

## Accepted Manuscript

Testing in generalized partially linear models: A robust approach

Graciela Boente, Ricardo Cao, Wenceslao González Manteiga, Daniela Rodríguez

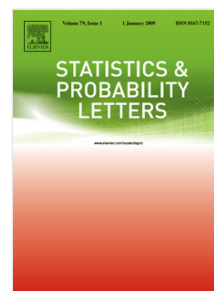
PII: S0167-7152(12)00336-7  
DOI: 10.1016/j.spl.2012.08.031  
Reference: STAPRO 6408

To appear in: *Statistics and Probability Letters*

Received date: 22 August 2011  
Revised date: 29 August 2012  
Accepted date: 30 August 2012

Please cite this article as: Boente, G., Cao, R., González Manteiga, W., Rodríguez, D., Testing in generalized partially linear models: A robust approach. *Statistics and Probability Letters* (2012), doi:10.1016/j.spl.2012.08.031

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



## Testing in generalized partially linear models: A robust approach

Graciela Boente<sup>a</sup>, Ricardo Cao<sup>b</sup>, Wenceslao González Manteiga<sup>c</sup>, Daniela Rodríguez<sup>\*,a</sup><sup>a</sup>*Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*<sup>b</sup>*Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain*<sup>c</sup>*Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Santiago de Compostela, Santiago de Compostela, Spain*

---

**Abstract**

In this paper, we introduce a family of robust statistics which allow to decide between a parametric model and a semiparametric one. More precisely, under a generalized partially linear model, i.e., when the observations satisfy  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  with  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^T \boldsymbol{\beta})$  and  $H$  a known link function, we want to test  $H_0 : \eta(t) = \alpha + \gamma t$  against  $H_1 : \eta$  is a nonlinear smooth function. A general approach which includes robust estimators based on a robustified deviance or a robustified quasi-likelihood is considered. The asymptotic behavior of the test statistic under the null hypothesis is obtained.

*Key words:* Generalized partially linear models, Kernel weights, Rate of convergence, Robust testing

---

**Abbreviated Title:** Robust test in generalized semiparametric models.

---

\*Corresponding Author

Instituto de Cálculo, FCEyN, UBA  
Ciudad Universitaria, Pabellón 2  
Buenos Aires, C1428EHA, Argentina  
FAX 54-11-45763375

*Email addresses:* gboente@dm.uba.ar (Graciela Boente), rcao@udc.es (Ricardo Cao),  
wenceslao.gonzalez@usc.es (Wenceslao González Manteiga), drodriguez@dm.uba.ar (Daniela Rodríguez)

*Preprint submitted to Elsevier*

*August 29, 2012*

## 1. Introduction

Semiparametric models contain both a parametric and a nonparametric component. Sometimes the nonparametric component plays the role of a nuisance parameter. A lot of research has been done on estimators of the parametric component in a general framework, aiming to obtain asymptotically efficient estimators. The aim of this paper is to consider semiparametric versions of the generalized linear models where the response  $y$  is to be predicted by covariates  $(\mathbf{x}, t)$ , where  $\mathbf{x} \in \mathbb{R}^p$  and  $t \in \mathcal{T} \subset \mathbb{R}$  with  $\mathcal{T}$  a compact set. Without loss of generality we will assume that  $\mathcal{T} = [0, 1]$ . It will also be assumed 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  $C$ . Then,  $\mu(\mathbf{x}, t) = \mathbb{E}(y|\mathbf{x}, t) = B'(\theta(\mathbf{x}, t))$ , with  $B'$  as the derivative of  $B$ . In generalized linear models (McCullagh and Nelder, 1989), which is a popular technique for modelling a wide variety of data, it is often assumed that the mean is modelled linearly through a known link function,  $g$ , i.e.,  $g(\mu(\mathbf{x}, t)) = \gamma + \mathbf{x}^T \boldsymbol{\beta} + \alpha t$ . For instance, an ordinary logistic regression model assumes that the observations  $(y_i, \mathbf{x}_i, t_i)$  are such that the responses are independent binomial variables  $y_i|\mathbf{x}_i, t_i \sim Bi(1, p_i)$  whose success probabilities depend on the explanatory variables through the relation  $g(p_i) = \gamma + \mathbf{x}_i^T \boldsymbol{\beta} + \alpha t_i$ , with  $g(u) = \log(u/(1-u))$ .

In many situations, the linear model is insufficient to explain the relationship between the response variable and its associated covariates. A natural generalization, which suffers from the *curse of dimensionality*, is to model the mean nonparametrically in the covariates. An alternative strategy is to allow most predictors to be modelled linearly while one or a small number of predictors enter in the model nonparametrically. This is the approach we will follow, so that the relationship will be given by the semiparametric generalized partially linear model

$$\mu(\mathbf{x}, t) = H\left(\eta(t) + \mathbf{x}^T \boldsymbol{\beta}\right) \quad (1)$$

where  $H = g^{-1}$  is a known link function,  $\boldsymbol{\beta} \in \mathbb{R}^p$  is an unknown parameter and  $\eta$  is an unknown smooth function.

In the context of hypothesis testing for regression models, that is, when  $H(t) = t$ , Gao (1997) established a large sample theory for testing  $H_0 : \boldsymbol{\beta} = 0$  and, in addition to this, Härdle *et al.* (2000) tested  $H_{0,\eta} : \eta = \eta_0$  too, while Härdle and Mammen (1993) considered the lack of fit problem  $H_0 : \eta \in \{\eta_\theta : \theta \in \Theta\}$ . Besides, González-Manteiga and Aneiros-Pérez (2003) studied the case of dependent errors and Koul and Ni (2004) considered the case of random design and heteroscedastic errors. These methods are based on a  $L^2$  distance comparison between a nonparametric estimator of the regression function and a smoothed parametric estimator, so they face the problem of selecting the smoothing parameter. An alternative approach is based on the empirical estimator of the integrated regression function. Goodness of fit tests based on empirical process for regression models with non-random design have been studied, for instance, by Koul and Stute (1998) and Diebolt (1995). On the other hand, under a purely nonparametric regression model with Berkson measurement errors, Koul and Song (2008) considered a marked empirical process of the calibrated residuals. Recently, Koul and Song (2010) proposed a test for the partial linear regression model based on the supremum of a martingale transform of a process of calibrated residuals, when both the covariates in the parametric and nonparametric components are subject to Berkson measurement errors.

On the other hand, for generalized partially linear models, hypothesis testing mainly focusses on comparing kernel based estimators with smoothed parametric estimators. For instance, Härdle *et al.* (1998) considered a test statistic to decide between a linear and a semiparametric model. Their proposal is based on the estimation procedure considered by Severini and Staniswalis (1994)

modified to deal with the smoothed and unsmoothed likelihoods. A comparative study of different procedures was performed by Müller (2001) while a different approach was considered in Rodríguez Campos *et al.* (1998).

As it is well known, such estimates fail to deal with outlying observations and so does the test statistic. 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  $\eta$ , leading to possibly wrong conclusions. In particular, as mentioned Hampel's comment on Stone (1977) paper "*If we believe in a smooth model without spikes, . . . , some robustification is possible. In this situation, a clear outlier will not be attributed to some sudden change in the true model, but to a gross error, and hence it may be deleted or otherwise made harmless*". Therefore, in this context robust procedures need to be developed to avoid wrong conclusions on the hypothesis to be tested (see Bianco *et al.* (2006) for a discussion).

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). The basic ideas from robust smoothing and from robust regression estimation have been adapted to deal with the case of independent observations following a partially linear regression model with  $H(t) = t$ ; we refer to Gao and Shi (1997), He *et al.* (2002) and Bianco and Boente (2004). Moreover, robust tests for a given alternative, under a partially linear regression model were studied in Bianco *et al.* (2006). Besides, a robust approach for testing the parametric form of a regression function versus an omnibus alternative, based on the centered asymptotic rank transformation, was considered by Wang and Qu (2007) when  $H(t) = t$  and  $\beta = 0$ , i.e., under the nonparametric model  $y_i = \eta(t_i) + \epsilon_i$ .

Under a generalized partially linear model (1), Boente *et al.* (2006) introduced a general profile-based two-step robust procedure to estimate the parameter  $\beta$  and the function  $\eta$  while Boente and Rodríguez (2010) (see also, Rodríguez, 2008) developed a three-step method to improve the computational time of the previous one. Beyond the importance of developing robust estimators in more general settings, the work on testing also deserves attention. An up-to-date review of robust hypothesis testing results can be found in He (2002). The aim of this paper is to propose a class of tests for the nonparametric component based on the three-step robust procedure proposed by Boente and Rodríguez (2010).

