Universidade de Vigo

Estimation of the Marginal Location Under a Partially Linear Model with Missing Responses

Ana Bianco, Graciela Bonete, Wenceslao González Manteiga and Ana Pérez González

Report 08/08

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://eioweb.uvigo.es/
E-mail: depc05@uvigo.es

Universidade de Vigo

Estimation of the Marginal Location Under a Partially Linear Model with Missing Responses

Ana Bianco, Graciela Bonete, Wenceslao González Manteiga and Ana Pérez González

Report 08/08

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita: 

Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

Estimation of the marginal location under a partially linear model with missing responses

Ana Bianco
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina
Graciela Boente
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina
Wenceslao González–Manteiga
Universidad de Santiago de Compostela, Spain
and
Ana Pérez–González
Universidad de Vigo, Spain

Abstract

In this paper, we consider a semiparametric partially linear regression model where missing data occur in the response. We propose robust Fisher–consistent estimators for the regression parameter, the regression function and for the marginal location parameter of response variable. A cross–validation method is discussed, even when the marginal estimators seem not to be sensitive to the bandwidth parameter. Finally, a Monte Carlo study is carried out to compare the performance of the robust proposed estimators among them and also with the classical ones, in normal and contaminated samples, under different missing data models.

Corresponding Author
Graciela Boente
Moldes 1855, 3^o A
Buenos Aires, C1428CRA, Argentina
email: gboente@dm.uba.ar

AMS Subject Classification 1990: Primary 62F35, Secondary 62H25.
Key words and phrases: Fisher–consistency, Kernel Weights, $M-$location Functionals, Missing at Random, Nonparametric Regression, Robust Estimation.

1 Introduction

Partially linear regression models assume that the regression function can be modeled linearly on some covariates, while it depends nonparametrically on some others. To be more precise, assume that we have a response $y_i \in I\!R$ and covariates or design points $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ such that $\mathbf{x}_i \in I\!R^p$, $t_i \in [0,1]$ satisfying

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

with the errors ϵ_i i.i.d., independent of $(\mathbf{x}_i^{\mathrm{T}}, t_i)$ such that $E(\epsilon_i) = 0$ and $\mathrm{VAR}\left(\epsilon_i\right) < \infty$. Note that m stands for the regression function which is modeled linearly on $\mathbf{x}$ and nonparametrically on t.

As it is well known, most of the statistical methods in nonparametric and semiparametric regression models are designed for complete data sets and problems arise when missing observations are present which is a common situation in biomedical or socioeconomic studies, for instance. Typical examples are found in the field of social sciences with the problem of non-response in sample surveys, in physics, in genetics (Meng, 2000), among others. 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. This situation arises in many biological experiments where the explanatory variables can be controlled. This pattern is common, for example, in the scheme of double sampling proposed by Neyman (1938), where first a complete sample is obtained and then some additional covariate values are computed since perhaps this is less expensive than to obtain more response values. In this paper, we will thus assume that missing occurs only on the responses variables.

In the regression setting, a common method is to impute the incomplete observations and then proceed to carry out the estimation of the conditional or unconditional mean of the response variable with the complete sample. The methods considered include linear regression (Yates, 1933), kernel smoothing (Cheng, 1994; Chu and Cheng, 1995) nearest neighbor imputation (Chen and Shao, 2000), semiparametric estimation (Wang *et al.*, 2004), nonparametric multiple imputation (Aerts et al., 2002), empirical likelihood over the imputed values (Wang and Rao, 2002), among others. For a nonparametric regression model, González–Manteiga and Pérez–Gonzalez (2004) considered an approach based on local polynomials to estimate the regression function when the response variable y is missing but the covariate $\mathbf{x}$ is totally observed. Wang, Linton and Härdle (2004) considered inference on the mean of y under regression imputation of missing responses based on the semiparametric regression model (1). Under the setting considered in this paper, the missingness of y is allowed to depend on $(\mathbf{x}^{\mathrm{T}}, t)$. All the proposals considered up to now are very sensitive to anomalous observations since they are based on a local least squares approach. Recently, Boente, González–Manteiga and Pérez–González (2008) introduced a robust proposal to estimate the regression function under missingness in the response.

The goal of this paper is to introduce resistant estimators for the marginal location of y, say θ, under the partially linear model (1), 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 introduces the robust semiparametric estimators. An algorithm to compute the given estimators is described in Section 3, while their consistency is discussed in Section 4. In Section 5, a robust cross–validation method is discussed and a simulation

study is described in Section 6. The concluding remarks are in Section 7 and finally, technical proofs are given in the Appendix.

2 Proposals

We will consider inference with an incomplete data set $\left(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i, \delta_i\right)$, $1 \leq i \leq n$ where $\delta_i = 1$ if y_i is observed and $\delta_i = 0$ if y_i is missing and

$$y_i \quad = \quad \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_0 + g_0\left(t_i\right) + \epsilon_i \qquad 1 \leq i \leq n \ , \tag{2}$$

with errors ϵ_i independent, identically distributed with symmetric distribution $F_0(\cdot/\sigma_0)$.

Let $(Y, \mathbf{X}^{\mathrm{T}}, T, \delta)$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i, t_i, \delta_i)$. Our aim is to estimate, with the data set at hand, the regression parameter and the regression function robustly to provide a robust estimator for the marginal location parameter. An ignorable missing mechanism will be imposed by assuming that Y is missing at random (MAR), i.e., δ and Y are conditionally independent given $(\mathbf{X}, T)$, i.e.,

$$P\left(\delta = 1 | (Y, \mathbf{X}, T)\right) = P\left(\delta = 1 | (\mathbf{X}, T)\right) = p\left(\mathbf{X}, T\right) \ . \tag{3}$$

We will denote by $p = E\left(p\left(\mathbf{X}, T\right)\right)$.

We will consider kernel smoothers weights for the nonparametric component which are given by

$$w_i(t) = \frac{K\left(\dfrac{t_i - t}{h_n}\right)\delta_i}{\displaystyle\sum_{j=1}^{n} K\left(\frac{t_j - t}{h_n}\right)\delta_j}, \tag{4}$$

with K a kernel function, i.e., a nonnegative integrable function on $\mathbb{R}$ and h_n the bandwidth parameter. Note that the kernel weights are modified multiplying by the indicator of the missing variables in order to adapt to the complete sample and avoid bias.

2.1 Estimators of the regression parameter and regression function

In the classical case, the regression estimators are defined by considering preliminary estimators, $\widehat{\boldsymbol{\eta}}_n(t)$ and $\widehat{\eta}_{0,n}(t)$, of the quantities $\boldsymbol{\eta}(t) = E(\delta \mathbf{X} | T = t)/E(\delta | T = t)$ and $\eta_0(t) = E(\delta Y | T = t)/E(\delta | T = t)$, respectively. Note that using (3), δ is conditionally independent of Y and so we have that $\eta_0(t) = E(Y | T = t)$. Then, using the fact that

$$\delta Y = \delta \mathbf{X}^{\mathrm{T}} \boldsymbol{\beta}_0 + \delta g_0(T) + \delta \epsilon$$

and taking conditional expectation, it is easily obtained that

$$\eta_0(t) = \boldsymbol{\eta}(t)^{\mathrm{T}} \boldsymbol{\beta}_0 + g_0(t) \ ,$$

which implies that

$$\delta_i \left(y_i - \eta_0(t_i)\right) \;=\; \delta_i \left(\mathbf{x}_i - \boldsymbol{\eta}(t_i)\right)^{\mathrm{T}} \boldsymbol{\beta}_0 + \delta_i \epsilon_i \quad 1 \leq i \leq n \, .$$

Then, the estimator of the regression parameter $\boldsymbol{\beta}_0$, introduced by Wang, Linton and Härdle (2004), is defined as the value

$$\widehat{\boldsymbol{\beta}} \;=\; \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} \delta_i \left\{ \left(y_i - \widehat{\eta}_{0,n}(t_i)\right) - \left(\mathbf{x}_i - \widehat{\boldsymbol{\eta}}_n(t_i)\right)^{\mathrm{T}} \boldsymbol{\beta} \right\}^2 \, .$$

An explicit expression for $\widehat{\boldsymbol{\beta}}$ is given by

$$\widehat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} \delta_i \left(\mathbf{x}_i - \widehat{\boldsymbol{\eta}}_n(t_i)\right) \left(\mathbf{x}_i - \widehat{\boldsymbol{\eta}}_n(t_i)\right)^{\mathrm{T}} \right]^{-1} \sum_{i=1}^{n} \delta_i \left(y_i - \widehat{\eta}_{0,n}(t_i)\right) \left(\mathbf{x}_i - \widehat{\boldsymbol{\eta}}_n(t_i)\right) \, . \tag{5}$$

Since these estimators are based on weighted means of the response variables, they are higly sensitive to anomalous data. This suggests that some resistant estimation scheme needs to be considered. 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. We will consider the approach given in Boente, González–Manteiga and Pérez–González (2008) to estimate the regression functions. The simplified local $M-$smoother defined therein (*SLMS*) uses the information at hand and defines the estimator with the complete observations only. The main problem is that if we proceed as in Bianco and Boente (2004) with the complete sample, the conditions needed to ensure Fisher–consistency entail that $p\left(\mathbf{X}, T\right) = p(T)$, which eliminates many situations arising in practice. Thus, a profile–likelihood approach will be considered.

Let ψ 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 $\|\rho\|_\infty = 1$.

To define a robust estimator, we can proceed as follows

- **Step 1** For each t and $\boldsymbol{\beta}$, define $g_{\boldsymbol{\beta}}(t)$ and its related estimate $\widehat{g}_{\boldsymbol{\beta}}(t)$ using the simplified local $M-$smoothers defined in Boente, González-Manteiga and Pérez-González (2008). That is, $g_{\boldsymbol{\beta}}(t)$ and $\widehat{g}_{\boldsymbol{\beta}}(t)$ are, respectively, the solutions of

$$E\left[\delta \psi\left(\frac{Y - \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} - g_{\boldsymbol{\beta}}(t)}{\sigma}\right) | T = t\right] \;=\; 0 \, , \tag{6}$$

$$\sum_{i=1}^{n} w_i(t) \psi\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \widehat{g}_{\boldsymbol{\beta}}(t)}{\widehat{s}(t)}\right) \;=\; 0 \, , \tag{7}$$

  where $\widehat{s}(t)$ is a preliminary robust consistent scale estimator.

- **Step 2** The functional $\boldsymbol{\beta}(F)$ where F is the distribution of $(\delta, Y, \mathbf{X}^{\mathrm{T}}, T)$ is defined as

$$\boldsymbol{\beta}(F) \;=\; \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, E\left[\delta\rho\left(\frac{Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(T)}{\sigma}\right)\upsilon\left(\mathbf{X}\right)\right]$$

and its related estimate as

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

with $\widehat{\sigma}$ a preliminary estimate of the scale σ_0, i.e., a robust $M-$scale computed using an initial (possible inefficient) estimate of $\boldsymbol{\beta}$ with high breakdown point. Therefore, if ρ' denotes the derivative of ρ, the estimator is the solution of

$$\sum_{i=1}^{n}\delta_i\rho'\left(\frac{y_i-\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}-\widehat{g}_{\widehat{\boldsymbol{\beta}}}(t_i)}{\widehat{\sigma}}\right)\upsilon\left(\mathbf{x}_i\right)\left(\mathbf{x}_i+\frac{\partial}{\partial\boldsymbol{\beta}}\widehat{g}_{\boldsymbol{\beta}}(t_i)\Big|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}\right)=0\;. \tag{8}$$

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

An estimator of the regression function m is thus given by $\widehat{m}(\mathbf{x},t)=\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}+\widehat{g}_n(t)$.

In the Appendix, it is shown that, under mild conditions, these functionals are Fisher–consistent since the errors ϵ are independent of (δ, T).

Remark 2.1.1. As in nonparametric regression without missing observations, the aim of a robust smoother, as the local $M-$ estimator, is to provide reliable estimations when outliers observations are present in the responses y_i. Indeed, the researcher is seeking for consistent estimators of the regression functions $g_{\boldsymbol{\beta}}(t)$ and $m(\mathbf{x},t)$ without requiring moment conditions on the errors ϵ_i. This includes the well–known α-contaminated neighborhood for the errors distribution. More precisely, in a robust framework, one looks for procedures that remain valid when $\epsilon_i \sim F_0 \in \mathcal{F}_\epsilon = \{G:\; G(y)=(1-\alpha)G_0(y)+\alpha H(y)\}$, with H any symmetric distribution and G_0 a central model with possible first or second moments. In fact, the same framework can be considered in this paper. In these neighborhoods, no moment conditions are required to the errors and outliers correspond to deviations on the errors distribution.

