Robust tests in semiparametric partly linear models*

Ana Bianco

Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Graciela Boente

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

Elena Martínez

Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Abstract

This paper focuses on the problem of testing the null hypothesis $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_o$ and $H_{0g} : g = g_o$, under a semiparametric partly linear regression model $y_i = \mathbf{x}_i'\boldsymbol{\beta} + g(t_i) + \epsilon_i$, $1 \le i \le n$ by using a three–step robust estimate for the regression parameter and the regression function. Two families of tests statistics are considered for $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_o$ and their asymptotic distributions are studied under the null hypothesis and under contiguous alternatives. A statistic is introduced to test the nonparametric component which turns out to behave more resistantly than the classical one. A Monte Carlo study is performed to compare the finite sample behavior of the proposed tests with the classical one.

Key words: hypothesis testing, partly linear models, robust estimation, smoothing techniques.

*This research was partially supported by Grants PICT # 03-00000-006277 from ANPCYT, X-094 from the Universidad de Buenos Aires, PIP 5505 from CONICET and a Grant from the Fundación Antorchas at Buenos Aires, Argentina.

1 Introduction

Let us assume that $(y_i, \mathbf{x}_i', t_i)'$ are independent observations that follow a partly linear regression model given by $y_i = \boldsymbol{\beta}'\mathbf{x}_i + g(t_i) + \epsilon_i$, $1 \le i \le n$, where $y_i \in I\!R$, $t_i \in I\!R$, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})' \in I\!R^p$ and the errors ϵ_i are independent and independent of $(\mathbf{x}_i', t_i)'$. As in Speckman (1988), Linton (1995), He *et al.* (2002) and González Manteiga & Aneiros Pérez (2003) we will assume that the covariates $(\mathbf{x}_i', t_i)'$ are nonparametrically related satisfying $x_{ij} = \phi_j(t_i) + z_{ij}$, $1 \le i \le n$, $1 \le j \le p$, where the errors z_{ij} are independent and independent of t_i.

Thus, the model that will be considered in this paper can be written as

$$\begin{cases} y_i &= \boldsymbol{\beta}'\mathbf{x}_i + g(t_i) + \epsilon_i \qquad 1 \le i \le n\,, \\ x_{ij} &= \phi_j(t_i) + z_{ij} \qquad 1 \le i \le n\,, \quad 1 \le j \le p\,, \end{cases} \tag{1}$$

where the errors ϵ_i are independent and independent of $(\mathbf{x}_i', t_i)'$ and the errors z_{ij} are independent and independent of t_i. We will assume that g and ϕ_j are smooth functions.

This model is a flexible generalization of the linear model since it includes a nonparametric component. Model (1) can be a suitable choice when one suspects that the response y depends linearly on $\mathbf{x}$, but that it is nonparametrically related to t. The components of $\boldsymbol{\beta}$ may have, for instance, interesting meaning and in that case, tests on the regression parameter may be of particular interest.

Several authors have studied model (1). See, for instance, Denby (1986), Robinson (1988), Green & Silverman (1995), Speckman (1988) who investigated some asymptotic results using smoothing splines or kernel techniques. In particular, Robinson (1988) explained why estimates of the regression parameter based on incorrect parametrization of the function g are generally inconsistent and proposed a least square estimator of $\boldsymbol{\beta}$ which will be root–n consistent by inserting nonparametric regression estimators in the nonlinear orthogonal projection on t. Estimates based on kernel weights were also considered by Severini & Wong (1992) for the independent setting. An extensive description of the different results obtained in partly linear regression models can be found in Härdle *et al.* (2000). A more general model for longitudinal data, is studied in Sun & Wu (2005), who considered a time–varying coefficient regression model. More precisely, their model includes another covariate w_i that appears multiplying the non–parametric regression function g. Furthermore, all random variables involved in the model are time–depending and observed on a compact interval time. Sun & Wu (2005) provide a kernel–based weighted least squares approach to the problem.

In the context of hypothesis testing, Gao (1997) established a large sample theory for testing $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \mathbf{0}$ in model (1) and, in addition to this, Härdle *et al.* (2000) tested $H_{0g} : g = g_0$ too. Recently, González Manteiga & Aneiros Pérez (2003) studied the case of dependent errors.

It is well known that, both in linear regression and in nonparametric regression, least squares estimators can be seriously affected by anomalous data. Brillinger, who discusses Stone's paper (1977) pointed out that $M-$estimates of the conditional expectation were desirable in order to achieve robustness against outliers, since the usual estimates, being a weighted average of the response variables, are very sensitive to large fluctuations of them, in particular when the independent variables t_i are close to the point t at which the regression function is to be estimated. This behavior was also described in Boente & Fraiman (1991a) where a review of some of the results obtained for $M-$smoothers can be found for the independent setting and for nonparametric time series (see also Robinson, 1984). As mentioned by Härdle (1990) *"From a data–analytic viewpoint, a nonrobust behavior of the smoother is sometimes undesirable.$\cdots$ Any erratic behavior of the non-parametric pilot estimate will cause biased parametric formulations"*. Robust estimates in a nonparametric setting can thus be defined as insensitive to a single wild spike outlier. In this sense, Hampel's comment on Stone (1977) paper is highlighting. In a smooth framework, as it is the case of the partly linear model we are considering, Hampel notes that *"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"*. For the regression model, Carroll & Ruppert (1988) described this idea as follows: *"Robust estimators can handle both data and model inadequacies. They will downweight and, in some cases, completely reject grossly erroneous data. In many situations, a simple model, will adequately fit all but a few unusual observations"*. The same statement holds for partly linear models, where large values of the response variable y_i can cause a peak on the estimates of the smooth function g in the neighborhood of t_i. Moreover, large values of the response variable y_i combined with high leverage points $\mathbf{x}_i$ produce also, as in linear regression, breakdown of the classical estimates of the regression parameter $\boldsymbol{\beta}$. To overcome this problem, Bianco & Boente (2004) considered a kernel–based three–step procedure to define robust estimates under the partly linear model (1). A different strategy was suggested by Bhattacharya & Zhao (1997), who defined a $\sqrt{n}-$consistent estimator of $\boldsymbol{\beta}$ when $p = 1$ and the carriers $\mathbf{x}$ lie in a compact set by a bandwidth–matched $M-$estimation procedure. Their estimators are based on differences of the observations with kernel weights and thus, Fisher–consistency is automatically ensured. When considering unbounded carriers a weight function depending on $\mathbf{x}_i - \mathbf{x}_j$, $i \neq j$, needs to be included to deal with high leverage points in the carriers $\mathbf{x}$. Another possibility could be to define bandwidth–matched $S-$estimators, for instance.

Spline–based estimators are an alternative to kernel methods. In particular, in partly linear models with longitudinal data, He *et al.* (2002) introduced $M-$estimators to estimate the regression parameter $\boldsymbol{\beta}$ and the spline coefficients. A weighted version of this procedure can also be defined to protect against outliers in the covariates $\mathbf{x}$. When the dimension of the covariates $\mathbf{x}$ is high, a different approach should be taken to guarantee a better breakdown point. An alternative is to consider a high–breakdown point regression procedure, such as S or $MM-$estimators, to estimate the regression parameter $\boldsymbol{\beta}$ and the spline coefficients. However, the study of the asymptotic properties of these new classes of estimators and of the test statistics derived from them deserve further research and is not investigated here.

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 based on the three–step robust procedure proposed by Bianco & Boente (2004). In Section 2, we remind the definition of the three–step robust estimates and their asymptotic properties. The test statistics for the regression parameter are introduced in Section 3, where their asymptotic behavior under the null hypothesis and contiguous alternatives is studied. Besides, in Section 4, we present a robust alternative to test hypothesis concerning the regression function g. In Section 5, we present the results of a Monte Carlo study and in Section 6, an application to a real data set. Finally, in Section 7 we briefly discuss a test for the nonparametric component and we give some final conclusions. Proofs are given in the Appendix.

2 The robust estimators

Let $(Y, \mathbf{X}', T)'$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i', t_i)'$, that is

$$Y = \boldsymbol{\beta}'\mathbf{X} + g(T) + \epsilon \quad \text{and} \quad X_j = \phi_j(T) + Z_j \;, \tag{2}$$

where ϵ has distribution $F(\cdot/\sigma_\epsilon)$ and is independent of $(\mathbf{X}', T)'$, with $\mathbf{X} = (X_1, \ldots, X_p)'$. The parameter σ_ϵ denotes a scale parameter for the errors which does not need to be equal to the square root of the variance, since we will not assume the existence of second moments as in the classical approach, where it is also assumed that $E(\epsilon) = 0$, $E(\mathbf{Z}) = \mathbf{0}$ and $E(\|\mathbf{Z}\|^2) < \infty$, where $\mathbf{Z} = (Z_1, \ldots, Z_p)'$. Model (2) states a structure on the regression variables that avoids the non–identifiability of the model (see Chen (1988) and Robinson (1988) for a discussion).

In the classical approach, $\phi_j(t) = E(X_j|T = t)$ and, thus, $g(t) = \phi_o(t) - \boldsymbol{\beta}'\boldsymbol{\phi}(t)$ where $\phi_o(t) = E(Y|T = t)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))'$. Hence, $Y - \phi_o(t) = \boldsymbol{\beta}'(\mathbf{X} - \boldsymbol{\phi}(t)) + \epsilon$, which suggests, as noted by Robinson (1988), that estimators of $\phi_o(t)$ and $\boldsymbol{\phi}(t)$, $\widehat{\phi}_o(t)$ and $\widehat{\boldsymbol{\phi}}(t)$, can be inserted prior to the estimation of the regression parameter to solve the problem under non–orthogonality. As mentioned by Chen & Shiau (1994), the least squares procedure proposed independently by Denby (1986) and Speckman (1988), can be related to the partial regression procedure in linear regression. As mentioned in the Introduction, the least squares estimators, used at each step, can be seriously affected by a small fraction of outliers, as in the purely parametric and nonparametric models. If the errors ϵ and Z_j have a symmetric distribution, $\phi_o(t)$ and $\boldsymbol{\phi}(t)$ can also be thought as robust conditional location functionals such as the conditional median, satisfying $\phi_o(t) = \boldsymbol{\beta}'\boldsymbol{\phi}(t) + g(t)$. So, it may be preferable to estimate these nonparametric regression functions through any robust smoothing and the regression parameter by a robust regression estimator. For a discussion regarding the choice of the score function leading to the conditional location functionals, see He *et al.* (2002).

Putting these ideas together, Bianco & Boente (2004) introduced a three–step robust procedure which can be described as follows:

- **Step 1:** Estimate $\phi_j(t)$, $0 \le j \le p$, through a robust smoothing, as local medians or local M–type estimates with kernel weights with bandwidth parameter b. Denote $\widehat{\phi}_j(t)$, $0 \leq j \leq p$, the obtained estimates and $\widehat{\boldsymbol{\phi}}(t) = \left(\widehat{\phi}_1(t), \ldots, \widehat{\phi}_p(t)\right)'$.

- **Step 2:** Estimate the regression parameter by applying any robust regression procedure to the residuals $y_i - \widehat{\phi}_o(t_i)$ and $\mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t_i)$. Let $\widehat{\boldsymbol{\beta}}$ denote the obtained estimator.

- **Step 3:** Define the estimate of the regression function g as $\widehat{g}(t, \widehat{\boldsymbol{\beta}}) = \widehat{\phi}_o(t) - \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\phi}}(t)$.

In **Step 3**, an alternative estimator of the regression function g can be obtained by robustly smoothing the residuals $y_i - \widehat{\boldsymbol{\beta}}'\mathbf{x}_i$. This can also be done using kernel weights. However, a different smoothing parameter h than the one used in **Step 1** may be preferable, as the residuals $y_i - \widehat{\boldsymbol{\beta}}'\mathbf{x}_i$ have a smaller variability than the original variables y_i. This is the approach we will follow in Section 4, where the dependence of the estimators on the smoothing parameter will be explicited.

As described in **Step 2**, the robust estimation of the regression parameter can be performed by applying to the residuals $\widehat{r}_i = y_i - \widehat{\phi}_o(t_i)$ and $\widehat{\mathbf{z}}_i = \mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t_i)$ any of the robust methods proposed for linear regression. Bianco & Boente (2004) studied the behavior of the estimate $\widehat{\boldsymbol{\beta}}$ defined as any solution of

$$\sum_{i=1}^{n} \psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\beta}}}{s_n} \right) w_2 \left(\|\widehat{\mathbf{z}}_i\| \right) \widehat{\mathbf{z}}_i = 0, \tag{3}$$

where ψ_1 and w_2 are a score and a weight function, respectively, and s_n is a robust consistent estimate of the residuals scale. This family of estimators includes, among others, GM, S and $MM-$estimators. These authors showed that, under model (1), when ψ_1 is an odd function and the errors have a symmetric distribution, if $s_n \xrightarrow{p} \sigma_0$, $0 < \sigma_0 < \infty$, then $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically normally distributed with asymptotic covariance matrix given by $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where

$$\mathbf{A} = E\left(\psi_1'\left(\frac{\epsilon}{\sigma_0}\right)\right) E\left(w_2\left(\|\mathbf{Z}\|\right)\mathbf{Z}\,\mathbf{Z}'\right) \tag{4}$$

$$\mathbf{B} = \sigma_0^2\, E\left(\psi_1^2\left(\frac{\epsilon}{\sigma_0}\right)\right) E\left(w_2^2\left(\|\mathbf{Z}\|\right)\mathbf{Z}\,\mathbf{Z}'\right). \tag{5}$$

This result extends straighforward if the oddness of the score function and the symmetry assumption on the errors distribution are replaced by $E\left(\psi_1\left(\epsilon/\sigma\right)\right) = 0$, for any $\sigma > 0$. This last condition is the one required all over this paper to allow a more bigger family of errors distribution. In practice, the robust scale estimator is calibrated to achieve asymptotically unbiased estimators of σ_ϵ under the central model. That is, if F_n denotes the empirical distribution function of the residuals and the scale estimator can be written as $s_n = S(F_n)$, with $S(G)$ a given scale functional, under mild assumptions, we have that $s_n \xrightarrow{p} \sigma_0 = S(F(\cdot/\sigma_\epsilon))$. Usually, the practitioner calibrates the scale functional S such that, at the normal distribution, $\sigma_0 = \sigma_\epsilon$.

