y_{i,c}, x_{i,c}, t_i)$, is defined as

Classical Estimator   Robust Estimator

$\text{EIF}(\widehat{\boldsymbol{\beta}})$

$\text{EIF}(\widehat{\eta})$
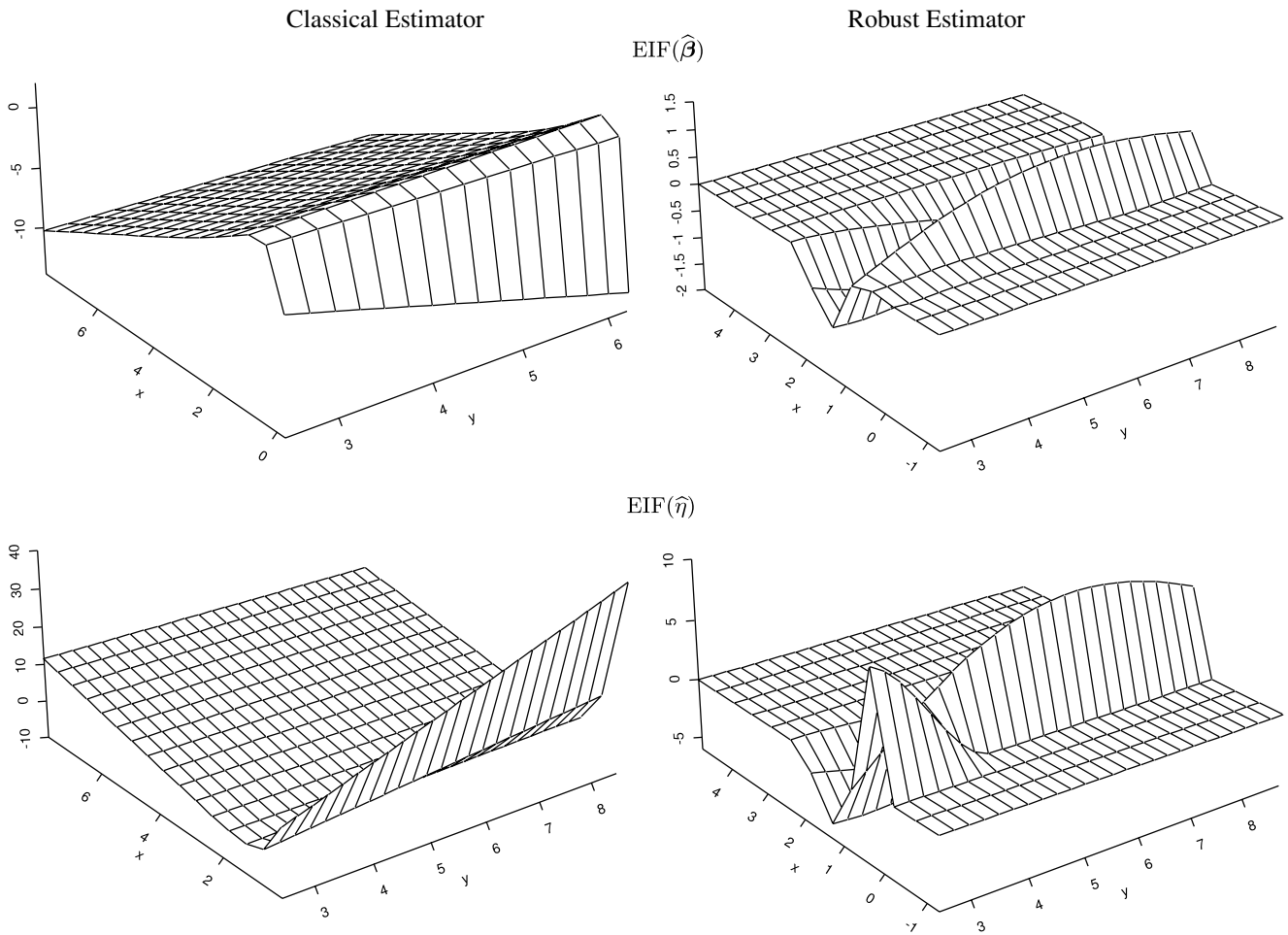
Fig. 5. Empirical influence function of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}$, under a Gamma model when $t = 0.6$.

Table 1
Summary results for the estimators of β_0 and η_0 under a logistic model.

	Estimator	Bias ($\hat{\beta}$)	SD ($\hat{\beta}$)	MSE ($\hat{\beta}$)	MSE ($\widetilde{\eta}$)	MSE ($\widehat{\eta}$)
C_0	QAL $_{\text{PRO}}$	0.0171	0.2175	0.0475		0.1164
	MOD $_{\text{BHZ}}$	0.0176	0.2222	0.0497		0.1197
	QAL	0.0144	0.2113	0.0426	0.1657	0.1158
	MOD	0.0217	0.2058	0.0428	0.2069	0.1190
C_1	QAL $_{\text{PRO}}$	−0.7194	0.0658	0.5219		0.2341
	MOD $_{\text{BHZ}}$	0.0212	0.2354	0.0559		0.1803
	QAL	−0.7198	0.0648	0.5222	0.2211	0.2277
	MOD	0.0260	0.2185	0.0484	0.2069	0.1992

follows $(y_{i,c}, x_{i,c}) = (y_i, x_i)$ if $u_i \leq 0.90$ and $(y_{i,c}, x_{i,c}) = (y_{i,\text{new}}, x_{i,\text{new}})$ if $u_i > 0.90$, where $x_{i,\text{new}}$ is a new observation from a $N(10, 1)$ and $y_{i,\text{new}}$ is a new observation from a $Bi(1, 0.05)$. Table 1 summarizes the results obtained. The estimators defined in Boente et al. (2006) are indicated as MOD_{BHZ} and those introduced in this paper MOD. Besides, the classical counterpart of the profile estimators considered in Boente et al. (2006) are indicated as QAL_{PRO}, since they correspond to the choice $\Psi(y, a) = -(y - H(a))H'(a)/V(H(a))$ and $w_1 = w_2 \equiv 1$. For the estimators of β_0, we have considered the following summary measures: bias, standard deviation (SD) and mean square errors (MSE) computed over replications. To study the performance of the estimators of the regression function η_0, denoted by $\widetilde{\eta}$ and $\widehat{\eta}$, we have considered the mean square error (MSE), i.e.,

$$\text{MSE}(\widetilde{\eta}) = \frac{1}{n} \sum_{i=1}^{n} [\widetilde{\eta}(t_i) - \eta(t_i)]^2.$$

The estimator defined of β_0 in Section 2 shows a larger bias than that considered in Boente et al. (2006) and also than the classical one. This fact may be explained by the fact that the design for the variables t is not equispaced. Thus, a bandwidth depending on the design may show its advantage under this model. However, the mean square error of the robust estimators