On the other hand, as in any regression model, leverage points in the explanatory variables $\mathbf{x}$, can cause breakdown. To overcome this problem, $GM-$ and $S-$estimators have been introduced, see for instance, Maronna, Martin and Yohai (2006). In **Step 2**, we have considered a score function ρ combined with a weight υ to include both families of estimators. Our proposal is thus resistant against outliers in the residuals and in the carriers $\mathbf{x}$ as well.

2.2 Estimation 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 $\lambda(a,\sigma)=E\psi_1\left((Y-a)/\sigma\right)=0$ for all σ, where ψ_1 is an odd and bounded

score funtion. When $\psi_1(u) = \text{sg}(u)$, θ is the median of Y. Note that the same score functions ψ and ψ_1 can be considered both in **Step 1** and when computing the marginal parameter estimators.

Denote by $\widehat{\sigma}$, $\widehat{\sigma}_{\text{W}}$ and $\widehat{\sigma}_{\text{MA}}$ robust consistent estimators of the marginal scale of the variables involved, such as the MAD. Since, we only have the responses at hand the unknown values can be predicted by $\mathbf{x}_i^{\text{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)$, where $\widehat{g}_n(t)$ and $\widehat{\boldsymbol{\beta}}$ are defined in Section 2.1. Besides 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 $p(\mathbf{x}, t)$, such as the nonparametric kernel estimator

$$p_n(\mathbf{x}, t) = \frac{\displaystyle\sum_{i=1}^{n} K_1\left(\frac{\mathbf{x}_i - \mathbf{x}}{\lambda_n}\right) K_2\left(\frac{t_i - t}{b_n}\right) \delta_i}{\displaystyle\sum_{j=1}^{n} K_1\left(\frac{\mathbf{x}_j - \mathbf{x}}{\lambda_n}\right) K_2\left(\frac{t_j - t}{b_n}\right)} ,$$

where $K_1 : I\!R^p \to I\!R$ and $K_2 : I\!R \to I\!R$ are kernel functions and λ_n and h_n denote the smoothing parameters. If a parametric model is assumed, other choices for estimating $p(\mathbf{x}, t)$ can be considered.

- **Weighted Simplified $M-$estimate**. This estimate uses the complete sample and is the solution, $\widehat{\theta}_{\text{S}}$, of

$$\sum_{i=1}^{n} \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)} \psi_1\left(\frac{y_i - \widehat{\theta}_{\text{S}}}{\widehat{\sigma}}\right) = 0 .$$

- **Averaged $M-$estimate**. This estimator uses the predicted values to compute the marginal parameter estimator. If the errors distribution is symmetric, as assumed, and $Z = m(\mathbf{X}, T) = \mathbf{X}^{\text{T}}\boldsymbol{\beta}_0 + g_0(T) = \theta + u$ with u with symmetric distribution, we have that the median of the distribution of Y equals the median of Z. The same will happen to M-estimators, that is, Y and Z will have the same $M-$location, and so we get the estimator, $\widehat{\theta}_{\text{MA}}$ as the solution of

$$\sum_{i=1}^{n} \psi_1\left(\frac{\mathbf{x}_i^{\text{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i) - \widehat{\theta}_{\text{MA}}}{\widehat{\sigma}_{\text{MA}}}\right) = 0 .$$

- **Weighted Imputed $M-$estimate**. This estimator combines the ideas of the previous ones by imputing the missing responses. The following estimate can also be considered. The estimate $\widehat{\theta}_{\text{W}}$ is the solution of

$$\sum_{i=1}^{n} \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)} \psi_1\left(\frac{y_i - \widehat{\theta}_{\text{W}}}{\widehat{\sigma}_{\text{W}}}\right) + \sum_{i=1}^{n} \left(1 - \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)}\right) \psi_1\left(\frac{\mathbf{x}_i^{\text{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i) - \widehat{\theta}_{\text{W}}}{\widehat{\sigma}_{\text{W}}}\right) = 0 . \quad (9)$$

The Fisher–consistency of the related functionals is derived in the Appendix.

2.2.1 On the strong robustness

In the classical setting, the target parameter is the mean $\theta = E(Y)$. When considering $\psi_1(t) = \text{sgn}(t) = I_{(0,\infty)}(t) - I_{(-\infty,0)}(t)$ the target is now the median of the response Y. For general score

functions ψ_1, the target is the robust location functional related to ψ_1, as introduced in Huber (1981).

It is worth noticing that the assumption of symmetry required to the error's distribution is needed if we want to guarantee that we are estimating the same quantity when using all robust location functionals. Furthermore, the weak continuity of these robust location functionals for bounded score functions can be seen in Huber (1981). Therefore, by applying this functional to weak consistent estimators of the distribution of Y, we obtain consistent and asymptotically strongly robust estimators of θ. These results can thus be applied in our missing setting.

Note that all the estimators introduced in Section 2.2 can be written as $M-$functionals applied to some modified empirical distribution.

In fact, we have

- $\widehat{\theta}_{\mathrm{S}}$, is the solution of $\widehat{\lambda}_n(a,\widehat{\sigma})=0$ with

$$\begin{aligned}\widehat{\lambda}_n(a,\sigma) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\psi_1\left(\frac{y_i-a}{\sigma}\right)=\int\psi_1\left(\frac{y-a}{\sigma}\right)d\widehat{F}_n(y)\,,\\ \widehat{F}_n(y) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}I_{(-\infty,y]}(y_i)\,.\end{aligned}$$

  From the above discussion and since $\widehat{F}_n(y)$ provides weak consistent estimators of $F(y)$ under mild conditions, the weighted simplified $M-$estimator provides asymptotically strongly robust estimators.

- Denote by $\widehat{m}_i$ the predicted values using the partially linear model (1), $\widehat{m}_i=\widehat{m}(\mathbf{x}_i,t_i)=\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}+\widehat{g}_n(t_i)$. Then, $\widehat{\theta}_{\mathrm{MA}}$ is the solution of $\widetilde{\lambda}_n(a,\widehat{\sigma}_{\mathrm{MA}})=0$

$$\begin{aligned}\widetilde{\lambda}_n(a,\sigma) &= \frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{\widehat{m}_i-a}{\sigma}\right)=\int\psi_1\left(\frac{z-a}{\sigma}\right)d\widetilde{F}_n(z)\,,\\ \widetilde{F}_n(z) &= \frac{1}{n}\sum_{i=1}^{n}I_{(-\infty,z]}(\widehat{m}_i)\,.\end{aligned}$$

  In this case, if $\widehat{\boldsymbol{\beta}}$ and $\widehat{g}$ are robust consistent estimators of $\boldsymbol{\beta}_0$ and g_0, $\widetilde{F}_n$ will be a weak consistent estimator of the distribution, F_Z, of $Z=m(\mathbf{X},T)$. Thus, the average $M-$estimators give a sequence of asymptotically strongly robust estimators.

- $\widehat{\theta}_{\mathrm{W}}$ is the solution of $\widehat{\widehat{\lambda}}_n(a,\widehat{\sigma}_{\mathrm{W}})=0$ with

$$\begin{aligned}\widehat{\widehat{\lambda}}_n(a,\sigma) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\psi_1\left(\frac{y_i-a}{\sigma}\right)+\left(1-\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\right)\psi_1\left(\frac{\widehat{m}_i-a}{\sigma}\right)=\int\psi_1\left(\frac{y-a}{\sigma}\right)d\widehat{\widehat{F}}_n(y)\,,\\ \widehat{\widehat{F}}_n(y) &= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}I_{(-\infty,y]}(y_i)+\left(1-\frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\right)I_{(-\infty,y]}(\widehat{m}_i)\right]\,.\end{aligned}$$

  As it will be shown $\widehat{\widehat{F}}_n$ is a weak consistent estimator of F if $p_n(\mathbf{x},t)$ is a consistent estimator of $p(\mathbf{x},t)=P\left(\delta=1|(\mathbf{X},T)=(\mathbf{x},t)\right)$. Thus, the weighted imputed $M-$estimators provide a sequence of asymptotically strongly robust estimators.

3 Algorithm

3.1 Computation of the parametric and nonparametric components

We will consider kernel smoothers weights for the nonparametric component which are given by (4). In this section, we describe an algorithm, which is a slight modification of the procedure described in Maronna, Martin and Yohai (2006, Chapter 5).

Let ρ_0 and ρ be two bounded *rho*–functions such that $\rho_0 \geq \rho$. The algorithm to compute the estimator $\widehat{\boldsymbol{\beta}}$ defined in **Step 2** can be defined as given below, when $\boldsymbol{\beta} \in I\!R^p$, $p = 1, 2$. Otherwise, as is usual in linear regression models with $S-$estimators, for instance, a subsampling scheme must be considered. This is mainly due to the fact, that due to the profile–likelihood approach, we cannot ensure that a reweighted procedure will decrease the objective function.

- **Step A0** Take a net $\boldsymbol{\beta}_j$ of possible values for $\boldsymbol{\beta}$, $j = 1, \ldots, J$.
- **Step A1** Fix $1 \leq j \leq J$.

  We first compute the regression function estimate $\widehat{g}_{\boldsymbol{\beta}}(t)$ for each $\boldsymbol{\beta} = \boldsymbol{\beta}_j$ of the net and each t_i and also, an estimator for the error's ϵ_i scale.

  $\star$ For any $1 \leq i \leq n$, evaluate $\widehat{g}_{j,i} = \widehat{g}_{\boldsymbol{\beta}_j}(t_i)$ using the simplified $M-$ estimator introduced by Boente, González-Manteiga and Pérez-González (2008) applied to $\{(y_k - \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta}_j, t_k, \delta_k)\}_{1 \leq k \leq n}$ i.e., as the solution of

$$\sum_{k=1}^{n} w_k(t_i)\psi\left(\frac{y_k - \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta}_j - \widehat{g}_{j,i}}{\widehat{s}(t_i)}\right) \;=\; 0\,, \tag{10}$$

  where $\widehat{s}(t_i)$ is a preliminary robust scale estimator, such as, the local MAD, i.e, $\widehat{s}(t) = \mathrm{mad}_{k \in \mathcal{I}(h_n)} \left| r_{k,j} - \mathrm{median}_{\ell \in \mathcal{I}(h_n)}\left(r_{\ell,j}\right)\right|$ with $r_{k,j} = y_k - \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta}_j$ and $\mathcal{I}(h_n) = \{\ell : 1 \leq \ell \leq n \text{ and } |t_\ell - t| \leq h_n\}$.

  $\star$ Compute

$$L_0(\boldsymbol{\beta}_j) \;=\; \underset{1 \leq i \leq n : \delta_i = 1}{\mathrm{median}} \left(\left(y_i - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_j - \widehat{g}_{j,i}\right)^2\right)\,.$$

- **Step A2** In order to define the residuals scale estimator, let $\widehat{\boldsymbol{\beta}}_{\mathrm{INI}} = \beta_{j_0}$ be the preliminar estimator of $\boldsymbol{\beta}$ such that

$$\widehat{\boldsymbol{\beta}}_{\mathrm{INI}} = \underset{j}{\mathrm{argmin}}\, L_0(\boldsymbol{\beta}_j) = L_0(\boldsymbol{\beta}_{j_0})\,,$$

  and let $\widehat{g}_{\mathrm{INI},i}$ be the solution of (10) when using $\widehat{\boldsymbol{\beta}}_{\mathrm{INI}}$. Note that there is no need to evaluate again the solution $\widehat{g}_{\mathrm{INI},i}$, since they were already computed in **Step A1** for all the values of $\boldsymbol{\beta}$ in the grid.

  The estimator of the scale σ, $\widehat{\sigma}$, is then defined as the solution of

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i \rho_0\left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\mathrm{INI}} - \widehat{g}_{\mathrm{INI},i}}{\widehat{\sigma}}\right) = \frac{1}{2}\,. \tag{11}$$

- **Step A3** To compute the final estimator of $\boldsymbol{\beta}$, let

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

where $\widehat{g}_{j,i}$ are obtained in **Step A1** as the solution of (10). Note that, as in **Step A2**, $\widehat{g}_{j,i}$ do not need to be computed again since we have already calculated them in **Step A1**. We can take $\upsilon \equiv 1$.

Let $\widehat{\boldsymbol{\beta}}$ be the value minimizing L over the grid, i.e., $\widehat{\boldsymbol{\beta}} = \mathrm{argmin}_{1 \le j \le J} L(\boldsymbol{\beta}_j)$.