The paper is organized as follows. In Section 2, we remind the definition of the general profile-based two-step estimators as well as the three-step robust estimates and their asymptotic properties. In Section 3, we present a robust alternative to test hypothesis concerning the nonparametric component  $\eta$ . Their asymptotic behavior is studied in Section 4 while a bootstrap procedure is discussed in Section 5. Section 6 reports the result of a Monte Carlo study conducted to evaluate the performance of the tests under the null hypothesis and under a set of alternatives. Finally, proofs are relegated to the Appendix.

## 2. Preliminaries: The estimation procedure

As mentioned in the Introduction, Boente *et al.* (2006) introduced a highly robust procedure under model (1) while Boente and Rodríguez (2010) introduced a local approach to improve the computational time. Let  $(y_i, \mathbf{x}_i, t_i)$  be independent observations such that  $y_i | (\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$  with  $\mu_i = H(\eta(t_i) + \mathbf{x}_i^T \beta)$  and  $\text{VAR}(y_i | (\mathbf{x}_i, t_i)) = V(\mu_i)$ . Let  $\eta_0(t)$  and  $\beta_0$  denote the true parameter values, and  $\mathbb{E}_0$  the expected value under the true model, so that  $\mathbb{E}_0(y_1 | (\mathbf{x}_1, t_1)) = H(\eta_0(t_1) + \mathbf{x}_1^T \beta_0)$ .

As in Robinson (1988), we will assume that the vector  $\mathbf{1}_n$  is not in the space spanned by the column vectors of  $(\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ , that is, we do not allow  $\beta_0$  to include an intercept so that the model is identifiable, i.e., if  $\mathbf{x}_i^\top \beta_1 + \eta_1(t_i) = \mathbf{x}_i^\top \beta_2 + \eta_2(t_i)$  for  $1 \leq i \leq n$ , then,  $\beta_1 = \beta_2$  and  $\eta_1 = \eta_2$ . Due to the generality of the semiparametric model (1), identifiability implies that only ‘‘slope’’ coefficients can be estimated.

Let  $w_1 : \mathbb{R}^p \rightarrow \mathbb{R}$  be a weight function to control leverage points on the carriers  $\mathbf{x}$ ,  $\rho : \mathbb{R}^2 \rightarrow \mathbb{R}$  a loss function and  $K : \mathbb{R} \rightarrow \mathbb{R}$  a kernel function. Define  $S(\beta, a, \tau) = \mathbb{E}_0 [\rho(y, \mathbf{x}^\top \beta + a) w_1(\mathbf{x}) | t = \tau]$  and  $S_n(\beta, a, t) = \sum_{i=1}^n W_i(t) \rho(y_i, \mathbf{x}_i^\top \beta + a) w_1(\mathbf{x}_i)$  where  $W_i(t)$  are the kernel weights on  $t_i$ , i.e.

$$W_i(t) = \left( \sum_{j=1}^n K((t - t_j)/h_n) \right)^{-1} K((t - t_i)/h_n).$$

Following the ideas of Severini and Staniswalis (1994), Boente *et al.* (2006) defined, for each fixed  $\beta$ , the function  $\eta_\beta(t)$  as the minimizer of  $S(\beta, a, t)$ . Since  $S_n(\beta, a, t)$  provides a consistent estimate of  $S(\beta, a, t)$ , the minimizer in  $a$ ,  $\hat{\eta}_\beta(t)$ , of  $S_n(\beta, a, t)$  estimates  $\eta_\beta(t)$ . These functions allow the above mentioned authors to define a two-step robust quasi-likelihood estimators of  $\beta_0$  and  $\eta_0$  as  $\hat{\beta} = \operatorname{argmin}_\beta S_n(\beta, \eta_\beta, t)$  and  $\hat{\eta}(t) = \hat{\eta}_{\hat{\beta}}(t)$ , respectively. Boente and Rodriguez (2010) introduced a new family of estimators of  $\beta_0$  and  $\eta_0$  that improve the computational results. Both proposals provide robust root- $n$  consistent estimators of the regression parameter  $\beta$ .

If the function  $\rho(y, u)$  is continuously differentiable and we denote  $\Psi(y, u) = \partial \rho(y, u) / \partial u$ , the functional  $\eta_\beta(t)$  and the estimates  $\hat{\eta}_\beta(t)$  will be a solution of the differentiated equations, i.e., they will be a solution of  $S^{(1)}(\beta, a, t) = 0$  and  $S_n^{(1)}(\beta, a, t) = 0$  respectively, where  $S^{(1)}(\beta, a, \tau) = \mathbb{E}_0 (\Psi(y, \mathbf{x}^\top \beta + a) w_1(\mathbf{x}) | t = \tau)$  and  $S_n^{(1)}(\beta, a, t) = \sum_{i=1}^n W_i(t) \Psi(y_i, \mathbf{x}_i^\top \beta + a) w_1(\mathbf{x}_i)$ . We refer to Boente *et al.* (2006) and Boente and Rodriguez (2010) for a discussion on the choice of the loss functions, where also conditions to ensure Fisher-consistency of the resulting estimators are stated. We only point out that, under a generalized linear model, two families of loss functions  $\rho$  have been considered in the literature, the first one bounds the deviances, as in our simulation study, while the second one introduced by Cantoni and Ronchetti (2001) is based on robustifying the quasi-likelihood by bounding the Pearson residuals.

### 3. Test statistics

A robust test statistic to test  $H_0 : \eta_0 \in \{\alpha + \gamma t, \alpha \in \mathbb{R}, \gamma \in \mathbb{R}\}$  can be defined by comparing the robust semiparametric estimators with the robust estimators obtained under a parametric model. We will give an approach which robustifies the test statistic defined in Härdle *et al.* (1998).

Denote  $\hat{\beta}$  a robust root- $n$  estimator of  $\beta_0$  and  $\hat{\eta}(t) = \hat{\eta}_{\hat{\beta}}(t)$  the estimates of  $\eta_0(t)$  solution of  $\hat{\eta}_{\hat{\beta}}(t) = \operatorname{argmin}_{a \in \mathbb{R}} S_n(\hat{\beta}, a, t)$ . As in Section 2, let  $w_2 : \mathbb{R}^p \rightarrow \mathbb{R}$  be a weight function that controls high leverage points on the covariates  $\mathbf{x}$ . Denote  $L(\beta, \alpha, \gamma) = \mathbb{E}_0 [\rho(y, \mathbf{x}^\top \beta + \alpha + \gamma t) w_2(\mathbf{x})]$  and

$$L_n(\beta, \alpha, \gamma) = \frac{1}{n} \sum_{i=1}^n \rho(y_i, \mathbf{x}_i^\top \beta + \alpha + \gamma t_i) w_2(\mathbf{x}_i),$$

which correspond to the robustified objective functions under a generalized linear regression model. Then, the robust estimates of the regression parameter under the generalized linear model can be

defined as the minimizer of  $L_n$

$$(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\alpha}_{H_0}, \hat{\gamma}_{H_0}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^p, \alpha \in \mathbb{R}, \gamma \in \mathbb{R}}{\operatorname{argmin}} L_n(\boldsymbol{\beta}, \alpha, \gamma). \quad (2)$$

To test  $H_0$ , a natural approach is to compare the predicted values  $\mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i)$  with those obtained under the null hypothesis,  $\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t_i$ . However, as it is well known, in nonparametric and semiparametric models, due to the bias of the kernel estimator of  $S(\boldsymbol{\beta}, a, t)$ , the smoothing bias of  $\hat{\eta}(t)$  is non-negligible, even under the linear hypothesis  $H_0$ , see, for instance, Härdle and Mammen (1993) and Härdle, *et al.* (1998) for a discussion, when considering the classical estimators. For that reason, a simple comparison between both estimators may be misleading and conduct wrong conclusions. To solve this problem, Härdle, *et al.* (1998) introduced a smoothing bias to  $\hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t$  to compensate that of  $\hat{\eta}(t)$ . It is worth noting that the smoothed estimators obtained under the null hypothesis may not belong to family of linear functions. However, they provide consistent estimators under the parametric model.

