UNIVERSIDADE DE
SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE
ESTATÍSTICA E INVESTIGACIÓN OPERATIVA

Asymptotic behaviour of robust estimators in partially linear models with missing responses: The effect of estimating the missing probability on the simplified marginal estimators.

A. Bianco, G. Boente, W. González-Manteiga, A. Pérez-González

Report 10-01

Reports in Statistics and Operations Research

Asymptotic behavior of robust estimators in partially linear models with missing responses: The effect of estimating the missing probability on the simplified marginal estimators.

Ana Bianco · Graciela Boente
Wenceslao González–Manteiga · Ana Pérez–González

Abstract In this paper, we consider a semiparametric partially linear regression model where missing data occur in the response. We derive the asymptotic behavior of the robust estimators for the regression parameter and of the weighted simplified location estimator introduced in Bianco, Boente, González–Manteiga and Pérez–González (2010). For the latter, the asymptotic distribution is derived when the missing probability is known and also when it is estimated.

Keywords Fisher–consistency · Kernel Weights · $M-$location Functionals · Missing at Random · Nonparametric Regression · Robust Estimation

A. Bianco
Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET
Ciudad Universitaria, Pabellón 2
Buenos Aires, 1428, Argentina
Tel.: +54-11-45763375
Fax: +54-11-45763375
E-mail: abianco@dm.uba.ar

G. Boente
Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET
Ciudad Universitaria, Pabellón 2
Buenos Aires, 1428, Argentina
Tel.: +54-11-45763375
Fax: +54-11-45763375
E-mail: gboente@dm.uba.ar

W. González–Manteiga
Departamento de Estatística e Investigación Operativa, Facultad de Matemáticas
Universidad de Santiago de Compostela, Campus Sur.
15706 Santiago de Compostela, Spain
Tel.: +34-981-563100 (ext.13204)
Fax: +34-981-597054
E-mail: wenceslao.gonzalez@usc.es

A. Pérez–González
Departamento de Estadística e Investigación Operativa
Universidad de Vigo, Campus Orense.
Campus Universitario As Lagoas s/n
32004 Ourense, Spain.
Tel: +34-988-368768
Fax: +34-988-368923
E-mail: anapg@uvigo.es

Mathematics Subject Classification (2000) MSC 62G35 · MSC 62F12

1 Introduction

Consider the partially linear regression model $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(t_i) + \epsilon_i$, $1 \le i \le n$, where the response $y_i \in \mathbb{R}$ and the covariates $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ are such that $\mathbf{x}_i \in \mathbb{R}^p$, $t_i \in \mathbb{R}$, while the errors ϵ_i are i.i.d., independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ satisfying $E(\epsilon_i) = 0$ and $\mathrm{VAR}(\epsilon_i) < \infty$. Partly linear models are more flexible than standard linear models since they have a parametric and a nonparametric component. They can be a suitable choice when one suspects that the response y linearly depends on $\mathbf{x}$, but that it is nonlinearly related to t. This model has gained attention in recent years. An extensive description of the different results obtained in partly linear regression models can be found in Härdle *et al.* (2000). He *et al.* (2002) considered $M-$type estimates for repeated measurements using $B-$splines, while Bianco and Boente (2004) considered a kernel–based three–step procedure to define robust estimates under the partly linear model.

In practice, some response variables may be missing, by design (as in two-stage studies) or by happenstance. As it is well known, the methods described above are designed for complete data sets and problems arise when missing observations are present. Even if there are many situations in which both the response and the explanatory variables are missing, we will focus our attention on those cases where missing data occur only in the responses. Actually, missingness of responses is very common in opinion polls, market research surveys, mail enquiries, social-economic investigations, medical studies and other scientific experiments. Wang *et al.* (2004) considered inference on the mean of y under regression imputation of missing responses based on the semiparametric regression model $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(t_i) + \epsilon_i$. The estimator of the regression parameter $\boldsymbol{\beta}_0$, introduced by Wang *et al.* (2004), is a least squares regression estimator defined by considering preliminary kernel estimators, of the quantities $E(\delta_1\mathbf{x}_1|t_1 = t)/E(\delta_1|t_1 = t)$ and $E(\delta_1 y_1|t_1 = t)/E(\delta_1|t_1 = t)$, where $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ if y_i is missing. Based on this estimator, estimators of the marginal mean of the responses y are defined using an imputation estimator and a number of propensity score weighting estimators. On the other hand, Wang and Sun (2007) considered estimators of the regression coefficients and the nonparametric function using either imputation, semiparametric regression surrogate or an inverse marginal probability weighted approach. These estimators are based on weighted means of the response variables and so, they are higly sensitive to anomalous data. This fact motivated the need of considering procedures resistant to outliers as those given in Bianco *et al.* (2010), who introduced robust estimators based on bounded score functions together with algorithms to compute them. Moreover, consistency of the marginal estimators was derived therein.

In this paper, we go further and we focus our attention on the asymptotic behavior of the robust estimators of the regression parameter and the marginal location y, say θ, when the response variable has missing observations, but the covariates $(\mathbf{x}^{\mathrm{T}}, t)$ are totally observed. The paper is organized as follows. Section 2 reviews the definition of the robust semiparametric estimators. The consistency and the asymptotic distribution of the regression parameter are derived in Section 3, while the asymptotic distribution of the marginal location estimator is studied in Section 4. For the marginal simplified location estimator, the asymptotic distribution is derived in the situation in which the missing probability is known and also when it is estimated under two different frameworks. In many situations, a parametric model can be assumed for the missingness probability and the influence of estimating the parameters of the model on the distribution of the marginal location estimators needs to be quantified. In particular, if a logistic model is assumed and the parameters are estimated using the maximum likelihood estimator a reduction in the variance is obtained with respect to the estimator computed with the true missingness probability, denoted $p(\mathbf{x}, t)$. On the other hand, if the parameters are estimated robustly we argue that a larger variance can be obtained. Besides, if a kernel estimator is used to estimate $p(\mathbf{x}, t)$, then a reduction of variance is always achieved and so, as recommended in Bianco *et al.* (2010), this estimator should be used whenever it is possible. Technical proofs are left to the Appendix.

2 The robust estimators

Suppose we obtain a random sample of incomplete data $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i)$, $1 \le i \le n$, of a partially linear model where $\delta_i = 1$ if y_i is observed, $\delta_i = 0$ if y_i is missing and

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + g_0(t_i) + \sigma_0 \epsilon_i \ , \qquad 1 \le i \le n \ , \tag{1}$$

with errors ϵ_i independent, identically distributed with symmetric distribution $F_0(\cdot)$, i.e., we assume that the error's scale equals 1 to identify the parameter σ_0. Moreover, ϵ_i are independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$.

Let $(y, \mathbf{x}^{\mathrm{T}}, t, \delta)$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i)$. As mentioned in the Introduction our aim is to study the asymptotic behavior of the robust estimators of the regression parameter and the marginal location. For that purpose, an ignorable missing mechanism will be imposed by assuming that y is missing at random (MAR), that is, δ and y are conditionally independent given $(\mathbf{x}, t)$, i.e., $P(\delta = 1|(y, \mathbf{x}, t)) = P(\delta = 1|(\mathbf{x}, t)) = p(\mathbf{x}, t)$. For the sake of completeness, we will briefly remind the definition of the estimators.

2.1 Estimators of the regression parameter and regression function

As mentioned in Bianco *et al.* (2010), the estimation of the robust location conditional functional related to each component of $\mathbf{x}_i$ causes no problem since the data set is complete, while that of the response y_i is problematic since there are missing responses. Therein, a profile–likelihood approach was considered by combining the $M-$smoothers defined in Boente *et al.* (2009) with robust regression estimators. Let ψ_1 be an odd and bounded score funtion and ρ be a *rho*–function as defined in Maronna, Martin and Yohai (2006, Chapter 2), i.e., a function ρ such that $\rho(x)$ is a nondecreasing function of $|x|$, $\rho(0) = 0$, $\rho(x)$ is increasing for $x > 0$ when $\rho(x) < \|\rho\|_\infty$. If ρ is bounded, it is also assumed that $\|\rho\|_\infty = 1$. We will consider kernel smoothers weights for the nonparametric component which are given by $w_i(\tau) = K((t_i - \tau)/h_n)\, \delta_i \left\{\sum_{j=1}^n K((t_j - \tau)/h_n)\, \delta_j\right\}^{-1}$, with K a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h_n the bandwidth parameter.

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

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

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

$$S_n^{(1)}(a, \boldsymbol{\beta}, \tau) = \sum_{i=1}^n w_i(\tau) \psi_1\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - a}{\widehat{s}_{\boldsymbol{\beta}}}\right) \upsilon(\mathbf{x}_i) \ , \tag{3}$$

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

Step 2. The functional $\boldsymbol{\beta}(F)$ where F is the distribution of $(y, \mathbf{x}^{\mathrm{T}}, t, \delta)$ is defined as $\boldsymbol{\beta}(F) = \mathrm{argmin}_{\boldsymbol{\beta}} H(\boldsymbol{\beta})$, where $H(\boldsymbol{\beta}) = E\left[\delta \rho\left((y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} - g_{\boldsymbol{\beta}}(t))/\sigma_0\right) \upsilon(\mathbf{x})\right]$. Its related estimate is defined as $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_n = \mathrm{argmin}_{\boldsymbol{\beta}} H_n(\boldsymbol{\beta})$, where

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

with $\widehat{\sigma}$ a preliminary estimate of the scale σ_0, i.e., a robust $M-$scale computed using an initial (possibly inefficient) estimate of $\boldsymbol{\beta}$ with high breakdown point.

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

As in any regression model, leverage points in the explanatory variables $\mathbf{x}$, can cause breakdown. To overcome this problem, $GM-$, $S-$ and $MM-$estimators have been introduced, see for instance, Maronna *et al.* (2006). In **Step 2**, a score function ρ combined with a weight υ is introduced to include both families of estimators. This proposal is thus resistant against outliers in the residuals and in the carriers $\mathbf{x}$ as well. In most situations, when considering $MM-$estimators, one chooses $\upsilon(\mathbf{x})\equiv 1$ since $MM-$estimators already control high–leverage points. An algorithm to compute these estimators is described in Bianco *et al.* (2010), where $MM-$estimators with initial $LMS-$estimators combined with $S-$estimators adapted to the partly linear setting are considered.

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

$$H_n^{(1)}(\widehat{\boldsymbol{\beta}})=\sum_{i=1}^n \delta_i\psi\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}-\widehat{g}_{\widehat{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\left(\mathbf{x}_i+\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{g}_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\right)=0\,. \tag{4}$$

2.2 Estimators of the marginal location

Let us denote by θ the marginal location of y, for instance, we are interested in the $M-$location parameter of y solution of $\lambda(a,\varsigma)=E\psi_2\left((y-a)/\varsigma\right)=0$ for all ς, where ψ_2 is an odd and bounded score funtion. When $\psi_2(u)=\mathrm{sg}(u)=I_{(0,\infty)}(u)-I_{(-\infty,0)}(u)$, θ is the median of y. The same score functions ψ_1 and ψ_2 can be considered both in **Step 1** and when computing the marginal parameter estimators defined below.

Denote by $\widehat{\varsigma}$ any robust consistent estimator of the marginal scale ς_0 of the responses y, such as the MAD. To correct the bias caused in the estimation by the missing mechanism, an estimator of the missingness probability needs to be considered. Denote by $p_n(\mathbf{x},t)$ any estimator of $p(\mathbf{x},t)$. The *weighted simplified $M-$estimate* was introduced in Bianco *et al.* (2010) where its consistency was derived. It is the solution, $\widehat{\theta}$, of $U_n(p_n,\widehat{\varsigma},\theta)=0$ with

$$U_n(q,\varsigma,\theta)=\sum_{i=1}^n\frac{\delta_i}{q(\mathbf{x}_i,t_i)}\psi_2\left(\frac{y_i-\theta}{\varsigma}\right)\,. \tag{5}$$

3 Asymptotic behavior of the regression parameter estimators

In this section, we will derive the strong consistency and the asymptotic normality of the regression parameter.

3.1 Consistency of $\widehat{\boldsymbol{\beta}}$

We will assume that $t\in\mathcal{T}\subset\mathbb{R}$, and let $\mathcal{T}_0\subset\mathcal{T}$ be a compact set. For any continuous function $v:\mathcal{T}\to\mathbb{R}$, we will denote $\|v\|_\infty=\sup_{t\in\mathcal{T}}|v(t)|$ and $\|v\|_{0,\infty}=\sup_{t\in\mathcal{T}_0}|v(t)|$. We will need the following set of assumptions

C1. The function ρ and ψ_1 are continuous and bounded. Moreover, the function ρ is Lipschitz and υ is bounded.
C2. The kernel $K:\mathbb{R}\to\mathbb{R}$ is an even, nonnegative, continuous and bounded function, with bounded variation, satisfying $\int K(u)du=1$, $\int u^2K(u)du<\infty$ $|u|K(u)\to 0$ as $|u|\to\infty$.