An alternative choice to the estimator given by (3) is to consider one–step high breakdown point regression estimates. More precisely, denoting by $\widehat{\boldsymbol{\beta}}_{\mathbf{I}}$ an initial regression estimate with high breakdown point and by $s_{\rm I} = \kappa \operatorname*{median}_{1\le i\le n} \left(|\widehat{r}_i - \widehat{\mathbf{z}}_i'\widehat{\boldsymbol\beta}_{\rm I}|\right)$ the related scale estimator with calibrating constant κ, we can define the one–step estimator as

$$\widehat{\boldsymbol\beta} = \widehat{\boldsymbol\beta}_{\rm I} + s_{\rm I} \left\{\sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widehat{\boldsymbol\beta}_{\rm I}}{s_{\rm I}}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i'\right\}^{-1} \left\{\sum_{i=1}^n \psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widehat{\boldsymbol\beta}_{\rm I}}{s_{\rm I}}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\right\} . \quad (6)$$

As in the location–scale and regression models (see, for instance, Bickel (1975) and Simpson *et al.* (1992)), the one–step estimator improves the order of convergence of the initial estimate and will have the same asymptotic behavior as the solution of (3).

3 Tests for the regression parameter

3.1 The statistics

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

$$D(\widehat{\boldsymbol\beta}, \widehat{\boldsymbol\Sigma}, H_{0\boldsymbol\beta}) = (\widehat{\boldsymbol\beta} - \boldsymbol\beta_o)' \, \widehat{\boldsymbol\Sigma}^{-1} (\widehat{\boldsymbol\beta} - \boldsymbol\beta_o) \, , \quad (7)$$

where $\widehat{\boldsymbol\Sigma}$ is an estimate of the asymptotic covariance matrix of $\widehat{\boldsymbol\beta}$. When considering the estimates defined through (3), as in Markatou & He (1994), two estimates of $\boldsymbol\Sigma$ may be considered. The first one is given by $\widehat{\boldsymbol\Sigma}_1 = \hat{\mathbf{A}}(\widehat{\boldsymbol\beta})^{-1}\hat{\mathbf{B}}(\widehat{\boldsymbol\beta})\hat{\mathbf{A}}(\widehat{\boldsymbol\beta})^{-1}$, where

$$\hat{\mathbf{A}}(\boldsymbol\beta) = \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\boldsymbol\beta}{s_n}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i' \quad (8)$$

$$\hat{\mathbf{B}}(\boldsymbol\beta) = s_n^2 \, \frac{1}{n} \sum_{i=1}^n \psi_1^2 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\boldsymbol\beta}{s_n}\right) w_2^2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i' \, , \quad (9)$$

and the second one by $\widehat{\boldsymbol\Sigma}_2 = \tilde{\mathbf{A}}(\widehat{\boldsymbol\beta})^{-1}\tilde{\mathbf{B}}(\widehat{\boldsymbol\beta})\tilde{\mathbf{A}}(\widehat{\boldsymbol\beta})^{-1}$, where

$$\tilde{\mathbf{A}}(\boldsymbol\beta) = \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\boldsymbol\beta}{s_n}\right) \frac{1}{n} \sum_{i=1}^n w_2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i' \quad (10)$$

$$\tilde{\mathbf{B}}(\boldsymbol\beta) = s_n^2 \, \frac{1}{n} \sum_{i=1}^n \psi_1^2 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\boldsymbol\beta}{s_n}\right) \frac{1}{n} \sum_{i=1}^n w_2^2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i' \, . \quad (11)$$

Note that, under general conditions, $\widehat{\gamma}(\mathbf{x}, t) = \mathbf{x}'\widehat{\boldsymbol\beta} + \widehat{g}(t, \widehat{\boldsymbol\beta})$ will be a consistent estimate for the regression function $\gamma(\mathbf{x}, t) = \mathbf{x}'\boldsymbol\beta + g(t)$. On the other hand, under the null parametric hypothesis $H_{0\boldsymbol\beta}$, the function $\gamma(\mathbf{x}, t)$ can be consistently estimated by $\widehat{\gamma}_o(\mathbf{x}, t) = \mathbf{x}'\boldsymbol\beta_o + \widehat{g}(t, \boldsymbol\beta_o)$. Therefore, we can consider the test statistic

$$S(\widehat{\gamma}, H_{0\boldsymbol\beta}) = \frac{1}{n} \sum_{i=1}^n \left(\widehat{\gamma}(\mathbf{x}_i, t_i) - \widehat{\gamma}_o(\mathbf{x}_i, t_i)\right)^2$$

that measures the difference between the null and the alternative hypothesis. When $w_2 \equiv 1$, i.e., for fixed covariates $\mathbf{x}_i$ or when we suspect that no leverage points are present, $D(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}_2, H_{0\boldsymbol{\beta}}) = c_n\, S(\widehat{\gamma}, H_{0\boldsymbol{\beta}})$, with c_n

$$c_n = s_n^2\, \frac{1}{n} \sum_{i=1}^{n} \psi_1^2 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i' \widehat{\boldsymbol{\beta}}}{s_n} \right) \left(\frac{1}{n} \sum_{i=1}^{n} \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i' \widehat{\boldsymbol{\beta}}}{s_n} \right) \right)^{-2} .$$

However, for random covariates $\mathbf{x}$, it is necessary to introduce a weight function w_2 in order to control possible leverage points.

Another possibility is to consider score–type tests, which were studied for regression models by Markatou & He (1994). Define the score

$$\mathbf{U}_n(\boldsymbol{\beta}) \quad = \quad \frac{1}{n} s_n \sum_{i=1}^{n} \psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i' \boldsymbol{\beta}}{s_n} \right) w_2 \left(\|\widehat{\mathbf{z}}_i\| \right) \widehat{\mathbf{z}}_i \ ,$$

where the function w_2 weights the influence of the predicted carriers $\widehat{\mathbf{z}}_i$, ψ_1 is a bounded score function and s_n denotes a consistent estimate of the residuals scale. When testing $H_{0\boldsymbol{\beta}}$, the score–type test statistic can be defined through the quadratic form

$$V_n(\widehat{\mathbf{C}}, H_{0\boldsymbol{\beta}}) = \mathbf{U}_n(\boldsymbol{\beta}_o)' \widehat{\mathbf{C}}^{-1} \mathbf{U}_n(\boldsymbol{\beta}_o) \ . \tag{12}$$

The matrix $\widehat{\mathbf{C}}$ denotes a consistent estimate of $\mathbf{B}$, the asymptotic covariance matrix of $\mathbf{U}_n(\boldsymbol{\beta}_o)$, and it can be chosen as the matrix $\hat{\mathbf{B}}(\widehat{\boldsymbol{\beta}})$ defined in (9) or the matrix $\tilde{\mathbf{B}}(\widehat{\boldsymbol{\beta}})$ defined in (11).

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

$$D_1(\widehat{\boldsymbol{\beta}}_{(1)}, \widehat{\boldsymbol{\Sigma}}, H_{0\boldsymbol{\beta}_{(1)}}) = (\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),o})'\, \widehat{\boldsymbol{\Sigma}}_{11}^{-1} (\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),o}) \, , \tag{13}$$

where $\widehat{\boldsymbol{\Sigma}}_{11}$ denotes the $q \times q$ submatrix of $\widehat{\boldsymbol{\Sigma}}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$. A score–type test statistic defined as

$$V_n^{(1)}(\widehat{\mathbf{C}}, H_{0\boldsymbol{\beta}_{(1)}}) = \mathbf{U}_n(\widehat{\boldsymbol{\beta}}^{(1)})'\, \widehat{\mathbf{C}}^{-1}\, \mathbf{U}_n(\widehat{\boldsymbol{\beta}}^{(1)}) \ , \tag{14}$$

can also be considered, where $\widehat{\boldsymbol{\beta}}^{(1)} = \left(\boldsymbol{\beta}_{(1),o}', \widehat{\boldsymbol{\beta}}_{(2)}' \right)'$ and $\widehat{\boldsymbol{\beta}}_{(2)}$ are the last $p - q$ coordinates of $\widehat{\boldsymbol{\beta}}$ defined in (3) or (6).

3.2 Asymptotic distribution of the test statistics

In this Section, we will state the asymptotic behavior of the test statistics based on the estimates of the regression parameter, defined through (3). In fact, combining the arguments used in Simpson *et al.* (1992) with those in Bianco & Boente (2004), it can be shown that, if $n^\tau \left(\widehat{\boldsymbol{\beta}}_{\rm I} - \boldsymbol{\beta}\right)$ is bounded in probability where $1/4 < \tau \le 1/2$, then the statistics based on the one–step estimate defined in (6) have the same behavior as those based on the solution of (3).

In order to derive the asymptotic distribution of the regression parameter estimates, Bianco & Boente (2004) required that $t_i \in [0,1]$ and assumptions **N1** to **N7** below.

N1. ψ_1 is a bounded and twice continuously differentiable function with bounded derivatives ψ_1' and ψ_1'', such that $\varphi_1(t) = t\psi_1'(t)$ and $\varphi_2(t) = t\psi_1''(t)$ are bounded.

N2. $E\left(w_2(\|\mathbf{Z}\|)\|\mathbf{Z}\|^2\right) < \infty$ and the matrix

$$\mathbf{A} = E\left(\psi_1'\left(\frac{\epsilon}{\sigma_0}\right)\right) E\left(w_2\left(\|\mathbf{Z}\|\right)\mathbf{Z}\,\mathbf{Z}'\right)$$

is non–singular.

N3. $w_2(u) = \psi_2(u)\,u^{-1} > 0$ is a bounded function, Lipschitz of order 1. Moreover, ψ_2 is also a bounded and continuously differentiable function with bounded derivative ψ_2', such that $\lambda_2(t) = t\psi_2'(t)$ is bounded.

N4. $E\left(w_2(\|\mathbf{Z}\|)\mathbf{Z}\right) = 0$.

N5. The functions $\phi_j(t)$, $0 \le j \le p$, are continuous with first derivative $\phi_j'(t)$ continuous in $[0,1]$, with $\phi_o(t) = \boldsymbol{\beta}'\boldsymbol{\phi}(t) + g(t)$.

N6. $\widehat{\phi}_j(t)$, $1 \le j \le p$, are such that $\widehat{\phi}_j(t)$ has first continuous derivative and

$$n^{\frac{1}{4}} \sup_{t\in[0,1]} |\widehat{\phi}_j(t) - \phi_j(t)| \xrightarrow{p} 0, \quad 1 \le j \le p \tag{15}$$

$$\sup_{t\in[0,1]} |\widehat{\phi}_j'(t) - \phi_j'(t)| \xrightarrow{p} 0, \quad 1 \le j \le p\,. \tag{16}$$

N7. $\widehat{\phi}_o(t)$ has first continuous derivative and

$$n^{\frac{1}{4}} \sup_{t\in[0,1]} |\widehat{\phi}_o(t) - \phi_o(t)| \xrightarrow{p} 0 \tag{17}$$

$$\sup_{t\in[0,1]} |\widehat{\phi}_o'(t) - \phi_o'(t)| \xrightarrow{p} 0\,, \tag{18}$$

with $\phi_o(t) = \boldsymbol{\beta}'\boldsymbol{\phi}(t) + g(t)$ when model (1) holds.

In order to study the asymptotic behavior of the test statistics under contiguous alternatives we will also require the following assumption.

N8. $\widehat{\phi}_o(t)$ has first continuous derivative and

$$n^{\frac{1}{4}} \sup_{t\in[0,1]} |\widehat{\phi}_o(t) - \phi_{o,n}(t)| \xrightarrow{p} 0 \tag{19}$$

$$\sup_{t\in[0,1]} |\widehat{\phi}_o'(t) - \phi_{o,n}'(t)| \xrightarrow{p} 0\,, \tag{20}$$

with $\phi_{o,n}(t) = \boldsymbol{\beta}_n'\boldsymbol{\phi}(t) + g(t)$ when model (1) holds for $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}$.

In the next Theorems we derive the asymptotic distribution of the Wald and score–type statistics under the null hypothesis and under a sequence of contiguous alternatives.