To define *smoothed* estimators under the null hypothesis, consider the pseudo-observations  $\tilde{y}_i$  corresponding to the parametric fit of the conditional expectation under the null hypothesis, that is,  $\tilde{y}_i = H(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t_i)$  and denote  $\tilde{\Psi}(\mu, \mathbf{x}^T \boldsymbol{\beta} + a) = \mathbb{E}(\Psi(y, \mathbf{x}^T \boldsymbol{\beta} + a) | (\mathbf{x}, t))$  where the conditional expectation is taken when  $y | (\mathbf{x}, t) \sim F(\cdot, \mu)$ .

The function  $\hat{\eta}_{H_0}$  is defined as follows. Since the pseudo-observations will not have outliers, in the sense of large Pearson residuals, but only leverage points could appear, it is quite natural to define  $\hat{\eta}_{H_0}(t)$  as the value solving  $\sum_{i=1}^n W_i(t) \tilde{\Psi}(\tilde{y}_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + a) w_1(\mathbf{x}_i) = 0$ , or equivalently as the value  $\hat{\eta}_{H_0}(t) = \operatorname{argmin}_{a \in \mathbb{R}} \tilde{S}_n(\hat{\boldsymbol{\beta}}_{H_0}, a, t)$ , where  $\tilde{S}_n(\boldsymbol{\beta}, a, t) = \sum_{i=1}^n W_i(t) \tilde{\rho}(\tilde{y}_i, \mathbf{x}_i^T \boldsymbol{\beta} + a) w_1(\mathbf{x}_i)$ , with  $(\partial \tilde{\rho}(\mu, a) / \partial a) = \tilde{\Psi}(\mu, a)$ . Note that under mild conditions  $\tilde{\rho}(\mu, a) = \mathbb{E}(\rho(y, a) | (\mathbf{x}, t))$  where the conditional expectation is taken when  $y | (\mathbf{x}, t) \sim F(\cdot, \mu)$ .

Then, the test statistic is defined using a goodness-of-fit measure, based on the quasi-likelihood function

$$\mathcal{T}_1 = -2 \sum_{i=1}^n Q \left( H \left( \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \hat{\eta}(t_i) \right), H \left( \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t_i) \right) \right) w_2(\mathbf{x}_i) w(t_i)$$

where  $Q(y, \mu) = \int_{\mu}^y (s - y) V^{-1}(s) ds$  is the quasi-likelihood. Since the quasi-likelihood is computed comparing predicted values for the responses based on robust estimators, large deviations of the predicted responses from its mean will not have large influence in the test statistics. However, outlying points in the explanatory variables may have large influence on the quasi-likelihood expression. Hence, in order to bound their effect, we introduce a weight function  $w_2(\mathbf{x}_i)$  in the test definition. We have also included a weight function  $w(t)$  to avoid boundary effects. The function  $w$  has a compact support  $\mathcal{T}_0 \subset \mathcal{T} = [0, 1]$ , in particular we have that for  $n$  large enough  $I_{[h_n, 1-h_n]}(t) \geq w(t)$ . This robust version of quasi-likelihood test is different from the robust likelihood ratio-type or score type tests as defined in Heritier and Ronchetti (1994) which still uses the responses  $y_i$  and compares the responses and the fits obtained under the restricted and unrestricted models.

#### 4. Asymptotic behavior

For the sake of simplicity, we denote  $\rho_n = h_n^2 + (n h_n)^{-\frac{1}{2}}$ ,  $\chi(y, a) = \partial \Psi(y, a) / \partial a$ ,  $\chi_1(y, a) = \partial^2 \Psi(y, a) / \partial a^2$ ,  $\hat{v}(\boldsymbol{\beta}, t) = \hat{\eta}_{\boldsymbol{\beta}}(t) - \eta_{\boldsymbol{\beta}}(t)$ ,  $\hat{v}_0(t) = \hat{v}(\boldsymbol{\beta}_0, t)$ ,  $\hat{v}_j(\boldsymbol{\beta}, t) = \partial \hat{v}(\boldsymbol{\beta}, t) / \partial \beta_j$  and  $\hat{v}_{j,0}(t) = \hat{v}_j(\boldsymbol{\beta}_0, t)$ .

We will need the following set of assumptions

- A1.** The density  $f$  of  $t_1$  is bounded on  $\mathcal{T}$ , twice continuously differentiable in the interior of  $\mathcal{T}$  with bounded derivatives
- A2.**  $\inf_{t \in [0,1]} f(t) > 0$
- A3.**  $\eta_0$  is twice continuously differentiable in the interior of  $\mathcal{T}$  with bounded derivatives on  $\mathcal{T}$ .
- A4.**  $r(t, \tau) = \mathbb{E}_0(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t)) w_1(\mathbf{x}_1) | t_1 = \tau)$  is uniformly continuous in the interior of  $\mathcal{T}$  and bounded in  $\mathcal{T}$ .
- A5.** The functions  $v_0(\tau) = \mathbb{E}_0(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(\tau)) w_1(\mathbf{x}_1) | t_1 = \tau)$  and  $v_1(\tau) = \mathbb{E}_0(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(\tau)) \mathbf{x}_1 w_1(\mathbf{x}_1) | t_1 = \tau)$  are uniformly continuous in the interior of  $\mathcal{T}$  and  $\mathcal{I}_{v_0} = \inf_{t \in [0,1]} |v_0(\tau)| > 0$ .
- A6.**  $\Psi, \chi, \chi_1, w, w_j$  and  $\psi_j(\mathbf{x}) = \mathbf{x} w_j(\mathbf{x})$  are bounded functions for  $j = 1, 2$ .
- A7.**  $K$  is a function of bounded variation with compact support  $[0, 1]$  and it satisfies  $\int K(u) du = 1$  and  $\int u K(u) du = 0$ .
- A8.** The bandwidth sequence satisfies  $nh_n^3 / \log(n) \rightarrow \infty$  and  $n^{\frac{1}{2}} h_n^4 \log(n) \rightarrow 0$ .

**Theorem 4.1.** Assume that **A1** to **A8** hold. Moreover, assume that

- a)  $\mathcal{G} = \{g(y, \mathbf{x}, u) = \chi(y, \mathbf{x}^\top \boldsymbol{\beta}_0 + a) w_1(\mathbf{x}) - \mathbb{E}_0(\chi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + a) w_1(\mathbf{x}_1) | t_1 = u), a \in \mathbb{R}\}$ , has covering number  $N(\epsilon, \mathcal{G}, L^1(\mathbb{Q})) \leq A\epsilon^{-W}$ , for any probability  $\mathbb{Q}$  and  $0 < \epsilon < 1$ .
- b)  $\psi_{1,2}(\mathbf{x}) = w_1(\mathbf{x}) \|\mathbf{x}\|^2$  is bounded or  $\sup_{t \in \mathcal{T}} \mathbb{E}_0(\psi_{1,2}(\mathbf{x}) | t) < \infty$ .

Then, under  $H_0 : \eta \in \{\alpha + \gamma t, \alpha \in \mathbb{R}, \gamma \in \mathbb{R}\}$ , we have that  $v_n^{-1}(\mathcal{T}_1 - m_n) \xrightarrow{w} N(0, 1)$ , with  $m_n = c_{1,\Psi} h_n^{-1} \int K^2(u) du$  and  $v_n^2 = 2c_{2,\Psi} h_n^{-1} \int (K * K(u))^2 du$ , where

$$\begin{aligned}
 c_{1,\Psi} &= \mathbb{E} \left( w(t_1) \mathbb{E} \left[ w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1)^2}{V(H(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1))} \middle| t_1 \right] \mathbb{E} \left[ w_1^2(\mathbf{x}_1) \sigma^2(\mathbf{x}_1, t_1) \middle| t_1 \right] v_0(t_1)^{-2} f(t_1)^{-1} \right) \\
 c_{2,\Psi} &= \mathbb{E} \left( \left[ \mathbb{E} \left\{ w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1)^2}{V(H(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1))} \middle| t_1 \right\} \right]^2 \left[ \mathbb{E} \left\{ w_1^2(\mathbf{x}_1) \sigma^2(\mathbf{x}_1, t_1) \middle| t_1 \right\} \right]^2 \frac{w^2(t_1)}{v_0(t_1)^4 f(t_1)} \right) \\
 \sigma^2(\mathbf{x}_0, t_0) &= \mathbb{E} \left\{ \left[ \Psi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t_1)) - \mathbb{E}_0(\Psi(y_1, \mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t_1)) | (\mathbf{x}_1, t_1)) \right]^2 | (\mathbf{x}_1, t_1) = (\mathbf{x}_0, t_0) \right\}.
 \end{aligned}$$

