Robust location estimation with missing data

Mariela SUED[1]* and Victor J. YOHAI[2]

[1] *Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires and CONICET, Buenos Aires, Argentina*
[2] *Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires and CONICET, Buenos Aires, Argentina*

Key words and phrases: Asymptotic distribution; breakdown point; L-location functional; median; missing at random; M-location functional.

MSC 2010: Primary 62F35; secondary 62F12.

Abstract: In a missing data setting, we have a sample in which a vector of explanatory variables $\mathbf{x}_i$ is observed for every subject i, while scalar responses y_i are missing by happenstance on some individuals. In this work we propose robust estimators of the distribution of the responses assuming missing at random (MAR) data, under a semiparametric regression model. Our approach allows the consistent estimation of any weakly continuous functional of the response's distribution. In particular, strongly consistent estimators of any continuous location functional, such as the median, L-functionals and M-functionals, are proposed. A robust fit for the regression model combined with the robust properties of the location functional gives rise to a robust recipe for estimating the location parameter. Robustness is quantified through the breakdown point of the proposed procedure. The asymptotic distribution of the location estimators is also derived. The proofs of the theorems are presented in Supplementary Material available online. *The Canadian Journal of Statistics* 41: 111–132; 2013 © 2012 Statistical Society of Canada

Résumé: Avec les données manquantes, nous avons un échantillon pour lequel les variables explicatives x_i sont observées pour chaque sujet i, tandis que les variables réponses y_i sont manquantes au hasard pour quelques individus. Dans ce travail, nous proposons des estimateurs robustes pour la fonction de distribution des variables réponses en supposant que les données soient manquantes au hasard (MAR), sous un modèle de régression non paramétrique. Notre approche permet l'estimation cohérente de n'importe quelle fonctionnelle faiblement continue de la distribution des variables réponses. Plus particulièrement, nous proposons des L- et M-fonctionnelles qui sont des estimateurs fortement cohérents de n'importe quelle fonctionnelle continue du paramètre de position (par exemple, la médiane). Une méthode d'ajustement robuste du modèle de régression combinée aux propriétés de robustesse des fonctionnelles de tendance centrale fournissent une méthode robuste pour l'estimation du paramètre de position. La robustesse de notre procédure est mesurée à l'aide du point de rupture. Nous obtenons aussi la fonction de distribution asymptotique des estimateurs du paramètre de position. Des suppléments, contenant les démonstrations des théorèmes, sont disponibles en ligne. *La revue canadienne de statistique* 41: 111–132; 2013 © 2012 Société statistique du Canada

1. INTRODUCTION

Suppose we have a sample of a population, such that for every subject i in the sample we observe a vector of explanatory variables $\mathbf{x}_i$ while a scalar response y_i is missing by happenstance on some individuals. A classical problem is to construct consistent estimators for the mean value of the response based on the observed data. In order to identify the parameter of interest in terms of the distribution of observed data, missing at random (MAR) is assumed.

Additional supporting information may be found in the online version of this article.
* *Author to whom correspondence may be addressed.*
E-mail: msued@dm.uba.ar

This hypothesis establishes that the value of the response does not provide additional information, on top of that given by the explanatory variables, to predict whether an individual will present a missing response (see Rubin, 1976). To be more rigorous, let us introduce a binary variable a_i such that $a_i = 1$ whenever the response is observed for subject i. In this way, MAR states that

$$\mathrm{P}(a_i = 1|\mathbf{x}_i, y_i) = \mathrm{P}(a_i = 1|\mathbf{x}_i). \tag{1}$$

This condition also implies that the conditional distribution of the responses given the vector of explanatory variables remains the same, regardless of the fact that the response is also observed: $y_i|\mathbf{x}_i \sim y_i|\mathbf{x}_i, a_i = 1$. Then $\mathrm{E}[y_i|\mathbf{x}_i] = \mathrm{E}[y_i|\mathbf{x}_i, a_i = 1]$. Since $\mathrm{E}[y_i] = \mathrm{E}[\mathrm{E}[y_i|\mathbf{x}_i]]$, a possible approach to estimate $\mathrm{E}[y_i]$ is based on a regression model (parametric or nonparametric) for $\mathrm{E}[y_i|\mathbf{x}_i] = g(\mathbf{x}_i)$, which is fitted using only the individuals for whom the response is observed. Then an estimator for $\mathrm{E}[y_i]$ is obtained by averaging $\widehat{g}(\mathbf{x}_i)$ over the whole sample, where $\widehat{g}$ is an estimator of g. A recent survey and discussion of this and other methods for dealing with this problem can be found in Kan & Schafer (2007) and Robins et al. (2007).

The estimation of the mean response under a missing at random assumption finds one of its most frequent applications in observational studies with medical or economic data. In particular, in the context of causal inference, to quantify the effect of two different treatments, say t_0 and t_1, on some response of interest, two random variables $y^{(0)}$ and $y^{(1)}$ are introduced. These variables represent responses in hypothetical worlds were all individuals are treated with t_0 and t_1, respectively. The average treatment effect is defined by $E[y^{(1)}] - E[y^{(0)}]$. Note that $y^{(j)}$, for $j = 0, 1$, is only observed in those individuals whose treatment level is $T = t_j$, and so it is considered a missing response for those individuals with treatment different from t_j. Since in observational studies the treatment assignment is in general not randomized, the estimation of $E[y^{(j)}]$ should be addressed using missing data techniques. This approach has been widely studied in the causal literature and examples of this methodology can be found in Dehejia & Wahba (1999) or in Bang & Robins (2005).

As is well known, the mean is not a robust location parameter, that is, a small change in the population distribution may have a large effect on this parameter. As a consequence of this, the mean does not admit consistent non-parametric robust estimators, except when strong properties on the distribution are assumed, as for example symmetry. For this reason, to introduce robustness in the present setting, we start by reformulating the statistical object of interest: instead of estimating the mean value of the response, we look for consistent estimators of $T_L(F_0)$, where T_L is a robust location functional and F_0 is the distribution of y_i. For example, if we are interested in estimating the median of F_0, we take $T_L(F_0) = \mathrm{med}(F_0)$. According to this, we say that a sequence of estimators $\widehat{\mu}_n$ is consistent for $T_L(F_0)$ if $\lim_{n\to\infty} \widehat{\mu}_n = T_L(F_0)$. Note that the naive estimator $T_L(F_n)$, where F_n is the empirical distribution of the non missing responses, in general is not consistent. In fact, $T_L(F_n) \to T_L(F_0^*)$ where F_0^* is the distribution of y_i conditionally on $a_i = 1$.

Bianco et al. (2010) obtain robust and consistent estimators of M-location functionals of the distribution of the responses. In their treatment they assumed a partially linear model to describe the relationship between y_i and $\mathbf{x}_i$, and also that the distributions of the response y_i and of the regression error under the true model are both symmetric.

In this paper we introduce a new estimator of any continuous location functional assuming that the relation between y_i and $\mathbf{x}_i$ is given by means of a semiparametric regression model. We show that, once the regression model is fitted using a robust estimator, we can define a consistent estimator of the distribution function of the response. Then, any location parameter of the response distribution defined throughout a weak continuous location functional may be also consistently estimated. This can be done by evaluating the functional at the estimated distribution function.

The consistency of this procedure does not require the symmetry assumptions used by Bianco et al. (2010).

A robust fit for the regression model combined with the robust properties of T_L gives rise to a robust recipe for estimating $T_L(F_0)$. Robustness is quantified by looking at the breakdown point of the proposed procedure. For this purpose, we introduce for the first time a definition of the breakdown point when there are missing data in the sample.

This work is organized as follows. In Section 2 we formalize the problem of the robust estimation of a location parameter with missing data. In Section 3 we present our proposal for estimating $T_L(F_0)$, when T_L is a weakly continuous location functional. In Section 4 we show that, under general conditions, the proposed estimators are strongly consistent and asymptotically normal. In Section 5 we study the breakdown point of the proposed estimators. In Sections 6 and 7 we introduce some possible robust regression and location functionals, respectively, and show that they satisfy the assumptions required for consistency and asymptotic normality of the proposed estimators. In Section 8 we discuss the results of a Monte Carlo study. These results confirm that the proposed estimators are highly robust under outlier contamination. In Section 9 we analyze an example with real data. The proofs of the theorems are presented in Supplementary Material available online.

2. DESCRIBING OUR SETTING: THE DATA, THE PROBLEM AND THE MODEL

We first introduce some notation. Henceforth $\mathrm{E}_G[h(\mathbf{z})]$ and $\mathrm{P}_G(A)$ will respectively denote the expectation of $h(\mathbf{z})$ and the probability that $\mathbf{z} \in A$, when $\mathbf{z}$ is distributed according to G. If $\mathbf{z}$ has distribution G we write $\mathbf{z} \sim G$ or $\mathcal{D}(\mathbf{z}) = G$. Weak convergence of distributions, convergence in probability and convergence in distribution of random variables or vectors are denoted by $G_n \to_w G$, $\mathbf{z}_n \to_p \mathbf{z}$ and $\mathbf{z}_n \to_d \mathbf{z}$, respectively. By an abuse of notation, we will write $\mathbf{z}_n \to_d G$ to denote $\mathcal{D}(\mathbf{z}_n) \to_w G$. We use $o_P(1)$ to denote a sequence that converges to zero in probability and $O_P(1)$ to denote a sequence bounded in probability. The complement and the indicator of the set A are denoted by A^c and $\mathbf{1}_A$, respectively. The scalar product of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^s$ is denoted by $\mathbf{a}'\mathbf{b}$ and $\mathbb{R}_{\geq 0}$ denotes the set of non-negative real numbers.

In this paper we use the expression *empirical distribution* of $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$, n points in $\mathbb{R}^k$, to denote the function $F_n : \mathbb{R}^k \to [0, 1]$ such that given $\mathbf{z} \in \mathbb{R}^k$, $F_n(\mathbf{z}) = m/n$, where m is the number of points $\mathbf{z}_i$ such that all its coordinates are smaller than or equal to the corresponding ones of $\mathbf{z}$.