- **Step A4** The estimator of the nonparametric component is the solution $\widehat{g}_n(t) = \widehat{g}_{\widehat{\boldsymbol{\beta}}}(t)$ of

$$\sum_{i=1}^{n} w_i(t) \psi \left(\frac{y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}} - \widehat{g}_n(t)}{\widehat{s}(t)} \right) = 0 \ .$$

3.2 Computation of the robust marginal location estimators

To compute $\widehat{\theta}_{\mathrm{S}}$ and $\widehat{\theta}_{\mathrm{MA}}$, any standard algorithm to compute $M-$estimators can be used. For instance, they can be computed iteratively using reweighting, as described in the location setting in Chapter 2 of Maronna, Martin and Yohai (2006).

On the other hand, the following algorithm can be used to compute $\widehat{\theta}_{\mathrm{W}}$. Using that $\widehat{\theta}_{\mathrm{W}}$ is the solution of (9) and denoting by $W_1(u) = \psi_1(u)/u$, $p_i = p_n(\mathbf{x}_i, t_i)$ and $\widehat{m}_i = \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)$, we get that

$$\sum_{i=1}^{n} \frac{\delta_i}{p_i} \psi_1 \left(\frac{y_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) + \sum_{i=1}^{n} \left(1 - \frac{\delta_i}{p_i} \right) \psi_1 \left(\frac{\widehat{m}_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) = 0$$

and so

$$\widehat{\theta}_{\mathrm{W}} = \frac{\displaystyle\sum_{i=1}^{n} \left[\frac{\delta_i}{p_i} W_1 \left(\frac{y_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) y_i + \left(1 - \frac{\delta_i}{p_i} \right) W_1 \left(\frac{\widehat{m}_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) \widehat{m}_i \right]}{\displaystyle\sum_{i=1}^{n} \left[\frac{\delta_i}{p_i} W_1 \left(\frac{y_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) + \left(1 - \frac{\delta_i}{p_i} \right) W_1 \left(\frac{\widehat{m}_i - \widehat{\theta}_{\mathrm{W}}}{\widehat{\sigma}_{\mathrm{W}}} \right) \right]} \ .$$

Let $\theta^{(0)} = \widehat{\theta}_{\mathrm{MA}}$ and $\widehat{\sigma}_{\mathrm{MA}} = \underset{1 \le i \le n}{\mathrm{mad}}(\widehat{m}_i)$ The algorithm can be defined as follows

- For $k = 0, 1, \ldots$, given $\theta^{(k)}$ define

$$\theta^{(k+1)} = \frac{\displaystyle\sum_{i=1}^{n} \left[\frac{\delta_i}{p_i} W_1 \left(\frac{y_i - \theta^{(k)}}{\widehat{\sigma}_{\mathrm{W}}} \right) y_i + \left(1 - \frac{\delta_i}{p_i} \right) W_1 \left(\frac{\widehat{m}_i - \theta^{(k)}}{\widehat{\sigma}_{\mathrm{W}}} \right) \widehat{m}_i \right]}{\displaystyle\sum_{i=1}^{n} \left[\frac{\delta_i}{p_i} W_1 \left(\frac{y_i - \theta^{(k)}}{\widehat{\sigma}_{\mathrm{W}}} \right) + \left(1 - \frac{\delta_i}{p_i} \right) W_1 \left(\frac{\widehat{m}_i - \theta^{(k)}}{\widehat{\sigma}_{\mathrm{W}}} \right) \right]}$$

- Iterate until convergence or for a fixed number of steps $k_{\max}$.

4 Main results

In this section, we will derive the strong consistency of the marginal location $M-$estimators under the following conditions:

A1 $\psi_1 : I\!R \to I\!R$ is a bounded, differentiable function with bounded derivative ψ_1', such that $\int |\psi_1'(u)|du < \infty$.

A2 $\inf_{(\mathbf{x},t)} p(\mathbf{x},t) = A > 0$.

A3 $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p(\mathbf{x},t)| \xrightarrow{a.s.} 0$.

Assumption **A1** is a standard condition on the score function ψ, while **A2** states that response variables are observed, which is a common assumption in the literature.

Theorem 4.1. *Assume that **A2** and **A3** hold. Then,*

a) $\|\widehat{F}_n - F\|_\infty = \sup_y |\widehat{F}_n(y) - F(y)| \xrightarrow{a.s.} 0$.

b) If in addition $\widehat{\sigma} \xrightarrow{a.s.} \sigma_0$, ***A1*** *holds and in a neighborhood of* θ, *the function* $\lambda(a, \sigma_0)$ *has a unique change of sign, there exists a solution* $\widehat{\theta}_{\mathrm{S}}$ *of* $\widehat{\lambda}_n(a, \widehat{\sigma}) = 0$, *such that* $\widehat{\theta}_{\mathrm{S}} \xrightarrow{a.s.} \theta$.

Theorem 4.2. *Denote by* $\Pi(Q, P)$ *the Prohorov distance between the probability measures* Q *and* P. *Let* $\widetilde{m}(\mathbf{x}, t)$ *be an estimator of* $m(\mathbf{x}, t)$ *such that for any compact set* $\mathcal{K}_1 \in I\!R^p$ $\mathcal{K}_2 \in I\!R$

$$\sup_{\substack{\mathbf{x}\in\mathcal{K}_1 \\ t\in\mathcal{K}_2}} |\widetilde{m}(\mathbf{x},t) - m(\mathbf{x},t)| \xrightarrow{a.s.} 0 \ .$$

Then,

a) $\Pi(\widetilde{P}_n, P_Z)| \xrightarrow{a.s.} 0$ *where* P_Z *is the probability measure induced by* $Z = m(\mathbf{X}, T)$ *and*

$$\widetilde{P}_n(A) = \frac{1}{n}\sum_{i=1}^n I_A(\widehat{m}(\mathbf{x}_i, t_i)) = \frac{1}{n}\sum_{i=1}^n I_A(\widehat{m}_i) \ .$$

b) If in addition $\widehat{\sigma}$ *is an estimator of the scale* σ_Z *of* Z *such that* $\widehat{\sigma} \xrightarrow{a.s.} \sigma_Z$, ***A1*** *holds and in a neighborhood of* θ, *the function* $\lambda(a, \sigma_Z)$ *has a unique change of sign, there exists a solution* $\widetilde{\theta}$ *of* $\widetilde{\widetilde{\lambda}}_n(a, \widehat{\sigma}) = 0$, *such that* $\widetilde{\theta} \xrightarrow{a.s.} \theta$ *where*

$$\begin{aligned}
\widetilde{\widetilde{\lambda}}_n(a,\sigma) &= \frac{1}{n}\sum_{i=1}^n \psi_1\left(\frac{\widetilde{m}_i - a}{\sigma}\right) = \int \psi_1\left(\frac{z-a}{\sigma}\right) d\widetilde{\widetilde{F}}_n(z) \\
\widetilde{m}_i &= \widetilde{m}(\mathbf{x}_i, t_i) \\
\widetilde{\widetilde{F}}_n(z) &= \frac{1}{n}\sum_{i=1}^n I_{(-\infty,z]}(\widetilde{m}_i) \ .
\end{aligned}$$

Note that Theorem 4.2 entails the following result.

Corollary 4.1. *Assume that*

i) $\widehat{\boldsymbol{\beta}} \xrightarrow{a.s.} \boldsymbol{\beta}_0$.

ii) For any compact set $\mathcal{K}$, $\sup_{t\in\mathcal{K}} |\widehat{g}_n(t) - g_0(t)| \xrightarrow{a.s.} 0$.

If in addition $\widehat{\sigma}_{\mathrm{MA}} \xrightarrow{a.s.} \sigma_Z$, ***A1*** *holds and in a neighborhood of* θ, *the function* $\lambda(a, \sigma_Z)$ *has a unique change of sign, there exists a solution* $\widehat{\theta}_{\mathrm{MA}}$ *of* $\widetilde{\lambda}_n(a, \widehat{\sigma}_{\mathrm{MA}}) = 0$, *such that* $\widehat{\theta}_{\mathrm{MA}} \xrightarrow{a.s.} \theta$.

Theorem 4.3. *Assume that* ***A2*** *and* ***A3*** *hold. Then,*

a) $\|\widehat{\widehat{F}}_n - F\|_\infty = \sup_y |\widehat{\widehat{F}}_n(y) - F(y)| \xrightarrow{a.s.} 0$.

b) If in addition $\widehat{\sigma}_{\mathrm{W}} \xrightarrow{a.s.} \sigma_0$, ***A1*** *holds and in a neighborhood of* θ, *the function* $\lambda(a, \sigma_0)$ *has a unique change of sign, there exists a solution* $\widehat{\theta}_{\mathrm{W}}$ *of* $\widehat{\widehat{\lambda}}_n(a, \widehat{\sigma}_{\mathrm{W}}) = 0$, *such that* $\widehat{\theta}_{\mathrm{W}} \xrightarrow{a.s.} \theta$.

4.1 Some comments

It is worth noticing that Theorem 4.3 entail that $\widehat{\theta}_{\mathrm{W}} \xrightarrow{a.s.} \theta$ even if the estimators of the regression function are not consistent when we estimate consistently the probability of missing. Obviously, the same happens with $\widehat{\theta}_{\mathrm{S}}$ that uses the observations at hand.

On the other hand, $\widehat{\theta}_{\mathrm{MA}}$ is consistent if the regression model is correct without any need of estimating the probability of missing.

One could try to combine both proposals in order to get the double–protected property in the sense of Scharfstein, Rotnitzky and Robins (1999). Let

$$\widehat{\theta} = \widehat{\theta}_{\mathrm{MA}} + \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)} \left(\widehat{\theta}_{\mathrm{S}} - \widehat{\theta}_{\mathrm{WMA}}\right)$$

with $\widehat{\theta}_{\mathrm{WMA}}$ the solution of

$$\sum_{i=1}^{n} \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)} \psi_1\left(\frac{\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i) - \widehat{\theta}_{\mathrm{WMA}}}{\widehat{\sigma}_{\mathrm{WMA}}}\right) = 0\,,$$

where $\widehat{\sigma}_{\mathrm{WMA}}$ is a preliminary robust scale estimator. It is clear that, if **A3** holds, $\widehat{\theta} \xrightarrow{a.s.} \theta$ since

- $\widehat{\theta}_{\mathrm{S}} \xrightarrow{a.s.} \theta$
- $\widehat{\theta}_{\mathrm{WMA}} \xrightarrow{a.s.} \theta_{\mathrm{WMA}}(F)$
- $\widehat{\theta}_{\mathrm{MA}} \xrightarrow{a.s.} \theta_{\mathrm{MA}}(F)$
- $\theta_{\mathrm{MA}}(F) = \theta_{\mathrm{WMA}}(F)$,

where $\theta_{\mathrm{MA}}(F)$ and $\theta_{\mathrm{WMA}}(F)$ stand for the functionals related to each proposal (see Section 8.3).

However, if $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p^\star(\mathbf{x},t)| \xrightarrow{a.s.} 0$ with $p^\star(\mathbf{x},t) \neq p(\mathbf{x},t)$ but $m(\mathbf{x},t) = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(t)$, from the consistency of $\widehat{\theta}_{\mathrm{MA}}$, we get that

$$\widehat{\theta} \xrightarrow{a.s.} \theta + E\left(\frac{p(\mathbf{X},T)}{p^\star(\mathbf{X},T)}\right)(\theta_{\mathrm{S}}(F) - \theta_{\mathrm{WMA}}(F))$$

and we cannot ensure that both location functionals, $\theta_{\mathrm{S}}(F)$ and $\theta_{\mathrm{WMA}}(F)$, will be equal. However, note that, when ψ_1 is the identity function, this equality holds due to the linearity of the expectation.

Moreover, assume that there exists a $M-$functional $\theta(F)$ such that if $\widehat{\theta}$ is the corresponding estimator then, $\widehat{\theta}$ satisfies the double–protected property. Hence, assuming that scale σ is known, we must have

a) If $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p(\mathbf{x},t)| \xrightarrow{a.s.} 0$, then $\theta(F)$ should be equal to $\theta_{\mathrm{S}}(F)$, i.e., it should satisfy

$$E\frac{\delta}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta(F)}{\sigma}\right) = 0$$

b) If $\sup_{(\mathbf{x},t)} |\widehat{m}_n(\mathbf{x},t) - m(\mathbf{x},t)| \xrightarrow{a.s.} 0$, then $\theta(F)$ should be equal to $\theta_{\mathrm{MA}}(F)$, i.e., it should satisfy

$$E\psi_1\left(\frac{m(\mathbf{X},T)-\theta(F)}{\sigma}\right) = 0$$

Thus, if one wants to obtain a robust and double–protected $M-$estimator, both equations should be satisfied. Clearly, when $\psi_1 \equiv \mathrm{id}$, this is fulfilled since the errors ϵ are independent of $(\mathbf{x},t)$ and have symmetric distribution.