C3. The bandwidth sequence h_n is such that $h_n \to 0$, $nh_n/\log(n) \to \infty$.
C4. The marginal density f_T of t is a bounded function. Moreover, given any compact set $\mathcal{T}_0 \subset \mathcal{T}$ there exists a positive constant $A_1(\mathcal{T}_0)$ such that $A_1(\mathcal{T}_0) < f_T(\tau)$ for all $\tau \in \mathcal{T}_0$.
C5. The function $S^{(1)}(a, \boldsymbol{\beta}, \tau)$ satisfies the following equicontinuity condition: for any $\epsilon > 0$ there exists $\delta > 0$ such that for any $\tau_1, \tau_2 \in \mathcal{T}_0$ and $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathcal{K}$, a compact set in $\mathbb{R}^p$,

$$|\tau_1 - \tau_2| < \delta \text{ and } \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| < \delta \Rightarrow \sup_{a \in \mathbb{R}} |S^{(1)}(a, \boldsymbol{\beta}_1, \tau_1) - S^{(1)}(a, \boldsymbol{\beta}_2, \tau_2)| < \epsilon \,.$$

C6. The function $S^{(1)}(a, \boldsymbol{\beta}, \tau)$ is continuous, and $g_{\boldsymbol{\beta}}(\tau)$ is a continuous function of $(\boldsymbol{\beta}, \tau)$.

Remark 3.1.1. If the conditional distribution of $\mathbf{x}|t = \tau$ is continuous with respect to τ, the continuity and boundness of ψ_1 stated in **C1** entail that $S^{(1)}(a, \boldsymbol{\beta}, \tau)$ is continuous. Assumption **C3** ensures that for each fixed a and $\boldsymbol{\beta}$ we have convergence of the kernel estimates to their mean, while **C5** guarantees that the bias term converges to 0. Assumption **C4** is a standard condition in semiparametric models. Assumption **C5** is fulfilled under **C1** if the following equicontinuity condition holds: for any $\epsilon > 0$ there exist compact sets $\mathcal{K}_1 \subset \mathbb{R}$ and $\mathcal{K}_p \subset \mathbb{R}^p$ such that for any $\tau \in \mathcal{T}_0$ $P((y, \mathbf{x}) \in \mathcal{K}_1 \times \mathcal{K}_p | t = \tau) > 1 - \epsilon$, which holds for instance if $x_{ij} = \phi_j(t_i) + u_{ij}$, $1 \le i \le n$, $1 \le j \le p$, where ϕ_j are continuous functions and u_{ij} are i.i.d and independent of t_i.

Theorem 3.1. *Let $\mathcal{K} \subset \mathbb{R}^p$ and $\mathcal{T}_0 \subset \mathcal{T}$ be compact sets such that $\mathcal{T}_\delta \subset \mathcal{T}$ where $\mathcal{T}_\delta$ is the closure of a δ neighborhood of $\mathcal{T}_0$. Assume that **C1** to **C6** and the following conditions hold*

i) ψ_1 is of bounded variation
*ii) $\inf_{\boldsymbol{\beta} \in \mathcal{K}} \sigma_{\boldsymbol{\beta}} > 0$ and $\sup_{\boldsymbol{\beta} \in \mathcal{K}} |\widehat{s}_{\boldsymbol{\beta}} - \sigma_{\boldsymbol{\beta}}| \xrightarrow{a.s.} 0$, where $\sigma_{\boldsymbol{\beta}}$ as defined in **Step 1**.*

Then, we have

a) $\sup_{\substack{\boldsymbol{\beta} \in \mathcal{K} \\ a \in \mathbb{R}}} \|S_n^{(1)}(a, \boldsymbol{\beta}, \cdot) - S^{(1)}(a, \boldsymbol{\beta}, \cdot)\|_{0,\infty} \xrightarrow{a.s.} 0$.
b) If, in addition, $S^{(1)}(a, \boldsymbol{\beta}, \tau) = 0$ has a unique root $g_{\boldsymbol{\beta}}(\tau)$, then $\sup_{\boldsymbol{\beta} \in \mathcal{K}} \|\widehat{g}_{\boldsymbol{\beta}} - g_{\boldsymbol{\beta}}\|_{0,\infty} \xrightarrow{a.s.} 0$.

The proof of Theorem 3.1 follows the same arguments of those used in Theorem 3.1 of Boente *et al.* (2006) using the fact that asumption ii) implies that the family of functions $\mathcal{F} = \{f(y, \mathbf{x}) = \psi_1\left((y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + a)/\sigma\right) \upsilon(\mathbf{x}), \boldsymbol{\beta} \in \mathcal{K}, a \in \mathbb{R}, \sigma > 0\}$ has covering number $N\left(\epsilon, \mathcal{F}, L^1(\mathbb{Q})\right) \le A\epsilon^{-W}$, for any probability $\mathbb{Q}$ and $0 < \epsilon < 1$. Besides, the condition that $S^{(1)}(a, \boldsymbol{\beta}, \tau) = 0$ has a unique root is fulfilled if ψ_1 is a nondecresing function and strictly increasing in a neighborhood of 0.

Theorem 3.2. *Let $\widehat{\boldsymbol{\beta}}$ be the minimizer of $H_n(\boldsymbol{\beta})$ where $H_n(\boldsymbol{\beta})$ is defined in **Step 2** with $\widehat{g}_{\boldsymbol{\beta}}$ satisfying $\sup_{\boldsymbol{\beta} \in \mathcal{K}} \|\widehat{g}_{\boldsymbol{\beta}} - g_{\boldsymbol{\beta}}\|_{0,\infty} \xrightarrow{a.s.} 0$ for any compact sets $\mathcal{K} \subset \mathbb{R}^p$ and $\mathcal{T}_0 \subset \mathcal{T}$. If **C1** holds and $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0$, then*

a) $\sup_{\boldsymbol{\beta} \in \mathcal{K}} |H_n(\boldsymbol{\beta}) - H(\boldsymbol{\beta})| \xrightarrow{a.s.} 0$,
b) If, in addition, there exists a compact set $\mathcal{K}_1$ such that $\lim_{m \to \infty} P\left(\bigcap_{n \ge m} \widehat{\boldsymbol{\beta}} \in \mathcal{K}_1\right) = 1$ and $H(\boldsymbol{\beta})$ has a unique minimum at $\boldsymbol{\beta}_0$, then $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$.

We also omit the proof of Theorem 3.2, since it follows as Theorem 3.2 of Boente *et al.* (2006).

Remark 3.2. Theorems 3.1 and 3.2 entail that $\|\widehat{g}_{\widehat{\boldsymbol{\beta}}} - g_0\|_{0,\infty} \xrightarrow{a.s.} 0$, for any compact set $\mathcal{T}_0 \subset \mathcal{T}$, since $g_{\boldsymbol{\beta}}(t)$ is continuous.

3.2 Asymptotic Normality of $\widehat{\boldsymbol{\beta}}$

From now on, $\mathcal{T}$ is assumed to be a compact set. The assumptions **N1** to **N6** under which the resulting estimates are asymptotically normally distributed are detailed in the Appendix.

Theorem 3.3. *Assume that t_1 is a random variable with distribution on a compact set $\mathcal{T}$. Assume that **N1** to **N6** in the Appendix hold and that $\widehat{\sigma} \xrightarrow{p} \sigma_0$, then for any consistent solution $\widehat{\boldsymbol{\beta}}$ of (4), we have*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \sigma_0^2 \mathbf{A}^{-1} \boldsymbol{\Sigma} \mathbf{A}^{-1}\right) ,$$

*where the symmetric matrix $\mathbf{A}$ is defined in **N3** and $\boldsymbol{\Sigma}$ is defined in **N4**.*

It is worth noticing that, when $\upsilon(\mathbf{x}) \equiv 1$, the efficiency of the robust estimator $\widehat{\boldsymbol{\beta}}$ with respect to its linear relative, i.e., the least square estimator, equals $\left[E\psi'(\epsilon)\right]^{-2} E\psi^2(\epsilon)$, which corresponds to the very well known efficiency of any robust location $M-$estimator. This situation includes, in particular, $MM-$estimators and so, the same asymptotic efficiency as in the regression model is obtained in this case.

4 Asymptotic Normality of $\widehat{\theta}$

In this section, we will derive the asymptotic distribution of the weighted simplified $M-$estimate, $\widehat{\theta}$, under different situations, i.e., when the missingness probability is assumed to be known or when it is estimated either parametrically or using a kernel approach. Different asymptotic variances are obtained in each situation. The goal is to see if we can validate theoretically the results observed in the simulation study performed in Bianco *et al.* (2010), i.e., if we can prove that estimating nonparametrically the missingness probability reduces the variance of the estimator. It is worth noticing that our results require consistency of the proposed estimators, i.e., that $\widehat{\theta} \xrightarrow{p} \theta$. Conditions that guarantee strongly consistent estimators are given in Theorem 4.1 of Bianco *et al.* (2010) and include among others, uniform consistency of the missingness probability, i.e., $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p(\mathbf{x},t)| \xrightarrow{a.s.} 0$, smootheness conditions to the score function ψ_2 and the assumption that $\inf_{(\mathbf{x},t)} p(\mathbf{x},t) = A > 0$, which states that some response variables are observed at each neighborhood of $(\mathbf{x},t)$.

Assumptions **NM1** to **NM8** under which the estimators are asymptotically normally distributed are stated in the Appendix. From now on, we will denote by $u = (y-\theta)/\varsigma_0$.

Theorem 4.1. *Let U_n be defined in (5). Assume that **NM1** to **NM3** in the Appendix hold and that $\widehat{\varsigma} \xrightarrow{p} \varsigma_0$. Denote by $\widehat{\theta}^{(1)}$, the solution of $U_n(p,\widehat{\varsigma},\theta) = 0$, i.e., the weighted simplified estimator assuming that the missingness probability is known. If $\widehat{\theta}^{(1)} \xrightarrow{p} \theta$, we have that $\sqrt{n}(\widehat{\theta}^{(1)} - \theta) \xrightarrow{\mathcal{D}} N(0, \upsilon^{(1)})$, where $\upsilon^{(1)} = E\left(\psi_2^2(u)/p(\mathbf{x},t)\right)\left(E\psi_2'(u)\right)^{-2}$.*

Note that in this situation, the efficiency with respect to the classical simplified estimator, i.e., when $\psi_2(u) = u$, is not the efficiency of the location estimator when no missing data are present, since a factor $1/p(\mathbf{x},t)$ depending on the missingness probability appears in the numerator's expectation. Therefore, the efficiency of the estimators depends on the proportion of missing data appearing in the sample.