Theorem 1
*Let $(y_i, \mathbf{x}_i', t_i)'$, $1 \le i \le n$ be independent random vectors satisfying (1), where ϵ_i are independent of $(\mathbf{x}_i', t_i)'$ such that $E(\psi_1(\epsilon/\sigma)) = 0$, for any $\sigma > 0$. Assume that t_i are random variables with distribution on $[0, 1]$. Denote by $\widehat{\mathbf{\Sigma}}$ any consistent estimate of $\mathbf{\Sigma}$. Then, if $s_n \xrightarrow{p} \sigma_0$, $\widehat{\boldsymbol{\beta}}$ is a consistent estimate of the regression parameter and **N1** to **N6** hold, we have that*

i) *under $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_o$, $\mathcal{W}_n = n\, D(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{\Sigma}}, H_{0\boldsymbol{\beta}}) \xrightarrow{D} \chi^2_p$, if **N7** holds,*

ii) *under $H_{1\boldsymbol{\beta}} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_o$, $\mathcal{W}_n \xrightarrow{p} \infty$, for any fixed $\boldsymbol{\beta}$, if **N7** holds,*

iii) *under $H_{1\boldsymbol{\beta}(\mathbf{c})} : \boldsymbol{\beta} = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$, $\mathcal{W}_n \xrightarrow{D} \chi^2_p(\theta)$ where $\theta = \mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c}$ if **N8** holds, for any $\mathbf{c} \in I\!R^p$.*

Lemma 1 in the Appendix shows that $\widehat{\mathbf{\Sigma}}_1$ or $\widehat{\mathbf{\Sigma}}_2$ are suitable choices for $\widehat{\mathbf{\Sigma}}$.

Theorem 2
*Let $(y_i, \mathbf{x}_i', t_i)'$, $1 \le i \le n$ be independent random vectors satisfying (1), where ϵ_i are independent of $(\mathbf{x}_i', t_i)'$ such that $E(\psi_1(\epsilon/\sigma)) = 0$, for any $\sigma > 0$. Assume that t_i are random variables with distribution on $[0, 1]$ and that ψ_1 is an increasing function. Then, if $s_n \xrightarrow{p} \sigma_0$, $\widehat{\mathbf{C}} \xrightarrow{p} \mathbf{B}$, $\widehat{\boldsymbol{\beta}}$ is a consistent estimate of the regression parameter and **N1** to **N6** hold, we have that*

i) *under $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \boldsymbol{\beta}_o$, $\mathcal{S}_n = n\, V_n(\widehat{\mathbf{C}}, H_{0\boldsymbol{\beta}}) \xrightarrow{D} \chi^2_p$, if **N7** holds,*

ii) *under $H_{1\boldsymbol{\beta}} : \boldsymbol{\beta} \neq \boldsymbol{\beta}_o$, $\mathcal{S}_n \xrightarrow{p} \infty$, for any fixed $\boldsymbol{\beta}$, if **N7** holds,*

iii) *under $H_{1\boldsymbol{\beta}(\mathbf{c})} : \boldsymbol{\beta} = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$, $\mathcal{S}_n \xrightarrow{D} \chi^2_p(\theta)$ where $\theta = \mathbf{c}'\mathbf{\Sigma}^{-1}\mathbf{c}$ if **N8** holds, for any $\mathbf{c} \in I\!R^p$.*

Remarks

1. When considering local M–smoothers in **Step 1** of the estimation procedure, following analogous arguments to those used in Boente & Fraiman (1991b), it can be shown that (15) and (17) hold under regularity conditions on the kernel for the optimal bandwidth. On the other hand, (16) and (18) can also be derived using similar arguments to those considered by Boente *et al.* (1997) in Proposition 2.1, for the fixed design setting. Assumption **N8** holds for local M–smoothers, for instance, if ϵ has a bounded density f since, in this case, $v(\boldsymbol{\beta}) = \epsilon + \mathbf{Z}'\boldsymbol{\beta}$ has a density majorized by $\|f\|_\infty$, independently of the value of $\boldsymbol{\beta}$. This entails that Assumption 3 (ii) and (iii) in Boente & Fraiman (1991b) hold uniformly in $\boldsymbol{\beta}$ and thus, **N8** can be derived using similar arguments to those considered therein.

2. It is worthwhile noticing that the condition $E\left(\psi_1\left(\epsilon/\sigma\right)\right)=0$, for any $\sigma>0$, is a common condition in robustness in order to guarantee Fisher–consistency of the regression parameter and is fulfilled for instance, when ψ_1 is an odd function and the errors have a symmetric distribution.

3. Note that, under the conditions of Theorem 1 (iii), similar arguments to those used in Lemma 1 and Theorem 1 in Bianco & Boente (2004) entail that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_o$, when $\boldsymbol{\beta}=\boldsymbol{\beta}_n=\boldsymbol{\beta}_o+\mathbf{c}n^{-1/2}$.

4. From Theorems 1 and 2, to test $H_{0\boldsymbol{\beta}}$ at a given significance level α, two possible consistent tests can be given:

   - the Wald–test which rejects $H_{0\boldsymbol{\beta}}$ if $\mathcal{W}_n>\chi^2_{p,\alpha}$,

   and

   - the score test that rejects $H_{0\boldsymbol{\beta}}$ when $\mathcal{S}_n>\chi^2_{p,\alpha}$.

Note also that, as mentioned in Section 3.1, the matrix $\widehat{\mathbf{C}}$ can be chosen as the matrix $\hat{\mathbf{B}}(\widehat{\boldsymbol{\beta}})$ defined in (9) or the matrix $\tilde{\mathbf{B}}(\widehat{\boldsymbol{\beta}})$ defined in (11). Their weak consistency, which is necessary for the results stated in Theorem 2, is derived in Lemma 1 in the Appendix.

Equivalent results to those given in the previous Theorems can be obtained when the null hypothesis involves only a subset of q parameters. In Theorem 3, we state the asymptotic distribution of the Wald–type statistic. Its proof is similar to that of Theorem 1. A similar result holds for the score–type statistic with increasing score function.

Theorem 3
Let $(y_i,\mathbf{x}_i',t_i)'$, $1\le i\le n$ be independent random vectors satisfying (1), where ϵ_i are independent of $(\mathbf{x}_i',t_i)'$ such that $E\left(\psi_1\left(\epsilon/\sigma\right)\right)=0$, for any $\sigma>0$. Assume that t_i are random variables with distribution on $[0,1]$. Denote by $\widehat{\boldsymbol{\Sigma}}$, the matrix $\widehat{\boldsymbol{\Sigma}}_1$ or $\widehat{\boldsymbol{\Sigma}}_2$. Then, if $s_n \xrightarrow{p} \sigma_0$, $\widehat{\boldsymbol{\beta}}$ is a consistent estimate of the regression parameter and ***N1*** *to* ***N6*** *hold, we have that*

i) under $H_{0\boldsymbol{\beta}_{(1)}}:\boldsymbol{\beta}_{(1)}=\boldsymbol{\beta}_{(1),o}$, $\mathcal{W}_{1,n}=n\,D_1(\widehat{\boldsymbol{\beta}}_{(1)},\widehat{\boldsymbol{\Sigma}},H_{0\boldsymbol{\beta}_{(1)}}) \xrightarrow{D} \chi^2_q$, if ***N7*** *holds,*

ii) under $H_{1\boldsymbol{\beta}_{(1)}}:\boldsymbol{\beta}_{(1)}\neq\boldsymbol{\beta}_{(1),o}$, $\mathcal{W}_{1,n} \xrightarrow{p} \infty$, if ***N7*** *holds,*

iii) under $H_{1\boldsymbol{\beta}_{(1)}(\mathbf{c}_{(1)})}:\boldsymbol{\beta}_{(1)}=\boldsymbol{\beta}_{(1),o}+\mathbf{c}_{(1)}n^{-1/2}$, $\mathcal{W}_{1,n} \xrightarrow{D} \chi^2_q(\theta_1)$ where $\theta_1=\mathbf{c}_{(1)}'\boldsymbol{\Sigma}_{11}^{-1}\mathbf{c}_{(1)}$ if ***N8*** *holds, for any $\mathbf{c}_{(1)}\in I\!R^q$.*

4 Tests for the regression function

Under model (1), the regression function $\gamma(\mathbf{x},t)$ equals $\boldsymbol{\beta}'\mathbf{x}+g(t)$. In this Section, we focus on testing the nonparametric component of γ, i.e., $H_{0g}:g=g_o$.

4.1 The test Statistic

To make explicit the dependence on the smoothing parameters, in this Section we will denote $\widehat{\boldsymbol{\beta}}(b)$ the estimator obtained in **Step 2**, while $\widehat{g}(t, \widehat{\boldsymbol{\beta}}(b), b)$ denotes the estimate defined in **Step 3** as $\widehat{g}(t, \widehat{\boldsymbol{\beta}}(b), b) = \widehat{\phi}_o(t) - \widehat{\boldsymbol{\phi}}(t)'\widehat{\boldsymbol{\beta}}(b)$.

As mentioned in Section 2, in **Step 3** an alternative estimator of the regression function g can be obtained by robustly smoothing the residuals $y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}(b)$ with a different smoothing parameter than the one used in **Step 1**, since the residuals $y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}(b)$ may have a smaller variability than the original variables y_i. When we use h as smoothing parameter, we denote the estimate $\widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)$, i.e., $\widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)$ solves

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi\left(\frac{y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}(b) - \widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)}{\widehat{\sigma}}\right) = 0\,, \tag{21}$$

where $\widehat{\sigma}$ is an estimate of the error's scale and ψ is a bounded differentiable score function.

Under the null nonparametric hypothesis H_{0g}, $\widehat{\gamma}_o^*(\mathbf{x}, t) = \mathbf{x}'\widehat{\boldsymbol{\beta}}(b) + g_o(t)$ is a consistent estimate of the regression function $\gamma(\mathbf{x}, t)$. Thus, since $\widehat{\gamma}_b(\mathbf{x}, t) = \mathbf{x}'\widehat{\boldsymbol{\beta}}(b) + \widehat{g}(t, \widehat{\boldsymbol{\beta}}(b), b)$ and $\widehat{\widehat{\gamma}}_h(\mathbf{x}, t) = \mathbf{x}'\widehat{\boldsymbol{\beta}}(b) + \widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)$ are consistent estimates of $\gamma(\mathbf{x}, t)$, natural test statistics for H_{0g} are

$$S(\widehat{\gamma}, H_{0g}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\gamma}_b(\mathbf{x}_i, t_i) - \widehat{\gamma}_o^*(\mathbf{x}_i, t_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{g}(t_i, \widehat{\boldsymbol{\beta}}(b), b) - g_o(t_i)\right)^2\,, \tag{22}$$

$$S(\widehat{\widehat{\gamma}}, H_{0g}) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\widehat{\gamma}}_h(\mathbf{x}_i, t_i) - \widehat{\gamma}_o^*(\mathbf{x}_i, t_i)\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\widehat{g}}(t_i, \widehat{\boldsymbol{\beta}}(b), h) - g_o(t_i)\right)^2\,. \tag{23}$$

We will focus our attention on $S(\widehat{\widehat{\gamma}}, H_{0g})$, since as discussed in González Manteiga and Aneiros Pérez (2003) for the classical test, to achieve consistent tests two different smoothing parameters are needed.

For the sake of simplicity, we denote $\widehat{\widehat{g}}(t) = \widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)$. Under mild conditions, it can be shown that, if $\epsilon \sim F(\cdot/\sigma_\epsilon)$,

a) under H_{0g}, we have that

$$\sqrt{n^2 h}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\widehat{g}}(t_i) - g_o(t_i)\right)^2 - e_\psi\frac{\int_0^1 K^2(u)du}{n\,h}\right] \xrightarrow{\mathcal{D}} N\left(0, \sigma_S^2\right)\,, \tag{24}$$

b) under contiguous alternatives of the form $H_{1g}: g(t) = g_o(t) + (n^2h)^{-\frac{1}{4}}g^*(t)$, we have that

$$\sqrt{n^2 h}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\widehat{g}}(t_i) - g_o(t_i)\right)^2 - e_\psi\frac{\int_0^1 K^2(u)du}{n\,h}\right] \xrightarrow{\mathcal{D}} N\left(\int [g^*(u)]^2 du, \sigma_S^2\right)\,, \tag{25}$$

where $\sigma_S^2 = 2\, e_\psi^2 \int (K * K)^2$ with $e_\psi = \sigma_\epsilon^2 E\psi^2\left(\epsilon/\sigma_\epsilon\right)\left\{E\psi'\left(\epsilon/\sigma_\epsilon\right)\right\}^{-2}$. By means of the asymptotic distribution given in (24), to test H_{0g} at a given significance level α, H_{0g} is rejected if

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{g}(t_i\widehat{\boldsymbol{\beta}}(b),h) - g_o(t_i)\right)^2 > \frac{\widehat{\sigma}_S z_\alpha}{\sqrt{n^2 h}} + \hat{e}_\psi \frac{\int_0^1 K^2(u)du}{n\,h}\,,$$

where $\hat{e}_\psi$ is an estimate of e_ψ and $\widehat{\sigma}_S^2 = 2\,\hat{e}_\psi^2 \displaystyle\int (K * K)^2$.

An estimate of e_ψ can be constructed as follows. Denote

$$\widehat{\epsilon}_i = y_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}(b) - \widehat{g}(t, \widehat{\boldsymbol{\beta}}(b), b) \tag{26}$$

and let $\hat{\sigma}$ be a robust scale estimator of σ_ϵ, as, for instance, $\hat{\sigma} = \kappa \operatorname*{median}_{1\le i\le n}(|\widehat{\epsilon}_i|)$. Then, we can define

$$\hat{e}_\psi = \frac{\hat{\sigma}^2 \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left[\psi^2\left(\frac{\widehat{\epsilon}_i}{\hat{\sigma}}\right)\right]}{\left\{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left[\psi'\left(\frac{\widehat{\epsilon}_i}{\hat{\sigma}}\right)\right]\right\}^2}\,.$$

4.2 Heuristics of the asymptotic behavior of the test statistic

We will outline the heuristics of the asymptotic distribution given in (24) and (25).