Throughout this work, we have a random sample of n subjects and for each subject i in the sample, $1 \leq i \leq n$, a vector of explanatory variables $\mathbf{x}_i$ is always observed, while the response y_i is missing for some subjects. Let a_i be the indicator of whether y_i is observed for subject i: $a_i = 1$ if y_i is observed and $a_i = 0$ if it is not.

We will be concerned with the estimation of a location functional of the distribution of the response. A location functional T_L, defined on a class of univariate distribution functions $\mathcal{G}$, assigns to each $F \in \mathcal{G}$ a real number $T_L(F)$ satisfying $T_L(F_{ay+b}) = aT_L(F_y) + b$, where F_y denotes the distribution of the random variable y.

Examples of location functionals are the mean and median. The M-location functionals form an important class of robust location functionals that includes, among others, the median. Another important class of location functionals is that of L-functionals. Both M- and L-location functionals are studied in Section 7.

A functional T is said to be weakly continuous at F if given a sequence $\{F_n\}$ of distribution functions that converges weakly to F ($F_n \to_w F$), then $T(F_n) \to T(F)$. In order to obtain a consistent estimator of a location parameter defined by means of a weakly continuous functional, it is sufficient to have a sequence of estimators $\widehat{F}_n$ that converges weakly to the distribution of the y_i's.

To be more precise, denote by F_0 the distribution of the outcomes y_i. Let T_L be a weakly continuous location functional at F_0. We are interested in estimating $\mu_0 = T_L(F_0)$. We assume a semiparametric regression model

$$y_i = g(\mathbf{x}_i, \beta_0) + u_i, \quad 1 \leq i \leq n, \tag{2}$$

with $y_i, u_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^p$, u_i independent of $\mathbf{x}_i$, $\beta_0 \in B \subset \mathbb{R}^q$, $g : \mathbb{R}^p \times B \to \mathbb{R}$. Furthermore, in order to guarantee the MAR condition, we assume that u_i is independent of $(\mathbf{x}_i, a_i)$. We denote by Q_0 and K_0 the distributions of $\mathbf{x}_i$ and u_i, respectively.

To identify β_0, without requiring that either (i) K_0 is symmetric around 0 or (ii) K_0 satisfies a centring condition (as, e.g., $\mathrm{E}_{K_0}[u] = 0$), we assume that

$$\mathrm{P}_{Q_0}(g(\mathbf{x}, \beta_0) = g(\mathbf{x}, \beta) + \alpha) < 1 \tag{3}$$

for all $\beta \neq \beta_0$ and for all α. To satisfy this condition it is required that, if there is an intercept, it should be included in the error term u_i instead of being a parameter of the regression function $g(\mathbf{x}, \beta)$. For a linear regression model, we have $g(\mathbf{x}, \beta) = \beta'\mathbf{x}$ and so, in this case, condition (3) means that the vector $\mathbf{x}_i$ is not concentrated on any hyperplane.

The function g can be selected by cross validation using different robust criteria as in Hubert & Engelen (2007), Boente & Rodriguez (2008), and Boente, González-Manteiga, & Pérez-González (2009). One possibility is to consider that g is a polynomial functional whose degree can be chosen by any robust selection criteria for linear models. Examples of such criteria for linear models are given in Ronchetti & Staudte (1994), Ronchetti, Field, & Blanchard (1997), Khan, van Aelst, & Zamar (2007), and Section 5.12 of Maronna, Martin, & Yohai (2006).

3. THE PROPOSED ESTIMATORS

Recall that K_0 denotes the distribution of u_i and let R_0 denote the distribution of $g(\mathbf{x}_i, \beta_0)$. Independence between $\mathbf{x}_i$ and u_i implies that F_0 is given by the convolution between R_0 and K_0. Then, by convoluting consistent estimators $\widehat{R}_n$ and $\widehat{K}_n$ of each of these distributions, we get a consistent estimator for F_0.

In order to estimate R_0 and K_0 we need to have a robust and strongly consistent estimator $\widehat{\beta}_n$ of β_0. This estimator may be, for example, an S-estimator (see Rousseeuw & Yohai, 1984) or an MM-estimator (see Yohai, 1987). Since u_i is independent of a_i, $\widehat{\beta}_n$ may be obtained by a robust fit of the model using the data for which y_i is available, that is, using the observations $(\mathbf{x}_i, y_i)$ with $a_i = 1$. Let $\widehat{R}_n$ be the empirical distribution of $g(\mathbf{x}_j, \widehat{\beta}_n)$, $1 \leq j \leq n$, defined by

$$\widehat{R}_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{g(\mathbf{x}_j, \widehat{\beta}_n)}, \tag{4}$$

where δ_s denotes the point mass distribution at s.

Let $A = \{i : a_i = 1\}$ and $m = \#A$. For $i \in A$ consider

$$\widehat{u}_i = y_i - g(\mathbf{x}_i, \widehat{\beta}_n).$$

The estimator $\widehat{K}_n$ of K_0 is defined as the empirical distribution of $\{\widehat{u}_i : i \in A\}$:

$$\widehat{K}_n = \frac{1}{m} \sum_{i \in A} \delta_{\widehat{u}_i} = \frac{1}{\sum_{i=1}^{n} a_i} \sum_{i=1}^{n} a_i \delta_{\widehat{u}_i}. \tag{5}$$

Then, we estimate F_0 by

$$\widehat{F}_n = \widehat{R}_n * \widehat{K}_n, \tag{6}$$

where $*$ denotes convolution. Note that $\widehat{F}_n$ is the empirical distribution of the nm points

$$\widehat{y}_{ij} = g(\mathbf{x}_j, \widehat{\beta}_n) + \widehat{u}_i, \quad 1 \le j \le n,\ i \in A,$$

and therefore we can also express $\widehat{F}_n$ as

$$\widehat{F}_n = \frac{1}{nm}\sum_{i\in A}\sum_{j=1}^{n}\delta_{\widehat{y}_{ij}} = \frac{1}{n\sum_{i=1}^{n} a_i}\sum_{i\in A}\sum_{j=1}^{n}\delta_{\widehat{y}_{ij}}. \tag{7}$$

Finally, we estimate μ_0 by

$$\widehat{\mu}_n = T_L(\widehat{F}_n). \tag{8}$$

Since we have assumed weak continuity of T_L at F_0, in order to prove that $\widehat{\mu}_n$ is a strongly consistent estimator of μ_0, we only need to prove that $\widehat{F}_n \to_w F_0$ a.s. Observe that

$$\mathrm{E}_{\widehat{F}_n}[h(y)] = \frac{1}{nm}\sum_{i\in A}\sum_{j=1}^{n} h(\widehat{y}_{ij}).$$

The right hand side of this equation was proposed by Müller (2009) to estimate $\mathrm{E}_{F_0}\left[h(y)\right]$.

A property that characterizes robust functionals is weak continuity. When a functional T is weakly continuous, a small change in the underlying distribution (e.g., when there is a small fraction of outliers) has a minor influence on the asymptotic value of the associated estimator. Assume that $\widehat{\beta}_n = \mathbf{T}_R(G_n^*)$, where G_n^* is the empirical distribution of the pairs $(\mathbf{x}_i, y_i)$ with $a_i = 1$, and $\mathbf{T}_R$ is a weakly continuous regression functional. Then, we will show that if T_L is weakly continuous and $g(\mathbf{x}, \beta)$ is continuous in β, the functional T^* associated with the proposed estimator is also weakly continuous. Note that this functional depends on M, the joint distribution of $(y, \mathbf{x}, a)$, and is defined as follows. Let $G^{(M)}$ be the marginal distribution of $(y, \mathbf{x})$ given $a = 1$, when $(y, \mathbf{x}, a)$ is distributed according to M. Let $R^{(M)}$ and $H^{(M)}$ be the distributions of $y - g(\mathbf{x}, T_R(G^{(M)}))$ given $a = 1$ and of $g(\mathbf{x}, T_R(G^{(M)}))$, respectively. Finally let $F^{(M)}$ be the convolution between $R^{(M)}$ and $H^{(M)}$. Then, the functional associated with our procedure can be written as $T^*(M) = T_L(F^{(M)})$. To prove the weak continuity of T^* we start by observing that if $M_n \to_w M_0$ then $G^{(M_n)} \to_w G^{(M_0)}$. Moreover, the continuity of g and T_R implies that $R^{(M)} \to_w R^{(M_0)}$ and $H^{(M)} \to_w H^{(M_0)}$. Then, since the convolution preserves weak convergence, (proved in Lemma 2 (i) in the Supplementary Material), we get that $F^{(M_n)} \to_w F^{(M_0)}$ and thus, by the weak continuity of T_L we obtain $T^*(M_n) \to T^*(M_0)$, proving the weak continuity of T^*.

We should emphasize that the procedure defined in this section can be applied to any continuous location functional, for example M-functionals, L-functionals, that is, functionals associated with estimators based on linear combination of order statistics, and R-functionals, that is, functionals associated with estimators based on ranks.

4. CONSISTENCY AND ASYMPTOTIC DISTRIBUTION

Let $(\mathbf{x}_i, y_i)$ and u_i satisfy model (2), with u_i independent of $(\mathbf{x}_i, a_i)$. Denote by G_0, Q_0 and K_0 the distributions of $(\mathbf{x}_i, y_i)$, $\mathbf{x}_i$ and u_i, respectively, and denote by G_0^* and Q_0^* the distributions of $(\mathbf{x}_i, y_i)$ and $\mathbf{x}_i$ conditioned on $a_i = 1$, respectively.