Theorem 4.2. *Let U_n be defined in (5). Assume that **NM1** to **NM5** in the Appendix hold and that $\widehat{\varsigma} \xrightarrow{p} \varsigma_0$. Moreover, assume that $p(\mathbf{x}_i,t_i) = G(\mathbf{x}_i,t_i,\boldsymbol{\lambda}_0)$, where $\boldsymbol{\lambda}_0 \in \mathbb{R}^q$, and let $p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i) = G(\mathbf{x}_i,t_i,\widehat{\boldsymbol{\lambda}})$, where $\widehat{\boldsymbol{\lambda}}$*

is an estimator of $\boldsymbol{\lambda}$ such that $\widehat{\boldsymbol{\lambda}} \xrightarrow{p} \boldsymbol{\lambda}_0$. Denote by $\widehat{\theta}^{(2)}$, the solution of $U_n(p_{n,\widehat{\boldsymbol{\lambda}}}, \widehat{\varsigma}, \theta) = 0$, i.e., assuming a parametric model for the missingness probability. If $\widehat{\theta}^{(2)} \xrightarrow{p} \theta$, we have that $\sqrt{n}(\widehat{\theta}^{(2)} - \theta) \xrightarrow{\mathcal{D}} N(0, \upsilon^{(2)})$, where $\upsilon^{(2)} = \gamma^2 \left(E\psi_2'(u)\right)^{-2}$ with

$$\begin{aligned}\gamma^2 &= E\left[\frac{\delta}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}\psi_2(u) - \boldsymbol{\eta}(\delta,\mathbf{x},t)^{\mathrm{T}} E\left(\frac{G'(\mathbf{x},t,\boldsymbol{\lambda}_0)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}\psi_2(u)\right)\right]^2 \\ &= E\frac{\psi_2^2(u)}{p(\mathbf{x},t)} + E\left(\psi_2(u)\frac{G'(\mathbf{x},t,\boldsymbol{\lambda}_0)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}\right)^{\mathrm{T}}\left\{\boldsymbol{\Sigma}\, E\left(\psi_2(u)\frac{G'(\mathbf{x},t,\boldsymbol{\lambda}_0)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}\right) - 2E\left[\frac{\delta\psi_2(u)\boldsymbol{\eta}(\delta,\mathbf{x},t)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}\right]\right\}\end{aligned}$$

*and $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ given in **NM5**.*

Remark 4.1. Denote by $F_{\mathrm{L}}(s) = (1+e^{-s})^{-1}$ the logistic distribution function and let us assume that the missingness probablity is given by the logistic model, i.e., that $p(\mathbf{x},t) = F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda}_0)$ and $G(\mathbf{x},t,\boldsymbol{\lambda}) = F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda})$ where $\mathbf{v} = (1, \mathbf{x}^{\mathrm{T}}, t)^{\mathrm{T}}$. Hence, $G'(\mathbf{x},t,\boldsymbol{\lambda}) = F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda})[1 - F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda})]\mathbf{v}$. Moreover, let us assume that $\widehat{\boldsymbol{\lambda}}$ is the maximum likelihood estimator. This estimator can be consider instead of a robust one, such as that defined in Croux and Haesbroeck (2003), if we suspect that no outliers are present in the covariates $\mathbf{x}$ or if we know that $p(\mathbf{x},t)$ only depends on t where no outliers appear, i.e., if in the above model, $\mathbf{v} = (1,t)^{\mathrm{T}}$. This last situation is also included in the sequel just by taking into account the new expression for $\mathbf{v}$. The calculations to be done include in particular, the classical estimators, for which up to our knowledge there are no results regarding the theoretical comparison of the asymptotic variances of the marginal location estimator when the missing probability is known and when it is parametrically estimated. In this situation, we have that

- $G'(\mathbf{x},t,\boldsymbol{\lambda}_0) = p(\mathbf{x},t)[1 - p(\mathbf{x},t)]\mathbf{v}$,
- $\boldsymbol{\eta}(\delta,\mathbf{x},t) = \mathbf{A}_1^{-1}(\delta - p(\mathbf{x},t))\mathbf{v}$ where $\mathbf{A}_1 = Ep(\mathbf{x},t)(1-p(\mathbf{x},t))\mathbf{v}\mathbf{v}^{\mathrm{T}}$, implying that $\boldsymbol{\Sigma} = \mathbf{A}_1^{-1}$,
- $E\left(G'(\mathbf{x},t,\boldsymbol{\lambda}_0)\psi_2(u)/p(\mathbf{x},t)\right) = E\left((1-p(\mathbf{x},t)\psi_2(u)\mathbf{v}\right)$.

Therefore, $\gamma^2 = E\left(\psi_2^2(u)/p(\mathbf{x},t)\right) + \nu$ with

$$\begin{aligned}\nu &= E\left((1-p(\mathbf{x},t))\psi_2(u)\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}_1^{-1}\left\{E\left((1-p(\mathbf{x},t))\psi_2(u)\mathbf{v}\right) - 2E\left[\psi_2(u)\frac{\delta(\delta - p(\mathbf{x},t))}{p(\mathbf{x},t)}\mathbf{v}\right]\right\} \\ &= -E\left((1-p(\mathbf{x},t))\psi_2(u)\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}_1^{-1}E\left((1-p(\mathbf{x},t))\psi_2(u)\mathbf{v}\right)\ ,\end{aligned}$$

where we have used that

$$E\left(\frac{\delta}{p(\mathbf{x},t)}(\delta - p(\mathbf{x},t))|(y,\mathbf{x},t)\right) = E\left(\frac{\delta}{p(\mathbf{x},t)}(\delta - p(\mathbf{x},t))|(\mathbf{x},t)\right) = 1 - p(\mathbf{x},t)\ . \tag{6}$$

Hence, $\nu \le 0$ which entails that $\upsilon^{(2)} \le \upsilon^{(1)}$ and equality holds if and only if $E\left((1-p(\mathbf{x},t))\psi_2(u)\mathbf{v}\right) = \mathbf{0}$ that happens obviously if there are no missing observations.

Remark 4.2. In some situations, the parameters of the logistic model need to be estimated robustly, for instance, if we suspect that high leverage points in the carriers $\mathbf{x}$ are present. We can carry on the robust estimation using, for instance, a weighted maximum likelihood estimator or the estimator defined in Croux and Haesbroeck (2003), i.e., $\widehat{\boldsymbol{\lambda}} = \operatorname{argmin}_{\boldsymbol{\lambda}} \sum_{i=1}^n w(\mathbf{x}_i)\varphi(\mathbf{v}_i^{\mathrm{T}}\boldsymbol{\lambda};\delta_i)$ where $\varphi(s;0) = \varphi(-s;1)$ and $\varphi(s;0) = \rho(-\ln(1 - F_{\mathrm{L}}(s))) + C(F_{\mathrm{L}}(s)) + C(1 - F_{\mathrm{L}}(s)) - C(1)$ and $C(s) = \int_0^s \rho'(-\ln u)du$. The weighted maximum likelihood estimator corresponds to the choice $\rho(s) = s$. Then, using the results in Bianco and Martínez (2009), we have that $\boldsymbol{\eta}(\delta,\mathbf{x},t) = -\mathbf{A}_{1,\mathrm{R}}^{-1}w(\mathbf{x})\Psi(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda}_0;\delta)\mathbf{v}$, where $\Psi(s;0) = \partial\varphi(s;0)/\partial s$, $\Psi(s;1) = -\Psi(-s;0)$ and

$$\mathbf{A}_{1,\mathrm{R}} = E\left\{w(\mathbf{x})\frac{\partial^2}{\partial s^2}\varphi(s;\delta)\Big|_{s=\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda}_0}\mathbf{v}\mathbf{v}^{\mathrm{T}}\right\}\ .$$

Straightforward calculations lead to

$$E\left(\frac{\delta\boldsymbol{\eta}(\delta,\mathbf{x},t)}{p(\mathbf{x},t)}\psi_2\left(u\right)\right) = \mathbf{A}_{1,\mathrm{R}}^{-1}E\left(\psi_2\left(u\right)w(\mathbf{x})\left(1-p(\mathbf{x},t)\right)D(\mathbf{x},t)\mathbf{v}\right)$$
$$\boldsymbol{\Sigma} = \mathbf{A}_{1,\mathrm{R}}^{-1}E\left(w^2(\mathbf{x})\left(1-p(\mathbf{x},t)\right)p(\mathbf{x},t)D^2(\mathbf{x},t)\mathbf{v}\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}_{1,\mathrm{R}}^{-1}\ ,$$

where $D(\mathbf{x},t) = (1-p(\mathbf{x},t))C'(p(\mathbf{x},t)) + p(\mathbf{x},t)C'(1-p(\mathbf{x},t))$. Therefore, $\gamma^2 = E\left(\psi_2^2\left(u\right)/p(\mathbf{x},t)\right) + \nu$ with

$$\nu = E\left((1-p(\mathbf{x},t))\psi_2\left(u\right)\mathbf{v}^{\mathrm{T}}\right)\left\{\boldsymbol{\Sigma}\ E\left((1-p(\mathbf{x},t))\psi_2\left(u\right)\mathbf{v}\right) - 2\mathbf{A}_{1,\mathrm{R}}^{-1}E\left(\psi_2\left(u\right)w(\mathbf{x})\left(1-p(\mathbf{x},t)\right)D(\mathbf{x},t)\mathbf{v}\right)\right\}\ .$$

In particular, if $w(\mathbf{x}) = w^2(\mathbf{x})$ which corresponds to a $0-1$ weight function and $\rho(s) = s$, i.e., when considering the weighted maximum likelihood, we have that $\mathbf{A}_{1,\mathrm{R}} = E\left\{w(\mathbf{x})p(\mathbf{x},t)(1-p(\mathbf{x},t))\mathbf{v}\mathbf{v}^{\mathrm{T}}\right\}$, $D(\mathbf{x},t)\equiv 1$ and so, $E\left(w^2(\mathbf{x})\left(1-p(\mathbf{x},t)\right)p(\mathbf{x},t)D^2(\mathbf{x},t)\mathbf{v}\mathbf{v}^{\mathrm{T}}\right) = \mathbf{A}_{1,\mathrm{R}}$ that implies

$$\begin{aligned}\nu &= E\left((1-p(\mathbf{x},t))\psi_2\left(u\right)\mathbf{v}^{\mathrm{T}}\right)\left\{\mathbf{A}_{1,\mathrm{R}}^{-1}\ E\left((1-p(\mathbf{x},t))\psi_2\left(u\right)\mathbf{v}\right) - 2\mathbf{A}_{1,\mathrm{R}}^{-1}E\left(\psi_2\left(u\right)w(\mathbf{x})\left(1-p(\mathbf{x},t)\right)\mathbf{v}\right)\right\}\\ &= \mathbf{b}^{\mathrm{T}}\mathbf{A}_{1,\mathrm{R}}^{-1}\mathbf{b} - 2\mathbf{b}^{\mathrm{T}}\mathbf{A}_{1,\mathrm{R}}^{-1}\mathbf{b}_w\ ,\end{aligned}$$

where $\mathbf{b} = E\left((1-p(\mathbf{x},t))\psi_2\left(u\right)\mathbf{v}\right)$ and $\mathbf{b}_w = E\left(\psi_2\left(u\right)w(\mathbf{x})\left(1-p(\mathbf{x},t)\right)\mathbf{v}\right)$. Depending on the choice of the weight function w, i.e., on the tuning constant selected to cutoff outliers, the inner product $\mathbf{b}^{\mathrm{T}}\mathbf{A}_{1,\mathrm{R}}^{-1}\mathbf{b}_w$ can be much smaller than the squared norm $\mathbf{b}^{\mathrm{T}}\mathbf{A}_{1,\mathrm{R}}^{-1}\mathbf{b}$, leading to a positive value of ν. In this situation, the variance of the robust marginal location estimator $\widehat{\theta}^{(2)}$ will be larger than that of the estimator $\widehat{\theta}^{(1)}$ computed with the true missingness probability. This fact is consistent with the simulation results obtained in Bianco *et al.* (2010) and opposite to the conclusions obtained when the parameters of the missing probability are estimated using the classical maximum likelihood estimator, leading to a larger loss of eficiency when robust estimators are used.

We will now study the asymptotic distribution of the weighted simplified estimator when the missingness probability is estimated using a kernel estimator

$$p_{n,b_n}(\mathbf{x},t) = \sum_{i=1}^{n} K_1\left(\frac{\mathbf{w}_i-\mathbf{w}}{b_n}\right)\delta_i\left\{\sum_{j=1}^{n} K_1\left(\frac{\mathbf{w}_j-\mathbf{w}}{b_n}\right)\right\}^{-1}\ , \tag{7}$$

where $K_1:\mathbb{R}^{p+1}\to\mathbb{R}$ is a kernel function, $\mathbf{w} = (\mathbf{x}^{\mathrm{T}},t)^{\mathrm{T}}$ and b_n denotes the smoothing parameter.

Theorem 4.3. *Let U_n be defined in (5). Assume that **NM1** to **NM3** and **NM6** to **NM8** in the Appendix hold and that $\widehat{\varsigma}\xrightarrow{p}\varsigma_0$. Let $p_{n,b_n}(\mathbf{x}_i,t_i)$ be the kernel estimator defined in (7). Denote by $\widehat{\theta}^{(3)}$, the solution of $U_n(p_{n,b_n},\widehat{\varsigma},\theta) = 0$, i.e., using the nonparametric estimator of the missingness probability. If $\widehat{\theta}^{(3)}\xrightarrow{p}\theta$, we have that $\sqrt{n}(\widehat{\theta}^{(3)}-\theta)\xrightarrow{\mathcal{D}}N(0,\upsilon^{(3)})$ where $\upsilon^{(3)} = \gamma_{\mathrm{S}}^2\left(E\psi_2'\left(u\right)\right)^{-2}$ and*

$$\gamma_{\mathrm{S}}^2 = E\left(\frac{\delta}{p(\mathbf{w})}\psi_2\left(\frac{y-\theta}{\varsigma_0}\right) - \frac{(\delta-p(\mathbf{w}))}{p(\mathbf{w})}r(\mathbf{w})\right)^2\ ,$$

with $r(\mathbf{w}) = E\left(\psi_2\left(u\right)|\mathbf{w}\right)$.