**Remark 4.1.** When considering the canonical exponential family described in the Introduction  $V(H(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t_1))) = H'(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \eta_0(t_1))$  and so

$$\begin{aligned}
 c_{1,\Psi} &= \mathbb{E} \left( w(t_1) \mathbb{E} \left[ w_2(\mathbf{x}_1) H'(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1) \middle| t_1 \right] \mathbb{E} \left[ w_1^2(\mathbf{x}_1) \sigma^2(\mathbf{x}_1, t_1) \middle| t_1 \right] v_0(t_1)^{-2} f(t_1)^{-1} \right) \\
 c_{2,\Psi} &= \mathbb{E} \left( w^2(t_1) \left[ \mathbb{E} \left\{ w_2(\mathbf{x}_1) H'(\mathbf{x}_1^\top \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1) \middle| t_1 \right\} \right]^2 \left[ \mathbb{E} \left\{ w_1^2(\mathbf{x}_1) \sigma^2(\mathbf{x}_1, t_1) \middle| t_1 \right\} \right]^2 \frac{1}{v_0(t_1)^4 f(t_1)} \right).
 \end{aligned}$$

## 5. A Montecarlo test

In this section, we develop a bootstrap procedure to implement the goodness-of-fit test for linearity. The need of bootstrapping has been studied by several authors such as Härdle and Mammen



(1993), Härdle *et al.* (1998). These authors applied a wild bootstrap procedure to construct the bootstrap samples. However, in the present setting due to the expensive computing time needed to compute the robust estimators, a linearised Montecarlo as defined in Zhu (2005) provides a better approach. This approach was also considered in Zhu and Zhang (2004) who propose a resampling procedure for approximating the  $p$ -value when considering a log-likelihood ratio test statistics for testing homogeneity. Rémillard and Scaillet (2009) and Kojadinovic and Yan (2011) applied this method to provide fast goodness-of-fit tests for copulas.

As it will be shown in the Appendix,  $\mathcal{T}_1 = R_n + O_p((n/h)^{\frac{1}{2}} \rho_n \log n)$ , under  $H_0 : \eta_0 \in \{\alpha + \gamma t, \alpha \in \mathbb{R}, \gamma \in \mathbb{R}\}$ , with

$$\begin{aligned} R_n &= \sum_{i=1}^n w(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i)^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i))} v_0(t_i)^{-2} f(t_i)^{-2} \times \\ &\quad \left\{ \sum_{j=1}^n W_{0,j}(t_i) w_1(\mathbf{x}_j) \left[ \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) - \mathbb{E}_0 \left( \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) \mid (\mathbf{x}_j, t_j, \mathbf{x}_i, t_i) \right) \right] \right\}^2 \\ &= n \int w(t) w_2(\mathbf{x}) \frac{H'(\mathbf{x}^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t)^2}{V(H(\mathbf{x}^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t))} v_0(t)^{-2} f(t)^{-2} \mathcal{W}_n^2(t) dF_n(\mathbf{x}, t), \end{aligned}$$

where  $\mathcal{W}_n(t) = \sum_{j=1}^n W_{0,j}(t) w_1(\mathbf{x}_j) \left[ \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t)) - \mathbb{E}_0 \left( \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) \mid (\mathbf{x}_j, t_j, t_i = t) \right) \right]$ . This suggests the following Montecarlo procedure

**Step B1** Given a sample  $\{(y_i, \mathbf{x}_i, t_i)\}_{1 \leq i \leq n}$  compute the estimators  $(\hat{\boldsymbol{\beta}}_{H_0}, \hat{\alpha}_{H_0}, \hat{\gamma}_{H_0})$  as in (2).

Define

- $\hat{v}_0(t) = \sum_{i=1}^n W_i(t) \chi(y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t)) w_1(\mathbf{x}_i)$  with  $\hat{\eta}_{H_0}(t) = \hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t$ .
- $\hat{\epsilon}_j(t) = \Psi(y_j, \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t)) - \mathbb{E}_0 \left( \Psi(y_j, \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t)) \mid (\mathbf{x}_j, t_j) \right)$

**Step B2** Generate  $n$  random variables  $\epsilon_1^* \dots \epsilon_n^*$ , independent of the sample  $\{(y_i, \mathbf{x}_i, t_i)\}_{1 \leq i \leq n}$  and such that  $\mathbb{E}(\epsilon_i^*) = 0$ ,  $\text{VAR}(\epsilon_i^*) = 1$  and  $\epsilon_i^*$  are bounded. For instance, we generate  $n$  observations from the two point distribution  $P^*(\epsilon^* = a) = p$  and  $P^*(\epsilon^* = b) = 1 - p$ , with  $a = (1 - \sqrt{5})/2$ ,  $b = (1 + \sqrt{5})/2$  and  $p = (5 + \sqrt{5})/10$ .

**Step B3** Define  $R_n^* = R_n^*(\hat{\boldsymbol{\beta}}_{H_0}, \hat{v}_0, \hat{\eta}_{H_0})$  with

$$R_n^* = \sum_{i=1}^n w(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t_i))^2}{V(H(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{H_0} + \hat{\eta}_{H_0}(t_i)))} \hat{v}_0(t_i)^{-2} \left\{ \sum_{j=1}^n W_j(t_i) w_1(\mathbf{x}_j) \hat{\epsilon}_j(t_i) \epsilon_j^* \right\}^2.$$

**Step B4** Repeat **Step B2** and **Step B3**  $N_{boot}$  times, to get  $N_{boot}$  values of  $R_n^*$ , say  $R_{n,i}^*$ ,  $1 \leq i \leq N_{boot}$ .

The  $(1 - \alpha)$ -quantiles of the distribution of  $R$  (and so of  $\mathcal{T}_1$ ) can be approximated by the  $(1 - \alpha)$ -quantiles of the conditional distribution of  $R^*$ . The  $p$ -value can be estimated by  $\hat{p} = k / (N_{boot} + 1)$  where  $k$  is the number of  $R_{n,i}^*$  which are larger or equal than  $\mathcal{T}_1$ .



## 6. Monte Carlo study

This section contains the results of a simulation study conducted with the aim of comparing the performance of the proposed testing procedure with the classical one. We consider a logistic partially linear model. The classical procedure corresponds to use the maximum likelihood estimators under the parametric model and the estimators defined in Carroll *et al.* (1997), which are an alternative to those, based on profile likelihood, considered in Severini and Staniswalis (1994), under the nonparametric one. To be more precise, we select the deviance as loss function and  $w_1 = w_2 \equiv 1$  in Sections 2 and 3. The robust estimators correspond to those controlling large values of the deviance and they are 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 are taken as the Tukey's biweight function with tuning constant  $c = 4.685$ . To be more precise, since  $x_i \in \mathbb{R}$ , we define  $w_1^2(x_i) = w_2^2(x_i) = \left(1 - [(x_i - M_n)/4.685]^2\right)^2$  when  $|x_i - M_n| \leq 4.685$  and 0 otherwise, with  $M_n$  the median of  $x_i$ . The central model denoted  $C_0$  in the figures corresponds to a logistic model where  $x_i \sim \mathcal{U}(-1, 1)$  and  $t_i \sim \mathcal{U}(0, 1)$ , independent each other. On the other hand, the responses are such that  $y_i|(x_i, t_i) \sim \text{Bi}(1, p(x_i, t_i))$  with  $\log(p(x, t)/(1 - p(x, t))) = \beta_0 x + \eta_0(t, \Delta)$ , with  $\beta_0 = 2$ ,  $\eta_0(t, \Delta) = (t - 0.5) + \Delta \cos(6\pi(t - 0.5))$ , that is,  $H(u) = \exp(u)/(1 + \exp(u))$ . The value  $\Delta = 0$  corresponds to the null hypothesis,  $H_0 : \eta_0 \in \{\alpha + \gamma t, \alpha \in \mathbb{R}, \gamma \in \mathbb{R}\}$ , while as alternatives we choose a grid of 10 equally spaced values of  $\Delta \in [0.2, 2.0]$ . We performed  $NR = 1000$  replications of samples of size  $n = 200$  and  $N_{boot} = 5000$  bootstrap samples. The Epanechnikov kernel  $K(t) = (3/4)(1 - t^2)I_{[-1,1]}(t)$  was selected for the smoothing procedure with bandwidth  $h = 0.1$ .