The MAR condition implies that under G_0^* model (2) is still satisfied with $\mathbf{x}_i^*$ and u_i^* independent, $\mathbf{x}_i^*$ with distribution Q_0^* and u_i^* with distribution K_0. We also assume that the regression function g satisfies the following assumption:

A0. The function $g(\mathbf{x}, \beta)$ is twice continuously differentiable with respect to β and there exists $\delta > 0$ such that

$$\mathrm{E}_{Q_0}\left[\sup_{\|\beta-\beta_0\|\leq\delta}\|\dot{g}(\mathbf{x}_1,\beta)\|^2\right] < \infty \quad \text{and} \quad \mathrm{E}_{Q_0}\left[\sup_{\|\beta-\beta_0\|\leq\delta}\|\ddot{g}(\mathbf{x}_1,\beta)\|\right] < \infty, \tag{9}$$

where $\dot{g}(\mathbf{x}, \beta)$ and $\ddot{g}(\mathbf{x}, \beta)$ denote, respectively, the vector of first derivatives and the matrix of second derivatives of g with respect to β, and for any matrix A, $||A||$ denotes its L_2 norm.

In order to prove the consistency and the asymptotic normality of $\widehat{\mu}_n$ the following assumptions on $\widehat{\beta}_n$ and T_L are required.

A1. $\{\widehat{\beta}_n\}$ is strongly consistent for β_0.

A2. The regression estimator $\widehat{\beta}_n$ satisfies

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = \frac{1}{n^{1/2}}\sum_{i=1}^{n} a_i I_R(\mathbf{x}_i, y_i) + o_P(1) \tag{10}$$

for some function $I_R(\mathbf{x}, u)$ with $\mathrm{E}[a_i I_R(\mathbf{x_i}, y_i)] = 0$ and finite second moments.

A3. T_L is weakly continuous at F_0.

A4. The following expansion holds:

$$\sqrt{n}\big(T_L(\widehat{F}_n) - T_L(F_0)\big) = \sqrt{n}\mathrm{E}_{\widehat{F}_n}\left[I_{T_L,F_0}(y)\right] + o_P(1), \tag{11}$$

where I_{T_L,F_0} is the influence function, see Hampel (1974), of T_L at F_0. We assume also that $\mathrm{E}_{F_0}[I_{T_L,F_0}(y)] = 0$, $\mathrm{E}_{F_0}[I^2_{T_L,F_0}(y)] < \infty$, and I_{T_L,F_0} is differentiable with $|I'_{T_L,F_0}(y)|$ bounded.

The following theorem shows the consistency of $\widehat{\mu}_n = T_L(\widehat{F}_n)$.

Theorem 1. *Let $\widehat{F}_n$ be defined as in (7) and assume that A0 and A1 hold. Then*

(a) $\{\widehat{F}_n\}$ converges weakly to F_0 a.s., that is,

$\mathrm{P}(\widehat{F}_n \to_w F_0) = 1.$

(b) Assume also that A3 holds; then $\widehat{\mu}_n = T_L(\widehat{F}_n)$ converges a.s. to $\mu_0 = T_L(F_0)$.

In order to find the asymptotic distribution of $\widehat{\mu}_n$, define $\eta = \mathrm{E}[a_1]$,

$$\begin{aligned}
\mathbf{c} &= \mathrm{E}\big[a_1\, I'_{T_L,F_0}(y_1 - g(\beta_0,\mathbf{x}_1) + g(\beta_0,\mathbf{x}_2))\,\{\dot{g}(\beta_0,\mathbf{x}_2) - \dot{g}(\beta_0,\mathbf{x}_1)\}\big],\\
e(\mathbf{x}_i, u_i, a_i) &= \mathrm{E}\big[a_i I_{T_L,F_0}(u_i + g(\mathbf{x}_j, \beta_0))|u_i, a_i\big]\\
&= a_i\mathrm{E}\big[I_{T_L,F_0}(u_i + g(\mathbf{x}_j, \beta_0))|u_i, a_i\big],\\
f(\mathbf{x}_j) &= \mathrm{E}\big[a_i I_{T_L,F_0}(u_i + g(\mathbf{x}_j, \beta_0))|\mathbf{x}_j\big],\\
\tau^2 &= \frac{1}{\eta^2}\mathrm{E}\big[\{e(\mathbf{x}_1, u_1, a_1) + f(\mathbf{x}_1) + a_1\mathbf{c}' I_R(\mathbf{x}_1, u_1)\}^2\big].
\end{aligned}$$

Then, the following theorem gives the asymptotic normality of the estimator $\widehat{\mu}_n$, defined in (8).

Theorem 2. *Assume A0–A4. Then*

$$n^{1/2}(\widehat{\mu}_n - \mu_0) \rightarrow_d N(0, \tau^2). \tag{12}$$

4.1. The Median as Location Parameter

The median is one of the most popular robust location functionals. However, since for this case A4 is not satisfied, Theorem 2 can not be applied to prove the asymptotic distribution of the median of $\widehat{F}_n$, where $\widehat{F}_n$ is defined at (7). In this subsection, we will prove consistency and asymptotic distribution for the median of $\widehat{F}_n$, assuming that $\{\widehat{\beta}_n\}$ satisfies A1 and A2.

The functional T_{med} is defined by

$$T_{\text{med}}(F) = \arg\min_{\mu} \mathrm{E}_F[|y - \mu|]. \tag{13}$$

When there is more than one value attaining the minimum, the functional is defined by choosing any of them. We have the following result, whose proof needs an extra argument to compensate for the absence of differentiability of $I_{T_{\text{med}},F_0}(y)$.

Theorem 3. *Assume that $\mu_0 = T_{\text{med}}(F_0)$ is well defined and let $\widehat{\mu}_n = T_{\text{med}}(\widehat{F}_n)$. Suppose that F_0 is continuous and strictly increasing at μ_0 and that A0–A1 holds. Then*

(a) We have $\widehat{\mu}_n \rightarrow \mu_0$ a.s.

(b) Assume also that A2 holds, that F_0 and K_0 have continuous and bounded densities f_0 and k_0 respectively, and that $f_0(\mu_0) > 0$. Then

$$n^{1/2}(\widehat{\mu}_n - \mu_0) \rightarrow_d N(0, \tau^2), \tag{14}$$

where τ^2 is as in Theorem 2, with $\mathbf{c}$ *replaced by*

$$\mathbf{c}^* = \frac{1}{\eta f_0(\mu_0)} \mathrm{E}[a_1 k_0(-g(\mathbf{x}_2, \beta_0) + \mu_0)\{\dot{g}(\mathbf{x}_2, \beta_0) - \dot{g}(\mathbf{x}_1, \beta_0)\}] \tag{15}$$

and $I_{T_L,F_0}(y)$ replaced by

$$I_{T_{\text{med}},F_0}(y) = \frac{\operatorname{sign}(y - \mu_0)}{2 f_0(\mu_0)}.$$

5. BREAKDOWN POINT

Consider first a dataset of n complete observations $\mathbf{Z} = \{\mathbf{z}_1, .., \mathbf{z}_n\}$, where $\mathbf{z}_i \in \mathbb{R}^j$, and let $\widehat{\theta}_n(\mathbf{Z})$ be an estimator of a parameter $\theta \in \mathbb{R}^k$ defined on all possible datasets. Donoho & Huber (1983) define the finite sample breakdown point (FSBP) of $\widehat{\theta}_n$ at $\mathbf{Z}$ by

$$\varepsilon^*(\widehat{\theta}_n, \mathbf{Z}) = \min\left\{ \frac{s}{n} : \sup_{\mathbf{Z}^* \in \mathcal{Z}_s} \|\widehat{\theta}_n(\mathbf{Z}^*)\| = \infty \right\},$$

where

$$\mathcal{Z}_s = \left\{ \mathbf{Z}^* = \{\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*\} : \sum_{i=1}^{n} I_{\{\mathbf{z}_i^* \neq \mathbf{z}_i\}} \leq s \right\}.$$

Then ε^* is the minimum fraction of outliers required to take the estimator beyond any bound. Now, we extend the notion of FSBP for samples with missing data. Let

$$\mathbf{W} = \{(\mathbf{x}_1, y_1, a_1), ...(\mathbf{x}_n, y_n, a_n)\} \tag{16}$$

be the set of all observations and missingness indicators, and let $A = \{i : 1 \leq i \leq n,\ a_i = 1\}$, $m = \#A$. Denote by $\mathcal{W}_{ts}$ the set of all samples obtained from $\mathbf{W}$ where no more than t points are replaced by outliers, with at most s of these replacements corresponding to the non missing observations. Then $\mathbf{W}^* = \{(\mathbf{x}_1^*, y_1^*, a_1), ...(\mathbf{x}_n^*, y_n^*, a_n)\}$ belongs to $\mathcal{W}_{ts}$ if

$$\sum_{i \in A} I_{\{(\mathbf{x}_i^*, y_i^*) \neq (\mathbf{x}_i, y_i)\}} + \sum_{i \in A^C} I_{\{\mathbf{x}_i^* \neq \mathbf{x}_i\}} \leq t$$

and

$$\sum_{i \in A} I_{\{(\mathbf{x}_i^*, y_i^*) \neq (\mathbf{x}_i, y_i)\}} \leq s.$$

Given an estimator $\widehat{\mu}_n$ of μ_0, we define

$$M_{ts} = \sup_{\mathbf{W}^* \in \mathcal{W}_{ts}} \left| \widehat{\mu}_n(\mathbf{W}^*) \right|$$

and

$$\kappa(t, s) = \max \left(\frac{t}{n}, \frac{s}{m} \right).$$

We define the finite sample breakdown point (FSBP) of an estimator $\widehat{\mu}_n$ at $\mathbf{W}$ by

$$\varepsilon^* = \min\{\kappa(t, s) : M_{ts} = \infty\}.$$

This definition means that ε^* is the minimum fraction of outliers in the complete sample or in the set of non missing observations required to take the estimator beyond any bound.