So, if $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p^\star(\mathbf{x},t)| \xrightarrow{a.s.} 0$ and $\sup_{(\mathbf{x},t)} |\widehat{m}(\mathbf{x},t) - m^\star(\mathbf{x},t)| \xrightarrow{a.s.} 0$, we need that

a) if $p^\star(\mathbf{x},t) = p(\mathbf{x},t)$, then $\theta(F)$ will satisfy

$$E\frac{\delta}{p^\star(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta(F)}{\sigma}\right) = E\frac{p^\star(\mathbf{X},T)}{p^\star(\mathbf{X},T)}\psi_1\left(\frac{m(\mathbf{X},T)+\epsilon-\theta(F)}{\sigma}\right) = 0\,,$$

or equivalently in this situation

$$E\psi_1\left(\frac{m(\mathbf{X},T)+\epsilon-\theta(F)}{\sigma}\right) = 0\,.$$

b) On the other hand, if $m^\star(\mathbf{x},t) = m(\mathbf{x},t)$, $\theta(F)$ should satisfy

$$E\psi_1\left(\frac{m(\mathbf{X},T)-\theta(F)}{\sigma}\right) = 0\,.$$

For the regular score functions used in robustness this seems difficult to be attained due to the non–linearity of ψ_1.

Another possibility could be to consider the solution $\widehat{\theta}$ of $\lambda_n^\star(\widehat{\theta}, \widehat{\sigma}_{\mathrm{WMA}}) = 0$ with

$$\lambda_n^\star(a,\sigma) = \widetilde{\lambda}_n(a,\sigma) + \left(\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\right)\frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\left[\psi_1\left(\frac{y_i-a}{\sigma}\right) - \psi_1\left(\frac{\widehat{m}_i-a}{\sigma}\right)\right] = 0\,,$$

where $\widehat{m}_i$ are the predicted values using the partially linear model (1), $\widehat{m}_i = \widehat{m}(\mathbf{x}_i,t_i) = \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)$ and $\widetilde{\lambda}_n(a,\sigma)$ was defined in Section 2.2.1 as

$$\widetilde{\lambda}_n(a,\sigma) = \frac{1}{n}\sum_{i=1}^n \psi_1\left(\frac{\widehat{m}_i-a}{\sigma}\right)\,.$$

Note that if $\sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p^\star(\mathbf{x},t)| \xrightarrow{a.s.} 0$ and $\sup_{(\mathbf{x},t)} |\widehat{m}(\mathbf{x},t) - m^\star(\mathbf{x},t)| \xrightarrow{a.s.} 0$, $\widehat{\theta}$ will be consistent to the solution $\theta^\star(F)$ of $\lambda^\star(a,\sigma) = 0$ with

$$\lambda^\star(a,\sigma) = \lambda_Z(a,\sigma) + \left(E\frac{p(\mathbf{X},T)}{p^\star(\mathbf{X},T)}\right)E\left\{\frac{p(\mathbf{X},T)}{p^\star(\mathbf{X},T)}\left[\psi_1\left(\frac{Y-a}{\sigma}\right) - \psi_1\left(\frac{Z-a}{\sigma}\right)\right]\right\}\,,$$

where $\lambda_Z(a,\sigma) = E\psi_1((Z-a)/\sigma)$, $Z = m^\star(\mathbf{X},T)$. Again,

a) If $p^\star(\mathbf{x},t) = p(\mathbf{x},t)$, then $\lambda^\star(a,\sigma) = E\psi_1((Y-a)/\sigma)$ and so $\theta^\star(F) = \theta_{\mathrm{S}}(F)$, attaining the desired Fisher–consistency.

b) If $m^\star(\mathbf{x},t) = m(\mathbf{x},t)$, then, if $R(\mathbf{X},T) = p(\mathbf{X},T)/p^\star(\mathbf{X},T)$, we have

$$\begin{aligned}\lambda^\star(a,\sigma) &= E\psi_1\left(\frac{m(\mathbf{X},T)-a}{\sigma}\right)\\ &+ (ER(\mathbf{X},T))\,E\left(R(\mathbf{X},T)\left[\psi_1\left(\frac{m(\mathbf{X},T)+\epsilon-a}{\sigma}\right) - \psi_1\left(\frac{m(\mathbf{X},T)-a}{\sigma}\right)\right]\right)\,.\end{aligned}$$

Thus, using that $m(\mathbf{X},T)$ has a symmetric distribution around θ, we obtain

$$\lambda^\star(\theta,\sigma) = (ER(\mathbf{X},T))\,E\left(R(\mathbf{X},T)\left[\psi_1\left(\frac{m(\mathbf{X},T)+\epsilon-\theta}{\sigma}\right) - \psi_1\left(\frac{m(\mathbf{X},T)-\theta}{\sigma}\right)\right]\right)\,.$$

So, if we want that $\theta^\star(F) = \theta_{\mathrm{MA}}(F) = \theta$ we need that

$$E\left(R(\mathbf{X},T)\left[\psi_1\left(\frac{m(\mathbf{X},T)+\epsilon-\theta}{\sigma}\right) - \psi_1\left(\frac{m(\mathbf{X},T)-\theta}{\sigma}\right)\right]\right) = 0\,. \tag{12}$$

Equation (12) will be fulfilled for instance if the ratio $R(\mathbf{X},T)$ is an even function of $m(\mathbf{X},T)-\theta$. This assumption seems unnatural and that is why these estimators where not considered in this paper. Again, if ψ_1 is the identity function, (12) is automatically fulfilled.

Analogous conclusions can be obtained if we define, as in Wang and Sun (2007), the $M-$location estimator of the weighted responses $(\delta_i/p_n(\mathbf{x}_i,t_i))\,y_i + (1-\delta_i/p_n(\mathbf{x}_i,t_i))\left(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)\right)$.

5 Selection of the smoothing parameter

As it is well known, the bias and variance of the marginal location estimators is less sensitive to the bandwidth than in other semiparametric settings, see for instance, Cheng (1994). This fact was also mentioned by Wang and Sun (2007) who pointed out that the bandwidth selection is not so critical when estimating the parametric component since we are dealing with global functional with root$-n$ rate of convergence. However, these authors recommend a least squares cross–validation scheme to choose the smoothing parameter.

The sensitivity to outliers of the classical methods for the selection of the smoothing parameter has been widely discussed in nonparametric regression, for independent observations and complete data sets. Because it is based on squared residuals, least squares cross–validation is very sensitive to outliers, even when it is used with local $M-$estimators. As noted by Wang and Scott (1994), in the presence of outliers, the least squares cross–validation function is nearly constant on its whole domain and thus, essentially worthless for the purpose of choosing a bandwidth. Moreover, it can be seen that just one outlier may cause the bandwidth (and so the estimate) to break down, in the sense that it often results in oversmoothing or undersmoothing. Boente and Fraiman (1991b) pointed out that robust cross–validation methods should be an alternative. Also, Wang and Scott (1994) proposed an L^1 cross–validation method in order to avoid the problems of L^2 cross–validation, while Cantoni and Ronchetti (2001) considered a resistant choice of the smoothing parameter for smoothing splines based on a robust version of C_p and of cross-validation. A similar proposal was suggested by Leung *et al.* (1993) for kernel $M-$smoothers. On the other hand, the classical plug–in bandwidth selector also breaks down in the presence of outliers. Boente *et al.* (1997) proposed a robust plug–in bandwidth selection procedure in noparametric regression. The problem of choosing robustly the bandwidth parameter under a partially linear regression model was studied by Boente and Rodriguez (2007), see also Bianco and Boente (2007) who considered the dependent setting. On the other hand, Boente, González–Manteiga and Pérez-González (2007) proposed a robust cross–validation method for a nonparametric regression model with missing responses.

We will adapt the ideas of robust cross–validation to the present situation. Let $\widehat{\theta}$ be the robust estimator to be considered, i.e., the average or the weighted imputed one, that depends on a previous smoothing. Denote by $\widehat{\theta}^{(i)}(h)$ the estimator computed with bandwidth h using all the data except $(y_i, \mathbf{x}_i^{\mathrm{T}}, t_i)$. Then, the classical least squares cross–validation method constructs an asymptotically optimal data–driven bandwidth and thus, adaptive data–driven estimators, by minimizing

$$CV_{\mathrm{LS}}(h) = n^{-1} \sum_{i=1}^{n} \delta_i \left(y_i - \widehat{\theta}^{(i)}(h) \right)^2 w^2 \left(t_i \right) \ ,$$

where the weight function w protects against boundary effects. In the classical setting, linear smoothers and least squares regression estimators are used, while if one tries to obtain resistant procedures, the proposals given in Section 2 should be considered but combined with a robust loss function. Taking into account that the classical cross–validation criterion tries to measure both bias and variance, it would be sensible to introduce, as in Bianco and Boente (2007) a new measure that establishes a trade–off between robust measures of bias and variance. Let μ_n and σ_n denote robust estimators of location and scale, respectively. A robust cross–validation criterion can be

defined by minimizing on h

$$RCV_{\mathrm{R}}(h) = \mu_n^2\left(\widehat{r}_i(h), w\left(t_i\right)\right) + \sigma_n^2\left(\widehat{r}_i(h), w\left(t_i\right)\right) \ ,$$

where $\widehat{r}_i(h) = y_i - \widehat{\theta}^{(i)}(h)$ are the residuals and $\mu_n\left(u_i, w_i\right)$ and $\sigma_n\left(u_i, w_i\right)$ indicates that to compute the robust location and scale, respectively, each observation u_i receives a weight w_i. For instance, when $w(t) = I_{\mathcal{C}}\left(t\right)$, we compute the robust location of the residuals $\widehat{r}_i(h)$ such that $t_i \in \mathcal{C}$, the other ones being discarded. As location estimator, μ_n, one can consider the median while σ_n can be taken as the bisquare a–scale estimator or the Huber $\tau-$scale estimator. For the situation we are dealing with, it is enough, to compute RCV_{R} with the observations at hand, i.e, to compute RCV_{R} we use only the observed residuals $\{\widehat{r}_i\}_{i:\delta_i=1}$ and discard the incomplete vectors.

6 Monte Carlo study

A simulation study was carried out when the regression parameter has dimension 1. The S–code is available upon request to the authors.

In all Tables and Figures WSE_{LS}, AE_{LS} and WIE_{LS} denote the classical estimates obtained using the weighted simplified estimate, the averaged estimate and the weighted imputed estimate, respectively. On the other hand, the corresponding robust estimates are denoted as WSE_{R}, AE_{R} and WIE_{R}.

The aims of this study are

- to compare the behavior of the classical and robust estimators under contamination and under normal samples, for different missing probabilities.
- to study the behavior of the robust proposals, the weighted simplified, the averaged and the weighted imputed estimators, between them and compared to that of the robust estimator (WSE_{R}) that would be computed if the complete data set were available. Note that this estimator, which corresponds to $p \equiv 1$, cannot be computed in practice. The aim is to detect which of the proposals would give mean square errors closer to those obtained if there were no missing responses.

In both, the clasical and robust smoothing procedures, we have used the gaussian kernel with standard deviation $\frac{0.25}{0.675} = 0.37$ such that the interquartile range is 0.5. The robust smoothing procedure used local $M-$estimates with score function ψ the bisquare function with tuning constant 4.685, and local medians as initial estimate. The chosen tuning constant for the local $M-$estimator gives a 95% efficiency with respect to its linear relative. The same score function was used to compute the marginal estimators, that is, we choose $\psi_1 = \psi$.

The robust estimator of the regression parameter $\boldsymbol{\beta}$ was computed as described in Section 3 using as *rho*–functions the bisquare function, that is,

$$\rho_0(x) = \rho_1\left(\frac{x}{c_0}\right) \quad \text{and} \quad \rho(x) = \rho_1\left(\frac{x}{c_1}\right)$$

with $c_0 = 1.56$, $c_1 \geq c_0$ and $\rho_1(x) = \min(1, 1-(1-x^2)^3)$. The value selected for c_0 ensures Fisher–consistency of the scale when the errors are gaussian, while $c_1 = 4.68$ guarantees that under a regression model the $MM-$estimates will have 95% efficiency.

