Robust Median Estimator in Logistic Regression*

T. Hobza[1]; L. Pardo[2] and I. Vajda[3]

[1]Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague.

[2]Department of Statistics and O. R., Complutense University of Madrid, Madrid.

[3]Institute of Information Theory and Automation, Czech Academy of Sciences, Prague.

Abstract

The paper introduces a median estimator of the logistic regression parameters. It is defined as the classical L_1-estimator applied to continuous data $Z_1, ..., Z_n$ obtained by a statistical smoothing of the original binary logistic regression observations $Y_1, ..., Y_n$. Consistency and asymptotic normality of this estimator are proved. A method called enhancement is introduced which in some cases increases the efficiency of this estimator. Sensitivity to contaminations and leverage points is studied by simulations and compared with the sensitivity of some robust estimators previously introduced to the logistic regression. The new estimator appears to be more robust for larger sample sizes and higher levels of contamination.

Key words: Logistic regression, MLE, Morgenthaler estimator, Bianco and Yohai estimator, Croux and Haselbroeck estimator, Median estimator, Consistency, Asymptotic normality, Robustness.

AMS 2000 Subject Classification: 62F10, 62F12, 62F35.

1 Introduction and basic concepts

In this paper we study estimation of Euclidean parameters $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in the logistic regression based on independent observations $Y_1 \sim Be(\pi_1), ..., Y_n \sim Be(\pi_n)$ where the Bernoulli parameters $\pi_i = \Pr(Y_i = 1) = EY_i$ depend on $\boldsymbol{\beta}_0$ and regressors $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ from $\mathbb{R}^d$ by the formula

$$\pi_i = \pi_i(\boldsymbol{\beta}_0) = \pi(\boldsymbol{x}_i^T \boldsymbol{\beta}_0). \tag{1.1}$$

Here and elsewhere in the paper, $\boldsymbol{x}^T\boldsymbol{\beta} = \sum_{j=1}^d x_j \beta_j$ denotes the scalar product of $\boldsymbol{x} = (x_1, ..., x_d)^T$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_d)^T$ and

$$\pi(t) = \frac{e^t}{1 + e^t} \tag{1.2}$$

*This work was supported by the grants MTM 2006-06872, AV ČR 1075403 and MŠMT 1M 0572

is the standard logistic function. The MLE $\boldsymbol{\beta}_n = \boldsymbol{\beta}_n(Y_1, ..., Y_n)$ of $\boldsymbol{\beta}_0$ minimizes the deviances (negative scores)

$$\mathcal{D}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} d_i(\boldsymbol{\beta})$$

of the sample $\boldsymbol{Y}_n = (Y_1, ..., Y_n)$ where

$$\begin{aligned} d_i(\boldsymbol{\beta}) &= -\ln\left[\pi_i(\boldsymbol{\beta})^{Y_i}(1-\pi_i(\boldsymbol{\beta}))^{1-Y_i}\right] \\ &= -Y_i \ln \pi_i(\boldsymbol{\beta}) - (1-Y_i)\ln(1-\pi_i(\boldsymbol{\beta})) \end{aligned} \tag{1.3}$$

are the deviances (negative scores) of individual observations Y_i. Thus

$$\boldsymbol{\beta}_n = \arg\min \mathcal{D}_n(\boldsymbol{\beta}) = \arg\min \sum_{i=1}^{n} d_i(\boldsymbol{\beta}). \tag{1.4}$$

Notice that the expected deviances are of the form

$$E\mathcal{D}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} Ed_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} [I(\pi_i(\boldsymbol{\beta}_0) \parallel \pi_i(\boldsymbol{\beta})) + H(\pi_i(\boldsymbol{\beta}_0))] \tag{1.5}$$

where

$$I(p_0 \parallel p) = p_0 \ln \frac{p_0}{p} + (1-p_0)\ln\frac{1-p_0}{1-p}$$

is the nonnegative information divergence of the Bernoulli models $Be(p_0)$ and $Be(p)$ which reduces to zero iff $p_0 = p$ and

$$H(p_0) = -p_0 \ln p_0 - (1-p_0)\ln(1-p_0)$$

is the entropy of the model $Be(p_0)$ which does not depend on p. Therefore $\boldsymbol{\beta}_0$ is the unique $\arg\min E\mathcal{D}_n(\boldsymbol{\beta})$ unless there is $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ with the property

$$\pi_i(\boldsymbol{\beta}) = \pi_i(\boldsymbol{\beta}_0) \qquad \left(\text{i.e. } \boldsymbol{x}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) = 0\right) \text{ for all } 1 \leq i \leq n.$$

This leads to the consistency of MLE unless all regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, ...$are on a hyperplane in $\mathbb{R}^d$ (cf. Andersen (1990), Agresti (2002), Pardo et al. (2006) and references therein).

However, this optimistic picture dramatically changes as soon as the true models $Be(\pi_i)$ for $\pi_i = \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)$ are positively ε-contaminated by some alternative Bernoulli models $Be(p_i)$, e.g. by $Be(1-\pi_i)$. Then the true models are

$$Be(\pi_i + \varepsilon(1-2\pi_i)) = (1-\varepsilon)Be(\pi_i) + \varepsilon Be(1-\pi_i) \tag{1.6}$$

for some $0 < \varepsilon < 1$ which differ from $Be(\pi_i)$ unless $\pi_i = 1/2$. Hence $\boldsymbol{\beta}_0$ remains to be the unique $\arg\min E\mathcal{D}_n(\boldsymbol{\beta})$ for all sufficiently large n only in the trivial case

$$\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) = 1/2 \qquad \left(\text{i.e. } \boldsymbol{x}_i^T\boldsymbol{\beta}_0 = 0\right) \qquad \text{for all } i \geq 1. \tag{1.7}$$

In the remaining cases even small contaminations $\varepsilon > 0$ may lead to large expected deviances $d_i(\boldsymbol{\beta}_0)$ for some $i \geq 1$, thus pushing the MLE's $\boldsymbol{\beta}_n = \arg\min \mathcal{D}_n(\boldsymbol{\beta})$ far away from the true values $\boldsymbol{\beta}_0$. Indeed, the expected deviances $d_i(\boldsymbol{\beta}_0)$ are in the contaminated models (1.6) given by the formula

$$Ed_i(\boldsymbol{\beta}_0) = I(\pi_i + \varepsilon(1-2\pi_i) \parallel \pi_i) + H(\pi_i + \varepsilon(1-2\pi_i)) \qquad \text{(cf. (1.5))}.$$

Hence for small π_i (in symbols $\pi_i \approx 0$) we get $1-\pi_i \approx 1$ and $\pi_i + \varepsilon(1-2\pi_i) \approx \varepsilon$. This means that for all $n \geq i$

$$E\mathcal{D}_n(\boldsymbol{\beta}_0) \geq Ed_i(\boldsymbol{\beta}_0) \approx \varepsilon \ln \frac{\varepsilon}{\pi_i} + (1-\varepsilon)\ln(1-\varepsilon) + H(\varepsilon)$$

where the right-hand term tends to infinity for $\pi_i \to 0$. At the same time $\boldsymbol{\beta}_1 = \mathbf{0}$ satisfies $\boldsymbol{x}_i^T\boldsymbol{\beta}_1 = \mathbf{0}$ and therefore $\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_1\right) = 1/2$ (cf. (1.7)) so that for $\boldsymbol{\beta}_0 \neq \mathbf{0}$

$$E\mathcal{D}_n(\boldsymbol{\beta}_1) = 0 < E\mathcal{D}_n(\boldsymbol{\beta}_0).$$

This means that, as stated above, the MLE $\boldsymbol{\beta}_n$ will be with a great probability far away from the true parameter $\boldsymbol{\beta}_0$.

In order to restrict the undesired influence of large deviances resulting from contaminations of logistic regression models $Be(\pi_i) = Be\left(\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right)$, previous authors replaced the deviances $d_i(\boldsymbol{\beta})$ in the definition (1.4) by appropriate functions $\varrho(d_i(\boldsymbol{\beta}))$ of deviances, or even by more general expressions $\phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right)$. This led to M-estimators $\boldsymbol{\beta}_n$ of the type

$$\boldsymbol{\beta}_n = \arg\min \sum_{i=1}^{n} \varrho(d_i(\boldsymbol{\beta})) \tag{1.8}$$

and

$$\boldsymbol{\beta}_n = \arg\min \sum_{i=1}^{n} \phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\right) \tag{1.9}$$

for $\varrho : (0,\infty) \to \mathbb{R}$ and $\phi : (0,\infty) \times (0,1) \to \mathbb{R}$, or to the related M-estimators $\boldsymbol{\beta}_n$ solving for suitable function $\boldsymbol{\psi} : \mathbb{R}^{2d+1} \to \mathbb{R}^d$ the equations

$$\sum_{i=1}^{n} \boldsymbol{\psi}(Y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) = \mathbf{0}. \tag{1.10}$$

The robust estimator $\boldsymbol{\beta}_n^{(0)}$ defined by (1.8) for a special function $\varrho(t)$ increasing with the rate $\sqrt{t}$ as $t \to \infty$ was proposed by Pregibon (1982). Morgenthaler (1992) proposed the robust estimator $\boldsymbol{\beta}_n^{(1)}$ defined by (1.9) for the function

$$\phi\left(Y, \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right) = \frac{\left|Y - \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right|}{\sqrt{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)}}.$$

In order to improve asymptotic properties he redefined $\boldsymbol{\beta}_n^{(1)}$ as the solution of (1.10) for the function

$$\boldsymbol{\psi}(Y, \boldsymbol{x}, \boldsymbol{\beta}) = \sqrt{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\left(1-\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)}\left(Y - \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)\boldsymbol{x}. \tag{1.11}$$

This estimator will be called Morg-*estimator* in the sequel.

Bianco and Yohai (1996) improved the asymptotic properties and also the robustness of the Pregibon's $\boldsymbol{\beta}_n^{(0)}$ by introducing the class of M-estimators defined as minimizers (1.9) for

$$\phi\left(Y_i, \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right) = \varrho\left(d_i\left(\boldsymbol{\beta}\right)\right) + \varrho_0\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right) \tag{1.12}$$

where $\varrho(t)$ is bounded and differentiable on $(0, \infty)$ with a derivative $\varrho'(t)$ and the compensator function ϱ_0 is of the form

$$\varrho_0(p) = \varrho_1(p) + \varrho_1(1-p) \tag{1.13}$$

for ϱ_1 depending on ϱ by the formula

$$\varrho_1(p) = \int_0^p \varrho'(-\ln t)\, dt, \qquad p \in (0,1)\,. \tag{1.14}$$

These authors found pleasing statistical properties of the particular estimator fr6m their family defined by

$$\varrho(0) = 0 \quad \text{and} \quad \varrho'(t) = (1-t)\, \boldsymbol{I}\,(0 < t < 1)\,. \tag{1.15}$$

This M-estimator is denoted as $\boldsymbol{\beta}_n^{(2)}$ and called briefly BY-*estimator* in the sequel. Croux and Haesbroeck (2003) found that even more pleasing statistical properties are obtained when the BY-estimator from (1.15) is replaced by the alternative estimator from the general Bianco-Yohai class defined by

$$\varrho(0) = 0 \quad \text{and} \quad \varrho'(t) = e^{-\sqrt{1/2}}\boldsymbol{I}\,(0 < t < 1/2) + e^{-\sqrt{t}} I\,(t \geq 1/2)\,. \tag{1.16}$$

This particular M-estimator is denoted as $\boldsymbol{\beta}_n^{(3)}$ and called CH-*estimator* in the sequel.

Extensions of the Morgenthaler-type M-estimators defined by equations (1.10) were studied later by several authors, e.g. Kordzakhia et al. (2001), Adimiri and Ventura (2001), Rousseeuw and Christmann (2003), Gervini (2005) and others cited there.

In this paper we propose a new robust M-estimator of the logistic regression parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ obtained by application of the classical robust L_1-method (cf. Hampel et al. (1986), Jurečková and Sen (1996) or Zwanzig (1997)) to the continuous data

$$Z_i = Y_i + U_i, \qquad 1 \leq i \leq n \tag{1.17}$$

obtained by adding mutually and on Y_i independent $U(0,1)$-distributed (i.e. uniformly on $(0,1)$ distributed) random variables U_i to the mutually independent above introduced observations

$$Y_i \sim Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right). \tag{1.18}$$

In other words, we define the estimator

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^n \left| Z_i - m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{1.19}$$

for Z_i given by (1.17), (1.18) and for the median function

$$m(p) = F_p^{-1}(1/2) = \inf\left\{z \in \mathbb{R} : F_p(z) \geq 1/2\right\}$$

corresponding to the class of distribution functions F_p of the random variables

$$Z = Be\,(p) + U(0,1)$$

when the parameter p varies in the closed interval $[0,1]$. Obviously, for each $p \in [0,1]$ and $z \in \mathbb{R}$

$$F_p\,(z) = (1-p)\,z\boldsymbol{I}\,(0 < z < 1) + (1 - 2p + pz))\,\boldsymbol{I}\,(1 < z \leq 2) + \boldsymbol{I}\,(z > 2) \qquad (1.20)$$

and the median function has the explicit form

$$m(p) = 1 + \frac{p - 1/2}{p \vee (1-p)}, \quad 0 \leq p \leq 1. \qquad (1.21)$$

Here and in the rest of the paper, $\boldsymbol{I}\,(\cdot)$ denotes the indicator function,

$$a \vee b = \max\{a,b\} \qquad \text{and} \qquad a \wedge b = \min\{a,b\}\,.$$

The graphs of functions $F_p\,(z)$ and $m(p)$ are presented in Figures 1.1 and 1.2. Figure 1.2 displays also the inverse median function

$$m^{-1}(z) = \frac{z - 1/2}{2 - z}\boldsymbol{I}(1/2 \leq z \leq 1) + \frac{1}{2(2-z)}\boldsymbol{I}(1 < z \leq 3/2) \qquad (1.22)$$

used in the sequel.

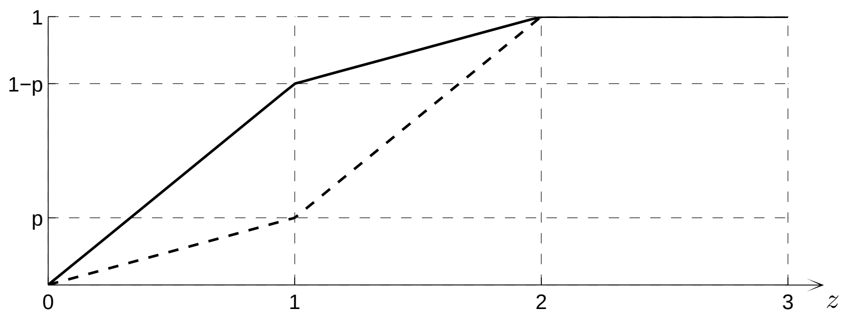

Figure 1.1: $F_p(z)$ full line, $F_{1-p}(z)$ dashed line.

Definition 1.1 The operation (1.17) is called a *statistical smoothing* of the discrete observations Y_i.

Definition 1.2 The estimator $\widehat{\boldsymbol{\beta}}_n$ defined by (1.17)-(1.19) is a *median estimator* of logistic regression parameters $\boldsymbol{\beta}_0$ called briefly *Med-estimator* in the sequel.

