Statistics and Probability Letters 97 (2015) 88–98

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Robust inference in partially linear models with missing responses

Ana M. Bianco [a,b,*], Graciela Boente [c,d], Wenceslao González-Manteiga [e], Ana Pérez-González [f]

[a] *Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, 1428, Buenos Aires, Argentina*

[b] *CONICET, Av. Rivadavia 1917, 1429, Buenos Aires, Argentina*

[c] *Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1428, Buenos Aires, Argentina*

[d] *IMAS, CONICET, Ciudad Universitaria, Pabellón 1, 1428, Buenos Aires, Argentina*

[e] *Departamento de Estatística e Investigación Operativa, Facultad de Matemáticas, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

[f] *Departamento de Estadística e Investigación Operativa, Facultad de Empresariales y Turismo, Universidad de Vigo, Campus de Orense, 32004 Orense, Spain*

ARTICLE INFO

Article history:
Received 13 March 2014
Received in revised form 22 October 2014
Accepted 7 November 2014
Available online 14 November 2014

MSC:
62G35
62F12
62F03

Keywords:
Kernel weights
Hypothesis testing
M-location functionals
Missing at random
Partly linear models
Robust estimation

ABSTRACT

We consider robust testing on the regression parameter of a partially linear regression model, where missing responses are allowed. We derive the asymptotic behavior of the proposed test statistic under the null and contiguous alternatives. A numerical study is performed.

1. Introduction

Non-parametric regression models suffer from the *curse of dimensionality* when the dimension of the covariates increases. Therefore, introducing some structure in the regression function the statistical analysis may become more efficient. Partially linear models (PLM) provide a solution to a large number of covariates by assuming that the regression function has two components: one depending linearly on some of the covariates, while the other one is non-parametric. In particular, PLM

* Correspondence to: Instituto de Cálculo, FCEyN, UBA, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina. Fax: +54 11 45763375.
E-mail addresses: abianco@dm.uba.ar (A.M. Bianco), gboente@dm.uba.ar (G. Boente), wenceslao.gonzalez@usc.es (W. González-Manteiga), anapg@uvigo.es (A. Pérez-González).

http://dx.doi.org/10.1016/j.spl.2014.11.004
0167-7152/

came to be more popular in the last years due to their flexibility, since the two components allow them to adapt to a wide class of situations. Sometimes, little is known about the relation among the response and some of the independent variables and hence, when the form of functional relation is unspecified, the use of a non-parametric component is recommended. In these situations, PLM are an appealing choice.

More formally, under a PLM, it is assumed that the response $y_i \in \mathbb{R}$ and the covariates $(\mathbf{x}_i^{\mathrm{T}}, t_i),\ \mathbf{x}_i \in \mathbb{R}^p,\ t_i \in \mathbb{R}$, are such that

$$y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \sigma\epsilon_i, \quad 1 \le i \le n, \tag{1}$$

where the errors ϵ_i are i.i.d., independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ with symmetric distribution $F_0(\cdot)$. That is, we assume that the error's scale equals 1 so as to identify the scale parameter as σ. We will not require any moment conditions on the errors distribution, but we only assume that the scale parameter for the errors equals 1. When the existence of seconds moments is assumed, as it is the case of the classical approach, these conditions imply that $\mathbb{E}(\epsilon_i) = 0$ and $\mathrm{VAR}(\epsilon_i) = 1$, which entails that, in this situation, σ represents the standard deviation of the responses conditional to the covariates.

Härdle et al. (2000, 2004) give an extensive description of different results obtained in PLM. In particular, in the context of hypothesis testing, Gao (1997) considers asymptotic test statistics for the problem $H_0 : \boldsymbol{\beta} = \mathbf{0}$, while González Manteiga and Aneiros Pérez (2003) studied the case of dependent errors. Classical procedures based on local polynomials and least squares estimation can be seriously damaged by a small fraction of anomalous observations. Robust estimates under the partly linear model were considered in He et al. (2002), where M-type estimates for repeated measurements using B-splines are introduced. On the other hand, Bhattacharya and Zhao (1997) define a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$ by taking differences of the observations and combining a bandwidth-matched M-estimation procedure with kernel weights, when $p = 1$ and the carriers $\mathbf{x}$ lie in a compact set. Bianco and Boente (2004) introduce a kernel-based three-step procedure in order to achieve robustness against anomalous data including high leverage points in $\mathbf{x}$.

Nevertheless, in practice, not all the responses may be available, this may be planned or unplanned. The methods described above are designed for complete data sets and problems arise when missing observations are present. In some cases, people may refuse to provide some kind of information, in others, the response variable may be very expensive or difficult to measure. Also, sometimes there may be loss of information in the registration process or the researcher may fail to collect the full information. There are many situations in which both the response and the explanatory variables have missing values, however we will focus our attention on those cases where missing data occur only in the responses.

Wang et al. (2004) considered regression imputation of missing responses based on partly linear regression model in order to make inference on the mean of y. The estimator of $\boldsymbol{\beta}$, introduced by Wang et al. (2004), is a least squares regression estimator defined by considering preliminary kernel estimators, of the quantities $\mathbb{E}(\delta_1\mathbf{x}_1|t_1 = t)/\mathbb{E}(\delta_1|t_1 = t)$ and $\mathbb{E}(\delta_1 y_1|t_1 = t)/\mathbb{E}(\delta_1|t_1 = t)$, where $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ if y_i is missing. Estimators of the marginal mean of the response y based on the obtained estimator of the regression parameter are defined using an imputation estimator and also propensity score weighting estimators. Wang and Sun (2007) studied estimators of the regression coefficients and the nonparametric function using either imputation, semiparametric regression surrogate or an inverse marginal probability weighted approach. Since these estimators are based on weighted means of the response variables, they are highly sensitive to outliers. The lack of robustness of weighted means procedures pushed on the search of procedures resistant to outliers as those given in Bianco et al. (2010), who introduced robust estimators based on bounded score functions together with algorithms to compute them. In this paper, we go further and we focus our attention on inference regarding the parametric component, when the response variable has missing observations, but the covariates $(\mathbf{x}^{\mathrm{T}}, t)$ are totally observed.

The rest of the paper is organized as follows. Section 2 reviews the definition of the robust semiparametric estimators defined in Bianco et al. (2010) and recalls some previous results. In Section 3, the Wald test statistics are introduced, while their asymptotic distribution is derived under the null hypothesis and under contiguous alternatives in Section 3.1. The results of a simulation study are reported in Section 4, while some final comments are given in Section 5. Technical proofs are left to the Appendix.

2. Preliminaries

Consider a random sample of incomplete data $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i\right)$, $1 \le i \le n$, of a partially linear model where $\delta_i = 1$ if y_i is observed, $\delta_i = 0$ if y_i is missing, and the responses y_i satisfy model (1).

As mentioned above, our goal is to introduce robust tests to check hypotheses that engage the regression parameter $\boldsymbol{\beta}$ in the case where responses are possibly missing, in particular when they are missing at random (MAR). This means that if $(y, \mathbf{x}^{\mathrm{T}}, t, \delta)$ has the same distribution as $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i\right)$, δ is conditionally independent of the response y given $(\mathbf{x}^{\mathrm{T}}, t)$. In other words, we assume an ignorable mechanism such that $\mathbb{P}\left(\delta = 1|(y, \mathbf{x}^{\mathrm{T}}, t)\right) = \mathbb{P}\left(\delta = 1|(\mathbf{x}^{\mathrm{T}}, t)\right) = p(\mathbf{x}, t)$.

One may wonder if, ignoring the vectors with missing responses, we will still obtain robust and consistent procedures. That is, if the robust estimators given in Bianco and Boente (2004) applied to the observations $\{\mathbf{z}_{i_1}, \ldots, \mathbf{z}_{i_N}\} = \{(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)^{\mathrm{T}}\}_{\delta_i=1}$, where $N = \sum_{i=1}^n \delta_i$, lead to asymptotically unbiased estimators so that, the tests defined through them in Bianco et al. (2006), turn out to be consistent. This is one of the conditions needed to successfully apply the transfer principle described in Koul et al. (2012). However, as mentioned in Bianco et al. (2010), a profile-likelihood procedure is needed to obtain consistent estimators for a wide class of situations when dealing with missing responses. Indeed, the robust estimators proposed in Bianco and Boente (2004) are not Fisher-consistent, unless the probability of missing responses is of the

form $p(\mathbf{x}, t) = p(t)$. This excludes interesting situations that may appear in practice. For that reason, since the transfer principle cannot be applied to the robust test defined in Bianco et al. (2006), we will consider the estimators addressed in Bianco et al. (2010) based on a profile-likelihood approach which combines the M-smoothers defined in Boente et al. (2009) with robust regression estimators. For the sake of clarity, we shortly remind the definition of these estimators.

2.1. Estimators of the regression parameter and regression function

Let ψ_1 be an odd and bounded score function and ρ be a *rho*-function as defined in Maronna et al. (2006, Chapter 2), i.e., a function ρ such that $\rho(x)$ is a nondecreasing function of $|x|$, $\rho(0) = 0$, $\rho(x)$ is increasing for $x > 0$ when $\rho(x) < \|\rho\|_\infty = \sup_x |\rho(x)|$. If ρ is bounded, it is also assumed that $\|\rho\|_\infty = 1$. We will consider kernel smoothers weights for the nonparametric component which are given by $w_i(\tau, h_n) = \delta_i K((t_i - \tau)/h_n) \left\{\sum_{j=1}^n \delta_j K\left((t_j - \tau)/h_n\right)\right\}^{-1}$, with K a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h_n the bandwidth parameter.