Since $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ at a $\sqrt{n}$–rate, we can assume that $\boldsymbol{\beta}$ and σ_ϵ are known. Then, $\widehat{\widehat{g}}(t) = \widehat{\widehat{g}}(t, \widehat{\boldsymbol{\beta}}(b), h)$ solves

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta} - \widehat{\widehat{g}}(t)}{\sigma_\epsilon}\right) = 0\,,$$

which is equivalent to

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi\left(\frac{g(t_i) - \widehat{\widehat{g}}(t) + \epsilon_i}{\sigma_\epsilon}\right) = 0\,. \tag{27}$$

Using a first order Taylor's expansion, we have that

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) + \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi'\left(\frac{\epsilon_i}{\sigma_\epsilon}\right)\frac{(g(t_i) - \widehat{\widehat{g}}(t))}{\sigma_\epsilon} \simeq 0\,,$$

which implies

$$\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\psi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) + \frac{E\psi'\left(\dfrac{\epsilon}{\sigma_\epsilon}\right)}{\sigma_\epsilon}\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)(g(t_i) - \widehat{\widehat{g}}(t)) \simeq 0\,. \tag{28}$$

From (28) we have the following approximation

$$\widehat{\widehat{g}}(t) \simeq \frac{\sigma_\epsilon}{E\psi'\left(\dfrac{\epsilon}{\sigma_\epsilon}\right)} \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{t_i - t}{h}\right)\left\{\psi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) + \frac{g(t_i)E\psi'\left(\dfrac{\epsilon}{\sigma_\epsilon}\right)}{\sigma_\epsilon}\right\} .$$

Denote $\tilde{u}_i$ the pseudo–observations

$$\tilde{u}_i = g(t_i) + \frac{\sigma_\epsilon}{E\psi'\left(\dfrac{\epsilon}{\sigma_\epsilon}\right)}\psi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) = g(t_i) + \tilde{\epsilon}_i \,.$$

Then, as in González Manteiga and Aneiros Pérez (2003), since $\widehat{\widehat{g}}(t)$ is a kernel estimate over the pseudo–observations, under H_{0g} we have that

$$\sqrt{n^2 h}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\widehat{g}}(t_i) - g_o(t_i)\right)^2 - \frac{Var(\tilde{\epsilon})}{nh}\int_0^1 K^2(u)du\right] \xrightarrow{\mathcal{D}} N\left(0, \sigma_S^2\right) ,$$

where

$$\sigma_S^2 = 2[Var(\tilde{\epsilon})]^2 \int (K * K)^2 \,.$$

Since

$$Var(\tilde{\epsilon}) = \frac{\sigma_\epsilon^2}{\left[E\left(\psi'(\frac{\epsilon}{\sigma_\epsilon})\right)\right]^2} E\left(\psi^2\left(\frac{\epsilon}{\sigma_\epsilon}\right)\right) ,$$

we get the desired result.

Using analogous arguments, we obtain (25), since under $H_{1g} : g(t) = g_o(t) + (n^2h)^{-\frac{1}{4}} g^*(t)$, the pseudo–observations $\tilde{u}_i$ are given by

$$\tilde{u}_i = g(t_i) + \frac{\sigma_\epsilon}{E\psi'\left(\dfrac{\epsilon}{\sigma_\epsilon}\right)}\psi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) = g_o(t_i) + (n^2h)^{-\frac{1}{4}} g^*(t_i) + \tilde{\epsilon}_i \,.$$

5 Monte Carlo study

5.1 Simulation study for $H_{0\boldsymbol{\beta}}$

5.1.1 General description

A simulation study was carried out in Splus, for the case $p = 1$. The S–code is available upon request to the authors. To compare the behavior of the proposed tests with respect to the classical ones, we have considered the tests based on:

- the Wald–type statistic computed using:

a) a $GM-$estimate with Huber functions with constants $\chi^2_{1,0.025}$ on the regression variables and 1.6 on the residuals

b) a one-step estimate defined in (6) with the same score functions as in a) and the least median of squares as initial estimate

c) the least squares estimate

- the score–type statistic where the residual scale and the matrix $\mathbf{B}$ are estimated using:

a) the $GM-$estimates

b) the one-step estimates defined in (6).

In all the Tables and Figures the procedures based on least squares, GM and one–step estimates will be denoted LS, GM and OS, respectively. The Wald– statistic will be indicated as $\mathcal{W}$, while the score–type statistic as $\mathcal{S}$.

The smoothing procedure uses local M–estimates based on the bisquare score function, with tuning constant 4.685, and local medians as initial estimate. We have used the standardized gaussian kernel with several bandwidth choices, to show the sensitivity of the tests, both in level and power, with respect to bandwidth selection. The bandwidths considered were $b = 0.008, 0.02, 0.03, 0.04, 0.08$ and 0.2. The bandwidth 0.08 corresponds to a choice near the asymptotically optimal one with respect to the mean square error of the least squares estimate of β (see Linton, 1995).

We performed 5000 replications generating independent samples of size $n = 100$ following the model

$$\begin{aligned} y_i &= \beta_o x_i + \frac{\pi}{4}\sin(\pi t_i) + \epsilon_i \quad 1 \le i \le n\,, \\ x_i &= 10(t_i - 0.5)^3 + z_i \quad 1 \le i \le n\,, \end{aligned}$$

where $\beta_o = 3$, $\{z_i\}$, $\{t_i\}$ and $\{\epsilon_i\}$ are independent, $t_i \sim \mathcal{U}(0,1)$, $z_i \sim N(0,\sigma_z^2)$ and $\epsilon_i \sim N(0,\sigma_\epsilon^2)$ with $\sigma_z = 0.125$ and $\sigma_\epsilon = 0.05$ in the non–contaminated case. To isolate the comparison between the competitors from any border effect, data were in fact generated at design points outside the interval $[0,1]$ as well.

The results for normal data sets will be indicated by C_0, while C_1 to C_3 will denote the following contaminations:

- C_1: $\epsilon_1,\ldots,\epsilon_n$ are i.i.d. distributed as 0.9 $N(0,\sigma_\epsilon^2)$ + 0.1 $\mathcal{C}(0,\sigma_\epsilon)$, where $\mathcal{C}(0,\sigma_\epsilon)$ denotes the Cauchy distribution centered at 0 with scale σ_ϵ. This contamination inflates the errors and will mainly affect the variance of the regression estimates and hence, both level and power.

- C_2: $\epsilon_1,\ldots,\epsilon_n$ are i.i.d. with distribution 0.9 $N(0,\sigma_\epsilon^2) + 0.1\ N(0,25\sigma_\epsilon^2)$. Artificially 10 observations of the carriers, but not of the response variables, were modified to be equal to 20 at equally spaced values of t. This case corresponds to introduce high–leverage points. The aim of this contamination is to see how the bias of the regression parameter estimates affects the level of the test.

- C_3: $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with distribution $0.9\ N(0, \sigma_\epsilon^2) + 0.1\ N(0, 25\sigma_\epsilon^2)$. Artificially one observation was modified, both in the carrier and in the response variable, and set equal to $(x_0, y_0) = (1.2, 4.1)$ at $t = 0.5$. The aim of this contamination is to breakdown power at the alternative $\beta = \beta_o + \Delta n^{-1/2}$ with $\Delta = 2.4$.

5.1.2 Results and comments

Tables 1 to 3 summarize the results of the simulations. In Table 1, we present, for normal errors, i.e., the non–contaminated case C_0, the observed frequencies of rejection under the null hypothesis for two different sample sizes and their corresponding optimal bandwidths. It is worth noticing that with $n = 100$ the observed frequencies are higher than the nominal values, but with $n = 500$ they are near the actual levels, in particular for $\alpha = 0.05$. This shows the slow rate of convergence to the asymptotic distribution of the tests statistics, perhaps due to the smoothing procedure. Note that $n = 500$ was the sample size considered by González Manteiga & Aneiros Pérez (2003) in their simulation study. For the remaining of this Monte Carlo study we considered samples of size $n = 100$ and nominal level $\alpha = 0.05$.

To study the dependence on the smoothing parameter, we computed the frequencies of rejection under the null hypothesis and at the alternative $\beta = \beta_o + \Delta n^{-1/2}$ with $\Delta = 2.4$, for several bandwidths and for the different tests statistics (Table 2). Since the results for $\mathcal{W}_{\mathrm{GM}}$ and $\mathcal{W}_{\mathrm{OS}}$ and for $\mathcal{S}_{\mathrm{GM}}$ and $\mathcal{S}_{\mathrm{OS}}$ were very similar, we only report the results for the first statistic and the last one. As expected, the bandwidth selection affects the frequencies of rejection. For $b = 0.2$, the observed frequencies under the null hypothesis are very much higher than the nominal levels. This can be explained by the oversmoothing which causes a bias in the estimation of β. In fact, this bandwidth shows to be useless under normality when we study the power for the different alternatives. For the robust statistics, the bandwidth $b = 0.04$ leads to observed frequencies under $H_{0\boldsymbol{\beta}}$ closer to α.

Figure 1 presents the relative frequencies of rejection for two bandwidth choices. The filled diamonds correspond to the values of the observed frequencies under C_0, while the triangles, circles and crosses to those observed under C_1, C_2 and C_3, respectively. The thick line is the asymptotic probability of rejection, π_{LS}, under the null hypothesis and under the contiguous alternatives for the classical procedure when the errors are normally distributed. This Figure shows that the classical test is non–informative under C_2 and that it is slightly sensitive under C_1. On the other hand, the robust tests are stable under C_1, while the inclusion of leverage points slightly affects their power. To explain this loss of power under C_2, Figure 2 gives the boxplots of the estimates of β under the alternative $\beta = \beta_o + \dfrac{\Delta}{\sqrt{n}}$ with $\Delta = 1.2$. These plots show the negative bias of the estimates, which explains the loss of power.

Table 2 shows that, except for $b = 0.2$, the level of the robust procedures does not breakdown. Besides, large values of the bandwidth lead to level breakdown under C_2 since the regression functions are oversmoothed. The first two lines of Table 3 give the asymptotic probability of rejection both for the classical procedure, π_{LS}, and for any of the robust ones, π_{R} under normal errors. Table 3 shows that the single outlier introduced in C_3 breaksdown the power of the classical test for $\Delta = 2.4$. On the other hand, the power of the robust tests seems stable in both cases, for this alternative. This stability is also illustrated in Table 2 and in Figure 1. It should be noted that the classical procedure breaksdown at $\Delta = 2.4$ for any choice of the bandwidth, except for $b = 0.2$ that produces oversmoothing and is meaningless even for the null hypothesis with normal errors. It is worth noticing that under C_2 and C_3, the robust tests reach lower power values due to the bias of the estimators, as mentioned above.

5.2 Simulation study for H_{0g}

Another simulation study was performed to compare the behavior of the proposed test for the regression function with respect to the classical one. We have considered the tests given in (23) for the case in which the estimates are computed through the robust three–step procedure and through the classical least squares method. In the case of the classical test we use kernel weights to estimate ϕ_o and ϕ_1 in **Step 1**, least squares in order to estimate the parametric component in **Step 2** and we solve equation (21) with $\psi(t) = t$ for the estimation of g.

As in González Manteiga and Aneiros Pérez (2003), we considered 500 independent samples of size $n = 500$ following the model

$$\begin{aligned} y_i &= \beta x_i + g(t) + \epsilon_i \quad 1 \le i \le n\,, & (29)\\ x_i &= t_i + z_i \qquad\qquad\quad 1 \le i \le n\,, & (30) \end{aligned}$$

where $\beta = 1$, $g(t) = \phi t^2$, $\phi = 0, 0.025, 0.05, 0.10$, $t_i = \dfrac{i - 0.5}{n}$, $z_i \sim U(-0.5, 0.5)$, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon = 0.01$ and $\{z_i\}$ and $\{\epsilon_i\}$ are independent. The situation $\phi = 0$ corresponds to the null hyporthesis, while $\phi = 0.025, 0.05, 0.10$ are the alternatives we considered in our simulation study.

The smoothing procedure used a local M–estimate with bisquare score function, with tuning constant 4.685, and local medians as initial estimate. In order to avoid boundary effects we used Gasser and Müller's weights with boundary kernels given by

$$w_{n,h}(t_i, t_j) = \begin{cases} h^{-1} \displaystyle\int_{(j-1)/n}^{j/n} K(\frac{t_i - u}{h}) du & \text{if } t_i \in [h, 1-h] \\ h^{-1} \displaystyle\int_{(j-1)/n}^{j/n} K_q(\frac{t_i - u}{h}) du & \text{if } t_i = qh \in [0, h) \\ h^{-1} \displaystyle\int_{(j-1)/n}^{j/n} K_q^*(\frac{t_i - u}{h}) du & \text{if } t_i = 1 - qh \in (1-h, 1]\,. \end{cases}$$

The Epanechnikov kernel was used in the interval $[h, 1-h]$, while in the boundary points we used the boundary kernels $K_q = (c_{2,q}x + c_{1,q}) I_{[-1,q]}(x)$ and $K_q^* = (-c_{2,q}x + c_{1,q}) I_{[-q,1]}(x)$, where $c_{1,q} = 4(q^3+1)(q+1)^{-4}$ and $c_{2,q} = 6(1-q^2)(q+1)^{-4}$. Boundary kernels were considered, in this case, to improve the performance of the regression function estimator.