Remark 1.1 Median function $m(p)$ defined by the continuous random variables $Z = Be\,(p) + U(0,1)$ is strictly increasing in $p \in [0,1]$. Since the logistic function is strictly increasing too, the argument $m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)$ in (1.19) detects every change of the product $\boldsymbol{x}^T\boldsymbol{\beta}$. Contrary to this, the median function $\widetilde{m}(p)$ defined in a similar manner by the discrete random variables $Y = Be(p)$ themselves is piecewise constant in $p \in (0,1)$ so that $\widetilde{m}\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right)$ is insensitive to small variations of the product $\boldsymbol{x}^T\boldsymbol{\beta}$. Therefore the

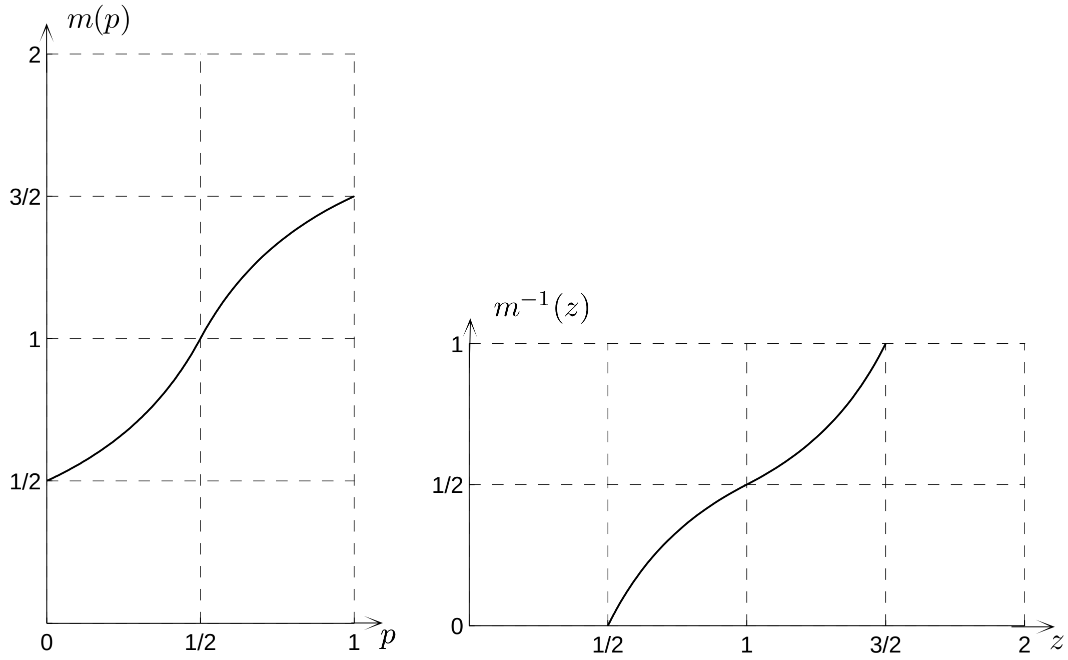

Figure 1.2: Median function $m(p)$ and its inverse $m^{-1}(z)$.

robust L_1-estimation cannot be applied directly to the logistic regression data Y_i, i.e., the estimator

$$\widetilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| Y_i - \widetilde{m}\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right) \right|$$

is not of a too much practical interest.

Remark 1.2 The operation of statistical smoothing $Z_i = T(Y_i)$ given by (1.17) is statistically sufficient because the image Z_i allows complete recovery of the original Y_i by applying the integer-part operation $[\cdot]$ to Z_i. It is equivalent to observation of discrete data through a semicontinuous channel of information theory. Such an observation procedure goes in the opposite direction than the statistical quantization frequently applied to continuous data. The quantization is usually accompanied with the loss of information so that it is not statistically sufficient.

Remark 1.3 The statistical smoothing (1.17) can be applied to integer valued observations Y_i also in others discrete statistical models. This opens much wider applicability of the methods developed for continuous models in the discrete statistics than is the one particular situation studied in this paper.

In the sections that follow we establish some desirable statistical properties of the Med-estimators such as the consistency and asymptotic normality. In simple situations we compare asymptotic variances of these estimators with the asymptotic variances of the above mentioned selected classical estimators. We verify the robustness of our Med-estimators by demonstrating their low sensitivity to high leverage outliers and also by demonstrating that they outperform the above mentioned classical robust estimators in certain special situations (e.g. heavy contaminations and large sample sizes). Our conclusions are based partly on simulations in the models used in the previous literature for mutual comparison of various estimators in logistic regression.

2 Asymptotic theory

A large class of statistical models assumes independent real valued observations $Z_1, ..., Z_n$ of the form

$$Z_i \sim F_{u\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(z), \qquad 1 \leq i \leq n. \tag{2.1}$$

Similarly as above, here $\boldsymbol{x}_i \in \mathbb{R}^d$ are vectors of explanatory variables (regressors), $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ is a vector of true parameters and $\boldsymbol{x}_i^T \boldsymbol{\beta}$ the scalar product. Further, $u : \mathbb{R} \to \Theta$ is a smooth mapping and $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$ a family of distribution functions on $\mathbb{R}$ with an interval parameter space $\Theta \subseteq \mathbb{R}$. The basic statistical problem related to these models is to find mappings $\widehat{\boldsymbol{\beta}}_n = \widehat{\boldsymbol{\beta}}_n (Z_1, ..., Z_n)$ from $\mathbb{R}^n$ into $\mathbb{R}^d$ which can be used to estimate the unknown parameters $\boldsymbol{\beta}_0$ on the basis of observations (2.1).

Various desirable asymptotic or non-asymptotic properties are usually required from estimators $\widehat{\boldsymbol{\beta}}_n$.Such properties are often found in the class of so-called *least absolute deviation estimators* (briefly L_1-estimators) defined by

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left| Z_i - m\left(u\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{2.2}$$

for a given function $m : \Theta \to \mathbb{R}$. From the extensive literature related to these estimators one can mention e.g. Koenker and Bassett (1978), Richardson and Bhattacharyya (1987), Yohai (1987), Pollard (1991), Morgenthaler (1992), Chen, Zhao and Wu (1993), Jurečková and Procházka (1994), Knight (1998), Arcones (2001), Liese and Vajda (1999, 2003, 2004) and Maronna et al. (2006).

In this section we study the asymptotics of the median estimator $\widehat{\boldsymbol{\beta}}_n$ from (1.19) which estimates the true parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ of the general logistic regression using the statistically smoothed responses

$$Z_i = Y_i + U_i \sim F_{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}(z) \qquad \text{(cf. (1.17))}$$

to the regressors $\boldsymbol{x}_i$ where $\pi(t)$ and $F_p(z)$ are given by (1.2) and (1.20). We see that our data Z_i as well as the estimator $\widehat{\boldsymbol{\beta}}_n$ are special cases of (2.1) and (2.2) for $\pi(t)$ and $F_p(z)$ given by (1.2) and (1.20) and for $m(p)$ given by (1.21).

Our results are based on what Liese and Vajda (1999, 2003, 2004) proved concerning the general median estimators (2.2) of parameters $\boldsymbol{\beta}_0$ in the general statistical models (2.1). We shall study and adapt to the present estimators (1.19) the following conditions **(c1) - (c8)** for consistency and asymptotic normality established by these authors.

(c1) The regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ are from a compact set $\mathcal{X} \subset \mathbb{R}^d$ and the probability measures

$$Q_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\boldsymbol{x}_i} \tag{2.3}$$

tend weakly for $n \to \infty$ to a probability measure Q on Borel subsets of $\mathcal{X}$.

Remark 2.1 If the regressors $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ are independently generated by a probability measure Q on the Borel subsets of a compact set $\mathcal{X} \subset \mathbb{R}^d$ then (**c1**) holds almost surely for these $\mathcal{X}$ and Q. For example, if the dimension $d = 1$ then, by the Glivenko theorem, the empirical probability measure (2.3) tends almost surely to Q in the Kolmogorov distance. But the convergence in this distance implies the weak convergence required by (**c1**).

(c2) The smallest eigenvalue of the matrix

$$\boldsymbol{\Sigma} = \int_{\mathcal{X}} \boldsymbol{x}\boldsymbol{x}^T dQ\left(\boldsymbol{x}\right) \tag{2.4}$$

is positive. Hence for every $\boldsymbol{\beta} \in \mathbb{R}^d$ different from $\boldsymbol{\beta}_0$

$$Q\left(\boldsymbol{x} \in \mathcal{X} : \boldsymbol{x}^T\left(\boldsymbol{\beta} - \boldsymbol{\beta}_0\right) \neq 0\right) > 0. \tag{2.5}$$

The following conditions **(c3)** - **(c5)** obviously hold for the distribution functions $F_p(z)$ under consideration and their densities

$$f_p\left(z\right) = \left(1-p\right)\boldsymbol{I}\left(0 < z \leq 1\right) + p\boldsymbol{I}\left(1 < z < 2\right), \qquad z \in \mathbb{R}. \tag{2.6}$$

(c3) Distributions functions $F_p\left(z\right)$ are continuous in both arguments $p \in (0,1)$ and $z \in (0,\infty)$. Moreover, for each $p \in (0,1)$

$$\int_{-\infty}^{+\infty} |z|\, f_p\left(z\right) dz = \frac{1}{2} + p < \infty. \tag{2.7}$$

(c4) Distributions functions F_p, $p \in (0,1)$ are increasing on interval $[0,2] \subseteq \mathbb{R}$ in the strict sense

$$F_\pi\left(z_1\right) < F_\pi\left(z_2\right) \text{ for } z_1 < z_2 \text{ from } [0,2] \tag{2.8}$$

and constant on the complement $\mathbb{R}-[0,2]$.

(c5) Distributions functions F_p, $p \in (0,1)$ are stochastically ordered in the sense that for every $0 < p_1 < p_2 < 1$ and $z \in \mathbb{R}$ it holds $F_{p_1}\left(z\right) \geq F_{p_2}\left(z\right)$ where

$$F_{p_1}\left(z\right) > F_{p_2}\left(z\right) \text{ if } z \in [0,2]. \tag{2.9}$$

The present conditions **(c1)** - **(c5)** imply the assumptions (E1+), (E2), (EM1), (EM2) and (M1)-(M4) of Theorem 2 and Lemmas 8 and 9 in Liese and Vajda (1999). For a detailed proof of this assertion we refer to Section 3 of Hobza et al (2005). We shall prove that in our model hold also the following less evident conditions of consistency and asymptotic normality.

(c6) for every $0 < p_1 < p_2 < 1$ there exists $a > 0$ such that the densities (2.6) and the median function $m(p)$ satisfy the condition

$$\Lambda(a) \equiv \inf_{|y| \le a} \left(\inf_{p_1 \le p \le p_2} f_p(m(p) + y) \right) > 0. \tag{2.10}$$

(c7) The quantile function $m(p)$ is differentiable on $(0,1)$ and the derivative $m'(p)$ is locally Lipschitz in the sense that for every $p_0 \in (0,1)$

$$|m'(p) - m'(p_0)| \le 2|p - p_0|.$$

(c8) The densities (2.6) satisfy for every $0 < p_1 < p_2 < 1$ the condition

$$\lim_{y \to 0} \sup_{p_1 \le p \le p_2} |f_p(m(p) + y) - f_p(m(p))| = 0. \tag{2.11}$$

Lemma 2.1 In the present model the conditions **(c6)** - **(c8)** hold.

Proof. (I) Condition (2.10) can be verified separately for $p_1 = 1/2 < p_2 < 1$ and $0 < p_1 < p_2 = 1/2$. We shall do this for $p_1 = 1/2 < p_2 < 1$ as the alternative can be verified similarly. Let $p > 1/2$ be arbitrary. By (1.21),

$$\frac{3}{2} > m(p) = 1 + \frac{2p-1}{2p} > 1$$

so that if $y \ne 0$ with $|y| \le 1/2$ is fixed then

$$f_p(m(p) + y) = f_p(m(p)) = p$$

unless $m(p) + y \le 1$ in which case $f_p(m(p) + y) = 1 - p$. Thus

$$\inf_{1/2 \le p \le p_2} f_p(m(p) + y) \ge 1 - p$$

so that (2.10) holds.

(II) The median function (1.21) is on the interval $(0,1)$ continuously differentiable with the positive derivative

$$m'(p) = \frac{1}{2[p \vee (1-p)]^2}. \tag{2.12}$$

bounded above by 2. Therefore **(c7)** holds too.

(III) Similarly as (2.10), the condition (2.11) can be verified separately for $p_1 = 1/2 < p_2 < 1$ and $0 < p_1 < p_2 = 1/2$. We shall do this for $p_1 = 1/2 < p_2 < 1$ as the alternative can be verified similarly. Let $y \ne 0$ with $|y| \le 1/2$ be arbitrary fixed. Then

$$\inf_{1/2 \le p \le p_2} f_p(m(p) + y) \ge 1 - p.$$

The absolute difference $|f_p(m(p)+y) - f_p(m(p))|$ is either zero or $p-(1-p)=2p-1$. This difference will be maximized if we take maximal p satisfying the inequality $m(p)+y \le 1$ for the fixed y under consideration. Since $m(p)$ is increasing in p, this means that

$$\sup_{1/2\le p<p_2} |f_p(m(p)+y) - f_p(m(p))| = f_{p_*}(m(p_*)+y) - f_{p_*}(m(p_*))$$

where p_* solves the equation $m(p)+y=1$. Solutions p_* exist only for $y<0$ (i.e. for $-1/2<y<0$) and then $p_* = 1/[2(1-|y|)]$. Thus we proved that

$$\sup_{1/2\le p<p_2} |f_p(m(p)+y) - f_p(m(p))| \le 2p_* - 1 = \frac{|y|}{1-|y|}$$

which implies (2.11) and completes the whole proof. ■

The median function $m(p)$ of (1.21) is bounded on $[0,1]$. By Lemma 8 in Liese and Vajda (1999), this means that the sufficient condition of Lemma 9 ibid. reduces to (2.5) assumed in **(c2).** Hence, by Theorem 2 and Lemmas 8, 9 in Liese and Vajda (1999), under **(c1)-(c5)** our Med-estimator $\widehat{\boldsymbol{\beta}}_n$ consistently estimates the true $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ provided the measure Q of **(c1)** defines the function

$$\boldsymbol{m}(\boldsymbol{\beta}) = \int_{\mathbb{R}}\int_{\mathcal{X}} \left|y - \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right| dF_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)}(y)\, \mathrm{d}Q(\boldsymbol{x}) \quad \text{for} \quad \varphi(t) = m(\pi(t)) \tag{2.13}$$

of variable $\boldsymbol{\beta} \in \mathbb{R}^d$ satisfying for every $\varepsilon > 0$ the condition

$$\inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\ge\varepsilon} \boldsymbol{m}(\boldsymbol{\beta}) > \boldsymbol{m}(\boldsymbol{\beta}_0) \tag{2.14}$$

of identifiability of true parameters $\boldsymbol{\beta}_0$. This important fact will be used in the proof of the following theorem.

Theorem 2.1 If the regressors of the model under consideration satisfy **(c1)**, **(c2)** then the Med-estimator $\widehat{\boldsymbol{\beta}}_n$ consistently estimates the model parameters $\boldsymbol{\beta}_0$.