In order to get a lower bound for the FSBP of the location estimator $\widehat{\mu}_n$ introduced in (8), we need to define the *uniform asymptotic breakdown point* ε_U^* of T_L as follows:

Definition 1. *Given a functional T_L, its uniform asymptotic breakdown point (UABP) $\varepsilon_U^*(T_L)$ is defined as the supremum of all $\varepsilon > 0$ satisfying the following property: for all $M > 0$ there exists $K > 0$ depending on M so that*

$$\mathrm{P}_F(|y| \leq M) > 1 - \varepsilon \Longrightarrow |T_L(F)| < K. \tag{17}$$

It is easy to show that for any location functional T_L we have that $\varepsilon_U^*(T_L) \leq 0.5$ and that $\varepsilon_U^*(T_{\text{med}}) = 0.5$. The following theorem gives a lower bound for the FSBP of the estimator $\widehat{\mu}_n$ defined in (8).

Theorem 4. *Let* $\mathbf{W}$ *be given by (16) and let* $\mathbf{Z} = \{(\mathbf{x}_i, y_i) : i \in A\}$. *Suppose that* $\widehat{\beta}_n = \widetilde{\beta}_m\,(\mathbf{Z})$, *where* $\widetilde{\beta}_m$ *is a regression estimator for samples of size m. Let* $\varepsilon_1 > 0$ *be a lower bound of the FSBP at* $\mathbf{Z}$ *of* $\widetilde{\beta}_m$ *and let* $\varepsilon_2 > 0$ *be a lower bound of the UABP of* T_L. *Then the FSBP* ε^* *of the estimator* $\widehat{\mu}_n$ *at* $\mathbf{W}$ *satisfies*

$$\varepsilon^* \geq \varepsilon_3 = \min\left(\varepsilon_1, 1 - \sqrt{1 - \varepsilon_2}\right). \tag{18}$$

In the next section we introduce MM-estimators of regression. The maximum value of ε_1 for an MM-estimator of regression is $(n - c(G_n^*))/(2n)$ (see Martin et al. 2006), where $c(G)$ is defined by (24). In Theorem 8 we show that M-location functionals may have $\varepsilon_2 = 0.5$. Then, if $c(G_n^*)/n$ is small, we can have ε_3 close to $1 - \sqrt{0.5} = 0.293$. A similar statement holds when T_L is the median. Instead, as we will see in Section 7.1, the value of ε_2 for location L-functionals is in general smaller than 0.5.

6. ESTIMATING THE REGRESSION PARAMETER: MM-REGRESSION FUNCTIONALS

In this section, we introduce robust regression estimators satisfying A1 and A2. Several robust estimators for the parameters of the regression model (2) based on complete data $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ have been proposed. In this paper we will consider the MM-estimator. However any other robust regression estimator satisfying A1 and A2 can be used.

The MM-estimators were introduced by Yohai (1987) for the linear model while Fasano et al. (2011) extended these estimators to the case of nonlinear regression. For linear regression, MM-estimators may combine the highest possible breakdown point with an arbitrarily high efficiency in the case of Gaussian errors. It will be convenient to present MM-estimators of β_0 in their functional form, that is, as a functional $\mathbf{T}_{\mathrm{MM},\beta}(G)$ defined on a set of distributions in $\mathbb{R}^{p+1}$, taking values in $\mathbb{R}^q$. Given a sample $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ the corresponding estimator of β_0 is given by $\widehat{\beta}_{MM} = \mathbf{T}_{\mathrm{MM},\beta}(G_n)$, where G_n is the empirical distribution of the sample. As we explained in the Introduction, we have excluded the intercept in model (2). However, to get consistent estimators of β_0 without requiring symmetric errors, it is necessary to estimate an additional parameter, which can be naturally interpreted as an intercept or a centre of the error distribution. For this purpose consider $\xi = (\beta, \alpha)$ with $\alpha \in \mathbb{R}$, and define $\underline{g}(\mathbf{x},\xi) = g(\mathbf{x}, \beta) + \alpha$.

We need the following definition

Definition 2. *A function* $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$ *is called a rho-function if (i)* ρ *is continuous, (ii)* ρ *is even, (iii)* $\rho(t)$ *is a non-decreasing function of* $|t|$ *and (iv)* $\rho(0) = 0$. *If* ρ *is bounded, without loss of generality, we will assume that* $\rho(\infty) = 1$.

To define a regression MM-functional $\mathbf{T}_{\mathrm{MM}}(G) = \left(\mathbf{T}_{\mathrm{MM},\beta}(G), T_{\mathrm{MM},\alpha}(G)\right)$, two bounded rho-functions, ρ_0 and ρ_1, are required. The function ρ_0 is used to define a dispersion functional $S(G)$ of the error distribution as follows. For any distribution G of $(\mathbf{x}, y)$ and $\xi = (\beta, \alpha)$, let $S^*(G, \xi)$ be defined by

$$\mathrm{E}_G\left[\rho_0\left(\frac{y - \underline{g}(\mathbf{x}, \xi)}{S^*(G, \xi)}\right)\right] = \delta, \tag{19}$$

where $\delta \in (0, 1)$. Then the dispersion functional $S(G)$ is defined by

$$S(G) = \min_{\xi \in B \times \mathbb{R}} S^*(G, \xi) \tag{20}$$

and the MM-estimating functional $\mathbf{T}_{\mathrm{MM}}(G) = \big(\mathbf{T}_{\mathrm{MM},\beta}(G), T_{\mathrm{MM},\alpha}(G)\big)$ by

$$\mathbf{T}_{\mathrm{MM}}(G) = \arg\min_{\xi\in B\times\mathbb{R}} \mathrm{E}_G\left[\rho_1\left(\frac{y-\underline{g}(\mathbf{x},\xi)}{S(G)}\right)\right]. \tag{21}$$

We can also consider another regression functional $\mathbf{T}_{\mathrm{S}}(G) = \big(\mathbf{T}_{\mathrm{S},\beta}(G), T_{\mathrm{S},\alpha}(G)\big)$, called the S-regression functional, as follows:

$$\mathbf{T}_{\mathrm{S}}(G) = \arg\min_{\xi\in B\times\mathbb{R}} \mathrm{E}_G\left[\rho_0\left(\frac{y-\underline{g}(\mathbf{x},\xi)}{S(G)}\right)\right], \tag{22}$$

where $S(G)$ is defined by (20).

In the case of a linear regression model, the asymptotic breakdown point of both $\mathbf{T}_{\mathrm{MM}}$ and $\mathbf{T}_{\mathrm{S}}$ is given by

$$\varepsilon^* = \min(\delta, 1-\delta-c(G)), \tag{23}$$

where

$$c(G) = \sup_{\gamma\neq 0,\gamma\in\mathbb{R}^{p+1}} \mathrm{P}_G(\gamma'(\mathbf{x}',1)'=0). \tag{24}$$

See, for example, Maronna, Martin, & Yohai (2006), Chapter 5. The maximum breakdown point occurs when $\delta = (1-c(G))/2$ and its value is $(1-c(G))/2$. It can be proved that this is the maximum possible breakdown point for equivariant regression functionals. In the case of nonlinear regression both $\mathbf{T}_{\mathrm{MM}}$ and $\mathbf{T}_{\mathrm{S}}$ have also the same breakdown point, but it is not given by a simple closed expression (see Fasano, 2009).

Yohai (1987) showed that MM-estimators for linear regression may combine the highest possible breakdown point $(1-c(G))/2$ with a Gaussian efficiency as high as desired. However, Hössjer (1992) showed that this is not possible for S-estimators. The maximum asymptotic Gaussian efficiency of an S-estimator with $\varepsilon^* = (1-c(G))/2$ is 0.33.

Let $(\mathbf{x}, y)$ and u satisfy model (2). Let $\{\mathrm{G}_n^*\}$ be the sequence of empirical distribution associated with observed pairs $(\mathbf{x}_i, y_i)$, that is, those pairs such that
$a_i = 1$. That is,

$$G_n^* = \frac{1}{\sum_{i=1}^n a_i}\sum_{i=i}^n a_i\delta_{(\mathbf{x}_i,y_i)}. \tag{25}$$

Then we can estimate β_0 by

$$\widehat{\beta}_n = \mathbf{T}_{\mathrm{MM},\beta}\big(G_n^*\big). \tag{26}$$

For the validity of assumptions A1 and A2, the rho-functions used to define the regression MM-functionals should satisfy assumptions R1 and R2 below.

R1. For some m, $\rho(u) = 1$ iff $|u| \geq m$, and $\log(1-\rho)$ is concave on $(-m, m)$.
R2. The function ρ is twice continuously differentiable.

A family of very popular bounded rho-functions satisfying R1 and R2 is Tukey's bisquare family:

$$\rho_{T,k}(u) = 1-\left(1-\left(\frac{u}{k}\right)^2\right)^3 I(|u|\leq k). \tag{27}$$

We denote by ψ_0 and ψ_1 the derivatives of ρ_0 and ρ_1, respectively. Let $\alpha_{01} = T_{\mathrm{MM},\alpha}(G_0^*)$, $\alpha_{00} = T_{\mathrm{S},\alpha}(G_0^*)$ and $\sigma_0 = S(G_0^*)$.

Regression MM- and S-functionals are studied in detail in Fasano et al. (2011). There we can find sufficient conditions for weak continuity and Fisher-consistency. Moreover, a weak differentiability notion involving the influence function of the functionals is also developed. This notion allows us to obtain asymptotic expansions, like the one required in (10). The following numbers will be used to derive the influence functions of the regression functionals:

$$a_{0i} = \mathrm{E}_{K_0}\left[\psi_i'\left((u - \alpha_{0i})/\sigma_0\right)\right], \quad i = 0, 1,$$

$$d_0 = \mathrm{E}_{K_0}\left[\psi_0\left((u - \alpha_{00})/\sigma_0\right)(u - \alpha_{00})/\sigma_0\right], \quad \mathbf{b}_0 = \mathrm{E}_{G_0^*}[\dot{g}(\mathbf{x}, \beta_0)].$$

We denote by A_0 the covariance matrix of $\dot{g}(\mathbf{x}, \beta_0)$ under Q_0^*.

The following theorem shows that conditions A1 and A2 are satisfied by MM-estimators of the regression parameter.

