Testing in generalized partly linear models: A robust approach

Graciela Boente[a], Ricardo Cao[b], Wenceslao González Manteiga[c], Daniela Rodriguez*[,a]

[a] *Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina*

[b] *Departamento de Matemáticas, Universidade da Coruña, A Coruña, Spain*

[c] *Departamento de Estadística e Investigación Operativa, Facultad de Matemáticas, Universidad de Santiago de Compostela, Santiago de Compostela, Spain*

Abstract

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

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

Abbreviated Title: Robust test in generalized semiparametric models.

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

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

1. Introduction

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

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

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

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

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

On the other hand, for generalized partly linear models, hypothesis testing mainly focusses on comparing kerneal based estimators with smoothed parametric estimators. For instance, Härdle *et al.* (1998) considered a test statistic to decide between a linear and a semiparametric model. Their proposal is based on the estimation procedure considered by Severini and Staniswalis (1994) modified to deal with the smoothed and unsmoothed likelihoods. A comparative study of different procedures was performed by Müller (2001) while a different approach was considered in Rodríguez Campos *et al.* (1998).

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

Robust procedures for generalized linear models have been considered among others by Stefanski *et al.* (1986), Künsch *et al.* (1989), Bianco and Yohai (1995), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco *et al.* (2005). The basic ideas from robust smoothing and from robust regression estimation have been adapted to deal with the case of independent observations following a partly linear regression model with $H(t) = t$; we refer to Gao and Shi (1997), He *et al.* (2002) and Bianco and Boente (2004). Moreover, robust tests for a given alternative, under a partly linear regression model were studied in Bianco *et al.* (2006). Besides, a robust approach for testing the parametric form of a regression function versus an omnibus alternative, based on the centered asymptotic rank transformation, was considered by Wang and Qu (2007) when $H(t) = t$ and $\boldsymbol{\beta} = 0$, i.e., under the nonparametric model $y_i = \eta(t_i) + \epsilon_i$.

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

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

2. Preliminaries: The estimation procedure

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

As in Robinson (1988), we will assume that the vector $\mathbf{1}_n$ is not in the space spanned by the column vectors of $(\mathbf{x}_1, \cdots, \mathbf{x}_n)^{\mathrm{T}}$, that is, we do not allow $\boldsymbol{\beta}_0$ to include an intercept so that the model is identifiable, i.e., if $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_1 + \eta_1(t_i) = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_2 + \eta_2(t_i)$ for $1 \leq i \leq n$, then, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and $\eta_1 = \eta_2$. Due to the generality of the semiparametric model (1), identifiability implies that only "slope" coefficients can be estimated.

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

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

Following the ideas of Severini and Staniswalis (1994), Boente *et al.* (2006) defined, for each fixed $\boldsymbol{\beta}$, the function $\eta_{\boldsymbol{\beta}}(t)$ as the minimizer of $S(\boldsymbol{\beta}, a, t)$. Since $S_n(\boldsymbol{\beta}, a, t)$ provides a consistent estimate of $S(\boldsymbol{\beta}, a, t)$, the minimizer in a, $\widehat{\eta}_{\boldsymbol{\beta}}(t)$, of $S_n(\boldsymbol{\beta}, a, t)$ estimates $\eta_{\boldsymbol{\beta}}(t)$. These functions allow the above mentioned authors to define a two–step robust quasi–likelihood estimators of $\boldsymbol{\beta}_0$ and η_0 as $\widehat{\boldsymbol{\beta}} = \operatorname{argmin}_{\boldsymbol{\beta}} S_n(\boldsymbol{\beta}, \eta_{\boldsymbol{\beta}}, t)$ and $\widehat{\eta}(t) = \widehat{\eta}_{\widehat{\boldsymbol{\beta}}}(t)$, respectively. Boente and Rodriguez (2010) introduced a new family of estimators of $\boldsymbol{\beta}_0$ and η_0 that improve the computational results. Both proposals provide robust root-n consistent estimators of the regression parameter $\boldsymbol{\beta}$.

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

3. Test statistics

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

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

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

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

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}, \widehat{\alpha}_{\mathrm{H}_0}, \widehat{\gamma}_{\mathrm{H}_0}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \alpha\in\mathbb{R}, \gamma\in\mathbb{R}}{\operatorname{argmin}} L_n(\boldsymbol{\beta}, \alpha, \gamma). \tag{2}$$

To test H_0, a natural approach is to compare the predicted values $\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_i)$ with those obtained under the null hypothesis, $\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \widehat{\alpha}_{\mathrm{H}_0} + \widehat{\gamma}_{\mathrm{H}_0} t_i$. However, as it is well known, in nonparametric and semiparametric models, due to the bias of the kernel estimator of $S(\boldsymbol{\beta}, a, t)$, the smoothing bias of $\widehat{\eta}(t)$ is non-negligible, even under the linear hypothesis H_0, see, for instance, Härdle and Mammen (1993) and Härdle, *et al.* (1998) for a discussion, when considering the classical estimators. For that reason, a simple comparison between both estimators may be misleading and conduct wrong conclusions. To solve this problem, Härdle, *et al.* (1998) introduced a smoothing bias to $\widehat{\alpha}_{\mathrm{H}_0} + \widehat{\gamma}_{\mathrm{H}_0} t$ to compensate that of $\widehat{\eta}(t)$. It is worth noting that the smoothed estimators obtained under the null hypothesis may not belong to family of linear functions. However, they provide consistent estimators under the parametric model.

To define *smoothed* estimators under the null hypothesis, consider the pseudo–observations $\widetilde{y}_i$ corresponding to the parametric fit of the conditional expectation under the null hypothesis, that is, $\widetilde{y}_i = H\left(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \widehat{\alpha}_{\mathrm{H}_0} + \widehat{\gamma}_{\mathrm{H}_0} t_i\right)$ and denote $\widetilde{\Psi}(\mu, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a) = \mathbb{E}\left(\Psi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right)|(\mathbf{x}, t)\right)$ where the conditional expectation is taken when $y|(\mathbf{x}, t) \sim F(\cdot, \mu)$.

The function $\widehat{\eta}_{\mathrm{H}_0}$ is defined as follows. Since the pseudo–observations will not have outliers, in the sense of large Pearson residuals, but only leverage points could appear, it is quite natural to define $\widehat{\eta}_{\mathrm{H}_0}(t)$ as the value solving $\sum_{i=1}^n W_i(t)\widetilde{\Psi}(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + a) w_1(\mathbf{x}_i) = 0$, or equivalently as the value $\widehat{\eta}_{\mathrm{H}_0}(t) = \operatorname{argmin}_{a\in\mathbb{R}} \widetilde{S}_n(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}, a, t)$, where $\widetilde{S}_n(\boldsymbol{\beta}, a, t) = \sum_{i=1}^n W_i(t)\widetilde{\rho}\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + a\right) w_1(\mathbf{x}_i)$, with $\left(\partial\widetilde{\rho}(\mu, a)\right)/\partial a = \widetilde{\Psi}(\mu, a)$. Note that under mild conditions $\widetilde{\rho}(\mu, a) = \mathbb{E}\left(\rho\left(y, a\right)|(\mathbf{x}, t)\right)$ where the conditional expectation is taken when $y|(\mathbf{x}, t) \sim F(\cdot, \mu)$.

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

$$\mathcal{T}_1 = -2\sum_{i=1}^n Q\left(H\left(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{\eta}(t_i)\right), H\left(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \widehat{\eta}_{\mathrm{H}_0}(t_i)\right)\right) w_2(\mathbf{x}_i) w(t_i)$$

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

4. Asymptotic behavior

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

We will need the following set of assumptions

A1. The density f of t_1 is bounded on $\mathcal{T}$, twice continuously differentiable in the interior of $\mathcal{T}$ with bounded derivatives

A2. $\inf_{t\in[0,1]} f(t) > 0$

A3. η_0 is twice continuously differentiable in the interior of $\mathcal{T}$ with bounded derivatives on $\mathcal{T}$.

A4. $r(t,\tau) = \mathbb{E}_0\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_1)|t_1 = \tau\right)$ is uniformly continuous in the interior of $\mathcal{T}$ and bounded in $\mathcal{T}$.