Proof. By what was said above, **(c1)-(c8)** hold. It suffices to prove that then (2.14) holds as well. Put for φ of (2.13)

$$\Delta = \Delta(\boldsymbol{x},\boldsymbol{\beta}) = \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) \tag{2.15}$$

and

$$Z = Y - \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right).$$

Then the density of Z is

$$g_{\boldsymbol{x}}(z) = f_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)}\left(z + \varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$

and

$$\boldsymbol{m}(\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{\beta}_0) = \int_{\mathcal{X}} \left[w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right] dQ(\boldsymbol{x}) \tag{2.16}$$

for

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) = E\left|Y-\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}\right)\right|$$

$$= E\left|Z+\Delta\left(\boldsymbol{x},\boldsymbol{\beta}\right)\right| \qquad \text{(cf. (2.15))}.$$

The difference

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) = E\left(\left|Z+\Delta\left(\boldsymbol{x},\boldsymbol{\beta}\right)\right| - |Z|\right) \tag{2.17}$$

will be estimated by using the generalized Taylor formula

$$|Z+\Delta| - |Z| = D^+|Z|\,\Delta + \mathcal{R}\left(Z,\Delta\right) \tag{2.18}$$

valid for all real Δ where

$$D^+|z| = \boldsymbol{I}\left(0 \le z < \infty\right) - \boldsymbol{I}\left(-\infty < z < 0\right) \tag{2.19}$$

is the right-hand derivative of the function $|z|$ for $z \in \mathbb{R}$ and $\mathcal{R}\left(z,\Delta\right) = \left(z+\Delta\right)\boldsymbol{I}(-\Delta < z < 0)$ is a remainder in the formula (2.18). This follows from the generalized Taylor expansion of arbitrary convex function established in (2.7) of Liese and Vajda (2003). Since $\text{med}(Z) = 0$, it holds $ED^+|Z| = 0$. Therefore we get from (2.17), (2.18)

$$\begin{aligned} w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) &= E\,\mathcal{R}\left(Z,\Delta\right) \\ &= \int \left(z+\Delta\right)\boldsymbol{I}\left(-\Delta < z < 0\right) g_{\boldsymbol{x}}\left(z\right)\mathrm{d}z \\ &= \int_0^\Delta \left(\Delta - z\right) g_{\boldsymbol{x}}\left(-z\right)\mathrm{d}z \\ &= \int_0^\Delta \left(\Delta - z\right) f_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - z\right)\mathrm{d}z. \end{aligned}$$

Since $\mathcal{X} \subset \mathbb{R}^d$ is bounded, the values

$$p_1 = \inf_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \qquad \text{and} \qquad p_2 = \sup_{\boldsymbol{x}\in\mathcal{X}} \pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)$$

are bounded away from 0 and 1. Thus, taking into account that $\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) = m\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$, we see from **(c6)** that we can find $a > 0$ such that

$$\inf_{|z|\le a}\inf_{\boldsymbol{x}\in\mathcal{X}} f_{\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) - z\right) \ge \Lambda\left(a\right) > 0.$$

This implies that if $0 < b < a$ then for every $\left|\Delta\left(\boldsymbol{x},\boldsymbol{\beta}\right)\right| > b$ it holds

$$w\left(\boldsymbol{x}^T\boldsymbol{\beta}\right) - w\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right) \ge \frac{b^2}{2}\Lambda\left(a\right).$$

Hence, by (2.16), for every $0 < b < a$ we get

$$\boldsymbol{m}\left(\boldsymbol{\beta}\right) - \boldsymbol{m}\left(\boldsymbol{\beta}_0\right) \ge \frac{b^2}{2}\Lambda\left(a\right) Q\left(\mathcal{X}_{b,\boldsymbol{\beta}}\right) \tag{2.20}$$

for the subset of regressors

$$\mathcal{X}_{b,\boldsymbol{\beta}} = \left\{\,\boldsymbol{x} \in \mathcal{X} : \left|\Delta\left(\,\boldsymbol{x},\boldsymbol{\beta}\right)\right| \ge b\right\}.$$

By **(c2),** the smallest eigenvalue $\lambda(\mathbf{\Sigma})$ of the matrix (2.4) is positive. Further, for every $\tau > 0$

$$\begin{aligned} \lambda(\mathbf{\Sigma}) \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 &\le (\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T \mathbf{\Sigma} (\boldsymbol{\beta}-\boldsymbol{\beta}_0) \\ &= \int_{\mathcal{X}} \left(\boldsymbol{x}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right)^2 dQ(\boldsymbol{x}) \\ &\le \|\mathcal{X}\| \,.\, \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) + \tau^2 \end{aligned}$$

where $\|\mathcal{X}\|$ stands for $\max\|\boldsymbol{x}\|$ on $\mathcal{X}$ and $\mathcal{X}^0_{\tau,\boldsymbol{\beta}} = \left\{ \boldsymbol{x} \in \mathcal{X} : \left|\boldsymbol{x}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right| > \tau\right\}$. From here we see that for all $\varepsilon > 0$ and all sufficiently small $\tau > 0$

$$\psi(\tau,\varepsilon) \equiv \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\ge\varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) > 0. \tag{2.21}$$

It is easy to see that $\varphi(t)$ of (2.13) is strictly increasing on $\mathbb{R}$ (for details see p. 11 in Hobza et al (2005)). Therefore the function

$$\phi(\tau) \equiv \inf_{\substack{|t|\le\|\mathcal{X}\|\cdot\|\boldsymbol{\beta}_0\| \\ |s-t|\ge\tau}} |\varphi(s)-\varphi(t)|$$

is positive in the domain $\tau > 0$ and, obviously,

$$\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}} \supseteq \mathcal{X}^0_{\tau,\boldsymbol{\beta}}.$$

Further, $\varphi(t)$ is continuous so that $\phi(\tau) < a$ for all sufficiently small $\tau > 0$. Consequently (2.20) implies for any $\varepsilon > 0$

$$\begin{aligned} \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\ge\varepsilon} [\boldsymbol{m}(\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{\beta}_0)] &\ge \tfrac{\phi(\tau)^2}{2}\Lambda(a) \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\ge\varepsilon} Q\left(\mathcal{X}_{\phi(\tau),\boldsymbol{\beta}}\right) \\ &\ge \tfrac{\phi(\tau)^2}{2}\Lambda(a) \inf_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\ge\varepsilon} Q\left(\mathcal{X}^0_{\tau,\boldsymbol{\beta}}\right) \\ &= \tfrac{\phi(\tau)^2}{2}\Lambda(a)\,\psi(\tau,\varepsilon). \end{aligned}$$

By (2.21), the last product is positive which proves the desired relation (2.14). ■

The function φ of (2.13) is continuously differentiable with the derivative $\varphi'(t) = m'(\pi(t))\pi'(t)$ where $\pi'(t) = \pi(t)(1-\pi(t))$. Let us introduce similar notation as in the proof of Theorem 2.1, namely let for $i = 1, 2, ...$

$$\begin{aligned} \Delta_i(\boldsymbol{\beta}) &= \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right) - \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right), \qquad \boldsymbol{\beta} \in \mathbb{R}^d, \\ Z_i &= Y_i - \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right), \end{aligned}$$

where

$$\widetilde{f}_i(z) = f_{\pi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)}\left(z + \varphi\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\right), \qquad z \in \mathbb{R},$$

is the probability density function of Z_i. The functions $\Delta_i(\boldsymbol{\beta})$ are continuously differentiable on $\mathbb{R}^d$ with gradients

$$grad\,(\Delta_i(\boldsymbol{\beta})) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)\boldsymbol{x}_i.$$

Therefore the linear term $\mathcal{L}_n(h)$ considered in (2.3) of Liese and Vajda (2004) is given here by

$$\mathcal{L}_n(\boldsymbol{h}) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} D^+ |Z_i| \, \varphi'\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right) \boldsymbol{x}_i^T \boldsymbol{h}, \qquad \boldsymbol{h} \in \mathbb{R}^d,$$

where $D^+|z|$ denotes the right-hand derivative (2.19). Since $E\, D^+|Z_i| = 0$, the variance of $\mathcal{L}_n(\boldsymbol{h})$ is $\boldsymbol{h}^T \Sigma_n \boldsymbol{h}$ for the matrix given in accordance with (2.5) of Liese and Vajda (2004) by

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^{n} E\left(D^+|Z_i|\right)^2 \left(\varphi'\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)^2 \boldsymbol{x}_i \boldsymbol{x}_i^T.$$

But $E\left(D^+|Z_i|\right)^2 = 1$ so that we can write the matrix Σ_n in the integral form

$$\Sigma_n = \int_{\mathcal{X}} \left(\varphi'\left(\boldsymbol{x}^T \boldsymbol{\beta}\right)\right)^2 \boldsymbol{x}^T \boldsymbol{x} \, dQ_n(\boldsymbol{x})$$

where Q_n is the empirical measure from **(c1).** Since $\varphi'\left(\boldsymbol{x}^T \boldsymbol{\beta}\right)$ is continuous and bounded on $\mathcal{X}$, it holds

$$\lim_{n\to\infty} \Sigma_n = \Sigma \equiv \int_{\mathcal{X}} \left(\varphi'\left(\boldsymbol{x}^T \boldsymbol{\beta}\right)\right)^2 \boldsymbol{x}^T \boldsymbol{x} \, dQ(\boldsymbol{x}) \tag{2.22}$$

where Q is the limit measure from **(c1)**.

The next step is evaluation of the matrices

$$\mathcal{Q}_n = \frac{1}{n} \sum_{i=1}^{n} g_i(0) \, \nabla\varphi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \left(\nabla\varphi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right)^T$$

where $g_i(t)$ denote derivatives of the functions $G_i(t) = E\mathcal{D}|Z_i + t|$ of variable $t \in \mathbb{R}$ introduced on p. 467 in Liese and Vajda (2003). By the definition of $D^+|z|$ in (2.19), for $\pi_i = \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$ and $\varphi_i = \varphi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$

$$\begin{aligned} G_i(t) &= EI(Z_i + t > 0) - EI(Z_i + t \le 0) \\ &= EI(Y_i > \varphi_i - t) - EI(Y_i \le \varphi_i - t) \\ &= 1 - 2F_{\pi_i}(\varphi_i - t). \end{aligned}$$

Thus $g_i(t) = 2f_{\pi_i}(\varphi_i - t)$ and

$$g_i(0) = 2f_{\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right).$$

Therefore the matrices $\mathcal{Q}_n$ may be represented as the integrals

$$\mathcal{Q}_n = 2 \int_{\mathcal{X}} f_{\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right) \left(\varphi'\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T \boldsymbol{x} \, dQ_n(\boldsymbol{x}).$$

Since $\varphi'\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)$ is continuous and bounded on $\mathcal{X}$ and, by **(c8),** the function

$$f_{\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)}\left(\varphi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right) = f_{\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)}\left(m\left(\pi\left(\boldsymbol{x}^T \boldsymbol{\beta}_0\right)\right)\right)$$

is continuous and bounded on $\mathcal{X}$ too, it holds

$$\lim_{n\to\infty} \mathcal{Q}_n = \mathcal{Q} \equiv 2\int_{\mathcal{X}} f_{\pi(\boldsymbol{x}^T\boldsymbol{\beta}_0)}\left(\varphi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)\left(\varphi'\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)^2 \boldsymbol{x}^T\boldsymbol{x}\,dQ\left(\boldsymbol{x}\right). \tag{2.23}$$

Finally, $D^+\rho\left(Z_i\right) = D^+\left|Z_i\right|$ is in the present situation bounded and $\nabla f_i\left(\boldsymbol{\beta}_0\right) = grad\left(\Delta_i\left(\boldsymbol{\beta}_0\right)\right) = -\varphi'\left(\boldsymbol{x}_i^T\boldsymbol{\beta}_0\right)\boldsymbol{x}_i$ is bounded uniformly for all possible $\boldsymbol{x}_i \in \mathcal{X}$. Consequently the Liapunov condition (2.6) of Liese and Vajda (2004) holds. Similarly, one can verify that the conditions **(C3), (C4)** of Liese and Vajda (2003) as well as (2.39), (2.40) ibid. hold. Thus, by Lemma 3 in Liese and Vajda (2003), **(C5)** and **(C6)** ibid. hold too.

To finalize the evaluation of the matrices Σ and $\mathcal{Q}$ given in (2.22),(2.23), take into account that (2.11) implies in the present situation

$$\varphi\left(t\right) = \begin{cases} \dfrac{3}{2} - \dfrac{e^{-t}}{2} & \text{if} \quad t \geq 0 \\ \dfrac{1}{2} + \dfrac{e^{t}}{2} & \text{if} \quad t < 0\,. \end{cases}$$

Therefore

$$\varphi'\left(t\right) = \frac{e^{-|t|}}{2} \quad \text{if} \quad t \in \mathbb{R}.$$

Further

$$\begin{aligned} f_{\pi(t)}\left(\varphi\left(t\right)\right) = {} & \pi\left(t\right) \vee \left(1 - \pi\left(t\right)\right) \\ = {} & \frac{1}{1+e^{-t}} \vee \frac{1}{1+e^{t}} = \frac{1}{1+e^{-|t|}} = \frac{e^{|t|}}{1+e^{|t|}}. \end{aligned}$$

Consequently,

$$f_{\pi(t)}\left(\varphi\left(t\right)\right)\left(\varphi'\left(t\right)\right)^2 = \frac{e^{-|t|}}{4\left(1+e^{|t|}\right)}$$

and by (2.22), (2.23)

$$\Sigma = \frac{1}{4}\int_{\mathcal{X}} e^{-2\left|\boldsymbol{x}^T\boldsymbol{\beta}_0\right|}\boldsymbol{x}\boldsymbol{x}^T dQ(\boldsymbol{x}) \tag{2.24}$$

and

$$\mathcal{Q} = \frac{1}{2}\int_{\mathcal{X}} \frac{e^{-\left|\boldsymbol{x}^T\boldsymbol{\beta}_0\right|}}{1+e^{\left|\boldsymbol{x}^T\boldsymbol{\beta}_0\right|}}\ \boldsymbol{x}\boldsymbol{x}^T dQ(\boldsymbol{x}). \tag{2.25}$$

Thus we can conclude that if (**c1**), (**c2**) hold then all assumptions of Theorem 1 in Liese and Vajda (2004) are satisfied and the following assertion follows from there.

Theorem 2.2 Let the regressors of the model under consideration satisfy (**c1**), (**c2**). If the limit matrix $\mathcal{Q}$ in (2.25) is positive definite then the Med-estimator $\widehat{\boldsymbol{\beta}}_n$ of the model parameters $\boldsymbol{\beta}_0$ is asymptotically normal in the sense that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(\mathbf{0},\ \mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1}\right) \tag{2.26}$$

for Σ given by (2.24).

Proof. See above. ■

Example 2.1. Most simple is the application of Theorem 2.2. to the univariate logistic regression

$$Y_i \sim Be(\pi(x_i\beta_0)), \quad 1 \leq i \leq n \tag{2.27}$$

with identical regressors $x_1 = x_2 = ... = 1$ and $\beta_0 \in \mathbb{R}$. In this model the Med-estimator $\widehat{\beta}_n \in \mathbb{R}$ is defined by

$$\widehat{\beta}_n = \arg\min_{\beta} \sum_{i=1}^{n} |Z_i - m(\pi(\beta))| \quad \text{for } Z_i = Y_i + U_i \qquad \text{(cf. (1.17)-(1.19))} \tag{2.28}$$

Then **(c1)** and **(c2)** hold for the Dirac measure $Q = \delta_1$ concentrated at the point $x_1 = x_2 = ... = 1$ from the singleton $\mathcal{X} = \{1\} \subset \mathbb{R}$. Therefore we get from (2.24) and (2.25)

$$\Sigma = \frac{1}{4}e^{-2|\beta_0|}, \qquad \mathcal{Q} = \frac{e^{-|\beta_0|}}{2(1+e^{|\beta_0|})}$$

so that $\mathcal{Q}^{-1}\Sigma\mathcal{Q}^{-1} = \left[1+e^{|\beta_0|}\right]^2$ and according (2.26)

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(0, \ \left[1+e^{|\beta_0|}\right]^2\right). \tag{2.29}$$

This asymptotic normality can be verified also directly from the central limit theorem applied to the explicit formula

$$\widehat{\beta}_n = \pi^{-1}\left(m^{-1}\left(Z_{(n/2)}\right)\right) \tag{2.30}$$

easily obtained from the definition (2.28). Here $Z_{(n/2)}$ denotes the median of $Z_1, ..., Z_n$, $m^{-1}(z)$ is the inverse function from (1.22) given also in Figure 1.2 and

$$\pi^{-1}(p) = \ln\frac{p}{1-p}, \qquad 0 < p < 1 \tag{2.31}$$

is the inverse to the logistic function $\pi(t)$ of (1.2). To verify (2.29), take first into account that if $p_0 = \pi(\beta_0)$ is the true Bernoulli parameter then, by p. 490 in Rényi (1970), the central limit theorem for $Z_{(n/2)}$ takes on the form

$$\sqrt{n}\frac{Z_{(n/2)} - m(p_0)}{1/\left[2f_{p_0}(m(p_0))\right]} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1) \tag{2.32}$$

where $f_{p_0}(z)$ is the probability density of Z_i given by (2.6) so that

$$f_{p_0}(m(p_0)) = p_0 \vee (1-p_0). \tag{2.33}$$

We first deduce from here the limit law

$$\sqrt{n}(\widehat{p}_n - p_0) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, [p_0 \vee (1-p_0)]^2\right) \tag{2.34}$$

useful here and also in the next section. The function $\phi(z) = m^{-1}(z)$ is continuously differentiable with

$$\phi'(z) = \frac{1}{m'(\phi(z))} = 2\left[\phi(z) \vee (1-\phi(z))\right]^2 \text{ (cf. (2.12).}$$

Therefore by the Taylor expansion of $\phi(z)$ around $z_0 = m(p_0)$ we get

$$\begin{aligned}\phi\left(Z_{(n/2)}\right) - \phi(m(p_0)) &= \phi'(m(p_0))\left(Z_{(n/2)} - m(p_0)\right) + o_p\left(Z_{(n/2)} - m(p_0)\right) \\ &= 2\left[p_0 \vee (1-p_0)\right]^2 \left(Z_{(n/2)} - m(p_0)\right) + o_p\left(Z_{(n/2)} - m(p_0)\right).\end{aligned}$$

Combining this with (2.32), (2.33) we obtain the desired result (2.34).