Classical Estimator　　　　　　　　　　　Robust Estimator

$\mathrm{EIF}(\widehat{\boldsymbol{\beta}})$

$\mathrm{EIF}(\widehat{\eta})$

Fig. 6. Empirical influence function of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\eta}$, under a Gamma model when $y = 3$.
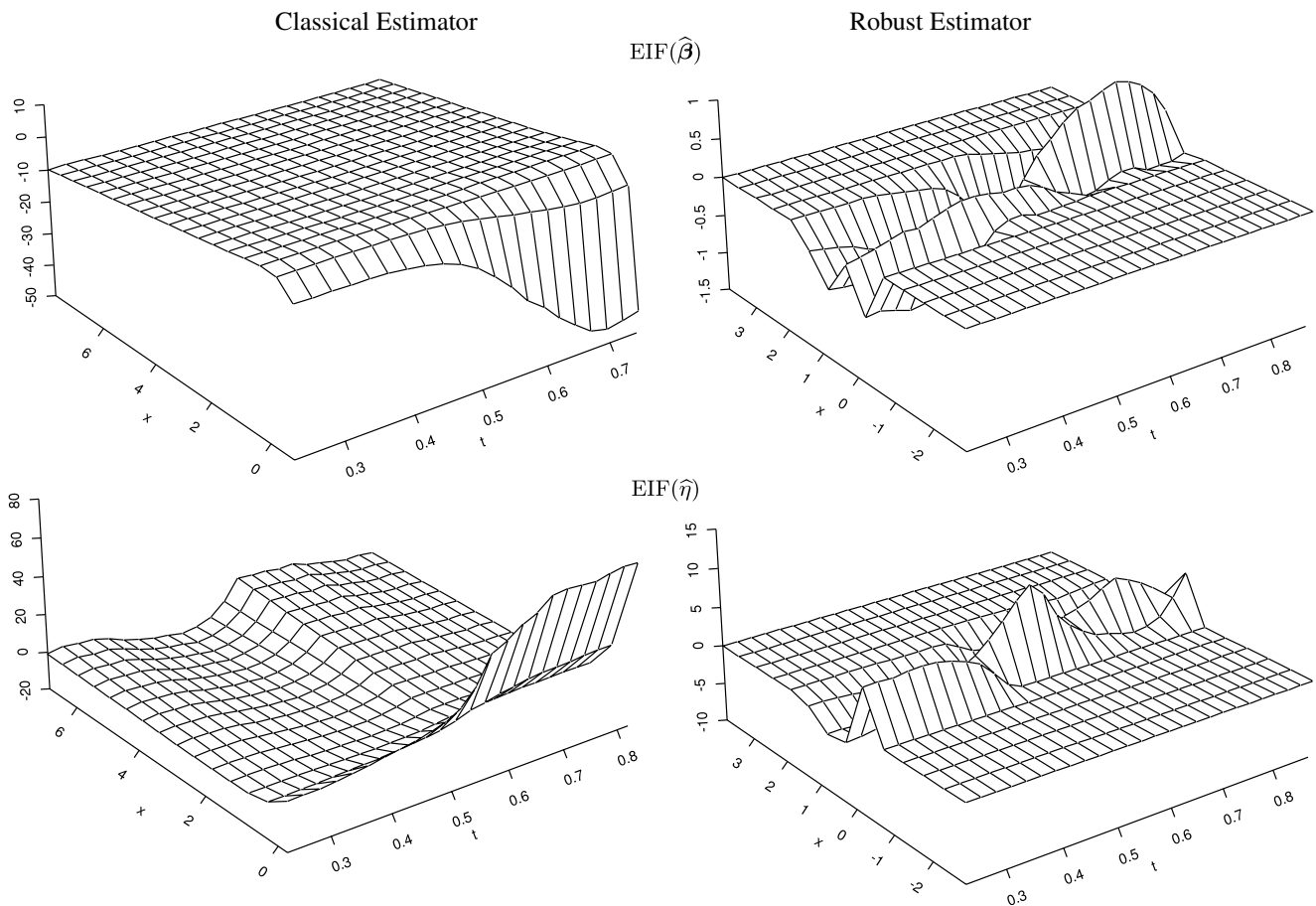
Classical Estimator　　　　　　　　　　　Robust Estimator
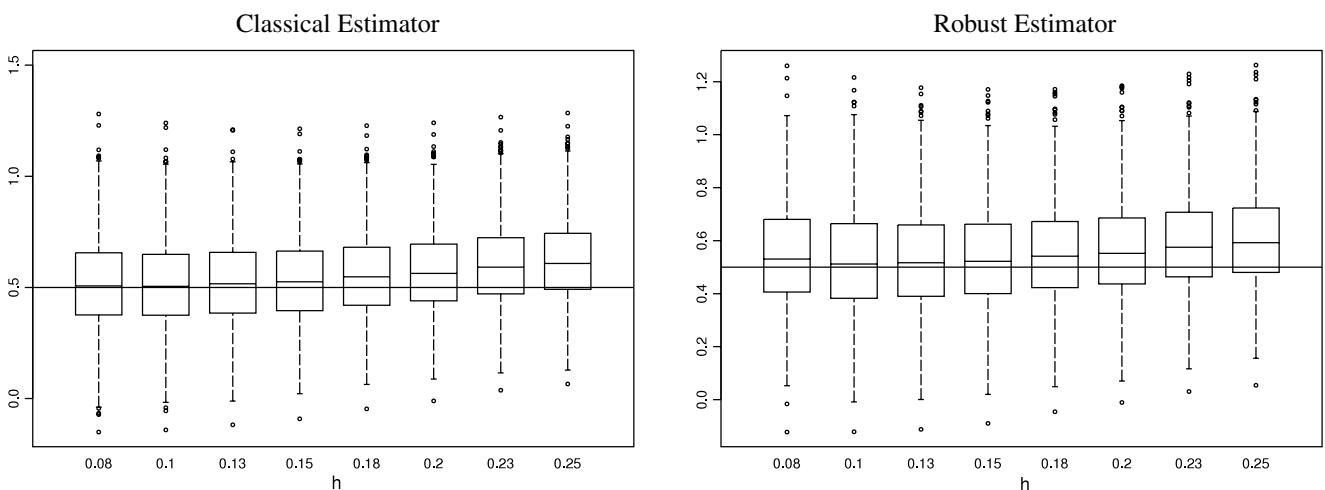
Fig. 7. Boxplots of $\widehat{\boldsymbol{\beta}}$ for different bandwidths h, for the logistic model.

is only slightly larger than that of the classical estimator under C_0, while under C_1, the situation is reversed being the mean square error of the classical procedure more than ten times larger than that of the robust ones. It is worth noting that the standard deviation of the quasi-likelihood estimator is reduced considerably under C_1, so that a test for the regression parameter will reject the null hypothesis $\boldsymbol{\beta}_0 = 0.5$, as we will see later. Finally, with respect to the estimators defined in Boente et al. (2006), the regression estimators introduced in Section 2 show a slight improvement both in variance and in mean square error, both under C_0 and C_1, while under C_0, they show a larger bias.

With respect to the estimation of η, the first step estimators show a poor behavior compared to those considered in Boente et al. (2006) since their mean square error is much larger (more than the double for the robust ones). It is worth noting that even if the target in this paper is the estimation of $\boldsymbol{\beta}$ and η can be considered as a nuisance parameter, we have

Table 2
Summary results for the estimators of β_0 and η_0 under a Gamma model with $c = 2$.

	Estimator	Bias ($\hat{\beta}$)	SD ($\hat{\beta}$)	MSE ($\hat{\beta}$)	MSE ($\tilde{\eta}$)	MSE ($\hat{\eta}$)
C_0	DEV	−0.0017	0.0618	0.0038	0.0288	0.0280
	MOD	−0.0015	0.0625	0.0039	0.0304	0.0287
C_1	DEV	−0.5600	0.2103	0.3578	0.0961	0.0638
	MOD	−0.0004	0.0741	0.0055	0.0659	0.0307
C_2	DEV	−1.4127	1.1147	3.2381	45.8904	54.1543
	MOD	0.0000	0.0675	0.0046	0.0325	0.0307
C_3	DEV	−1.8129	0.3661	3.4206	14.6368	20.5997
	MOD	0.0023	0.0745	0.0056	0.0333	0.0312

introduced Step 3 to improve the performance of the first step estimator. Effectively, $\hat{\eta}$ improves the mean square error of $\tilde{\eta}$ and is less expensive computationally than the procedure introduced in Boente et al. (2006). Besides, it is worth noting that, all procedures are quite stable to estimate the nonparametric component since the size of the outlying observation is bounded in this case.

6.1.2. Gamma model

In this section we summarize the results of a simulation study designed to compare the performance of the proposed estimators with the classical ones under a model with unbounded responses such as the Gamma model. In Tables and Figures, the estimators in this paper are indicated as MOD while their classical counterparts are indicated as DEV, since they correspond to the estimators based on the deviance. To be more precise, the robust estimators correspond to those controlling large values of the deviance as described in Bianco et al. (2005) and they were computed using Tukey's biweight score function with two tuning constants $c = 1.5$ and $c = 2$. The weight functions w_1 and w_2 used to control high leverage points were taken as in the previous section. On the other hand, the classical estimators correspond to the choice $\phi(t) = t$ in (13) and $w_1 = w_2 \equiv 1$. We have performed $NR = 1000$ replications with samples of size $n = 100$ and we have used as bandwidth $h = 0.15$. Other bandwidth values were tested and they give quite similar results.

The central model denoted C_0 in Tables and Figures corresponds to select (x_i, t_i) independent of each other such that $x_i \sim N(0, 1)$, $t_i \sim \mathcal{U}(0, 1)$. The response variable was generated as $y_i|(x_i, t_i) \sim \Gamma(3, \lambda_i)$, where $E(y_i|(x_i, t_i)) = 3/\lambda_i = \exp\{\beta_0 x_i + \eta_0(t_i)\}$, with $\beta_0 = 2$, $\eta_0(t) = \sin(2\pi t)$. As in Section 5, we have considered the Epanechnikov kernel in the smoothing procedure with $h = 0.15$.

For each sample generated we have considered three contaminations labeled C_1, C_2 and C_3 in Tables and Figures that lead to contaminated samples $(y_{i,c}, x_{i,c}, t_i)$. We have first generated a sample $u_i \sim \mathcal{U}(0, 1)$ for $1 \leq i \leq n$ and then, we have considered the following contamination scheme

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

$$x_{i,c} = \begin{cases} x_i & \text{if } u_i \leq 0.90 \\ \text{a new observation } x_i^\star \text{ from a N}\left(5, \dfrac{1}{16}\right) & \text{if } u_i > 0.90. \end{cases}$$

- C_2 introduces outlying observations in the responses generated according to the model but with an incorrect carrier x.

$$y_{i,c} = \begin{cases} y_i & \text{if } u_i \leq 0.90 \\ \text{a new observation } y_i^\star \text{ such that } y_i^\star \sim \Gamma(3, \lambda_i^\star) & \text{if } u_i > 0.90 \end{cases}$$

where $3/\lambda_i^\star = \exp\{\beta_0 x_i^\star + \eta_0(t_i)\}$ and $x_i^\star$ is a new observation from a N $\left(5, \frac{1}{16}\right)$. Note that the carriers are not contaminated in this situation, i.e., $x_{i,c} = x_i$.

- C_3 corresponds to increasing the variance of the carriers x and also to introduce large values on the responses

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

and

$$y_{i,c} = \begin{cases} y_i & \text{if } u_i \leq 0.90 \\ \text{a new observation from a } \Gamma(3, 3/1000) & \text{if } u_i > 0.90. \end{cases}$$

Table 2 summarize the results obtained when $c = 2$. We have considered the same summary measures as in the logistic case. As expected the robust estimators computed with $c = 2$ are more efficient and they lead to similar results under both contaminations than those computed with $c = 1.5$. Therefore, they should be preferred. The behavior of the estimators when $c = 1.5$ can be found in Boente and Rodriguez (2008).

The classical estimator shows its sensitivity under all contaminations, the effect being worst in this case on the estimation of the regression function η when contaminating the responses as in C_2 or C_3. For these two contamination the mean square errors of the classical estimators of η are more than one thousand or five hundred times, respectively, those obtained by

Table 3
Observed frequencies of rejection under the null hypothesis under C_0 for the logistic model.

h	0.08	0.1	0.13	0.15	0.18	0.2	0.23	0.25
QAL	0.059	0.054	0.052	0.053	0.052	0.053	0.061	0.072
MOD	0.045	0.039	0.040	0.039	0.041	0.043	0.053	0.062

Table 4
Observed frequencies of rejection under the null hypothesis under C_0 for the Gamma model when $c = 2$.

		0.10	0.15	0.20	0.25
C_0	DEV	0.086	0.079	0.075	0.083
	MOD	0.079	0.070	0.073	0.073

the robust procedure that are quite close to the corresponding ones under C_0. On the other hand, contaminating only on the carriers triplicates of the mean square error of the classical estimators $\widehat{\eta}$. Therefore, as expected large responses affect the estimators of the nonparametric component more than leverage points. It is worth noting that for the studied Gamma model, both the bias and the standard deviation of the classical estimators of β are increased under C_2 and so, they both enlarge the mean square error. On the other hand, the increased mean square error obtained under C_3 is mainly due to the bias.