A5. The functions $\mathrm{v}_0(\tau) = \mathbb{E}_0\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\tau)\right) w_1(\mathbf{x}_1)|t_1 = \tau\right)$ and $\mathrm{v}_1(\tau) = \mathbb{E}_0\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\tau)\right) \mathbf{x}_1 w_1(\mathbf{x}_1)|t_1 = \tau\right)$ are uniformly continuous in the interior of $\mathcal{T}$ and $\mathcal{I}_{\mathrm{v}_0} = \inf_{t\in[0,1]} |\mathrm{v}_0(\tau)| > 0$.

A6. Ψ, χ, χ_1, w, w_j and $\psi_j(\mathbf{x}) = \mathbf{x}w_j(\mathbf{x})$ are bounded functions for $j = 1, 2$.

A7. K is a function of bounded variation with compact support $[0,1]$ and it satisfies $\int K(u)du = 1$ and $\int u\, K(u)du = 0$.

A8. The bandwidth sequence satisfies $nh_n^3/\log(n) \to \infty$ and $n^{\frac{1}{2}} h_n^4 \log(n) \to 0$.

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

a) $\mathcal{G} = \{g(y,\mathbf{x},u) = \chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + a\right) w_1(\mathbf{x}) - \mathbb{E}_0\left(\chi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + a\right) w_1(\mathbf{x}_1)|t_1 = u\right),\ a \in \mathbb{R}\}$*, has covering number* $N(\epsilon, \mathcal{G}, L^1(\mathbb{Q})) \le A\epsilon^{-W}$*, for any probability* $\mathbb{Q}$ *and* $0 < \epsilon < 1$.

b) $\psi_{1,2}(\mathbf{x}) = w_1(\mathbf{x})\|\mathbf{x}\|^2$ *is bounded or* $\sup_{t\in\mathcal{T}} \mathbb{E}_0\left(\psi_{1,2}(\mathbf{x})|t\right) < \infty$.

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

$$c_{1,\Psi} = \mathbb{E}\left(w(t_1)\mathbb{E}\left[w_2(\mathbf{x}_i)\frac{H'\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)^2}{V\left(H\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)\right)}\Big| t_1\right]\mathbb{E}\left[w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)\Big|t_1\right]\mathrm{v}_0(t_1)^{-2}f(t_1)^{-1}\right)$$

$$c_{2,\Psi} = \mathbb{E}\left(\left[\mathbb{E}\left\{w_2(\mathbf{x}_i)\frac{H'\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)^2}{V\left(H\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)\right)}\Big| t_1\right\}\right]^2\left[\mathbb{E}\left\{w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)\Big|t_1\right\}\right]^2\frac{w^2(t_1)}{\mathrm{v}_0(t_1)^4 f(t_1)}\right)$$

$$\sigma^2(\mathbf{x}_0,t_0) = \mathbb{E}\left\{\left[\Psi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)\right) - \mathbb{E}_0\left(\Psi\left(y_1, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)\right)|(\mathbf{x}_1,t_1)\right)\right]^2 |(\mathbf{x}_1,t_1) = (\mathbf{x}_0,t_0)\right\}.$$

Remark 4.1. When considering the canonical exponential family described in the Introduction $V\left(H\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)\right)\right) = H'\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_1)\right)$ and so

$$c_{1,\Psi} = \mathbb{E}\left(w(t_1)\mathbb{E}\left[w_2(\mathbf{x}_1)H'\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)\Big|t_1\right]\mathbb{E}\left[w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)\Big|t_1\right]\mathrm{v}_0(t_1)^{-2}f(t_1)^{-1}\right)$$

$$c_{2,\Psi} = \mathbb{E}\left(w^2(t_1)\left[\mathbb{E}\left\{w_2(\mathbf{x}_1)H'\left(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_1\right)\Big|t_1\right\}\right]^2\left[\mathbb{E}\left\{w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)\Big|t_1\right\}\right]^2\frac{1}{\mathrm{v}_0(t_1)^4 f(t_1)}\right).$$

5. A Montecarlo test

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

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

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

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

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

Step B1 Given a sample $\{(y_i, \mathbf{x}_i, t_i)\}_{1\le i\le n}$ compute the estimators $(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}, \widehat{\alpha}_{\mathrm{H}_0}, \widehat{\gamma}_{\mathrm{H}_0})$ as in (2). Define

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

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

Step B3 Define $R_n^\star = R_n^\star(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}, \widehat{\mathrm{v}}_0, \widehat{\eta}_{H_0})$ with

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

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

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

6. Monte Carlo study

This section contains the results of a simulation study conducted with the aim of comparing the performance of the proposed testing procedure with the classical one. We consider a logistic partly linear model. The robust estimators correspond to those controlling large values of the deviance and they are computed using the score function defined in Croux and Haesbroeck (2002) with tuning constant $c = 0.5$. The weight functions w_1 and w_2 used to control high leverage points are taken as the Tukey's biweight function with tuning constant $c = 4.685$. To be more precise, since $x_i \in \mathbb{R}$, we define $w_1^2(x_i) = w_2^2(x_i) = \left(1 - [(x_i - M_n)/4.685]^2\right)^2$ when $|x_i - M_n| \leq 4.685$ and 0 otherwise, with M_n the median of x_i. The central model denoted C_0 in the figures corresponds to a logistic model where $x_i \sim \mathcal{U}(-1,1)$ and $t_i \sim \mathcal{U}(0,1)$, independent each other. On the other hand, the responses are such that $y_i|(x_i,t_i) \sim Bi(1, p(x_i,t_i))$ with $\log\left(p(x,t)/\left(1 - p(x,t)\right)\right) = \beta_0 x + \eta_0(t,\Delta)$, with $\beta_0 = 2$, $\eta_0(t,\Delta) = (t - 0.5) + \Delta\cos(6\pi(t - 0.5))$, that is, $H(u) = \exp(u)/(1 + \exp(u))$. The value $\Delta = 0$ corresponds to the null hypothesis, $H_0 : \eta_0 \in \{\alpha + \gamma\ t,\ \alpha \in \mathbb{R}, \gamma \in \mathbb{R}\}$, while as alternatives we choose a grid of 10 equally spaced values of $\Delta \in [0.2, 2.0]$. We performed $NR = 1000$ replications of samples of size $n = 200$ with bandwidth $h = 0.1$ and $N_{boot} = 5000$ bootstrap samples. The Epanechinikov kernel $K(t) = (3/4)(1 - t^2)I_{[-1,1]}(t)$ was selected for the smoothing procedure.

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

a) b)

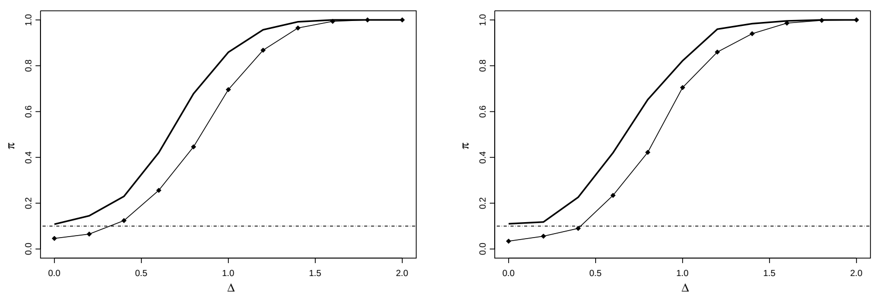

Figure 1: Frequency of rejection π of the asymptotic test, plotted with filled diamonds, and the Montecarlo test plotted with a solid line. a) Classical test b) Robust test.

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

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

$y_{i,new} \sim Bi(1, p_{i,new})$. Then, $(y_{i,c}, x_{i,c}) = (y_i, x_i)$ if $u_i \leq 0.90$ and $(y_{i,c}, x_{i,c}) = (y_{i,new}, x_i)$ if $u_i > 0.90$.

Figures 2 and 3 give the frequency of rejection both for the classical and robust procedure for the contaminated samples. Figure 2 reports the frequencies of rejection for both the asymptotic and Montecarlo procedure, on the other hand, only the results for the Montecarlo test are reported for C_2 since the asymptotic ones behave similarly. The results show that the classical test seem to be quite insensitive to high leverage points if the model is adequate, while its power is sensitive to a misleading model.