Due to the expensive computing time when evaluating the robust estimators, we only performed 500 replications generating independent samples of size $n = 100$ following the model

$$z_i = \boldsymbol{\beta}\mathbf{x}_i + 2\sin(4\pi(t_i - 0.5)) + \epsilon_i \quad 1 \leq i \leq n\,,$$

where $\boldsymbol{\beta} = 2$, $\left(\mathbf{x}_i^{\mathrm{T}}, t_i\right)^{\mathrm{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0, \frac{1}{2})^{\mathrm{T}}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/\left(6\sqrt{3}\right) \\ 1/\left(6\sqrt{3}\right) & 1/36 \end{pmatrix}$, and $\epsilon_i \sim N(0, \sigma^2)$ with $\sigma^2 = 0.25$ in the non–contaminated case.

The results for normal data sets will be indicated by C_0 in the Tables, while C_1 to C_4 will denote the following contaminations:

- C_1: $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 25\sigma^2)$. This contamination corresponds to inflating the errors and thus, will only affect the variance of the location estimates.
- C_2: $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 25\sigma^2)$ and artificially 10 observations of the response z_i but not of the carriers $\mathbf{x}_i$, were modified to be equal to 20 at equally spaced values of t. This case corresponds to introduce outliers with high–residuals. The aim of this contamination is to study changes in bias in the estimation of the location parameter.
- C_3: $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 25\sigma^2)$ and artificially 10 observations of the carriers $\mathbf{x}_i$ but not of the response z_i, were modified to be equal to 20 at equally spaced values of t. This case corresponds to introduce high–leverage points. The aim of this contamination is to study changes in bias in the estimation of the location parameter when using the averaged and the weighted imputed estimates, since this contamination affects mainly the estimation of the regression parameter.
- C_4: $\epsilon_1, \ldots, \epsilon_n$, are i.i.d. $0.9N(0, \sigma^2) + 0.1N(0, 25\sigma^2)$ and artificially 5 observations of the carriers $\mathbf{x}_i$ and 5 of the response z_i, were modified to be equal to 20 and -20, respectively at equally spaced values of t. The modified observations at the response were not allocated at the same t as those of the carriers. This case corresponds to introduce both high–leverage points and high–residuals. The aim of this contamination is to study changes in bias in the different estimators of the location parameter since this contamination affects the regression parameter and also the marginal one.

Let $\pi(u) = 0.4 + 0.5(\cos(2u + 0.4))^2$. We then define $y_i = z_i$, if $\delta_i = 1$, and missing otherwise to obtain the missing responses, where the missing probability considered (see (3)) are:

a) $p(x, t) \equiv 1$ that corresponds to the complete data situation. As in nonparametric regression with missing responses, in this case, the WSE_{LS} and WIE_{LS} give the same results. However, even if WSE_{R} and WIE_{R} also should be identical, they provide slightly different results since the algorithms used to compute them where not identical. To compute WSE_{R} we have used the S–plus routine *location.m* with 20 iterations while to compute the WIE_{R}, we used the

reweighted method described in Section 3.2 with $k_{\max} = 10$, i.e., only 10 iterations were performed.

b) $p(x,t) \equiv 0.8$, corresponding to data missing completely at random.

c) $p(x,t) = \pi(t)$.

d) $p(x,t) = \pi(x)$.

e) $p(x,t) = \pi(x\,t)$.

In all cases, we have not estimated the missing probabilities but we have taken $p_n(x,t) = p(x,t)$ to avoid increasing the computational time and to increase biases due to the estimation of the missing probability (see, for instance, Chen *et al.* (2006)). Moreover, as discussed in Wang *et al.* (1998), it seems natural to argue that the weighted estimators using the estimated probabilities are at least as efficient as those using the true model.

In order to study the sensitivity of the resulting estimator to the bandwidth, Table 1 shows the minimum, mean and maximum values of CV_{LS} as a function of h for the missing models b) to e), for the average and weighted average estimators and for one of the samples generated as above. We have generated the sample according to C_0 and to C_2 to show the sensitivity of the least squares procedure and we have computed the cross–validation errors for a grid of 20 equally spaced values of h between 0.05 and 1. As it can be seen, when considering both the classical or the robust procedures, for non–contaminated samples, the cross–validation error of all the estimators is almost constant on its domain showing the lack of sensitivity of the marginal estimators to the smoothing parameter. It is also clear that the least–squares cross–validation error is highly sensitive to anomalous data, since its values are almost 7 times those obtained with the non–contaminated samples. The robust procedure has a behavior similar to that described for the least squares method, under C_0. For instance, when considering $p(x,t) = 0.4 + 0.5(\cos(2t + 0.4))^2$, the minimum (m) and maximum values (M) of the robust cross–validation function related to AE_{R} and WIE_{R} are $m_{AE_{\text{R}}} = 7.3511$, $M_{AE_{\text{R}}} = 7.3953$ and $m_{WIE_{\text{R}}} = 7.3601$ and $M_{WIE_{\text{R}}} = 7.3943$, respectively. The robust cross–validation $RCV_{\text{R}}(h)$ is much more stable under contamination. Effectively, under C_2, the minimum (m) and maximum values (M) of the robust cross–validation function are $m_{AE_{\text{R}}} = 12.0666$, $M_{AE_{\text{R}}} = 12.1572$ and $m_{WIE_{\text{R}}} = 11.9976$ and $M_{WIE_{\text{R}}} = 12.0888$ and the shape of the function is almost the same as in the non–contaminated situation.

Considering the above discussion, a robust cross–validation procedure was not performed in this preliminary study, taking into account that it is very expensive computationally when it is combined with the robust profile procedure and since it was clear from the results obtained that, in all situations, the bandwidth choice did not seem crucial for the estimation of the marginal location parameter. Even when we have performed the simulation with bandwidths $h = 0.1$, 0.2 and 0.4, we only present in this paper the results for $h = 0.2$. In fact, all the considered bandwidths lead to the same conclusions.

The perfomance of the location estimators was measured using the bias, standard deviation and mean square error (bias, sd and MSE, respectively) in Tables 2 to 6. Also, boxplots are given in Figures 1 to 5.

The results reported in Tables 2 to 6 show that when there is no contamination, the linear estimators perform better than the robust ones (both in bias and mean square error) showing the loss of efficiency related to the score function used to compute the location $M-$estimators. On the other hand, the robust estimators show their advantage over the classical ones when outliers with high residuals are present having a similar performance when contaminating the errors. In fact, the MSE of the classical estimators is more than 20 times larger than the one observed under no contamination and also than the MSE of the robust estimators which are much more stable under C_2 or C_4. Note that this is mainly due to the increased bias of the classical estimators using any of the methods (weighted simplified, averaged or weighted imputed one). This explains the better efficiency of the robust estimators under C_2 and C_4. This is also reflected in Figures 1 to 5. It is worth noticing that the classical estimators seem stable with respect to contaminating only the carriers with high leverage points (C_3). This is natural when using the weighted simplified estimator since responses with large residuals were not included in this contamination, but it could seem unnatural when using the average or weighted imputed estimators, since it could be expected that large values of the covariates $\mathbf{x}$ would lead the classical estimate to explode. However, the good performance observed is mainly due to the fact that the least squares regression parameter estimates β almost as 0, in all the missingness schemes. This behavior is illustrated by Figure 6 which shows than under C_3 the median of $\widehat{\boldsymbol{\beta}}_{\text{LS}}$ is close to 0 while $\widehat{\boldsymbol{\beta}}_{\text{R}}$ is close to 2 as it should be. The same behavior is observed also under C_4, where the classical estimates of $\boldsymbol{\beta}$ exhibit a large bias which decrease the influence of the leverage points when using the average or the weighted imputed estimators. In both cases, the classical regression parameter estimators are useless to estimate the regression parameter and in this sense, the least squares procedures seem not reliable to estimate the marginal parameter.

It is worth noticing that, when there is no contamination, except for the complete data estimators, the weighted imputed estimators (linear or robust) perform better than the two other competitors leading to smaller mean square errors. Their advantage over the weighted simplified is specially reflected when the missing probability depends on x, t or in both variables. In this situation, see Tables 4 to 6, the weighted simplified estimators have almost twice mean square errors than the weighted imputed. The worst situation for the weighted simplified procedure is when $p(x,t)$ only depends on t. This fact can be explained since it gives the larger proportion of missing data in each sample, near 70%, while in the two other situations the proportion of missing data is about 65%. The same conclusion holds under C_1. However, under C_2 to C_4 a different behavior is observed for the classical and robust estimators. As expected, the weighted imputed $M-$estimators perform much better than the weighted simplified method leading to almost the same ratios between the mean square error of the weighted simplified $M-$estimator and the weighted imputed $M-$estimator as in the non–contaminated situation. On the contrary, when using the linear estimators, the mean square errors of the three estimators are almost the same due to the bias of all procedures. Note also that, under C_2 and C_4, when using the robust estimators the smallest mean square errors are attained by the averaged $M-$estimators, even if they are almost of the same order than the weighted imputed $M-$estimator.

7 Final comments

We have introduced three robust procedures to estimate the marginal location parameter under a partially linear model when there are missing observations in the response variable and it can be suspected that anomalous observations are present in the sample. All procedures are Fisher–consistent and thus they lead to strongly consistent estimators.

Under the contaminations considered, they show their advantage over the classical estimators. Moreover, the average and weighted imputed $M-$estimators, even if they are computationally more expensive, should be used since they perform better than the weighted simplified $M-$estimator in all situations. Both the classical and robust procedures do not seem to be very sensitive to the choice of the smoothing parameter and so an exhaustive bandwidth search can be avoided.

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 PID 5505 from CONICET, PAV 120 and PICT 21407 from ANPCYT, X-094 from the Universidad de Buenos Aires at Buenos Aires, Argentina and also by the DGICYT Spanish Grant MTM2005-00820 (European FEDER support included).

8 Appendix

In Section 8.1 we give the proofs of the Theorems stated in Section 4, while in Sections 8.2 and 8.3, we will study the Fisher–consistency of the given proposals.

8.1 Proofs

PROOF OF THEOREM 4.1. a) It is enough to show that for any borelian set B, $\widehat{\phi}(B) \xrightarrow{a.s.} P(Y \in B)$ where

$$\widehat{\phi}(B) = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{p_n(\mathbf{x}_i, t_i)} I_B(y_i) \; .$$

Note that $\widehat{\phi}(B) = S_{1n} + S_{2n}$ where

$$\begin{aligned} S_{1n} &= \frac{1}{n}\sum_{i=1}^{n} \left[\frac{1}{p_n(\mathbf{x}_i, t_i)} - \frac{1}{p(\mathbf{x}_i, t_i)}\right] \delta_i \; I_B(y_i) \\ S_{2n} &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{p(\mathbf{x}_i, t_i)} \delta_i \; I_B(y_i) \; . \end{aligned}$$

Using **A2** and **A3**, we have that $|S_{1n}| \xrightarrow{a.s.} 0$. On the other hand, using the strong law of large numbers and the MAR assumption, we have that $S_{2n} \xrightarrow{a.s.} P(Y \in B)$, concluding the proof.

The proof of b) follows easily using a), the following bound

$$\sup_{\substack{a\in\mathbb{R}\\ \sigma\in[\sigma_0/2,2\sigma_0]}} |\widehat{\lambda}(a,\sigma)-\lambda(a,\sigma)| \leq \int |\psi_1'(u)|du \, \|\widehat{F}_n - F\|_\infty \, ,$$

the continuity of $\lambda(a,\sigma)$ as a function of σ and the fact that in a neighborhood of θ, the function $\lambda(a,\sigma_0)$ has a unique change of sign. □

PROOF OF THEOREM 4.2. a) It is enough to show that $\Pi(\widetilde{P}_n, \widetilde{\widetilde{P}}_n)| \xrightarrow{a.s.} 0$ with

$$\widetilde{\widetilde{P}}_n(A) = \frac{1}{n}\sum_{i=1}^n I_A(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + g_0(t_i)) \, .$$

This result follows if we show that for any bounded and continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$ we have that

$$\left| E_{\widetilde{\widetilde{P}}_n}(f) - E_{\widetilde{P}_n}(f) \right| \xrightarrow{a.s.} 0$$

which follows using analogous arguments to those considered in Lemma 1 of Bianco and Boente (2004).