Remark 4.3. Using (6) and after some algebra, we get that

$$\gamma_{\mathrm{S}}^2 = E\frac{\psi_2^2\left(u\right)}{p(\mathbf{w})} - E\left(\frac{(1-p(\mathbf{w}))}{p(\mathbf{w})}r^2(\mathbf{w})\right)\ .$$

Hence, $\upsilon^{(3)} \le \upsilon^{(1)}$ and so, the marginal location estimator $\widehat{\theta}^{(3)}$ using the kernel estimator for the missing probability is more efficient than $\widehat{\theta}^{(1)}$. Note that both estimators have equal variance if and only if $E\left((1-p(\mathbf{w}))\, r^2(\mathbf{w})/p(\mathbf{w})\right) = \mathbf{0}$, i.e., if and only if there are no missing observations, since $E\left(\psi_2\left(u\right)|\mathbf{w}\right) = 0$ a.e. holds only if $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta} + g(t)$ is constant, which is a situation to be discarded in practice.

Remark 4.4. As in Remark 4.1 let us assume that the missingness probablity is given by the logistic model, i.e., that $p(\mathbf{x},t) = F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda}_0)$ and $G(\mathbf{x},t,\boldsymbol{\lambda}) = F_{\mathrm{L}}(\mathbf{v}^{\mathrm{T}}\boldsymbol{\lambda})$ where $\mathbf{v} = (1, \mathbf{x}^{\mathrm{T}}, t)^{\mathrm{T}}$ and F_{L} is the logistic function. Moreover, let us assume that $\widehat{\boldsymbol{\lambda}}$ is estimated using the maximum likelihood estimator. In this case, we can compare the asymptotic variances of the marginal location estimator when the missing probability is estimated parametrically or using a kernel estimator. We want to show that $\gamma_{\mathrm{S}}^2 \le \gamma^2$ and hence, $\upsilon^{(3)} \le \upsilon^{(2)}$ which means that the nonparametric estimator of the missing probability gives whenever it is possible to compute the smallest asymptotic variance. As above, our conclusions include in particular, the classical estimators, for which up to our knowledge there are no results regarding the theoretical comparison of the asymptotic variances a parametric or a nonparametric approach is used to estimate the missing probability. Let us recall that for the parametric situation, the asymptotic variance is given by $\gamma^2 = E\left(\psi_2^2\left(u\right)/p(\mathbf{x},t)\right) + \nu$ with $\nu = -E\left((1-p(\mathbf{w}))\psi_2\left(u\right)\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}_1^{-1}E\left((1-p(\mathbf{w}))\psi_2\left(u\right)\mathbf{v}\right)$, where $\mathbf{A}_1 = Ep(\mathbf{x},t)(1-p(\mathbf{x},t))\mathbf{v}\mathbf{v}^{\mathrm{T}}$. Note that $E\left((1-p(\mathbf{w}))\psi_2\left(u\right)\mathbf{v}^{\mathrm{T}}\right) = E\left((1-p(\mathbf{w}))r(\mathbf{w})\mathbf{v}^{\mathrm{T}}\right)$, and so, in order to compare the asymptotic variances and using the expression given in Remark 4.3, we only need to compare, the quantities

$$\begin{aligned}
\nu_{\mathrm{P}} &= E\left((1-p(\mathbf{w}))r(\mathbf{w})\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}_1^{-1}E\left((1-p(\mathbf{w}))r(\mathbf{w})\mathbf{v}\right) \\
\nu_{\mathrm{S}} &= E\left(\frac{(1-p(\mathbf{w}))}{p(\mathbf{w})}r^2(\mathbf{w})\right) .
\end{aligned}$$

Clearly, if $\nu_{\mathrm{S}} = 0$ then $\nu_{\mathrm{P}} = 0$, so we can assume that $\nu_{\mathrm{S}} > 0$. Let $\mathbf{A}_1 = \mathbf{C}_1\mathbf{C}_1^{\mathrm{T}}$ and denote by $\mathbf{z} = \mathbf{C}_1^{-1}(\delta - p(\mathbf{w}))\mathbf{v}$ and by $\xi = (\delta - p(\mathbf{w}))p(\mathbf{w})^{-1}r(\mathbf{w})$. Then, $E(\mathbf{z}) = 0$, $E(\xi) = 0$, $E(\mathbf{z}\mathbf{z}^{\mathrm{T}}) = \mathbf{I}$, $\nu_{\mathrm{P}} = \|E\left(\xi\mathbf{z}\right)\|^2$ while $\nu_{\mathrm{S}} = E(\xi^2) = Var(\xi)$. If we denote by $\boldsymbol{\rho} = E\left(\xi\mathbf{z}\right)$ and $\boldsymbol{\Sigma}^{\star} = E\left(\mathbf{s}\mathbf{s}^{\mathrm{T}}\right)$ with $\mathbf{s} = (\xi, \mathbf{z}^{\mathrm{T}})^{\mathrm{T}}$, we have that $\boldsymbol{\Sigma}^{\star} = \begin{pmatrix} \nu_{\mathrm{S}} & \boldsymbol{\rho}^{\mathrm{T}} \\ \boldsymbol{\rho} & \mathbf{I} \end{pmatrix}$ is a non–negative definite matrix. Note that since $\det\left(\boldsymbol{\Sigma}^{\star}\right) = \nu_{\mathrm{S}} \det\left(\mathbf{I} - \nu_{\mathrm{S}}^{-1}\, \boldsymbol{\rho}\boldsymbol{\rho}^{\mathrm{T}}\right) \ge 0$, the eigenvalue $1 - \nu_{\mathrm{S}}^{-1}\, \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho}$ of $\mathbf{I} - \nu_{\mathrm{S}}^{-1}\, \boldsymbol{\rho}\boldsymbol{\rho}^{\mathrm{T}}$ is non–negative and so, $\nu_{\mathrm{P}} = \|\boldsymbol{\rho}\|^2 \le \nu_{\mathrm{S}}$, as desired.

5 Concluding Remarks

Under a partially linear model when there are missing observations in the response variable, but the covariates $(\mathbf{x}^{\mathrm{T}}, t)$ are totally observed, the classical procedures fail to give reliable estimations when it can be suspected that anomalous observations are present in the sample. Robust procedures to estimate the regression parameter and the marginal location y were introduced in Bianco *et al.* (2010). These methods lead to strongly consistent estimators. Moreover, in this paper, we derive their asymptotic distribution. In particular, for the weighted simplified $M-$estimate, $\widehat{\theta}$, we obtain the asymptotic distribution when the missingness probability is assumed to be known or when it is estimated either parametrically or using a kernel approach. Different asymptotic variances are obtained in each situation.

The obtained theoretical results validate the numerical results observed in the simulation study performed in Bianco *et al.* (2010), since they allow to show that estimating nonparametrically the missingness probability reduces the variance of the marginal estimator either when the probability is known or when it is estimated parametrically using the maximum likelihood estimator under a logistic missingness model. This counterintuitive phenomenon was also observed by several authors, such as, Pierce (1982), Rosenbaum (1987), Robins *et al.* (1994, 1995), Wang *et al.* (1998) and the references given therein. When the covariates are missing, Wang *et al.* (1997) discussed the gain of efficiency of the estimators of θ via adjustment of the missing probability. An heuristical argument justifying this behavior for general parameter estimation with missing covariates was given in Robins *et al.* (1994). The same arguments can be applied for missing responses. When the missing probability is modeled parametrically and the unknown quantities are estimated using maximum likelihood estimators, the gain of efficiency is related to the linear expansion given in the Appendix together with the joint asymptotic distribution of $\sum_{i=1}^n \delta_i p^{-1}(\mathbf{w}_i)\psi_2(u_i)/\sqrt{n}$ and of $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0\right)$ and so, the optimality arguments used in Pierce (1982) can be considered to explain the effect of replacing estimators for the true paramaters.

On the other hand, when the parameters are estimated robustly using a weighted maximum likelihood method with weight function w, the robust estimators of the marginal location $\widehat{\theta}^{(2)}$ may have a higher loss of efficiency. To be more precise, depending on the tuning constant selected to cutoff outliers, the variance of the robust marginal location estimator $\widehat{\theta}^{(2)}$ may be larger than that of the estimator $\widehat{\theta}^{(1)}$ computed with the true missingness probability and so, larger than that of $\widehat{\theta}^{(3)}$, the estimator based on a kernel approach. In this sense, we recommend using a smooth estimator of the missing probabilities instead of a parametric one, if the dimension of the covariates and the number of observations allow to compute the kernel estimator.

6 Appendix

6.1 Proof of the asymptotic normality of the regression estimates

For the sake of simplicity, we denote ψ' and ψ'' the first and second derivatives of ψ. Moreover, let $\mathbf{z} = \mathbf{z}(\boldsymbol{\beta}_0) = \mathbf{x} + (\partial g_{\boldsymbol{\beta}}(t)/\partial\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, $\mathbf{z}_i = \mathbf{z}_i(\boldsymbol{\beta}_0) = \mathbf{x}_i + (\partial g_{\boldsymbol{\beta}}(t_i)/\partial\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and

$$\widehat{\gamma}(\boldsymbol{\beta},\tau) = \widehat{g}_{\boldsymbol{\beta}}(\tau) - g_{\boldsymbol{\beta}}(\tau) \qquad \widehat{\gamma}_0(\tau) = \widehat{\gamma}(\boldsymbol{\beta}_0,\tau) \tag{8}$$

$$\widehat{\mathrm{v}}_j(\boldsymbol{\beta},\tau) = \frac{\partial\widehat{\gamma}(\boldsymbol{\beta},\tau)}{\partial\beta_j} \qquad \widehat{\mathrm{v}}_{j,0}(\tau) = \widehat{\mathrm{v}}_j(\boldsymbol{\beta}_0,\tau)\ . \tag{9}$$

We list the conditions needed for the asymptotic normality of the regression parameter estimators, followed by general comments on those conditions. The first condition is on the preliminary estimate of $g_{\boldsymbol{\beta}}(\tau)$, while the other ones concern the score functions and the underlying model distributions.

N1. a) The functions $\widehat{g}_{\boldsymbol{\beta}}(\tau)$ and $g_{\boldsymbol{\beta}}(\tau)$ are continuously differentiable with respect to $(\boldsymbol{\beta},\tau)$ and twice continuously differentiable with respect to $\boldsymbol{\beta}$ such that $(\partial^2 g_{\boldsymbol{\beta}}(\tau)/\partial\beta_j\partial\beta_\ell)|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is bounded. Furthermore, for any $1\le j,\ell\le p$, $\partial^2 g_{\boldsymbol{\beta}}(\tau)/\partial\beta_j\partial\beta_\ell$ satisfies the following equicontinuity condition:

$$\forall\epsilon>0,\ \exists\delta>0:\ |\boldsymbol{\beta}_1-\boldsymbol{\beta}_0|<\delta \Rightarrow \left\| \left.\frac{\partial^2}{\partial\beta_j\partial\beta_\ell} g_{\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_1} - \left.\frac{\partial^2}{\partial\beta_j\partial\beta_\ell} g_{\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \right\|_\infty < \epsilon\ .$$

b) $\left\|\widehat{g}_{\widehat{\boldsymbol{\beta}}} - g_0\right\|_\infty \xrightarrow{p} 0$, for any consistent estimate $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$.
c) For each $\tau\in\mathcal{T}$ and $\boldsymbol{\beta}$, $\widehat{\gamma}(\boldsymbol{\beta},\tau) \xrightarrow{p} 0$. Moreover, $n^{1/4}\,\|\widehat{\gamma}_0\|_\infty \xrightarrow{p} 0$ and $n^{1/4}\,\|\widehat{\mathrm{v}}_{j,0}\|_\infty \xrightarrow{p} 0$ for all $1\le j\le p$.
d) There exists a neighborhood of $\boldsymbol{\beta}_0$ with closure $\mathcal{K}$ such that for any $1\le j,\ell\le p$, $\sup_{\boldsymbol{\beta}\in\mathcal{K}}\left(\|\widehat{\mathrm{v}}_j(\boldsymbol{\beta},\cdot)\|_\infty + \|\partial\widehat{\mathrm{v}}_j(\boldsymbol{\beta},\cdot)/\partial\beta_\ell\|_\infty\right) \xrightarrow{p} 0$.
e) $\|\partial\widehat{\gamma}_0/\partial\tau\|_\infty + \|\partial\widehat{\mathrm{v}}_{j,0}/\partial\tau\|_\infty \xrightarrow{p} 0$ for any $1\le j\le p$.