To define a robust estimator, Bianco et al. (2010) proceed as follows:

Step 1. For each τ and $\mathbf{b}$, define $g_{\mathbf{b}}(\tau)$ and its related estimate $\widehat{g}_{\mathbf{b}}(\tau)$ as the solutions of $S^{(1)}(g_{\mathbf{b}}(\tau), \mathbf{b}, \tau) = 0$ and $S_n^{(1)}(\widehat{g}_{\mathbf{b}}(\tau), \mathbf{b}, \tau) = 0$, respectively, where

$$S^{(1)}(a, \mathbf{b}, \tau) = \mathbb{E}\left[\delta\psi_1\left(\frac{y - \mathbf{x}^{\mathrm{T}}\mathbf{b} - a}{\sigma_{\mathbf{b}}}\right)\upsilon(\mathbf{x})\,|t = \tau\right], \tag{2}$$

$$S_n^{(1)}(a, \mathbf{b}, \tau) = \sum_{i=1}^n w_i(\tau, h_n)\psi_1\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - a}{\widehat{s}_{\mathbf{b}}}\right)\upsilon(\mathbf{x}_i),$$

with $\widehat{s}_{\mathbf{b}}$ a preliminary robust consistent scale estimator of $\sigma_{\mathbf{b}}$, the scale of $y - \mathbf{x}^{\mathrm{T}}\mathbf{b} - g_{\mathbf{b}}(\tau)$, and υ a weight function.

Step 2. The functional $\boldsymbol{\beta}(F)$, where F is the distribution of $(y, \mathbf{x}^{\mathrm{T}}, t, \delta)$, is defined as $\boldsymbol{\beta}(F) = \operatorname{argmin}_{\mathbf{b}} H(\mathbf{b})$, with $H(\mathbf{b}) = \mathbb{E}\left[\delta\rho\left((y - \mathbf{x}^{\mathrm{T}}\mathbf{b} - g_{\mathbf{b}}(t))/\sigma\right)\upsilon(\mathbf{x})\right]$. Its related estimate is defined as $\widehat{\boldsymbol{\beta}} = \operatorname{argmin}_{\mathbf{b}} H_n(\mathbf{b})$, where $H_n(\mathbf{b}) = \sum_{i=1}^n \delta_i\rho\left((y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \widehat{g}_{\mathbf{b}}(t_i))/\widehat{\sigma}\right)\upsilon(\mathbf{x}_i)/n$, with $\widehat{\sigma}$ a preliminary estimate of the scale σ, i.e., a robust M-scale computed using an initial (possibly inefficient) estimate of $\boldsymbol{\beta}$ with high breakdown point.

Step 3. Then, the functional $g(\tau, F)$ is defined as $g(\tau, F) = g_{\boldsymbol{\beta}(F)}(\tau)$, while the estimate of the nonparametric component is $\widehat{g}_n(\tau) = \widehat{g}_{\widehat{\boldsymbol{\beta}}}(\tau)$.

Let $\psi = \rho'$ be the derivative of the loss function ρ. It is worth noticing that the regression estimator defined in **Step 2** is the solution of

$$H_n^{(1)}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \widehat{g}_{\widehat{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\left(\mathbf{x}_i + \frac{\partial}{\partial\mathbf{b}}\widehat{g}_{\mathbf{b}}(t_i)\Big|_{\mathbf{b}=\widehat{\boldsymbol{\beta}}}\right) = 0. \tag{3}$$

As is well-known, leverage points in the covariates $\mathbf{x}$ may cause breakdown in regression models. For this reason, GM-, S- and MM-estimators have been introduced (see for instance, Maronna et al., 2006). By means of a score function ρ combined with a weight υ in **Step 2**, we include these robust families of estimators. Hence, the proposal is resistant against outliers in the residuals and in the carriers $\mathbf{x}$, as well. Usually, when computing MM-estimators, since they already control high-leverage points, the practitioner takes $\upsilon(\mathbf{x}) \equiv 1$. Bianco et al. (2010) described an algorithm to compute these estimators, where MM-estimators with initial LMS-estimators combined with S-estimators adapted to the partly linear setting are considered. If ψ_1 is chosen as the identity function, ρ is taken as the square function and $\upsilon \equiv 1$, this procedure will lead to the estimators introduced in Wang et al. (2004), which are non resistant to the presence of outlying observations. If in addition, $p \equiv 1$, i.e., when there are no missing responses, these estimators correspond to those defined in Speckman (1988) and studied in Robinson (1988).

2.2. Asymptotic distribution

In this section, we state the asymptotic behavior of the estimator $\widehat{\boldsymbol{\beta}}$ defined above, which was derived in Bianco et al. (2011). This result will be helpful to obtain the asymptotic distribution of the test statistic under the null hypothesis.

Assume that $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i\right)$, $1 \le i \le n$ are as above, i.e., $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + g(t_i) + \sigma\epsilon_i$ for $1 \le i \le n$. Denote ψ' and ψ'' the first and second derivatives of ψ. Moreover, let $\mathbf{z} = \mathbf{z}(\boldsymbol{\beta})$ with $\mathbf{z}(\mathbf{b}_0) = \mathbf{x} + (\partial g_{\mathbf{b}}(t)/\partial\mathbf{b})\,|_{\mathbf{b}=\mathbf{b}_0}$, $\mathbf{z}_i = \mathbf{z}_i(\boldsymbol{\beta})$ with $\mathbf{z}_i(\mathbf{b}_0) = \mathbf{x}_i + (\partial g_{\mathbf{b}}(t_i)/\partial\mathbf{b})\,|_{\mathbf{b}=\mathbf{b}_0}$ and

$$\widehat{\gamma}(\mathbf{b}, \tau) = \widehat{g}_{\mathbf{b}}(\tau) - g_{\mathbf{b}}(\tau) \qquad \widehat{\gamma}(\tau) = \widehat{\gamma}(\boldsymbol{\beta}, \tau) \tag{4}$$

$$\widehat{\upsilon}_j(\mathbf{b}, \tau) = \frac{\partial\widehat{\gamma}(\mathbf{b}, \tau)}{\partial b_j} \qquad \widehat{\upsilon}_j(\tau) = \widehat{\upsilon}_j(\boldsymbol{\beta}, \tau). \tag{5}$$

Furthermore, for any function $m : \mathcal{T} \to \mathbb{R}$ denote $\|m\|_\infty = \sup_{t\in\mathcal{T}} |m(t)|$. The first condition below states a MAR assumption, the second one is a condition on the preliminary estimate of $g_{\mathbf{b}}(\tau)$, while the other ones state requirements to the score and weight functions and to the underlying model distributions.

N0. δ and y are conditionally independent given $(\mathbf{x}^{\mathrm{T}}, t)$, that is, $\mathbb{P}(\delta = 1|(y, \mathbf{x}^{\mathrm{T}}, t)) = \mathbb{P}(\delta = 1|(\mathbf{x}^{\mathrm{T}}, t)) = p(\mathbf{x}, t)$.

N1. The functions $\widehat{g}_{\mathbf{b}}(\tau)$ and $g_{\mathbf{b}}(\tau)$ are continuously differentiable with respect to $(\mathbf{b}, \tau)$, twice continuously differentiable with respect to $\mathbf{b}$ and such that $(\partial^2 g_{\mathbf{b}}(\tau)/\partial b_j \partial b_\ell)|_{\mathbf{b}=\boldsymbol{\beta}}$ is bounded. Furthermore, for any $1 \le j, \ell \le p$, $\partial^2 g_{\mathbf{b}}(\tau)/\partial b_j \partial b_\ell$ satisfies the following equicontinuity condition:

$$\forall \epsilon > 0,\ \exists \delta > 0 : |\mathbf{b}_1 - \mathbf{b}_0| < \delta \Rightarrow \left\| \frac{\partial^2}{\partial b_j \partial b_\ell} g_{\mathbf{b}} \bigg|_{\mathbf{b}=\mathbf{b}_1} - \frac{\partial^2}{\partial b_j \partial b_\ell} g_{\mathbf{b}} \bigg|_{\mathbf{b}=\mathbf{b}_0} \right\|_\infty < \epsilon.$$

N2. The functions υ and $\Upsilon(\mathbf{x}) = \mathbf{x}\upsilon(\mathbf{x})$ are bounded and continuous. The function $\psi = \rho'$ is an odd, bounded and twice continuously differentiable function with bounded derivatives ψ' and ψ'', such that $\varphi_1(s) = s\psi'(s)$ and $\varphi_2(s) = s\psi''(s)$ are bounded. Moreover, the function ψ_1 is a bounded and continuously differentiable function with bounded derivative ψ_1'.

N3. The matrix $\mathbf{A}(\boldsymbol{\beta}) = \mathbb{E}\psi'(\epsilon)\ \mathbb{E}\left(\upsilon(\mathbf{x})p(\mathbf{x}, t)\mathbf{z}(\boldsymbol{\beta})\mathbf{z}(\boldsymbol{\beta})^{\mathrm{T}}\right)$ is non-singular.

N4. The matrix $\mathbf{B}(\boldsymbol{\beta}) = \mathbb{E}\psi^2(\epsilon)\ \mathbb{E}\left(\upsilon^2(\mathbf{x})p(\mathbf{x}, t)\mathbf{z}(\boldsymbol{\beta})\mathbf{z}(\boldsymbol{\beta})^{\mathrm{T}}\right)$ is positive definite.