6.2. Performance of the tests

We have also performed a simulation study to compare the behavior of the proposed tests with respect to the classical ones, both based on the three-step procedure described in Section 2. For the two models considered below, we are interested in testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$, where $\beta_0 = 0.5$ and $\beta_0 = 2$ for the logistic and Gamma models, respectively. We studied the behavior of the tests under the null hypothesis and under contiguous alternatives by reporting the relative frequencies of rejection.

6.2.1. Logistic model
The simulation conditions were analogous to those described in Section 6.1.1, but we have also considered a second contamination indicated as C_2 and defined as follows. Let $p_i^\star = H(-0.5x_i + \eta_0(t_i))$. As above, we have generated a sample $u_i \sim \mathcal{U}(0, 1)$ and then, we have defined the contaminated sample $(y_{i,c}, x_{i,c}, t_i)$ as

$$x_{i,c} = \begin{cases} x_i & \text{if } u_i \leq 0.90 \\ \text{a new observation from a } N(5, 1) & \text{if } u_i > 0.90, \end{cases}$$

$$y_{i,c} = \begin{cases} y_i & \text{if } u_i \leq 0.90 \\ \text{a new observation from a } Bi(1, p_i^\star) & \text{if } u_i > 0.90. \end{cases}$$

In Table 3, we present the observed frequencies of rejection under the null hypothesis the non-contaminated case C_0, for the classical and robust procedure using different smoothing parameters. It is worth noting that, for the test based on the classical estimators the observed frequencies are higher than the nominal values, while this phenomenon is not observed for the robust Wald statistic. For the remaining of this Monte Carlo study we considered a nominal level $\alpha = 0.05$ and the bandwidth equal to $h = 0.18$ and $h = 0.23$, for the classical and robust procedures, respectively.

Fig. 8 presents the relative frequencies of rejection π for the classical and robust Wald test. The thick line correspond to the values of the observed frequencies under C_0, while the filled diamonds and the triangles to those observed under C_1 and C_2, respectively. The selected alternatives correspond to $\beta = \beta_0 + \Delta n^{-1/2}$ where $\beta_0 = 0.5$ and Δ taking values over a non-equidistant grid of points, i.e., the null hypothesis was $H_0 : \beta = 0.5$. The grid chose was $\Delta = 0, 0.2, 0.4, 0.8, 1.2$, from 2.4 to 33.6 the grid has a step of 1.2 and finally, the values $\Delta = 42, 50.4, 58.8$ were also included. This Figure shows that the classical test is non-informative under C_1 and that it is also extremely sensitive under C_2, leading to a decreasing frequency of rejection as Δ increases. On the other hand, the robust Wald test is stable under C_1 and C_2.

6.2.2. Gamma model
The simulation conditions were analogous to those described in Section 6.1.2 and we present the results for the robust estimators computed with $c = 2$. As for the logistic model, Table 4 present the observed frequencies of rejection under the null hypothesis the non-contaminated case C_0, for the classical and robust procedure using different smoothing parameters. It is worth noting that, for the both tests the observed frequencies are higher than the nominal values and this can be explained by the sample size which is $n = 100$ while for the logistic model we considered $n = 200$. For the remaining of this Monte Carlo study we considered a nominal level $\alpha = 0.05$ and the bandwidth equal to $h = 0.15$ both for the classical and robust procedure.

Fig. 9 presents the relative frequencies of rejection π for the classical and robust Wald test. The thick line correspond to the values of the observed frequencies under C_0, while the filled diamonds, triangles and stars to those observed under C_1, C_2 and C_3, respectively, as described in Section 6.1.2. The selected alternatives correspond to $\beta = \beta_0 \pm \Delta n^{-1/2}$ where $\beta_0 = 2$
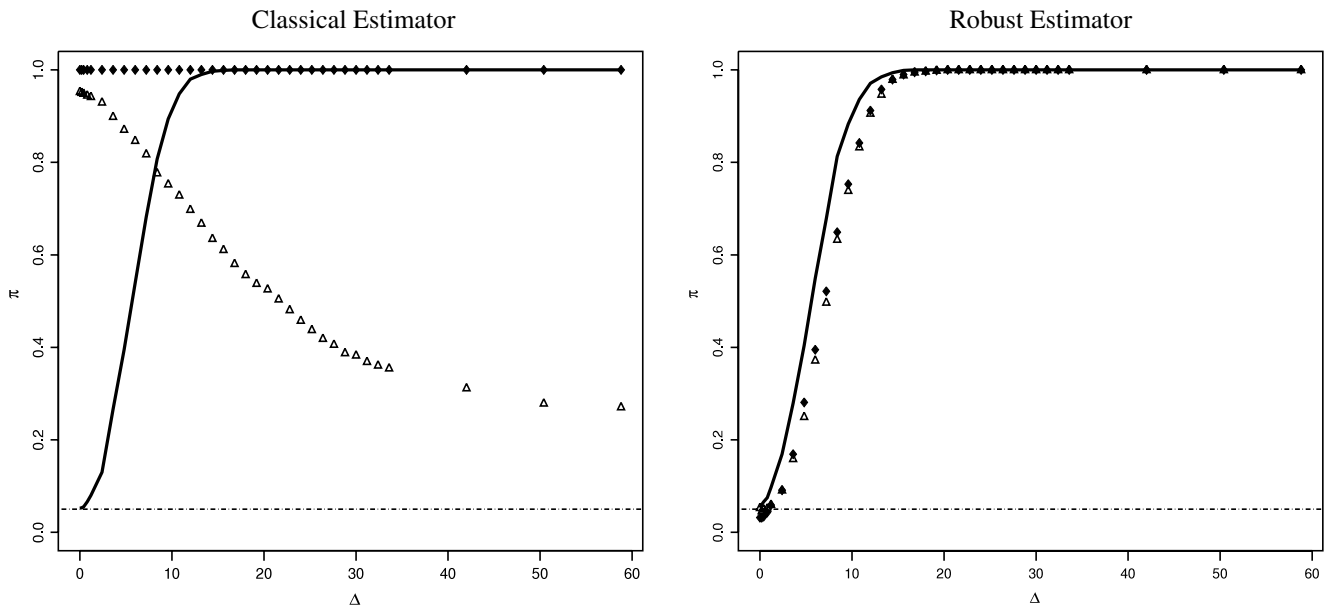
Fig. 8. Relative frequencies of rejection π for the classical and robust Wald test, under C_0 (thick line), C_1 (diamonds) and C_2 (triangles). The horizontal lines indicate the nominal level $\alpha = 0.05$.
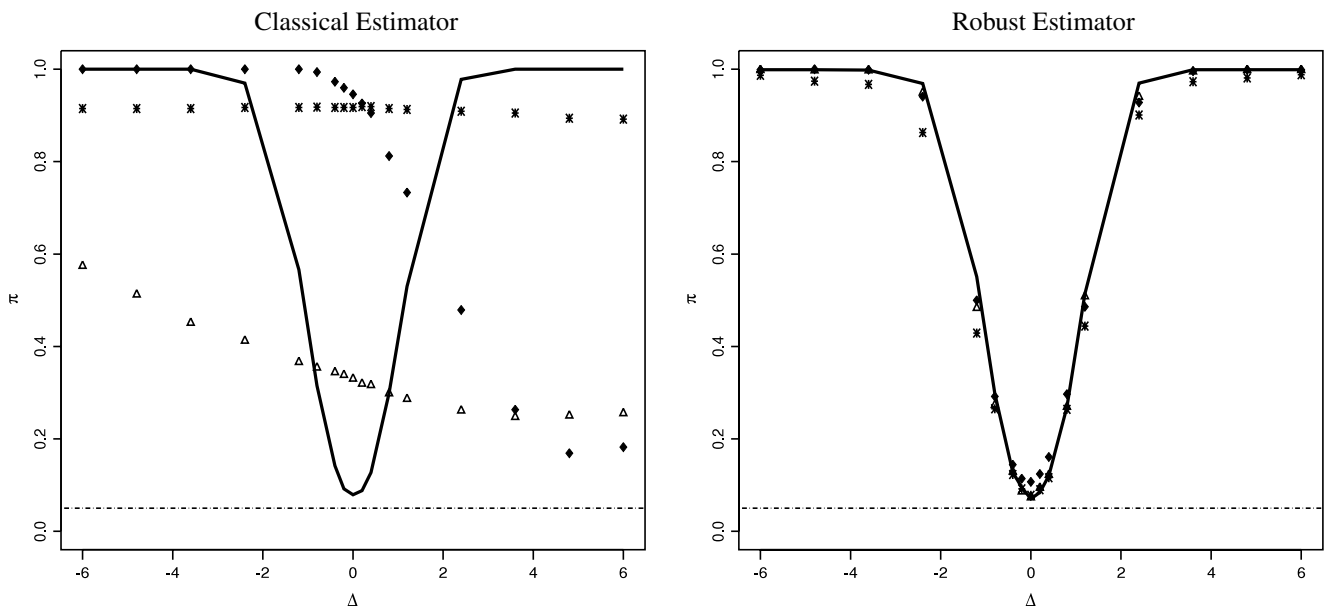

Fig. 9. Relative frequencies of rejection π for the classical and robust Wald test for the Gamma model, under C_0 (thick line), C_1 (diamonds), C_2 (triangles) and C_3 (stars). The horizontal lines indicates the nominal level $\alpha = 0.05$.

and Δ takes values 0, 0.2, 0.4, 0.8, 1.2 while from 2.4 to 6.0 the grid has a step of 1.2, i.e., the null hypothesis was $H_0 : \beta = 2$. Fig. 9 shows that the classical test has different behaviors depending on the contamination considered. Contamination C_3 has the same effect for the Gamma model than that observed for the logistic model under contamination C_1, the test is non-informative. Under C_3, the test leads to a decreasing frequency of rejection as Δ increases. Besides, under C_2 the classical test losses his power while the empirical level under C_2 is increased. The observed behavior under the null hypothesis for this contamination can be explained by the fact that, as observed in Table 2, not only the bias but also the standard deviation of the classical estimator is increased and so we do not observe the same behavior as in the other two contaminations.