Theorem 5. *Assume that A0 holds and let ρ_0 and ρ_1 be bounded rho-functions satisfying R1, with $\rho_1 \leq \rho_0$. Assume that K_0 has a strongly unimodal density and that (3) holds replacing Q_0 by Q_0^*. We will consider that either (a) B is compact or (b) $g(\mathbf{x}, \beta) = \beta'\mathbf{x}$ and $\delta < 1 - c(G_0^*)$. Then*

(i) $\lim_{n\to\infty} \mathbf{T}_{\mathrm{MM},\beta}(G_n^*) = \beta_0$ *a.s. and therefore* $\widehat{\beta}_n$ *satisfies A1. Moreover,* $\lim_{n\to\infty} \mathbf{T}_{\mathrm{MM},\alpha}(G_n^*) = \alpha_{01}$.

(ii) Assume also that a_{00}, a_{01} *and* d_0 *are different from* 0 *and that* ρ_0 *and* ρ_1 *satisfy R2. Then (10) holds with* $I_R(\mathbf{x}, y) = I_{\mathbf{T}_{\mathrm{MM},\beta},G_0^*}(\mathbf{x}, y)/\mathrm{E}[a_1]$, *where* $I_{\mathbf{T}_{\mathrm{MM},\beta},G_0^*}(\mathbf{x},y)$ *is the influence function of* $\mathbf{T}_{\mathrm{MM},\beta}$ *at* G_0^*, *given by*

$$I_{\mathbf{T}_{\mathrm{MM},\beta},G_0^*}(\mathbf{x},y) = \frac{\sigma_0}{a_{01}}\psi_1\left(\frac{y - \underline{g}(\mathbf{x}, (\beta_0, \alpha_{01}))}{\sigma_0}\right) A_0^{-1}(\dot{g}(\mathbf{x}, \beta_0) - \mathbf{b}_0). \tag{28}$$

Then A2 holds.

Note that, according to Theorem 5, $\widehat{\beta}_n$ converges to β_0 without assuming symmetry for the error distribution. Instead, in general the value of α_{01} is different from $\mathrm{E}[u]$, except in the case that u has a symmetric distribution. However, since as established in Theorem 1, the consistency of $T_L(\widehat{F}_n)$ only requires the consistency of $\widehat{\beta}_n$, this is not a problem.

When ρ_0 and ρ_1 are taken in the bisquare family, we have to choose the values of the corresponding tuning constants k_0 and k_1. To get MM-estimators with breakdown point 0.5, we should set k_0=1.55 and $\delta = 0.5$. Maronna, Martin, & Yohai (2006) recommend setting $k_1 = 3.44$ as a good trade off between robustness and efficiency. This value corresponds to an asymptotic Gaussian efficiency of 85% with respect to the LS-estimator. Larger values of k_1 allow for a larger efficiency, at the expense of sacrificing robustness.

The MM-estimator obtained with these values has a relatively high efficiency compared to the least squares (LS) estimator for a large variety of non Gaussian distributions, including asymmetric ones. In Table 1 we show the asymptotic efficiency of this MM-estimator with respect to the LS-estimator for some asymmetric distributions.

We note that in the case of the chi-squared distribution with one degree of freedom, the efficiency of the MM-estimate is very high. This is due to the fact that this distribution has a very heavy tail.

TABLE 1: Asymptotic relative efficiencies of the MM-estimator for some asymmetric distributions.

Distribution	W	χ_1^2	χ_2^2	χ_3^2	χ_4^2	$\log W$	$\log\chi_1^2$	$\log\chi_2^2$	$\log\chi_3^2$	$\log\chi_4^2$
% Efficiency	84	187	97	86	85	91	97	91	88	87

W, Weibull distribution with shape parameter equal to 10; χ_p^2, Chi squared with p degrees of freedom.

7. LOCATION FUNCTIONALS

We can apply the procedure described in Section 3 to estimate $\mu_0 = T_L(F_0)$, for any weakly continuous location functional T_L. The most popular ones are the L- and M-functionals. For this reason we will give here their influence functions that, according to Theorem 2, are necessary to compute the asymptotic variance of the estimator defined in (8). In this section we also prove that, under general conditions, The L- and M- functionals satisfy assumptions A3 and A4.

7.1. L-Functionals

The L-functionals are defined by

$$T(F) = \int_0^1 F^{-1}(v)W(v)\,\mathrm{d}v, \tag{29}$$

where $F^{-1}(u) = \inf\{y : F(y) \geq u\}$, $W : [0, 1] \to \mathbb{R}_{\geq 0}$ is a symmetric function around 0.5, non-increasing for $v \geq 0.5$, satisfying $\int_0^1 W(v)\,\mathrm{d}v = 1$. Given a sample $y_1, \ldots, y_n$, let F_n be the corresponding empirical distribution and $y_{(i)}$, $1 \leq i \leq n$, the order statistics. Then

$$T(F_n) = \sum_{i=1}^{n} w_{i,n} y_{(i)},$$

where $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ and $w_{i,n} = \int_{(i-1)/n}^{i/n} W(v)\,\mathrm{d}v$. Let $\alpha_0 = \inf\{\alpha : W(\alpha) > 0\}$. It is straightforward to show that the uniform asymptotic breakdown point of an L-functional is α_0. The influence function of an L-functional is given by (see Huber & Ronchetti, 2009)

$$I_{T,F}(y) = \int_{-\infty}^{y} W(F(u)\,\mathrm{d}u - \int_{-\infty}^{\infty} (1 - F(u))W(F(u))\,\mathrm{d}u.$$

One of the most popular location L-functionals is the α-trimmed mean. For this functional $W(u) = I_{[\alpha,1-\alpha]}(u)/(1 - 2\alpha)$, for $0 \leq \alpha < 0.5$, and $\alpha_0 = \alpha$.

The following theorem shows that L-functionals satisfy A3 and A4 under very general conditions.

Theorem 6. *Suppose that W is bounded and continuous a.e. Lebesgue. Assume also (i)* $\alpha_0 > 0$, *(ii)* K_0 *and* F_0 *have bounded densities, (iii)* **x** *has bounded support and (iv)* $n^{1/2}(\beta_n - \beta_0) = O_p(1)$. *Then the L-functional given by (29) satisfies A3 and A4.*

More information on L-functionals can be found, for example, in Huber & Ronchetti (2009).

7.2. M-Functionals of Location

An M-functional of location is defined by

$$T_{\mathrm{M}}(F) = \arg\min_{\mu} \mathrm{E}\left[\rho_1^*\left(\frac{x-\mu}{S(F)}\right)\right], \tag{30}$$

where ρ_1^* is a rho-function and $S(F)$ is a dispersion functional satisfying (i) $S(F) \geq 0$, and (ii) $S(F_{\sigma y+\mu}) = |\sigma| S(F_y)$.

The condition for the robustness of an M-functional defined by (30) is that $\psi_1^* = \rho_1^{*\prime}$ must be bounded and $S(F)$ has to be robust. However, contrary to what is required in the regression case with random covariables, it is not necessary for ρ_1^* to be bounded. For example ρ_1^* may be in the unbounded Huber family

$$\rho_k^H(u) = \begin{cases} u^2 & \text{if} \quad |u| \leq k, \\ 2k|u| - k^2 & \text{if} \quad |u| > k, \end{cases} \tag{31}$$

where $0 < k < \infty$. Note that when $k \to 0$, the function $\rho_k(u)/k \to |u|$ and therefore the corresponding functional approaches the median. However, when $k \to \infty$, the function $\rho_k(u)$ approaches u^2 and the corresponding M-functional approaches the mean. Then, for $0 < k < \infty$, the corresponding M-functional can be interpreted as an intermediate location measure between the mean and the median. We can also use as ρ_1^* a bounded function such as those in the bisquare family given in (27).

The dispersion functional S can be defined simultaneously with or separately of $T_{\mathrm{M}}(F)$. An example of an M-estimator where S is defined simultaneously with T_{M} is proposal 2 of Huber (1964). However, when $S(F)$ is defined simultaneously, the breakdown point of the location functional is smaller than 0.5. For example, Maronna, Martin, & Yohai (2006) show in pp. 60–61 that proposal 2 of Huber with a Gaussian efficiency of 95% has a breakdown point equal to 0.33. For this reason we consider here only M-functionals with $S(F)$ obtained separately. A convenient way to define the dispersion functional $S(F)$ is, as in the regression case, by means of an S-functional. For this purpose, let ρ_0^* be a bounded rho-function satisfying R1. For any distribution F of y and $\mu \in \mathbb{R}$ let $S^*(F, \mu)$ be defined by

$$\mathrm{E}_F\left[\rho_0^*\left(\frac{y-\mu}{S^*(F,\mu)}\right)\right] = \delta, \tag{32}$$

where $0 < \delta < 1$. Then the dispersion functional $S(F)$ is defined by

$$S(F) = \min_{\mu \in \mathbb{R}} S^*(F, \mu). \tag{33}$$

Note that we can also define an S-functional of location by

$$T_{\mathrm{S}}(F) = \arg\min_{\mu} S^*(F, \mu). \tag{34}$$

The breakdown point of the dispersion functional $S(F)$ is given by $\min(\delta, 1-\delta)$ and therefore its maximum is 0.5, which is attained when $\delta = 0.5$.

We will consider here two types of M-location functionals that may have simultaneously a breakdown point equal to 0.5 and high Gaussian efficiency. These two types of M-functionals use the dispersion functional $S(F)$ given in (33) with $\delta = 0.5$.

Convex ρ_1^* and bounded ψ_1^*: The functional is defined by (30), where ρ_1^* is a differentiable and convex rho-function with bounded $\psi_1^* = \rho_1^{*\prime}$. For example ρ_1^* may be in the Huber family given by (31).

Bounded ρ_1^* (MM-estimator): In this case, the functional is defined by (30) with a bounded rho-function ρ_1^*, such that $\rho_1^*(u) \leq \rho_0^*(u)$. For example, we can take $\rho_i^* = \rho_{T,k_i}$ with $k_0 \leq k_1$, where $\rho_{T,k}$ is defined in (27).