N5. $\mathbb{E}\left(p(\mathbf{x}, t)\upsilon(\mathbf{x})\,\|\mathbf{z}(\boldsymbol{\beta})\|^2\right) < \infty$.

N6. $\mathbb{E}(\psi_1'(\epsilon)) \neq 0$ and $\mathbb{E}(\psi'(\epsilon)) \neq 0$.

N7. (a) $\left\|\widehat{g}_{\widehat{\boldsymbol{\beta}}} - g\right\|_\infty \xrightarrow{p} 0$, for any $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$.

(b) For each $\tau \in \mathcal{T}$ and $\mathbf{b}$, $\widehat{\gamma}(\mathbf{b}, \tau) \xrightarrow{p} 0$. Moreover, $n^{1/4}\,\|\widehat{\gamma}\|_\infty \xrightarrow{p} 0$ and $n^{1/4}\left\|\widehat{\upsilon}_j\right\|_\infty \xrightarrow{p} 0$ for all $1 \le j \le p$.

(c) There exists a neighborhood of $\boldsymbol{\beta}$ with closure $\mathcal{K}$ such that for any $1 \le j, \ell \le p$, $\sup_{\mathbf{b}\in\mathcal{K}}(\|\widehat{\upsilon}_j(\mathbf{b}, \cdot)\|_\infty + \|\partial\widehat{\upsilon}_j(\mathbf{b}, \cdot)/\partial b_\ell\|_\infty) \xrightarrow{p} 0$.

(d) $\|\partial\widehat{\gamma}/\partial\tau\|_\infty + \left\|\partial\widehat{\upsilon}_j/\partial\tau\right\|_\infty \xrightarrow{p} 0$ for any $1 \le j \le p$.

Remark 2.1. Using that $S^{(1)}(g_{\mathbf{b}}(\tau), \mathbf{b}, \tau) = 0$ for any $\mathbf{b} \in \mathbb{R}^p$ and that the errors have a symmetric distribution and are independent of the covariates, we obtain that **N6** implies

$$\mathbb{E}\left[\left(\mathbf{x} + \frac{\partial}{\partial \mathbf{b}} g_{\mathbf{b}}(\tau)\bigg|_{\mathbf{b}=\boldsymbol{\beta}}\right)\upsilon(\mathbf{x})p(\mathbf{x}, \tau)|t = \tau\right] = 0, \tag{6}$$

which ensures that $\widehat{g}_{\mathbf{b}}$ and its first derivative with respect to $\mathbf{b}$ can be replaced by the true functions.

The convergence requirements in **N7** are similar to those stated in Severini and Staniswalis (1994) and are needed to obtain root-n regression estimators. In particular, the continuity of $g_{\mathbf{b}}(\tau)$ with respect to $(\mathbf{b}, \tau)$ and Theorem 3.1 in Bianco et al. (2011) entail **N7**(a). For a discussion on the validity of **N7**(b)–(d), see Remark 6.2 of the above mentioned paper, where more comments on the remaining assumptions can be found.

Proposition 2.1. *Assume that t_1 is a random variable with distribution on a compact set and that the errors have a symmetric distribution and are independent of the covariates. If* ***N0*** to ***N7*** *hold and $\widehat{\sigma} \xrightarrow{p} \sigma$, then for any consistent solution $\widehat{\boldsymbol{\beta}}$ of (3), we have that $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{D} N\left(\mathbf{c}, \sigma^2\mathbf{A}^{-1}(\boldsymbol{\beta})\mathbf{B}(\boldsymbol{\beta})\mathbf{A}^{-1}(\boldsymbol{\beta})\right)$, where the symmetric matrices $\mathbf{A}(\boldsymbol{\beta})$ and $\mathbf{B}(\boldsymbol{\beta})$ are defined in* ***N3*** *and* ***N4****, respectively.*

Proposition 2.1 is used in Section 3 to define the Wald test statistic for the simple null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ and to derive its asymptotic distribution when H_0 holds.

3. Robust testing

In this Section, we mainly focus on testing hypotheses of the form $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ vs. $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ through a Wald-type test statistic based on the robust estimator $\widehat{\boldsymbol{\beta}}$ defined in Section 2.1.

In order to construct the Wald-type test statistic, we need to estimate the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$. Let $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i\right)$, $1 \le i \le n$, be a random sample satisfying (1). Define

$$\mathbf{A}(\mathbf{b}) = \mathbb{E}\left(\psi'(\epsilon(\mathbf{b}))\ \upsilon(\mathbf{x})p(\mathbf{x}, t)\mathbf{z}(\mathbf{b})\mathbf{z}(\mathbf{b})^{\mathrm{T}}\right) \quad \text{and} \quad \mathbf{B}(\mathbf{b}) = \mathbb{E}\left(\psi^2(\epsilon(\mathbf{b}))\ \upsilon^2(\mathbf{x})p(\mathbf{x}, t)\mathbf{z}(\mathbf{b})\mathbf{z}(\mathbf{b})^{\mathrm{T}}\right) \tag{7}$$

with $\epsilon(\mathbf{b}) = (y - \mathbf{x}^{\mathrm{T}}\mathbf{b} - g_{\mathbf{b}}(t))/\sigma$. Note that $\epsilon(\boldsymbol{\beta}) = \epsilon$, thus we obtain the matrices defined in **N3** and **N4**. These matrices involve the quantity $(\partial g_{\mathbf{b}}(t)/\partial \mathbf{b})\,|_{\mathbf{b}=\boldsymbol{\beta}}$, so its estimation is required. Since $S^{(1)}(g_{\mathbf{b}}(\tau), \mathbf{b}, \tau) = 0$, for all $\mathbf{b} \in \mathbb{R}^p$, differentiating with respect to $\mathbf{b}$ we get that

$$0 = \frac{-1}{\sigma_{\boldsymbol{\beta}}^2}\mathbb{E}\left[p(\mathbf{x}, t)\psi_1'\left(\frac{y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} - g_{\boldsymbol{\beta}}(\tau)}{\sigma_{\boldsymbol{\beta}}}\right)\upsilon(\mathbf{x})\left[\left(\mathbf{x} + \frac{\partial g_{\mathbf{b}}(\tau)}{\partial \mathbf{b}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}}\right)\sigma_{\boldsymbol{\beta}}\frac{\partial \sigma_{\mathbf{b}}}{\partial \mathbf{b}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}}\left(y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} - g_{\boldsymbol{\beta}}(\tau)\right)\right]|t = \tau\right].$$

As the observations satisfy (1), we have that $g_{\boldsymbol{\beta}} = g$ and $\sigma_{\boldsymbol{\beta}} = \sigma$, so we obtain that

$$0 = \mathbb{E}\left[\psi_1'(\epsilon)\,p(\mathbf{x}, t)\upsilon(\mathbf{x})\left(\mathbf{x} + \frac{\partial g_{\mathbf{b}}(\tau)}{\partial \mathbf{b}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}}\right)\sigma|t = \tau\right] + \mathbb{E}\left[\epsilon\psi_1'(\epsilon)\,p(\mathbf{x}, t)\upsilon(\mathbf{x})\frac{\partial \sigma_{\mathbf{b}}}{\partial \mathbf{b}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}}|t = \tau\right]. \tag{8}$$

Using the independence between the errors and the covariates, the symmetry of F_0 and the fact that the oddness of ψ_1 entails that $u\psi_1'(u)$ is an odd function, we get that $\mathbb{E}\left[\epsilon\psi_1'(\epsilon)\right] = 0$ implying that the right hand side term in (8) equals 0. Thus,

$$\left.\frac{\partial g_{\mathbf{b}}(\tau)}{\partial \mathbf{b}}\right|_{\mathbf{b}=\boldsymbol\beta} = -\frac{\mathbb{E}\left[\psi_1'\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta-g_0(\tau)}{\sigma}\right)\delta\upsilon(\mathbf{x})\,\mathbf{x}|t=\tau\right]}{\mathbb{E}\left[\psi_1'\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta-g_0(\tau)}{\sigma}\right)\delta\upsilon(\mathbf{x})\,|t=\tau\right]}.$$

It is worth noting that (6) entails that $(\partial g_{\mathbf{b}}(t)/\partial\mathbf{b})\,|_{\mathbf{b}=\boldsymbol\beta} = -\mathbb{E}\left[\delta\upsilon(\mathbf{x})\,\mathbf{x}|t=\tau\right]\{\mathbb{E}\left[\delta\upsilon(\mathbf{x})\,|t=\tau\right]\}^{-1}$. Hence, $(\partial g_{\mathbf{b}}(t)/\partial\mathbf{b})\,|_{\mathbf{b}=\boldsymbol\beta}$ does not depend on the score function ψ_1. However, in the estimation procedure we will use the score function ψ_1 in order to bound the effect of bad leverage points. Effectively, $(\partial g_{\mathbf{b}}(t)/\partial\mathbf{b})\,|_{\mathbf{b}=\boldsymbol\beta}$ will be estimated as

$$\widehat{g}_{\widehat{\boldsymbol\beta}}^{(\mathbf{b})}(\tau) = -\frac{\sum_{i=1}^{n} w_i(\tau, h_{\mathrm{DER}})\psi_1'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\widehat{g}_n(\tau)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\mathbf{x}_i}{\sum_{i=1}^{n} w_i(\tau, h_{\mathrm{DER}})\psi_1'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\widehat{g}_n(\tau)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)}, \tag{9}$$

where $w_i(\tau, h) = \delta_i K\left((t_i-\tau)/h\right)\left\{\sum_{j=1}^{n}\delta_j K\left((t_j-\tau)/h\right)\right\}^{-1}$ and the bandwidth h_{DER} used to estimate the partial derivative $(\partial g_{\mathbf{b}}(t)/\partial\mathbf{b})\,|_{\mathbf{b}=\boldsymbol\beta}$ may be different from that used in the estimation of $g_{\mathbf{b}}$. Note that when computing the estimator $\widehat{g}_{\widehat{\boldsymbol\beta}}^{(\mathbf{b})}$, we bound the effect of large residuals through the score function ψ_1. We may also control bad leverage points, even without using a weight function υ, choosing $\psi_1 = \rho_1'$, with ρ_1 a redescending loss function. The estimator $\widehat{g}_{\widehat{\boldsymbol\beta}}^{(\mathbf{b})}$ relies on the assumption that $\mathbb{E}\left[\delta\upsilon(\mathbf{x})\,|t=\tau\right] = \mathbb{E}\left[p(\mathbf{x},\tau)\upsilon(\mathbf{x})\,|t=\tau\right] \neq 0$, which means that there are enough responses at each neighborhood of t, since we already require that $\mathbb{E}\psi_1'(\epsilon) \neq 0$ to obtain the correct rate of convergence.