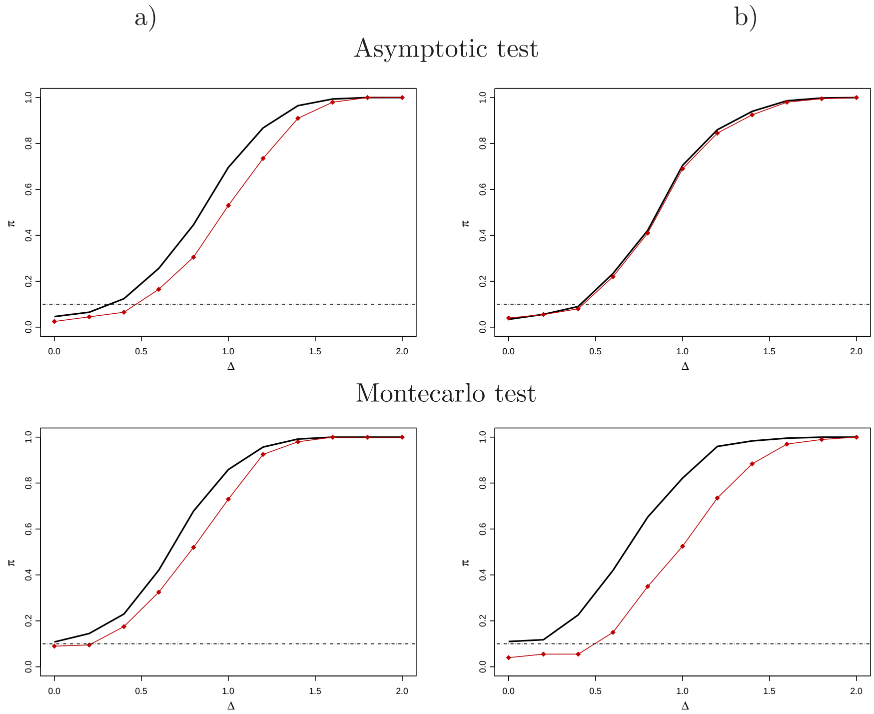

Figure 2: Frequency of rejection π of the asymptotic and Montecarlo test, under C_0 in solid lines and under C_1 in lines with diamonds. a) Classical test b) Robust test

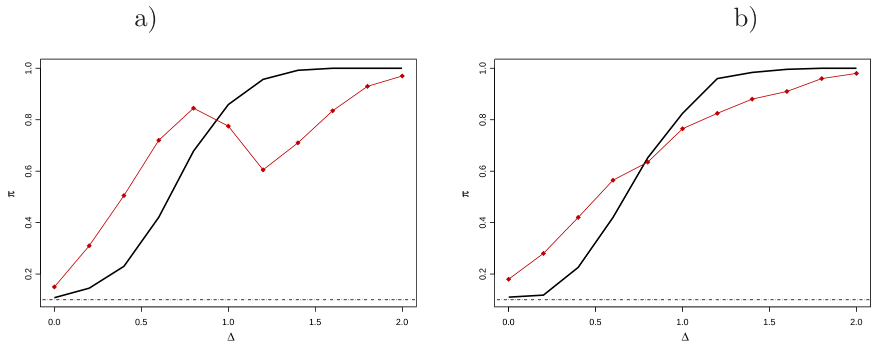

Figure 3: Frequency of rejection π of the Montecarlo test, under C_0 in solid lines and under C_2 in lines with diamonds. a) Classical test b) Robust test

P. Appendix. Proofs

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

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

Proof. Using that χ and ψ_1 are bounded functions it is easy to see that

$$\sup_{\substack{t\in[2h_n,1-2h_n]\\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\rho_n}}\left|S_n^{(0,1)}(\boldsymbol{\beta},\eta_0(t),t)-S_n^{(0,1)}(\boldsymbol{\beta}_0,\eta_0(t),t)\right|\le\|\chi\|_\infty\|\psi_1\|_\infty\rho_n\sup_{t\in[2h_n,1-2h_n]}\sum_{i=1}^n|W_{0,i}(t)|=O_p\left(\rho_n\right).$$

Note that $\mathbb{E}_0\left(\Psi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_i)\right)w_1(\mathbf{x}_i)|t_i\right)=0$. Denote

$$\begin{aligned} v_{0,n}^1(t) &= \sum_{i=1}^n W_{0,i}(t)\left[\Psi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_i)-\mathbb{E}_0\left(\Psi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_i)|t_i\right)\right]\\ A_n^1(t) &= \sum_{i=1}^n W_{0,i}(t)\mathbb{E}_0\left(\Psi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_i)|t_i\right)\end{aligned}$$

with $W_{0,i}(t)$ defined in (**??**). It will be enough to show that

$$\sup_{t\in[2h_n,1-2h_n]}\left|v_{0,n}^1(t)\right| = O_p(h_n^2)=O_p\left(\rho_n\sqrt{\log n}\right) \tag{P.1}$$

$$\sup_{t\in[2h_n,1-2h_n]}\left|A_n^1(t)\right| = O_p(h_n^2)=O_p\left(\rho_n\sqrt{\log n}\right) \tag{P.2}$$

We begin by proving (P.2). Define $m(\tau)=\mathbb{E}_0\left(\Psi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}+\eta_0(t)\right)w_1(\mathbf{x}_1)|t_1=\tau\right)$ and $g(u)=m(u)f(u)$. Then, we have that $m(t)=0$ and we can write $A_n^1(t)=\sum_{i=1}^n W_{0,i}(t)m(t_i)$. On the other hand, using that K has bounded support, $\int uK(u)du=0$ and $t\in[2h_n,1-2h_n]$, we get that

$$\mathbb{E}(A_n^1(t))=\frac{1}{h_n}\int K\left(\frac{u-t}{h_n}\right)g(u)du=g(t)\int K\left(u\right)du+h_ng'(t)\int K\left(u\right)udu+h_n^2\int K\left(u\right)u^2g''(\xi_n)du\,,$$

which implies that $\sup_{t\in[2h_n,1-2h_n]}|\mathbb{E}(A_n^1(t))|=O(h_n^2)$. Now, since K is a bounded variation kernel and f and r are bounded, Theorem 37 in Pollard (1984) with $\mathcal{F}_n=\{f_{t,h_n}(u)=K\left((u-t)/h_n\right)\ m(u),\ t\in[2h_n,1-2h_n]\}$, $\delta_n^2=h_n$, $\alpha_n=h_n$ and the fact that $nh_n^3/\log(n)\to\infty$, entails that $\sup_{t\in[2h_n,1-2h_n]}|A_n^1(t)-\mathbb{E}(A_n^1(t))|=o_p(h_n^2)$

Let us prove (P.1). Note that $\mathbb{E}_0\left(v_{0,n}^1(t)\right)=0$. Denote

$$\begin{aligned}\widetilde{\mathcal{F}}_n &= \{f_{t,h_n}(y,\mathbf{x},u)=(2\|K\|_\infty\|\Psi\|_\infty\|w_1\|_\infty)^{-1}K\left(\frac{u-t}{h_n}\right)\left[\Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x})\right.\\ &- \left.\mathbb{E}_0\left(\Psi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_1)|t_1=u\right)\right],\quad t\in[2h_n,1-2h_n]\}\,.\end{aligned}$$