Further, by combining the Taylor expansion of $\pi^{-1}(p)$ around $p_0 = \pi(\beta_0)$ with (2.34) and using

$$\frac{d}{dp}\pi^{-1}(p) = \left[p(1-p)\right]^{-1} \tag{2.35}$$

we get

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \left[\frac{p_0 \vee (1-p_0)}{p_0(1-p_0)}\right]^2\right) = N\left(0, \left[p_0 \wedge (1-p_0)\right]^{-2}\right).$$

Equivalently,

$$\sqrt{n}\left(\widehat{\beta}_n - \beta_0\right) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \left[\pi(\beta_0) \wedge (1-\pi(\beta_0))\right]^{-2}\right) \tag{2.36}$$

where

$$\pi(\beta_0) \wedge (1-\pi(\beta_0)) = \frac{1}{e^{-\beta_0}+1} \wedge \frac{1}{e^{\beta_0}+1} = \frac{1}{e^{|\beta_0|}+1}.$$

In other words, the asymptotic normality (2.29) obtained from Theorem 2.2 follows also directly from the explicit formula (2.30) and from the central limit theorem (2.32).

Notice that $\left[\pi(\beta_0) \wedge (1-\pi(\beta_0))\right]^2$ in (2.36) is strictly smaller than the Fisher information

$$\mathcal{J}(\beta_0) = \pi(\beta_0)(1-\pi(\beta_0))$$

in the statistical model $Be(\pi(\beta_0))$. Thus the Med-estimator $\widehat{\beta}_n$ of (2.28) or (2.30) is not asymptotically efficient. Its subefficiency can to some extent be suppressed by the enhancing introduced in the next Section 3. In Section 4 we shall see that this subefficiency is compensated by the positive property of robustness with respect to contaminations of the model $Be(\pi(\beta_0))$.

3 Enhancement

As illustrated on the lines above, application of the L_1-estimators (2.2) in discrete statistical models with observations Y_i, $1 \le i \le n$ statistically smoothed into the continuous form (2.1) is usually accompanied by a loss of efficiency achievable in the original discrete models.

An example of L_1-estimator which is even simpler than the Med-estimator (2.28), (2.30) is the median estimator

$$\widehat{p}_n = \arg\min_p \sum_{i=1}^{n} |Z_i - m(p)| = m^{-1}\left(Z_{(n/2)}\right) \tag{3.1}$$

of the Bernoulli parameter $p_0 \in (0,1)$ based on the smoothed versions $Z_i = Y_i + U_i$ (cf. (1.17)) of the original discrete observations $Y_i \sim Be(p_0)$. In (3.1), $m(p)$ is the median function (1.21), $m^{-1}(z)$ is its inverse (1.22) and $Z_{(n/2)}$ is the median of $Z_1, ..., Z_n$. As well known, the Fisher information in the model $Be(p_0)$ is $\mathcal{J}(p_0) = 1/\left[p_0(1-p_0)\right]$ and the MLE

$$p_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{3.2}$$

of p_0 is asymptotically efficient in the sense

$$\sqrt{n}\,(p_n - p_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, 1/\mathcal{J}(p_0)) = N\left(0, p_0(1-p_0)\right).$$

On the other hand, by (2.34) the Med-estimator $\widehat{p}_n$ is asymptotically normal in the sense

$$\sqrt{n}\,(\widehat{p}_n - p_0) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, [p_0 \vee (1-p_0)]^2) \tag{3.3}$$

where

$$\left[p_0 \vee (1-p_0)\right]^2 \geq p_0(1-p_0).$$

Since this inequality is strict unless $p_0 = 1/2$, the Med-estimator $\widehat{p}_n$ is asymptotically less efficient than the MLE p_n except the special case $p_0 = 1/2$.

The set of statistically smoothed data $Z_i = Y_i + U_i$, $1 \leq i \leq n$ can be expanded by considering for $k > 1$ the matrix of data

$$Z_{ij} = Y_i + U_{ij}, \qquad 1 \leq i \leq n, \qquad 1 \leq j \leq k \tag{3.4}$$

where U_{ij} are mutually and also on $Y_1, ..., Y_n$ independent $U(0,1)$-distributed random variables. If a method of processing the data $Z_1, ..., Z_n$ is statistically optimal in an appropriate sense, like e.g. the MLE

$$\begin{aligned} \widetilde{p}_n(Z_1, ..., Z_n) &= \arg\max_p \prod_{i=1}^{n} f_p(Z_i) \\ &= \arg\max_p \prod_{i=1}^{n} p^{Y_i}(1-p)^{1-Y_i} \end{aligned}$$

coinciding with the classical Bernoulli MLE $p_n = p_n(Y_1, ..., Y_n)$ introduced in (3.2), then its performance cannot be improved by expanding the sufficient statistic $(Z_1, ..., Z_n)$. For example, it is easy to see that the MLE

$$\widetilde{p}_n(Z_{11}, ..., Z_{nn}) = \arg\max_p \prod_{j=1}^{k}\prod_{i=1}^{n} f_p(Z_{ij})$$

"enhanced" by utilizing the expanded data set (3.4) coincides with the previous MLE $\widetilde{p}_n(Z_1,...,Z_n) \equiv p_n$. On the other hand, if the method is suboptimal like e.g. the median estimator $\widehat{p}_n = \widehat{p}_n(Z_1,...,Z_n)$ introduced in (3.1), then its performance can be improved by passing to the expanded data set (3.4).

The following theorem motivates this section. It deals with the *k-enhanced median estimator*

$$\widehat{p}_{n*k} = \arg\min_p \sum_{j=1}^{k}\sum_{i=1}^{n} |Z_{ij} - m(p)| = m^{-1}\left(Z_{(nk/2)}\right) \tag{3.5}$$

of the Bernoulli parameter $p_0 \in (0,1)$ where $m(p)$ is the same as in (3.1), $Z_{11},...,Z_{kn}$ are the smoothed observations (3.4) and $Z_{(nk/2)}$ is the median of all these observations. More precisely, it deals with the variance

$$\sigma^2(\widehat{p}_{n*k}) = E(\widehat{p}_{n*k} - p_0)^2 \tag{3.6}$$

of the estimator $\widehat{p}_{n*k}$ and compares it with the variance

$$\sigma^2(p_n) = E(p_n - p_0)^2 = \frac{p_0(1-p_0)}{n} \tag{3.7}$$

of the classical MLE p_n minimizing the variance in the class of all unbiased estimators (see e. g. Example 31 on p. 322 in Mood et al.(1974)).

Theorem 3.1 The k-enhanced median estimator $\widehat{p}_{n*k}$ is asymptotically optimal in the sense that for each $n \geq 1$

$$\sigma^2(\widehat{p}_{n*k}) \underset{k\to\infty}{\longrightarrow} \frac{p_0(1-p_0)}{n} \qquad (\text{cf } (3.7)). \tag{3.8}$$

Proof. Let $n \geq 1$ be arbitrary fixed and consider for every $0 \leq r \leq n$ the random event

$$\mathcal{E}_r = \left\{\sum_{i=1}^{n} Y_i = r\right\}$$

where $Y_i \sim Be(p_0)$ are the observations assumed in (3.4). If $\mathcal{E}_r$ takes place then $k(n-r)$ observations (3.4) are of the form $Z_{ij} = U_{ij}$ and the remaining kr observations are of the form $Z_{ij} = 1 + U_{ij}$. Therefore under $\mathcal{E}_r$ the random data (3.4) are generated by the distribution function $F_{r/n}(z)$ defined by (1.20) for $p = r/n$ (cf. also Figure 1.1). Hence by the Glivenko theorem (p. 100 in Rényi (1970)), the empirical distribution function G_{kn} of the data (3.4) satisfies under $\mathcal{E}_r$ the relation

$$\sup_z \left|G_{kn}(z) - F_{r/n}(z)\right| \xrightarrow[k\to\infty]{a.s.} 0.$$

Consequently, under $\mathcal{E}_r$ the median of $G_{kn}(z)$ tends a.s. to $m(r/n)$,

$$\text{med } G_{kn}(z) \xrightarrow[k\to\infty]{a.s.} m(r/n),$$

and the bounded sequence of random variables $\widehat{p}_{n*1}, \widehat{p}_{n*2}, ...$satisfies the limit relation

$$\widehat{p}_{n*k} = m^{-1}(\text{med } G_{kn}(z)) \xrightarrow[k\to\infty]{a.s.} \frac{r}{n}. \tag{3.9}$$

The events $\mathcal{E}_r$, and hence the relations (3.9) take place with the binomial probabilities $\binom{n}{r}p_0^r\,(1-p_0)^{n-r}$. From here we deduce for each fixed $k \geq 1$ the equality

$$E\left(\widehat{p}_{n*k}-p_0\right)^2=\sum_{r=0}^{n} E\left(\left(\widehat{p}_{n*k}-p_0\right)^2 \mid \mathcal{E}_r\right)\binom{n}{r}p_0^r\,(1-p_0)^{n-r}$$

where (3.9) implies for each fixed $0 \leq r \leq n$

$$E\left(\left(\widehat{p}_{n*k}-p_0\right)^2 \mid \mathcal{E}_r\right) \underset{k\to\infty}{\longrightarrow} \left(\frac{r}{n}-p_0\right)^2. \tag{3.10}$$

Consequently,

$$E\left(\widehat{p}_{n*k}-p_0\right)^2 \underset{k\to\infty}{\longrightarrow} \sum_{r=0}^{n}\left(\frac{r}{n}-p_0\right)^2\binom{n}{r}p_0^r\,(1-p_0)^{n-r}$$

where the right-hand side is $\sigma^2(p_n)=p_0(1-p_0)/n$ which completes the proof. ■

The method of statistical smoothing and the subsequent median estimation originally introduced for continuous statistical models is studied in Table 3.1, together with the loss of efficiency and its compensation by means of the enhancement procedure introduced above. Presented are the *mean absolute errors*

$$\mathrm{MAE}=\mathrm{MAE}\left(\widetilde{p}_n\right)=\frac{1}{1000}\sum_{l=1}^{1000}\left|\widetilde{p}_n^{(l)}-p_0\right| \tag{3.11}$$

and the *standard deviations* (roots of the *mean squared errors*)

$$\mathrm{STD}=\mathrm{STD}\left(\widetilde{p}_n\right)=\left(\frac{1}{1000}\sum_{l=1}^{1000}\left(\widetilde{p}_n^{(l)}-p_0\right)^2\right)^{1/2} \tag{3.12}$$

obtained from 1000 independent realizations $\widetilde{p}_n^{(l)}$ of the estimators

$$\widetilde{p}_n \in \left\{p_n,\ \widehat{p}_n=\widehat{p}_{n*1},\ \text{and } \widehat{p}_{n*k} \text{ for } k>1\right\}. \tag{3.13}$$

For sample sizes $n \in \{10, 20, 50, 100\}$ and several true values p_0 are printed in bold the minimal errors MAE and STD achieved by the studied estimators $\widetilde{p}_n$.

From (3.7) we see that if the true parameters are $p_0 \in \{0.1,\ 0.2,\ 0.5\}$ then $\sqrt{n}\sigma\,(p_n)=\sqrt{p_0\,(1-p_0)}$ takes on the values $\{0.3,\ 0.4,\ 0.5\}$. In it is easy to verify that the values of $\sqrt{n}\sigma\,(p_n)$ are well approximated by the products $\sqrt{n}$STD obtained by simulations. Table 3.1 also clearly indicates that for all fixed $10 \leq n \leq 100$ the STD's of the k-enhanced suboptimal median estimators $\widehat{p}_{n*k}$ tend for increasing k to the STD's of the optimal MLE p_n as predicted by Theorem 3.1. Finally, this table demonstrates that the level k of enhancing needed to achieve STD($\widehat{p}_{n*k}$) comparable to the optimal STD(p_n) decreases with increasing n and p_0. E.g. if $p_0 = 0.5$ and $n \geq 20$ then the optimal values STD(p_n) are achieved by STD($\widehat{p}_{n*k}$) already for $k = 50$. If $p_0 = 0.2$ then we need $k = 100$.

Remark 3.1 Obviously, the modification

$$E\left(\left|\widehat{p}_{n*k}-p_0\right| \mid \mathcal{E}_r\right) \underset{k\to\infty}{\longrightarrow} \left|\frac{r}{n}-p_0\right|$$

		$n=10$		$n=20$		$n=50$		$n=100$	
p_0	$\widetilde{p}_n$	MAE	STD	MAE	STD	MAE	STD	MAE	STD
0.1	p_n	**0.071**	**0.093**	**0.052**	**0.068**	**0.032**	**0.041**	**0.025**	**0.031**
	$\widehat{p}_n$	0.231	0.344	0.170	0.238	0.101	0.132	0.073	0.092
	$\widehat{p}_{n*5}$	0.123	0.160	0.091	0.114	0.055	0.070	0.039	0.049
	$\widehat{p}_{n*10}$	0.103	0.129	0.073	0.094	0.047	0.059	0.032	0.040
	$\widehat{p}_{n*50}$	0.084	0.102	0.059	0.074	0.036	0.045	0.026	0.033
	$\widehat{p}_{n*100}$	0.079	0.098	0.055	0.069	0.035	0.043	0.026	0.032
0.2	p_n	**0.097**	**0.127**	**0.071**	**0.090**	**0.045**	**0.057**	**0.032**	**0.040**
	$\widehat{p}_n$	0.214	0.326	0.153	0.214	0.091	0.119	0.065	0.082
	$\widehat{p}_{n*5}$	0.133	0.171	0.093	0.118	0.058	0.073	0.039	0.049
	$\widehat{p}_{n*10}$	0.118	0.149	0.082	0.104	0.052	0.066	0.035	0.044
	$\widehat{p}_{n*50}$	0.105	0.132	0.074	0.092	0.047	0.059	0.033	0.041
	$\widehat{p}_{n*100}$	0.104	0.131	0.073	**0.090**	0.046	0.058	**0.032**	**0.040**
0.5	p_n	**0.118**	**0.153**	0.091	**0.116**	**0.056**	**0.070**	**0.039**	**0.050**
	$\widehat{p}_n$	0.169	0.240	0.120	0.168	0.069	0.091	0.045	0.058
	$\widehat{p}_{n*5}$	0.131	0.177	0.095	0.124	0.058	0.074	0.040	0.051
	$\widehat{p}_{n*10}$	0.123	0.162	0.093	0.119	0.057	0.072	0.040	0.051
	$\widehat{p}_{n*50}$	0.119	0.155	**0.090**	**0.116**	**0.056**	**0.070**	0.040	**0.050**
	$\widehat{p}_{n*100}$	0.119	0.155	0.091	**0.116**	**0.056**	**0.070**	**0.039**	**0.050**

Table 3.1: Analysis of the proposed smoothing method in the Bernoulli model $Be(p_0)$. Compared are two estimators $\tilde{p}_n$, namely the MLE p_n and the Med-estimator $\hat{p}_n$ together with its enhanced versions $\widehat{p}_{n*k}$ for selected $k > 1$. The table presents for given sample sizes n the errors defined by (3.11)-(3.12).

of (3.10) holds too. Therefore Theorem 3.1 can be extended in the sense that for each $n \geq 1$ the expected absolute errors

$$\mathsf{e}\left(\widehat{p}_{n*k}\right) = E\left|\widehat{p}_{n*k} - p_0\right|$$

of the enhanced Med-estimators $\widehat{p}_{n*k}$ tend for $k \to \infty$ to the expected error

$$\mathsf{e}\left(p_n\right) = E\left|p_n - p_0\right| = \frac{1}{n}\sum_{r=0}^{n}\left|r - np_0\right|\binom{n}{r}p_0^r\left(1-p_0\right)^{n-r}$$

of the MLE p_n. This convergence is illustrated in Table 3.1 too. Restrict ourselves for simplicity to $n = 10$. We see that then the expected absolute errors $\mathsf{e}\left(p_n\right) \in \{0.070,\ 0.097,\ 0.123\}$ of the MLE p_n corresponding to the errors $p_0 \in \{0.1,\ 0.2,\ 0.5\}$ are well approximated by the bold printed values of MAE(p_n) from the column $n = 10$ obtained by simulations. At the same time the values of MAE($\hat{p}_{n*k}$) from this column seem to tend for increasing k to the bold printed values of MAE(p_n).

Motivated by Theorem 3.1, Remark 3.1 and their experimental verification by the results presented in Table 3.1, we extend Definition 1.2 as follows.