Denote $\widehat{\mathbf{z}}_i(\boldsymbol\beta) = \mathbf{x}_i + \widehat{g}_{\boldsymbol\beta}^{(\mathbf{b})}(t_i)$ and $\widehat{\epsilon}_i(\mathbf{b}) = (y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \widehat{g}_{\mathbf{b}}(t_i))/\widehat{\sigma}$, then estimators of $\mathbf{A}(\boldsymbol\beta)$ and $\mathbf{B}(\boldsymbol\beta)$ can be defined as $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol\beta})$ and $\widehat{\mathbf{B}} = \widehat{\mathbf{B}}(\widehat{\boldsymbol\beta})$, where

$$\widehat{\mathbf{A}}(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\,\psi'\left(\widehat{\epsilon}_i(\mathbf{b})\right)\;\upsilon(\mathbf{x}_i)\,\widehat{\mathbf{z}}_i(\mathbf{b})\widehat{\mathbf{z}}_i(\mathbf{b})^{\mathrm{T}} \quad\text{and}\quad \widehat{\mathbf{B}}(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\,\psi^2\left(\widehat{\epsilon}_i(\mathbf{b})\right)\;\upsilon^2(\mathbf{x}_i)\,\widehat{\mathbf{z}}_i(\mathbf{b})\widehat{\mathbf{z}}_i(\mathbf{b})^{\mathrm{T}}. \tag{10}$$

Lemma 6.1 in Bianco et al. (2011) entails that, for any fixed $\boldsymbol\beta$, under **N0**, **N1**, **N2**, **N5** and **N7**(a) the matrices $\widehat{\mathbf{A}}(\widehat{\boldsymbol\beta})$ and $\widehat{\mathbf{B}}(\widehat{\boldsymbol\beta})$ provide consistent estimators of $\mathbf{A}(\boldsymbol\beta)$ and $\mathbf{B}(\boldsymbol\beta)$, respectively. This result together with Proposition 2.1 suggests the following Wald test statistic to test $H_0 : \boldsymbol\beta = \boldsymbol\beta_0$

$$\widehat{\mathcal{W}}_n = \frac{n\,(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)^{\mathrm{T}}\left(\widehat{\mathbf{A}}\widehat{\mathbf{B}}^{-1}\widehat{\mathbf{A}}\right)(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0)}{\widehat{\sigma}^2}.$$

Lemma A.1 generalizes the above mentioned Lemma to deal with contiguous alternatives, since it allows to derive the consistency of the matrices $\widehat{\mathbf{A}}(\widetilde{\boldsymbol\beta})$ and $\widehat{\mathbf{B}}(\widetilde{\boldsymbol\beta})$ to $\mathbf{A}(\boldsymbol\beta_0)$ and $\mathbf{B}(\boldsymbol\beta_0)$, respectively, when model (1) holds for $\boldsymbol\beta = \boldsymbol\beta_n = \boldsymbol\beta_0 + \mathbf{c}\,n^{-1/2}$ and $\widetilde{\boldsymbol\beta} \xrightarrow{p} \boldsymbol\beta_0$.

When there are no missing responses in the sample, Bianco et al. (2006) also considered a score type test. In our setting, a score type test can also be considered, but based on the profile estimators $\widehat{\boldsymbol\beta}$. However, this approach is beyond the scope of this paper.

3.1. Asymptotic behavior of the test statistics

The asymptotic behavior under the null and local alternatives of the Wald statistic is derived in this Section. As mentioned above, under $H_0 : \boldsymbol\beta = \boldsymbol\beta_0$, the asymptotic distribution of the test statistic, given in Theorem 3.1, follows from Proposition 2.1 and the convergence of $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ to $\mathbf{A}(\boldsymbol\beta_0)$ and $\mathbf{B}(\boldsymbol\beta_0)$, given in Lemma 6.1 of Bianco et al. (2011).

Theorem 3.1. *Assume that t_1 is a random variable with distribution on a compact set $\mathcal{T}$ and that $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i)$ satisfy model (1) for $\boldsymbol\beta = \boldsymbol\beta_0$, i.e., $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta_0 + g(t_i) + \sigma\epsilon_i$, where ϵ_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ and have symmetric distribution. If **N0**–**N7** hold for $\boldsymbol\beta = \boldsymbol\beta_0$ and $\widehat{\sigma} \xrightarrow{p} \sigma$, we have that $\widehat{\mathcal{W}}_n \xrightarrow{D} \chi^2_p$.*

Note that the test statistic is asymptotically χ^2_p distributed under the null hypothesis, which is the same asymptotic distribution of the classical test based on local means and least squares estimation.

In order to state the asymptotic behavior under local alternatives, we must generalize assumption **N7** to the case of contiguous alternatives of the form $\boldsymbol\beta_n = \boldsymbol\beta_0 + \mathbf{c}n^{-\frac{1}{2}}$.

N8. When $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n + g(t_i) + \epsilon_i,\ 1 \le i \le n$, with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}$, if $\widehat{\gamma}_n(\tau) = \widehat{\gamma}(\boldsymbol{\beta}_n, \tau)$ and $\widehat{\upsilon}_{j,n}(\tau) = \widehat{\upsilon}_j(\boldsymbol{\beta}_n, \tau)$, it holds that

(a) $\left\|\widehat{g}_{\widehat{\boldsymbol{\beta}}} - g\right\|_\infty \xrightarrow{p} 0$, for any $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$.

(b) For each $\tau \in \mathcal{T}$ and $\mathbf{b}$, $\widehat{\gamma}(\mathbf{b}, \tau) \xrightarrow{p} 0$. Moreover, $n^{1/4}\,\|\widehat{\gamma}_n\|_\infty \xrightarrow{p} 0$ and $n^{1/4}\left\|\widehat{\upsilon}_{j,n}\right\|_\infty \xrightarrow{p} 0$ for all $1 \le j \le p$.

(c) There exists a neighborhood of $\boldsymbol{\beta}_0$ with closure $\mathcal{K}$ such that $\sup_{\mathbf{b}\in\mathcal{K}}(\|\widehat{\upsilon}_j(\mathbf{b}, \cdot)\|_\infty + \|\partial\widehat{\upsilon}_j(\mathbf{b}, \cdot)/\partial b_\ell\|_\infty) \xrightarrow{p} 0$, for any $1 \le j, \ell \le p$.

(d) $\|\partial\widehat{\gamma}_n/\partial\tau\|_\infty + \left\|\partial\widehat{\upsilon}_{j,n}/\partial\tau\right\|_\infty \xrightarrow{p} 0$ for any $1 \le j \le p$.

It is worth noticing that **N8** is analogous to **N7**, but under a sequence of contiguous models. Hence, the validity of **N8** follows under similar conditions to those considered for **N7**.

Theorem 3.2 gives the asymptotic distribution of the test statistic under contiguous alternatives. Its proof is an immediate consequence of Lemmas A.1 and A.2 in the Appendix. Note that the non-centrality parameter depends on the loss function ρ used through its derivative ψ, so that some loss of power may be expected due to the balance between robustness and efficiency.

Theorem 3.2. *Let t_1 be a random variable with distribution on a compact set $\mathcal{T}$. Assume that $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i)$, $1 \le i \le n$, satisfy model (1) with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$, i.e., $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n + g\,(t_i) + \sigma\epsilon_i$, where ϵ_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ and have symmetric distribution. Assume that **N0**–**N6** hold for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. If in addition, **N8** holds and $\widehat{\sigma} \xrightarrow{p} \sigma$, we have that under $H_1^n : \boldsymbol{\beta} = \boldsymbol{\beta}_n$, $\widehat{\mathcal{W}}_n \xrightarrow{D} \chi_p^2(\theta)$, where $\theta = \mathbf{c}^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\mathbf{c}/\sigma^2$ with $\boldsymbol{\Sigma}_0 = \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}$, for $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}_0 = \mathbf{B}(\boldsymbol{\beta}_0)$ defined in (7).*

Similar results to those given in Theorem 3.1 can be obtained when the null hypothesis involves only a subset of q components of the regression parameter, by adapting assumptions **N3**–**N5** and also **N7** or **N8** to the actual null hypothesis. This is one of the most frequent hypothesis testing problems in regression. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^{\mathrm{T}}, \boldsymbol{\beta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_{(1)}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\beta}_{(1)} \in \mathbb{R}^q$. In order to test $H_0 : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\boldsymbol{\beta}_{(2)}$ unspecified, one may use the statistic $\widehat{\mathcal{W}}_{1,n} = n(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{11}^{-1}(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})/\widehat{\sigma}^2$ where $\widehat{\boldsymbol{\Sigma}}_{11}$ denotes the $q \times q$ submatrix of the matrix $\widehat{\boldsymbol{\Sigma}} \in \mathbb{R}^{p\times p}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$, $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}$, $\widehat{\mathbf{A}} = \widehat{\mathbf{A}}(\widehat{\boldsymbol{\beta}})$ and $\widehat{\mathbf{B}} = \widehat{\mathbf{B}}(\widehat{\boldsymbol{\beta}})$ defined in (10).