Using that χ and η_0 are bounded, we obtain that the family of functions

$$\mathcal{G}_n = \{g_t(y,\mathbf{x},u) = \Psi\left(y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right) w_1(\mathbf{x}) - \mathbb{E}_0\left(\Psi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right) w_1(\mathbf{x}_1)|t_1=u\right), \quad t\in[2h_n,1-2h_n]\}\,,$$

has covering number $N(\epsilon,\mathcal{G}_n,L^1(\mathbb{Q}))\le A\epsilon^{-W}$, for any probability $\mathbb{Q}$ and $0<\epsilon<1$ which together with the fact that K has bounded variation and Problem 27 in Pollard (1984) allow to conclude that $N(\epsilon,\widetilde{\mathcal{F}}_n,L^1(\mathbb{Q}))\le B\epsilon^{-C}$, for some constants B and C not depending on n. On the other hand, $|f_{t,h_n}(y,\mathbf{x},u)|\le 1$ and $\mathrm{V{\small AR}}(f_{t,h_n}(y_1,\mathbf{x}_1,t_1))\le h_n\|K\|_\infty^{-2}\|f\|_\infty\int K^2(u)du$. Therefore, Theorem 37 in Pollard (1984) with $\delta_n^2=h_n$, $\alpha_n=h_n$ and the fact that $nh_n^3/\log(n)\to\infty$, concludes the proof of (P.1). $\square$

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

Proof. Using that $S_n^{(0,1)}(\boldsymbol{\beta},\widehat{\eta}_{\boldsymbol{\beta}}(t),t)=0$, a Taylor's expansion of order one, leads to

$$\widehat{\eta}_{\boldsymbol{\beta}}(t)-\eta_0(t) = -S_n^{(0,1)}(\boldsymbol{\beta},\eta_0(t),t)\left[\frac{\partial}{\partial a}S_n^{(0,1)}(\boldsymbol{\beta},a,t)\Big|_{a=\xi(t)}\right]^{-1}$$

with $\xi(t)$ an intermediate point between $\widehat{\eta}_{\boldsymbol{\beta}}(t)$ and $\eta_0(t)$. Note that

$$\frac{\partial}{\partial a}S_n^{(0,1)}(\boldsymbol{\beta},a,t) = \sum_{i=1}^n W_{0,i}(t)\chi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}+a\right)w_1(\mathbf{x}_i).$$

Using that χ_1 and ψ_1 are bounded functions analogous arguments to those given in Lemma P.1, allow to show that

$$\sup_{\substack{t\in[2h_n,1-2h_n],a\in\mathbb{R}\\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\rho_n}}\left|\sum_{i=1}^n W_{0,i}(t)\left[\chi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}+a\right)-\chi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+a\right)\right]w_1(\mathbf{x}_i)\right| = O_p(\rho_n)$$

and

$$\sup_{\substack{t\in[2h_n,1-2h_n]\\ a\in\mathbb{R}}}\left|\sum_{i=1}^n W_{0,i}(t)\left[\chi\left(y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+a\right)w_1(\mathbf{x}_i)-\mathbb{E}_0\left(\chi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+a\right)w_1(\mathbf{x}_1)|t_1=t_i\right)\right]\right| = o_p(1)$$

$$\sup_{\substack{t\in[2h_n,1-2h_n]\\ a\in\mathbb{R}}}\left|\sum_{i=1}^n W_i(t)\mathbb{E}_0\left(\chi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+a\right)w_1(\mathbf{x}_1)|t_1=t_i\right)-\mathbb{E}_0\left(\chi\left(y_1,\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+a\right)w_1(\mathbf{x}_1)|t_1=t\right)\right| = o_p(1)$$

which together with **A2** and **A5** entail that

$$\inf_{\substack{t\in[2h_n,1-2h_n],a\in\mathbb{R}\\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\rho_n}}\left|\frac{\partial}{\partial a}S_n^{(0,1)}(\boldsymbol{\beta},a,t)\Big|_{a=\xi(t)}\right| > \inf_{t\in[0,1]} f(t)\,\mathcal{I}_{\mathrm{V}_0}/2\,,$$

with probability converging to 1. The conclusion follows now from Lemma P.1. $\square$

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

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

$$\widehat{\widehat{\eta}}(t) = \eta_0(t) - \{\mathrm{v}_0(t) f(t)\}^{-1} \sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i) \Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) \tag{P.3}$$

$$\widehat{R}(\boldsymbol{\beta},t) = \mathrm{v}_0(t)^{-1}\mathrm{v}_1(t)^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) . \tag{P.4}$$

PROOF. $\widehat{\eta}_{\boldsymbol{\beta}}(t)$ satisfies $S_n^{(0,1)}(\boldsymbol{\beta}, \widehat{\eta}_{\boldsymbol{\beta}}(t), t) = 0$ and so, using a Taylor expansion, we get

$$\begin{aligned}
0 &= \sum_{i=1}^n W_{0,i}(t)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + \widehat{\eta}_{\boldsymbol{\beta}}(t)\right) w_1(\mathbf{x}_i) = \sum_{i=1}^n W_{0,i}(t)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i) \\
&+ \sum_{i=1}^n W_{0,i}(t)\chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right] \\
&+ \frac{1}{2}\sum_{i=1}^n W_{0,i}(t)\chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right]^2
\end{aligned}$$

with $\boldsymbol{\beta}^*$ and $\eta^*(t)$ intermediate points. Then, we have that

$$\begin{aligned}
\mathrm{v}_{0,n}(t) f_n(t)\left(\widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right) &= -\sum_{i=1}^n W_{0,i}(t)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i) - f_n(t)\, \mathrm{v}_{1,n}(t)^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) \\
&- \frac{1}{2}\sum_{i=1}^n W_{0,i}(t)\chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right]^2
\end{aligned}$$

with $\mathrm{v}_{0,n}(t) = \sum_{i=1}^n W_i(t)\chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i)$, $\mathrm{v}_{1,n}(t) = \sum_{i=1}^n W_i(t)\chi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i)\mathbf{x}_i$ and $f_n(t) = \sum_{i=1}^n W_{0,i}(t)$, which implies that

$$\begin{aligned}
\widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t) &= -\mathrm{v}_{0,n}(t)^{-1}\left[f_n(t)^{-1}\sum_{i=1}^n W_{0,i}(t)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) + \mathrm{v}_{1,n}(t)^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) w_1(\mathbf{x}_i)\right] \\
&- \frac{1}{2}\,\mathrm{v}_{0,n}(t)^{-1}\sum_{i=1}^n W_i(t)\chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right]^2 .
\end{aligned}$$

Using that χ_1 is bounded, $\psi_{1,2}(\mathbf{x}) = w_1(\mathbf{x})\|\mathbf{x}\|^2$ is bounded or $\sup_{t\in\mathcal{T}} \mathbb{E}_0\left(\psi_{1,2}(\mathbf{x})|t\right) < \infty$, Lemma P.2 entails that

$$\begin{aligned}
&\sup_{\substack{t\in[2h_n,1-2h_n]\\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\rho_n}} \left|\sum_{i=1}^n W_i(t)\chi_1\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_0\right) + \widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right]^2\right| \\
\le\ & 2\|\chi_1\|_\infty \sup_{\substack{t\in[2h_n,1-2h_n]\\ \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\le\rho_n}} \left[\sum_{i=1}^n |W_i(t)| w_1(\mathbf{x}_i)\|\mathbf{x}_i\|^2\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 + \|w_1\|_\infty\sum_{i=1}^n |W_i(t)|\left[\widehat{\eta}_{\boldsymbol{\beta}}(t) - \eta_0(t)\right]^2\right] \\
=\ & O_p(\rho_n^2 \log n).
\end{aligned}$$

Then, using that **A2** holds and the fact that

$$\inf_{t\in[2h_n,1-2h_n]} |f_n(t)| \ge \inf_{t\in[2h_n,1-2h_n]} f(t) + \sup_{t\in[2h_n,1-2h_n]} |f_n(t) - f(t)|$$

it will be enough to show that

$$\sup_{t\in[2h_n,1-2h_n]} |\mathrm{v}_{0,n}(t) - \mathrm{v}_0(t)| = O_p(\rho_n \sqrt{\log n}) \tag{P.5}$$

$$\sup_{t\in[2h_n,1-2h_n]} |\mathrm{v}_{1,n}(t) - \mathrm{v}_1(t)| = O_p(\rho_n \sqrt{\log n}) \tag{P.6}$$

$$\sup_{t\in[2h_n,1-2h_n]} |\sum_{i=1}^n W_{0,i}(t) w_1(\mathbf{x}_i)\Psi\left(y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right)| = O_p(\rho_n \sqrt{\log n}) \tag{P.7}$$

$$\sup_{t\in[2h_n,1-2h_n]} |f_n(t) - f(t)| = O_p(h_n^2) = O_p(\rho_n \sqrt{\log n}) \tag{P.8}$$

where $\mathrm{v}_0(t)$ is defined in **A5**. Equation (P.7) follows from Lemma P.1. On the other hand, (P.5), (P.6) and (P.8) follow using analogous arguments to those considered to prove (P.1) and (P.2) together with **A5** and the fact that $\int u K(u)du = 0$ and $\int K(u)du = 1$. Therefore, we have that (**??**) holds concluding the proof. □

The following lemma follows using similar arguments as those considered in Lemma P.3, however, we give its proof for the sake of completeness.