Definition 3.1 For every $k \geq 1$ we define the *k-enhanced median estimator* (briefly, *k-Med estimator*) $\widehat{\boldsymbol{\beta}}_{n*k}$ of the parameters of logistic regression $\boldsymbol{\beta}_0$ by the condition

$$\widehat{\boldsymbol{\beta}}_{n*k} = \arg\min_{\beta} \sum_{j=1}^{k} \sum_{i=1}^{n} \left| Z_{ij} - m\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)\right| \tag{3.14}$$

where Z_{ij} are defined by (3.4) for the same logistic $Y_i \sim Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right)$ as considered in the special case $k = 1$ in Definition 1.2.

Obviously, if $k = 1$ then the estimator introduced by Definition 3.1 coincides with the median estimator of Definition 1.2.

Example 3.1 Let us consider the simple logistic regression model of Example 2.1 with an unknown parameter $\beta_0 \in \mathbb{R}$ and the Med-estimator $\widehat{\beta}_n$ satisfying (2.29). The MLE in this example is

$$\beta_n = \pi^{-1}(p_n) \quad \text{for} \quad p_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad \text{(cf. (3.2))} \tag{3.15}$$

where π^{-1} is the inverse function to $\pi(t)$ of (1.2) and

$$Y_i \sim Be(p_0) \quad \text{for} \quad p_0 = \pi(\beta_0). \tag{3.16}$$

Combining the Taylor expansion of $\pi^{-1}(p)$ around $p_0 = \pi(\beta_0)$ with (3.3) and using (2.35), we get

$$\sqrt{n}\,(\beta_n - \beta_0) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \left[\pi(\beta_0)(1-\pi(\beta_0))\right]^{-1}\right) \quad \text{(cf. (2.36))} \tag{3.17}$$

where the asymptotic variance

$$\left[\pi(\beta_0)(1-\pi(\beta_0))\right]^{-1} = \left(e^{-\beta_0}+1\right)\left(e^{\beta_0}+1\right) \tag{3.18}$$

of the MLE β_n is minimal in the class of all unbiased estimators of β_0. In particular, it is for all $\beta_0 \neq 0$ smaller than the asymptotic variance $\left(e^{|\beta_0|}+1\right)^2$ of the Med-estimator $\widehat{\beta}_n$ found in (2.29).

We shall prove that the k-Med estimators $\widehat{\beta}_{n*k}$ satisfy for every $k \geq 1$ the limit law

$$\sqrt{n}\left(\widehat{\beta}_{n*k} - \beta_0\right)^2 \xrightarrow[n\to\infty]{\mathcal{D}} N(0, s_k^2(\beta_0)) \tag{3.19}$$

where $s_k^2(\beta_0)$ is an asymptotic variance of $\widehat{\beta}_{n*k}$ for fixed k and $n \to \infty$, tending for fixed n and $k \to \infty$ to the minimal asymptotic variance (3.18). Thus in the model of present example the enhancing enables to bring the asymptotic performance of the *Med-estimator* arbitrarily close to the optimal MLE β_n. The relation (3.19) as well as the desired convergence

$$s_k^2(\beta_0) \xrightarrow[k\to\infty]{} \left[\pi(\beta_0)(1-\pi(\beta_0))\right]^{-1} \tag{3.20}$$

of the variances appearing in (3.19) follow from the fact that, similarly as in (2.34), the k-Med estimator $\widehat{p}_{n*k}$ considered in Theorem 3.1 satisfies for each $k > 1$ and $p_0 = \pi(\beta_0)$ the limit law

$$\sqrt{n}\,(\widehat{p}_{n*k} - p_0) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \sigma_k^2(p_0)\right) \tag{3.21}$$

where $\sigma_k^2(p_0) \underset{k\to\infty}{\longrightarrow} p_0\,(1-p_0)$ (cf. Theorem 3.1). Indeed, by Definition 3.1 $\widehat{\beta}_{n*k} = \pi^{-1}\left(\widehat{p}_{n*k}\right)$. Using the Taylor expansion of the function $\pi^{-1}(p)$ around p_0 and employing (2.35), we get both the desired convergences (3.19) and (3.20) from (3.21) by putting

$$s_k^2(\beta_0) = \frac{\sigma_k^2(p_0)}{\left[p_0\,(1-p_0)\right]^2} = \frac{\sigma_k^2(\pi\,(\beta_0))}{\left[\pi\,(\beta_0)\,(1-\pi\,(\beta_0))\right]^2}.$$

A systematic general theory of $k-$enhanced median estimators in the logistic regression exceeds the scope of the present paper. We restrict ourselves to a simulation study which in the next section demonstrates that the enhancement improves the efficiency of median estimators but in some situations deteriorates their robustness.

4 Robustness

The median estimator $\widehat{\boldsymbol{\beta}}_n$ of logistic regression parameters $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ was defined in Section 1.1 by means of the least absolute deviations principle (the L_1-method principle) which is motivated in the classical statistical literature by the robustness of the corresponding statistical procedures (cf. Hampel et al. (1986) or Jurečková and Sen (1996)). Therefore we expected that this estimator will be robust and this robustness was in fact our main motivation for this research. A secondary motivation was to demonstrate that the general theory of consistency and asymptotic normality of M-estimators developed by Liese and Vajda (1999, 2003, 2004) is applicable in concrete particular situations.

The primary purpose of this section is verification of the desired robustness of $\widehat{\boldsymbol{\beta}}_n$. The Med-estimator $\widehat{\boldsymbol{\beta}}_n$ is in this respect compared with several robust estimators known from the previous literature and discussed in Section 1.1, namely with the Morgenthaler's Morg-estimator $\boldsymbol{\beta}_n^{(1)}$, the Bianco and Yohai's BY-estimator $\boldsymbol{\beta}_n^{(2)}$ and the Croux and Haesbroeck's the CH-estimator $\boldsymbol{\beta}_n^{(3)}$. For the sake of completeness, in our comparisons is included also the MLE $\boldsymbol{\beta}_n$. The robustness is compared by means of simulated performances of all these estimators in the logistic models $Be\left(\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$ ε-contaminated at the levels $0 \le \varepsilon \le 0.3$ by the alternative data source $Be\left(1-\pi\left(\boldsymbol{x}^T\boldsymbol{\beta}_0\right)\right)$, or contaminated at the same levels ε by the leverage points from logistic models $Be\left(\pi\left(\tilde{\boldsymbol{x}}^T\boldsymbol{\beta}_0\right)\right)$ with strongly distorted regressors $\tilde{\boldsymbol{x}}$ at the place of $\boldsymbol{x}$. A secondary purpose of this section is to verify the effect of the enhancing on performance of Med-estimator $\widehat{\boldsymbol{\beta}}_n$.

Our simulations are at first generated by a simple model where the correctness of simulations can be verified by theoretical means such as the central limit theorem, and where also the eventual inaccuracies of computational algorithms can be detected and excluded. As the next step, they are generated by a more realistic model used formerly by Bianco and Yohai (1996) for verification of the robustness of their BY-estimator $\boldsymbol{\beta}_n^{(2)}$.

(I) The first model is $Be\,(\pi\,(x\beta_0))$ for a real valued parameter β_0 and regressors x achieving at time $1 \le i \le n$ mutually independent binary random values $x_i \in \mathcal{X} = \{-1, 1\}$ with equal probabilities $\Pr(x_i = -1) = \Pr(x_i = 1) = 1/2$. We use the concrete parameter value $\beta_0 = -\ln 4$ for which $p_0 = \pi\,(\beta_0) = 0.2$. For the levels $\varepsilon \in \{0,\ 0.1,\ 0.2,\ 0,3\}$ and $1 \le i \le n$ we consider the contaminated logistic regression data

$$Y_i \sim (1-\varepsilon)Be\,(\pi\,(x_i\beta_0)) + \varepsilon Be\,(1-\pi\,(x_i\beta_0)) \tag{4.1}$$

i.e. instead of the assumed Bernoulli data $Y_i \sim Be\left(\pi\left(x_i\beta_0\right)\right)$ we generate the distorted Bernoulli data

$$Y_i \sim Be\left(x_i \pi\left(\beta_\varepsilon\right)\right) \text{ for } \beta_\varepsilon = \pi^{-1}\left(\pi\left(\beta_0\right) + \varepsilon\left(1 - 2\pi\left(\beta_0\right)\right)\right) \tag{4.2}$$

(cf. (1.6) and (1.2)). In this simulation model the above mentioned estimators $\boldsymbol{\beta}_n^{(1)} - \boldsymbol{\beta}_n^{(3)}$ and $\widehat{\boldsymbol{\beta}}_n$ take on the univariate forms $\beta_n^{(1)} - \beta_n^{(3)}$ and $\widehat{\beta}_n$. Moreover, their formulas considerably simplify. For derivation of these formulas it is useful to consider the sets

$$A_n^+ = \{1 \le i \le n : x_i = 1\}, \qquad A_n^- = \{1, ..., n\} - A_n^+ \tag{4.3}$$

and the statistics

$$\widetilde{Y}_i = \begin{cases} Y_i & \text{if } i \in A_n^+ \\ 1 - Y_i & \text{if } i \in A_n^- \end{cases} . \tag{4.4}$$

Clearly,

$$\widetilde{Y}_i \sim Be\left(\pi\left(\beta_\varepsilon\right)\right) \quad \text{for all } 1 \le i \le n \text{ and } \beta_\varepsilon \text{ given by (4.2).} \tag{4.5}$$

By (1.10) and (1.11), the Morg-estimatotr $\beta_n^{(1)}$ solves the equation

$$\sum_{i \in A_n^+} \left(Y_i - \pi\left(\beta\right)\right) - \sum_{i \in A_n^-} \left(Y_i - 1 + \pi\left(\beta\right)\right) = 0,$$

i.e. coincides with the MLE β_n specified in the model of (4.5) by

$$\beta_n = \pi^{-1}\left(\widetilde{p}_n\right) \quad \text{for} \quad \widetilde{p}_n = \frac{1}{n}\sum_{i=1}^n \widetilde{Y}_i \tag{4.6}$$

for $\widetilde{Y}_i$ given in (4.5).

As to the BY-estimator $\beta_n^{(2)}$, one can deduce from (1.3), (1.9) and (1.12)-(1.14) that $\beta_n^{(2)} = \pi^{-1}\left(\alpha_n\right)$ for

$$\alpha_n = \arg\min_{\alpha \in (0.1)} L\left(p_n, \alpha\right) \tag{4.7}$$

where

$$L\left(p_n, \alpha\right) = \widetilde{p}_n \rho\left(-\ln \alpha\right) + \left(1 - \widetilde{p}_n\right)\rho\left(-\ln\left(1 - \alpha\right)\right) + \rho_0\left(\alpha\right)$$

for $\widetilde{p}_n$ given by (4.6) and the functions ρ, ρ_0 appearing in (1.12). It is easy to see that

$$\frac{d}{d\alpha} L\left(\widetilde{p}_n, \alpha\right) = \left(\alpha - \widetilde{p}_n\right)\left[\frac{\rho'\left(-\ln \alpha\right)}{\alpha} + \frac{\rho'\left(-\ln\left(1 - \alpha\right)\right)}{1 - \alpha}\right]$$

where the derivative $\rho'(t)$ is by (1.15) nonnegative for $t > 0$ and positive for $0 < t < 1$. Therefore $\alpha_n = \widetilde{p}_n$ is the unique argmin in (4.7) which means that $\beta_n^{(2)}$ coincides with the estimator β_n given in (4.6).

The CH-estimator $\beta_n^{(3)}$ differs from $\beta_n^{(2)}$ by a somewhat modified function $\rho'(t)$ and an argument similar to that given above leads to the conclusion that in the model under consideration $\beta_n^{(3)}$ coincides with the estimator β_n of (4.6) too.

Let us now turn attention to our Med-estimator $\widehat{\beta}_n$ and its k-enhanced versions $\widehat{\beta}_{n*k}$. By Definition 3.1,

$$\widehat{\beta}_n = \pi^{-1}\left(\widehat{p}_n\right) = \ln \frac{\widehat{p}_n}{1-\widehat{p}_n} \tag{4.8}$$

where

$$\begin{aligned} \widehat{p}_n &= \arg\min_p \left(\sum_{i\in A_n^+} |Z_i - m(p)| + \sum_{i\in A_n^-} |Z_i - m(1-p)| \right) \\ &= \arg\min_p \left(\sum_{i\in A_n^+} |Y_i + U_i - m(p)| + \sum_{i\in A_n^-} |Y_i + U_i - 2 + m(p)| \right). \end{aligned}$$

The second equality holds because from (1.19) we deduce that $m(1-p) = 2 - m(p)$. But

$$|Y_i + U_i - 2 + m(p)| = \left|\widetilde{Y}_i + \widetilde{U}_i - m(p)\right|$$

where $\widetilde{Y}_i$ is defined by (4.4) and $\widetilde{U}_i = 1 - U_i$ is the same uniform $U(0,1)$-variable as U_i. Therefore

$$\widehat{p}_n = \arg\min_p \left(\sum_{i=1}^n \left|\widetilde{Z}_i - m(p)\right| \right) = m^{-1}(\widetilde{Z}_{(n/2)}) \tag{4.9}$$

for

$$\widetilde{Z}_i = \widetilde{Y}_i + U_i \quad \text{(cf. (4.5), (1.17))} \quad \text{and} \quad \widetilde{Z}_{(n/2)} = \text{median}\left(\widetilde{Z}_1, ..., \widetilde{Z}_n\right). \tag{4.10}$$

Similarly,

$$\widehat{\beta}_{n*k} = \pi^{-1}\left(\widehat{p}_{n*k}\right) = \ln \frac{\widehat{p}_{n*k}}{1-\widehat{p}_{n*k}} \tag{4.11}$$

where

$$\widehat{p}_{n*k} = \arg\min_p \left(\sum_{j=1}^k \sum_{i=1}^n \left|\widetilde{Z}_{ij} - m(p)\right| \right) = m^{-1}\left(\widetilde{Z}_{nk/2}\right) \tag{4.12}$$

for

$$\widetilde{Z}_{ij} = \widetilde{Y}_i + U_{ij} \quad \text{(cf. (4.5))} \quad \text{and} \quad \widetilde{Z}_{(nk/2)} = \text{median}\left(\widetilde{Z}_{11}, ..., \widetilde{Z}_{kn}\right), \tag{4.13}$$

where U_{ij} are independent and uniformly distributed on $(0,1)$.

In Table 4.1 are evaluated from 1000 realizations $\widetilde{\beta}_n^{(1)}, ..., \widetilde{\beta}_n^{(1000)}$ the *mean absolute errors*

$$\text{MAE} = \frac{1}{1000} \sum_{l=1}^{1000} \left|\widetilde{\beta}_n^{(l)} - \beta_0\right|, \tag{4.14}$$

standard deviations (roots of the *mean squared errors*)

$$\text{STD} = \left(\frac{1}{1000} \sum_{l=1}^{1000} \left(\widetilde{\beta}_n^{(l)} - \beta_0\right)^2 \right)^{1/2} \tag{4.15}$$

		$n=50$			$n=100$			$n=500$			$n=1000$		
ε	$\widetilde{\beta}_n$	MAE	STD	RR%	MAE	STD	RR%	MAE	STD	RR%	MAE	STD	RR%
0	β_n	**0.273**	**0.350**	0	**0.195**	**0.248**	0	**0.088**	**0.110**	0	**0.065**	**0.081**	0
	$\widehat{\beta}_n$	0.579	0.854	7	0.436	0.617	0	0.180	0.246	0	0.121	0.153	0
	$\widehat{\beta}_{n*5}$	0.371	0.504	1	0.256	0.338	0	0.114	0.142	0	0.082	0.104	0
	$\widehat{\beta}_{n*10}$	0.324	0.424	0	0.233	0.297	0	0.102	0.128	0	0.073	0.091	0
	$\widehat{\beta}_{n*50}$	0.291	0.376	0	0.201	0.256	0	0.091	0.114	0	0.067	0.083	0
0.1	β_n	**0.393**	**0.463**	0	**0.358**	**0.411**	0	0.343	**0.357**	0	0.339	**0.346**	0
	$\widehat{\beta}_n$	0.563	0.708	3	0.448	0.617	0	0.346	0.380	0	**0.335**	0.354	0
	$\widehat{\beta}_{n*5}$	0.432	0.509	0	0.373	0.432	0	**0.338**	**0.357**	0	0.338	0.348	0
	$\widehat{\beta}_{n*10}$	0.417	0.498	0	0.365	0.418	0	0.341	**0.357**	0	0.337	**0.346**	0
	$\widehat{\beta}_{n*50}$	0.399	0.467	0	0.360	0.413	0	0.342	**0.357**	0	0.339	**0.346**	0
0.2	β_n	0.632	**0.695**	0	0.617	**0.654**	0	0.631	**0.638**	0	0.630	0.634	0
	$\widehat{\beta}_n$	0.685	0.783	1	0.624	0.685	0	**0.623**	0.640	0	**0.626**	**0.633**	0
	$\widehat{\beta}_{n*5}$	0.631	0.702	0	**0.615**	0.659	0	0.630	0.639	0	0.629	0.634	0
	$\widehat{\beta}_{n*10}$	**0.629**	0.700	0	0.616	0.657	0	0.632	0.639	0	0.629	**0.633**	0
	$\widehat{\beta}_{n*50}$	0.633	0.696	0	0.619	0.656	0	0.631	**0.638**	0	0.630	0.634	0
0.3	β_n	0.890	0.941	0	0.895	0.920	0	0.894	0.899	0	0.896	0.899	0
	$\widehat{\beta}_n$	**0.874**	0.950	0	**0.879**	0.920	0	**0.885**	**0.894**	0	**0.894**	**0.897**	0
	$\widehat{\beta}_{n*5}$	0.883	**0.940**	0	0.892	0.920	0	0.893	0.899	0	0.896	0.898	0
	$\widehat{\beta}_{n*10}$	0.887	**0.940**	0	0.893	**0.919**	0	0.892	0.897	0	0.896	0.899	0
	$\widehat{\beta}_{n*50}$	0.889	**0.940**	0	0.895	0.920	0	0.893	0.898	0	0.896	0.899	0

Table 4.1: Mean absolute errors (4.14), standard deviations (4.15) and rejection rates RR for selected estimators $\widetilde{\beta}_n$ of the true parameter $\beta_0 = -\ln 4$ of the simple logistic regression model $Be(\pi(x_i\beta_0))$ for random independent x_i with $Pr(x_i = 1) = Pr(x_i = -1) = 1/2$. Compared is the common value β_n of the Morg-, CH- and BY-estimators with the median estimators $\widehat{\beta}_n$ and their k-enhancements $\widehat{\beta}_{n*k}$. The errors are evaluated for the contaminated data $Y_i \sim (1-\varepsilon)Be(\pi(x_i\beta_0)) + \varepsilon Be(1-\pi(x_i\beta_0))$.