Clearly, the test statistics considered depend on the choice of  $h$ . For the classical procedure, Härdle *et al.* (1998) pointed the sensitivity of the test level to the selection of the bandwidth parameter. In particular, they suggested to apply the test for different choices of  $h$  to have an impression on how the function  $\eta_0$  differs significantly from linearity. The selected value  $h = 0.1$  was chosen so that, under  $H_0$ , the observed frequencies of rejection for the bootstrap robust and classical test reached values close to the nominal level  $\alpha = 0.1$ .

Figure 1 gives the frequency of rejection both for the classical and robust procedure for the uncontaminated samples. The nominal level was 0.10. The frequency of rejection of the asymptotic test is plotted in lines combined with filled diamonds while that of the Montecarlo test corresponds to the solid line. As expected the Montecarlo test improves the performance of the asymptotic ones, for the sample size considered.

For each generated sample, we also consider the following contaminations labelled  $C_1$  and  $C_2$ . We first generate a sample  $u_i \sim \mathcal{U}(0, 1)$  for  $1 \leq i \leq n$  and then, the contaminated sample, denoted  $(y_{i,c}, x_{i,c}, t_i)$ , is defined as follows for each contamination scheme

- Contamination  $C_1$  introduces *bad* high leverage points in the carriers  $x$ , without changing the responses already generated, i.e.,  $(y_{i,c}, x_{i,c}) = (y_i, x_i)$  if  $u_i \leq 0.90$  and  $(y_{i,c}, x_{i,c}) = (y_i, x_{i,new})$  if  $u_i > 0.90$ , where  $x_{i,new}$  is a new observation from a  $N(10, 1)$ .
- Contamination  $C_2$  includes outlying observations in the responses generated according to an incorrect model. Let  $\tilde{\eta}(t, \Delta) = \Delta \cos(6\pi(t - 0.5))$  and  $p_{i,new} = H(\tilde{\eta}(t_i, 20(1 - \Delta)))$ , define  $y_{i,new} \sim \text{Bi}(1, p_{i,new})$ . Then,  $(y_{i,c}, x_{i,c}) = (y_i, x_i)$  if  $u_i \leq 0.90$  and  $(y_{i,c}, x_{i,c}) = (y_{i,new}, x_i)$  if  $u_i > 0.90$ .

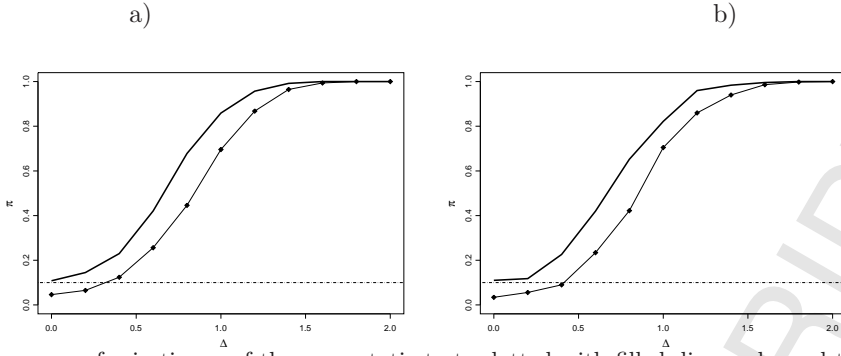


Figure 1: Frequency of rejection  $\pi$  of the asymptotic test, plotted with filled diamonds, and the Montecarlo test plotted with a solid line. a) Classical test b) Robust test.

Figures 2 and 3 give the frequency of rejection both for the classical and robust procedure for the contaminated samples. Figure 2 reports the frequencies of rejection for both the asymptotic and Montecarlo procedure, on the other hand, only the results for the Montecarlo test are reported for  $C_2$  since the asymptotic ones behave similarly. The results show that the classical test seem to be quite insensitive to high leverage points if the model is adequate as in  $C_1$ , while its power is sensitive to a misleading model. It is worth noting that under  $C_1$ , the Montecarlo classical test outperforms the robust one, since it leads to a lower loss of power. Quite surprisingly, under the present setting, the classical test adapts for the effect of high leverage points, if the model is correct, since the same disturbance is produced both in the parametric and nonparametric estimators. However, the classical procedure suffers from the inclusion of outliers under a misleading model, while the robust procedure is more stable. In this sense, the robust tests should be preferred. In particular, the robust asymptotic test avoids extra computing time at the expense of some level loss, leading to a conservative test.

## P. Appendix. Proofs

In this section we will give the proof of Theorem 4.1. From now on, let  $S_n^{(0,1)}(\beta, a, t) = \sum_{i=1}^n W_{0,i}(t) \Psi(y_i, \mathbf{x}_i^T \beta + a) w_1(\mathbf{x}_i)$  where  $W_{0,i}(t) = 1/(nh)K((t_i - t)/h_n)$ . Besides, define the family of functions  $\mathcal{G} = \{g(y, \mathbf{x}, u) = \chi(y, \mathbf{x}^T \beta_0 + a) w_1(\mathbf{x}) - \mathbb{E}_0(\chi(y_1, \mathbf{x}_1^T \beta_0 + a) w_1(\mathbf{x}_1) | t_1 = u), a \in \mathbb{R}\}$  and let  $N(\epsilon, \mathcal{G}, L^1(\mathbb{Q}))$  stand for its  $L^1$ -covering number. Denote also by  $\mathcal{K}_n = \{(t, \beta) : t \in [2h_n, 1 - 2h_n], \|\beta - \beta_0\| \leq \rho_n\}$ . We will need the following lemmas available in Boente *et al.* (2012).

**Lemma P.1.** Assume that **A1** to **A4**, **A6** and **A7** hold and that  $nh_n^3/\log(n) \rightarrow \infty$ . Then, we have that  $\sup_{(t, \beta) \in \mathcal{K}_n} |S_n^{(0,1)}(\beta, \eta_0(t), t)| = O_p(\rho_n \sqrt{\log n})$ .

**Lemma P.2.** Assume that **A1** to **A7** hold and that  $nh_n^3/\log(n) \rightarrow \infty$ . If  $N(\epsilon, \mathcal{G}, L^1(\mathbb{Q})) \leq A\epsilon^{-W}$ , for any probability  $\mathbb{Q}$  and  $0 < \epsilon < 1$ , we have that  $\sup_{(t, \beta) \in \mathcal{K}_n} |\hat{\eta}_\beta(t) - \eta_0(t)| = O_p(\rho_n \sqrt{\log n})$ .

**Lemma P.3.** Assume that **A1** to **A7** hold and that  $nh_n^3/\log(n) \rightarrow \infty$ . If, in addition,  $\psi_{1,2}(\mathbf{x}) = w_1(\mathbf{x})\|\mathbf{x}\|^2$  is bounded or  $\sup_{t \in \mathcal{T}} \mathbb{E}_0(\psi_{1,2}(\mathbf{x}) | t) < \infty$  and  $N(\epsilon, \mathcal{G}, L^1(\mathbb{Q})) \leq A\epsilon^{-W}$ , for any probability

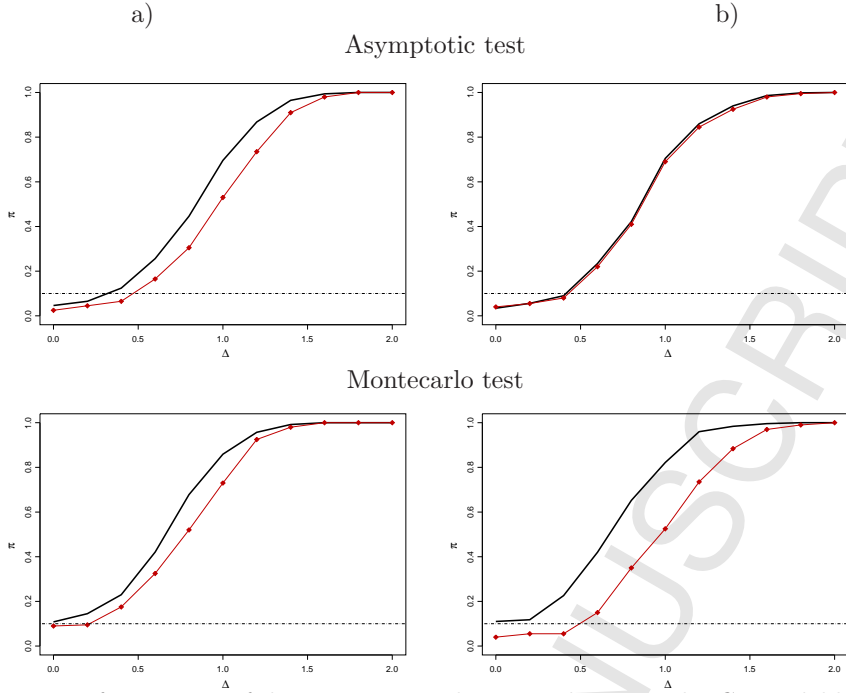


Figure 2: Frequency of rejection  $\pi$  of the asymptotic and Montecarlo test, under  $C_0$  in solid lines and under  $C_1$  in lines with diamonds. a) Classical test b) Robust test.