The following theorem states the asymptotic distribution of the Wald-type statistic $\widehat{\mathcal{W}}_{1,n}$. Its proof is similar to that of Theorem 3.1, so it is omitted.

Theorem 3.3. *Let t_1 be a random variable with distribution on a compact set $\mathcal{T}$ and $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i)$, $1 \le i \le n$, be i.i.d. observations satisfying (1) and **N0**, where the errors are independent of the covariates and have symmetric distribution. Assume that $\widehat{\sigma} \xrightarrow{p} \sigma$ and that, for any $\boldsymbol{\beta}_{(2)}$, **N1**–**N7** hold when $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1),0}^{\mathrm{T}}, \boldsymbol{\beta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$. Then, we have that*

(a) *Under $H_0 : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\widehat{\mathcal{W}}_{1,n} \xrightarrow{D} \chi_q^2$.*

(b) *Under $H_1^n : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0} + \mathbf{c}_{(1)}n^{-1/2}$, $\widehat{\mathcal{W}}_{1,n} \xrightarrow{D} \chi_q^2(\theta_1)$, with $\theta_1 = \mathbf{c}_{(1)}^{\mathrm{T}}\boldsymbol{\Sigma}_{0,11}^{-1}\mathbf{c}_{(1)}/\sigma^2$, where $\boldsymbol{\Sigma}_0 = \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}$, if in addition **N8** holds taking $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{(1),0}^{\mathrm{T}}, \boldsymbol{\beta}_{(2)}^{\mathrm{T}})^{\mathrm{T}}$.*

4. Monte Carlo study

A simulation study was carried out in order to assess the performance of the proposed test and also to compare its behavior with that of the classical one under contamination and under normal samples, for different missing probability schemes.

For both, the classical and robust smoothing procedures, we use the Gaussian kernel. For the robust smoothing procedure, we compute the robust local M-estimates with the bisquare function as score function ψ_1 with bandwidth h. For the computation of $\widehat{g}_{\boldsymbol{\beta}}^{(\mathbf{b})}(\tau)$, we consider its derivative ψ_1' with bandwidth h_{DER}, as described in (9). We choose as tuning constant for the bisquare function the value 4.685, which gives a 95% efficiency with respect to its linear relative. To compute the local M-estimates, local medians are selected as initial estimates in the iterative procedure.

The robust estimator of the regression parameter $\boldsymbol{\beta}$ is computed as described in Section 3 of Bianco et al. (2010) using as *rho*-function the bisquare function, that is, choosing $\rho_0(x) = \rho_{\mathrm{TUK}}\,(x/c_0)$ and $\rho(x) = \rho_{\mathrm{TUK}}\,(x/c_1)$, with $c_0 = 1.56$, $c_1 \ge c_0$ and $\rho_{\mathrm{TUK}}(x) = \min(1, 1 - (1 - x^2)^3)$. The value selected for c_0 ensures Fisher-consistency of the scale when the errors are Gaussian, while $c_1 = 4.68$ guarantees that under a regression model the resulting estimates will achieve 95% efficiency.

In a first step, we generate observations (z_i, x_i, t_i) according to the model $z_i = \beta x_i + \sin(2\pi(t_i - 0.5)) + \sigma\epsilon_i,\ 1 \le i \le n$, where $\beta = 2$ and $\sigma^2 = 0.25$ in the non-contaminated case, which we identify as C_0. Besides, the covariates (x_i, t_i) are such that $x_i \sim N(0, 1)$ and $t_i \sim \mathcal{U}(0, 1)$ independent of each other, while the errors are $\epsilon_i \sim N(0, 1)$. Then, missing responses are introduced using different missing schemes to be described below, that is, we define $y_i = z_i$ if $\delta_i = 1$ and missing otherwise.

For each of the situations to be considered below, we perform 1000 replications generating independent samples. For each replication, we test the null hypothesis $H_0 : \beta = 2$ through the test statistic $\widehat{\mathcal{W}}_n$ and the classical Wald-type statistics $\widehat{\mathcal{W}}_{n,\mathrm{LS}}$ based on the least squares estimator, that is the estimator defined in Wang et al. (2004).

Table 1
Observed frequencies of rejection at $\beta = 2$, for nominal levels $\alpha = 0.05$ and $\alpha = 0.10$, when $n = 100$, 200 and 500 under C_0.

		$\alpha = 0.05$				$\alpha = 0.10$			
		$\widehat{\mathcal{W}}_{n,\text{LS}}$	$\widehat{\mathcal{W}}_n$			$\widehat{\mathcal{W}}_{n,\text{LS}}$	$\widehat{\mathcal{W}}_n$		
			h_{DER}				h_{DER}		
n	h		0.04	0.075	0.1		0.04	0.075	0.1
100	0.05	0.060	0.062	0.077	0.082	0.108	0.103	0.129	0.136
	0.075	0.057	0.044	0.055	0.058	0.099	0.081	0.103	0.109
	0.10	0.052	0.046	0.056	0.058	0.091	0.075	0.09	0.093
	0.20	0.040	0.039	0.048	0.051	0.087	0.068	0.089	0.094
200	0.05	0.062	0.059	0.069	0.070	0.105	0.109	0.119	0.120
	0.075	0.053	0.048	0.055	0.055	0.097	0.093	0.100	0.100
	0.10	0.043	0.044	0.050	0.053	0.090	0.086	0.095	0.097
	0.20	0.041	0.042	0.046	0.049	0.078	0.075	0.088	0.089
500	0.05	0.049	0.060	0.061	0.061	0.111	0.114	0.119	0.119
	0.075	0.044	0.060	0.060	0.061	0.108	0.108	0.117	0.118
	0.10	0.035	0.042	0.047	0.048	0.095	0.097	0.099	0.101
	0.20	0.036	0.048	0.050	0.051	0.083	0.105	0.108	0.113

Even if a full study on the level dependence on the bandwidths h and h_{DER} is beyond the scope of the paper, in a first stage, our concern is the level of the tests and how it may be influenced by the choice of the smoothing parameters. For that purpose, we consider three sample sizes $n = 100$, 200 and 500 and different values of the bandwidths, more precisely, we choose $h = 0.05, 0.075, 0.10$ and 0.20 and $h_{\text{DER}} = 0.04, 0.075$ and 0.1.

We first describe the results for the situation in which there are no missing responses which corresponds to the complete data case, that is, $p(x, t) \equiv 1$ and $y_i = z_i$. Table 1 gives, in this situation for the non-contaminated case C_0, the observed frequencies of rejection under the null hypothesis for the different sample sizes and bandwidths and for two nominal levels $\alpha = 0.05$ and 0.10. In most cases, for the bandwidth choices $h = h_{\text{DER}} = 0.075$ and $h = 0.10$, $h_{\text{DER}} = 0.075$, the robust test based on $\widehat{\mathcal{W}}_n$ reaches the closest values to the nominal levels. Besides, the classical test based on $\widehat{\mathcal{W}}_{n,\text{LS}}$ also attains observed frequencies of rejection very close to the nominal values of α for these smoothing parameters. Hence, from now on we consider these bandwidth parameters. On the other hand, since we consider below missing schemes with at least 30% of missing responses, a sample size of $n = 100$ may be not large enough. Besides, $n = 200$ seems a good compromise between a moderate sample size and the required number of observations to achieve the desired level α. For these reasons, from now on we only report the results when $n = 200$ and $\alpha = 0.05$. Similar results were obtained for the nominal level $\alpha = 0.10$.

In a second stage, we take into account four contamination schemes in order to evaluate their impact on the level and power of the classical and robust tests. The considered contaminations are

- $C_1 : \epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, 1) + 0.1N(0, 25)$. In this contamination only the errors are inflated and it is expected that it will affect moderately both level and power.
- $C_2 : \epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, 1) + 0.1N(0, 25)$ and artificially 20 observations of the response z_i, but not of the carriers x_i, are modified to be equal to 20 at equally spaced values of t. This contamination introduces 10% of outliers with high-residuals, so that it will have influence on the test power.
- $C_3 : \epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, 1) + 0.1N(0, 25)$ and artificially 20 observations of the carriers x_i, but not of the response z_i, are modified to be equal to 20 at equally spaced values of t. In this case, high-leverage points are introduced to assess how the bias of the regression parameter estimates affects the level of the test.
- $C_4 : \epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, 1) + 0.1N(0, 25)$ and artificially 10 observations of the carriers x_i and 10 of the response z_i, are modified to be equal to 20 and -20, respectively at equally spaced values of t. The outlying responses are not allocated at the same t than the outlying carriers. This case corresponds to introduce both high-leverage points and high-residuals.