Lemma P.4. *Assume that H_0 holds, i.e., $\eta_0(t) = \alpha_0 + \gamma_0 t$. Denote $\widetilde{y}_{i,0} = H\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \alpha_0 + \gamma_0 t_i\right)$ and $\nu(\tau) = \mathbb{E}\left(w_1(\mathbf{x}_1)\widetilde{\zeta}\left(\widetilde{y}_{1,0}, \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(\tau)\right) H'\left(H^{-1}\left(\widetilde{y}_{1,0}\right)\right) | t_1 = \tau\right)$ where $\widetilde{\zeta}(y,a) = \partial\widetilde{\Psi}\left(y,a\right)/\partial y$. Under **A1** to **A7** if in addition $nh_n^3/\log(n) \to \infty$, we have that*

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

Proof. $\widehat{\eta}_{\mathrm{H}_0}(t)$ is the solution of $\sum_{i=1}^n W_i(t)\widetilde{\Psi}(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + a) w_1(\mathbf{x}_i) = 0$ or equivalently, it satisfies $\sum_{i=1}^n W_{0,i}(t)\widetilde{\Psi}(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \widehat{\eta}_{\mathrm{H}_0}(t)) w_1(\mathbf{x}_i) = 0$ and so, using a Taylor expansion, we get

$$\begin{aligned} 0 &= \sum_{i=1}^n W_{0,i}(t)\widetilde{\Psi}\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} + \widehat{\eta}_{\mathrm{H}_0}(t)\right) w_1(\mathbf{x}_i) = \sum_{i=1}^n W_{0,i}(t)\widetilde{\Psi}\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i) \\ &+ \sum_{i=1}^n W_{0,i}(t)\widetilde{\chi}\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i) \left[\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) + \widehat{\eta}_{\mathrm{H}_0}(t) - \eta_0(t)\right] \\ &+ \frac{1}{2}\sum_{i=1}^n W_{0,i}(t)\widetilde{\chi}_1\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i) \left[\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) + \widehat{\eta}_{\mathrm{H}_0}(t) - \eta_0(t)\right]^2 \end{aligned}$$

with $\widetilde{\chi}(\mu, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a) = \mathbb{E}\left(\chi\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) | (\mathbf{x}, t)\right)$, $\widetilde{\chi}_1(\mu, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a) = \mathbb{E}\left(\chi_1\left(y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a\right) | (\mathbf{x}, t)\right)$, where the conditional expectation are taken when $y|(\mathbf{x}, t) \sim F(\cdot, \mu)$, and $\boldsymbol{\beta}^*$ and $\eta^*(t)$ intermediate points. Then,

$$\begin{aligned} \widetilde{\mathrm{v}}_{0,n}(t) f_n(t) \left(\widehat{\eta}_{\mathrm{H}_0}(t) - \eta_0(t)\right) &= -\sum_{i=1}^n W_{0,i}(t)\widetilde{\Psi}\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t)\right) w_1(\mathbf{x}_i) - f_n(t)\, \widetilde{\mathrm{v}}_{1,n}(t)^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) \\ &- \frac{1}{2}\sum_{i=1}^n W_{0,i}(t)\widetilde{\chi}_1\left(\widetilde{y}_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^* + \eta^*(t)\right) w_1(\mathbf{x}_i) \left[\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) + \widehat{\eta}_{\mathrm{H}_0}(t) - \eta_0(t)\right]^2 \end{aligned}$$

with $\widetilde{\mathrm{v}}_{0,n}(t)=\sum_{i=1}^{n}W_i(t)\widetilde{\chi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_i)$, $\widetilde{\mathrm{v}}_{1,n}(t)=\sum_{i=1}^{n}W_i(t)\widetilde{\chi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)w_1(\mathbf{x}_i)\mathbf{x}_i$ and $f_n(t)=\sum_{i=1}^{n}W_{0,i}(t)$, which implies that

$$\begin{aligned}\widehat{\eta}_{\mathrm{H}_0}(t)-\eta_0(t) &= -\widetilde{\mathrm{v}}_{0,n}(t)^{-1}\left[f_n(t)^{-1}\sum_{i=1}^{n}W_{0,i}(t)w_1(\mathbf{x}_i)\widetilde{\Psi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)+\widetilde{\mathrm{v}}_{1,n}(t)^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}-\boldsymbol{\beta}_0\right)\right]\\ &- \frac{1}{2}\,\widetilde{\mathrm{v}}_{0,n}(t)^{-1}\sum_{i=1}^{n}W_i(t)\widetilde{\chi}_1\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*+\eta^*(t)\right)w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}-\boldsymbol{\beta}_0\right)+\widehat{\eta}_{\mathrm{H}_0}(t)-\eta_0(t)\right]^2.\end{aligned}$$

Using that χ_1 is bounded, $\psi_{1,2}(\mathbf{x})=w_1(\mathbf{x})\|\mathbf{x}\|^2$ is bounded or $\sup_{t\in\mathcal{T}}\mathbb{E}_0\left(\psi_{1,2}(\mathbf{x})|t\right)<\infty$, and that $\|\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}-\boldsymbol{\beta}_0\|\le\rho_n$, similar arguments to those considered in Lemma P.2 allow to show that $\sup_{t\in[2h_n,1-2h_n]}|\widehat{\eta}_{\mathrm{H}_0}(t)-\eta_0(t)|=O_p\left(\rho_n\sqrt{\log n}\right)$ which entails that

$$\begin{aligned}&\sup_{t\in[2h_n,1-2h_n]}\left|\sum_{i=1}^{n}W_i(t)\widetilde{\chi}_1\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*+\eta^*(t)\right)w_1(\mathbf{x}_i)\left[\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}-\boldsymbol{\beta}_0\right)+\widehat{\eta}_{\mathrm{H}_0}(t)-\eta_0(t)\right]^2\right|\\ \le\ & 2\|\chi_1\|_\infty\sup_{t\in[2h_n,1-2h_n]}\left[\sum_{i=1}^{n}|W_i(t)|w_1(\mathbf{x}_i)\|\mathbf{x}_i\|^2\|\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0}-\boldsymbol{\beta}_0\|^2+\|w_1\|_\infty\sum_{i=1}^{n}|W_i(t)|\left[\widehat{\eta}_{\mathrm{H}_0}(t)-\eta_0(t)\right]^2\right]\\ =\ & O_p(\rho_n^2\log n).\end{aligned}$$

Then, using **A2**, the fact that $\inf_{t\in[2h_n,1-2h_n]}|f_n(t)|\ge\inf_{t\in[2h_n,1-2h_n]}f(t)+\sup_{t\in[2h_n,1-2h_n]}|f_n(t)-f(t)|$ and (P.8) it will be enough to show that

$$\sup_{t\in[2h_n,1-2h_n]}|\widetilde{\mathrm{v}}_{0,n}(t)-\mathrm{v}_0(t)| = O_p(\rho_n\sqrt{\log n}) \tag{P.9}$$

$$\sup_{t\in[2h_n,1-2h_n]}|\widetilde{\mathrm{v}}_{1,n}(t)-\mathrm{v}_1(t)| = O_p(\rho_n\sqrt{\log n}) \tag{P.10}$$

$$\sup_{t\in[2h_n,1-2h_n]}\left|\sum_{i=1}^{n}W_{0,i}(t)w_1(\mathbf{x}_i)\widetilde{\Psi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)\right| = O_p(\rho_n\sqrt{\log n}) \tag{P.11}$$

where $\mathrm{v}_0(t)$ is defined in **A5**. Equation (P.11) follows from Lemma P.1. On the other hand, (P.9), (P.10) and (P.8) follow using similar arguments to those considered to prove (P.1) to (P.2) together with the fact that $\int uK(u)du=0$ and $\int K(u)du=1$. Therefore, we have that

$$\sup_{t\in[2h_n,1-2h_n]}|\widehat{\eta}_{\mathrm{H}_0}(t)-\overline{\eta}(t)| = O_p(\rho_n^2\log n)$$

where

$$\overline{\eta}(t)=\eta_0(t)-\mathrm{v}_0(t)^{-1}f(t)^{-1}\sum_{i=1}^{n}W_{0,i}(t)w_1(\mathbf{x}_i)\widetilde{\Psi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)$$