		$n = 50$			$n = 100$			$n = 500$			$n = 1000$		
ε	$\widetilde{\beta}_n$	MAE	STD	RR%	MAE	STD	RR%	MAE	STD	RR%	MAE	STD	RR%
0	β_n	**0.273**	**0.350**	0	**0.195**	**0.248**	0	**0.088**	**0.110**	0	**0.065**	**0.081**	0
	$\widehat{\beta}_n$	0.579	0.854	7	0.436	0.617	2	0.180	0.246	0	0.121	0.153	0
	$\widehat{\beta}_{n*5}$	0.371	0.504	1	0.256	0.338	0	0.114	0.142	0	0.082	0.104	0
	$\widehat{\beta}_{n*10}$	0.324	0.424	0	0.233	0.297	0	0.102	0.128	0	0.073	0.091	0
	$\widehat{\beta}_{n*50}$	0.291	0.376	0	0.201	0.256	0	0.091	0.114	0	0.067	0.083	0
0.1	β_n	**0.454**	**0.520**	0	0.441	**0.478**	0	0.441	**0.449**	0	0.443	**0.447**	0
	$\widehat{\beta}_n$	0.575	0.718	1	0.474	0.541	0	**0.425**	0.451	0	**0.436**	0.449	0
	$\widehat{\beta}_{n*5}$	0.470	0.539	0	0.441	0.489	0	0.439	0.450	0	0.443	0.449	0
	$\widehat{\beta}_{n*10}$	0.461	0.529	0	**0.440**	0.484	0	0.442	0.452	0	0.442	**0.447**	0
	$\widehat{\beta}_{n*50}$	0.457	0.522	0	0.443	0.481	0	0.441	**0.449**	0	0.443	**0.447**	0
0.2	β_n	0.791	0.820	0	0.809	0.823	0	0.811	0.815	0	0.811	0.813	0
	$\widehat{\beta}_n$	**0.774**	0.835	0	**0.783**	**0.821**	0	**0.806**	**0.813**	0	**0.808**	**0.811**	0
	$\widehat{\beta}_{n*5}$	0.779	**0.818**	0	0.803	0.822	0	0.810	0.814	0	0.811	0.813	0
	$\widehat{\beta}_{n*10}$	0.791	0.824	0	0.806	0.823	0	0.811	0.814	0	0.811	0.813	0
	$\widehat{\beta}_{n*50}$	0.791	0.821	0	0.809	0.824	0	0.812	0.816	0	0.811	0.813	0
0.3	β_n	1.135	1.151	0	1.141	1.149	0	1.143	1.145	0	1.143	1.144	0
	$\widehat{\beta}_n$	**1.092**	**1.126**	0	**1.119**	**1.134**	0	**1.138**	**1.141**	0	**1.140**	**1.142**	0
	$\widehat{\beta}_{n*5}$	1.127	1.147	0	1.134	1.144	0	1.142	1.144	0	1.143	1.143	0
	$\widehat{\beta}_{n*10}$	1.129	1.147	0	1.140	1.148	0	1.142	1.144	0	1.143	1.144	0
	$\widehat{\beta}_{n*50}$	1.135	1.151	0	1.141	1.149	0	1.143	1.145	0	1.143	1.144	0

Table 4.2: The same simulations as in Table 4.1 evaluated for $n(1-\varepsilon)$ standard logistic regression data $Y_i \sim Be(\pi(x_i\beta_0))$ for $1 \leq i \leq n(1-\varepsilon)$ and $n\varepsilon$ leverage points $Y_i \sim Be(\pi(-10x_i\beta_0))$ for $n(1-\varepsilon)+1 \leq i \leq n$.

and also the *rejection rates* RR, i.e. the percentages of those data vectors $\widetilde{\boldsymbol{Z}}_n = \left(\widetilde{Z}_1, ..., \widetilde{Z}_n\right)$ or data matrices $\widetilde{\boldsymbol{Z}}_{k,n} = \left(\widetilde{Z}_{11}, ..., \widetilde{Z}_{kn}\right)$ for which the corresponding estimator $\widehat{\beta}_n$ or $\widehat{\beta}_{n*k}$ was undefined because the median $\widetilde{Z}_{(n/2)}$ or $\widetilde{Z}_{(nk/2)}$ was outside of the definition domain $(1/2, 3/2)$ of the inverse median function $m^{-1}(z)$ (cf. Figure 1.2).

From the first (noncontamined) sector of Table 4.1 one can verify that $\sqrt{n}\mathrm{STD}(\beta_n)$ agrees very well already for not too large n with the asymptotic standard deviation

$$\sigma\left(\mathrm{MLE}\right) = \frac{1}{\sqrt{\pi\left(\beta_0\right)\left(1-\pi\left(\beta_0\right)\right)}} = \frac{1}{\sqrt{0.2 \times 0.8}} = 2.5$$

obtained from the limit theorem of Example 2.1. E.g., $\sqrt{50}\mathrm{STD}(\beta_{50}) = 2.47$, or $\sqrt{100}\cdot \mathrm{STD}(\beta_{100}) = 2.48$. Similarly one can verify that for the Med-estimator the scaled standard deviation $\sqrt{n}\mathrm{STD}\left(\widehat{\beta}_n\right)$ tends with increasing n to the corresponding theoretic value

$$\sigma\left(\mathrm{Med}\right) = \frac{1}{\pi\left(\beta_0\right)} = \frac{1}{0.2} = 5,$$

e.g. $\sqrt{100}\mathrm{STD}\left(\hat{\beta}_{100}\right) = 6.17$ and $\sqrt{500}\mathrm{STD}\left(\hat{\beta}_{500}\right) = 5.5$.

Similarly, from the ε-contaminated sector of Table 4.1 one can verify that $\sqrt{n}\mathrm{STD}(\beta_n)$ and $\sqrt{n}\mathrm{STD}\left(\widehat{\beta}_n\right)$ approximate the theoretical standard deviations

$$\sigma_n\,(\mathrm{MLE}) = \sqrt{n\,(\beta_\varepsilon - \beta_0)^2 + 1/[\pi\,(\beta_\varepsilon)\,(1 - \pi\,(\beta_\varepsilon))]}$$

and

$$\sigma_n\,(\mathrm{Med}) = \sqrt{n\,(\beta_\varepsilon - \beta_0)^2 + (e^{|\beta_\varepsilon|} + 1)^2}\,.$$

E.g., for $\varepsilon = 0.2$ we get $\beta_\varepsilon = -0.7538$ and $(\beta_\varepsilon - \beta_0)^2 = 0.4000$ so that we compute $\sigma_n\,(\mathrm{MLE}) = 6.68$ while in the 0.2-contaminated sector of Table 4.1 we find $\sqrt{100}\mathrm{STD}(\beta_{100}) = 6.54$. Similarly we compute $\sigma_n\,(\mathrm{Med}) = 7.05$ while in the table we find $\sqrt{100}\mathrm{STD}\left(\hat{\beta}_{100}\right) = 6.85$ which is a satisfactory agreement.

All sectors of Table 4.1 also clearly indicate that for fixed n the error measures $\mathrm{STD}\left(\widehat{\beta}_{n*k}\right)$ and $\mathrm{MAE}\left(\widehat{\beta}_{n*k}\right)$ tend with increasing k to the respective values $\mathrm{STD}(\beta_n)$ and $\mathrm{MAE}(\beta_n)$, in accordance with what is asserted by Theorem 3.1.

However the main message of Table 4.1 is that for larger sample sizes n our Med-estimators $\widehat{\boldsymbol{\beta}}_n$ better resist to higher levels of contamination than the remaining four estimators known from the previous literature.

Table 4.2 in some sense even more strongly supports the observations and conclusions drawn from Table 4.1 above. In this table the ε-fractions of the data Y_i were generated by the false regressors $\tilde{x}_i = -10x_i$ so that they represent leverage points in the common sense of the regression analysis.

(II) Let us now turn to the second model and the corresponding results in Tables 4.3 - 4.10. These tables present similar characteristics as Tables 4.1, 4.2, just being based on simulations in more realistic logistic regression models with bivariate parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)^T \in \mathbb{R}^2$ and bivariate regressors $\boldsymbol{x} = (x_0, x_1)^T \in \mathbb{R}^2$. This means among other that in Tables 4.3 - 4.10 the MAE formula (4.14) is replaced by

$$\mathrm{MAE} = \frac{1}{2000}\sum_{l=1}^{1000}\left(\left|\widetilde{\beta}_{n0}^{(l)} - \beta_{00}\right| + \left|\widetilde{\beta}_{n1}^{(l)} - \beta_{01}\right|\right) \tag{4.16}$$

obtained from 1000 simulated realizations $\widetilde{\boldsymbol{\beta}}_n^{(l)} = \left(\widetilde{\beta}_{n0}^{(l)}, \widetilde{\beta}_{n1}^{(l)}\right)^T$ of an estimator $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of true parameters $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T$.

In order to achieve an ideal comparability, all selected estimators $\widetilde{\boldsymbol{\beta}}_n$ are evaluated for the same simulated data vectors $\boldsymbol{Y}_n = (Y_1, \ldots, Y_n)$ and the related smoothed data vectors or matrices

$$\boldsymbol{Z}_n = (Y_1 + U_1, \ldots, Y_n + U_n) \quad \text{or} \quad \boldsymbol{Z}_{k,n} = (Y_1 + U_{11}, \ldots, Y_n + U_{kn}) \quad \text{cf. } ((1.17), (3.4)).$$

If numerical evaluation of one of these estimators for a given $\boldsymbol{Y}_n$, $\boldsymbol{Z}_n$ or $\boldsymbol{Z}_{k,n}$ fails then this $\boldsymbol{Y}_n$ is rejected and replaced by a new independent realization. This procedure is repeated until the computer successfully numerically evaluates 1000 realizations $\widetilde{\boldsymbol{\beta}}_n^{(1)}, \ldots, \widetilde{\boldsymbol{\beta}}_n^{(1000)}$ for each of the selected estimators $\widetilde{\boldsymbol{\beta}}_n$. The *rejection rate* RR then specifies for each such estimator the percentage of the data vectors $\boldsymbol{Y}_n$ rejected during evaluation of the corresponding 1000 realizations.

Let us explain what is meant by the numerical evaluation of estimates $\widetilde{\boldsymbol{\beta}}_n^{(l)} = (\widetilde{\beta}_{n0}^{(l)}, \widetilde{\beta}_{n1}^{(l)})^T$ used in the formulas (4.16) and leading to the characteristics appearing in the columns and rows of Tables 4.3 - 4.10. This is the evaluation in accordance with the corresponding definition given above, using the iteration procedures presented in the *IMSL C Numerical Libraries*, version 3.0. The minimization of a function of two variables uses there a quasi-Newton method (for details see Appendix A of Dennis and Schnabel (1983)), and systems of equations is solved using a modified Powell hybrid algorithm (for further description see Moré et al. (1980)). The initial iteration seeds for the MLE $\boldsymbol{\beta}_n = (\beta_{n0}, \beta_{n1})^T$ were the true parameters $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T$ and the initial iteration seeds for all the remaining estimates $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ were the MLE's $\boldsymbol{\beta}_n = (\beta_{n0}, \beta_{n1})^T$.

The results of the first four Tables 4.3 - 4.6 were obtained for the simulated data $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$ with

$$Y_i \sim (1-\varepsilon)\, Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) \quad \text{(cf. (4.1))} \tag{4.17}$$

used previously by Bianco and Yohai (1996) to demonstrate the robustness of their BY-estimator denoted above as $\boldsymbol{\beta}_n^{(2)}$. To be precise, the data Y_i were generated using concrete random regressors

$$\boldsymbol{x}_i = (x_{i0} \equiv 1,\ x_{i1} \sim N(0,1))^T = (1, N(0,1)_i)^T \tag{4.18}$$

and the true parameters $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T$ from the following concrete set

$$\left\{(-2.82,\ 2.82)^T;\ (-2.16,\ 3.71)^T;\ (-1.16,\ 4.20)^T;\ (0.00,\ 4.36)^T\right\} \tag{4.19}$$

leading to the corresponding probabilities

$$\Pr(Y_i = 1) \equiv E\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \in \{0.2;\ 0.3;\ 0.4;\ 0.5\}, \tag{4.20}$$

exactly as in the mentioned paper of Bianco and Yohai.

The results of the last four Tables 4.7 - 4.10 were obtained for the simulated data $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)$ with the contaminated data source (4.17) replaced by

$$Y_i \sim (1-\varepsilon)\, Be\left(\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right) + \varepsilon Be\left(\pi\left(\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\beta}_0\right)\right) \tag{4.21}$$

with the random regressors $\boldsymbol{x}_i$ and true parameters $\boldsymbol{\beta}_0$ given by the same formulas (4.18) and (4.19) as in the previous tables, but with the random regressors $\widetilde{\boldsymbol{x}}_i$ different, given by

$$\widetilde{\boldsymbol{x}}_i = \left(1, \widetilde{x}_{i1} = \beta_{00} + 4 sign\left[-\frac{\beta_{00}}{\beta_{01}} - x_{i1}\right]\beta_{01}\right), \quad x_{i1} \sim N(0,1) \quad \text{(cf. (4.18))}. \tag{4.22}$$

We see that in (4.21) the source $Be\left(1 - \pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right)\right)$ of contaminating data is replaced by the source $Be\left(\pi\left(\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\beta}_0\right)\right)$ of leverage points where the regressors $\widetilde{\boldsymbol{x}}_i = \widetilde{\boldsymbol{x}}_i(\boldsymbol{\beta}_0)$ given by (4.22) are characterized by the property

$$\pi\left(\boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) > 1/2 \quad \text{implies} \quad \pi\left(\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\beta}_0\right) \approx 0$$

and

$$\pi\left(\, \boldsymbol{x}_i^T \boldsymbol{\beta}_0\right) \le 1/2 \quad \text{implies} \quad \pi\left(\widetilde{\boldsymbol{x}}_i^T \boldsymbol{\beta}_0\right) \approx 1.$$

We see that the results in Tables 4.3 - 4.10 basically agree with those in Tables 4.1, 4.2. An obvious difference between these two sets of tables is that in the present regression models (4.17) - (4.20) the MLE $\boldsymbol{\beta}_n$, the Morg-estimator $\boldsymbol{\beta}_n^{(1)}$, the BY-estimator $\boldsymbol{\beta}_n^{(2)}$ and the CH-estimator $\boldsymbol{\beta}_n^{(3)}$ mutually differ. Therefore the Tables 4.3 - 4.10 compare our Med-estimator $\widehat{\boldsymbol{\beta}}_n$ and its k-enhanced versions $\widehat{\boldsymbol{\beta}}_{n*k}$ with four different estimators. This means that the reading and interpretation of these tables is more complicated than in case of Tables 4.1, 4.2.