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

N3. The matrix $\mathbf{A} = E\psi'(\epsilon)\; E\left(\upsilon(\mathbf{x})p(\mathbf{x},t)\mathbf{z}(\boldsymbol{\beta}_0)\mathbf{z}(\boldsymbol{\beta}_0)^{\mathrm{T}}\right)$ is non-singular.
N4. The matrix $\boldsymbol{\Sigma} = E\psi^2(\epsilon)\; E\left(\upsilon^2(\mathbf{x})p(\mathbf{x},t)\mathbf{z}(\boldsymbol{\beta}_0)\mathbf{z}(\boldsymbol{\beta}_0)^{\mathrm{T}}\right)$ is positive definite.
N5. $E(\psi_1'(\epsilon)) \neq 0$ and $E(\psi'(\epsilon)) \neq 0$.
N6. $E\left(p(\mathbf{x},t)\upsilon(\mathbf{x})\,\|\mathbf{z}(\boldsymbol{\beta}_0)\|^2\right) < \infty$.

Remark 6.1.1. Condition **N1**b) follows from the continuity of $g_{\boldsymbol{\beta}}(\tau) = g(\boldsymbol{\beta},\tau)$ with respect to $(\boldsymbol{\beta},\tau)$ and Theorem 3.1 that leads to $\sup_{\boldsymbol{\beta}\in\mathcal{K}} \|\widehat{g}_{\boldsymbol{\beta}} - g_{\boldsymbol{\beta}}\|_\infty \xrightarrow{a.s.} 0$. Conditions **N1**a) and d) entail that for any consistent estimator $\widetilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$, we have that

$$\max_{1\le j\le p}\left\|\frac{\partial \widehat{g}_{\boldsymbol{\beta}}}{\partial\beta_j}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} - \frac{\partial g_{\boldsymbol{\beta}}}{\partial\beta_j}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right\|_\infty \xrightarrow{p} 0 \text{ and } \max_{1\le j,\ell\le p}\left\|\frac{\partial^2 \widehat{g}_{\boldsymbol{\beta}}}{\partial\beta_j\partial\beta_\ell}|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}} - \frac{\partial^2 g_{\boldsymbol{\beta}}}{\partial\beta_j\partial\beta_\ell}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right\|_\infty \xrightarrow{p} 0\,.$$

Remark 6.1.2. When the kernel K is continuously differentiable with bounded derivative K' and with bounded variation, the uniform convergence required in **N1**d) and e) can be derived through analogous arguments to those considered in Theorem 3.1 by using that

$$\frac{\partial}{\partial\tau}\widehat{g}_{\boldsymbol{\beta}}(\tau) = \frac{\left(nh_n{}^2\right)^{-1}\sum_{i=1}^n K'\left(\frac{\tau-t_i}{h_n}\right)\delta_i\psi_1\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\widehat{g}_{\boldsymbol{\beta}}(\tau)}{\widehat{s}_{\boldsymbol{\beta}}}\right)\upsilon(\mathbf{x}_i)}{(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{\tau-t_i}{h_n}\right)\delta_i\psi_1'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\widehat{g}_{\boldsymbol{\beta}}(\tau)}{\widehat{s}_{\boldsymbol{\beta}}}\right)\upsilon(\mathbf{x}_i)}$$

$$\frac{\partial}{\partial\beta_j}\widehat{g}_{\boldsymbol{\beta}}(\tau) = -\frac{\sum_{i=1}^n K\left(\frac{\tau-t_i}{h_n}\right)\left[\delta_i\psi_1'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\widehat{g}_{\boldsymbol{\beta}}(\tau)}{\widehat{s}_{\boldsymbol{\beta}}}\right)\upsilon(\mathbf{x}_i)\right]\left(x_{ij}+\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\widehat{g}_{\boldsymbol{\beta}}(\tau)}{\widehat{s}_{\boldsymbol{\beta}}}\frac{\partial}{\partial\beta_j}\widehat{s}_{\boldsymbol{\beta}}\right)}{\sum_{i=1}^n K\left(\frac{\tau-t_i}{h_n}\right)\delta_i\psi_1'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\widehat{g}_{\boldsymbol{\beta}}(\tau)}{\widehat{s}_{\boldsymbol{\beta}}}\right)\upsilon(\mathbf{x}_i)}$$

and requiring that $u\psi_1'(u)$ is a bounded function and

$$\sup_{\tau\in\mathcal{T}} E\left(\sup_{\boldsymbol{\beta}\in\mathcal{K},\sigma\in\mathcal{K}_\sigma}|\psi_1'\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(\tau)}{\sigma}\right)\|\mathbf{x}\|\;|t=\tau\right) < \infty$$

$$\sup_{\tau\in\mathcal{T}} E\left(\sup_{\boldsymbol{\beta}\in\mathcal{K},\sigma\in\mathcal{K}_\sigma}|\psi_1''\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(\tau)}{\sigma}\right)\|\mathbf{x}\|\;|t=\tau\right) < \infty$$

$$\inf_{\substack{\boldsymbol{\beta}\in\mathcal{K},\sigma\in\mathcal{K}_\sigma\\ \tau\in\mathcal{T}}}|E\left(\psi_1'\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(\tau)}{\sigma}\right)|t=\tau\right)| > 0\,.$$

The uniform convergence rates required in **N1**c) are fulfilled when $\widehat{g}_{\boldsymbol{\beta}}$ is defined in **Step 1** using kernel weights and a rate-optimal bandwidth is used for the kernel.

Remark 6.1.3. Note that if $P(\upsilon(\mathbf{x}) > 0) = 1$ and $E\psi'(\epsilon) \neq 0$, **N3** holds, i.e., $\mathbf{A}$ will be non–singular unless $P\left(\mathbf{a}^{\mathrm{T}}\mathbf{z}(\boldsymbol{\beta}_0) = 0\right) = 1$, for some $\mathbf{a} \in \mathbb{R}^p$, that is, unless there is a linear combination of $\mathbf{x}$ which can be completely determined by t. The condition $E\psi'(\epsilon) \neq 0$ is a standard requirement in robust regression in order to get root-n estimators of $\boldsymbol{\beta}$.

Again, if **N4** is fulfilled the columns of $\mathbf{x} + (\partial g_{\boldsymbol{\beta}}(t)/\partial\boldsymbol{\beta})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ will not be collinear. It is necessary not to allow $\mathbf{x}$ to be predicted by t to get root-n regression estimates.

N5 is a standard condition in robustness in order to get root$-n$ estimators. It is worth noticing that **N5** entails

$$E\left[\left(\mathbf{x}+\frac{\partial}{\partial\boldsymbol{\beta}}g_{\boldsymbol{\beta}}(\tau)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\upsilon(\mathbf{x})p(\mathbf{x},\tau)|t=\tau\right] = 0\,. \tag{10}$$

Effectively, since $g_{\boldsymbol{\beta}}(\tau)$ satisfies (2) for each τ differentiating with respect to $\boldsymbol{\beta}$, we get

$$E\left[\delta\psi_1'\left(\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(\tau)}{\sigma_{\boldsymbol{\beta}}}\right)\left(\mathbf{x}+\frac{\partial}{\partial\boldsymbol{\beta}}g_{\boldsymbol{\beta}}(\tau)+\frac{y-\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(\tau)}{\sigma_{\boldsymbol{\beta}}}\frac{\partial}{\partial\boldsymbol{\beta}}\sigma_{\boldsymbol{\beta}}\right)\upsilon(\mathbf{x})\middle|\, t=\tau\right]=0\quad\forall\boldsymbol{\beta}\,.$$

Thus, specializing at $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, we get that

$$\begin{aligned}
0 &= E\left[\delta\psi_1'\left(\epsilon\right)\left(\mathbf{x}+\frac{\partial}{\partial\boldsymbol{\beta}}g_{\boldsymbol{\beta}}(\tau)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}+\epsilon\frac{\partial}{\partial\boldsymbol{\beta}}\sigma_{\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\upsilon(\mathbf{x})\middle|\, t=\tau\right]\\
&= E(\psi_1'\left(\epsilon\right))\; E\left[p(\mathbf{x},t)\left(\mathbf{x}+\frac{\partial}{\partial\boldsymbol{\beta}}g_{\boldsymbol{\beta}}(\tau)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\upsilon(\mathbf{x})\middle|\, t=\tau\right]+E(\epsilon\psi_1'\left(\epsilon\right))E\left[p(\mathbf{x},t)\upsilon(\mathbf{x})|\, t=\tau\right]\frac{\partial}{\partial\boldsymbol{\beta}}\sigma_{\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\\
&= E(\psi_1'\left(\epsilon\right))\; E\left[p(\mathbf{x},t)\left(\mathbf{x}+\frac{\partial}{\partial\boldsymbol{\beta}}g_{\boldsymbol{\beta}}(\tau)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right)\upsilon(\mathbf{x})\middle|\, t=\tau\right]\;,
\end{aligned}$$

where the last equality holds since ψ_1' is an even function and ϵ has a symmetric distribution. Thus, (10) holds.

Assumption **N6** is used to ensure the consistency of the estimates of $\mathbf{A}$ based on preliminary estimates of the regression parameter $\boldsymbol{\beta}$ and of the functions $g_{\boldsymbol{\beta}}$.

Lemma 6.1.1. *Let $(y_i,\mathbf{x}_i^{\mathrm{T}},t_i)$ be independent observations satisfying (1). Assume that t_i are random variables with distribution on a compact set $\mathcal{T}$ and that **N1** to **N3** and **N6** hold. Let $\widetilde{\boldsymbol{\beta}}$ be such that $\widetilde{\boldsymbol{\beta}}\xrightarrow{p}\boldsymbol{\beta}_0$ and $\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})=\mathbf{x}_i+(\partial\widehat{g}_{\boldsymbol{\beta}}(t_i)/\partial\boldsymbol{\beta})\,|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}$. Then, $\mathbf{A}_n\xrightarrow{p}\mathbf{A}$ where $\mathbf{A}$ is given in **N3***

$$\mathbf{A}_n=\frac{1}{n}\sum_{i=1}^{n}\left(\psi'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})^{\mathrm{T}}+\psi\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\;\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\widehat{g}_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\right)\delta_i\upsilon(\mathbf{x}_i)$$

Proof. Note that $\mathbf{A}_n$ can be written as $\mathbf{A}_n=\sum_{j=1}^{6}\mathbf{A}_n^{(j)}$ where

$$\begin{aligned}
\mathbf{A}_n^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-g_0(t_i)}{\widehat{\sigma}}\right)\mathbf{z}_i\,\mathbf{z}_i^{\mathrm{T}}\upsilon(\mathbf{x}_i)\\
\mathbf{A}_n^{(2)} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-g_0(t_i)}{\widehat{\sigma}}\right)\;\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}g_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}^{\mathrm{T}}\upsilon(\mathbf{x}_i)\\
\mathbf{A}_n^{(3)} &= \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\delta_i\psi''\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\xi_{i,1}}{\widehat{\sigma}}\right)\widehat{\mathrm{w}}_0(t_i)\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\upsilon(\mathbf{x}_i)\\
\mathbf{A}_n^{(4)} &= \frac{1}{\widehat{\sigma}}\frac{1}{n}\sum_{i=1}^{n}\delta_i\psi'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\xi_{i,2}}{\widehat{\sigma}}\right)\widehat{\mathrm{w}}_0(t_i)\;\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}g_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}^{\mathrm{T}}\upsilon(\mathbf{x}_i)\\
\mathbf{A}_n^{(5)} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\left[\widehat{\mathbf{w}}(t_i)\mathbf{z}_i^{\mathrm{T}}+\mathbf{z}_i\widehat{\mathbf{w}}(t_i)^{\mathrm{T}}+\widehat{\mathbf{w}}(t_i)\;\widehat{\mathbf{w}}(t_i)^{\mathrm{T}}\right]\upsilon(\mathbf{x}_i)\\
\mathbf{A}_n^{(6)} &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\psi\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}-\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\;\widehat{\mathbf{V}}(t_i)^{\mathrm{T}}\upsilon(\mathbf{x}_i)\;,
\end{aligned}$$

where $\xi_{i,1}$ and $\xi_{i,2}$ are intermediate points and $\mathbf{z}_i=\mathbf{z}_i(\boldsymbol{\beta}_0)$, $\widehat{\mathrm{w}}_0(t)=\widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t)-g_0(t)$ and

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

Using **N1**a), b) and d), **N6**, the boundness of ψ, ψ', ψ'', υ and Υ and the fact that $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$, it follows easily that $\mathbf{A}_n^{(j)} \xrightarrow{p} 0$ for $3 \le j \le 6$. From **N6**, the consistency of $\widetilde{\boldsymbol{\beta}}$ and the continuity of ψ and ψ', we get easily that $\mathbf{A}_n^{(1)} + \mathbf{A}_n^{(2)} \xrightarrow{p} \mathbf{A}$. $\square$