As for the logistic model, the robust Wald test is stable under contamination, showing only a small loss of power under C_3.

7. Concluding remarks

The problem of estimating robustly the regression parameter and the nonparametric component under a generalized partially linear model has been considered recently. In this paper, we have introduced a family of estimators that allows us to define a resistant procedure to test hypotheses on the parametric component. Our proposal tends to overcome the

sensitivity of the classical procedure by considering a robust Wald test based on estimators that penalize observations with large Pearson residuals or large deviances and also high leverage points. The tests statistics have a limiting χ^2-distribution under the null hypothesis and under contiguous alternatives.

The simulation study confirms the expected inadequate behavior of the classical Wald test in the presence of outliers. All methods are sensitive to the choice of the smoothing parameter. Under a partially linear regression model, this fact was also noticed by González Manteiga and Aneiros Pérez (2003), who deal with classical procedures under dependent errors and by Bianco et al. (2006) who considered a robust test for the regression parameter. As mentioned by these authors, more research in this direction is necessary.

The proposed robust procedure for the regression parameter are quite stable under contamination, leading to almost the same level and power either under the central model or under the contaminations studied.

We introduced an empirical influence measure, related to the notion introduced by Mallows (1974), that allows us to evaluate on a given data set the sensitivity of the regression parameter and regression function estimators to anomalous data. It turns out that, under a generalized partially linear model, the classical approach is not robust, since it leads to unbounded empirical influence functions. On the other hand, our proposals have bounded empirical influence.

Acknowledgements

Appendix

Proof of (23). For the sake of simplicity, we will denote $\widehat{\boldsymbol{\beta}}_{0,\varepsilon}$ as $\widehat{\boldsymbol{\beta}}_{\varepsilon}$ keeping in mind that we have a new sample with the original data set representing an $1 - \varepsilon$ proportion and the new observation an ε proportion. Since $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) = \left(\partial \widehat{\boldsymbol{\beta}}_{\varepsilon} / \partial \varepsilon \right) |_{\varepsilon=0}$, in order to derive $\mathrm{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0))$ for the proposed estimator let us compute $\left(\partial \widehat{\boldsymbol{\beta}}_{\varepsilon} / \partial \varepsilon \right) |_{\varepsilon=0}$. Note that $\widehat{\boldsymbol{\beta}}_{\varepsilon}$ satisfies

$$\frac{(1-\varepsilon)}{n} \sum_{i=1}^{n} \Psi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\varepsilon} + \widetilde{\eta}_{0,\varepsilon}(t_i) \right) w_2(\mathbf{x}_i)\mathbf{x}_i + \varepsilon \Psi \left(y_0, \mathbf{x}_0^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_{\varepsilon} + \widetilde{\eta}_{0,\varepsilon}(t_0) \right) w_2(\mathbf{x}_0)\mathbf{x}_0 = 0, \tag{A.1}$$

where $(\widetilde{\boldsymbol{\beta}}_{0,\varepsilon}(\tau), \widetilde{\eta}_{0,\varepsilon}(\tau))$ are the estimators obtained with the new sample in Step 1, i.e., we have that

$$0 = \frac{1-\varepsilon}{n} \sum_{i=1}^{n} K \left(\frac{\tau - t_i}{h} \right) \Psi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widetilde{\boldsymbol{\beta}}_{0,\varepsilon}(\tau) + \widetilde{\eta}_{0,\varepsilon}(\tau) \right) w_1(\mathbf{x}_i)\mathbf{z}_i$$

$$+ \varepsilon K \left(\frac{\tau - t_0}{h} \right) \Psi \left(y_0, \mathbf{x}_0^{\mathsf{T}} \widetilde{\boldsymbol{\beta}}_{0,\varepsilon}(\tau) + \widetilde{\eta}_{0,\varepsilon}(\tau) \right) w_1(\mathbf{x}_0)\mathbf{z}_0. \tag{A.2}$$

Hence, differentiating (A.1) with respect to ε, evaluating at $\varepsilon = 0$ and using that $(\widehat{\boldsymbol{\beta}}, \widetilde{\eta})$ solve (12), we get

$$0 = -\frac{1}{n} \sum_{i=1}^{n} \Psi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i) \right) w_2(\mathbf{x}_i)\mathbf{x}_i + \Psi \left(y_0, \mathbf{x}_0^{\mathsf{T}} \widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_0) \right) w_2(\mathbf{x}_0)\mathbf{x}_0$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \chi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i) \right) w_2(\mathbf{x}_i)\mathbf{x}_i[\mathbf{x}_i^{\mathsf{T}}\mathrm{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) + \mathrm{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(t_i)]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \chi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i) \right) w_2(\mathbf{x}_i)\mathbf{x}_i[\mathbf{x}_i^{\mathsf{T}}\mathrm{EIF}(\widehat{\boldsymbol{\beta}}; (y_0, \mathbf{x}_0, t_0)) + \mathrm{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(t_i)]$$

$$+ \Psi \left(y_0, \mathbf{x}_0^{\mathsf{T}} \widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_0) \right) w_2(\mathbf{x}_0)\mathbf{x}_0. \tag{A.3}$$

Therefore, to compute the empirical influence function of $\widehat{\boldsymbol{\beta}}$ we need an expression for $\mathrm{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(t_i)$, $1 \leq i \leq n$. To that purpose, let us differentiate (A.2) with respect to ε and evaluate at $\varepsilon = 0$. We easily obtain that

$$0 = \frac{1}{n} \sum_{i=1}^{n} K \left(\frac{\tau - t_i}{h} \right) \chi \left(y_i, \mathbf{x}_i^{\mathsf{T}} \widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau) \right) w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^{\mathsf{T}} \left(\begin{array}{c} \mathrm{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(\tau) \\ \\ \left. \frac{\partial}{\partial \varepsilon} \widetilde{\boldsymbol{\beta}}_{0,\varepsilon}(\tau) \right|_{\varepsilon=0} \end{array} \right)$$

$$+ K \left(\frac{\tau - t_0}{h} \right) \Psi \left(y_0, \mathbf{x}_0^{\mathsf{T}} \widetilde{\boldsymbol{\beta}}(\tau) + \widetilde{\eta}(\tau) \right) w_1(\mathbf{x}_0)\mathbf{z}_0,$$

and so, $\mathrm{EIF}(\widetilde{\eta}; (y_0, \mathbf{x}_0, t_0))(\tau)$ is the first element of the vector $v_0(\tau)$ defined in (24). The result follows now easily using (A.3). □

In order to prove Lemma 4.1, we will need the following lemma.

Lemma A.1. *Let $(t_1, Z_1), \ldots, (t_n, Z_n)$ be i.i.d. random vectors, where Z_i are scalar random variables such that $|Z_i| \le A$. Let K be a bounded positive function, satisfying a Lipschitz condition of order one and such that $\int K(u)\mathrm{d}u = 1$. Assume that $\mathcal{T}$, the support of t, is compact and that the density f_T of t is bounded.*

(a) *For any $\epsilon > 0$ and n large enough such that $\theta_n \le 1$, we have that*

$$P\left(\theta_n^{-1} \sup_{\tau \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^n K_h(\tau - t_i)Z_i - \mathrm{E}\left(K_h(\tau - t_i)Z_i\right) \right| > 2\epsilon \right) \le 2C_1 \delta_n^{-1} \exp\{-(4C_2)^{-1} A_1^2 \epsilon^2 n h_n \theta_n^2\},$$

where $\delta_n < \epsilon \theta_n h_n^2 / (2AC_K)$, $C_2 = 2\left(1 + \frac{1}{3}\epsilon A_1\right) A^2 \|K\|_\infty \|f_T\|_\infty$, $C_1 = \mathrm{diam}(\mathcal{T}) + 1$ and C_K denotes the Lipschitz constant of the kernel K.

(b) *If, in addition, $nh^4 \to 0$ and $nh_n^2 / \log^2(1/h) \to \infty$, then*

$$n^{1/4} \sup_{\tau \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^n K_h(\tau - t_i)Z_i - \mathrm{E}\left(K_h(\tau - t_i)Z_i\right) \right| \overset{p}{\longrightarrow} 0.$$

It is worth noting that Lemma A.1 allows us to show that $\sup_{\tau \in \mathcal{T}} \left| \sum_{j=1}^n K\left((\tau - t_j)/h\right)/(nh) - f_T(\tau) \right| \overset{p}{\longrightarrow} 0$, if $nh/\log(1/h) \to \infty$. However, it is well known that using results in Dvoretzky et al. (1956), this result can be derived under the assumption that $nh_n^2 \to \infty$; see for instance, Theorem 2.1.3 in Prakasa Rao (1983).