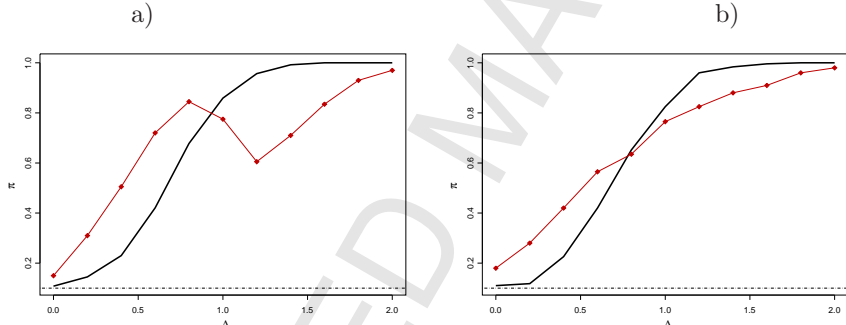


Figure 3: Frequency of rejection  $\pi$  of the Montecarlo test, under  $C_0$  in solid lines and under  $C_2$  in lines with diamonds. a) Classical test b) Robust test.

$\mathbb{Q}$  and  $0 < \epsilon < 1$ , we have that  $\sup_{(t, \beta) \in \mathcal{K}_n} |\hat{\eta}_\beta(t) - \hat{\eta}(t) - \hat{R}(\beta, t)| = O_p(\rho_n^2 \log n)$ , with

$$\hat{\eta}(t) = \eta_0(t) - \{v_0(t)f(t)\}^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \Psi(y_i, \mathbf{x}_i^T \beta_0 + \eta_0(t)) \quad (\text{P.1})$$

$$\hat{R}(\beta, t) = v_0(t)^{-1} v_1(t)^T (\beta - \beta_0). \quad (\text{P.2})$$

**Lemma P.4.** Assume that  $H_0$  holds, i.e.,  $\eta_0(t) = \alpha_0 + \gamma_0 t$ . Denote  $\tilde{y}_{i,0} = H(\mathbf{x}_i^T \beta_0 + \alpha_0 + \gamma_0 t_i)$

and  $\nu(\tau) = \mathbb{E} \left( w_1(\mathbf{x}_1) \tilde{\zeta}(\tilde{y}_{1,0}, \mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(\tau)) H' \left( H^{-1}(\tilde{y}_{1,0}) \right) | t_1 = \tau \right)$  where  $\tilde{\zeta}(y, a) = \partial \tilde{\Psi}(y, a) / \partial y$ . Under **A1** to **A7** if in addition  $nh_n^3 / \log(n) \rightarrow \infty$ , we have that

$$\sup_{t \in [2h_n, 1-2h_n]} \left| \hat{\eta}_{H_0}(t) - (\hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t - \alpha_0 - \gamma_0 t) v_0(t)^{-1} \nu(t) - \eta_0(t) + v_0(t)^{-1} f(t)^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \tilde{\Psi}(\tilde{y}_{i,0}, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) \right| = O_p(\rho_n^2 \log n).$$

PROOF OF THEOREM 4.1. In order to derive an expansion for the test statistic note that, uniformly on  $t \in [2h_n, 1-2h_n]$  we have

$$\begin{aligned} \hat{\eta}(t) - \hat{\eta}_{H_0}(t) &= \hat{\eta}(t) + \hat{R}(\hat{\boldsymbol{\beta}}, t) - (\hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t - \alpha_0 - \gamma_0 t) v_0(t)^{-1} \nu(t) - \eta_0(t) \\ &\quad + v_0(t)^{-1} f(t)^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \tilde{\Psi}(\tilde{y}_{i,0}, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) + O_p(\rho_n^2 \log n) \end{aligned}$$

with  $\hat{\eta}(t)$  and  $\hat{R}(\boldsymbol{\beta}, t)$  defined in (P.1) and (P.2), respectively. Hence,

$$\begin{aligned} \hat{\eta}(t) - \hat{\eta}_{H_0}(t) &= -v_0(t)^{-1} f(t)^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \left[ \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) - \tilde{\Psi}(\tilde{y}_{i,0}, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) \right] \\ &\quad + \hat{R}(\hat{\boldsymbol{\beta}}, t) - (\hat{\alpha}_{H_0} + \hat{\gamma}_{H_0} t - \alpha_0 - \gamma_0 t) v_0(t)^{-1} \nu(t) + O_p(\rho_n^2 \log n) \\ &= -v_0(t)^{-1} f(t)^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \left[ \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) - \mathbb{E}_0 \left( \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t)) | (\mathbf{x}_i, t_i) \right) \right] \\ &\quad + O_p(\sqrt{n}) + O_p(\rho_n^2 \log n). \end{aligned}$$

Therefore, we have the following expression for the test statistic  $\mathcal{T}_1 = R + O_p((n/h)^{\frac{1}{2}} \rho_n \log n)$  with

$$\begin{aligned} R &= \sum_{i=1}^n \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i)^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i))} \left[ \hat{\eta}(t_i) - \mathbb{E} \left( \hat{\eta}(t_i) | \mathbf{x}_1, t_1, \dots, \mathbf{x}_n, t_n \right) \right]^2 w(t_i) \\ &= \sum_{i=1}^n w(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i)^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i))} v_0(t_i)^{-2} f(t_i)^{-2} \mathcal{W}_n^2(t_i) \\ \mathcal{W}_n(t_i) &= \sum_{j=1}^n W_{0,j}(t_i) w_1(\mathbf{x}_j) \left[ \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) - \mathbb{E}_0 \left( \Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) | (\mathbf{x}_j, t_j, \mathbf{x}_i, t_i) \right) \right] \end{aligned}$$

which is a  $U$ -statistic. Therefore, using standard arguments as in Härdle and Mammen (1993) it follows that  $v_n^{-1} (\mathcal{T}_1 - m_n) \xrightarrow{w} N(0, 1)$ , with  $v_n^2 = 2c_{2,\Psi} h_n^{-1} \int (K * K(u))^2 du$  and  $m_n = c_{1,\Psi} h_n^{-1} \int K^2(u) du$  where  $c_1(\Psi)$ ,  $c_2(\Psi)$  and  $\sigma^2(\mathbf{x}_0, t_0)$  are given in Theorem 4.1.

Let us verify the expressions for  $m_n$  and  $v_n$ . Denote  $V_{j,i} = w_1(\mathbf{x}_j) [\Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) - \mathbb{E}_0(\Psi(y_j, \mathbf{x}_j^T \boldsymbol{\beta}_0 + \eta_0(t_i)) | (\mathbf{x}_j, t_j, \mathbf{x}_i, t_i))]$ , then

$$R = \frac{1}{n^2} \sum_{i=1}^n w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} \sum_{j=1}^n \sum_{\ell=1}^n K_{h_n}(t_j - t_i) K_{h_n}(t_\ell - t_i) V_{j,i} V_{\ell,i}$$

$$\begin{aligned}
 R &= \frac{K^2(0)}{n^2 h_n^2} \sum_{i=1}^n w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} V_{i,i}^2 \\
 &+ \frac{2K(0)}{n^2 h_n} \sum_{i=1}^n w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} V_{i,i} \sum_{\ell \neq i} K_{h_n}(t_\ell - t_i) V_{\ell,i} \\
 &+ \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i} \sum_{\ell \neq i} w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} K_{h_n}(t_j - t_i) K_{h_n}(t_\ell - t_i) V_{j,i} V_{\ell,i} \\
 &+ \frac{1}{n^2} \sum_{i=1}^n w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} \sum_{j \neq i} K_{h_n}^2(t_j - t_i) V_{j,i}^2 \\
 &= R_1 + R_2 + R_3 + R_4.
 \end{aligned}$$

Using that  $nh_n^2 \rightarrow \infty$  and that

$$\frac{1}{n} \sum_{i=1}^n w(t_i) w_2(\mathbf{x}_i) v_0^{-2}(t_i) f^{-2}(t_i) \frac{H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i)))} V_{i,i}^2 \xrightarrow{p} \mathbb{E}_0 \left( \frac{w(t_1) w_2(\mathbf{x}_1)}{v_0^2(t_1) f^2(t_1)} \frac{H'(\mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1))^2}{V(H(\mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1)))} V_{1,1}^2 \right)$$

we get that  $R_1 \xrightarrow{p} 0$  and so,  $h_n^{1/2} R_1 \xrightarrow{p} 0$ .