There is however one additional difference between the present Tables 4.3 - 4.10 and the previous Tables 4.1, 4.2 which is visible at the first sight. Namely, in the present tables the MAE of $\widehat{\boldsymbol{\beta}}_{n*k}$ does not seem to converge for increasing k to the MAE of the MLE $\boldsymbol{\beta}_n$. But the MAE($\widehat{\boldsymbol{\beta}}_{n*k}$) still preserves another properties observed in Tables 4.1, 4.2: For $\varepsilon = 0$ it monotonically decreases when $50 \le n \le 1000$ remains fixed and k increases. For $\varepsilon > 0$ the MAE($\widehat{\boldsymbol{\beta}}_{n*k}$) with increasing k first decreases and then increases, and this phenomenon is the more evident the larger is the sample size n. Another feature shared with Tables 4.1, 4.2 is that the conclusions drawn from the classical contaminated models in Tables 4.3 - 4.6 are even more evidently supported by Tables 4.7 - 4.10 obtained from the models contaminated by leverage points.

But the main message in Tables 4.3 - 4.10 remains the same as before: Our Med-estimator $\widehat{\boldsymbol{\beta}}_n$ is more resistent to the distortions of logistic regression models by contaminations and leverage points than the remaining four estimators when the levels of distortions are higher and the sample sizes are medium (around $n \approx 100$) or large (around $n \approx 1000$). For the medium sample sizes it is convenient to use the k-enhancing for $k \approx 10$. The preferences between the remaining estimators $\boldsymbol{\beta}_n$ and $\boldsymbol{\beta}_n^{(1)}$- $\boldsymbol{\beta}_n^{(3)}$ which can be drawn from Tables 4.3 - 4.10 agree with what has been previously published in the literature, in particular with what was stated in Bianco and Yohai (1996) and Croux and Haesbroeck (2003).

		n = 50		n = 100		n = 500		n = 1000	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.862	1	**0.579**	0	**0.235**	0	**0.159**	0
	Morg	0.951	3	0.639	0	0.256	0	0.176	0
	CH	**0.840**	1	0.582	0	0.247	0	0.171	0
	BY	1.527	8	0.942	1	0.312	0	0.212	0
	Med	3.267	18	2.554	9	0.925	0	0.497	0
	5-Med	2.786	22	1.637	10	0.521	0	0.327	0
	10-Med	2.496	24	1.730	10	0.464	0	0.306	0
	50-Med	2.502	28	1.579	13	0.420	0	0.284	0
0.05	MLE	1.067	0	0.979	0	1.047	0	1.048	0
	Morg	1.062	1	**0.834**	0	0.769	0	0.771	0
	CH	**0.984**	0	0.850	0	0.843	0	0.843	0
	BY	1.482	5	0.915	1	0.553	0	0.527	0
	Med	2.715	13	2.524	8	0.848	0	0.587	0
	5-Med	2.093	18	1.544	8	0.579	0	0.487	0
	10-Med	2.098	19	1.444	9	0.547	0	0.475	0
	50-Med	2.403	26	1.475	13	**0.526**	0	**0.470**	0
0.1	MLE	1.425	0	1.473	0	1.510	0	1.525	0
	Morg	1.347	0	1.290	0	1.320	0	1.338	0
	CH	**1.334**	0	1.333	0	1.357	0	1.373	0
	BY	1.446	3	**1.220**	0	1.072	0	1.100	0
	Med	2.701	11	2.293	4	0.977	0	**0.921**	0
	5-Med	1.907	14	1.589	5	0.920	0	0.931	0
	10-Med	1.819	16	1.569	6	0.909	0	0.934	0
	50-Med	2.189	22	1.623	10	**0.908**	0	0.939	0
0.2	MLE	1.989	0	2.011	0	2.029	0	2.034	0
	Morg	1.939	0	1.956	0	1.975	0	1.979	0
	CH	1.945	0	1.957	0	1.973	0	1.977	0
	BY	**1.897**	0	1.891	0	1.941	0	1.948	0
	Med	2.381	5	1.988	2	**1.683**	0	**1.710**	0
	5-Med	2.029	5	1.879	2	1.721	0	1.732	0
	10-Med	2.042	8	**1.735**	3	1.725	0	1.734	0
	50-Med	2.194	10	1.808	4	1.731	0	1.739	0
0.3	MLE	2.335	0	2.336	0	2.354	0	2.356	0
	Morg	2.323	0	2.323	0	2.342	0	2.344	0
	CH	2.321	0	2.320	0	2.338	0	2.340	0
	BY	2.306	0	2.316	0	2.339	0	2.341	0
	Med	2.464	3	2.271	1	**2.231**	0	**2.245**	0
	5-Med	2.269	2	2.239	0	2.245	0	2.252	0
	10-Med	2.289	3	2.225	0	2.248	0	2.253	0
	50-Med	**2.261**	4	**2.207**	1	2.249	0	2.253	0

Table 4.3: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-2.82, 2.82)^T$ in the ε-contaminated logistic regression model (4.17) with $Pr(Y = 1) = 0.2$.

		$n=50$		$n=100$		$n=500$		$n=1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.948	1	**0.574**	0	**0.242**	0	**0.170**	0
	Morg	0.993	3	0.649	0	0.269	0	0.187	0
	CH	**0.900**	1	0.596	0	0.259	0	0.181	0
	BY	1.358	7	0.944	1	0.334	0	0.222	0
	Med	3.125	17	2.607	9	1.079	0	0.542	0
	5-Med	2.482	22	1.660	11	0.567	0	0.345	0
	10-Med	2.381	24	1.647	11	0.499	0	0.322	0
	50-Med	2.433	27	1.569	12	0.450	0	0.293	0
0.05	MLE	1.128	0	1.079	0	1.127	0	1.136	0
	Morg	1.114	1	**0.842**	0	0.774	0	0.779	0
	CH	**1.040**	0	0.891	0	0.870	0	0.876	0
	BY	1.447	5	0.951	1	0.547	0	0.521	0
	Med	2.899	14	2.282	7	0.854	0	0.605	0
	5-Med	2.238	18	1.571	8	0.565	0	0.483	0
	10-Med	2.033	20	1.508	8	0.537	0	0.476	0
	50-Med	2.350	27	1.338	13	**0.518**	0	**0.469**	0
0.1	MLE	1.556	0	1.584	0	1.634	0	1.647	0
	Morg	1.447	0	1.349	0	1.382	0	1.400	0
	CH	**1.440**	0	1.407	0	1.439	0	1.454	0
	BY	1.533	3	**1.170**	0	1.080	0	1.111	0
	Med	2.569	12	2.182	5	1.035	0	**0.928**	0
	5-Med	2.210	15	1.606	6	0.948	0	0.942	0
	10-Med	1.960	18	1.541	7	0.941	0	0.947	0
	50-Med	2.050	21	1.627	10	**0.935**	0	0.958	0
0.2	MLE	2.129	0	2.129	0	2.170	0	2.167	0
	Morg	2.059	0	2.053	0	2.097	0	2.093	0
	CH	2.070	0	2.059	0	2.097	0	2.093	0
	BY	**2.012**	1	1.954	0	2.040	0	2.040	0
	Med	2.314	6	2.041	3	**1.729**	0	**1.786**	0
	5-Med	2.097	7	1.863	2	1.782	0	1.803	0
	10-Med	2.109	10	**1.817**	3	1.793	0	1.805	0
	50-Med	2.207	13	1.860	4	1.798	0	1.806	0
0.3	MLE	2.465	0	2.467	0	2.484	0	2.486	0
	Morg	2.449	0	2.448	0	2.467	0	2.470	0
	CH	2.448	0	2.445	0	2.463	0	2.465	0
	BY	2.432	0	2.436	0	2.463	0	2.466	0
	Med	2.481	3	2.345	0	**2.337**	0	**2.352**	0
	5-Med	**2.370**	3	**2.291**	0	2.354	0	2.360	0
	10-Med	2.428	4	2.293	0	2.354	0	2.359	0
	50-Med	2.484	5	2.304	0	2.355	0	2.360	0

Table 4.4: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-2.16, 3.71)^T$ in the ε-contaminated logistic regression model (4.17) with $Pr(Y=1) = 0.3$.

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.865	1	**0.560**	0	**0.235**	0	**0.167**	0
	Morg	0.921	2	0.603	0	0.254	0	0.184	0
	CH	**0.845**	1	0.567	0	0.247	0	0.178	0
	BY	1.289	8	0.842	2	0.304	0	0.213	0
	Med	2.724	19	2.338	10	0.891	0	0.518	0
	5-Med	2.228	22	1.651	10	0.509	0	0.320	0
	10-Med	2.164	24	1.559	12	0.477	0	0.297	0
	50-Med	2.171	28	1.377	14	0.429	0	0.276	0
0.05	MLE	1.051	0	1.003	0	1.045	0	1.057	0
	Morg	1.071	1	0.817	0	0.715	0	0.720	0
	CH	**0.970**	0	0.856	0	0.805	0	0.811	0
	BY	1.382	4	**0.813**	0	0.513	0	0.488	0
	Med	2.450	15	1.969	6	0.781	0	0.538	0
	5-Med	1.991	18	1.346	8	0.542	0	0.466	0
	10-Med	1.999	21	1.211	7	0.516	0	0.455	0
	50-Med	1.995	28	1.223	13	**0.492**	0	**0.442**	0
0.1	MLE	1.429	0	1.459	0	1.509	0	1.513	0
	Morg	**1.321**	1	1.255	0	1.271	0	1.276	0
	CH	1.329	0	1.305	0	1.325	0	1.329	0
	BY	1.340	3	**1.083**	0	0.996	0	1.012	0
	Med	2.120	12	1.680	6	0.966	0	**0.863**	0
	5-Med	1.857	15	1.342	6	0.841	0	0.873	0
	10-Med	1.737	18	1.230	7	0.843	0	0.870	0
	50-Med	1.780	24	1.160	11	**0.839**	0	0.874	0
0.2	MLE	1.968	0	1.951	0	1.994	0	1.994	0
	Morg	1.904	0	1.880	0	1.920	0	1.922	0
	CH	1.910	0	1.884	0	1.921	0	1.922	0
	BY	**1.849**	1	1.788	0	1.861	0	1.869	0
	Med	2.190	6	1.925	2	**1.572**	0	**1.623**	0
	5-Med	2.009	8	1.773	3	1.622	0	1.643	0
	10-Med	1.889	9	1.769	3	1.621	0	1.647	0
	50-Med	1.895	14	**1.639**	3	1.630	0	1.648	0
0.3	MLE	2.248	0	2.273	0	2.276	0	2.277	0
	Morg	2.228	0	2.256	0	2.259	0	2.261	0
	CH	2.228	0	2.253	0	2.255	0	2.256	0
	BY	2.217	0	2.246	0	2.255	0	2.257	0
	Med	2.279	5	**2.098**	1	**2.135**	0	**2.144**	0
	5-Med	**2.139**	4	2.099	1	2.146	0	2.153	0
	10-Med	2.172	4	2.100	0	2.149	0	2.153	0
	50-Med	2.182	7	2.119	1	2.149	0	2.154	0

Table 4.5: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-1.16, 4.20)^T$ in the ε-contaminated logistic regression model (4.17) with $Pr(Y = 1) = 0.4$.

		$n=50$		$n=100$		$n=500$		$n=1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.906	1	**0.563**	0	**0.231**	0	**0.160**	0
	Morg	0.939	2	0.616	0	0.256	0	0.176	0
	CH	**0.873**	1	0.584	0	0.249	0	0.170	0
	BY	1.210	8	0.821	1	0.303	0	0.207	0
	Med	2.735	20	2.095	10	0.863	0	0.457	0
	5-Med	1.982	23	1.472	10	0.497	0	0.314	0
	10-Med	1.888	25	1.325	12	0.459	0	0.290	0
	50-Med	1.845	31	1.308	12	0.419	0	0.275	0
0.05	MLE	0.981	0	0.919	0	0.901	0	0.898	0
	Morg	1.003	1	**0.761**	0	0.636	0	0.630	0
	CH	**0.921**	0	0.781	0	0.707	0	0.702	0
	BY	1.156	5	0.803	1	0.472	0	0.445	0
	Med	2.224	15	1.891	8	0.715	0	0.505	0
	5-Med	1.835	19	1.427	8	0.515	0	0.428	0
	10-Med	1.725	21	1.419	9	0.482	0	0.421	0
	50-Med	1.811	27	1.179	13	**0.460**	0	**0.417**	0
0.1	MLE	1.302	0	1.284	0	1.270	0	1.265	0
	Morg	1.227	0	1.108	0	1.075	0	1.074	0
	CH	**1.218**	0	1.152	0	1.119	0	1.115	0
	BY	1.321	2	**1.026**	0	0.855	0	0.860	0
	Med	2.179	11	1.686	6	0.877	0	0.762	0
	5-Med	1.863	14	1.266	6	0.769	0	**0.747**	0
	10-Med	1.798	16	1.175	7	**0.744**	0	0.752	0
	50-Med	1.925	23	1.161	12	0.748	0	0.753	0
0.2	MLE	1.718	0	1.685	0	1.658	0	1.652	0
	Morg	1.665	0	1.627	0	1.598	0	1.594	0
	CH	1.672	0	1.630	0	1.599	0	1.594	0
	BY	**1.610**	1	1.561	0	1.552	0	1.552	0
	Med	1.960	7	1.737	2	**1.364**	0	**1.369**	0
	5-Med	1.815	9	1.537	3	1.365	0	1.377	0
	10-Med	1.717	12	1.529	3	1.368	0	1.379	0
	50-Med	1.744	15	**1.509**	3	1.368	0	1.379	0
0.3	MLE	1.958	0	1.914	0	1.892	0	1.879	0
	Morg	1.942	0	1.899	0	1.879	0	1.866	0
	CH	1.941	0	1.896	0	1.875	0	1.862	0
	BY	1.932	0	1.890	0	1.875	0	1.862	0
	Med	2.102	4	1.927	1	**1.777**	0	**1.776**	0
	5-Med	1.994	3	1.817	1	1.784	0	1.779	0
	10-Med	1.976	4	1.797	1	1.786	0	1.779	0
	50-Med	**1.953**	6	**1.787**	0	1.787	0	1.779	0

Table 4.6: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (0, 4.36)^T$ in the ε-contaminated logistic regression model (4.17) with $Pr(Y=1) = 0.5$.

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.862	1	**0.579**	0	**0.235**	0	**0.159**	0
	Morg	0.951	3	0.639	0	0.256	0	0.176	0
	CH	**0.840**	1	0.582	0	0.247	0	0.171	0
	BY	1.527	8	0.942	1	0.312	0	0.212	0
	Med	3.267	18	2.554	9	0.925	0	0.497	0
	5-Med	2.786	22	1.637	10	0.521	0	0.327	0
	10-Med	2.496	24	1.730	10	0.464	0	0.306	0
	50-Med	2.502	28	1.579	13	0.420	0	0.284	0
0.05	MLE	1.074	0	1.000	0	1.062	0	1.072	0
	Morg	1.048	1	**0.851**	0	0.818	0	0.822	0
	CH	**0.995**	0	0.871	0	0.881	0	0.886	0
	BY	1.362	4	0.915	1	0.623	0	0.611	0
	Med	2.659	12	2.358	7	0.832	0	0.617	0
	5-Med	2.004	17	1.777	6	0.638	0	0.586	0
	10-Med	1.953	17	1.546	8	0.615	0	0.585	0
	50-Med	2.166	24	1.386	13	**0.601**	0	**0.581**	0
0.1	MLE	1.484	0	1.500	0	1.551	0	1.550	0
	Morg	1.417	1	1.339	0	1.395	0	1.395	0
	CH	**1.411**	0	1.375	0	1.424	0	1.423	0
	BY	1.435	3	**1.171**	0	1.213	0	1.216	0
	Med	2.460	10	1.715	3	1.105	0	**1.109**	0
	5-Med	1.948	12	1.458	4	**1.101**	0	1.130	0
	10-Med	1.958	15	1.356	5	1.102	0	1.133	0
	50-Med	2.231	19	1.252	8	1.112	0	1.134	0
0.2	MLE	2.041	0	2.052	0	2.060	0	2.069	0
	Morg	2.006	0	2.009	0	2.019	0	2.029	0
	CH	2.010	0	2.010	0	2.018	0	2.027	0
	BY	1.966	0	1.966	0	1.997	0	2.009	0
	Med	2.222	5	1.996	2	**1.835**	0	**1.879**	0
	5-Med	**1.956**	5	1.896	2	1.875	0	1.890	0
	10-Med	2.037	5	1.837	1	1.879	0	1.889	0
	50-Med	2.054	8	**1.826**	2	1.881	0	1.891	0
0.3	MLE	2.374	0	2.374	0	2.389	0	2.389	0
	Morg	2.367	0	2.365	0	2.381	0	2.381	0
	CH	2.367	0	2.363	0	2.379	0	2.379	0
	BY	2.357	0	2.361	0	2.380	0	2.380	0
	Med	2.359	2	**2.279**	0	**2.330**	0	**2.336**	0
	5-Med	**2.312**	1	2.286	0	2.336	0	2.340	0
	10-Med	2.318	2	2.290	0	2.337	0	2.340	0
	50-Med	2.320	1	2.291	0	2.338	0	2.341	0

Table 4.7: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-2.82, 2.82)^T$ in the logistic regression model (4.21) ε-contaminated by leverage points and preserving $Pr(Y = 1) = 0.2$.