Using that H is twice continuously differentiable and $\widetilde{\Psi}(y,a)$ is twice continuously differentiable with respect to y with derivatives $\widetilde{\zeta}(y,a)=\partial\widetilde{\Psi}\left(y,a\right)/\partial y$ and $\widetilde{\zeta}_1(y,a)=\partial^2\widetilde{\Psi}\left(y,a\right)/\partial y^2$ we get that

$$\begin{aligned}\overline{\eta}(t) &= \eta_0(t)-\mathrm{v}_0(t)^{-1}f(t)^{-1}\sum_{i=1}^{n}W_{0,i}(t)w_1(\mathbf{x}_i)\widetilde{\Psi}\left(\widetilde{y}_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)\\ &= \eta_0(t)-\mathrm{v}_0(t)^{-1}f(t)^{-1}\left[\sum_{i=1}^{n}W_{0,i}(t)w_1(\mathbf{x}_i)\widetilde{\Psi}\left(\widetilde{y}_{i,0},\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t)\right)\right.\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i=1}^{n} W_{0,i}(t) w_1(\mathbf{x}_i) \widetilde{\zeta}\left(\widetilde{y}_{i,0}, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_0(t)\right) H'\left(H^{-1}\left(\widetilde{y}_{i,0}\right)\right)\left(\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) + \widehat{\alpha}_{\mathrm{H}_0} - \alpha_0 + \left(\widehat{\gamma}_{\mathrm{H}_0} - \gamma_0\right) t_i\right) \\
&- \frac{1}{2} \sum_{i=1}^{n} W_{0,i}(t) w_1(\mathbf{x}_i) \left\{\widetilde{\zeta}\left(\widetilde{y}_i^*, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_0(t)\right) H''\left(H^{-1}\left(\widetilde{y}_i^*\right)\right)\right. \\
&+ \left.\widetilde{\zeta}_1\left(\widetilde{y}_i^*, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_0(t)\right)\left[H''\left(H^{-1}\left(\widetilde{y}_i^*\right)\right)\right]^2\right\}\left(\mathbf{x}_i^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{H}_0} - \boldsymbol{\beta}_0\right) + \widehat{\alpha}_{\mathrm{H}_0} - \alpha_0 + \left(\widehat{\gamma}_{\mathrm{H}_0} - \gamma_0\right) t_i\right)^2 \Bigg]
\end{aligned}
$$

with $\widetilde{y}_i^* = H\left(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}^* + \alpha^* + \gamma^* t_i\right)$ concluding the proof. □

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

$$
\begin{aligned}
\widehat{\eta}(t) - \widehat{\eta}_{\mathrm{H}_0}(t) &= \widehat{\widehat{\eta}}(t) + \widehat{R}(\widehat{\boldsymbol{\beta}}, t) - \left(\widehat{\alpha}_{\mathrm{H}_0} + \widehat{\gamma}_{\mathrm{H}_0} t - \alpha_0 - \gamma_0 t\right) \mathrm{v}_0(t)^{-1} \nu(t) - \eta_0(t) \\
&+ \mathrm{v}_0(t)^{-1} f(t)^{-1} \sum_{i=1}^{n} W_{0,i}(t) w_1(\mathbf{x}_i) \widetilde{\Psi}\left(\widetilde{y}_{i,0}, \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + \eta_0(t)\right) + O_p(\rho_n^2 \log n)
\end{aligned}
$$

with $\widehat{\widehat{\eta}}(t)$ and $\widehat{R}(\boldsymbol{\beta}, t)$ defined in (P.3) and (P.4), respectively. Hence,

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

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

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

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

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

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

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

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

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

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

On the other hand, using that $\mathbb{E}(V_{\ell,i}|(\mathbf{x}_\ell,t_\ell,\mathbf{x}_i,t_i))=0$ and $\mathbb{E}(V_{i,i}|(\mathbf{x}_i,t_i))=0$ and that $V_{\ell,i}$ and $V_{i,i}$ are conditionally independent, for $\ell\neq i$, we get that $\mathbb{E}(R_2)=0$.

On the other hand, let $Z_i=w(t_i)v_0^{-2}(t_i)f^{-2}(t_i)w_2(\mathbf{x}_i)V_{i,i}(H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_i))^2)/V(H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_i)))$, then, we have that $R_2=(2K(0))/(n^2h)\sum_{i\neq\ell}Z_iK_{h_n}(t_\ell-t_i)V_{\ell,i}$, and so,

$$\begin{aligned}
\mathrm{VAR}(R_2) &= \frac{2K^2(0)}{n^4h_n^2}\left\{\sum_{i=1}^n\sum_{\ell\neq i}\sum_{s\neq i}\mathrm{Cov}(Z_iK_{h_n}(t_\ell-t_i)V_{\ell,i},Z_iK_{h_n}(t_s-t_i)V_{s,i})\right.\\
&+ \left.\sum_{i=1}^n\sum_{\ell\neq i}\sum_{j\neq i}\mathrm{Cov}(Z_iK_{h_n}(t_\ell-t_i)V_{\ell,i},Z_jK_{h_n}(t_i-t_j)V_{i,j})\right\}\\
&= \frac{2K^2(0)}{n^4h_n^2}\sum_{\ell\neq 1}\mathrm{Cov}(Z_1K_{h_n}(t_\ell-t_1)V_{\ell,1},Z_1K_{h_n}(t_\ell-t_1)V_{\ell,1})\\
&+ \frac{2K^2(0)}{n^4h_n^2}\sum_{\ell\neq 1}\mathrm{Cov}(Z_1K_{h_n}(t_\ell-t_1)V_{\ell,1},Z_\ell K_{h_n}(t_1-t_\ell)V_{1,\ell})\\
&= \frac{2K^2(0)n(n-1)}{n^4h_n^2}(C_{1,h}+C_{2,h})
\end{aligned}$$

with $C_{1,h}=\mathbb{E}(Z_1^2K_{h_n}^2(t_2-t_1)V_{2,1}^2)$ and $C_{2,h}=\mathrm{Cov}(Z_1K_{h_n}(t_2-t_1)V_{2,1},Z_2K_{h_n}(t_1-t_2)V_{1,2})$. Note that,

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

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

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

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

then, $\mathbb{E}(R_4) = (1/n)\sum_{j\neq 1}\mathbb{E}(W_1 K_{h_n}(t_j - t_1)V_{j,1}^2) = ((n-1)/n)\mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1)\mathbb{E}(V_{2,1}^2|(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2)))$. Using the fact that $\mathbb{E}(V_{2,1}|(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2)) = 0$, we get that

$$\mathbb{E}(V_{2,1}^2|(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) \;=\; \mathrm{VAR}(V_{2,1}|(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) = w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2, \mathbf{x}_1, t_1) = w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2, t_1)\ .$$

Let $R_{4,1} = ((n-1)/n)\mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1)w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2, t_2))$ and $R_{4,2} = ((n-1)/n)\mathbb{E}(W_1 K_{h_n}^2(t_2 - t_1)w_1^2(\mathbf{x}_2)[\sigma^2(\mathbf{x}_2, t_2, t_1) - \sigma^2(\mathbf{x}_2, t_2)])$, then, $\mathbb{E}(R_4) = R_{4,1} + R_{4,2}$. Using that $\sigma^2(\mathbf{x}_2, t_2, t_1)$ is Lipschitz, we obtain that

$$|\sigma^2(\mathbf{x}_2, t_2, t_1) - \sigma^2(\mathbf{x}_2, t_2)| < |t_1 - t_2| < h_n\ . \tag{P.12}$$