We compute the observed frequencies of rejection at $\beta = 2 + \Delta n^{-1/2}$, $n = 200$ for $\Delta = 0, 0.25, 0.5, 0.75, 1, 1.5$ and 2 and we summarize the obtained results in Table 2.

As expected, under C_3 and C_4, the classical test $\widehat{\mathcal{W}}_{n,\text{LS}}$ becomes non-informative since its estimated power function equals 1. Besides, under C_2 the test $\widehat{\mathcal{W}}_{n,\text{LS}}$ leads to a power function which decreases with Δ, leading to wrong conclusions. Contamination C_1 seems to be the less harmful for $\widehat{\mathcal{W}}_{n,\text{LS}}$ since both its level and power are only slightly modified. Its major effect is a loss of power. Only the scenario without contamination, C_0, is favorable to the classical test $\widehat{\mathcal{W}}_{n,\text{LS}}$. On the other hand, the robust test $\widehat{\mathcal{W}}_n$ is stable under all contaminations, leading to reliable results for both choices $h = h_{\text{DER}} = 0.075$ and $h = 0.10$ combined with $h_{\text{DER}} = 0.075$.

In a third stage, we introduce missing at random responses according to different patterns. As mentioned above, we define $y_i = z_i$, if $\delta_i = 1$, and missing otherwise, where δ_i are generated as Bernoulli random variables using the following missing data models: (i) $P_1 : p(x, t) = 0.4 + 0.5(\cos(2(x + 0.2)))^2$, (ii) $P_2 : p(x, t) = 0.4 + 0.5(\cos(2(t + 0.2)))^2$, (iii) $P_3 : p(x, t) = 0.4+0.5(\cos(2(xt+0.2)))^2$ and (iv) $P_4 : p(x, t) = 1/(1+\exp(-2x-12(t-0.5)))$, which lead to an approximated proportion of missing responses of 0.3494, 0.4572, 0.2951 and 0.5006, respectively.

Table 2
Observed frequencies of rejection at $\beta = 2 + \Delta n^{-1/2}$, for $n = 200$, with nominal level $\alpha = 0.05$, $h_{\text{DER}} = 0.075$ and $h = 0.075$ and 0.1 when $p(x, y) \equiv 1$.

	$h = 0.075$							$h = 0.10$						
	Δ							Δ						
	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2
	C_0													
$\widehat{W}_{n,\text{LS}}$	0.053	0.088	0.165	0.305	0.516	0.861	0.984	0.043	0.082	0.157	0.290	0.482	0.845	0.976
$\widehat{W}_n$	0.055	0.090	0.154	0.294	0.485	0.824	0.972	0.050	0.078	0.173	0.280	0.455	0.796	0.970
	C_1													
$\widehat{W}_{n,\text{LS}}$	0.047	0.063	0.086	0.147	0.220	0.417	0.617	0.044	0.061	0.087	0.138	0.218	0.413	0.611
$\widehat{W}_n$	0.062	0.086	0.155	0.253	0.408	0.746	0.918	0.054	0.082	0.141	0.237	0.394	0.709	0.909
	C_2													
$\widehat{W}_{n,\text{LS}}$	0.069	0.065	0.060	0.060	0.061	0.059	0.062	0.068	0.065	0.064	0.062	0.059	0.058	0.056
$\widehat{W}_n$	0.065	0.081	0.141	0.244	0.380	0.686	0.882	0.050	0.069	0.132	0.228	0.367	0.670	0.878
	C_3													
$\widehat{W}_{n,\text{LS}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\widehat{W}_n$	0.059	0.074	0.147	0.246	0.377	0.687	0.893	0.057	0.069	0.136	0.226	0.356	0.674	0.877
	C_4													
$\widehat{W}_{n,\text{LS}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\widehat{W}_n$	0.059	0.074	0.147	0.246	0.377	0.687	0.893	0.057	0.069	0.136	0.226	0.356	0.674	0.877

Table 3
Observed frequencies of rejection at $\beta = 2 + \Delta n^{-1/2}$, for $n = 200$, with nominal level $\alpha = 0.05$, $h_{\text{DER}} = 0.075$ and $h = 0.075$ and 0.1 when $p(x, t) = 0.4 + 0.5(\cos(2(x + 0.2)))^2$.

	$h = 0.075$							$h = 0.10$						
	Δ							Δ						
	0	0.25	0.5	0.75	1	1.5	2	0	0.25	0.5	0.75	1	1.5	2
	C_0													
$\widehat{W}_{n,\text{LS}}$	0.058	0.083	0.145	0.242	0.375	0.698	0.904	0.048	0.077	0.135	0.232	0.355	0.678	0.895
$\widehat{W}_n$	0.062	0.091	0.165	0.243	0.355	0.648	0.861	0.055	0.077	0.150	0.222	0.333	0.628	0.852
	C_1													
$\widehat{W}_{n,\text{LS}}$	0.063	0.074	0.100	0.136	0.190	0.311	0.506	0.064	0.067	0.100	0.139	0.187	0.302	0.496
$\widehat{W}_n$	0.074	0.088	0.125	0.214	0.303	0.552	0.796	0.060	0.079	0.122	0.186	0.269	0.528	0.759
	C_2													
$\widehat{W}_{n,\text{LS}}$	0.067	0.062	0.061	0.060	0.059	0.055	0.057	0.069	0.063	0.060	0.058	0.055	0.058	0.056
$\widehat{W}_n$	0.066	0.096	0.137	0.195	0.273	0.506	0.739	0.061	0.069	0.114	0.180	0.271	0.473	0.719
	C_3													
$\widehat{W}_{n,\text{LS}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\widehat{W}_n$	0.062	0.083	0.131	0.196	0.284	0.514	0.739	0.052	0.067	0.114	0.178	0.266	0.483	0.710
	C_4													
$\widehat{W}_{n,\text{LS}}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\widehat{W}_n$	0.066	0.091	0.135	0.195	0.279	0.511	0.739	0.053	0.071	0.116	0.179	0.277	0.494	0.712

Table 3 summarizes the results corresponding to the missing probability P_1. Besides, in the supplementary file available online (see Appendix B), Table S.1 corresponds to P_2, Table S.2 to P_3, while the results from the logistic missing probability P_4 are given in Table S.3.

With respect to the effect of the missing schemes, a loss of power is observed for both the classical and robust tests, under C_0. In particular, Tables S.1 and S.3 in the supplementary file (see Appendix B) show that the largest loss of power is attained for the missing probability schemes P_2 and P_4. This behavior can be explained by the fact that, in average, almost half of the observations are lost in these two cases.

When analyzing the effect of the contaminations on the classical test $\widehat{W}_{n,\text{LS}}$, the same conclusions obtained for the complete case remain valid for all the missing schemes. On the other hand, for the robust procedure $\widehat{W}_n$, some loss of level is observed for the considered missing schemes in particular, when $p(x, t) = 0.4 + 0.5(\cos(2(t + 0.2)))^2$ and $p(x, t) = 1/(1 + \exp(-2x - 12(t - 0.5)))$ (see the supplementary file, Appendix B).

We also observe some loss of power under the missing data model P_4, where the percentage of missing observations is close to 50%. However, the proposed test is stable and still informative in all the contaminated situations.

5. Final comments

In this paper, we have introduced a Wald type test statistic, based on a robust three step estimation procedure, for linear hypotheses related to the regression parameter. The robust test statistic involves the selection of tuning constants for the

rho-functions allowing to compute the regression parameter estimate and for the score function used in **Step 1**. As in our simulation study, in most cases, these constants are selected by the user to attain a desired efficiency for Gaussian errors.

On the other hand, the test statistic depends on the bandwidth parameters used to estimate the nonparametric component and its derivative. As shown in Section 4, for both the classical and robust Wald statistics, the choice of the smoothing parameters is important to study the performance in terms of power and level. However, this relevant topic is beyond the scope of this paper and is still an open problem even for goodness of fit tests based on linear estimators. Some interesting discussions regarding the choice of the regularization parameters for the classical estimators can be found in González-Manteiga and Crujeiras (2013) and Sperlich (2014).

Acknowledgments

This work began while Graciela Boente was visiting the Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela. This research was partially supported by Grants PIP 112-201101-00339 from CONICET, PICT 2011-0397 from ANPCYT and w276 and 20120130100241BA from the Universidad de Buenos Aires at Buenos Aires, Argentina and also by the Spanish Project MTM2008-03010 from the Ministry of Science and Innovation. The authors thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper.

Appendix A

Using similar arguments to those considered in Lemma 6.1 from Bianco et al. (2011) we obtain the following result. Recall that $\widehat{\mathbf{A}}(\mathbf{b})$ and $\widehat{\mathbf{B}}(\mathbf{b})$ are defined in (10), while $\mathbf{A} = \mathbf{A}(\boldsymbol{\beta})$ and $\mathbf{B} = \mathbf{B}(\boldsymbol{\beta})$ are given in (7). Let $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}_0 = \mathbf{B}(\boldsymbol{\beta}_0)$.