When ρ is convex, existence and uniqueness of the functional defined in (30) are guaranteed if (i) the support of F is a finite or infinite interval I where F is strictly increasing and (ii) $\psi_1^*(u) > 0$ for $u > 0$. When ρ_1^* is bounded the functional is well defined if ρ_1^* satisfies R1 and F has an unimodal density (see Theorem 7 (i) in Fasano et al., 2011).

It is easy to prove that for M-functionals the following equation holds

$$T_{\mathrm{M}}(F) = \mathrm{E}_F \left[y w \left(\frac{y - T_{\mathrm{M}}(F)}{S(F)} \right) \right], \tag{35}$$

where $w(u) = \psi(u)/u$ is even and non-increasing for $u > 0$. Then $T_{\mathrm{M}}(F)$ can be interpreted as a weighted mean, where the weights decrease with the distance to the centre $T_{\mathrm{M}}(F)$.

Let $\mu_{0\mathrm{M}} = T_{\mathrm{M}}(F_0)$, $\mu_{0\mathrm{S}} = T_{\mathrm{S}}(F_0)$, and $\sigma_0^* = S(F_0)$. Then define

$$\begin{aligned} a_{0i}^* &= \mathrm{E}_{F_0} \left[\psi_i^{*\prime} \left((y - \mu_{0i})/\sigma_0 \right) \right], \quad i = 0, 1, \\ e_0^* &= \mathrm{E}_{F_0} \left[\psi_0^{*\prime} \left((y - \mu_{0i})/\sigma_0 \right) (y - \mu_{0i})/\sigma_0 \right], \\ d_0^* &= \mathrm{E}_{F_0} \left[\psi_0^{*} \left((y - \mu_{00})/\sigma_0 \right) (y - \mu_{00})/\sigma_0 \right]. \end{aligned}$$

In both cases, ρ_1^* convex or bounded, the influence function of T_{M} is given by

$$I_{T_{\mathrm{M}},F_0}(y) = \frac{\sigma_0^*}{a_{01}^*} \psi_1^* \left(\frac{y - \mu_{0M}^*}{\sigma_0^*} \right) - \frac{e_{01}^* \sigma_0^*}{a_{01}^* d_0^*} \left(\rho_0^* \left(\frac{y - \mu_{0S}^*}{\sigma_0^*} \right) - \delta \right). \tag{36}$$

When F_0 is symmetric with respect to ν_0 we have $e_0^* = 0$, $\mu_{0M}^* = \mu_{0,S}^* = \nu$ and

$$I_{T_{\mathrm{M}},F_0}(y) = \frac{\sigma_0^*}{a_{01}^*} \psi_1^* \left(\frac{y - \nu_0}{\sigma_0^*} \right).$$

The following theorem establishes that, under general conditions, M-location functionals satisfy A3 and A4.

Theorem 7. *Assume that ρ_0^* is a bounded rho-function and that ρ_1^* is either a bounded or convex rho-function. In both cases assume that ρ_1^* is differentiable and that ψ_1^* is bounded. Let T_{M} be an M-location functional defined by (30), with $S(F)$ given by (33). Assume also that $T_{\mathrm{M}}(F_0)$ and $T_{\mathrm{S}}(F_0)$ are uniquely defined. Then*

(i) The functional T_{M} is weakly continuous at F_0, and so assumption A3 holds.
(ii) Assume also that ρ_0^* and ρ_1^* satisfy R2, and that $n^{1/2}(\widehat{\beta}_n - \beta_0) = O_P(1)$. Then the functional T_{M} satisfies assumption A4 with $I_{T_M,F_0}(y)$ given by (36).

The following theorem gives a lower bound for the uniform asymptotic breakdown point of the two types of M-location functionals proposed in this section. In both cases, the bound is 0.5 when $\delta = 0.5$.

Theorem 8. *Let T_{M} be an M-location functional defined by (30), with $S(F)$ given by (33). Then, under the same conditions as in Theorem 7, the uniform asymptotic breakdown point ε_U^* of T_{M} satisfies*

(i) $\varepsilon_U^* \geq \min(0.5, \delta)$, *if* ρ_1^* *is convex and* ψ_1^* *is bounded.*
(ii) $\varepsilon_U^* \geq \min(1-\delta, \delta)$, *if* ρ_1^* *is bounded and* $\rho_1^* \leq \rho_0^*$.

8. MONTE CARLO STUDY

We performed a Monte Carlo study to compare the classical procedure that uses as $\widehat{\beta}_n$ the LS-estimator and as T_L the mean functional, with the robust proposal presented in this work and that introduced by Bianco et al. (2010). We consider three regression models:

Model 1 The variable y is generated as $y = 5x_1 + x_2 + x_3 + 4v + 9$, where $x_{1,}x_2, x_3$, and v are independent random variables with distribution $N(0, 1)$. Note that in this model both y and the regression error have a symmetric distribution.

Model 2 The variable y is generated as $y = 5x_1 + x_2 + x_3 + 4v + 4$, where x_1, x_2, and x_3 have a chi-squared distribution with one degree of freedom, v has a standard normal distribution and the four variables are independent. In this case the regression error has a symmetric distribution but the distribution of y is asymmetric.

Model 3 In this case y is generated by the nonlinear model $y = 5\exp(-0.5x_1) + x_2 + x_3 + v$, where x_1, x_2, x_3, and v are independent random variables with a chi-squared distribution with one degree of freedom. In this case both the regression error and the distribution of y are asymmetric.

For the three models, the variable a that indicates when y is observed is generated so that

$$\log \frac{\mathrm{P}(a=1|x_1,x_2,x_3)}{1-\mathrm{P}(a=1|x_1,x_2,x_3)} = 0.15(x_1+x_2+x_3).$$

This mechanism together with the distribution of x_1, x_2, and x_3 gives $\mathrm{P}(a=1) = 0.605$ for models 2 and 3, and 0.50 for Model 1.

For the three models we study 62 cases. The first one corresponds to the central model, without outlier contaminations. Then, we consider 61 cases where 10% of the observations ($\mathbf{x}_i$ y_i) are replaced by the same values ($\mathbf{x}^*$, y^*), where $\mathbf{x}^* = (2, 0, 0)$ and with y^* varying in a grid of 61 equally spaced values in the interval $[-20, 40]$. For each of the 62 simulations we performed 1,000 replications using samples of size 100.

We considered 5 location functionals: the mean (MEAN), the median (MED), an M-location functional with ρ_1^* in the Tukey family, defined at (27) with $k_1 = 3.44$ (TU), an M-location functional with ρ_1^* in the Huber family with $k = 1.37$ (HU) and the 0.1-trimmed mean functional (TR10). For both M-location functionals we use the dispersion functional $S(G)$ defined by (33), with ρ_0^* in the Tukey's family with $k = 1.57$ and $\delta = 0.5$. Table 2 gives exact values of the 5 functionals for Model 1 and approximated ones for Models 2 and 3. The approximated values of the functionals for Models 2 and 3 were computed with one sample of size 100,000. Since in Model 1 the distribution of y is symmetric, the values of the five functionals are the same and coincide with the centre of symmetry. However, for Models 2 and 3 the five functionals take different values.

To estimate the mean, we use as $\widehat{\beta}_n$ the LS-estimator. However, for the four robust location functionals, we use as $\widehat{\beta}_n$ an MM-estimator with $\rho_i = \rho_{T,k_i}$, k_0=1.57, $k_1 = 3.44$, and $\delta = 0.5$.

For each of the four robust location functionals we consider two estimators: the one proposed in this work, defined at (8), and the one proposed by Bianco et al. (2010), which is given by

TABLE 2: Values of the location functionals.

Model	Location functional				
	MEAN	MED	TU	HU	TR10
1	9.00	9.00	9.00	9,00	9.00
2	11.00	9.47	9.29	10.02	10.05
3	6.54	6.16	6.14	6.32	6.33

TABLE 3: Monte Carlo results for Model 1 under the true model (lines 1–3) and with 10% of outliers (lines 4–5).

Estim.	MEAN	MED_1	TU_1	HU_1	TM10_1	MED_2	TU_2	HU_2	TM10_2
MSE_0	0.50	0.65	0.64	0.62	0.61	0.84	0.76	0.72	0.73
Bias $\widehat{\mu}_n$	−0.002	−0.003	0.000	−0.001	−0.001	−0.005	−0.001	−0.002	−0.003
As. bias	−0.000	−0.000	0.000	−0.000	0.000	0.000	0.000	0.000	0.000
$\text{MSE}_{\max}$	9.21	2.69	2.59	3.14	4.01	3.07	2.85	2.56	2.67
$y_{\max}$	40	25	24	40	40	22	22	23	23

$T_L(\widehat{F}_n^*)$, where $\widehat{F}_n^*$ is the empirical distribution of $g(\mathbf{x}_j, \widehat{\beta}_n) + \widehat{\alpha}_n$. The results of the simulations for Models 1–3 are shown in Tables 3–5, respectively. In these tables the results for the estimators that we propose are denoted with the subscript 1 and those corresponding to Bianco et al. (2010) with the subscript 2. The first line of these tables shows the mean squared errors without outliers (MSE_0). The second line (Bias $\widehat{\mu}_n$) shows estimates of the bias of the estimators, obtained as the mean of the 1,000 replications minus the true value of the corresponding functional. The third line gives the asymptotic bias of the estimators, defined as the asymptotic value of the estimators minus the true values of the corresponding functionals. The fourth line ($\text{MSE}_{\max}$) contains the maximum mean squared error under outlier contamination. The maximum is taken along the 61 values of y^*. The fifth line contains the value of y^* where the maximum of the fourth line is attained.

Line 3 of these tables shows that, as expected, the Bianco et al. (2010) estimators are consistent for the corresponding location functionals only for Model 1, where both y and the regression errors have symmetric distributions. However, for Models 2 and 3 these estimators have an important

TABLE 4: Monte Carlo results for Model 2 under the true model (lines 1–3) and with 10% of outliers (lines 4–5).