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

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

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

Let $a(t_1) = w(t_1)v_0^{-2}(t_1)f^{-2}(t_1)$ and $b(t_1) = \mathbb{E}\left(w_2(\mathbf{x}_i)H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i))^2 \left\{V(H(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)))\right\}^{-1}\Big|\, t_1\right)$, then

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

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

$$\begin{aligned} h_n^{1/2}R_{4,1} &= \frac{n-1}{nh_n^2}h_n^{1/2}\int K^2\left(\frac{t_2 - t_1}{h_n}\right)a(t_1)b(t_1)c(t_2)f(t_1)f(t_2)dt_1dt_2 \\ &= \frac{n-1}{nh_n}h_n^{1/2}\int K^2(u)du\int a(t_2)b(t_2)c(t_2)f^2(t_2)dt_2 + o(1). \end{aligned}$$

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

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

where

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

Finally, we will study the asymptotic behavior of R_3. The expected value of R_3 is equal 0, and so it is enough to study its variance.

$$\mathrm{VAR}(R_3) = \frac{1}{n^4}\sum_{1\leq i,s\leq n}\ \sum_{\substack{j\neq i, j\neq \ell, \ell\neq i \\ m\neq s, m\neq r, r\neq i}} \mathrm{COV}\left(a(t_i)w_2(\mathbf{x}_i)\frac{H'(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + \eta_0(t_i)))}K_{h_n}(t_j - t_i)K_{h_n}(t_\ell - t_i)\ V_{j,i}V_{\ell,i},\right.$$

$$
\begin{aligned}
& a(t_s)w_2(\mathbf{x}_i)\frac{H'(\mathbf{x}_s^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_s))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_s)))}K_{h_n}(t_m-t_s)K_{h_n}(t_r-t_s)V_{m,s}V_{r,s}\Big) \\
= & \frac{n}{n^4}\sum_{s=1}^{n}\sum_{\substack{j\neq 1, j\neq \ell, \ell\neq 1\\ m\neq s, m\neq r, r\neq 1}}\mathbb{E}\left(a(t_1)w_2(\mathbf{x}_i)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}K_{h_n}(t_j-t_1)K_{h_n}(t_\ell-t_1)\ V_{j,1}V_{\ell,1}\times\right. \\
& \left. a(t_s)w_2(\mathbf{x}_i)\frac{H'(\mathbf{x}_s^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_s))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_s)))}K_{h_n}(t_m-t_s)K_{h_n}(t_r-t_s)V_{m,s}V_{r,s}\right) \\
= & \frac{1}{n^3}\sum_{j\neq 1, j\neq \ell, \ell\neq 1}\mathbb{E}\left(a^2(t_1)w_2^2(\mathbf{x}_1)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^4}{V^2(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}K_{h_n}^2(t_j-t_1)K_{h_n}^2(t_\ell-t_1)\times\right. \\
& \left. w_1^2(\mathbf{x}_j)w_1^2(\mathbf{x}_\ell)\sigma^2(\mathbf{x}_j,t_j,t_1)\sigma^2(\mathbf{x}_\ell,t_\ell,t_1)\right) \\
+ & \frac{n-1}{n^3}\sum_{\substack{j\neq 1, j\neq \ell, \ell\neq 1\\ m\neq 1, m\neq r, r\neq 2}}\mathbb{E}\left(a(t_1)a(t_2)w_2(\mathbf{x}_1)w_2(\mathbf{x}_2)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}\frac{H'(\mathbf{x}_2^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2)))}\times\right. \\
& K_{h_n}(t_j-t_1)K_{h_n}(t_\ell-t_1)K_{h_n}(t_m-t_2)K_{h_n}(t_r-t_2)\times \\
& \left.\mathbb{E}(V_{j,1}V_{\ell,1}V_{m,2}V_{r,2}|(\mathbf{x}_j,t_j,\mathbf{x}_\ell,t_\ell,\mathbf{x}_m,t_m,\mathbf{x}_r,t_r,\mathbf{x}_1,t_1,\mathbf{x}_2,t_2))\right) \\
= & \frac{(n-1)(n-2)}{n^3}\mathbb{E}\left(a^2(t_1)w_2^2(\mathbf{x}_1)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^4}{V^2(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}K_{h_n}^2(t_2-t_1)K_{h_n}^2(t_3-t_1)\times\right. \\
& \left. w_1^2(\mathbf{x}_3)w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2,t_2,t_1)\sigma^2(\mathbf{x}_3,t_3,t_1)\right) \\
+ & 2\frac{(n-1)^2(n-2)}{n^3}\mathbb{E}\left(a(t_1)a(t_2)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}w_2(\mathbf{x}_1)w_2(\mathbf{x}_2)\frac{H'(\mathbf{x}_2^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2)))}\times\right. \\
& \left. K_{h_n}(t_3-t_1)K_{h_n}(t_4-t_1)K_{h_n}(t_3-t_2)K_{h_n}(t_4-t_2)\mathbb{E}(V_{3,1}V_{3,2}|(\mathbf{x}_3,t_3,t_2,t_1))\mathbb{E}(V_{4,1}V_{4,2}|(\mathbf{x}_4,t_4,t_2,t_1))\right).
\end{aligned}
$$

Hence, $\mathrm{VAR}(R_3) = A_1 + A_2 + A_3$ where

$$
\begin{aligned}
A_1 = & \frac{(n-1)(n-2)}{n^3}\mathbb{E}\left(a^2(t_1)w_2^2(\mathbf{x}_1)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^4}{V^2(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}K_{h_n}^2(t_2-t_1)K_{h_n}^2(t_3-t_1)\times\right. \\
& \left. w_1^2(\mathbf{x}_3)w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2,t_2)\sigma^2(\mathbf{x}_3,t_3)\right) \\
A_2 = & \frac{(n-1)(n-2)}{n^3}\mathbb{E}\left(a^2(t_1)w_2^2(\mathbf{x}_1)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^4}{V^2(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}K_{h_n}^2(t_2-t_1)K_{h_n}^2(t_3-t_1)\right. \\
& \left. w_1^2(\mathbf{x}_3)w_1^2(\mathbf{x}_2)[\sigma^2(\mathbf{x}_2,t_2,t_1)\sigma^2(\mathbf{x}_3,t_3,t_1)-\sigma^2(\mathbf{x}_2,t_2)\sigma^2(\mathbf{x}_3,t_3)]\right) \\
A_3 = & 2\frac{(n-1)^2(n-2)}{n^3}\mathbb{E}\left(a(t_1)a(t_2)w_2(\mathbf{x}_1)w_2(\mathbf{x}_2)\frac{H'(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}\frac{H'(\mathbf{x}_2^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2))^2}{V(H(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_2)))}\times\right. \\
& \left. K_{h_n}(t_3-t_1)K_{h_n}(t_4-t_1)K_{h_n}(t_3-t_2)K_{h_n}(t_4-t_2)\mathbb{E}(V_{3,1}V_{3,2}|(\mathbf{x}_3,t_3,t_2,t_1))\mathbb{E}(V_{4,1}V_{4,2}|(\mathbf{x}_4,t_4,t_2,t_1))\right).
\end{aligned}
$$

Let $b_H(t_1) = \mathbb{E}\left(w_2^2(\mathbf{x}_1)\frac{H'^4(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1))}{V^2(H(\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_0+\eta_0(t_1)))}|t_1\right)$ and $\sigma^2(t_1) = \mathbb{E}(w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)|t_1)$ thus

$$
h_nA_1 = \frac{h_n}{n}\mathbb{E}\left(a^2(t_1)b_H(t_1)\sigma^2(t_2)\sigma^2(t_3)K_{h_n}^2(t_2-t_1)K_{h_n}^2(t_3-t_1)\right)
$$