		$n=50$		$n=100$		$n=500$		$n=1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.948	1	**0.574**	0	**0.242**	0	**0.170**	0
	Morg	0.993	3	0.649	0	0.269	0	0.187	0
	CH	**0.900**	1	0.596	0	0.259	0	0.181	0
	BY	1.358	7	0.944	1	0.334	0	0.222	0
	Med	3.125	17	2.607	9	1.079	0	0.542	0
	5-Med	2.482	22	1.660	11	0.567	0	0.345	0
	10-Med	2.381	24	1.647	11	0.499	0	0.322	0
	50-Med	2.433	27	1.569	12	0.450	0	0.293	0
0.05	MLE	1.137	0	1.110	0	1.154	0	1.169	0
	Morg	1.122	1	0.889	0	0.844	0	0.860	0
	CH	**1.059**	0	0.933	0	0.927	0	0.942	0
	BY	1.420	4	**0.881**	0	0.632	0	0.636	0
	Med	2.822	13	2.118	7	0.823	0	0.649	0
	5-Med	2.270	17	1.465	6	0.668	0	0.614	0
	10-Med	1.906	19	1.389	7	0.629	0	**0.610**	0
	50-Med	2.279	25	1.354	12	**0.609**	0	0.611	0
0.1	MLE	1.578	0	1.628	0	1.668	0	1.668	0
	Morg	**1.445**	0	1.435	0	1.459	0	1.456	0
	CH	1.467	0	1.481	0	1.504	0	1.501	0
	BY	1.518	2	**1.282**	0	1.223	0	1.230	0
	Med	2.333	10	1.894	5	1.139	0	**1.126**	0
	5-Med	2.082	12	1.567	5	1.145	0	1.144	0
	10-Med	2.130	15	1.456	6	**1.111**	0	1.148	0
	50-Med	2.166	19	1.564	8	1.114	0	1.155	0
0.2	MLE	2.152	0	2.173	0	2.190	0	2.199	0
	Morg	2.095	0	2.111	0	2.133	0	2.144	0
	CH	2.102	0	2.115	0	2.133	0	2.143	0
	BY	2.032	0	2.038	0	2.094	0	2.110	0
	Med	2.409	6	2.021	2	**1.920**	0	**1.961**	0
	5-Med	2.103	5	1.922	2	1.948	0	1.975	0
	10-Med	**2.082**	6	1.921	1	1.949	0	1.977	0
	50-Med	2.165	10	**1.910**	2	1.954	0	1.977	0
0.3	MLE	2.489	0	2.509	0	2.516	0	2.519	0
	Morg	2.478	0	2.496	0	2.505	0	2.509	0
	CH	2.477	0	2.494	0	2.502	0	2.505	0
	BY	2.464	0	2.490	0	2.503	0	2.507	0
	Med	2.527	3	2.405	0	**2.439**	0	**2.451**	0
	5-Med	2.454	1	**2.398**	0	2.446	0	2.456	0
	10-Med	**2.406**	2	2.437	0	2.448	0	2.456	0
	50-Med	2.442	2	2.439	0	2.450	0	2.456	0

Table 4.8: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-2.16, 3.71)^T$ in the logistic regression model (4.21) ε-contaminated by leverage points and preserving $Pr(Y=1) = 0.3$.

		$n = 50$		$n = 100$		$n = 500$		$n = 1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.865	1	**0.560**	0	**0.235**	0	**0.167**	0
	Morg	0.921	2	0.603	0	0.254	0	0.184	0
	CH	**0.845**	1	0.567	0	0.247	0	0.178	0
	BY	1.289	8	0.842	2	0.304	0	0.213	0
	Med	2.724	19	2.338	10	0.891	0	0.518	0
	5-Med	2.228	22	1.651	10	0.509	0	0.320	0
	10-Med	2.164	24	1.559	12	0.477	0	0.297	0
	50-Med	2.171	28	1.377	14	0.429	0	0.276	0
0.05	MLE	1.068	0	0.988	0	1.071	0	1.085	0
	Morg	1.065	1	**0.806**	0	0.775	0	0.791	0
	CH	**0.992**	0	0.839	0	0.855	0	0.869	0
	BY	1.294	3	0.893	0	0.583	0	0.590	0
	Med	2.371	14	1.970	6	0.756	0	0.628	0
	5-Med	1.925	17	1.439	5	0.599	0	0.569	0
	10-Med	1.967	19	1.275	7	0.592	0	**0.567**	0
	50-Med	1.922	26	1.317	10	**0.573**	0	0.569	0
0.1	MLE	1.469	0	1.492	0	1.540	0	1.542	0
	Morg	1.353	0	1.319	0	1.336	0	1.343	0
	CH	1.376	0	1.360	0	1.381	0	1.385	0
	BY	**1.299**	2	**1.165**	0	1.113	0	1.137	0
	Med	2.127	11	1.802	4	1.033	0	**1.033**	0
	5-Med	1.653	13	1.400	6	**1.022**	0	1.058	0
	10-Med	1.580	15	1.307	6	1.025	0	1.057	0
	50-Med	1.645	21	1.288	7	1.033	0	1.059	0
0.2	MLE	1.983	0	1.989	0	2.016	0	2.022	0
	Morg	1.934	0	1.933	0	1.959	0	1.967	0
	CH	1.940	0	1.937	0	1.959	0	1.967	0
	BY	**1.891**	0	1.870	0	1.920	0	1.933	0
	Med	2.107	5	1.879	2	**1.759**	0	**1.792**	0
	5-Med	2.005	5	1.801	1	1.786	0	1.805	0
	10-Med	1.917	6	1.889	1	1.789	0	1.808	0
	50-Med	1.967	11	**1.757**	2	1.791	0	1.808	0
0.3	MLE	2.290	0	2.292	0	2.309	0	2.307	0
	Morg	2.278	0	2.280	0	2.299	0	2.296	0
	CH	2.278	0	2.277	0	2.296	0	2.293	0
	BY	2.264	0	2.273	0	2.296	0	2.294	0
	Med	2.455	2	2.237	0	**2.236**	0	**2.242**	0
	5-Med	2.258	2	**2.184**	0	2.245	0	2.245	0
	10-Med	2.260	3	2.198	0	2.244	0	2.244	0
	50-Med	**2.230**	5	2.193	0	2.245	0	2.244	0

Table 4.9: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (-1.16, 4.20)^T$ in the logistic regression model (4.21) ε-contaminated by leverage points and preserving $Pr(Y = 1) = 0.4$.

		$n=50$		$n=100$		$n=500$		$n=1000$	
ε	$\widetilde{\boldsymbol{\beta}}_n$	MAE	RR%	MAE	RR%	MAE	RR%	MAE	RR%
0	MLE	0.906	1	**0.563**	0	**0.231**	0	**0.160**	0
	Morg	0.939	2	0.616	0	0.256	0	0.176	0
	CH	**0.873**	1	0.584	0	0.249	0	0.170	0
	BY	1.210	8	0.821	1	0.303	0	0.207	0
	Med	2.735	20	2.095	10	0.863	0	0.457	0
	5-Med	1.982	23	1.472	10	0.497	0	0.314	0
	10-Med	1.888	25	1.325	12	0.459	0	0.290	0
	50-Med	1.845	31	1.308	12	0.419	0	0.275	0
0.05	MLE	0.989	0	0.938	0	0.913	0	0.911	0
	Morg	0.992	1	**0.779**	0	0.670	0	0.675	0
	CH	**0.927**	0	0.804	0	0.736	0	0.737	0
	BY	1.103	4	0.788	0	**0.511**	0	0.518	0
	Med	2.140	15	1.689	6	0.737	0	0.560	0
	5-Med	1.735	18	1.214	7	0.550	0	0.516	0
	10-Med	1.762	20	1.218	8	0.538	0	0.508	0
	50-Med	1.710	26	1.122	11	0.521	0	**0.506**	0
0.1	MLE	1.321	0	1.314	0	1.290	0	1.285	0
	Morg	1.220	0	1.173	0	1.130	0	1.126	0
	CH	1.240	0	1.207	0	1.164	0	1.159	0
	BY	**1.191**	2	**1.057**	0	0.963	0	0.963	0
	Med	2.094	11	1.592	5	0.912	0	**0.890**	0
	5-Med	1.751	13	1.309	5	**0.903**	0	0.910	0
	10-Med	1.790	16	1.227	5	0.905	0	0.916	0
	50-Med	1.825	21	1.212	7	0.908	0	0.914	0
0.2	MLE	1.745	0	1.712	0	1.682	0	1.671	0
	Morg	1.704	0	1.669	0	1.636	0	1.626	0
	CH	1.708	0	1.670	0	1.636	0	1.625	0
	BY	**1.685**	0	1.624	0	1.606	0	1.598	0
	Med	2.063	5	1.610	2	**1.490**	0	**1.501**	0
	5-Med	1.896	5	**1.533**	1	1.504	0	**1.501**	0
	10-Med	1.862	7	1.564	2	1.505	0	**1.501**	0
	50-Med	1.763	8	**1.533**	3	1.508	0	**1.501**	0
0.3	MLE	1.972	0	1.947	0	1.910	0	1.903	0
	Morg	1.963	0	1.937	0	1.901	0	1.894	0
	CH	1.963	0	1.935	0	1.898	0	1.891	0
	BY	**1.956**	0	1.932	0	1.899	0	1.892	0
	Med	2.057	2	**1.863**	0	1.859	0	**1.854**	0
	5-Med	2.110	3	1.891	0	**1.858**	0	**1.854**	0
	10-Med	2.007	3	1.891	1	1.859	0	1.855	0
	50-Med	1.991	4	1.880	0	1.860	0	1.855	0

Table 4.10: Mean absolute errors (4.16) and rejection rates RR for selected estimators $\widetilde{\boldsymbol{\beta}}_n = (\widetilde{\beta}_{n0}, \widetilde{\beta}_{n1})^T$ of the true parameter $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{01})^T = (0, 4.36)^T$ in the logistic regression model (4.21) ε-contaminated by leverage points and preserving $Pr(Y=1) = 0.5$.

Acknowledgement. The authors thank the referees and the associate editor for careful reading of the paper and many interesting improvements. Special thanks are due to the referee who proposed the procedure called statistical enhancing in this paper. This procedure proved to be able to increase the efficiency of our estimators, in some cases up to the level achieved by the MLE. It seems that it will remain to be useful also in applications of the concept of statistical smoothing in another discrete statistical models.

References

[1] G. Adimari and L. Ventura (2001) Robust inference for generalized linear models with application to logistic regression. *Statistics and Probability Letters,* **55**, 4, 413–419.

[2] A. Agresti (2002). *Categorical Data Analysis* (Second Edition). Wiley, New York.

[3] E. B. Andersen (1990). *The Statistical Analysis of Categorical data.* Springer, New York.

[4] M. A. Arcones (2001), Asymptotic distribution of regression M-estimators. *Journal of Statistical Planning and Inference,* **97,** 235-261.

[5] A. M. Bianco and V. J. Yohai (1996), Robust estimation in the logistic regression model. *Robust Statistics, Data Analysis, and Computer Intensive Methods* (Schloss Thurnau, 1994), 17-34. Lecture Notes in Statistics, 109, Springer, New York.

[6] X. R. Chen, L. Zhao and Y. Wu (1993), On conditions of consistency of ML_1N estimates. *Statistica Sinica,* **3,** 9-18.

[7] C. Croux, and G. Haesbroeck (2003), Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics and Data Analysis*, **44**, 273–295.

[8] J.E. Dennis, Jr. and Robert B. Schnabel (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, New Jersey.

[9] D. Gervini (2005), Robust adaptive estimators for binary regression models. *Journal of Statistical Planning and Inference,* **131**, 297–311.

[10] F. R. Hampel, P. J. Rousseeuw, E. M. Ronchetti, W. A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

[11] T. Hobza, L. Pardo and I. Vajda (2005), *Median Estimators in Generalized Logistic Regression.* Research Report DAR-UTIA 2005/40. Institute of Information Theory, Prague (available at http://xxxxxxxxxxxxxxxxxxxxxxxxxxx).

[12] J. Jurečková and B. Procházka (1994), Regression quantiles and trimmed least squares estimator in nonlinear regression model. *Nonparametric Statistics,* **3,** 201-222.

[13] J. Jurečková and P. K. Sen (1996), *Robust Statistical Procedures.* Wiley, New York.

[14] K. Knight (1998), Limiting distributions for L_1 regression estimators under general conditions, *Annals Statistics,* **26**, 755-770.

[15] R. Koenker and G. Basset (1987), Regression quantiles. *Econometrica,* **46,** 33-50.

[16] N. Kordzakhia, G. D. Mishra and L. Reiersølmoen (2001) Robust estimation in the logistic regression model, *Journal of Statistical Planning and Inference,* **98**, 211–223.

[17] F. Liese and I. Vajda (1999), M-estimators of structural parameters in pseudolinear models. *Applications of Mathematics,* **44,** 245-270.

[18] F. Liese and I. Vajda (2003), A general asymptotic theory of M-estimators I. *Mathematical Methods of Statistics,* **12,** 454-477.

[19] F. Liese and I. Vajda (2004), A general asymptotic theory of M-estimators II. *Mathematical Methods of Statistics,* **13,** 82-95.

[20] R. A. Maronna, R. D. Martin and V. J. Yohai (2006), *Robust Statistics. Theory and Methods.* Wiley, New York.

[21] A. M. Mood, F. A. Graybill and D. C. Boes (1974). *Introduction to the Theory of Statistics.* Mc GrawHill. New York.

[22] J. Moré, G. Burton, and H. Kenneth (1980), *User Guide for MINPACK-1*, Argonne National Laboratory Report ANL-80-74, Argonne, Illinois.

[23] S. Morgenthaler (1992), Least-absolute-deviations fits for generalized linear models. *Biometrika,* **79**, 747-754.

[24] J. A. Pardo, L. Pardo and M. C. Pardo (2006), Testing in logistic regression models based on ϕ-divergences measures. *Journal of Statistical Planning and Inference*, **136,** 982-1006.

[25] D. Pregibon (1982), Resistant lits for some commonly used logistic models with medical applications, Biometrics, **38,** 485-498.

[26] D. Pollard (1991), Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7,** 186-199.

[27] A. Rényi (1970), *Probability Theory.* North-Holland, Amsterdam.

[28] G. D. Richardson and B. B. Bhattacharyya (1987), Consistent L_1-estimates in non-linear regression for a noncompact parameter space. *Sankhya,* **49,** ser. A, 377-387.

[29] P. J. Rousseeuw and A. Christmann (2003), Robustness against separation and outliers in logistic regression. *Computational Statistics and Data Analysis,* **43**, 315–332.

[30] V. J. Yohai (1987), High breakdown point high efficiency robust estimates for regression. *Annals Statistics,* **15,** 692–656.

[31] S. Zwanzig (1997), *On L_1-norm Estimators in Nonlinear Regression and in Nonlinear Error-in-Variables Models.* IMS Lecture Notes 31, 101-118, Hayward.