Proof of Lemma A.1. (a) Let us denote by $Y_i(\tau) = \{K_h(\tau - t_i)Z_i - \mathrm{E}\left(K_h(\tau - t_i)Z_i\right)\}/n$ and $U_n(\tau) = \sum_{i=1}^n Y_i(\tau)$. Then, we have that $|Y_i(\tau)| \le 2A\|K\|_\infty / (nh_n) = M_n$ and

$$\mathrm{VAR}\left(Y_i(\tau)\right) \le \frac{\mathrm{E}(K_h^2(\tau - t_i)Z_i^2)}{n^2} \le \frac{A^2}{n^2 h_n^2} \int K^2\left(\frac{\tau - u}{h_n}\right) f_T(u)\mathrm{d}u$$

$$\le \frac{A^2 \|K\|_\infty \|f_T\|_\infty}{n^2 h_n^2} \int K\left(\frac{\tau - u}{h_n}\right)\mathrm{d}u = \frac{A^2 \|K\|_\infty \|f_T\|_\infty}{n^2 h_n},$$

which implies that, $\sum_{i=1}^n \mathrm{VAR}\left(Y_i(\tau)\right) \le A^2 \|K\|_\infty \|f_T\|_\infty / (nh_n) = V_n$. Therefore, using Bernstein's inequality, we get easily that

$$P\left(|U_n(\tau)| > \epsilon \theta_n\right) \le 2\exp\left\{ -\frac{1}{2}(\epsilon \theta_n)^2 \left(V_n + \frac{1}{3}\epsilon \theta_n M_n\right)^{-1} \right\}.$$

Note that $M_n V_n^{-1} = 2\|K\|_\infty / (A\|f_T\|_\infty) = A_1$ therefore, using that $\theta_n \le 1$, we obtain

$$P\left(|U_n(\tau)| > \epsilon \theta_n\right) \le 2\exp\left\{ -\frac{1}{2}(\epsilon \theta_n)^2 V_n^{-1} \left(1 + \frac{1}{3}\epsilon \theta_n A_1\right)^{-1} \right\}$$

$$\le 2\exp\left\{ -\frac{1}{2\left(1 + \frac{1}{3}\epsilon A_1\right)} (\epsilon \theta_n)^2 V_n^{-1} \right\} = 2\exp\left\{ -\frac{1}{C_2}\epsilon^2 \theta_n^2 n h_n \right\}. \tag{A.4}$$

Denote by $\mathcal{V}_i, 1 \le i \le \ell_n$ a finite covering of $\mathcal{T}$ where $\mathcal{V}_i$ are closed balls centered at points $\tau_i \in \mathcal{T}$ with radius $\delta_n < \epsilon \theta_n h_n^2 / (2AC_K)$. Hence, $\ell_n \le C_1 \delta_n^{-1}$ and for any $\tau \in \mathcal{V}_j$, we have that

$$|U_n(\tau) - U_n(\tau_j)| \le \frac{A}{h}\left(\frac{1}{n} \sum_{i=1}^n \left| K\left(\frac{\tau - t_i}{h}\right) - K\left(\frac{\tau_j - t_i}{h}\right) \right| + \mathrm{E}\left| K\left(\frac{\tau - t_1}{h}\right) - K\left(\frac{\tau_j - t_1}{h}\right) \right| \right)$$

$$\le 2AC_K \frac{\delta_n}{h_n^2} < \epsilon \theta_n,$$

which implies that

$$\sup_{\tau \in \mathcal{T}} |U_n(\tau)| \le \max_{1 \le j \le \ell_n} |U_n(\tau_j)| + \max_{1 \le j \le \ell_n} \sup_{\tau \in \mathcal{V}_j} |U_n(\tau) - U_n(\tau_j)| \le \epsilon \theta_n + \max_{1 \le j \le \ell_n} |U_n(\tau_j)|,$$

and so, using (A.4), we get

$$P\left(\theta_n^{-1} \sup_{\tau \in \mathcal{T}} |U_n(\tau)| > 2\epsilon \right) \le P\left(\theta_n^{-1} \max_{1 \le j \le \ell_n} |U_n(\tau_j)| > \epsilon \right) \le 2\ell_n \exp\left\{ -\frac{1}{C_2}\epsilon^2 \theta_n^2 n h_n \right\},$$

concluding the proof of (a).

(b) The proof of (b) follows easily by applying (a) with $\theta_n = n^{-1/4}$. Effectively, given $\epsilon > 0$, we have that $\delta_n = h_n^3 < \epsilon \theta_n h_n^2/(2AC_K)$ for n large enough, since $nh^4 \to 0$. Then,

$$P\left(n^{1/4} \sup_{\tau \in \mathcal{T}} \left| \frac{1}{n} \sum_{i=1}^n K_h(\tau - t_i)Z_i - \mathrm{E}\left(K_h(\tau - t_i)Z_i\right) \right| > 2\epsilon\right) \le 2C_1 \frac{1}{h_n^3} \exp\left\{-(4C_2)^{-1}A_1^2\epsilon^2 nh_n n^{-\frac{1}{2}}\right\}$$

$$\le 2C_1 \exp\left\{3\log(1/h_n) - (4C_2)^{-1}A_1^2\epsilon^2 n^{\frac{1}{2}}h_n\right\},$$

which concludes the proof since $n^{\frac{1}{2}}h_n/\log(1/h_n) \to \infty$. $\quad \square$

Proof of Lemma 4.1. Let $\Delta_n(\tau) = \begin{pmatrix} \widetilde{\eta}(\tau) - \eta_0(\tau) \\ \widetilde{\beta}(\tau) - \beta_0 \end{pmatrix}$, then, we have that $S_n^1(\widetilde{\beta}(\tau), \widetilde{\eta}(\tau), \tau) = 0$. A Taylor expansion of order two at $(\beta_0, \eta_0(\tau))$ lead to $0 = W_n(t) + \widetilde{\mathbf{A}}_n(t)\Delta_n(t)$ where $\widetilde{\mathbf{A}}_n(t) = \mathbf{A}_n(t) + \mathbf{B}_n(t)$

$$W_n(t) = \frac{1}{n}\sum_{i=1}^n \Psi(y_i, \mathbf{x}_i^\mathsf{T}\beta_0 + \eta_0(t))K_h(t - t_i)w_1(\mathbf{x}_i)\mathbf{z}_i,$$

$$\mathbf{A}_n(t) = \frac{1}{n}\sum_{i=1}^n \chi(y_i, \mathbf{x}_i^\mathsf{T}\beta_0 + \eta_0(t))K_h(t - t_i)w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^\mathsf{T},$$

$$\mathbf{B}_n(t) = \frac{1}{2n}\sum_{i=1}^n \chi_1(y_i, \xi_i(t))(\Delta_n(t)^\mathsf{T}\mathbf{z}_i)K_h(t - t_i)w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^\mathsf{T},$$

with $\xi_i(\tau)$ an intermediate point between $\mathbf{z}_i^\mathsf{T}\begin{pmatrix} \widetilde{\eta}(\tau) \\ \widetilde{\beta}(\tau) \end{pmatrix}$ and $\mathbf{z}_i^\mathsf{T}\begin{pmatrix} \eta_0(\tau) \\ \beta_0 \end{pmatrix}$.

Let us show that,

(i) $\|\mathbf{B}_n\|_\infty = O_P(\|\Delta_n\|_\infty)$,

(ii) $\sup_{\tau \in \mathcal{T}} |\mathbf{A}_n(\tau) - f_T(\tau)\mathbf{A}(\tau)| = O_P(h^2) + o_p(n^{-1/4})$,

where $\mathbf{A}(\tau) = \mathrm{E}_0\left(\chi(y, \mathbf{x}^\mathsf{T}\beta_0 + \eta_0(t))w_1(\mathbf{x})\mathbf{z}\mathbf{z}^\mathsf{T}|t = \tau\right)$.

(i) Using N2, the fact that $\lambda_1(\mathbf{x}) = \|\mathbf{x}\|^3 w_1(\mathbf{x})$ is bounded, we obtain that

$$|\mathbf{B}_n(\tau)| \le \frac{1}{2}\|\Delta_n(\tau)\| \, \|\chi_1\|_\infty \max\{\|w_1\|_\infty, \|\lambda_1\|_\infty\} \frac{1}{nh}\sum_{i=1}^n K\left(\frac{t_i - t}{h}\right).$$

Then, as $\mathcal{T}$ is a compact set, f_T is bounded and $nh_n^2 \to \infty$, we get that

$$\sup_{\tau \in \mathcal{T}} \frac{1}{nh}\sum_{i=1}^n K\left(\frac{\tau - t_i}{h}\right) = O_p(1),$$

since $\sup_{\tau \in \mathcal{T}} \left| \sum_{j=1}^n K\left((\tau - t_j)/h\right)/(nh) - f_T(\tau)\right| \xrightarrow{p} 0$, which concludes the proof of (i).

To derive (ii), applying Lemma A.1 to each component of $\mathbf{A}_n(\tau)$, we get that $\mathbf{A}_n(\tau) = \mathrm{E}_0(\mathbf{A}_n(\tau)) + o_P(n^{-1/4})$, uniformly for $\tau \in \mathcal{T}$. Let us compute $\mathrm{E}_0(\mathbf{A}_n(\tau))$,

$$\mathrm{E}_0(\mathbf{A}_n(\tau)) = \frac{1}{nh}\sum_{i=1}^n \mathrm{E}_0\left(K\left(\frac{\tau - t_i}{h}\right)\chi(y_i, \mathbf{x}_i^\mathsf{T}\beta_0 + \eta_0(\tau))w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^\mathsf{T}\right)$$

$$= \frac{1}{nh}\sum_{i=1}^n \mathrm{E}_0\left(K\left(\frac{\tau - t_i}{h}\right)\mathrm{E}_0\left(\chi(y_i, \mathbf{x}_i^\mathsf{T}\beta_0 + \eta_0(\tau))w_1(\mathbf{x}_i)\mathbf{z}_i\mathbf{z}_i^\mathsf{T}|t_i\right)\right)$$

$$= \mathrm{E}\left(\frac{1}{h}K\left(\frac{\tau - t}{h}\right)\mathbf{A}(t)\right).$$

Through a change of variables in the integrand, using a Taylor expansion and the fact that the kernel is even, straightforward calculations lead to $\sup_{\tau \in \mathcal{T}} |\mathrm{E}\left(K\left((\tau - t)/h\right)\mathbf{A}(t)\right)/h - \mathbf{A}(t)f_T(\tau)| = O(h^2)$ and so, $\mathbf{A}_n(\tau) = f_T(\tau)\mathbf{A}(\tau) + O_P(h^2) + o_p(n^{-1/4})$, concluding the proof of (ii). In particular, $\mathbf{A}_n(\tau) = f_T(\tau)\mathbf{A}(\tau) + o_p(1)$.