Lemma A.1. *Let t_1 be a random variable with distribution on a compact set $\mathcal{T}$. Moreover, assume that, for $1 \le i \le n$, $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n + g(t_i) + \sigma\epsilon_i$, where $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$ and ϵ_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ and have symmetric distribution. Assume that* **N0**, **N1**, **N2** *and* **N5** *hold for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. If in addition,* **N8**(a) *holds, $\widehat{\sigma} \xrightarrow{p} \sigma$ and $\widetilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, then we have that $\widehat{\mathbf{A}}(\widetilde{\boldsymbol{\beta}}) \xrightarrow{p} \mathbf{A}_0$ and $\widehat{\mathbf{B}}(\widetilde{\boldsymbol{\beta}}) \xrightarrow{p} \mathbf{B}_0$. Moreover, we have that $\widehat{\mathbf{C}} \xrightarrow{p} \mathbf{A}_0$ where*

$$\widehat{\mathbf{C}} = (1/n)\sum_{i=1}^{n}\left(\psi'\left(\widehat{\epsilon}_i(\widetilde{\boldsymbol{\beta}})\right)\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})^{\mathrm{T}} + \psi\left(\widehat{\epsilon}_i(\widetilde{\boldsymbol{\beta}})\right)\left\{\partial^2\widehat{g}_{\mathbf{b}}(t_i)/\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}\right\}\Big|_{\mathbf{b}=\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\right)\delta_i\upsilon(\mathbf{x}_i)$$

with $\widehat{\epsilon}_i(\widetilde{\boldsymbol{\beta}}) = (y_i - \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i))/\widehat{\sigma}$.

Proof. We will only show that $\widehat{\mathbf{C}} \xrightarrow{p} \mathbf{A}_0$, since similar arguments lead to the consistency of $\widehat{\mathbf{A}}(\widetilde{\boldsymbol{\beta}})$ and $\widehat{\mathbf{B}}(\widetilde{\boldsymbol{\beta}})$.

Denote by $\widetilde{\widetilde{\boldsymbol{\beta}}} = \widetilde{\boldsymbol{\beta}} - \mathbf{c}n^{-1/2}$, $y_{i,0} = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + g(t_i) + \epsilon_i$ and by ξ an intermediate point between $\epsilon_i + \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta}_0 - \widetilde{\widetilde{\boldsymbol{\beta}}})$ and $\widehat{\epsilon}(\widetilde{\boldsymbol{\beta}}) = \epsilon_i + \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta}_0 - \widetilde{\widetilde{\boldsymbol{\beta}}}) + (g(t_i) - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i))$, $\mathbf{z}_i = \mathbf{x}_i + (\partial g_{\mathbf{b}}(t_i)/\partial\mathbf{b})\,|_{\mathbf{b}=\boldsymbol{\beta}_0}$. As in Lemma 6.1 of Bianco et al. (2011), we have that a Taylor expansion of first order and some algebra lead us to $\widehat{\mathbf{C}} = \sum_{j=1}^{6}\widehat{\mathbf{C}}_n^{(j)}$ with $\widehat{\mathbf{C}}_n^{(1)} = \sum_{i=1}^{n}\delta_i\psi'([y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - g(t_i)]/\widehat{\sigma})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\upsilon(\mathbf{x}_i)/n$, $\widehat{\mathbf{C}}_n^{(3)} = \sum_{i=1}^{n}\delta_i\psi''([y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - \xi_{i,1}]/\widehat{\sigma})\widehat{w}_0(t_i)\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\upsilon(\mathbf{x}_i)/(n\widehat{\sigma})$, $\widehat{\mathbf{C}}_n^{(6)} = \sum_{i=1}^{n}\delta_i\psi([y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)]/\widehat{\sigma})\,\widehat{\mathbf{V}}(t_i)^{\mathrm{T}}\upsilon(\mathbf{x}_i)/n$ and

$$\widehat{\mathbf{C}}_n^{(2)} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi\left(\frac{y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - g(t_i)}{\widehat{\sigma}}\right)\frac{\partial^2}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}g_{\mathbf{b}}(t_i)\Big|_{\mathbf{b}=\boldsymbol{\beta}_0}^{\mathrm{T}}\upsilon(\mathbf{x}_i)$$

$$\widehat{\mathbf{C}}_n^{(4)} = \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\delta_i\psi'\left(\frac{y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - \xi_{i,2}}{\widehat{\sigma}}\right)\widehat{w}_0(t_i)\,\frac{\partial^2}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}g_{\mathbf{b}}(t_i)\Big|_{\mathbf{b}=\boldsymbol{\beta}_0}^{\mathrm{T}}\upsilon(\mathbf{x}_i)$$

$$\widehat{\mathbf{C}}_n^{(5)} = \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi'\left(\frac{y_{i,0} - \mathbf{x}_i^{\mathrm{T}}\widetilde{\widetilde{\boldsymbol{\beta}}} - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\left[\widehat{\mathbf{w}}(t_i)\mathbf{z}_i^{\mathrm{T}} + \mathbf{z}_i\widehat{\mathbf{w}}(t_i)^{\mathrm{T}} + \widehat{\mathbf{w}}(t_i)\,\widehat{\mathbf{w}}(t_i)^{\mathrm{T}}\right]\upsilon(\mathbf{x}_i),$$

where $\xi_{i,1}$ and $\xi_{i,2}$ are intermediate points, $\mathbf{z}_i = \mathbf{z}_i(\boldsymbol{\beta}_0)$, $\widehat{w}_0(t) = \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t) - g(t)$ and

$$\widehat{\mathbf{w}}(t) = \frac{\partial}{\partial\mathbf{b}}\widehat{g}_{\mathbf{b}}(t)\Big|_{\mathbf{b}=\widetilde{\boldsymbol{\beta}}} - \frac{\partial}{\partial\mathbf{b}}g_{\mathbf{b}}(t)\Big|_{\mathbf{b}=\boldsymbol{\beta}_0} \qquad \widehat{\mathbf{V}}(t) = \frac{\partial^2}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}\widehat{g}_{\mathbf{b}}(t_i)\Big|_{\mathbf{b}=\widetilde{\boldsymbol{\beta}}} - \frac{\partial^2}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}g_{\mathbf{b}}(t_i)\Big|_{\mathbf{b}=\boldsymbol{\beta}_0}.$$

As in Lemma 1 in Bianco and Boente (2002), we have that $\widehat{\mathbf{C}}_n^{(1)} + \widehat{\mathbf{C}}_n^{(2)} \xrightarrow{p} \mathbf{A}_0$, since $\widetilde{\widetilde{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{\beta}_0$. Using **N1**, **N2**, the consistency of $\widehat{\sigma}$, the Strong Law of Large Numbers and the fact that $\sup_{t\in\mathcal{T}}|\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t) - g(t)| \xrightarrow{p} 0$, we get that $\widehat{\mathbf{C}}_n^{(j)} \xrightarrow{p} 0,\ 3 \le j \le 6$. □

Lemma A.2. *Assume that t_1 is a random variable with distribution on a compact set $\mathcal{T}$ and that $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n + g(t_i) + \sigma\epsilon_i$ for $1 \le i \le n$, where $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$ and ϵ_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ and have symmetric distribution. Moreover, assume that **N0**–**N6** hold for $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. If in addition, **N8** holds and $\widehat{\sigma} \xrightarrow{p} \sigma$, then for any consistent solution $\widehat{\boldsymbol{\beta}}$ of (3), we have that $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N\left(\mathbf{c}, \sigma^2\mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}\right)$, with $\mathbf{A}_0 = \mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}_0 = \mathbf{B}(\boldsymbol{\beta}_0)$.*

Proof. To derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ it will be enough to show that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n) \xrightarrow{D} N\big(\mathbf{0}, \sigma^2\mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}\big)$. Let $\widehat{\boldsymbol{\beta}}$ be a solution of $H_n^{(1)}(\mathbf{b}) = 0$ defined in (3) and denote by $\boldsymbol{\zeta}_i(\mathbf{b}) = \mathbf{x}_i + (\partial\widehat{g}_{\mathbf{b}}(t_i)/\partial\mathbf{b})$. Using a Taylor's expansion of order one, we get

$$0 = H_n^{(1)}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - \widehat{g}_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\boldsymbol{\zeta}_i(\boldsymbol{\beta}_n) - \frac{1}{\widehat{\sigma}}\widehat{\mathbf{C}}(\widetilde{\boldsymbol{\beta}})\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n\right),$$

where $\widehat{\mathbf{C}}(\widetilde{\boldsymbol{\beta}}) = -(\widehat{\sigma}/n)\sum_{i=1}^n \delta_i\left(\partial\left\{\psi\left(\widehat{\epsilon}_i(\mathbf{b})\right)\boldsymbol{\zeta}_i(\mathbf{b})\right\}/\partial\mathbf{b}\right)|_{\mathbf{b}=\widetilde{\boldsymbol{\beta}}}\,\upsilon(\mathbf{x}_i)$, so that

$$\widehat{\mathbf{C}}(\widetilde{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^n\left(\psi'\left(\widehat{\epsilon}_i(\widetilde{\boldsymbol{\beta}})\right)\boldsymbol{\zeta}_i(\widetilde{\boldsymbol{\beta}})\boldsymbol{\zeta}_i(\widetilde{\boldsymbol{\beta}})^{\mathrm{T}} - \psi\left(\widehat{\epsilon}_i(\widetilde{\boldsymbol{\beta}})\right)\;\frac{\partial^2}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}\widehat{g}_{\mathbf{b}}(t_i)\bigg|_{\mathbf{b}=\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\right)\delta_i\upsilon(\mathbf{x}_i),$$

with $\widetilde{\boldsymbol{\beta}}$ an intermediate point between $\boldsymbol{\beta}_n$ and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\epsilon}_i(\mathbf{b}) = (y_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{b} - \widehat{g}_{\mathbf{b}}(t_i))/\widehat{\sigma}$. Using Lemma A.1, we have that $\widehat{\mathbf{C}}(\widetilde{\boldsymbol{\beta}}) \xrightarrow{p} \mathbf{A}_0$. Therefore, in order to obtain the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ it will be enough to derive the asymptotic behavior of

$$\widehat{\mathbf{L}}_n = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - \widehat{g}_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\boldsymbol{\zeta}_i(\boldsymbol{\beta}_n).$$