Proof of Theorem 3.3. Let $\widehat{\boldsymbol{\beta}}$ is a solution of $H_n^{(1)}(\boldsymbol{\beta}) = 0$ defined in (4) and denote by $\widehat{\mathbf{z}}_i(\boldsymbol{\beta}) = \mathbf{x}_i + (\partial \widehat{g}_{\boldsymbol{\beta}}(t_i)/\partial \boldsymbol{\beta})\,|_{\boldsymbol{\beta}}$. Using a Taylor's expansion of order one we get

$$\begin{aligned}0 &= \sum_{i=1}^n \delta_i \psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \widehat{g}_{\widehat{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\widehat{\mathbf{z}}_i(\widehat{\boldsymbol{\beta}}) \\ &= \sum_{i=1}^n \delta_i \psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \widehat{g}_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\widehat{\mathbf{z}}_i(\boldsymbol{\beta}_0) - \frac{1}{\widehat{\sigma}} n\mathbf{A}_n\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\ ,\end{aligned}$$

where

$$\begin{aligned}\mathbf{A}_n &= -\frac{\widehat{\sigma}}{n}\sum_{i=1}^n \delta_i \frac{\partial}{\partial\beta}\left\{\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \widehat{g}_{\boldsymbol{\beta}}(t_i)}{\widehat{\sigma}}\right)\widehat{\mathbf{z}}_i(\boldsymbol{\beta})\right\}\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}\upsilon(\mathbf{x}_i) \\ &= \frac{1}{n}\sum_{i=1}^n\left(\psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})\widehat{\mathbf{z}}_i(\widetilde{\boldsymbol{\beta}})^{\mathrm{T}} - \psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}} - \widehat{g}_{\widetilde{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\ \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}}\widehat{g}_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\widetilde{\boldsymbol{\beta}}}^{\mathrm{T}}\right)\delta_i\upsilon(\mathbf{x}_i)\ ,\end{aligned}$$

with $\widetilde{\boldsymbol{\beta}}$ an intermediate point between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. From Lemma 6.1.1, we have that $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$, where $\mathbf{A}$ is defined in **N3**. Therefore, in order to obtain the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$ it will be enough to derive the asymptotic behavior of

$$\widehat{L}_n = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \widehat{g}_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\widehat{\mathbf{z}}_i(\boldsymbol{\beta}_0)\ .$$

Let

$$L_n = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\mathbf{z}_i(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \delta_i\psi\left(\frac{\epsilon_i\sigma_0}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\mathbf{z}_i(\boldsymbol{\beta}_0)\ ,$$

since $g_{\boldsymbol{\beta}_0} = g_0$. Using that ψ is odd and the errors have a symmetric distribution and are independent of the carriers, we have that $E\left[\psi\left(\epsilon_i\sigma_0/\sigma\right)|(\mathbf{x}_i, t_i)\right] = E\psi\left(\epsilon_i\sigma_0/\sigma\right) = 0$, for all σ. Then, the consistency of $\widehat{\sigma}$ and standard tightness arguments entail that L_n is asymptotically normally distributed with covariance matrix $\boldsymbol{\Sigma}$. Therefore, it remains to show that $L_n - \widehat{L}_n \xrightarrow{p} 0$.

We have the following expansion $\widehat{L}_n - L_n = -\widehat{\sigma}^{-2}L_n^1 + \widehat{\sigma}^{-1}L_n^2 - \widehat{\sigma}^{-1}L_n^3 + \widehat{\sigma}^{-2}L_n^4$, with

$$\begin{aligned}L_n^1 &= n^{-1/2}\widehat{\sigma}\sum_{i=1}^n \delta_i\psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\mathbf{z}_i(\boldsymbol{\beta}_0)\upsilon(\mathbf{x}_i)\widehat{\gamma}_0(t_i) \\ L_n^2 &= n^{-1/2}\widehat{\sigma}\sum_{i=1}^n \delta_i\psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\widehat{\mathbf{v}}_0(t_i) \\ L_n^3 &= n^{-1}\sum_{i=1}^n \delta_i\psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_{\boldsymbol{\beta}_0}(t_i)}{\widehat{\sigma}}\right)\upsilon(\mathbf{x}_i)\ \left(n^{1/4}\widehat{\mathbf{v}}_0(t_i)\right)\ \left(n^{1/4}\widehat{\gamma}_0(t_i)\right) \\ L_n^4 &= (2n)^{-1}\sum_{i=1}^n \delta_i\psi''\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - \xi_i(t_i)}{\widehat{\sigma}}\right)\mathbf{z}_i(\boldsymbol{\beta}_0)\upsilon(\mathbf{x}_i)\left(n^{1/4}\widehat{\gamma}_0(t_i)\right)^2\ ,\end{aligned}$$

where $\widehat{\gamma}_0(\tau) = \widehat{g}_{\boldsymbol{\beta}_0}(\tau) - g_0(\tau)$, $\widehat{\mathbf{v}}_0(\tau) = (\widehat{\mathrm{v}}_{1,0}(\tau), \dots, \widehat{\mathrm{v}}_{p,0}(\tau))^{\mathrm{T}} = \partial\widehat{\gamma}(\boldsymbol{\beta}, \tau)/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is defined in (9), $\widehat{\gamma}$ is defined in (8) and $\xi(t_i)$ an intermediate point between $\widehat{g}_{\boldsymbol{\beta}_0}(t_i)$ and $g_0(t_i)$. It is easy to see that $L_n^3 \xrightarrow{p} 0$ and $L_n^4 \xrightarrow{p} 0$ follow from **N1**c) and **N2**.

To complete the proof, we will show that $L_n^j \xrightarrow{p} 0$ for $j = 1, 2$ which will follow from **N1**c) to e) and (10), using similar arguments to those considered in Bianco and Boente (2004).

Effectively, fix the coordinate j, $1 \le j \le p$. For any function γ and any positive fixed number σ, if $x_{i,j}$ and β_j denote the $j-$th coordinate of $\mathbf{x}_i$ and $\boldsymbol{\beta}$ respectively, we define

$$J_{n,1}(\gamma, \sigma) = n^{-1/2}\sigma \sum_{i=1}^{n} \psi'\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_0(t_i)}{\sigma}\right)\left[x_{i,j} + \frac{\partial}{\partial\beta_j} g_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \upsilon(\mathbf{x}_i)\, \gamma(t_i)$$
$$J_{n,2}(\gamma, \sigma) = n^{-1/2}\sigma \sum_{i=1}^{n} \psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 - g_0(t_i)}{\sigma}\right) \upsilon(\mathbf{x}_i)\, \gamma(t_i)$$

where we have omitted the subscript j for the sake of simplicity.

Let $\mathcal{V} = \{\gamma \in \mathcal{C}^1(\mathcal{T}) : \|\gamma\|_\infty \le 1 \quad \|\gamma'\|_\infty \le 1\}$. Note that, for any probability measure $\mathbb{Q}$, the bracketing number $N_{[\,]}\left(\epsilon, \mathcal{V}, L^2(\mathbb{Q})\right)$, and so the covering number $N\left(\epsilon, \mathcal{V}, L^2(\mathbb{Q})\right)$, satisfy

$$\log N\left(\epsilon/2, \mathcal{V}, L^2(\mathbb{Q})\right) \le \log N_{[\,]}\left(\epsilon, \mathcal{V}, L^2(\mathbb{Q})\right) \le A\epsilon^{-1} ,$$

for $0 < \epsilon < 2$, where the constant A is independent of the probability measure $\mathbb{Q}$ (see Corollary 2.7.2 in van der Vaart and Wellner (1996)).

Denote $\mathcal{I} = [\sigma_0/2, 2\sigma_0]$. Consider the classes of functions

$$\mathcal{F}_1 = \left\{ f_{1,\gamma,\sigma}(y, \mathbf{x}, t) = \sigma\psi'\left(\frac{y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 - g_0(t)}{\sigma}\right)\left[x_j + \frac{\partial}{\partial\beta_j} g_{\boldsymbol{\beta}}(t)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right] \upsilon(\mathbf{x})\, \gamma(t) \ , \ \gamma \in \mathcal{V} \ , \ \sigma \in \mathcal{I} \right\}$$
$$\mathcal{F}_2 = \left\{ f_{2,\gamma,\sigma}(y, \mathbf{x}, t) = \sigma\psi\left(\frac{y - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 - g_0(t)}{\sigma}\right) \upsilon(\mathbf{x})\, \gamma(t) \ , \ \gamma \in \mathcal{V} \ , \ \sigma \in \mathcal{I} \right\} .$$

$\mathcal{F}_1$ and $\mathcal{F}_2$ have as envelopes the constants

$$A_1 = 2\sigma_0\|\psi'\|_\infty \left[\|\Upsilon\|_\infty + \left\|(\partial g_{\boldsymbol{\beta}})/\partial\beta_j|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}\right\|_\infty \|\upsilon\|_\infty\right] = 2\sigma_0\|\psi'\|_\infty A_0$$

and $A_2 = 2\sigma_0\|\psi\|_\infty\|\upsilon\|_\infty$, respectively. On the other hand, (10), the oddness of ψ and the symmetry of the error's distribution imply that, for any $f \in \mathcal{F}_1 \cup \mathcal{F}_2$, $E\, f\,(y_i, \mathbf{x}_i, t_i) = 0$.

Write $\psi_\sigma(r) = \sigma\,\psi(r/\sigma)$ and $\psi'_\sigma(r) = \sigma\,\psi'(r/\sigma)$. From **N2**, we have that φ_1 and φ_2 are bounded, which entails that

$$|\psi'_{s_1}(r) - \psi'_{s_2}(r)| \le (\|\psi'\|_\infty + \|\varphi_2\|_\infty)\ |s_1 - s_2| \quad \text{and} \quad |\psi_{s_1}(r) - \psi_{s_2}(r)| \le (\|\psi\|_\infty + \|\varphi_1\|_\infty)\ |s_1 - s_2| .$$

Let $B_1 = A_0\left(\|\psi'\|_\infty(3 + 2\sigma_0) + 3\|\varphi_2\|_\infty\right)$ and $B_2 = \|\upsilon\|_\infty\left(\|\psi\|_\infty(3 + 2\sigma_0) + 3\|\varphi_1\|_\infty\right)$. Denote $\|f\|_{\mathbb{Q},2} = \left(E_{\mathbb{Q}}(f^2)\right)^{1/2}$. It is easy to see that, for any $\gamma \in \mathcal{V}$, $\sigma \in \mathcal{I}$ and $0 < \epsilon < 2$, $\|\gamma_s - \gamma\|_{\mathbb{Q},2} < \epsilon$ and $|\sigma_\ell - \sigma| < \epsilon$, entail that $\|f_{1,\gamma_s,\sigma_\ell} - f_{1,\gamma,\sigma}\|_{\mathbb{Q},2} \le B_1\,\epsilon$ which allows to conclude that

$$N\left(\epsilon\, B_1, \mathcal{F}_1, L^2(\mathbb{Q})\right) \le N\left(\epsilon, \mathcal{V}, L^2(\mathbb{Q})\right)\ N\left(\epsilon, \mathcal{I}, |\cdot|\right) .$$

Similarly, we get that $N\left(\epsilon\, B_2, \mathcal{F}_2, L^2(\mathbb{Q})\right) \le N\left(\epsilon, \mathcal{V}, L^2(\mathbb{Q})\right)\ N\left(\epsilon, \mathcal{I}, |\cdot|\right)$. Therefore, these classes of functions have finite uniform–entropy.

For any class of functions $\mathcal{F}$, denote by $\mathcal{J}(\delta, \mathcal{F}) = \sup_{\mathbb{Q}} \int_0^\delta \sqrt{1 + \log\left(N\left(\epsilon\,\|F\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q})\right)\right)}\, d\epsilon$, where the supremum is taken over all discrete probability measures $\mathbb{Q}$ with $\|F\|_{\mathbb{Q},2} > 0$ and F is the envelope of $\mathcal{F}$. The function $\mathcal{J}$ is increasing, $\mathcal{J}(0,\mathcal{F})=0$ and $\mathcal{J}(1,\mathcal{F})<\infty$ and $\mathcal{J}(\delta,\mathcal{F})\to 0$ as $\delta\to 0$ for classes of functions $\mathcal{F}$ which satisfies the uniform–entropy condition. Moreover, if $\mathcal{F}_0\subset\mathcal{F}$ and the envelope F is used for $\mathcal{F}_0$, then $\mathcal{J}(\delta,\mathcal{F}_0)\leq\mathcal{J}(\delta,\mathcal{F})$.