Several bandwidths were selected to investigate the sensitivity of the tests, both in level and power, with respect to bandwidth choice. The bandwidths considered in **Step 1** were $b = 0.04, 0.06, 0.08, 0.10, 0.12$, while those chosen in **Step 3** were $h = 0.004, 0.006, 0.008$.

For both the classical and the robust test, to compute the residual scale estimator, $\widehat{\sigma}$, pilot bandwidths $b_o = h_o = 0.25$ were used.

In order to illustrate the level and power behavior of the tests in the presence of outliers, we considered two contamination schemes:

- C_1: $\epsilon_1, \dots, \epsilon_n$ are i.i.d. distributed as $0.9N(0, \sigma_\epsilon^2) + 0.1\mathcal{C}(0, \sigma_\epsilon)$, where $\mathcal{C}(0, \sigma_\epsilon)$ denotes the Cauchy distribution centered at 0 with scale σ_ϵ. This contamination was also considered for the regression parameter.

- C_4: Artificially 53 observations at equally spaced values of t were generated following the model
$$y_i = \beta x_i + 5t_i^2 + \epsilon_i \, .$$
This case corresponds to introduce points that lie far from the central model with the aim of breaking down the level of the test.

As above, we will identify the non–contaminated case given in (29) and (30) as C_0.

The nominal level was fixed at $\alpha = 0.10$.

In Tables 4 to 8, we present the observed frequencies of rejection under the null hypothesis $H_{0g} : g(t) = 0$ and under alternatives of the form $g(t) = \phi t^2$, corresponding to $g^*(t) = 10\sqrt{5}h^{\frac{1}{4}}\phi t^2$.

Table 4 shows the observed frequencies of rejection of the classical test under the null hypothesis and under different alternatives, when we consider the non–contaminated case C_0. In order to make the comparison easier, in the third column we show the asymptotic probability of rejection for the classical test, π_{LS}. Analogous results for the robust test based on the statistic $S(\widehat{\widehat{\gamma}}, H_{0g})$ are given in Table 5.

Tables 4 and 5 show that the observed frequencies of rejection for both tests reach values very close to the asymptotic value π_{LS}. They also exemplify the sensitivity of both tests to the selection of the bandwidths. The selection of the smoothing parameters deserves more attention and may be the subject of future works.

In Tables 6 and 7 we display the observed frequencies of rejection under the null hypothesis $H_{0g} : g(t) = 0$ and under the alternatives with $\phi = 0.025, 0.05, 0.10$ in the case of contamination C_1.

Figure 3 shows the plot of the rejection frequencies of both tests for the particular bandwidth choice $b = 0.04$ and $h = 0.008$ and for equispaced alternatives of the form $g(t) = \phi t^2$, with $\phi = k \cdot 0.0125$, $1 \le k \le 8$. We represent in black the asymptotic power, π_{LS}, in blue the rejection frequencies of both tests under the non-contaminated case C_0

and in red the corresponding ones under the contamination C_1. In order to distinguish the curves, we plot squares for the robust test and circles for the classical one.

The scheme contamination C_1 affects the power of the classical test. In fact the observed percentages of rejection are all below 51 %, instead the observed powers of the robust test, specially for high values of the second bandwidth h, are much more stable, in the sense that they behave as in the non–contaminated case C_0. This becomes also evident from Figure 3, in which the curves for the LS and the robust procedure lie very close under C_0, while under C_1 the observed frequencies lie very far away one from the other.

Finally, Table 8 shows how the contamination scheme C_4 affects the level of the classical test, while the proposed robust test has a much more stable behavior.

6 An example

Daniel & Wood (1980) studied a data set obtained in a process variable study of a refinery unit. The response variable y is the octane number of the final product, while the covariates represent the feed compositions ($\mathbf{x} = (x_1, x_2, x_3)'$) and the logarithm of a combination of process conditions scaled to $[0, 1]$ (t). We have performed the test for the hypothesis $H_{0\boldsymbol{\beta}_{(1)}} : \beta_3 = 0$ with bandwidth $b = 0.06$. In order to avoid boundary effects we used Gasser and Müller's weights with boundary kernels, as in González Manteiga & Aneiros Pérez (2003) (see formula (7) therein). The Epanechnikov kernel was used in the interval $[h, 1 - h]$, while in the boundary points $t_i = qh$ or $t_i = 1 - qh$, $0 \leq q < 1$, we used, respectively, the boundary kernels $K_q(x) = (c_{2,q}x + c_{1,q})I_{[-1,q]}(x)$ and $K_q^*(x) = (-c_{2,q}x+c_{1,q})I_{[-q,1]}(x)$, where $c_{1,q} = 4(q^3+1)(q+1)^{-4}$ and $c_{2,q} = 6(1-q^2)(q+1)^{-4}$. Boundary kernels were considered, in this case, to improve the performance of the regression function estimator. The values of the estimates of $\boldsymbol{\beta}$ and the $p-$values corresponding to the test statistics are given in Table 9. All test statistics reject $H_{0\boldsymbol{\beta}_{(1)}}$ at level 0.05. Note that the $p-$value of the classical test is quite near to the stated level, while the robust tests remain significant at level 0.01. Daniel & Wood (1980) discussed the presence of three anomalous observations (labeled 75 to 77) which correspond to high values of octanes associated with low values of the first component of the feed composition. These observations extend the range of both variables (x_1 and y) and thus correspond to outliers having large residuals associated with high leverage points. We repeat the analysis excluding these three observations and the results, given in Table 9, show that now all statistics reject the null hypothesis even at level 0.01. The change in the decision for the classical test can be explained by the fact that the variances of the errors z_{ij}, $j = 1, 2, 3$, decrease when removing the anomalous observations. In particular, the variance of z_{i1} decreases from 90.8322 to 30.5141. Similar conclusions are obtained, for instance, with $b = 0.1$.

7 Final Conclusions

We have introduced two resistant procedures to test hypothesis on the parametric component in a partly linear model. The test statistics are robust versions of the classical Wald and score–type statistics, already studied in the linear regression model.

Even when the tests statistics have a limiting χ^2-distribution under the null hypothesis and under contiguous alternatives, the simulation study illustrates the slow convergence to the asymptotic distribution. Bootstrapping techniques could be implemented in order to improve the convergence rate, but this task deserves further research that will be the subject of a forthcoming work.

The simulation study also confirms the expected inadequate behavior of the classical Wald test in the presence of outliers. All methods are very sensitive to the choice of the smoothing parameter. This was also noticed by González Manteiga & Aneiros Pérez (2003), who deal with the classical procedures under dependent errors. As mentioned by these authors, more research in this direction is necessary. The proposed robust procedures for the regression parameter perform quite similarly both in level and power, either under normal errors or under the contaminations studied.

On the other hand, a robust alternative to the classical statistic to test simple hypothesis on the nonparametric component was described. Under the null hypothesis and under contiguous alternatives of order $(n^2\,h)^{-\frac{1}{4}}$, the test statistic is asymptotically normally distributed, after bias correction and both, its asymptotic bias and its asymptotic variance, depend on the score function. The simulation study seems encouraging, since the robust test performs quite stable under the contamination schemes considered.

Appendix

From now on and for the sake of simplicity, we will denote y_i^o the observations of model (1) when $\boldsymbol{\beta} = \boldsymbol{\beta}_o$, $r_i^o = y_i^o - \phi_o(t_i)$ with $\phi_o(t)$ defined as in **N7** with $\boldsymbol{\beta} = \boldsymbol{\beta}_o$ and by $r_i = y_i - \phi_{o,n}(t_i)$ where y_i follows model (1) when $\boldsymbol{\beta} = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$.

For any matrix $\mathbf{B} \in I\!R^{p\times p}$, let $|\mathbf{B}| = \max_{1\le \ell,j\le p} |b_{\ell\, j}|$.

The following Lemmas allow to derive the asymptotic behavior of the test statistic $D(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}, H_{0\boldsymbol{\beta}})$ under contiguous alternatives. The asymptotic results under the null hypothesis are obtained taking $\mathbf{c} = \mathbf{0}$, in these Lemmas.

Lemma 1
Let $(y_i, \mathbf{x}_i', t_i)'$, $1 \le i \le n$ be independent random vectors satisfying (1) with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$, $\mathbf{c} \in I\!R^p$, and ϵ_i independent of $(\mathbf{x}_i', t_i)'$ with distribution $F(\cdot/\sigma_\epsilon)$. Assume that t_i are random variables with distribution on $[0,1]$. Let $\widehat{\phi}_j(t)$, $1 \le j \le p$, be estimates of $\phi_j(t)$ such that

$$\sup_{t\in[0,1]} |\widehat{\phi}_j(t) - \phi_j(t)| \xrightarrow{p} 0, \quad 1 \le j \le p$$

and $\widehat{\phi}_o(t)$ *such that*

$$\sup_{t\in[0,1]} |\widehat{\phi}_o(t) - \phi_{o,n}(t)| \xrightarrow{p} 0 \,,$$

where $\phi_{o,n}(t) = \boldsymbol{\beta}_n'\boldsymbol{\phi}(t) + g(t)$. *Assume that* $\widetilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_o$ *and* $s_n \xrightarrow{p} \sigma_0$. *Then, under* ***N1*** *to* ***N3***, $\hat{\mathbf{A}}\left(\widetilde{\boldsymbol{\beta}}\right) \xrightarrow{p} \mathbf{A}$ *and* $\hat{\mathbf{B}}\left(\widetilde{\boldsymbol{\beta}}\right) \xrightarrow{p} \mathbf{B}$, *where* $\hat{\mathbf{A}}(\boldsymbol{\beta})$ *and* $\hat{\mathbf{B}}(\boldsymbol{\beta})$ *are given in (8) and (9) and* $\mathbf{A}$ *and* $\mathbf{B}$ *are given in (4) and (5), respectively.*

Proof. Denote by $\widetilde{\boldsymbol{\beta}}^* = \widetilde{\boldsymbol{\beta}} - n^{-1/2}\mathbf{c}$ and by ξ_i intermediate points between $r_i - \mathbf{z}_i'\widetilde{\boldsymbol{\beta}} = r_i^o - \mathbf{z}_i'\widetilde{\boldsymbol{\beta}}^*$ and $\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}$. Let $\widehat{\eta}_j(t) = \widehat{\phi}_j(t) - \phi_j(t)$, $1 \le j \le p$, $\widehat{\eta}_o(t) = \widehat{\phi}_o(t) - \phi_{o,n}(t)$, and $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1(t), \ldots, \widehat{\eta}_p(t))'$. A Taylor expansion of first order and some algebra lead us to $\hat{\mathbf{A}}(\widetilde{\boldsymbol{\beta}}) = \mathbf{A}_n^1 + \mathbf{A}_n^2 + \mathbf{A}_n^3 + \mathbf{A}_n^4$, where

$$\begin{aligned}
\mathbf{A}_n^1 &= \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{r_i^o - \mathbf{z}_i'\widetilde{\boldsymbol{\beta}}}{s_n}\right) w_2\left(\|\mathbf{z}_i\|\right)\mathbf{z}_i\mathbf{z}_i' \\
\mathbf{A}_n^2 &= -\frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{s_n}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right)\left[\widehat{\boldsymbol{\eta}}(t_i)\mathbf{z}_i' + \widehat{\mathbf{z}}_i\,\widehat{\boldsymbol{\eta}}(t_i)'\right] \\
\mathbf{A}_n^3 &= -\frac{1}{n}\sum_{i=1}^n \psi_1''\left(\frac{\xi_i}{s_n}\right)\left(\frac{\widehat{\eta}_o(t_i) - \widehat{\boldsymbol{\eta}}(t_i)'\widetilde{\boldsymbol{\beta}}}{s_n}\right) w_2\left(\|\mathbf{z}_i\|\right)\mathbf{z}_i\mathbf{z}_i' \\
\mathbf{A}_n^4 &= \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{s_n}\right)\left[w_2\left(\|\widehat{\mathbf{z}}_i\|\right) - w_2\left(\|\mathbf{z}_i\|\right)\right]\mathbf{z}_i\mathbf{z}_i'\,.
\end{aligned}$$

As in Lemma 2 in Bianco and Boente (2004), we have that $\mathbf{A}_n^1 \xrightarrow{p} \mathbf{A}$, since $\widetilde{\boldsymbol{\beta}}^* \xrightarrow{p} \boldsymbol{\beta}_o$. Using **N2**, **N3**, the consistency of s_n and $\widetilde{\boldsymbol{\beta}}$, the Law of Large Numbers and the fact that $\max_{0\le j\le p}\sup_{t\in[0,1]} |\widehat{\eta}_j(t)| \xrightarrow{p} 0$, we get that $\mathbf{A}_n^j \xrightarrow{p} 0$ for $j = 2, 3, 4$.

Similar arguments lead to the consistency of $\hat{\mathbf{B}}\left(\widetilde{\boldsymbol{\beta}}\right)$. □

An analogous result holds for the matrices $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{B}}$ defined in (10) and (11), respectively.