Let us begin by proving (a). From (i) and (ii), and the fact that $\inf_{\tau \in \mathcal{T}} \lambda_{p+1}(\mathbf{A}(\tau)) > 0$, we obtain that $\widetilde{\mathbf{A}}_n(\tau)$ is non-singular and thus to show that $n^{\frac{1}{4}}\|\Delta_n\|_\infty \xrightarrow{p} 0$, it suffices to show that $n^{\frac{1}{4}}\|W_n\|_\infty \xrightarrow{p} 0$. Applying Lemma A.2 to the bounded variables $Z_i = \Psi(y_i, \mathbf{x}_i^\mathsf{T}\beta_0 + \eta_0(t))w_1(\mathbf{x}_i)(\mathbf{z}_i)_j$, where $(\mathbf{z}_i)_j$ denotes the jth component of $\mathbf{z}_i$, we get that $n^{\frac{1}{4}}\|W_n - E(W_n)\|_\infty \xrightarrow{p} 0$.

Besides, note that, since N5 implies that $E_0 \Psi(y, \mathbf{x}^T\boldsymbol{\beta}_0 + \eta_0(t))K_h(\tau - t)w_1(\mathbf{x})\mathbf{z} = 0$, using a Taylor's expansion of order one we get

$$E(W_n(\tau)) = E_0 \Psi(y, \mathbf{x}^T\boldsymbol{\beta}_0 + \eta_0(\tau))K_h(\tau - t)w_1(\mathbf{x})\mathbf{z}$$
$$= E_0 \chi(y, \mathbf{x}^T\boldsymbol{\beta}_0 + \xi(\tau))(\eta_0(\tau) - \eta_0(t))K_h(\tau - t)w_1(\mathbf{x})\mathbf{z},$$

where $\xi(\tau)$ is an intermediate point between $\eta_0(t)$ and $\eta_0(\tau)$. The Lipschitz continuity of η_0 implies that

$$\sup_{\tau \in \mathcal{T}} |n^{\frac{1}{4}} E(W_n(\tau))| \le C_{\eta_0} \|\chi\|_\infty \|\psi_1\|_\infty \|\eta_0'\|_\infty n^{\frac{1}{4}} \sup_{\tau \in \mathcal{T}} E|(\tau - t)K_h(\tau - t)|,$$

where C_{η_0} stands for the Lipschitz constant of η_0. Denote by $K_1(u) = |u|K(u)$, then, we get that $n^{\frac{1}{4}} E|(\tau - t)K_h(\tau - t)| \le n^{\frac{1}{4}} h \int K_1(u) f_T(\tau - uh) \mathrm{d}u \le \|f_T\|_\infty n^{\frac{1}{4}} h \int K_1(u) \mathrm{d}u$ which together with the fact that $nh^4 \to 0$ concludes the proof of (a). Similarly, to obtain (b) we notice that using (i) and (ii), we get that

$$0 = W_n(\tau) + f_T(\tau)\mathbf{A}(\tau)\Delta_n(\tau)[1 + O_P(h^2) + o_p(n^{-1/4})] + O_P(\|\Delta_n(\tau)\|^2),$$

which implies that

$$\Delta_n(\tau) + [f_T(\tau)\mathbf{A}(\tau)]^{-1}W_n(\tau) = \Delta_n(\tau)(O_P(h^2) + o_p(n^{-1/4})) + O_P(\|\Delta_n(\tau)\|^2).$$

Therefore, using that from (a), $n^{1/4}\|\Delta_n\|_\infty \xrightarrow{p} 0$ and the fact that $nh^4 \to 0$, we conclude the proof. $\square$

The following lemma is needed for the proof of Theorem 4.1.

Lemma A.2. *Let us assume that t_i have compact support $\mathcal{T}$ and that* N2 *and* N6 *hold. Let $\widehat{\boldsymbol{\beta}}$ be any consistent estimator of $\boldsymbol{\beta}$ and $\widetilde{\eta}$ be uniform consistent estimators of η, i.e., $\|\widetilde{\eta} - \eta\|_\infty \xrightarrow{p} 0$. Then, we have that*

$$\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widetilde{\eta}) \xrightarrow{p} \mathbf{A},$$

where $\widehat{\mathbf{A}}(\boldsymbol{\beta}, \eta)$ is defined in (18).

Proof of Lemma A.2. It is easy to see that $\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widetilde{\eta})$ can be written as

$$\widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}}, \widetilde{\eta}) = \frac{1}{n}\sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \eta_0(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T + \frac{1}{n}\sum_{i=1}^n \chi_1\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \xi_{ni}\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T(\widetilde{\eta}(t_i) - \eta_0(t_i))$$
$$= \mathbf{A}_n^{(1)} + \mathbf{A}_n^{(2)},$$

with $\xi_{ni} = \theta_{ni}\widetilde{\eta}(t_i) + (1 - \theta_{ni})\eta_0(t_i)$ and intermediate point, $0 \le \theta_{ni} \le 1$. N2, N6 and the fact that $\|\widetilde{\eta} - \eta\|_\infty \xrightarrow{p} 0$ entails that $\mathbf{A}_n^{(2)} \xrightarrow{a.s.} 0$. On the other hand, N2 implies that $\lambda(\boldsymbol{\beta}) = E_0(\chi(y, \mathbf{x}^T\boldsymbol{\beta} + \eta_0(t))w_2(\mathbf{x})\mathbf{x}\mathbf{x}^T)$ is a continuous function, and thus, from the fact that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$ we obtain that

$$E_0\left[\chi\left(y, \mathbf{x}^T\widehat{\boldsymbol{\beta}} + \eta_0(t)\right) w_2(\mathbf{x})\mathbf{x}\mathbf{x}^T\right] - E_0\left[\chi\left(y, \mathbf{x}^T\boldsymbol{\beta}_0 + \eta_0(t)\right) w_2(\mathbf{x})\mathbf{x}\mathbf{x}^T\right] \xrightarrow{p} 0.$$

Therefore, it is enough to show that

$$\frac{1}{n}\sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \eta_0(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T - E_0\left[\chi\left(y, \mathbf{x}^T\widehat{\boldsymbol{\beta}} + \eta_0(t)\right) w_2(\mathbf{x})\mathbf{x}\mathbf{x}^T\right] \xrightarrow{p} 0.$$

Define the following class of functions $\mathcal{H} = \{\chi(y, \mathbf{x}^T\boldsymbol{\beta} + \eta_0(t)) w_2(\mathbf{x})\mathbf{x}\mathbf{x}^T, \boldsymbol{\beta} \in \mathcal{K}\}$ with $\mathcal{K}$ a compact neighborhood of $\boldsymbol{\beta}_0$. Using analogous arguments to those considered in Lemma 1 from Bianco and Boente (2002), we obtain that $\mathbf{A}_n^{(1)} \xrightarrow{p} \mathbf{A}$. $\square$

Proof of Theorem 4.1. Let $\widehat{\boldsymbol{\beta}}$ be a solution of $L_n^1(\boldsymbol{\beta}, \widetilde{\eta}) = 0$ defined in (12). Using a Taylor expansion of order one, we get

$$0 = \frac{1}{n}\sum_{i=1}^n \Psi\left(y_i, \mathbf{x}_i^T\widehat{\boldsymbol{\beta}} + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i$$
$$= \frac{1}{n}\sum_{i=1}^n \Psi\left(y_i, \mathbf{x}_i^T\boldsymbol{\beta}_0 + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i + \frac{1}{n}\sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\boldsymbol{\beta}^* + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right),$$

with $\boldsymbol{\beta}^*$ an intermediate point between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. Lemma A.2 entails that

$$\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^n \chi\left(y_i, \mathbf{x}_i^T\boldsymbol{\beta}^* + \widetilde{\eta}(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T \xrightarrow{p} \mathbf{A},$$

$$= \frac{n-1}{\sqrt{n}} \int_{\mathcal{T}} \boldsymbol{\gamma}(\tau_1) f_T(\tau_1) \int_{\mathcal{T}} m_1(\tau_1, \tau_2) \frac{1}{h} K\left(\frac{\tau_1 - \tau_2}{h}\right) f_T(\tau_2) \mathrm{d}\tau_2 \mathrm{d}\tau_1.$$

Using N7, N8 and N9, we get that

$$\int m_1(\tau, \tau_2) \frac{1}{h} K\left(\frac{\tau - \tau_2}{h}\right) f_T(\tau_2) \mathrm{d}\tau_2 = \int m_1(\tau, \tau - uh) K(u) f_T(\tau - uh) \mathrm{d}u$$

$$= \int [-uhm_1'(\tau, \tau) + u^2 h^2 m_1''(\tau, \xi_1)] K(u)[f_T(\tau) - uhf_T'(\xi_2)] \mathrm{d}u$$

$$= h^2 \left[m_1'(\tau, \tau) \int u^2 K(u) f_T'(\xi_2) \mathrm{d}u + f_T(\tau) \int u^2 K(u) m_1''(\tau, \xi_1) \mathrm{d}u - h \int u^3 K(u) m_1''(\tau, \xi_1) f_T'(\xi_2) \mathrm{d}u \right]$$

$$= O(h^2),$$

which entails that $\|E_0(\mathbf{B}_n^{(7)})\| = ((n-1)/\sqrt{n})O(h^2) = O(n^{1/2}h^2) \to 0$ since N2 holds and $\mathcal{T}$ is a compact set. To compute the variance of each component of $\mathbf{B}_n^{(7)}$, the derivations are quite similar. First notice that

$$\mathbf{B}_n^{(7)} = \frac{1}{n^{3/2}} \sum_{i \neq j} \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right) \boldsymbol{\gamma}(t_i)[G_i(t_j) - G_i(t_i)] = \frac{1}{n^{3/2}} \sum_{i \neq j} \mathbf{V}_{ij},$$

then,

$$\mathrm{Cov}\,(\mathbf{B}_n^{(7)}, \mathbf{B}_n^{(7)}) = \frac{1}{n^3} \sum_{i \neq j} \mathrm{Cov}\,(\mathbf{V}_{ij}, \mathbf{V}_{ij}) + \frac{1}{n^3} \sum_{i \neq j} \sum_{l \neq s} \mathrm{Cov}\,(\mathbf{V}_{ij}, \mathbf{V}_{ls})$$

$$= \frac{n(n-1)}{n^3} \mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{12}) + \frac{n^2(n-1)}{n^3}[\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{13}) + \mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{32})].$$