The proof of b) is derived as in Theorem 4.1 using the following bound

$$\sup_{\substack{a\in\mathbb{R}\\ \sigma\in[\sigma_0/2,2\sigma_0]}} |\widetilde{\widetilde{\lambda}}_n(a,\sigma) - \lambda_Z(a,\sigma)| \leq 2\|\psi_1'\|_\infty \Pi(\widetilde{P}_n, P_Z)$$

with $\lambda_Z(a,\sigma) = E\psi_1\left((Z-a)/\sigma\right)$. □

PROOF OF THEOREM 4.3. a) It is enough to show that for any borelian set B, $\widehat{\widehat{\phi}}(B) \xrightarrow{a.s.} P(Y \in B)$ where

$$\widehat{\widehat{\phi}}(B) = \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{p_n(\mathbf{x}_i,t_i)} I_B(y_i) + \frac{1}{n}\sum_{i=1}^n \left[\left(1 - \frac{\delta_i}{p_n(\mathbf{x}_i,t_i)}\right) I_B(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i))\right] \, .$$

Note that $\widehat{\widehat{\phi}}(B) = \widehat{\phi}(B) + S_{1n} + S_{2n}$ where $\widehat{\phi}(B)$ is defined in the proof of Theorem 4.1 and

$$\begin{aligned} S_{1n} &= \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{p_n(\mathbf{x}_i,t_i)} - \frac{1}{p(\mathbf{x}_i,t_i)}\right] \delta_i \, I_B(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)) \\ S_{2n} &= \frac{1}{n}\sum_{i=1}^n \left[1 - \frac{\delta_i}{p(\mathbf{x}_i,t_i)}\right] \, I_B(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} + \widehat{g}_n(t_i)) \, . \end{aligned}$$

In the proof of Theorem 4.1 it was shown that $\widehat{\phi}(B) \xrightarrow{a.s.} P(Y \in B)$. Using **A2**, we have that

$$|S_{1n}| \leq \frac{1}{A \inf\limits_{(\mathbf{x},t)} p_n(\mathbf{x},t)} \sup_{(\mathbf{x},t)} |p_n(\mathbf{x},t) - p(\mathbf{x},t)|$$

which together with **A3** entail that $S_{1n} \xrightarrow{a.s.} 0$. Besides,

$$|S_{2n}| \leq \frac{1}{n}\sum_{i=1}^n \left[1 - \frac{\delta_i}{p(\mathbf{x}_i,t_i)}\right]$$

and so, using the strong law of large numbers, we have that $S_{2n} \xrightarrow{a.s.} 0$, concluding the proof.

The proof of b) follows as in Theorem 4.1. □

8.2 Fisher–consistency of the parametric and nonparametric components

We first consider the functionals defined in Section 2.1.

Note that

$$E\left[\delta\psi\left(\frac{Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(t)}{\sigma}\right)|T=t\right]=E\left[p(\mathbf{X},t)\psi\left(\frac{Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(t)}{\sigma}\right)|T=t\right]\ .$$

Thus, it is easy to see that the independence between ϵ and $(\mathbf{X},T)$ entails that $g_{\boldsymbol{\beta}_0}(t)\equiv g_0(t)$. On the other hand, we get easily that

$$\begin{aligned}E\left[\delta\rho\left(\frac{Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(T)}{\sigma}\right)\upsilon\left(\mathbf{X}\right)\right] &= E\left[p(\mathbf{X},T)\rho\left(\frac{Y-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(T)}{\sigma}\right)\upsilon\left(\mathbf{X}\right)\right]\\ &= E\left[p(\mathbf{X},T)\rho\left(\frac{\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}_0+g_0\left(T\right)+\epsilon-\mathbf{X}^{\mathrm{T}}\boldsymbol{\beta}-g_{\boldsymbol{\beta}}(T)}{\sigma}\right)\upsilon\left(\mathbf{X}\right)\right]\\ &= E\left[p(\mathbf{X},T)\upsilon\left(\mathbf{X}\right)E\left\{\rho\left(\frac{\epsilon+\mathbf{X}^{\mathrm{T}}\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)+g_0(T)-g_{\boldsymbol{\beta}}(T)}{\sigma}\right)\Big|(\mathbf{X},T)\right\}\right]\end{aligned}$$

Using that ρ is a *rho*–function as defined in Maronna, Martin and Yohai (2006), from the location case, the symmetry of the errors distribution and the independence mentioned above, we get

$$E\left\{\rho\left(\frac{\epsilon+\mathbf{x}^{\mathrm{T}}\left(\boldsymbol{\beta}_0-\boldsymbol{\beta}\right)+g_0(t)-g_{\boldsymbol{\beta}}(t)}{\sigma}\right)\Big|(\mathbf{X},T)=(\mathbf{x},t)\right\}\geq E\left\{\rho\left(\frac{\epsilon}{\sigma}\right)\Big|(\mathbf{X},T)=(\mathbf{x},t)\right\}=E\left\{\rho\left(\frac{\epsilon}{\sigma}\right)\right\}$$

and so $\boldsymbol{\beta}(F)=\boldsymbol{\beta}_0$, which concludes the proof.

8.3 Fisher–consistency of the marginal location functionals

Throughout this section we assume that the errors distribution is symmetric and $\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0+g_0\left(t_i\right)=\theta+u_i$ where u_i has a symmetric distribution too. The functionals related to the proposed estimators are given by

- **Weighted Simplified functional** This functional is the solution, $\theta_{\mathrm{S}}(F)$, of

$$E\frac{\delta}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\mathrm{S}}(F)}{\sigma}\right)=0\ .$$

  Note that, by taking conditional expectation and using that we have a MAR missingness scheme, we have

$$\begin{aligned}E\frac{\delta}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\mathrm{S}}(F)}{\sigma}\right) &= E\frac{p(\mathbf{X},T)}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\mathrm{S}}(F)}{\sigma}\right)\\ &= E\psi_1\left(\frac{Y-\theta_{\mathrm{S}}(F)}{\sigma}\right)\ ,\end{aligned}$$

  and so $\theta_{\mathrm{S}}(F)=\theta$ if $u+\epsilon$ has a symmetric distribution.

- **Averaged $M-$functional** The functional $\theta_{\rm MA}(F)$ is the solution of

$$E\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}(F)+g_{\boldsymbol{\beta}(F)}(T)-\theta_{\rm MA}(F)}{\sigma}\right)=0\ .$$

Using that ϵ has a symmetric distribution, from Section 8.2 we get $\boldsymbol{\beta}(F)=\boldsymbol{\beta}_0$ and $g_{\boldsymbol{\beta}(F)}=g_0$. Thus,

$$\begin{aligned} E\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}(F)+g_{\boldsymbol{\beta}(F)}(T)-\theta_{\rm MA}(F)}{\sigma}\right) &= E\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}_0+g_0(T)-\theta_{\rm MA}(F)}{\sigma}\right) \\ &= E\psi_1\left(\frac{u+\theta-\theta_{\rm MA}(F)}{\sigma}\right)\ . \end{aligned}$$

Since we have assumed that u has a symmetric distribution, we obtain that $\theta_{\rm MA}=\theta$.

- **Weighted Imputed functional**. The functional $\theta_{\rm W}(F)$ solves

$$E\left[\frac{\delta}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\rm W}(F)}{\sigma}\right)+\left(1-\frac{\delta}{p(\mathbf{X},T)}\right)\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}(F)+g_{\boldsymbol{\beta}(F)}(T)-\theta_{\rm W}(F)}{\sigma}\right)\right]=0\ .$$

As above, we get

$$\begin{aligned} & E\left[\frac{\delta}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\rm W}(F)}{\sigma}\right)+\left(1-\frac{\delta}{p(\mathbf{X},T)}\right)\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}(F)+g_{\boldsymbol{\beta}(F)}(T)-\theta_{\rm W}(F)}{\sigma}\right)\right] \\ = \ & E\left[\frac{p(\mathbf{X},T)}{p(\mathbf{X},T)}\psi_1\left(\frac{Y-\theta_{\rm W}(F)}{\sigma}\right)+\left(1-\frac{p(\mathbf{X},T)}{p(\mathbf{X},T)}\right)\psi_1\left(\frac{\mathbf{X}^{\rm T}\boldsymbol{\beta}(F)+g_{\boldsymbol{\beta}(F)}(T)-\theta_{\rm W}(F)}{\sigma}\right)\right] \\ = \ & E\psi_1\left(\frac{Y-\theta_{\rm W}(F)}{\sigma}\right) \end{aligned}$$

and so $\theta_{\rm W}(F)=\theta$.

References

Aerts, M.; Claeskens, G.; Hens, N. and Molenberghs, G., 2002. Local multiple imputation. *Biometrika* **89**, 2, 375–388.

Bianco, A. and Boente, G., 2004. Robust estimators in semiparametric partly linear regression models. *J. Statist. Planning Inf.*, **122**, 229-252.

Bianco, A. and Boente, G., 2007. Robust estimators under a semiparametric partly linear autoregression model: Asymptotic behavior and bandwidth selection. *J. Time Ser. Anal.*, **28**, 274-306.

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

Boente, G., Fraiman, R. and Meloche, J., 1997. Robust plug–in bandwidth estimators in nonparametric regression. *J. Statist. Planning and Inference*, **57**, 109-142.

Boente, G., González–Manteiga, W. and Pérez–González, A. (2008). Robust nonparametric estimation with missing data. To appear in *J. Statist. Plann. Inference*. Available at `http://www.ic.fcen.uba.ar/preprints/boegonper.pdf`

Boente, G. and Rodriguez, D., 2007. Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. *Comp. Statist. Data Anal.*, **52**, 2808-2828.

Cantoni, E. and Ronchetti, E., 2001. Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141-146.

Chen, J. , Fan, J., Li, K. and Zhou, H., 2006. Local quasi–likelihood estimation with data missing at random. *Statistica Sinica* **16**, 1071-1100.

Chen, J. and Shao, J., 2000. Nearest neighbor imputation for survey data. *J. Official Statist.* **16**, 113-131.

Cheng, P. E., 1994. Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* **89**, 81-87.

Cheng, P. E. and Chu, C.K., 1996. Kernel estimation of distribution functions and quantiles with missing data. *Statist. Sinica* **6**, 63-78.

González–Manteiga, W. and Pérez–González, A., 2004. Nonparametric mean estimation with missing data. *Comm. Statist. Theory Methods* **33**, 277-303.

Huber, P., 1981. *Robust statistics*, Wiley, New York.

Leung, D. H. Y., Marriott, F. H. C. and Wu, E. K. H., 1993. Bandwidth selection in robust smoothing. *J. Nonpar. Statist*, **2**, 333-339.

Maronna, R., Martin, D. and Yohai, V., 2006. *Robust Statistics: Theory and Methods* , Wiley, New York.

Meng, X.-L., 2000. Missing data: Dial M for ???. *J. Amer. Statist. Assoc.* **95**, 452, 1325-1330.

Neyman, J., 1938. Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.* **33** 101-116.

Scharfstein, D.O., Rotnitzky, A., and Robins, J., 1999. Adjusting for nonignroable drop out in semiparametric nonresponse models (with discussion). *J. Amer. Statist. Assoc.*, **94**, 1096-1146.

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.

Wang, F. and Scott, D., 1994. The L_1 method for robust nonparametric regression. *J. Amer. Stat. Assoc.*, **89**, 65-76.

Wang, Q.; Linton, O. and Härdle, W., 2004. Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.*, **99**, 466, 334–345.

Wang, Q. and Sun, Z., 2007. Estimation in partially linear models with missing responses at random. *J. Multiv. Anal.*, **98**, 1470-1493.

Wang, W. and Rao, J. N. K., 2002. Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30**, 896-924.

Yates, F., 1933. The analysis of replicated experiments when the field results are incomplete. *Emporium J. Exp. Agriculture* **1**, 129-142.

$p(\mathbf{x},t)$	0.8		$\pi(t)$		$\pi(x)$		$\pi(xt)$		
	AE_{LS}	WIE_{LS}	AE_{LS}	WIE_{LS}	AE_{LS}	WIE_{LS}	AE_{LS}	WIE_{LS}	
Minimum	5.6703	5.6702	6.7913	6.7911	6.0852	6.0846	5.7455	5.7442	
Mean	5.6713	5.6709	6.7972	6.8010	6.0880	6.0863	5.7496	5.7491	C_0
Maximum	5.6717	5.6713	6.7992	6.8056	6.0891	6.0871	5.7516	5.7512	
Minimum	34.7509	34.7516	40.9682	40.9641	35.0603	35.1356	36.3458	36.3825	
Mean	34.7542	34.7540	41.0222	41.0219	35.0668	35.1434	36.3476	36.3922	C_2
Maximum	34.7554	34.7550	41.5475	41.5549	35.0831	35.1623	36.3497	36.4062	

Table 1: Minimum, mean value and maximum of least squares cross–validation error, under C_0 and C_2.