Using that $g_{\boldsymbol{\beta}_n} = g$, so that $y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - g_{\boldsymbol{\beta}_n}(t_i) = \epsilon_i\sigma$, we get that

$$\begin{aligned}\mathbf{L}_n &= n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - g_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\mathbf{z}_i(\boldsymbol{\beta}_n) = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\mathbf{z}_i(\boldsymbol{\beta}_n)\\ &= n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\mathbf{z}_i(\boldsymbol{\beta}_0) + n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\frac{\partial^2 g_{\mathbf{b}}(t_i)}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}\bigg|_{\mathbf{b}=\boldsymbol{\beta}^\star}\frac{\mathbf{c}}{\sqrt{n}},\end{aligned}$$

where $\boldsymbol{\beta}^\star$ is an intermediate point between $\boldsymbol{\beta}_n$ and $\boldsymbol{\beta}_0$, so that $\mathbf{L}_n = \mathbf{L}_n^{(1)} + \mathbf{L}_n^{(2)} + \mathbf{L}_n^{(3)}$ with

$$\mathbf{L}_n^{(1)} = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\mathbf{z}_i(\boldsymbol{\beta}_0), \qquad \mathbf{L}_n^{(2)} = n^{-1}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\,\frac{\partial^2 g_{\mathbf{b}}(t_i)}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}|_{\mathbf{b}=\boldsymbol{\beta}_0}\mathbf{c}$$

$$\mathbf{L}_n^{(3)} = n^{-1}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\left(\frac{\partial^2 g_{\mathbf{b}}(t_i)}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}|_{\mathbf{b}=\boldsymbol{\beta}^\star} - \frac{\partial^2 g_{\mathbf{b}}(t_i)}{\partial\mathbf{b}\partial\mathbf{b}^{\mathrm{T}}}|_{\mathbf{b}=\boldsymbol{\beta}_0}\right)\mathbf{c}.$$

The fact that ψ is odd and the errors have a symmetric distribution and are independent of the carriers implies that $\mathbb{E}\left[\psi(\epsilon_i\sigma/s)\,|(\mathbf{x}_i, t_i)\right] = \mathbb{E}\psi(\epsilon_i\sigma/s) = 0$, for all $s > 0$. Then, the consistency of $\widehat{\sigma}$ and standard tightness arguments entail that $\mathbf{L}_n^{(1)}$ is asymptotically normally distributed with covariance matrix $\mathbf{B}$. Besides, $\mathbf{L}_n^{(2)} \xrightarrow{p} 0$ since $\mathbb{E}\left[\psi(\epsilon_i\sigma/s)\,|(\mathbf{x}_i, t_i)\right] = \mathbb{E}\psi(\epsilon_i\sigma/s) = 0$, for all $s > 0$ and $\mathbf{L}_n^{(3)} \xrightarrow{p} 0$ using **N1** and **N2** and the fact that $\boldsymbol{\beta}^\star \to \boldsymbol{\beta}_0$.

Therefore, it remains to show that $\mathbf{L}_n - \widehat{\mathbf{L}}_n \xrightarrow{p} 0$. We have the following expansion $\widehat{\mathbf{L}}_n - \mathbf{L}_n = -\widehat{\sigma}^{-2}\mathbf{L}_{n,1} + \widehat{\sigma}^{-1}\mathbf{L}_{n,2} - \widehat{\sigma}^{-1}\mathbf{L}_{n,3} + \widehat{\sigma}^{-2}\mathbf{L}_{n,4}$, with

$$\mathbf{L}_{n,1} = n^{-1/2}\widehat{\sigma}\sum_{i=1}^n \delta_i\psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - g_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\mathbf{z}_i(\boldsymbol{\beta}_n)\upsilon(\mathbf{x}_i)\widehat{\gamma}_n(t_i)$$

$$\mathbf{L}_{n,2} = n^{-1/2}\widehat{\sigma}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - g_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\widehat{\mathbf{v}}_n(t_i)$$

$$\mathbf{L}_{n,3} = n^{-1}\sum_{i=1}^n \delta_i\psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - g_{\boldsymbol{\beta}_n}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\;\left(n^{1/4}\widehat{\mathbf{v}}_n(t_i)\right)\;\left(n^{1/4}\widehat{\gamma}_n(t_i)\right)$$

$$\mathbf{L}_{n,4} = (2n)^{-1}\sum_{i=1}^n \delta_i\psi''\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_n - \xi_i(t_i)}{\widehat{\sigma}}\right)\mathbf{z}_i(\boldsymbol{\beta}_n)\upsilon(\mathbf{x}_i)\left(n^{1/4}\widehat{\gamma}_n(t_i)\right)^2,$$

where $\widehat{\gamma}_n(\tau) = \widehat{g}_{\boldsymbol{\beta}_n}(\tau) - g_{\boldsymbol{\beta}_n}(\tau)$, $\widehat{\mathbf{v}}_n(\tau) = \left(\widehat{v}_{1,n}(\tau), \ldots, \widehat{v}_{p,n}(\tau)\right)^{\mathrm{T}} = \partial \widehat{\gamma}(\mathbf{b}, \tau)/\partial \mathbf{b}|_{\mathbf{b}=\boldsymbol{\beta}_n}$ is defined in (5), $\widehat{\gamma}$ is defined in (4) and $\xi(t_i)$ an intermediate point between $\widehat{g}_{\boldsymbol{\beta}_n}(t_i)$ and $g_{\boldsymbol{\beta}_n}(t_i)$. It is easy to see that **N8** and **N2** entail that $\mathbf{L}_{n,3} \xrightarrow{p} 0$ and $\mathbf{L}_{n,4} \xrightarrow{p} 0$.

To complete the proof, it remains to show that $\mathbf{L}_{n,j} \xrightarrow{p} 0$ for $j = 1, 2$ which will follow from **N8**(b)–(d) and the fact that **N6** implies that $\mathbb{E}\left[\left(\mathbf{x} + \{\partial g_{\mathbf{b}}(\tau)/\partial \mathbf{b}\}\Big|_{\mathbf{b}=\boldsymbol{\beta}_n}\right) \upsilon(\mathbf{x}) p(\mathbf{x}, \tau) | t = \tau\right] = 0$ (see (6)), using similar arguments to those considered in Bianco et al. (2011). □

Appendix B. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.spl.2014.11.004.

References

Bhattacharya, P.K., Zhao, P.L., 1997. Semiparametric inference in a partial linear model. Ann. Statist. 25, 244–262.
Bianco, A., Boente, G., 2002. On the asymptotic behavior of one-step estimates in heteroscedastic regression models. Statist. Probab. Lett. 60, 33–47.
Bianco, A., Boente, G., 2004. Robust estimators in semiparametric partly linear regression models. J. Statist. Plann. Inference 122, 229–252.
Bianco, A., Boente, G., González-Manteiga, W., Pérez-González, A., 2010. Estimation of the marginal location under a partially linear model with missing responses. Comput. Statist. Data Anal. 54, 546–564.
Bianco, A., Boente, G., González-Manteiga, W., Pérez-González, A., 2011. Asymptotic behavior of robust estimators in partially linear models with missing responses: The effect of estimating the missing probability on the simplified marginal estimators. TEST 20, 524–548.
Bianco, A., Boente, G., Martínez, E., 2006. Robust tests in semiparametric partly linear models. Scand. J. Statist. 33, 435–450.
Boente, G., González-Manteiga, W., Pérez-González, A., 2009. Robust nonparametric estimation with missing data. J. Statist. Plann. Inference 139, 571–592.
Gao, J., 1997. Adaptive parametric test in a semiparametric regression model. Comm. Statist. Theory Methods 26, 787–800.
González Manteiga, W., Aneiros Pérez, G., 2003. Testing in partial linear regression models with dependent errors. J. Nonparametr. Stat. 15, 93–111.
González-Manteiga, W., Crujeiras, R., 2013. An updated review of goodness-of-fit tests for regression models. TEST 22, 361–447.
Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer-Verlag.
Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and Semiparametric Models. Springer.
He, X., Zhu, Z., Fung, W., 2002. Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. Biometrika 89, 579–590.
Koul, H.L., Müller, U., Schick, A., 2012. The transfer principle: a tool for complete case analysis. Ann. Statist. 40, 3031–3049.
Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics: Theory and Methods. Wiley, New York.
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.
Speckman, P., 1988. Kernel smoothing in partial linear models. J. R. Stat. Soc. Ser. B 50, 413–436.
Sperlich, S., 2014. On the choice of regularization parameters in specification testing: a critical discussion. Empir. Econom. 47, 427–450.
Wang, Q., Linton, O., Härdle, W., 2004. Semiparametric regression analysis with missing response at random. J. Amer. Statist. Assoc. 99 (466), 334–345.
Wang, Q., Sun, Z., 2007. Estimation in partially linear models with missing responses at random. J. Multivariate Anal. 98, 1470–1493.