Estim.	MEAN	MED_1	TU_1	HU_1	TM10_1	MED_2	TU_2	HU_2	TM10_2
MSE_0	0.78	0.61	0.63	0.64	0.65	1.43	1.57	1.08	0.84
Bias $\widehat{\mu}_n$	0.04	0.00	0.00	0.03	0.02	−0.82	−0.90	−0.60	−0.37
As. bias	0.00	0.00	0.00	0.00	0.00	−0.79	−0.93	−0.61	−0.39
$\text{MSE}_{\max}$	11.10	2.75	2.94	3.13	4.19	6.77	7.37	5.56	4.41
$y_{\max}$	−20	−5	−5	40	−20	−3	−3	−3	−3

TABLE 5: Monte Carlo Results for Model 3 under the true model (lines 1–3) and with 10% of outliers (lines 4–53.

Estim.	MEAN	MED$_1$	TU$_1$	HU$_1$	TM10$_1$	MED$_2$	TU $_2$	HU$_2$	TM10$_2$
MSE$_0$	0.088	0.065	0.073	0.075	0.075	0.242	0.294	0.346	0.345
Bias $\widehat{\mu}_n$	0.001	0.001	−0.005	−0.002	0.004	−0.423	−0.466	−0.525	−0.524
As. bias	0.000	0.000	0.000	0.000	0.000	−0.466	−0.439	−0.569	−0.570
MSE$_{\max}$	12.45	0.42	0.50	0.86	1.34	0.59	0.91	1.09	1.14
$y_{\max}$	40	−19	−2	−19	−20	2	1	1	1

asymptotic bias and therefore they are not consistent. The second lines show that the bias of the estimators using samples of size 100 is very close to the asymptotic bias. Line 1 of Table 3 shows that, for Model 1, the classical estimator is the most efficient one in the absence of outliers, as would be expected for normal variables. We also observe that the robust estimators obtained by the procedure proposed in this work are more efficient than those obtained through the Bianco et al. (2010) procedure. For Models 2 and 3 the robust estimators obtained with the procedure proposed here are more efficient than the estimator of the mean. The reason is that for these two cases the distribution of y has heavy tails.

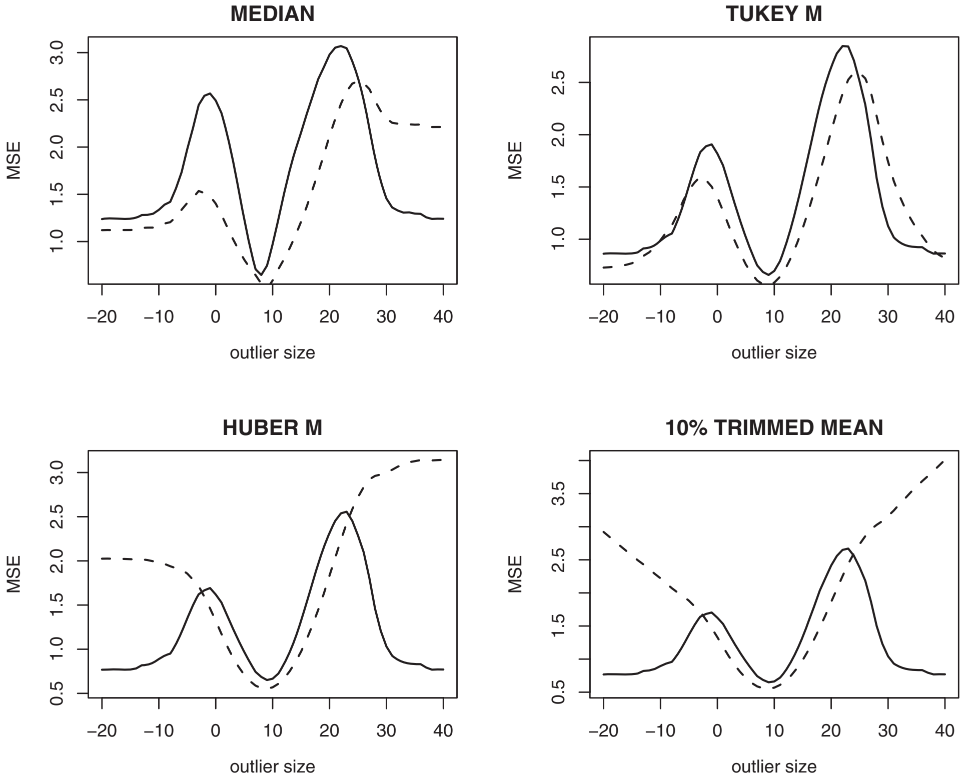

FIGURE 1: Comparison of the two estimators for Model 1 under outlier contamination. The solid line corresponds to the Bianco et al. (2010) estimator and the dashed line to the estimator proposed here.

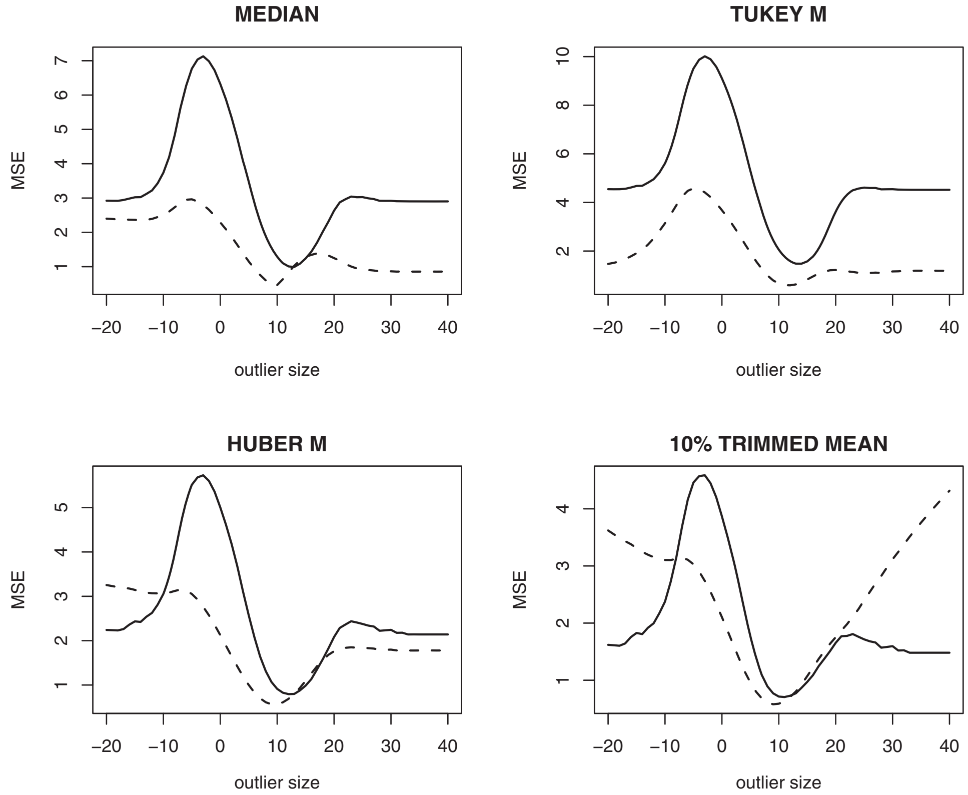

FIGURE 2: Comparison of the two estimators for Model 2 under outlier contamination. The solid line corresponds to the Bianco et al. (2010) estimator and the dashed line to the estimator proposed here.

The results in line 4 show that, under outlier contamination, the more robust estimators are MED_1 and TU_1 and, in the third place, HU_1. However, MEAN and TR10_1 break down since the mean squared error of each of these estimators goes to ∞ when the value of y^* goes to $-\infty$ or ∞. Note that, since the 0.1-trimmed mean has a breakdown point equal to 0.10 and the regression estimator has a breakdown point of 0.5, according to (18), the breakdown point of the estimator TR10_1 is 0.0523. This fact explains why TR10_1 breaks down when the sample has 10% of large outliers.

In Figures 1–3 we compare the behavior of the two estimates of each robust location functional under outlier contamination. For this purpose, we plot the mean squared errors as a function of the outlier size y^*. These figures show that for the median and the Tukey M-functionals, our proposal seems clearly preferable to that of Bianco et al. (2010) for the three models. For the Huber M-functional, our proposal seems preferable in the case of Models 2 and 3 while in the case of Model 1, the Bianco et al. (2010) estimator seems to be the best choice. Our proposed estimator breaks down for large values of y^* for the 0.1-trimmed mean functional for the reason mentioned above.