For any $\epsilon>0$ and $0<\delta<1$, consider the subclasses

$$\mathcal{F}_{1,\delta}=\{f_{1,\gamma,\sigma}(y,\mathbf{x},t)\in\mathcal{F}_1 \text{ with } \|\upsilon\|_\infty<\delta\} \quad \text{and} \quad \mathcal{F}_{2,\delta}=\{f_{2,\gamma,\sigma}(y,\mathbf{x},t)\in\mathcal{F}_2 \text{ with } \|\upsilon\|_\infty<\delta\}\ .$$

Remind that $\widehat{\gamma}_0(\tau)=\widehat{g}_{\boldsymbol{\beta}_0}(\tau)-g_0(\tau)$ and $\widehat{\mathrm{v}}_{j,0}(\tau)=(\partial\widehat{\gamma}(\boldsymbol{\beta},\tau))/\partial\beta_j|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$. Using that **N1**c) and e) entail that $\sup_{\tau\in\mathcal{T}}|\widehat{\gamma}_0(\tau)|\xrightarrow{p}0$, $\sup_{\tau\in\mathcal{T}}|\partial\widehat{\gamma}_0(\tau)/\partial\tau|\xrightarrow{p}0$, $\sup_{\tau\in\mathcal{T}}|\widehat{\mathrm{v}}_{j,0}(\tau)|\xrightarrow{p}0$, $\sup_{\tau\in\mathcal{T}}|\partial\widehat{\mathrm{v}}_{j,0}(\tau)/\partial\tau|\xrightarrow{p}0$ and the consistency of $\widehat{\sigma}$, we have that, for $1\leq j\leq p$ and n large enough,

$$\begin{aligned}P\left(\widehat{\sigma}\in\mathcal{I}\,,\widehat{\gamma}_0\in\mathcal{V}\text{ and }\|\widehat{\gamma}_0\|_\infty<\delta\right)&>1-\delta/2\\ P\left(\widehat{\mathrm{v}}_{j,0}\in\mathcal{V}\text{ and }\|\widehat{\mathrm{v}}_{j,0}\|_\infty<\delta\right)&>1-\delta/2\ .\end{aligned}$$

It is clear that $\sup_{f\in\mathcal{F}_{1,\delta}}\sum_{i=1}^n f^2(r_i,\mathbf{z}_i,t_i)/n\leq A_1^2\delta^2$ and $\sup_{f\in\mathcal{F}_{2,\delta}}\sum_{i=1}^n f^2(r_i,\mathbf{z}_i,t_i)/n\leq A_2^2\,\delta^2$. Therefore, the maximal inequality for covering numbers entails that, for any $0\leq\ell\leq p$,

$$\begin{aligned}P\left(|J_{n,1}\left(\widehat{\gamma}_0,\widehat{\sigma}\right)|>\epsilon\right)&\leq P\left(|J_{n,1}\left(\widehat{\gamma}_0,\widehat{\sigma}\right)|>\epsilon\,,\ \widehat{\sigma}\in\mathcal{I},\ \widehat{\gamma}_0\in\mathcal{V}\text{ and }\|\widehat{\gamma}_0\|_\infty<\delta\right)+\delta\\&\leq P\left(\sup_{f\in\mathcal{F}_{1,\delta}}\left|n^{-1/2}\sum_{i=1}^n f(y_i,\mathbf{x}_i,t_i)\right|>\epsilon\right)+\delta\\&\leq\epsilon^{-1}D_1\,A_1\,\mathcal{J}\left(\delta,\mathcal{F}_1\right)+\delta\ ,\end{aligned}$$

where D_1 is a constant not depending on n.

Similarly, $P\left(|J_{n,2}\left(\widehat{\mathrm{v}}_{j,0},\widehat{\sigma}\right)|>\epsilon\right)\leq\epsilon^{-1}D_2\,A_2\,\mathcal{J}\left(\delta,\mathcal{F}_2\right)+\delta$. Using that the classes $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy the uniform–entropy condition, we get $\lim_{\delta\to 0}\mathcal{J}\left(\delta,\mathcal{F}_1\right)=0$ and $\lim_{\delta\to 0}\mathcal{J}\left(\delta,\mathcal{F}_2\right)=0$. Thus, we have that $L_n^1=J_{n,1}\left(\widehat{\gamma}_0\right)\xrightarrow{p}0$ and $L_n^2=\left(J_{n,2}\left(\widehat{\mathrm{v}}_{1,0}\right),\ldots,J_{n,2}\left(\widehat{\mathrm{v}}_{p,0}\right)\right)^{\mathrm{T}}\xrightarrow{p}0$, as desired. □

6.2 Proof of the asymptotic distribution of the marginal estimators

When estimating the marginal location, we will assume, without loss of generality, that the marginal scale ς_0 is known and so we will replace $\widehat{\varsigma}$ by ς_0. Recall that $u=(y-\theta)/\varsigma_0$ and denote $u_i=(y_i-\theta)/\varsigma_0$.

NM1. The function ψ_2 is twice continuously differentiable with bounded derivatives.
NM2. $A(\psi_2)=E\left[\delta\psi_2'\left(u\right)/p(\mathbf{x},t)\right]=E\psi_2'\left(u\right)\neq 0$.
NM3. $\inf_{(\mathbf{x},t)}p(\mathbf{x},t)=\iota(p)>0$.
NM4. The missingness probability $p(\mathbf{x},t)=G(\mathbf{x},t,\boldsymbol{\lambda}_0)$, $\boldsymbol{\lambda}_0\in\mathbb{R}^q$, is such that
a) the family of functions $\mathcal{G}=\{G(\mathbf{x},t,\boldsymbol{\lambda}):\boldsymbol{\lambda}\in\mathbb{R}^q\}$ has finite entropy.
b) $G(\mathbf{x},t,\boldsymbol{\lambda})$ is twice continuously differentiable with respect to $\boldsymbol{\lambda}$. We will denote by $G'(\mathbf{x},t,\boldsymbol{\lambda})$ and $G''(\mathbf{x},t,\boldsymbol{\lambda})$ the gradient and hessian matrix of $G(\mathbf{x},t,\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$.
c) $E\left(|G_j'(\mathbf{x},t,\boldsymbol{\lambda}_0)|\psi_2'\left(u\right)/p(\mathbf{x},t)\right)<\infty$ for $1\leq j\leq q$.
d) For some $\Lambda>0$, $E\left(\sup_{\|\boldsymbol{\lambda}-\boldsymbol{\lambda}_0\|<\Lambda}|G_{j\ell}''(\mathbf{x},t,\boldsymbol{\lambda})\psi_2'\left(u\right)|/p(\mathbf{x},t)\right)<\infty$ for $1\leq j,\ell\leq q$.

NM5. $\widehat{\boldsymbol{\lambda}}$ admits a Bahadur expansion given by $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0\right)=(1/\sqrt{n})\sum_{i=1}^n\boldsymbol{\eta}(\delta_i,\mathbf{x}_i,t_i)+o_p(1)$ where $E\boldsymbol{\eta}(\delta_i,\mathbf{x}_i,t_i)=\mathbf{0}$ and $E\|\boldsymbol{\eta}(\delta_i,\mathbf{x}_i,t_i)\|^2<\infty$. We will denote by $\boldsymbol{\Sigma}=E\boldsymbol{\eta}(\delta,\mathbf{x},t)\boldsymbol{\eta}(\delta,\mathbf{x},t)^{\mathrm{T}}$ the asymptotic covariance matrix of $\widehat{\boldsymbol{\lambda}}$.

Assumption **NM4** holds in most parametric situations such as the logistic missingness model.

Proof of Theorem 4.1. A Taylor's expansion of order two leads to

$$0=\sqrt{n}\,U_n(p,\varsigma_0,\widehat{\theta}^{(1)})=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)-\sqrt{n}(\widehat{\theta}^{(1)}-\theta)A_n(\psi_2)\,,$$

where

$$A_n(\psi_2)=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2'\left(u_i\right)+\frac{1}{2}(\widehat{\theta}^{(1)}-\theta)\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2''\left(\frac{y_i-\xi_n}{\varsigma_0}\right)$$

and ξ_n is an intermediate point between $\widehat{\theta}^{(1)}$ and θ. Using that $A_n(\psi_2)\xrightarrow{p}A(\psi_2)$ and $\sqrt{n}\,U_n(p,\varsigma_0,\theta)\xrightarrow{\mathcal{D}}N(0,\gamma_0^2)$ with $\gamma_0^2=E\left(\delta\psi_2^2\left(u\right)/p(\mathbf{x},t)^2\right)=E\psi_2^2\left(u\right)/p(\mathbf{x},t)$, the proof follows. $\square$

Proof of Theorem 4.2. As in the proof of Theorem 4.1, using a Taylor's expansion of order two, we get that $0=(1/\sqrt{n})\sum_{i=1}^n(\delta_i/p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i))\psi_2\left(u_i\right)-\sqrt{n}(\widehat{\theta}^{(2)}-\theta)A_n^{(2)}(\psi_2)$, where

$$A_n^{(2)}(\psi_2)=\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i)}\psi_2'\left(u_i\right)+\frac{1}{2}(\widehat{\theta}^{(2)}-\theta)\frac{1}{n}\sum_{i=1}^n\frac{\delta_i}{p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i)}\psi_2''\left(\frac{y_i-\xi_n}{\varsigma_0}\right)$$

and ξ_n is an intermediate point between $\widehat{\theta}^{(2)}$ and θ. Using **NM3**, it follows that $A_n^{(2)}(\psi_2)\xrightarrow{p}A(\psi_2)$. Therefore, it is enough to show that $B_n=(1/\sqrt{n})\sum_{i=1}^n\left(\delta_i/p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i)\right)\psi_2\left(u_i\right)\xrightarrow{\mathcal{D}}N(0,\upsilon^2)$. Note that

$$B_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{p(\mathbf{x}_i,t_i)}{p_{n,\widehat{\boldsymbol{\lambda}}}(\mathbf{x}_i,t_i)}-1\right)\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)\,.$$

Denote by

$$R_n(\boldsymbol{\lambda})=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\frac{G(\mathbf{x}_i,t_i,\boldsymbol{\lambda}_0)}{G(\mathbf{x}_i,t_i,\boldsymbol{\lambda})}-1\right)\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\left[\psi_2\left(u_i\right)-r(\mathbf{x}_i,t_i)\right]\,,$$

where $r(\mathbf{x},t)=E\psi_2\left(u\right)|(\mathbf{x},t)$. Then, using **NM4**a), the fact that $\widehat{\boldsymbol{\lambda}}\xrightarrow{p}\boldsymbol{\lambda}_0$ and standard empirical processes arguments as those considered in the proof of Theorem 3.3, we get easily that $R_n(\widehat{\boldsymbol{\lambda}})\xrightarrow{p}0$ and so, $B_n=B_{1,n}+B_{2,n}+B_{3,n}+o_p(1)$ where

$$B_{1,n}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)\,,\quad B_{2,n}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\boldsymbol{\lambda}_0-\widehat{\boldsymbol{\lambda}}\right)^{\mathrm{T}}\frac{G'(\mathbf{x}_i,t_i,\boldsymbol{\lambda}_0)}{G(\mathbf{x}_i,t_i,\widehat{\boldsymbol{\lambda}})}\frac{\delta_i}{p(\mathbf{x}_i,t_i)}r(\mathbf{x}_i,t_i)$$

$$B_{3,n}=\frac{1}{2}\left(\boldsymbol{\lambda}_0-\widehat{\boldsymbol{\lambda}}\right)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^nG''(\mathbf{x}_i,t_i,\boldsymbol{\xi})\frac{1}{G(\mathbf{x}_i,t_i,\widehat{\boldsymbol{\lambda}})}\frac{\delta_i}{p(\mathbf{x}_i,t_i)}r(\mathbf{x}_i,t_i)\sqrt{n}\left(\boldsymbol{\lambda}_0-\widehat{\boldsymbol{\lambda}}\right)\,.$$

The Bahadur expansion given in **NM5** implies that $\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0\right)=O_p(1)$, thus, using **NM4**d) we obtain that $B_{3,n}\xrightarrow{p}0$. Therefore, since $B_{2,n}=-\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0\right)^{\mathrm{T}}E\left((G'(\mathbf{x},t,\boldsymbol{\lambda}_0)/G(\mathbf{x},t,\boldsymbol{\lambda}_0))r(\mathbf{x},t)\right)+o_p(1)$, to derive the asymptotic distribution of B_n it is enough to study that of

$$\begin{aligned}C_n&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)-\sqrt{n}\left(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0\right)^{\mathrm{T}}E\left(\frac{G'(\mathbf{x},t,\boldsymbol{\lambda}_0)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}r(\mathbf{x},t)\right)+o_p(1)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\frac{\delta_i}{p(\mathbf{x}_i,t_i)}\psi_2\left(u_i\right)-\boldsymbol{\eta}(\delta_i,\mathbf{x}_i,t_i)^{\mathrm{T}}E\left(\frac{G'(\mathbf{x},t,\boldsymbol{\lambda}_0)}{G(\mathbf{x},t,\boldsymbol{\lambda}_0)}r(\mathbf{x},t)\right)\right]+o_p(1)\,,\end{aligned}$$

where the last equality follows from **NM5**. The proof follows now from the Central Limit Theorem. □

To derive the asymptotic distribution of $\widehat{\theta}^{(3)}$ we will need the following additional assumptions. For the sake of simplicity we will denote by $\mathbf{w} = (\mathbf{x}^{\mathrm{T}}, t)^{\mathrm{T}}$.