Lemma 2
Let $(y_i, \mathbf{x}_i', t_i)'$, $1 \le i \le n$ *be independent random vectors satisfying (1) with* $\boldsymbol{\beta}_n = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$ *and* ϵ_i *independent of* $(\mathbf{x}_i', t_i)'$ *with distribution* $F(\cdot/\sigma_\epsilon)$ *such that* $E\left(\psi_1\left(\epsilon/\sigma\right)\right) = 0$, *for any* $\sigma > 0$. *Assume that* t_i *are random variables with distribution on* $[0, 1]$. *Then, if* $s_n \xrightarrow{p} \sigma_0$ *and* $\widehat{\boldsymbol{\beta}}$ *is a consistent estimate of the regression parameter satisfying (3), under* ***N1*** *to* ***N6*** *and* ***N8***, $n^{-1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o\right) \xrightarrow{D} N\left(\mathbf{c}, \boldsymbol{\Sigma}\right)$, *where* $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ *and* $\mathbf{A}$ *and* $\mathbf{B}$ *are given in (4) and (5), respectively.*

Proof. It will be enough to show that $n^{\frac{1}{2}}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n\right) \xrightarrow{D} N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$. Write

$$\begin{aligned}
\mathbf{L}_n\left(\sigma, \mathbf{b}\right) &= \frac{\sigma}{n}\sum_{i=1}^n \psi_1\left(\frac{r_i - \mathbf{z}_i'\mathbf{b}}{\sigma}\right) w_2\left(\|\mathbf{z}_i\|\right)\mathbf{z}_i \\
\widehat{\mathbf{L}}_n\left(\sigma, \mathbf{b}\right) &= \frac{\sigma}{n}\sum_{i=1}^n \psi_1\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\mathbf{b}}{\sigma}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right)\widehat{\mathbf{z}}_i\,.
\end{aligned}$$

Using a first order Taylor expansion around $\widehat{\boldsymbol{\beta}}$, we get

$$\widehat{\mathbf{L}}_n(\sigma, \boldsymbol{\beta}_n) = \frac{\sigma}{n}\sum_{i=1}^n \psi_1\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\beta}}}{\sigma}\right) w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i + \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{\sigma}\right) w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n),$$

with $\widetilde{\boldsymbol{\beta}}$ an intermediate point between $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_n$ and thus $\widetilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_o$. This implies that

$$\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}_n) = 0 + \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{s_n}\right) w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n)$$

and so, we get $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n) = \mathbf{A}_n^{-1}\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}_n)$ with $\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{s_n}\right) w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i'$. Using that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_o$, Lemma 1 implies that $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ and therefore, from **N2** it will be enough to show that

a) $n^{\frac{1}{2}}\mathbf{L}_n(\sigma_\epsilon, \boldsymbol{\beta}_n) \xrightarrow{D} N(\mathbf{0}, \mathbf{B})$.

b) $n^{\frac{1}{2}}\left[\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}_n) - \mathbf{L}_n(s_n, \boldsymbol{\beta}_n)\right] \xrightarrow{p} 0$

c) $n^{\frac{1}{2}}\left[\mathbf{L}_n(s_n, \boldsymbol{\beta}_n) - \mathbf{L}_n(\sigma_\epsilon, \boldsymbol{\beta}_n)\right] \xrightarrow{p} 0$

a) Follows inmediately from the Central Limit Theorem, since $r_i - \mathbf{z}_i\boldsymbol{\beta}_n' = \epsilon_i$.

b) Denote by ξ_i intermediate points between $r_i - \mathbf{z}_i'\widetilde{\boldsymbol{\beta}}$ and $\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}$. Let $\widehat{\eta}_j(t) = \widehat{\phi}_j(t) - \phi_j(t)$, $1 \le j \le p$, $\widehat{\eta}_o(t) = \widehat{\phi}_o(t) - \phi_{o,n}(t)$ and $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1(t), \ldots, \widehat{\eta}_p(t))'$. Using a second order Taylor expansion, we have that $\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}_n) = \mathbf{L}_n(s_n, \boldsymbol{\beta}_n) + \widehat{\mathbf{L}}_{n,1} + \widehat{\mathbf{L}}_{n,2} + \widehat{\mathbf{L}}_{n,3} + \widehat{\mathbf{L}}_{n,4} + \widehat{\mathbf{L}}_{n,5}$, where

$$\begin{aligned}
\widehat{\mathbf{L}}_{n,1} &= \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{r_i - \mathbf{z}_i'\boldsymbol{\beta}_n}{s_n}\right)[\widehat{\boldsymbol{\eta}}'(t_i)\boldsymbol{\beta}_n - \widehat{\eta}_o(t_i)]\; w_2(\|\mathbf{z}_i\|)\;\mathbf{z}_i \\
&= \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{r_i^o - \mathbf{z}_i'\boldsymbol{\beta}_o}{s_n}\right)[\widehat{\boldsymbol{\eta}}'(t_i)\boldsymbol{\beta}_o - \widehat{\eta}_o(t_i)]\; w_2(\|\mathbf{z}_i\|)\;\mathbf{z}_i \\
\widehat{\mathbf{L}}_{n,2} &= \frac{s_n}{n}\sum_{i=1}^n \psi_1\left(\frac{r_i - \mathbf{z}_i'\boldsymbol{\beta}_n}{s_n}\right)\left[w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i - w_2(\|\mathbf{z}_i\|)\;\mathbf{z}_i\right] \\
&= \frac{s_n}{n}\sum_{i=1}^n \psi_1\left(\frac{r_i^o - \mathbf{z}_i'\boldsymbol{\beta}_o}{s_n}\right)\left[w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i - w_2(\|\mathbf{z}_i\|)\;\mathbf{z}_i\right] \\
\widehat{\mathbf{L}}_{n,3} &= \frac{s_n}{n}\sum_{i=1}^n \left[\psi_1\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\boldsymbol{\beta}_n}{s_n}\right) - \psi_1\left(\frac{r_i - \mathbf{z}_i'\boldsymbol{\beta}_n}{s_n}\right)\right] w_2(\|\widehat{\mathbf{z}}_i\|)\;(\widehat{\mathbf{z}}_i - \mathbf{z}_i) \\
\widehat{\mathbf{L}}_{n,4} &= \frac{1}{2\,s_n}\frac{1}{n}\sum_{i=1}^n \psi_1''\left(\frac{\xi_i}{s_n}\right)[\widehat{\boldsymbol{\eta}}'(t_i)\boldsymbol{\beta}_n - \widehat{\eta}_o(t_i)]^2\; w_2(\|\widehat{\mathbf{z}}_i\|)\;\mathbf{z}_i \\
\widehat{\mathbf{L}}_{n,5} &= \frac{1}{n}\sum_{i=1}^n \psi_1'\left(\frac{r_i - \mathbf{z}_i'\boldsymbol{\beta}_n}{s_n}\right)[\widehat{\boldsymbol{\eta}}'(t_i)\boldsymbol{\beta}_n - \widehat{\eta}_o(t_i)]\;[w_2(\|\widehat{\mathbf{z}}_i\|) - w_2(\|\mathbf{z}_i\|)]\;\mathbf{z}_i\;.
\end{aligned}$$

Since, **N3** entails $|w_2(\|\widehat{\mathbf{z}}_i\|) - w_2(\|\mathbf{z}_i\|)| \le C\dfrac{\|\widehat{\boldsymbol{\eta}}(t_i)\|}{\|\mathbf{z}_i\|}$, where $C = \|w_2\|_\infty + C_{\psi_2}$, and

$$\begin{aligned}
n^{\frac{1}{2}}\|\widehat{\mathbf{L}}_{n,3}\| &\le p\,\|w_2\|_\infty \|\psi_1'\|_\infty n^{\frac{1}{2}} \left[\max_{0\le j\le p}\sup_{t\in[0,1]}|\widehat{\eta}_j(t)|\right]^2 (1+p\,\|\boldsymbol{\beta}_n\|)\\
n^{\frac{1}{2}}\|\widehat{\mathbf{L}}_{n,4}\| &\le \frac{1}{2}\frac{1}{s_n}\|\psi_1''\|_\infty n^{\frac{1}{2}} \left[\max_{0\le j\le p}\sup_{t\in[0,1]}|\widehat{\eta}_j(t)|\right]^2 (1+p\,\|\boldsymbol{\beta}_n\|)^2 \Bigg(\|\psi_2\|_\infty +\\
&\quad + p\,\|w_2\|_\infty \max_{0\le j\le p}\sup_{t\in[0,1]}|\widehat{\eta}_j(t)|\Bigg)\\
n^{\frac{1}{2}}\|\widehat{\mathbf{L}}_{n,5}\| &\le p\,C\|\psi_1'\|_\infty (1+p\,\|\boldsymbol{\beta}_n\|)\, n^{\frac{1}{2}} \left[\max_{0\le j\le p}\sup_{t\in[0,1]}|\widehat{\eta}_j(t)|\right]^2 ,
\end{aligned}$$

using (15), (17) and the consistency of s_n, we get that for $3 \le j \le 5$, $n^{\frac{1}{2}}\|\widehat{\mathbf{L}}_{n,j}\| \xrightarrow{p} 0$.

It remains to show that $n^{\frac{1}{2}}\widehat{\mathbf{L}}_{n,j} \xrightarrow{p} 0$ for $j = 1, 2$, that is,

$$\begin{aligned}
n^{\frac{1}{2}}\frac{s_n}{n}\sum_{i=1}^n \psi_1'\left(\frac{r_i^o - \mathbf{z}_i'\boldsymbol{\beta}_o}{s_n}\right)\widehat{\eta}_\ell(t_i)\, w_2(\|\mathbf{z}_i\|)\, \mathbf{z}_i &\xrightarrow{p} 0 \qquad\qquad 0\le \ell \le p\\
n^{\frac{1}{2}}\frac{s_n}{n}\sum_{i=1}^n \psi_1\left(\frac{r_i^o - \mathbf{z}_i'\boldsymbol{\beta}_o}{s_n}\right)\Big[w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i - w_2(\|\mathbf{z}_i\|)\,\mathbf{z}_i\Big] &\xrightarrow{p} 0\,.
\end{aligned}$$

which follows from the proof of Theorem 2 in Bianco and Boente (2004).

c) Since

$$n^{\frac{1}{2}}\left[\mathbf{L}_n(s_n,\boldsymbol{\beta}_n) - \mathbf{L}_n(\sigma_o,\boldsymbol{\beta}_n)\right] = n^{-\frac{1}{2}}\sum_{i=1}^n \left[\psi_{1,s_n}(r_i - \mathbf{z}_i'\boldsymbol{\beta}_n) - \psi_{1,\sigma_n}(r_i - \mathbf{z}_i'\boldsymbol{\beta}_n)\right] w_2(\|\mathbf{z}_i\|)\,\mathbf{z}_i\,,$$

we get the desired result using **N1**, the boundness of ψ_2 and the maximal inequality for covering numbers, as in b). □

Lemma 3
*Let $(y_i, \mathbf{x}_i', t_i)'$, $1 \le i \le n$ be independent random vectors satisfying (1) with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_o + \mathbf{c}n^{-1/2}$, $\mathbf{c} \in I\!R^p$, and ϵ_i independent of $(\mathbf{x}_i', t_i)'$ with symmetric distribution $F(\cdot/\sigma_\epsilon)$ such that $E(\psi_1(\epsilon/\sigma)) = 0$, for any $\sigma > 0$. Assume that t_i are random variables with distribution on $[0,1]$. Then, if $s_n \xrightarrow{p} \sigma_0$ under **N1** to **N6** and **N8**, $n^{1/2}\mathbf{U}_n(\boldsymbol{\beta}_o) \xrightarrow{D} N(\mathbf{Ac},\mathbf{B})$, where $\mathbf{B}$ is given in (5).*

Proof. Define

$$\widehat{\mathbf{L}}_n(\sigma,\mathbf{b}) = \frac{\sigma}{n}\sum_{i=1}^n \psi_1\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\mathbf{b}}{\sigma}\right) w_2(\|\widehat{\mathbf{z}}_i\|)\,\widehat{\mathbf{z}}_i\,.$$

Then, we have that $\mathbf{U}_n(\boldsymbol{\beta}_n) = \widehat{\mathbf{L}}_n(s_n,\boldsymbol{\beta}_n)$. Following similar arguments to those considered in the proof of Theorem 2 of Bianco & Boente (2004), it can be shown that $n^{1/2}\widehat{\mathbf{L}}_n(s_n,\boldsymbol{\beta}_n) \xrightarrow{D} N(\mathbf{0},\mathbf{B})$. Therefore, the proof will be complete if we show that

$n^{1/2}\left(\mathbf{U}_n(\boldsymbol{\beta}_n) - \mathbf{U}_n(\boldsymbol{\beta}_o)\right) \xrightarrow{p} -\mathbf{A}\mathbf{c}$. Using a first order Taylor expansion around $\boldsymbol{\beta}_o$ and if $\widetilde{\boldsymbol{\beta}}$ denotes an intermediate point, we have

$$\widehat{\mathbf{L}}_n\left(\sigma, \boldsymbol{\beta}_n\right) - \widehat{\mathbf{L}}_n\left(\sigma, \boldsymbol{\beta}_o\right) = \frac{1}{n}\sum_{i=1}^{n} \psi_1'\left(\frac{\widehat{r}_i - \widehat{\mathbf{z}}_i'\widetilde{\boldsymbol{\beta}}}{\sigma}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right) \widehat{\mathbf{z}}_i\widehat{\mathbf{z}}_i'\left(\boldsymbol{\beta}_o - \boldsymbol{\beta}_n\right) = -n^{-\frac{1}{2}}\hat{\mathbf{A}}\left(\widetilde{\boldsymbol{\beta}}\right)\mathbf{c} \,,$$

which entails that $n^{1/2}\left(\mathbf{U}_n(\boldsymbol{\beta}_n) - \mathbf{U}_n(\boldsymbol{\beta}_o)\right) = -\hat{\mathbf{A}}\left(\widetilde{\boldsymbol{\beta}}\right)\mathbf{c}$. Hence, the proof follows from Lemma 1. □

Proof of Theorem 1. i) Follows inmediately from Lemma 2., with $\mathbf{c} = \mathbf{0}$.

ii) Denote $\mathcal{W}(\boldsymbol{\beta}) = n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Thus,

$$\mathcal{W}_n = n(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o)'\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o) = \mathcal{W}(\boldsymbol{\beta}) + n(\boldsymbol{\beta} - \boldsymbol{\beta}_o)'\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o + \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \,.$$

Taking $\mathbf{c} = \mathbf{0}$, Lemma 2 and the consistency of $\widehat{\boldsymbol{\Sigma}}$ entail $\mathcal{W}(\boldsymbol{\beta}) \xrightarrow{D} \chi^2_p$.