	Estimator						
	WSE_{LS}	AE_{LS}	WIE_{LS}	WSE_{R}	AE_{R}	WIE_{R}	
bias	-0.0072	-0.0077	-0.0072	-0.0135	-0.0137	-0.0132	
sd	0.2486	0.2515	0.2486	0.2612	0.2653	0.2608	C_0
MSE	0.0618	0.0633	0.0618	0.0684	0.0706	0.0682	
bias	-0.0081	-0.0091	-0.0081	-0.0171	-0.0161	-0.0166	
sd	0.2636	0.2666	0.2636	0.2767	0.2702	0.2762	C_1
MSE	0.0696	0.0712	0.0696	0.0769	0.0733	0.0766	
bias	1.9922	2.0318	1.9922	-0.01287	-0.0078	-0.0126	
sd	0.2517	0.2574	0.2517	0.2888	0.2703	0.2884	C_2
MSE	4.0322	4.1946	4.0322	0.0836	0.0731	0.0833	
bias	0.0017	0.0010	0.0017	-0.0058	-0.0119	-0.0056	
sd	0.2601	0.2621	0.2601	0.2730	0.2831	0.2724	C_3
MSE	0.0676	0.0687	0.0676	0.0746	0.0803	0.0742	
bias	-1.0070	-1.0502	-1.0070	-0.0145	-0.0081	-0.0141	
sd	0.2563	0.2551	0.2563	0.2803	0.2763	0.2800	C_4
MSE	1.0796	1.1681	1.0796	0.0788	0.0764	0.0786	

Table 2: Biases, standard deviations and mean square errors of the classical and robust procedures, under C_0 to C_4, when $h = 0.2$ and $p(\mathbf{x},t) = 1$.

	Estimator						
	WSE_{LS}	AE_{LS}	WIE_{LS}	WSE_{R}	AE_{R}	WIE_{R}	
bias	-0.0032	-0.0062	-0.0054	-0.0103	-0.0126	-0.0124	
sd	0.2786	0.2511	0.2481	0.2957	0.2646	0.2605	C_0
MSE	0.0776	0.0631	0.0616	0.0875	0.0701	0.0680	
bias	-0.0036	-0.0071	-0.0059	-0.0136	-0.0166	-0.0161	
sd	0.2981	0.2704	0.2673	0.3152	0.2702	0.2793	C_1
MSE	0.0889	0.0731	0.0715	0.0996	0.0733	0.0783	
bias	1.9818	2.0232	1.9846	-0.0102	-0.00756	-0.0145	
sd	0.4017	0.4016	0.3954	0.3280	0.2712	0.2933	C_2
MSE	4.0889	4.2547	4.0952	0.1077	0.0736	0.0862	
bias	0.0066	0.0101	0.0101	-0.0021	-0.0118	-0.0019	
sd	0.2945	0.2840	0.2819	0.3110	0.2844	0.2781	C_3
MSE	0.0868	0.0807	0.0796	0.0967	0.0811	0.0773	
bias	-0.9914	-1.0299	-0.9888	-0.0112	-0.0079	-0.0121	
sd	0.3764	0.3776	0.3716	0.3200	0.2773	0.2891	C_4
MSE	1.1246	1.2033	1.1158	0.1025	0.0769	0.0837	

Table 3: Biases, standard deviations and mean square errors of the classical and robust procedures, under C_0 to C_4, when $h = 0.2$ and $p(\mathbf{x}, t) = 0.80$.

	Estimator						
	WSE_{LS}	AE_{LS}	WIE_{LS}	WSE_{R}	AE_{R}	WIE_{R}	
bias	-0.0016	-0.0227	-0.0102	-0.0104	-0.0313	-0.0183	
sd	0.3750	0.2618	0.2579	0.4009	0.2748	0.2703	C_0
MSE	0.1406	0.0691	0.0666	0.1608	0.0765	0.0734	
bias	-0.0054	-0.0285	-0.0159	-0.0145	-0.0369	-0.0250	
sd	0.4018	0.2933	0.2899	0.4269	0.2850	0.3006	C_1
MSE	0.1614	0.0868	0.0843	0.1821	0.0826	0.0910	
bias	1.982	1.9777	1.9753	-0.0096	-0.0288	-0.0260	
sd	0.8037	0.8111	0.8079	0.4435	0.2889	0.3223	C_2
MSE	4.5747	4.5691	4.5547	0.1968	0.0843	0.1046	
bias	0.0046	0.0022	0.0271	-0.0035	-0.0325	-0.0092	
sd	0.4008	0.4080	0.4041	0.4254	0.3018	0.3130	C_3
MSE	0.1607	0.1665	0.1640	0.1810	0.0921	0.0981	
bias	-0.9985	-0.9497	-0.9112	-0.0162	-0.0291	-0.0258	
sd	0.6778	0.8168	0.8146	0.4340	0.2952	0.3247	C_4
MSE	1.4564	1.5692	1.4937	0.1887	0.0880	0.1061	

Table 4: Biases, standard deviations and mean square errors of the classical and robust procedures, under C_0 to C_4, when $h = 0.2$ and $p(\mathbf{x}, t) = 0.4 + 0.5\cos^2(2t + 0.4)$.

	Estimator						
	WSE_{LS}	AE_{LS}	WIE_{LS}	WSE_{R}	AE_{R}	WIE_{R}	
bias	-0.0040	-0.0090	-0.0068	-0.0139	-0.0155	-0.0137	
sd	0.3336	0.2547	0.2539	0.3551	0.2684	0.2668	C_0
MSE	0.1113	0.0650	0.0645	0.1263	0.0723	0.0713	
bias	-0.0042	-0.0095	-0.0085	-0.0173	-0.0202	-0.0170	
sd	0.3545	0.2784	0.2789	0.3754	0.2769	0.2904	C_1
MSE	0.1257	0.0776	0.0778	0.1412	0.0771	0.0846	
bias	1.9946	2.0265	1.9953	-0.0093	-0.0114	-0.0125	
sd	0.6059	0.5389	0.5776	0.3835	0.2784	0.3055	C_2
MSE	4.3456	4.3970	4.3148	0.1471	0.0777	0.0935	
bias	0.0082	0.0011	0.0088	-0.0023	-0.0152	-0.0021	
sd	0.3423	0.3073	0.3096	0.3623	0.2921	0.2880	C_3
MSE	0.1172	0.0944	0.0959	0.1313	0.0856	0.0829	
bias	-0.9857	-1.0313	-0.9883	-0.0112	-0.0115	-0.0132	
sd	0.5026	0.4605	0.4896	0.3731	0.2846	0.2998	C_4
MSE	1.2241	1.2757	1.2165	0.1393	0.0811	0.0900	

Table 5: Biases, standard deviations and mean square errors of the classical and robust procedures, under C_0 to C_4, when $h = 0.2$ and $p(\mathbf{x}, t) = 0.4 + 0.5\cos^2(2x + 0.4)$.

	Estimator						
	WSE_{LS}	AE_{LS}	WIE_{LS}	WSE_{R}	AE_{R}	WIE_{R}	
bias	-0.0124	-0.0203	-0.0086	-0.0308	-0.0283	-0.0158	
sd	0.3238	0.2535	0.2511	0.3481	0.2666	0.2637	C_0
MSE	0.1050	0.0647	0.0631	0.1221	0.0719	0.0698	
bias	-0.0131	-0.0229	-0.0098	-0.0347	-0.0323	-0.0189	
sd	0.3421	0.2725	0.2719	0.3647	0.2732	0.2834	C_1
MSE	0.1172	0.0748	0.0739	0.1342	0.0757	0.0807	
bias	1.9982	1.9659	1.9995	-0.0262	-0.0235	-0.0202	
sd	0.5411	0.4862	0.5206	0.3757	0.2744	0.2983	C_2
MSE	4.2855	4.1010	4.2690	0.1419	0.0759	0.0894	
bias	-0.0036	-0.1115	0.0021	-0.0215	-0.0269	-0.0062	
sd	0.3364	0.2917	0.3122	0.3583	0.2871	0.2857	C_3
MSE	0.1132	0.0975	0.0975	0.1288	0.0831	0.0817	
bias	-1.0132	-1.1413	-1.0053	-0.0309	-0.0233	-0.0179	
sd	0.4784	0.4513	0.4862	0.3697	0.2799	0.2946	C_4
MSE	1.2555	1.5063	1.2470	0.1376	0.0789	0.0871	

Table 6: Biases, standard deviations and mean square errors of the classical and robust procedures, under C_0 to C_4, when $h = 0.2$ and $p(\mathbf{x}, t) = 0.4 + 0.5\cos^2(2xt + 0.4)$.

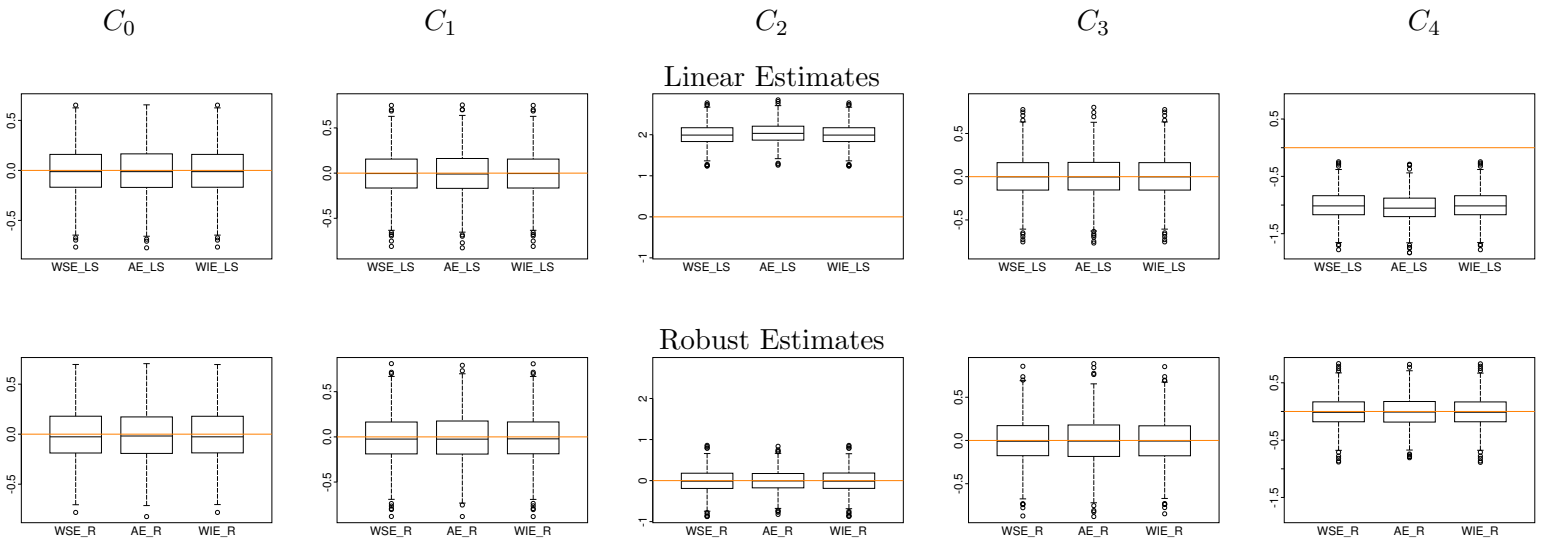

Figure 1: Boxplots for the estimates of the marginal location parameter for the classical and robust proposals with bandwidth $h = 0.2$ when $p(\mathbf{x}, t) \equiv 1$.