On the other hand, using that  $\mathbb{E}(V_{\ell,i} | (\mathbf{x}_\ell, t_\ell, \mathbf{x}_i, t_i)) = 0$  and  $\mathbb{E}(V_{i,i} | (\mathbf{x}_i, t_i)) = 0$  and that  $V_{\ell,i}$  and  $V_{i,i}$  are conditionally independent, for  $\ell \neq i$ , we get that  $\mathbb{E}(R_2) = 0$ . On the other hand, let  $Z_i = w(t_i) v_0^{-2}(t_i) f^{-2}(t_i) w_2(\mathbf{x}_i) V_{i,i} (H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2 / V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))))$ , then, we have that  $R_2 = (2K(0)) / (n^2 h) \sum_{i \neq \ell} Z_i K_{h_n}(t_\ell - t_i) V_{\ell,i}$ , and so,

$$\text{VAR}(R_2) = \frac{2K^2(0)n(n-1)}{n^4 h_n^2} (C_{1,h} + C_{2,h})$$

with  $C_{1,h} = \mathbb{E}(Z_1^2 K_{h_n}^2(t_2 - t_1) V_{2,1}^2)$  and  $C_{2,h} = \text{COV}(Z_1 K_{h_n}(t_2 - t_1) V_{2,1}, Z_2 K_{h_n}(t_1 - t_2) V_{1,2})$ . Note that,

$$C_{1,h} = \frac{1}{h_n^2} \mathbb{E} \left( Z_1^2 K^2 \left( \frac{t_2 - t_1}{h_n} \right) V_{2,1}^2 \right) = \frac{1}{h_n} \int R(t_1, t_1 + u h_n) K^2(u) f(t_1) f(t_1 + u h) du dt_1.$$

Hence,  $C_{1,h} = O(1)/h_n$ . In a similar way, we get that  $C_{2,h} = O(1)/h_n$ , which implies that  $h_n \text{VAR}(R_2) \rightarrow 0$  as  $n \rightarrow \infty$ , therefore,  $h_n^{1/2} R_2 \xrightarrow{p} 0$ .

Write  $R_4 = (1/n^2) \sum_{i=1}^n \sum_{j \neq i} W_i K_{h_n}^2(t_j - t_i) V_{j,i}$  with

$$W_i = w(t_i) w_2(\mathbf{x}_i) H'(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))^2 \left\{ v_0^2(t_i) f^2(t_i) V(H(\mathbf{x}_i^T \boldsymbol{\beta}_0 + \eta_0(t_i))) \right\}^{-1}$$

then,  $\mathbb{E}(R_4) = (1/n) \sum_{j \neq 1} \mathbb{E}(W_1 K_{h_n}(t_j - t_1) V_{j,1}^2) = ((n-1)/n) \mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1) \mathbb{E}(V_{2,1}^2 | (\mathbf{x}_1, t_1, \mathbf{x}_2, t_2)))$  and it is easy to see that  $\mathbb{E}(V_{2,1}^2 | (\mathbf{x}_1, t_1, \mathbf{x}_2, t_2)) = w_1^2(\mathbf{x}_2) \sigma^2(\mathbf{x}_2, t_2, t_1)$ . Let  $R_{4,1} = ((n-1)/n) \mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1) w_1^2(\mathbf{x}_2) \sigma^2(\mathbf{x}_2, t_2, t_1))$  and  $R_{4,2} = ((n-1)/n) \mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1) w_1^2(\mathbf{x}_2) [\sigma^2(\mathbf{x}_2, t_2, t_1) - \sigma^2(\mathbf{x}_2, t_2)])$ , then,  $\mathbb{E}(R_4) = R_{4,1} + R_{4,2}$ . Using that  $\sigma^2(\mathbf{x}_2, t_2, t_1)$  is Lipschitz, we obtain that

$$|\sigma^2(\mathbf{x}_2, t_2, t_1) - \sigma^2(\mathbf{x}_2, t_2)| < |t_1 - t_2| < h_n. \quad (\text{P.3})$$

Now, using that  $K$  has compact support in  $[-1, 1]$ , we get that

$$|R_{4,2}| \leq \frac{n-1}{n h_n} \mathbb{E} \left( |W_1| K^2 \left( \frac{t_2 - t_1}{h_n} \right) \right) = \frac{n-1}{n} O(1),$$

and so,  $h_n^{1/2}A_2 \rightarrow 0$ .

Let  $a(t_1) = w(t_1)v_0^{-2}(t_1)f^{-2}(t_1)$  and  $b(t_1) = \mathbb{E} \left( w_2(\mathbf{x}_i)H'(\mathbf{x}_i^\top\boldsymbol{\beta}_0 + \eta_0(t_i))^2 \{V(H(\mathbf{x}_i^\top\boldsymbol{\beta}_0 + \eta_0(t_i)))\}^{-1} \middle| t_1 \right)$ , then

$$h_n^{1/2}R_{4,1} = \frac{n-1}{nh_n^2}h_n^{1/2}\mathbb{E} \left( K^2 \left( \frac{t_2 - t_1}{h_n} \right) w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2)a(t_1)b(t_1) \right)$$

Denote by  $c(t_2) = \mathbb{E}(w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2)|t_2)$ . Thus,

$$h_n^{1/2}R_{4,1} = \frac{n-1}{nh_n}h_n^{1/2} \int K^2(u)du \int a(t_2)b(t_2)c(t_2)f^2(t_2)dt_2 + o(1).$$

Using analogous arguments to those considered previously when studying the convergence of  $R_2$ , one can easily obtain that  $h_n \text{VAR}(R_4) \rightarrow 0$ . Then,

$$h_n^{1/2} \left[ R_4 - \frac{1}{h_n} \int K^2(u)du \mathbb{E}(a(t_2)b(t_2)c(t_2)f(t_2)) \right] \xrightarrow{p} 0,$$

where

$$\mathbb{E}(a(t_2)b(t_2)c(t_2)f(t_2)) = \mathbb{E} \left( \mathbb{E}(w_1^2(\mathbf{x})\sigma^2(\mathbf{x}, t)|t)w(t)v_0^{-2}(t)f^{-1}(t)\mathbb{E} \left( w_2(\mathbf{x}) \frac{H'(\mathbf{x}^\top\boldsymbol{\beta}_0 + \eta_0(t))^2}{V(H(\mathbf{x}^\top\boldsymbol{\beta}_0 + \eta_0(t)))} \middle| t \right) \right).$$

Finally, we will study the asymptotic behavior of  $R_3$ . The expected value of  $R_3$  is equal 0, and so it is enough to study its variance. Straightforward calculations lead to write  $\text{VAR}(R_3) = A_1 + A_2 + A_3$  where

$$\begin{aligned} A_1 &= \frac{(n-1)(n-2)}{n^3} \mathbb{E} \left( a^2(t_1)w_2^2(\mathbf{x}_1) \frac{H'(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1))^4}{V^2(H(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1)))} K_{h_n}^2(t_2 - t_1)K_{h_n}^2(t_3 - t_1) \times \right. \\ &\quad \left. w_1^2(\mathbf{x}_3)w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2)\sigma^2(\mathbf{x}_3, t_3) \right) \\ A_2 &= \frac{(n-1)(n-2)}{n^3} \mathbb{E} \left( a^2(t_1)w_2^2(\mathbf{x}_1) \frac{H'(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1))^4}{V^2(H(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1)))} K_{h_n}^2(t_2 - t_1)K_{h_n}^2(t_3 - t_1) \right. \\ &\quad \left. w_1^2(\mathbf{x}_3)w_1^2(\mathbf{x}_2)[\sigma^2(\mathbf{x}_2, t_2, t_1)\sigma^2(\mathbf{x}_3, t_3, t_1) - \sigma^2(\mathbf{x}_2, t_2)\sigma^2(\mathbf{x}_3, t_3)] \right) \\ A_3 &= 2 \frac{(n-1)^2(n-2)}{n^3} \mathbb{E} \left( a(t_1)a(t_2)w_2(\mathbf{x}_1)w_2(\mathbf{x}_2) \frac{H'(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1))^2}{V(H(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1)))} \frac{H'(\mathbf{x}_2^\top\boldsymbol{\beta}_0 + \eta_0(t_2))^2}{V(H(\mathbf{x}_2^\top\boldsymbol{\beta}_0 + \eta_0(t_2)))} \times \right. \\ &\quad \left. K_{h_n}(t_3 - t_1)K_{h_n}(t_4 - t_1)K_{h_n}(t_3 - t_2)K_{h_n}(t_4 - t_2)\mathbb{E}(V_{3,1}V_{3,2}|\mathbf{x}_3, t_3, t_2, t_1)\mathbb{E}(V_{4,1}V_{4,2}|\mathbf{x}_4, t_4, t_2, t_1) \right). \end{aligned}$$