Besides, from the consistency of $\widehat{\boldsymbol{\Sigma}}$ and the fact that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, we get that $(\boldsymbol{\beta} - \boldsymbol{\beta}_o)'\,\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_o + \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{p} (\boldsymbol{\beta} - \boldsymbol{\beta}_o)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_o) > 0$ and so, since $\mathcal{W}(\boldsymbol{\beta}) > 0$ with probability converging to 1, the result follows inmediately.

iii) Is an inmediate consequence of Lemma 2. □

Proof of Theorem 2. i) and iii) follow inmediately from Lemma 3 and the consistency of $\widehat{\mathbf{C}}$.

ii) As in Lemma 2 in Bianco & Boente (2004), for any $\boldsymbol{\beta} \neq \boldsymbol{\beta}_o$, it is easy to show that $\mathbf{U}_n(\boldsymbol{\beta}_o) \xrightarrow{p} \mathbf{U}$ with

$$\mathbf{U} = \sigma_0 E\left(\psi_1\left(\frac{\epsilon + \mathbf{Z}'\left(\boldsymbol{\beta} - \boldsymbol{\beta}_o\right)}{\sigma_0}\right) w_2\left(\|\mathbf{Z}\|\right)\mathbf{Z}\right) \,.$$

Therefore, since ψ_1 is increasing, $\mathbf{U}'\left(\boldsymbol{\beta} - \boldsymbol{\beta}_o\right) > 0$ and hence we get that $\|n^{1/2}\mathbf{U}_n(\boldsymbol{\beta})\| \xrightarrow{p} \infty$, which entails the result using that $\widehat{\mathbf{C}} \xrightarrow{p} \mathbf{B}$, with is positive definite. □

References

Bhattacharya, P. K. & Zhao, P. L. (1997). Semiparametric inference in a partial linear model. *Ann. Statist.* **25**, 244-262.

Bianco, A. & Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Plann. Inference* **122**, 229-252.

Bickel, P., 1975. One–Step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428-434.

Boente, G. & Fraiman, R. (1991a). A functional approach to robust nonparametric regression. In: *Directions in robust statistics and diagnostics*, Werner Stahel and Sanford Weisberg (ed.). Proceedings of the IMA Institute, USA, **33**, Part I, pp. 35-46.

Boente, G. & Fraiman, R. (1991b). Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. *Internat. Statist. Rev.* **59**, 355-372.

Boente, G., Fraiman, R. & Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *J. Statist. Plann. Inference* **57**, 109-142.

Carroll, R. & Ruppert, D. (1988). . Chapman and Hall.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16**, 136-146.

Chen, H. & Shiau, J. (1994). Data-driven efficient estimates for partially linear models. *Ann. Statist.***22**, 211-237.

Daniel, C. & Wood, F. (1980). *Fitting equations to data: Computer analysis of multifactor data*. Wiley, New York.

Denby, L. (1986). Smooth regression functions. *Statistical Research Report* **26**, AT&T Bell Laboratories, Murray Hill.

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. Statist.* **15**, 93-111.

Green, P. & Silverman, B. (1995). *Nonparametric regression and generalized linear models: A roughness penalty approach*, Monographs on Statist. Appl. Probab. No. 58. Chapman and Hall, London.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press.

Härdle, W., Liang, H. & Gao, J. (2000). *Partially linear models.* Springer-Verlag.

He, X. (2002). Robust tests in statistics – a review of the past decade. *Estadística* **54**, 29-46.

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

Linton, O. (1995). Second order approximations in the partially linear regression model. *Econometrica* **63**, 1079-1112.

Markatou, M. & He, X. (1994). Bounded influence and high breakdown point testing procedures in linear models, *J. Amer. Statist. Assoc.* **89**, 543-549.

Robinson, P. (1984). Robust nonparametric autoregression. In: *Robust and nonlinear time series analysis*, Franke, Härdle & Martin (eds.). Heidelberg: Springer–Verlag.

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

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

Simpson, D., Ruppert, D. & Carroll, R.J. (1992). On one-step $GM-$estimates and stability of inferences in linear regression. *J. Amer. Statist. Assoc.* **87**, 439-450.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Ser. B* **50**, 413-436.

Stone, C. (1977). Consistent nonparametric regression. *Ann. Statist.* **5**, 595-645.

Sun, Y. & Wu, H. (2005). Semiparametric time–varying coefficients regression model for longitudinal data. *Scand. J. Statist.* **32**, 21-47.

α	$n=100,\ b=0.08$ $\mathcal{W}_{\text{LS}}$	$\mathcal{W}_{\text{GM}}$	$\mathcal{W}_{\text{OS}}$	$\mathcal{S}_{\text{GM}}$	$\mathcal{S}_{\text{OS}}$	$n=500,\ b=0.065$ $\mathcal{W}_{\text{LS}}$	$\mathcal{W}_{\text{GM}}$	$\mathcal{W}_{\text{OS}}$	$\mathcal{S}_{\text{GM}}$	$\mathcal{S}_{\text{OS}}$
0.025	0.0556	0.0582	0.0626	0.056	0.0552	0.0306	0.0312	0.0326	0.0322	0.0322
0.050	0.0990	0.0958	0.1054	0.094	0.0940	0.0588	0.0614	0.0604	0.0610	0.0610

Table 1: Observed frequencies of rejection under the null hypothesis with normal errors (C_0).

	$\Delta=0$											
	C_0						C_2					
b	0.008	0.02	0.03	0.04	0.08	0.2	0.008	0.02	0.03	0.04	0.08	0.2
$\mathcal{W}_{\text{LS}}$	0.1320	0.0776	0.0652	0.0614	0.0990	0.2416	1	1	1	1	1	1
$\mathcal{W}_{\text{GM}}$	0.1220	0.0860	0.0726	0.0666	0.0958	0.4906	0.0712	0.0486	0.046	0.0438	0.1544	1
$\mathcal{S}_{\text{OS}}$	0.1174	0.0840	0.0690	0.0652	0.0940	0.4914	0.0690	0.0466	0.0438	0.0430	0.1512	1
	$\Delta=2.4$											
	C_0						C_3					
b	0.008	0.02	0.03	0.04	0.08	0.2	0.008	0.02	0.03	0.04	0.08	0.2
$\mathcal{W}_{\text{LS}}$	0.9984	1	1	1	0.9996	0.9996	0.1698	0.0346	0.0208	0.0168	0.0546	0.9932
$\mathcal{W}_{\text{GM}}$	0.9962	0.9992	0.9994	0.9998	0.9994	0.9932	0.8532	0.9154	0.9498	0.9626	0.9730	0.9776
$\mathcal{S}_{\text{OS}}$	0.9962	0.9992	0.9994	0.9998	0.9992	0.9930	0.8554	0.9152	0.9472	0.9616	0.9696	0.9766

Table 2: Observed frequencies of rejection at $\beta=3+\Delta n^{-1/2}$, $\Delta=0$ and 2.4 for different values of the bandwidth under normal errors and under contaminations C_2 and C_3, respectively.

		Δ						
		0	0.1	0.2	0.4	0.8	1.2	2.4
	π_{LS}	0.0500	0.0572	0.0791	0.1701	0.5160	0.8508	1
	π_{R}	0.0500	0.0570	0.0783	0.1666	0.5046	0.8406	1
	$\mathcal{W}_{\mathrm{LS}}$	0.0614	0.0694	0.0942	0.1836	0.5050	0.8302	1
C_0	$\mathcal{W}_{\mathrm{GM}}$	0.0666	0.0740	0.0956	0.1864	0.4892	0.8126	0.9998
	$\mathcal{S}_{\mathrm{OS}}$	0.0652	0.0730	0.0944	0.1826	0.4846	0.8082	0.9998
	$\mathcal{W}_{\mathrm{LS}}$	0.0646	0.0678	0.0748	0.1192	0.2788	0.4718	0.7726
C_1	$\mathcal{W}_{\mathrm{GM}}$	0.0624	0.0680	0.0826	0.1494	0.3960	0.6918	0.9892
	$\mathcal{S}_{\mathrm{OS}}$	0.0610	0.0662	0.0808	0.1448	0.3922	0.6856	0.9882
	$\mathcal{W}_{\mathrm{LS}}$	1	1	1	1	1	1	1
C_2	$\mathcal{W}_{\mathrm{GM}}$	0.0438	0.0492	0.0604	0.1006	0.2566	0.487	0.9566
	$\mathcal{S}_{\mathrm{OS}}$	0.0430	0.0504	0.0596	0.0990	0.2536	0.4840	0.9540
	$\mathcal{W}_{\mathrm{LS}}$	0.9464	0.9358	0.9202	0.8858	0.7762	0.6292	0.1300
C_3	$\mathcal{W}_{\mathrm{GM}}$	0.0810	0.0616	0.0544	0.0674	0.1926	0.4334	0.9626
	$\mathcal{S}_{\mathrm{OS}}$	0.0786	0.0588	0.0514	0.0664	0.1910	0.4302	0.9616

Table 3: Observed frequencies of rejection at $\beta = 3 + \Delta n^{-1/2}$, for $b = 0.04$ under normal errors and under contamination.

Classical test							
$\phi = 0$		π_{LS}			b		
			0.04	0.06	0.08	0.10	0.12
	0.004	0.10	0.068	0.06	0.06	0.064	0.062
h	0.006	0.10	0.134	0.128	0.13	0.128	0.128
	0.008	0.10	0.194	0.184	0.176	0.182	0.176
$\phi = 0.025$		π_{LS}			b		
			0.04	0.06	0.08	0.10	0.12
	0.004	0.196	0.21	0.196	0.198	0.196	0.194
h	0.006	0.223	0.32	0.302	0.302	0.306	0.304
	0.008	0.248	0.38	0.382	0.376	0.380	0.372
$\phi = 0.05$		π_{LS}			b		
			0.04	0.06	0.08	0.10	0.12
	0.004	0.661	0.53	0.530	0.522	0.528	0.532
h	0.006	0.787	0.67	0.660	0.672	0.668	0.666
	0.008	0.868	0.73	0.728	0.728	0.730	0.738
$\phi = 0.10$		π_{LS}			b		
			0.04	0.06	0.08	0.10	0.12
	0.004	1	0.99	0.988	0.988	0.988	0.988
h	0.006	1	0.99	0.996	0.996	0.994	0.994
	0.008	1	1	0.998	0.998	0.998	0.998

Table 4: Observed frequencies of rejection of the classical test under the null hypothesis and under alternatives $g(t) = \phi t^2$ under C_0. π_{LS} denotes the corresponding asymptotic power.

Robust test						
$\phi = 0$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.14	0.130	0.128	0.124	0.122
h	0.006	0.19	0.180	0.184	0.186	0.188
	0.008	0.22	0.222	0.226	0.222	0.220
$\phi = 0.025$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.23	0.226	0.232	0.228	0.230
h	0.006	0.31	0.308	0.302	0.308	0.308
	0.008	0.37	0.356	0.358	0.354	0.356
$\phi = 0.05$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.55	0.554	0.554	0.560	0.562
h	0.006	0.63	0.638	0.646	0.650	0.650
	0.008	0.70	0.716	0.712	0.702	0.702
$\phi = 0.10$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.99	0.990	0.990	0.990	0.990
h	0.006	0.99	0.994	0.994	0.994	0.994
	0.008	1	0.998	0.998	0.996	0.996

Table 5: Observed frequencies of rejection of the proposed test under the null hypothesis and under alternatives $g(t) = \phi t^2$ under C_0.

Classical test						
$\phi = 0$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.08	0.076	0.076	0.070	0.074
h	0.006	0.13	0.134	0.138	0.132	0.138
	0.008	0.18	0.178	0.168	0.166	0.164
$\phi = 0.025$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.10	0.100	0.096	0.098	0.102
h	0.006	0.17	0.166	0.164	0.158	0.162
	0.008	0.20	0.194	0.198	0.194	0.198
$\phi = 0.05$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.15	0.160	0.164	0.164	0.160
h	0.006	0.25	0.250	0.248	0.248	0.252
	0.008	0.29	0.298	0.300	0.302	0.300
$\phi = 0.10$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.35	0.356	0.348	0.350	0.350
h	0.006	0.47	0.466	0.462	0.462	0.460
	0.008	0.52	0.504	0.510	0.506	0.504

Table 6: Observed frequencies of rejection of the classical test under the null hypothesis and under alternatives $g(t) = \phi t^2$ under C_1.