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

then, we have that $h_nA_1\to 0$. On the other hand, using (P.12), we get the following bound

$$\begin{aligned}
|A_2| &\le \frac{2}{nh_n^4}\mathbb{E}\left(a^2(t_1)\frac{H'(\mathbf{x}_1^{\mathrm T}\boldsymbol\beta_0+\eta_0(t_1))^4}{V^2(H(\mathbf{x}^{\mathrm T}\boldsymbol\beta_0+\eta_0(t_1)))}K^2((t_2-t_1)/h_n)K^2((t_3-t_1)/h_n)h_n\right)\\
&\le \frac{2}{nh_n^3}\int a^2(t_1)b(t_1)K^2((t_2-t_1)/h_n)K^2((t_3-t_1)/h_n)f(t_1)f(t_2)f(t_3)dt_1dt_2dt_3\\
&\le \frac{2}{nh_n}\int a^2(t_1)b(t_1)K^2(u)K^2(v)f(t_1)f(uh_n+t_1)f(vh_n+t_1)dt_1\,du\,dv = \frac{1}{nh_n}O(1)
\end{aligned}$$

which implies that $h_nA_2\to 0$. Finally, straightforward calculations lead to

$$\begin{aligned}
h_nA_3 &= 2h_n\mathbb{E}\left(a(t_1)a(t_2)b(t_1)b(t_2)c(t_1,t_2,t_3)c(t_1,t_2,t_4)K_{h_n}(t_3-t_1)K_{h_n}(t_4-t_1)K_{h_n}(t_3-t_2)K_{h_n}(t_4-t_2)\right)\\
&= \frac{2}{h_n^3}\int a(t_1)a(t_2)b(t_1)b(t_2)c(t_1,t_2,t_3)c(t_1,t_2,t_4)K((t_3-t_1)/h_n)\\
&\qquad K((t_4-t_1)/h_n)K((t_3-t_2)/h_n)K((t_4-t_2)/h_n)f(t_1)f(t_2)f(t_3)f(t_4)dt_1\,dt_2\,dt_3\,dt_4\\
&= \frac{2}{h_n}\int a(t_1)a(t_2)b(t_1)b(t_2)c(t_1,t_2,uh_n+t_1)c(t_1,t_2,vh_n+t_1)K(u)K(v)K((uh_n+t_1-t_2)/h_n)\\
&\qquad K((v+h_n+t_1-t_2)/h_n)f(t_1)f(t_2)f(uh_n+t_1)f(uh_n+t_1)dt_1\,dt_2\,du\,dv\\
&= \frac{2}{h_n}\int a(h_nz+t_2)a(t_2)b(h_nz+t_2)b(t_2)c(h_nz+t_2,t_2,uh_n+h_nz+t_2)c(h_nz+t_2,t_2,vh_n+h_nz+t_2)\\
&\times\quad K(u)K(v)K((uh_n+h_nz+t_2)/h_n)K((v+h_n+h_nz+t_2)/h_n)f(h_nz+t_2)\\
&\times\quad f(t_2)f(uh_n+t_1)f(uh_n+t_1)dz\,dt_2\,du\,dv
\end{aligned}$$

and so, h_nA_3 converges to $2\int a^2(t_2)b^2(t_2)c^2(t_2,t_2,t_2)K(u)K(v)K(u+z)K(v+z)f^4(t_2)dt_2du\,dv\,dz = 2\,\mathbb{E}(a^2(t)b^2(t)c^2(t,t,t)f^3(t))\int[K*K(u)]^2du$.

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

$$\mathbb{E}(a^2(t_2)b^2(t_2)c^2(t_2,t_2,t_2)f^3(t_2)) = \mathbb{E}(a^2(t_2)b^2(t_2)\mathbb{E}(w_1^2(\mathbf{x}_2)\sigma^2(\mathbf{x}_2,t_2)|t_2)f^3(t_2))$$

$$= \mathbb{E}\left(\frac{w^2(t_1)}{v_0^4(t_1)f(t_1)}\left[\mathbb{E}\left(w_2(\mathbf{x}_1)\frac{H'(\mathbf{x}_1^{\mathrm T}\boldsymbol\beta_0+\eta_0(t_1))}{V(H(\mathbf{x}_1^{\mathrm T}\boldsymbol\beta_0+\eta_0(t_1)))}|t_1\right)\right]^2\mathbb{E}(w_1^2(\mathbf{x}_1)\sigma^2(\mathbf{x}_1,t_1)|t_1)\right) = c_{2,\Psi}\ ,$$

concluding the proof. □

Acknowledgements

This research was partially supported by Grants 20020100100276 and 20020100300057 from the Universidad de Buenos Aires, PIP 112-200801-00216 from CONICET, PICT 0821 from ANPCYT at Argentina and also by Spanish Grant MTM2008-03010/MTM of the Ministerio Español de Ciencia e Innovación and XUNTA Grant PGIDIT07PXIB207031PR. This work began while Graciela Boente and Daniela Rodriguez were visiting the Departamento de Estadística e Investigación Operativa from the Universidad de Santiago de Compostela.

References

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

Bianco, A., Boente, G. and Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scand. J. Statist.*, **33**, 435-450.

Bianco, A. García Ben, M. and Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canad. J. Statist.*, **33**, 1-18.

Bianco, A. and Yohai, V. (1995). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17-34. Springer–Verlag, New York.

Boente, G., Cao, R., González–Manteiga, W. and Rodriguez, D. (2012). Testing in generalized partly linear models: A robust approach. Technical Report, available at `http://www.ic.fcen.uba.ar/preprints.pdf`

Boente, G., He, X. and Zhou, J. (2006). Robust Estimates in Generalized Partially Linear Models. *Ann. Statist.*, **34**, 2856-2878.

Boente, G. and Rodriguez, D. (2010). Robust inference in generalized partially linear models. *Comp. Statist. Data Anal.*, **54**, 2942-2966.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *J. Amer. Statist. Assoc.*, **96**, 1022-1030.

Croux, C. and Haesbroeck, G. (2002). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273-295.

Diebolt, J. (1995). A nonparametric test for the regression function: Asymptotic theory. *J. Statist. Plan. Inference*, **44**, 1-17.

Gao, J. (1997). Adaptive parametric test in a semiparametric regression model. *Comm. Statist. Theory Methods*, **26**, 787-800.

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

González–Manteiga, W. and Aneiros–Pérez, G. (2003). Testing in partial linear regression models with dependent errors. *J. Nonparametr. Statist.*, **15**, 93-111.

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

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926-1947.

Härdle, W., Mammen, E. and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc.*, **93**, 1461-1474.

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

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

Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.* **89**, 897-904.

Kojadinovic, I. and Yan, J. (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, **21**, 17-30.

Koul, H. L. and Song, W. (2008). Regression model checking with Berkson measurement errors. *J. Statist. Plan. Inference*, **138**, 1615–1628.

Koul, H. L. and Song, W. (2010). Model checking in partial linear regression models with Berkson measurement errors. *Statistica Sinica*, **20**, 1551-1579.

Koul, H. L. and Stute, W. (1998). Lack of fit tests in regression with non–random design. *Applied Statistical Science*, **3**, 53-69.

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

Mc Cullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. London: Chapman and Hall. 2nd edition.

Müller, M. (2001). Estimation and testing in generalized partial linear modelsA comparative study. *Statistics and Computing*, **11**, 299309.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer–Verlag.

Rémillard, B. and Scaillet, O. (2009) Testing for equality between two copulas. *J. Multivariate Anal.*, **100**, 377386.

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

Rodriguez, D. (2008). *Estimación robusta en modelos parcialmente lineales generalizados*. Ph.D. thesis at the University of Buenos Aires. Available at `http://cms.dm.uba.ar/doc/tesisdanielarodriguez.pdf`

Rodríguez Campos, M. C., González–Manteiga, W. and Cao, R. (1998). Testing the hypothesis of a generalized linear regression model using nonparametric regression estimation. *J. Statist. Plan. Inference*, **67**, 99-122.

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

Stefanski, L., Carroll, R. and Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413-424.

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

Wang, L. and Qu, A. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *J. Amer. Statist. Assoc.*, **102**, 347-358.

Zhu, L. (2005). *Nonparametric Monte Carlo tests and their applications*. Lecture notes in Statistics, **182**.

Zhu, L. and Zhang, H. (2004). Hypothesis testing in mixture regression models. *J. Royal Statist. Soc., Ser. B*, **66**, 3-16.