On the other hand, using that $m_{is}(u_1, u_2, u, u) = 0$ and $m_{is}(u, u, u_3, u_4) = 0$, we have that

$$\mathrm{tr}(\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{12})) \leq \frac{1}{h^2} E_0\left(K^2\left(\frac{t_1 - t_2}{h}\right) \mathrm{tr}(\boldsymbol{\gamma}(t_1)\boldsymbol{\gamma}^{\mathrm{T}}(t_1))[G_1(t_1) - G_1(t_2)]^2 \right)$$

$$= \frac{1}{h^2} E_0\left(K^2\left(\frac{t_1 - t_2}{h}\right) \mathrm{tr}(\boldsymbol{\gamma}(t_1)\boldsymbol{\gamma}^{\mathrm{T}}(t_1)) m_{11}(t_1, t_2, t_1, t_2) \right)$$

$$= \frac{1}{h} \int K^2(z) \mathrm{tr}(\boldsymbol{\gamma}(u)\boldsymbol{\gamma}^{\mathrm{T}}(u)) m_{11}(u, u - zh, u, u - zh) f_T(u) f_T(u - zh) \, \mathrm{d}u \, \mathrm{d}z$$

$$= \frac{1}{2h} \int K^2(z) \mathrm{tr}(\boldsymbol{\gamma}(u)\boldsymbol{\gamma}^{\mathrm{T}}(u)) m_{11,44}''(u, u - zh, u, \xi) h^2 z^2 f_T(u) f_T(u - zh) \, \mathrm{d}u \, \mathrm{d}z,$$

which entails that $\mathrm{tr}(\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{12})) = O(h)$. Therefore, $n(n-1)\mathrm{tr}(\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{12}))/n^3 \to 0$. Let us compute $\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{13})$. Note that

$$E_0(\mathbf{V}_{12}) = \int \frac{1}{h} K\left(\frac{u - v}{h}\right) \boldsymbol{\gamma}(u) m_1(u, v) f_T(u) f_T(v) \, \mathrm{d}u \, \mathrm{d}v$$

$$= \int K(z) \boldsymbol{\gamma}(u) m_1(u, u - zh) f_T(u) f_T(u - zh) \, \mathrm{d}u \, \mathrm{d}z$$

$$= O(h),$$

$$\mathrm{tr}(E_0(\mathbf{V}_{12}\mathbf{V}_{13}^{\mathrm{T}})) = \int \frac{1}{h^2} K\left(\frac{u - v}{h}\right) K\left(\frac{u - z}{h}\right) \mathrm{tr}(\boldsymbol{\gamma}(u)\boldsymbol{\gamma}^{\mathrm{T}}(u)) m_{11}(u, v, u, z) f_T(u) f_T(v) f_T(z) \, \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}z$$

$$= \int K(z) K(v) \mathrm{tr}(\boldsymbol{\gamma}(u)\boldsymbol{\gamma}^{\mathrm{T}}(u)) m_{11}(u, u - hv, u, u - hz) f_T(u) f_T(u - hv) f_T(u - hz) \, \mathrm{d}u \, \mathrm{d}v \, \mathrm{d}z$$

$$= O(h^2),$$

and so, we obtain that $\mathrm{tr}(\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{13})) = O(h^2)$. In an analogous way, we get that $\mathrm{tr}(\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{32})) = O(h^2)$ which implies that

$$\mathrm{tr}(\mathrm{Cov}\,(\mathbf{B}_n^{(7)}, \mathbf{B}_n^{(7)})) = \frac{n-1}{n^2} \mathrm{tr}(\mathrm{VAR}\,(\mathbf{V}_{12})) + \frac{n-1}{n} \mathrm{tr}[\mathrm{Cov}\,(\mathbf{V}_{12}, \mathbf{V}_{13}) + \mathrm{Cov}\,(\mathbf{V}_{12,r}, \mathbf{V}_{32})] = O(hn^{-1} + h^2).$$

The convergence of $\mathbf{B}_n^{(5)}$ and $\mathbf{B}_n^{(6)}$ are obtained straightforwardly using similar arguments to those considered with $\mathbf{B}_n^{(7)}$. Let us compute their expectation.

$$E_0(\mathbf{B}_n^{(6)}) = \frac{n(n-1)}{n\sqrt{n}} E_0 \left(\frac{1}{h} K \left(\frac{t_1 - t_2}{h} \right) [\boldsymbol{\gamma}(t_2) - \boldsymbol{\gamma}(t_1)] G_1(t_2) \right)$$

$$= \frac{(n-1)}{\sqrt{n}} E \left(\frac{1}{h} K \left(\frac{t_1 - t_2}{h} \right) [\boldsymbol{\gamma}(t_1) - \boldsymbol{\gamma}(t_2)] m_1(t_1, t_2) \right)$$

$$= \frac{n-1}{\sqrt{n}} \int \frac{1}{h} K \left(\frac{u-v}{h} \right) [\boldsymbol{\gamma}(u) - \boldsymbol{\gamma}(v)] m_1(u, v) f_T(u) f_T(v) \mathrm{d}u \mathrm{d}v.$$

Through a Taylor expansion of order one and using that $m_1(u, u) = 0$, we get

$$E_0(\mathbf{B}_n^{(6)}) = \frac{n-1}{\sqrt{n}} K(z) [\boldsymbol{\gamma}(u) - \boldsymbol{\gamma}(u - hz)] m_1(u, u - hz) f_T(u) f_T(u - zh) \mathrm{d}u\mathrm{d}z$$

$$= \frac{n-1}{\sqrt{n}} \int f_T(u) \int K(z) \boldsymbol{\gamma}'(\xi_1) hz \left(-hzm_1'(u, \xi_2) \right) f_T(u - hz) \mathrm{d}z\mathrm{d}u.$$

Thus, using N8, we get that $E_0(\mathbf{B}_n^{(6)}) = O(n^{1/2}h^2)$. Using analogous arguments to those considered above, it is easy to show that the trace of its covariance matrix converges to 0, which entails that $\mathbf{B}_n^{(6)} \xrightarrow{p} 0$.

Finally, let us show that $E_0 \left(\mathbf{B}_n^{(5)} \right) = 0$ since $R(y, \mathbf{x}, t) = \chi \left(y, \mathbf{x}^T \boldsymbol{\beta}_0 + \eta_0(t) \right) w_2(\mathbf{x}) \mathbf{x} f_T^{-1}(t)$ and $\gamma(\tau) = E_0(R(y, \mathbf{x}, t)|t) = \tau$. Effectively,

$$E_0(\mathbf{B}_n^{(5)}) = \frac{n(n-1)}{n\sqrt{n}} E_0 \left(\frac{1}{h} K \left(\frac{t_1 - t_2}{h} \right) [R(y_2, \mathbf{x}_2, t_2) - \boldsymbol{\gamma}(t_2)] G_1(t_2) \right)$$

$$= \frac{n-1}{\sqrt{n}} E_0 \left(\frac{1}{h} K \left(\frac{t_1 - t_2}{h} \right) G_1(t_2) E_0(R(y_2, \mathbf{x}_2, t_2) - \boldsymbol{\gamma}(t_2)|\mathbf{x}_1, y_1, t_1, t_2) \right)$$

$$= \frac{n-1}{\sqrt{n}} E_0 \left(\frac{1}{h} K \left(\frac{t_1 - t_2}{h} \right) G_1(t_2) [E_0(R(y_2, \mathbf{x}_2, t_2)|t_2) - \boldsymbol{\gamma}(t_2)] \right) = 0.$$

Using analogous arguments, it is easy to see that the covariance terms appearing in the expansion of the variance are 0 and that $\mathrm{tr}(\mathrm{VAR}(\mathbf{B}_n^{(5)})) = O((nh)^{-1})$, concluding the proof. $\square$

Proof of Theorem 4.2. Analogous arguments to those considered in Lemma A.2 allow us to show that, when $\boldsymbol{\beta} = \boldsymbol{\beta}_1$, $\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{\Sigma}_{\boldsymbol{\beta}_1}$ and so, (i) and (ii) follow easily from Theorem 4.1.

In order to prove (iii), we will use Theorem 6.6 in Van der Vaart (1998). Therefore, we need to obtain the asymptotic distribution of $(\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \ln(q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})/p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})))$, where $p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})$ is the joint density under the null hypothesis and $q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})$ is the corresponding one under the alternative, $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{t} = (t_1, \ldots, t_n)^T$.