Robust test						
$\phi = 0$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.40	0.388	0.396	0.392	0.388
h	0.006	0.26	0.256	0.260	0.256	0.252
	0.008	0.27	0.272	0.272	0.264	0.264
$\phi = 0.025$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.50	0.494	0.488	0.486	0.484
h	0.006	0.39	0.382	0.378	0.380	0.378
	0.008	0.40	0.390	0.386	0.390	0.388
$\phi = 0.05$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.73	0.730	0.722	0.722	0.724
h	0.006	0.68	0.672	0.666	0.670	0.676
	0.008	0.69	0.672	0.672	0.672	0.670
$\phi = 0.10$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.98	0.984	0.984	0.982	0.984
h	0.006	0.99	0.988	0.988	0.986	0.986
	0.008	0.99	0.994	0.994	0.992	0.992

Table 7: Observed frequencies of rejection of the proposed test under the null hypothesis and under alternatives $g(t) = \phi t^2$ under C_1.

Classical Test						
$\phi = 0$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.48	0.482	0.482	0.494	0.496
h	0.006	0.56	0.560	0.560	0.560	0.564
	0.008	0.87	0.876	0.876	0.876	0.876
Robust Test						
$\phi = 0$		b				
		0.04	0.06	0.08	0.10	0.12
	0.004	0.004	0.004	0.004	0.004	0.004
h	0.006	0.01	0.010	0.008	0.008	0.008
	0.008	0.03	0.030	0.032	0.026	0.028

Table 8: Observed frequencies of rejection of the proposed test under the null hypothesis for contamination C_4.

	Estimated values			$p-$values		
	$\widehat{\boldsymbol{\beta}}_{\mathrm{LS}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{GM}}$	$\widehat{\boldsymbol{\beta}}_{\mathrm{OS}}$	$\mathcal{W}_{\mathrm{LS}}$	$\mathcal{W}_{\mathrm{GM}}$	$\mathcal{S}_{\mathrm{OS}}$
Original Data Set	-0.0982	-0.1067	-0.1105			
	-0.1255	-0.1184	-0.1410			
	-0.0308	-0.0506	-0.0475	0.0456	0.0028	0.0014
Data Set excluding observations 75 to 77	-0.1139	-0.1100	-0.1081			
	-0.1112	-0.1223	-0.1201			
	-0.0563	-0.0522	-0.0479	0.0007	0.0018	0.0026

Table 9: Results for the refinery data

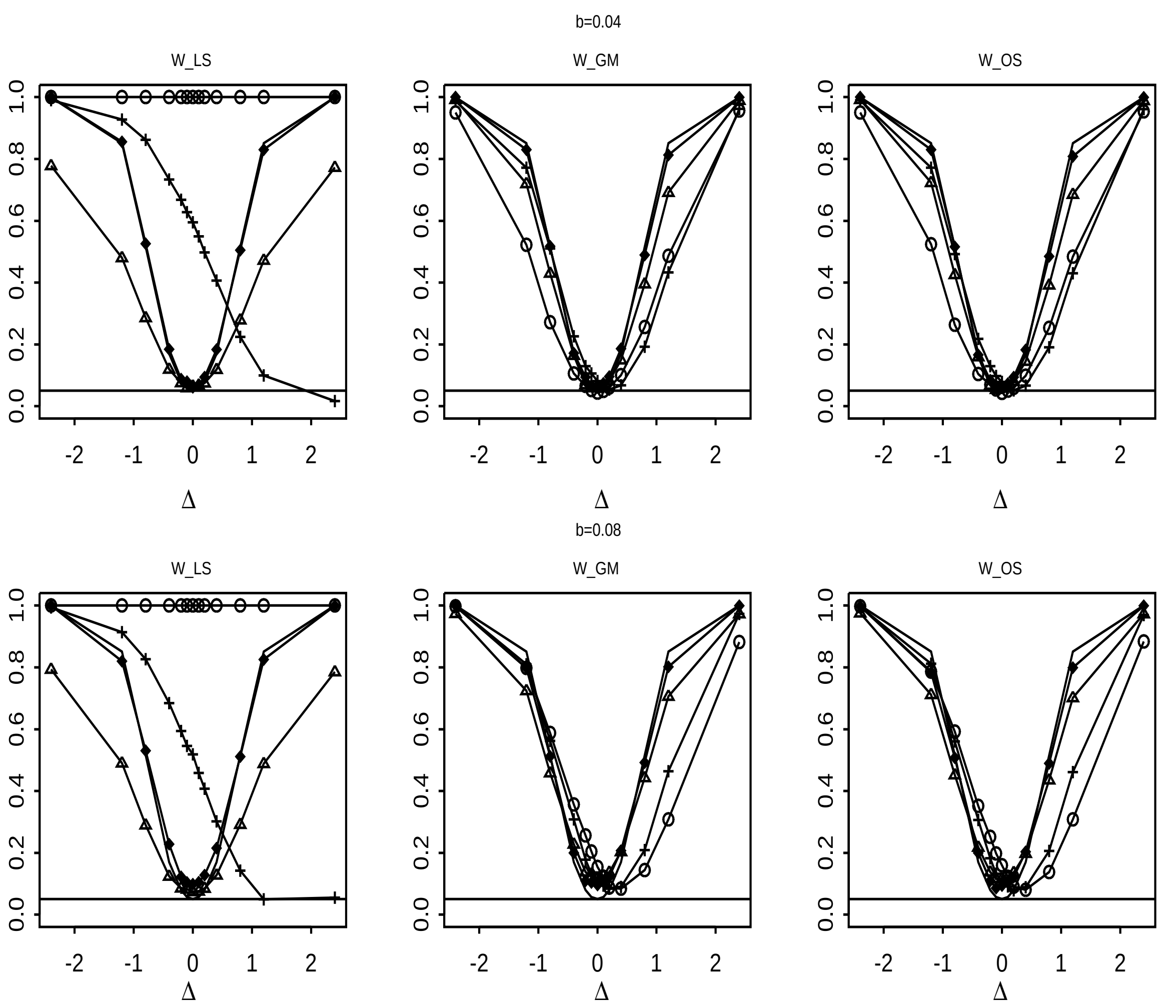

Figure 1: Frequencies of rejection. The values of the observed frequencies under C_0 are plotted with filled diamonds, while triangles, circles and crosses correspond to C_1, C_2 and C_3, respectively. The thick line is the asymptotic probability of rejection, π_{LS}.

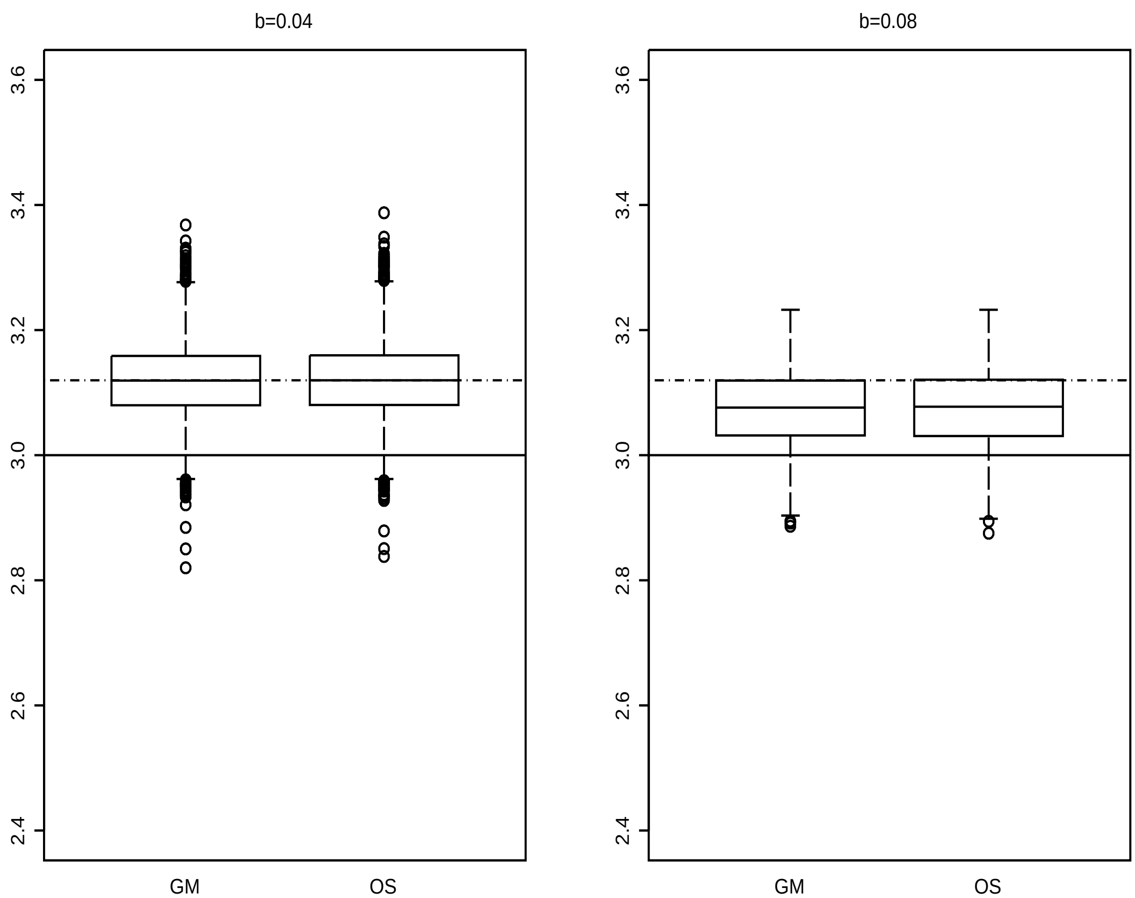

Figure 2: Boxplots for $\widehat{\beta}$ under $H_{1\boldsymbol{\beta}}$ with $\Delta = 1.2$ for contamination C_2

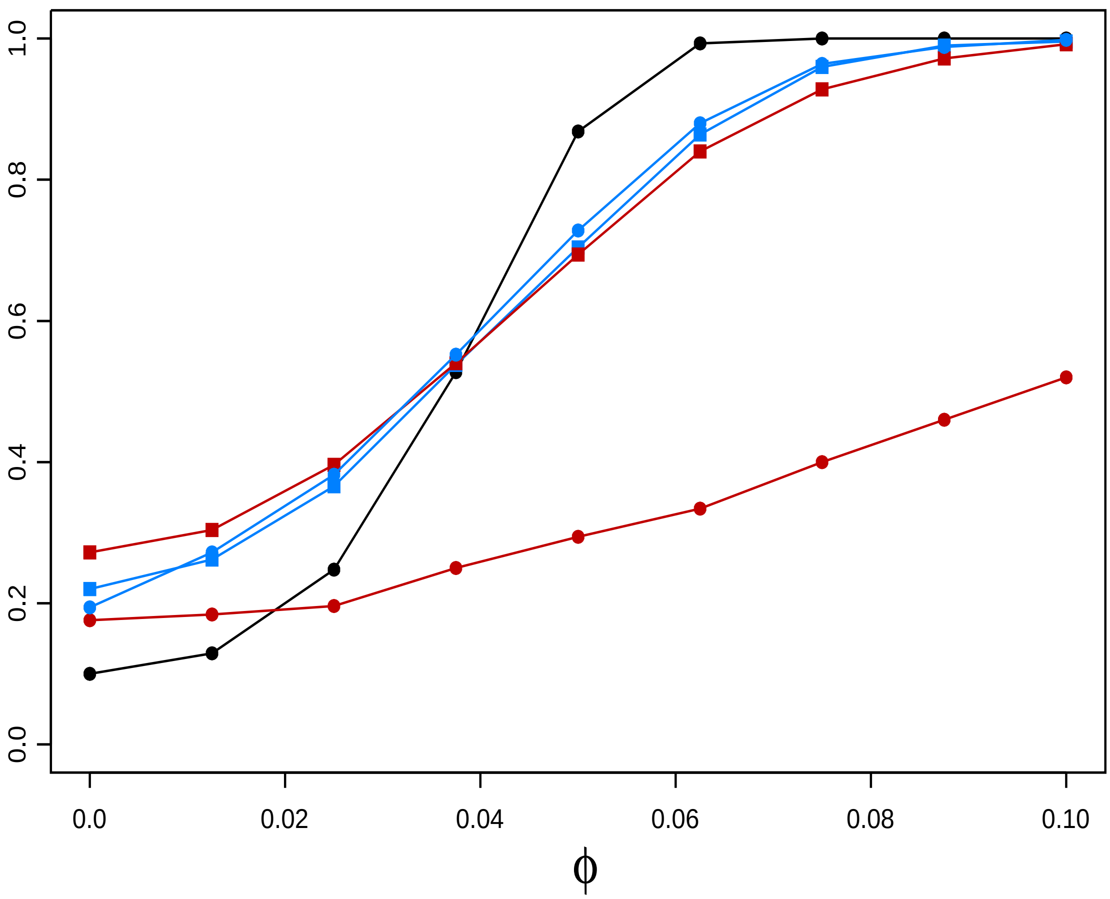

Figure 3: In black the asymptotic power, π_{LS}, in blue the rejection frequencies of both test under the model C_0 and in red the corresponding ones under the contaminated model C_1: squares for the robust test and circles for the classical one.