NM6. The missingness probability $p(\mathbf{w})$ is twice continuously differentiable.

NM7. The bandwidth b_n satisfies that $\rho_n^2 = \left\{nb_n^4 + (nb_n^{2(p+1)})^{-1}\right\} \to 0$

NM8. The kernel $K_1 : \mathbb{R}^{p+1} \to \mathbb{R}$ is bounded, nonnegative, has compact support and $\int K_1(\mathbf{u})d\mathbf{u} > 0$, $\int u_j K_1(\mathbf{u})d\mathbf{u} = 0$, for $1 \le j \le p+1$, $\int \|\mathbf{u}\|^2 K_1(\mathbf{u})d\mathbf{u} > 0$ and $\int u_j^2 K_1(\mathbf{u})d\mathbf{u} > 0$.

For the sake of simplicity, we will assume that $\int K_1(\mathbf{u})d\mathbf{u} = 1$.

Proof of Theorem 4.3. As in the proof of Theorem 4.2, using a Taylor's expansion of order two, we get that $0 = (1/\sqrt{n}) \sum_{i=1}^n (\delta_i / p_{n,b_n}(\mathbf{x}_i, t_i))\, \psi_2\left((y_i - \theta)/\varsigma_0\right) - \sqrt{n}(\widehat{\theta}^{(3)} - \theta) A_n^{(3)}(\psi_2)$ where

$$A_n^{(3)}(\psi_2) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{p_{n,b_n}(\mathbf{x}_i, t_i)}\psi_2'(u_i) + \frac{1}{2}(\widehat{\theta}^{(3)} - \theta)\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{p_{n,b_n}(\mathbf{x}_i, t_i)}\psi_2''\left(\frac{y_i - \xi_n}{\varsigma_0}\right)$$

and ξ_n is an intermediate point between $\widehat{\theta}^{(3)}$ and θ. Using **NM3**, it is easy to see that $A_n^{(3)}(\psi_2) \xrightarrow{p} A(\psi_2)$. Therefore, it is enough to show that $B_n = (1/\sqrt{n})\sum_{i=1}^n (\delta_i / p_{n,b_n}(\mathbf{x}_i, t_i))\psi_2(u_i) \xrightarrow{\mathcal{D}} N(0, \gamma_{\mathrm{S}}^2)$. Note that

$$B_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\delta_i}{p(\mathbf{w}_i)}\psi_2(u_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{p(\mathbf{w}_i)}{p_{n,b_n}(\mathbf{w}_i)} - 1\right)\frac{\delta_i}{p(\mathbf{w}_i)}\psi_2(u_i) = B_n^{(1)} + B_n^{(2)}\,,$$

where $p_{n,b_n}(\mathbf{w})$ is defined in (7). Denote by $f_n(\mathbf{w}) = \sum_{i=1}^n K_1\left((\mathbf{w}_i - \mathbf{w})/b_n\right)/\left(nb_n^{p+1}\right)$. As in Wang *et al.* (1997), we have that

$$\begin{aligned}
B_n^{(2)} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(p(\mathbf{w}_i) - p_{n,b_n}(\mathbf{w}_i)\right)\frac{\delta_i}{p^2(\mathbf{w}_i)}\psi_2(u_i) + O_p(\rho_n) \\
&= -\frac{1}{\sqrt{n}}\frac{1}{nb_n^{p+1}}\sum_{i=1}^n \frac{\sum_{j=1}^n K_1\left(\frac{\mathbf{w}_j - \mathbf{w}_i}{b_n}\right)\left(\delta_j - p(\mathbf{w}_i)\right)\delta_i}{f_n(\mathbf{w}_i)p^2(\mathbf{w}_i)}\psi_2(u_i) + O_p(\rho_n) \\
&= -\frac{1}{\sqrt{n}}\frac{1}{nb_n^{p+1}}\sum_{i=1}^n \frac{\sum_{j=1}^n K_1\left(\frac{\mathbf{w}_j - \mathbf{w}_i}{b_n}\right)\left(\delta_j - p(\mathbf{w}_i)\right)\left(\delta_i - p(\mathbf{w}_i)\right)}{f_n(\mathbf{w}_i)p^2(\mathbf{w}_i)}\psi_2(u_i) \\
&\quad - \frac{1}{\sqrt{n}}\frac{1}{nb_n^{p+1}}\sum_{i=1}^n \frac{\sum_{j=1}^n K_1\left(\frac{\mathbf{w}_j - \mathbf{w}_i}{b_n}\right)\left(\delta_j - p(\mathbf{w}_i)\right)}{f_n(\mathbf{w}_i)p(\mathbf{w}_i)}\psi_2(u_i) + O_p(\rho_n) = -B_{1,n} - B_{2,n} + O_p(\rho_n)\,.
\end{aligned}$$

Besides,

$$\begin{aligned}
B_{2,n} &= \frac{1}{\sqrt{n}}\frac{1}{nb_n^{p+1}}\sum_{j=1}^n\sum_{i=1}^n \frac{K_1\left(\frac{\mathbf{w}_j - \mathbf{w}_i}{b_n}\right)\left(\delta_j - p(\mathbf{w}_i)\right)}{f_n(\mathbf{w}_i)p(\mathbf{w}_i)}\left(\psi_2(u_i) - r(\mathbf{w}_i)\right) \\
&\quad + \frac{1}{\sqrt{n}}\frac{1}{nb_n^{p+1}}\sum_{j=1}^n\sum_{i=1}^n \frac{K_1\left(\frac{\mathbf{w}_j - \mathbf{w}_i}{b_n}\right)\left(\delta_j - p(\mathbf{w}_i)\right)r(\mathbf{w}_i)}{f_n(\mathbf{w}_i)p(\mathbf{w}_i)}\,.
\end{aligned}$$

Then, arguing as in Wang *et al.* (1997) and using standard $U-$statistics arguments, we get that $B_{2,n} = (1/\sqrt{n})\sum_{j=1}^{n} r(\mathbf{w}_j)\left(\delta_j - p(\mathbf{w}_j)\right)/p(\mathbf{w}_j) + O_p(\rho_n)$. On the other hand, using the same arguments as in Wang *et al.* (1997), we get that $B_{1,n} = O_p(\rho_n)$. Hence, we obtain that

$$B_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\delta_j}{p(\mathbf{w}_j)}\psi_2\left(\frac{y_j-\theta}{\varsigma_0}\right) - \frac{(\delta_j - p(\mathbf{w}_j))}{p(\mathbf{w}_j)}r(\mathbf{w}_j) + O_p(\rho_n)$$

and so, the Central Limit Theorem entails that $B_n \xrightarrow{\mathcal{D}} N(0,\gamma_{\mathrm{S}}^2)$ concluding the proof. $\square$

Acknowledgment. This work began while Graciela Boente and Ana Bianco were visiting the Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela. This research was partially supported by Grants PIP 112-200801-00216 from CONICET, PICT 0821 from ANPCYT and X-018 from the Universidad de Buenos Aires at Buenos Aires, Argentina and also by the DGICYT Spanish Grant MTM2008-03010 (European FEDER support included). We also wish to thank two anonymous referees for valuable comments which led to an improved version of the original paper.

References

1. Bianco A and Boente G (2004). Robust estimators in semiparametric partly linear regression models. *J. Statist. Planning Inf.*, **122**: 229-252.
2. Bianco A, Boente G, González–Manteiga W and Pérez–González A (2010). Estimation of the marginal location under a partially linear model with missing responses. *Comp. Statist. Data Anal.*, **54**,546-564.
3. Bianco A and Martínez E (2009). Robust testing in the logistic regression model. *Comp. Statist. Data Anal.*, **53**, 4095-4105.
4. Boente G, González–Manteiga W and Pérez–González A (2009). Robust nonparametric estimation with missing data. *J. Statist. Plann. Inference*, **139**: 571-592.
5. Boente G, He X and Zhou J (2006). Robust estimates in generalized partially linear models. *Ann. Statist.*, **34**, 2856-2878.
6. Croux, C and Haesbroeck, G (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273-295.
7. Härdle W, Liang H and Gao J (2000). *Partially linear models.* Springer-Verlag.
8. He X , Zhu Z and Fung W (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579-590.
9. Maronna R, Martin D and Yohai V (2006). *Robust statistics: Theory and methods* , Wiley, New York.
10. Pierce D (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**: 475-478.
11. Robins J, Rotnitzky A and Zhao L (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**: 846-866.
12. Robins J, Rotnitzky A and Zhao L (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**: 106-121.
13. Rosenbaum P (1987). Model–based direct adjustment. *J. Amer. Statist. Assoc.* **82**: 387-394.
14. van der Vaart A and Wellner J. (1996). *Weak Convergence and empirical processes. With applications to statistics.* Springer–Verlag, New York.
15. Wang C, Wang S, Zhao L and Ou S (1997). Weighted semiparametric estimation in regression analysis regression with missing covariates data. *J. Amer. Statist. Assoc.* **92**: 512-525.
16. Wang C, Wang S, Gutierrez R and Carroll R (1998). Local linear regression for generalized linear models with missing data. *Ann. Statist.* **26**: 1028-1050.
17. Wang Q, Linton O and Härdle W (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.*, **99**, 466, 334–345.
18. Wang Q and Sun Z (2007). Estimation in partially linear models with missing responses at random. *J. Multiv. Anal.*, **98**, 1470-1493.

Reports in Statistics and Operations Research

2005

05-01 SiZer Map for Evaluating a Bootstrap Local Bandwidth Selector in Nonparametric Additive Models. *M. D. Martínez-Miranda, R. Raya-Miranda, W. González-Manteiga and A. González-Carmona.*

05-02 The Role of Commitment in Repeated Games. *I. García Jurado, Julio González Díaz*.

05-03 Project Games. *A. Estévez Fernández, P. Borm, H. Hamers*

05-04 Semiparametric Inference in Generalized Mixed Effects Models. *M. J. Lombardía, S. Sperlich*

2006

06-01 A unifying model for contests: effort-prize games. J. González Díaz

06-02 The Harsanyi paradox and the "right to talk" in bargaining among coalitions. J. J. Vidal Puga

06-03 A functional analysis of NOx levels: location and scale estimation and outlier detection. M. Febrero, P. Galeano, W. González-Manteiga

06-04 Comparing spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga

06-05 On the spectral simulation of spatial dependence structures. R. M. Crujeiras, R. Fernández-Casal

06-06 An L_2-test for comparing spatial spectral densities. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.

2007

07-01 Goodness-of-fit tests for the spatial spectral density. R. M. Crujeiras, R. Fernández-Casal, W. González-Manteiga.

07-02 Presmothed estimation with left truncated and right censores data. M. A. Jácome, M. C. Iglesias-Pérez

07-03 Robust nonparametric estimation with missing data. G. Boente, W. González-Manteiga, A. Pérez-González

07-04 k-Sample test based on the common area of kernel density estimators, P. Martínez-Camblor, J. de Uña Álvarez, N. Corral-Blanco

07-05 A bootstrap based model checking for selection-biased data, J. L. Ojeda, W. González-Manteiga, J . A. Cristobal

07-06 The Gaussian mixture dynamic conditional correlation model: Bayesian estimation, value at risk calculation and portfolio selection, P. Galeano, M. C. Ausín

2008

08-01 ROC curves in nonparametric location-scale regression models, W. González-Manteiga, J. C. Pardo Fernández, I. Van Keilegom

08-02 On the estimation of α-convex sets, B. Pateiro-López, A. Rodríguez-Casal.

2009

09-01 Lasso Logistic Regression, GSoft and the Cyclyc Coordinate Descent Algorithm. Application to Gene Expression Data

2010

10-01 Asymptotic behaviour of robust estimators in partially linear models with missing responses: The effect of estimating the missing probability on simplified marginal estimators

Previous issues (2001 – 2004):
http://eio.usc.es/pub/reports.html