Let $\theta_n(\mathbf{x}_i, t_i) = \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i) + \mathbf{x}_i^T \mathbf{c} n^{-1/2} = \theta(\mathbf{x}_i, t_i) + \mathbf{x}_i^T \mathbf{c} n^{-1/2}$ and we consider

$$\frac{q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})}{p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})} = \prod_{i=1}^{n} \frac{\exp \left\{ y_i \theta(\mathbf{x}_i, t_i) + y_i \mathbf{x}_i^T \mathbf{c} n^{-1/2} - B(\theta(\mathbf{x}_i, t_i) + \mathbf{x}_i^T \mathbf{c} n^{-1/2}) + C(y_i) \right\}}{\exp \left\{ y_i \theta(\mathbf{x}_i, t_i) - B(\theta(\mathbf{x}_i, t_i)) + C(y_i) \right\}}$$

$$= \prod_{i=1}^{n} \exp \left\{ y_i \mathbf{x}_i^T \mathbf{c} n^{-1/2} + B(\theta(\mathbf{x}_i, t_i)) - B(\theta(\mathbf{x}_i, t_i) + \mathbf{x}_i^T \mathbf{c} n^{-1/2}) \right\},$$

then

$$\ln \frac{q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})}{p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i \mathbf{x}_i^T \mathbf{c} - \sum_{i=1}^{n} B'(\theta(\mathbf{x}_i, t_i)) \mathbf{x}_i^T \mathbf{c} n^{-1/2} - \frac{1}{2n} \sum_{i=1}^{n} B''(\theta(\mathbf{x}_i, t_i))(\mathbf{x}_i^T \mathbf{c})^2$$

$$- \frac{1}{2n} \sum_{i=1}^{n} \left[B''(\xi_i) - B''(\theta(\mathbf{x}_i, t_i)) \right] (\mathbf{x}_i^T \mathbf{c})^2.$$

Since $B'(\theta(\mathbf{x}_i, t_i)) = H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))$ we have that,

$$\ln \frac{q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})}{p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)) \right] \mathbf{x}_i^T \mathbf{c} - \frac{1}{2n} \sum_{i=1}^{n} H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))(\mathbf{x}_i^T \mathbf{c})^2$$

$$- \frac{1}{2n} \sum_{i=1}^{n} \left[B''(\xi_i) - B''(\theta(\mathbf{x}_i, t_i)) \right] (\mathbf{x}_i^T \mathbf{c})^2.$$

It is easy to see that $\ln(q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})/p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})) \xrightarrow{\mathcal{D}} N(-\sigma^2/2, \sigma^2)$ with $\sigma^2 = \mathbf{c}^T E(H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)) \mathbf{x}_i \mathbf{x}_i^T) \mathbf{c}$.

In the proof of Theorem 4.1 we obtained that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}^{-1}\mathbf{C}_n + o_p(1)$, where $\mathbf{A}$ is defined in N3 and

$$\mathbf{C}_n = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} [\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)\right) w_2(\mathbf{x}_i)\mathbf{x}_i + \boldsymbol{\gamma}(t_i)G_i(t_i)f_T(t_i)]$$

$$= \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)\right) [w_2(\mathbf{x}_i)\mathbf{x}_i + w_1(\mathbf{x}_i)D_1(\mathbf{x}_i, t_i)\mathbf{b}(t_i)],$$

with $D_1(\mathbf{x}_i, t_i)$ the first component of $\mathbf{A}(t_i)^{-1}\mathbf{z}_i$ and $\mathbf{b}(t_i) = \mathrm{E}_0(\chi(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i))\mathbf{x}w_2(\mathbf{x})|t = t_i)$.

Then, to derive the joint asymptotic distribution of $(\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^{\mathrm{T}}, \ln(q_n(\mathbf{y}, \mathbf{X}, \mathbf{t})/p_n(\mathbf{y}, \mathbf{X}, \mathbf{t})))^{\mathrm{T}}$, it is enough to compute the covariance between $-\mathbf{C}_n$ and $R_1 = \sum_{i=1}^{n} [y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i))]\mathbf{x}_i^{\mathrm{T}}\mathbf{c}/\sqrt{n}$. Using N5, we get that

$$\mathrm{Cov}\left(-\mathbf{C}_n, R_1\right) = \mathrm{Cov}\left(-\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)\right) [w_2(\mathbf{x}_i)\mathbf{x}_i + w_1(\mathbf{x}_i)D_1(\mathbf{x}_i, t_i)\mathbf{b}(t_i)], [y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i))]\mathbf{x}_i^{\mathrm{T}}\mathbf{c}\right)$$

$$= -\mathrm{E}_0[(y_i - H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)))\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)\right) [w_2(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} + w_1(\mathbf{x}_i)D_1(\mathbf{x}_i, t_i)\mathbf{b}(t_i)\mathbf{x}_i^{\mathrm{T}}]]\mathbf{c}.$$

It is easy to see that N5 entails that

$$\mathrm{E}_0[(y_1 - H(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)))\Psi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)\right)|(\mathbf{x}_1, t_1)] = -\mathrm{E}_0\chi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)),$$

and so, $\mathrm{Cov}\left(-\mathbf{C}_n, R_1\right) = \mathbf{Ac} + \mathbf{Bc}$ with $\mathbf{B} = \mathrm{E}_0[\chi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1))w_1(\mathbf{x}_1)D_1(\mathbf{x}_1, t_1)\mathbf{b}(t_1)\mathbf{x}_1^{\mathrm{T}}]$. Denote by $\mathbf{a}^{(1)}(t_1), \ldots, \mathbf{a}^{(p)}(t_1)$ the rows of $\mathbf{A}(t_1)^{-1}$. Therefore, we have that

$$\mathrm{E}_0\left(\chi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1))w_1(\mathbf{x}_1)D_1(\mathbf{x}_1, t_1)\mathbf{b}(t_1)\mathbf{x}_1^{\mathrm{T}}\right) = \mathrm{E}_0\left(\mathbf{b}(t_1)\mathbf{a}^{(1)}(t_1)\mathrm{E}\left(\chi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1))w_1(\mathbf{x}_1)\mathbf{z}_i\mathbf{x}_1^{\mathrm{T}}\Big|t_1\right)\right).$$

The proof follows using straightforward calculations that lead to $\mathbf{a}^{(1)}(t_1)\mathrm{E}\left(\chi(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)w_1(\mathbf{x}_1)\mathbf{z}_i\mathbf{x}_1^{\mathrm{T}}|t_1)\right) = 0$ and so, $\mathbf{B} = 0$. $\quad\square$

References

Aït Sahalia, Y., 1995. The delta method for nonparametric kernel functionals. Ph.D. Dissertation. University of Chicago.

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

Bianco, A., Boente, G., 2004. Robust estimators in semiparametric partly linear regression models. J. Statist. Plan. Inference 122, 229–252.

Bianco, A., Boente, G., Martínez, E., 2006. Robust tests in semiparametric partly linear models. Scand. J. Statist. 33, 435–450.

Bianco, A., García Ben, M., Yohai, V., 2005. Robust estimation for linear regression with asymmetric errors. Canad. J. Statist. 33, 511–528.

Bianco, A., Martínez, E., 2009. Robust testing in the logistic regression model. Comput. Statist. Data Anal. 53, 4095–4105.

Bianco, A., Yohai, V., 1995. Robust estimation in the logistic regression model. In: Lecture Notes in Statistics, vol. 109. Springer-Verlag, New York, pp. 17–34.

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

Boente, G., Rodriguez, D., 2008. Robust inference in generalized partially linear models. http://www.ic.fcen.uba.ar/preprints/boente_rodriguez_2008.pdf.

Cantoni, E., Ronchetti, E., 2001. Robust inference for generalized linear models. J. Amer. Statist. Assoc. 96, 1022–1030.

Carroll, R., Fan, J., Gijbels, I., Wand, M., 1997. Generalized partially linear single-index models. J. Amer. Statist. Assoc. 92, 477–489.

Chang, X.W., Qu, L., 2004. Wavelet estimation of partially linear models. Comput. Statist. Data Anal. 47, 31–48.

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

Dvoretzky, A., Kiefer, J., Wolfowitz, J., 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27, 642–669.

Gao, J., Shi, P., 1997. *M*-type smoothing splines in nonparametric and semiparametric regression models. Statist. Sinica 7, 1155–1169.

González Manteiga, W., Aneiros Pérez, G., 2003. Testing in partial linear regression models with dependent errors. J. Nonparametr. Stat. 15, 93–111.

Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Amer. Statist. Assoc. 69, 383–394.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.

Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Physica-Verlag.

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2006. Nonparametric and Semiparametric Models. Springer-Verlag.

He, X., Fung, W., Zhu, Z., 2005. Robust estimation in generalized partial linear models for clustered data. J. Amer. Statist. Assoc. 100, 1176–1184.

He, X., Zhu, Z., Fung, W., 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika 89, 579–590.

Künsch, H., Stefanski, L., Carroll, R., 1989. Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. J. Amer. Statist. Assoc. 84, 460–466.

Liang, H., 2006. Estimation in partially linear models and numerical comparisons. Comput. Statist. Data Anal. 50, 675–687.

Mack, Y., Silverman, B., 1982. Weak and strong uniform consistency of kernel regression estimates. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61, 405–415.

Mallows, C., 1974. On some topics in robustness. In: Memorandum, Bell Laboratories. Murray Hill, NJ.

Manchester, L., 1996. Empirical influence for robust smoothing. Austral. J. Statist. 38, 275–296.

McCullagh, P., Nelder, J., 1989. Generalized Linear Models, 2nd ed.. Champman and Hall, London.

Opsomer, J.D., Ruppert, D., 1999. A root-*n* consistent backfitting estimator for semiparametric additive modelling. J. Comput. Graph. Statist. 8, 715–732.

Prakasa Rao, B.L.S., 1983. Nonparametric Functional Estimation. Academic Press, New York.

Qin, G., Zhu, Z., 2007. Robust estimation in generalized semiparametric mixed models for longitudinal data. J. Multivariate Anal. 98, 1658–1683.

Robinson, P., 1988. Root-*n* consistent semiparametric regression. Econometrica 56, 931–954.

Severini, T., Staniswalis, J., 1994. Quasi-likelihood estimation in semiparametric models. J. Amer. Statist. Assoc. 89, 501–511.

Severini, T., Wong, W., 1992. Profile likelihood and conditionally parametric models. Ann. Statist. 20 (4), 1768–1802.

Stefanski, L., Carroll, R., Ruppert, D., 1986. Bounded score functions for generalized linear models. Biometrika 73, 413–424.

Tamine, J., 2002. Smoothed influence function: another view at robust nonparametric regression. Discussion Paper 62. Sonderforschungsbereich 373. Humboldt-Universität zu Berlin.

Tukey, J., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.

Van der Vaart, A., 1998. Asymptotic Statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.