9. AN EXAMPLE

We considered a real example with complete data, and we have generated a sample with artificially missing responses by removing some of them using an MAR mechanism. In this way, we were able to compare the estimators of different location functionals using the whole set of original

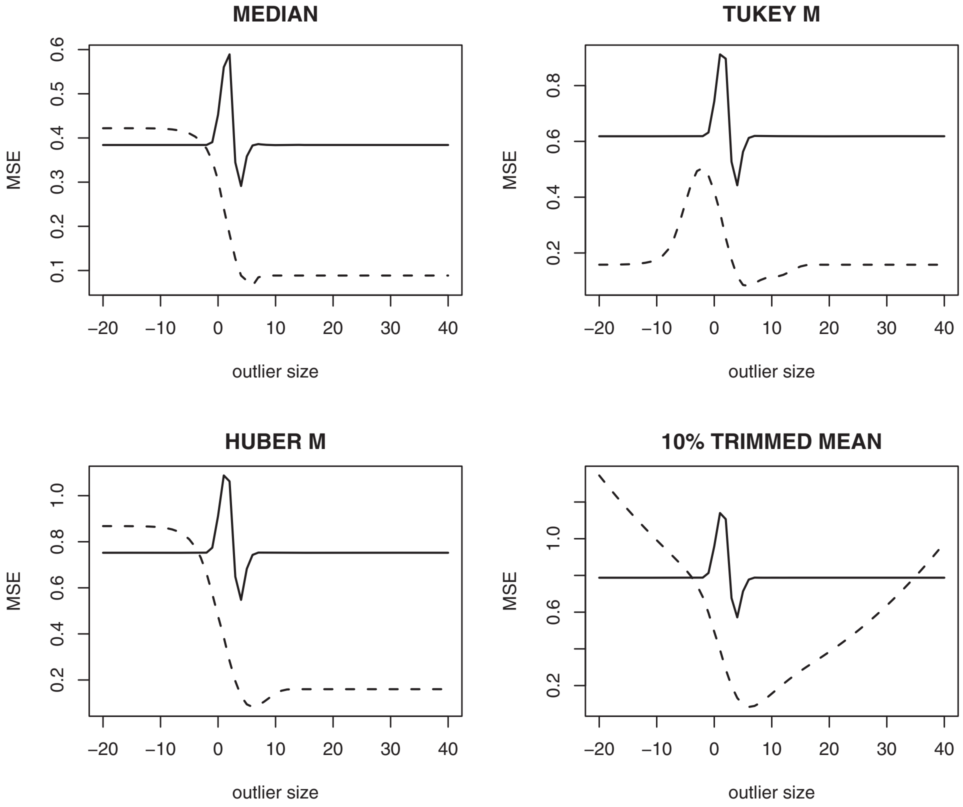

FIGURE 3: Comparison of the two estimators for Model 3 under outlier contamination. The solid line corresponds to the Bianco et al. (2010) estimator and the dashed line to the estimator proposed here.

responses with the estimators proposed for the case where there are missing data. The data we have considered were first studied by LaLonde (1986), and also analyzed by many other authors. Among them we can cite Dehejia & Wahba (1999). These data were collected to compare the annual salaries of individuals that followed an employment training program, with those that did not.

In this work, we consider the data corresponding to one of the groups, consisting of 297 individuals who were involved in a training program (National Supported Work). These data can be downloaded from http://www.nber.org/∼rdehejia/nsw_treated.txt. The variable of interest y is the annual salary corresponding to 1978. The data set also contains information about seven variables, that we use as a vector $\mathbf{x}$ of covariables. These variables are: education (in years), race condition (black–white), Hispanic condition (yes–no), educational level (no-degree indicator), married status, and earnings corresponding to year 1975. We generated the observed indicator a according to the following mechanism:

$$\log \frac{\mathrm{P}(a = 1|\mathbf{x})}{1 - \mathrm{P}(a = 1|\mathbf{x})} = 0.001x_7, \tag{37}$$

where, as defined above, x_7 represents the earnings corresponding to 1975. This mechanism produces 26% of missing responses.

The boxplot of the 297 values of y in Figure 4 shows several outliers and justifies the use of robust methods. We consider the same location functionals used for the Monte Carlo study:

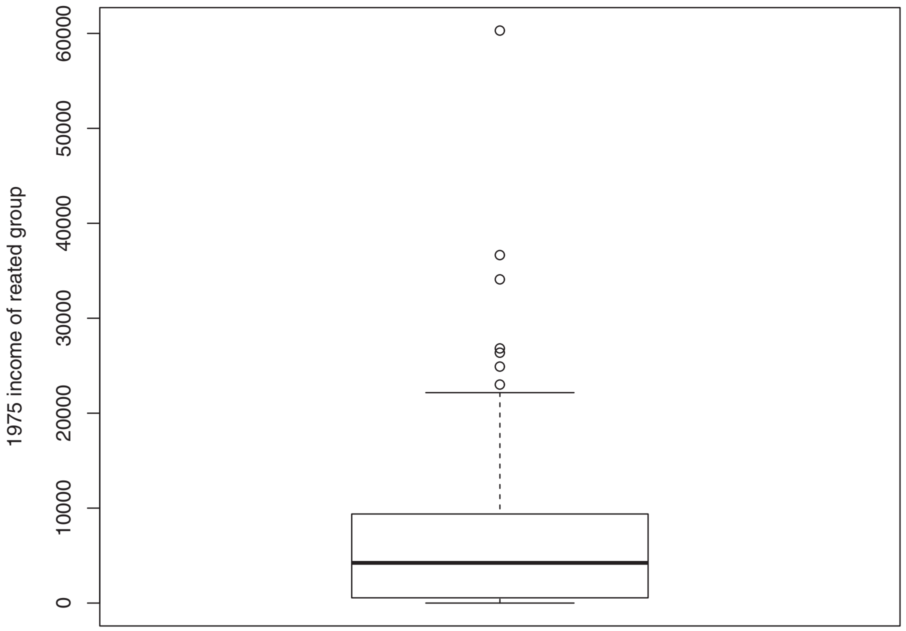

FIGURE 4: Boxplot of the 1975 income of the treated group.

the mean, the median, the Tukey M-location functional, the Huber M-location functional, and a trimmed mean functional with $\alpha = 0.1$. The values of these functionals for the empirical distribution of the whole sample of y are shown in the first row of Table 6.

Based on the sample with missing responses, we compute our proposed estimators for the five location functionals. That is, the values of these functionals for the empirical distributions of the $\widehat{y}_{ij}$'s defined in Section 3 are presented in row 2. In line 3 we present the ratio between rows 2 and 1. Row 4 contains the estimators proposed by Bianco et al. (2010) for samples with missing observations: the same functionals evaluated for the empirical distribution of the predicted values $\widehat{\alpha}_n + \widehat{\beta}'_n \mathbf{x}_i$, $1 \leq i \leq 297$. Finally, the ratio between rows 4 and 1 is presented in row 5.

We observe that, as expected, the means of the second and fourth rows are very close to each other. Moreover, in comparing rows 3 and 5, we note that, for all functionals except the mean, the estimators proposed in this paper are closer to the values of the corresponding estimates for the complete sample than the estimators proposed in Bianco et al. (2010). This may be explained by the fact that the distribution of the y_i's is not symmetric, and in this case the latter estimators are not consistent.

TABLE 6: Estimates for the NSW data.

Sample	Estimators				
	MEAN	MED	TU	HU	TR10
y_i: $1 \leq i \leq 297$	5976.35	4232.31	4234.29	5007.08	4910.59
$\widehat{y}_{ij}$	6136.84	4070.76	4195.01	4921.18	4918.68
Line 2/line 1	1.03	0.96	0.99	0.98	1.00
$\widehat{\alpha}_n + \widehat{\beta}'_n \mathbf{x}_i$	6136.38	3975.19	3956.13	4015.19	4024.39
Line 4/line 1	1.03	0.94	0.93	0.80	0.82

ACKNOWLEDGEMENTS

We are grateful to the anonymous referees and the editor for their comments that have largely contributed to improve the original manuscript, shedding light on several issues. We also thank Graciela Boente for valuable discussions.

This research was partially supported by grants from the University of Buenos Aires, the National Council for Science and Technology of Argentina (CONICET) and the National Agency for Scientific Promotion (ANPCyT).

BIBLIOGRAPHY

Bang, H. & Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61, 962–972.

Bianco, A., Boente, G., González-Manteiga, W., & Pérez-González, A. (2010). Estimation of the marginal location under a partially linear model with missing responses. *Computational Statistics and Data Analysis*, 54, 546–564.

Boente, G. & Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: Monte Carlo study and influential analysis. *Computational Statistics and Data Analysis*, 52, 2808–2828.

Boente, G., González–Manteiga, W., & Pérez–González, A. (2009). Robust nonparametric estimation with missing data. *Journal of Statistical Planning and Inference*, 139, 571–592.

Dehejia, R. H. & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association,* 94, 1053–1062.

Donoho, D. L. & Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for E. L. Lehmann*, Bickel, P. J., Doksum, K. A., & Hodges, J. L., editors. Wadsworth, Belmont, CA, pp. 157–184.

Fasano, M. V. (2009). Robust estimation in nonlinear regression. Ph.D. Thesis, University of La Plata. Available at http://www.mate.unlp.edu.ar/tesis/tesis_fasano_v.pdf.

Fasano, M. V., Maronna, R. A., Sued, M., & Yohai, V. J. (2011). Continuity and differentiability of regression M functionals. *Bernoulli* (in press).

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 383–393.

Hössjer, O. (1992). On the optimality of S-estimators, *Statistics and Probability Letters*, 14, 413–419.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35, 73–101.

Huber, P. J. & Ronchetti, E. M. (2009). *Robust Statistics*, 2nd ed., Wiley, New York.

Hubert, M. & Engelen, S. (2007). Fast cross-validation of high-breakdown resampling algorithms for PCA. *Computational Statistics and Data Analysis*, 51, 5013–5024.

Kang, J. D. Y & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22, 523–539.

Khan, J. A., Van Aelst, S., & Zamar, R. H. (2007). Robust linear model selection based on least angle regression *Journal of the American Statistical Association*, 102, 1289–1299.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, 76, 604–620.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*, Wiley, Chichester.

Müler, U. U. (2009). Estimating linear functionals in nonlinear regression with responses missing at random. *Annals of Statistics*, 37, 2245–2277.

Robins, J., Sued, M., Lei-Gomez, Q., & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when inverse probability weights are highly variable. *Statistical Science*, 22, 544–559.

Ronchetti, E., Field, R., & Blanchard, W. (1997). Robust linear model selection by cross-validation. *Journal of the American Statistical Association*, 92, 1017–1023.

Ronchetti, E. & Staudte, R. G. (1994). A robust version of Mallows's Cp. *Journal of the American Statistical Association*, 89, 550–559.

Rousseeuw, P. J. & Yohai, V. J. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series*. Vol. 26, Franke, J., Hardle, W., & Martin, R. D., editors. Lecture Notes in Statistics, Springer, New York, pp. 256–272.

Rubin, D. (1976). Inference and missing data. *Biometrika*, 63, 581–592.

Yohai, V. J. (1987). High breakdown-point and high efficiency estimates for regression. *The Annals of Statistics*, 15, 642–656.

Received 22 August 2011
Accepted 13 July 2012