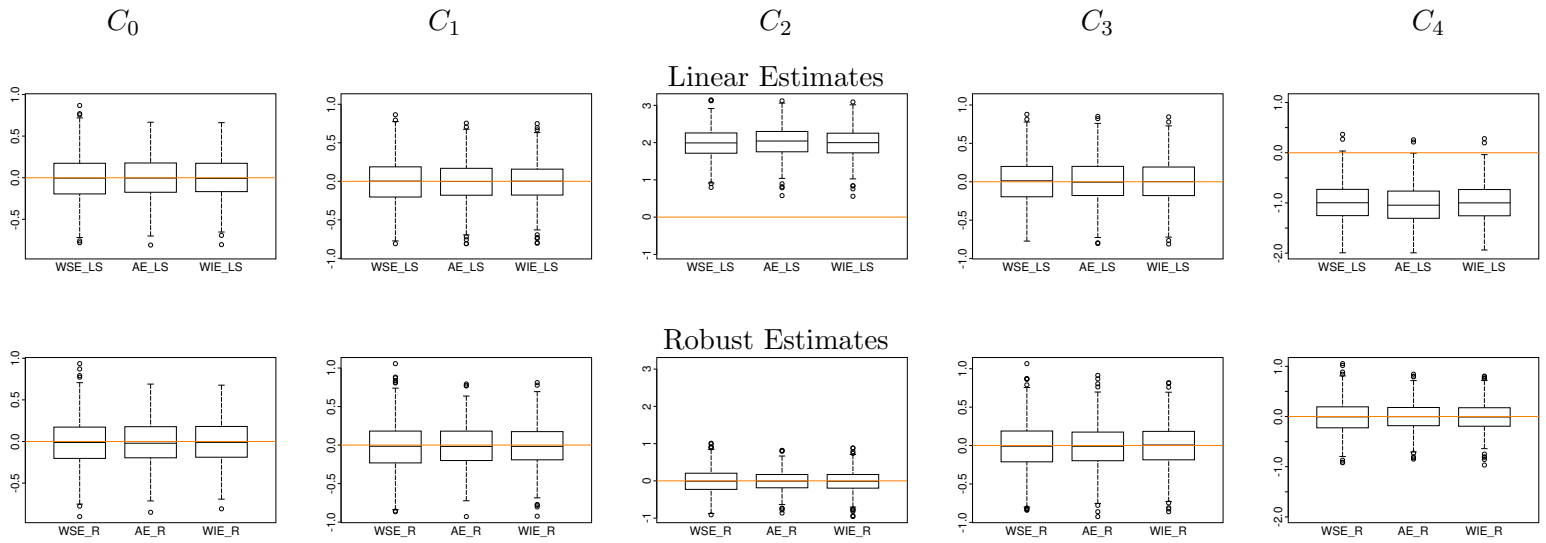

Figure 2: Boxplots for the estimates of the marginal location parameter for the classical and robust proposals with bandwidth $h = 0.2$ when $p(\mathbf{x}, t) \equiv 0.8$.

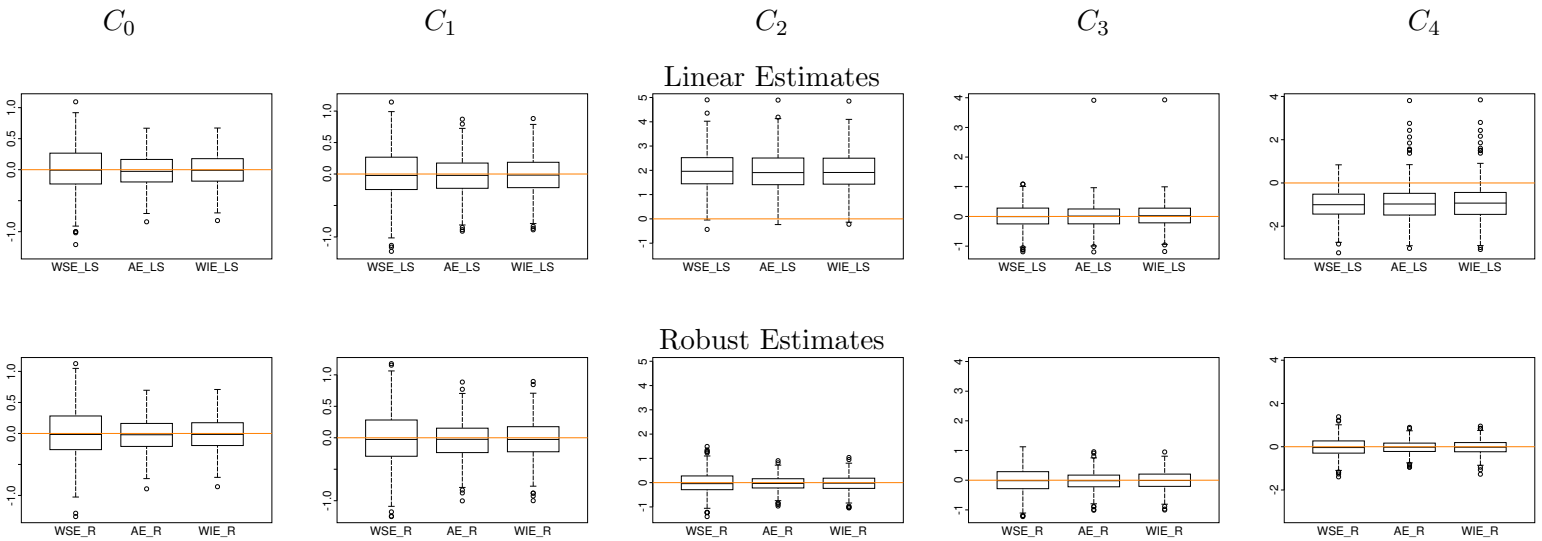

Figure 3: Boxplots for the estimates of the marginal location parameter for the classical and robust proposals with bandwidth $h = 0.2$ when $p(\mathbf{x}, t) \equiv 0.4 + 0.5\cos^2(2t + 0.4)$.

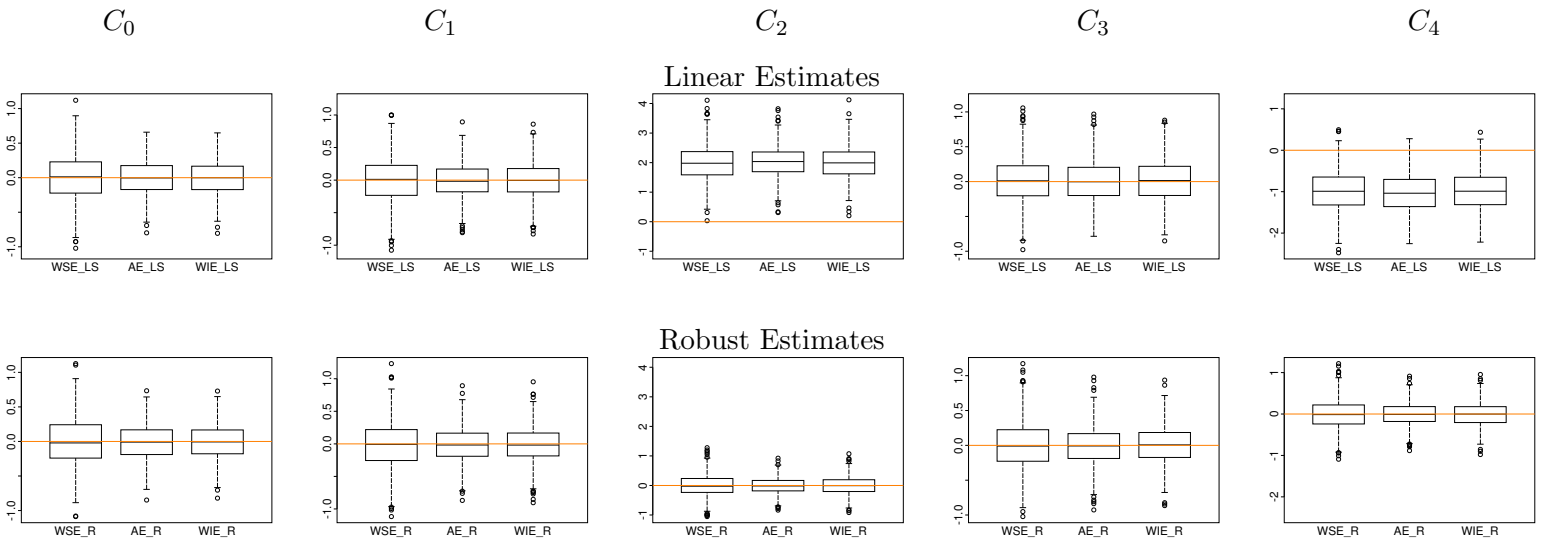

Figure 4: Boxplots for the estimates of the marginal location parameter for the classical and robust proposals with bandwidth $h = 0.2$ when $p(\mathbf{x}, t) \equiv 0.4 + 0.5\cos^2(2x + 0.4)$.

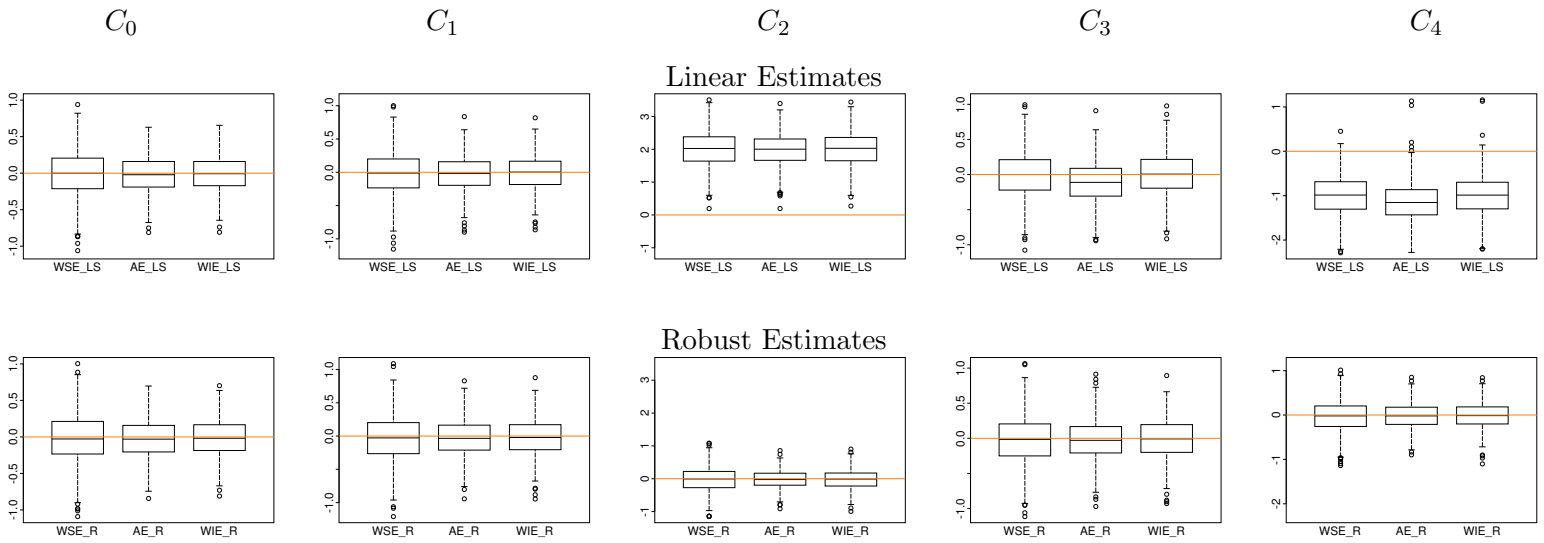

Figure 5: Boxplots for the estimates of the marginal location parameter for the classical and robust proposals with bandwidth $h = 0.2$ when $p(\mathbf{x}, t) = 0.4 + 0.5\cos^2(2x\,t + 0.4)$.

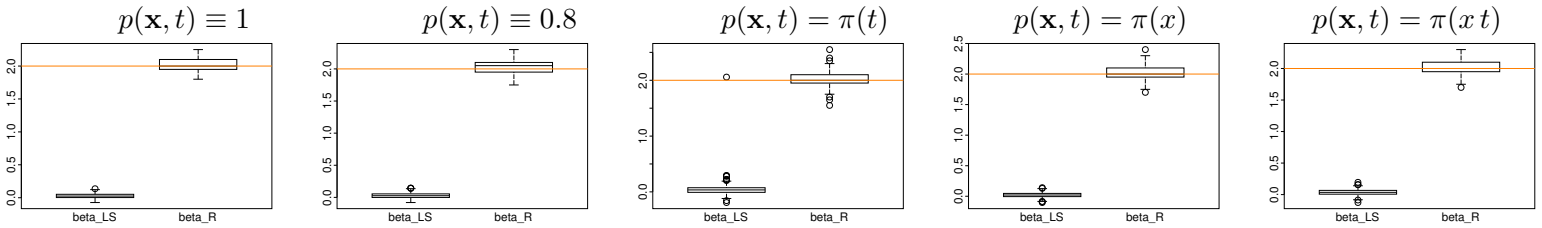

Figure 6: Boxplots for the estimates of the regression parameter $\boldsymbol{\beta}$ with bandwidth $h = 0.2$.