Let  $b_H(t_1) = \mathbb{E} \left( w_2^2(\mathbf{x}_1) \frac{H'^4(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1))}{V^2(H(\mathbf{x}_1^\top\boldsymbol{\beta}_0 + \eta_0(t_1)))} \middle| t_1 \right)$  and  $\sigma^2(t_1) = \mathbb{E}(w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1, t_1)|t_1)$  thus

$$h_n A_1 = \frac{h_n}{nh_n^2} \int a^2(t_1)b_H(t_1)\sigma^2(t_1 + uh_n)\sigma^2(vh_n + t_1)K^2(u)K^2(v)du dv dt_1 = \frac{1}{nh_n}O(1)$$

then, we have that  $h_n A_1 \rightarrow 0$ . On the other hand, using (P.3), we get the following bound

$$|A_2| \leq \frac{2}{nh_n} \int a^2(t_1)b(t_1)K^2(u)K^2(v)f(t_1)f(uh_n + t_1)f(vh_n + t_1)dt_1 du dv = \frac{1}{nh_n}O(1)$$

which implies that  $h_n A_2 \rightarrow 0$ . Finally, straightforward calculations lead to  $h_n A_3$  converges to  $2 \mathbb{E}(a^2(t)b^2(t)c^2(t, t, t)f^3(t)) \int [K * K(u)]^2 du$ .

Using that  $c^2(t_2, t_2, t_2) = \mathbb{E}(\mathbb{E}(V_{2,2}^2 | (\mathbf{x}_2, t_2)) | t_2) = \mathbb{E}(w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2) | t_2)$ , we get that

$$\begin{aligned} & \mathbb{E}(a^2(t_2)b^2(t_2)c^2(t_2, t_2, t_2)f^3(t_2)) = \mathbb{E}(a^2(t_2)b^2(t_2)\mathbb{E}(w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2) | t_2)f^3(t_2)) \\ = & \mathbb{E}\left(\frac{w^2(t_1)}{v_0^4(t_1)f(t_1)} \left[ \mathbb{E}\left(w_2(\mathbf{x}_1) \frac{H'(\mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1))}{V(H(\mathbf{x}_1^T \boldsymbol{\beta}_0 + \eta_0(t_1)))} \middle| t_1 \right) \right]^2 \mathbb{E}(w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1, t_1) | t_1)\right) = c_{2, \Psi}, \end{aligned}$$

concluding the proof.  $\square$

## Acknowledgements

This research was partially supported by Grants 20020100100276 and 20020100300057 from the Universidad de Buenos Aires, PIP 112-200801-00216 from CONICET, PICT 0821 from ANPCYT at Argentina and also by Spanish Grant MTM2008-03010/MTM of the Ministerio Español de Ciencia e Innovación and XUNTA Grant PGDIT07PXIB207031PR. This work began while Graciela Boente and Daniela Rodríguez were visiting the Departamento de Estadística e Investigación Operativa from the Universidad de Santiago de Compostela. The authors wish to thank the Editor, the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper.

## References

- Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plan. Inference*, **122**, 229–252.
- Bianco, A., Boente, G. and Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scand. J. Statist.*, **33**, 435–450.
- Bianco, A. García Ben, M. and Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canad. J. Statist.*, **33**, 1–18.
- Bianco, A. and Yohai, V. (1995). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17–34. Springer-Verlag, New York.
- Boente, G., Cao, R., González-Manteiga, W. and Rodríguez, D. (2012). Testing in generalized partly linear models: A robust approach. Technical Report, available at <http://www.ic.fcen.uba.ar/preprints.pdf>
- Boente, G., He, X. and Zhou, J. (2006). Robust Estimates in Generalized Partially Linear Models. *Ann. Statist.*, **34**, 2856–2878.
- Boente, G. and Rodríguez, D. (2010). Robust inference in generalized partially linear models. *Comp. Statist. Data Anal.*, **54**, 2942–2966.
- Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.*, **96**, 1022–1030.
- Carroll, R., Fan, J., Gijbels, I. and Wand, M. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.*, **92**, 477–489.
- Croux, C. and Haesbroeck, G. (2002). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273–295.
- Diebolt, J. (1995). A nonparametric test for the regression function: Asymptotic theory. *J. Statist. Plan. Inference*, **44**, 1–17.
- Gao, J. (1997). Adaptive parametric test in a semiparametric regression model. *Comm. Statist. Theory Methods*, **26**, 787–800.
- Gao, J. and Shi, P. (1997). M-type smoothing splines in nonparametric and semiparametric regression models. *Statist. Sinica*, **7**, 1155–1169.



- González-Manteiga, W. and Aneiros-Pérez, G. (2003). Testing in partial linear regression models with dependent errors. *J. Nonparametr. Statist.*, **15**, 93–111.
- Härdle, W., Liang, H. and Gao, J. (2000). *Partially linear models*. Springer-Verlag.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.
- Härdle, W., Mammen, E. and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc.*, **93**, 1461–1474.
- He, X. (2002). Robust tests in statistics – a review of the past decade. *Estadística*, **54**, 29–46.
- He, X., Zhu, Z. and Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579–590.
- Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.*, **89**, 897–904.
- Kojadinovic, I. and Yan, J. (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, **21**, 17–30.
- Koul, H. and Ni, P. (2004). Minimum distance regression model checking. *J. Statist. Plan. Inference*, **119**, 109–142.
- Koul, H. L. and Song, W. (2008). Regression model checking with Berkson measurement errors. *J. Statist. Plan. Inference*, **138**, 1615–1628.
- Koul, H. L. and Song, W. (2010). Model checking in partial linear regression models with Berkson measurement errors. *Statistica Sinica*, **20**, 1551–1579.
- Koul, H. L. and Stute, W. (1998). Lack of fit tests in regression with non-random design. *Applied Statistical Science*, **3**, 53–69.
- Künsch, H., Stefanski, L. and 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.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. London: Chapman and Hall. 2nd edition.
- Müller, M. (2001). Estimation and testing in generalized partial linear models A comparative study. *Statistics and Computing*, **11**, 299–309.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.
- Rémillard, B. and Scaillet, O. (2009) Testing for equality between two copulas. *J. Multivariate Anal.*, **100**, 377–386.
- Robinson, P. (1988). Root-n-consistent Semiparametric regression. *Econometrica*, **56**, 931–954.
- Rodríguez, D. (2008). *Estimación robusta en modelos parcialmente lineales generalizados*. Ph.D. thesis at the University of Buenos Aires. Available at <http://cms.dm.uba.ar/doc/tesis/danielarodriguez.pdf>
- Rodríguez Campos, M. C., González-Manteiga, W. and Cao, R. (1998). Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. *J. Statist. Plan. Inference*, **67**, 99–122.
- Severini, T. and Staniswalis, J. (1994). Quasi-likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.*, **89**, 501–511.
- Stefanski, L., Carroll, R. and Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413–424.
- Stone, C. (1977). Consistent nonparametric regression. *Ann. Statist.*, **5**, 595–645.
- Wang, L. and Qu, A. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *J. Amer. Statist. Assoc.*, **102**, 347–358.
- Zhu, L. (2005). *Nonparametric Monte Carlo tests and their applications*. Lecture notes in Statistics, **182**.
- Zhu, L. and Zhang, H. (2004). Hypothesis testing in mixture regression models. *J. Royal Statist. Soc., Ser. B*, **66**, 3–16.