



Robust tests in generalized linear models with missing responses

Ana M. Bianco^{a,*}, Graciela Boente^{a,*}, Isabel M. Rodrigues^b

^a Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina

^b Departamento de Matemática and CEMAT, Instituto Superior Técnico, Technical University of Lisbon (TULisbon), Lisboa, Portugal

ARTICLE INFO

Article history:

Received 14 August 2011

Received in revised form 6 May 2012

Accepted 10 May 2012

Available online 23 May 2012

Keywords:

Fisher-consistency

Generalized linear models

Influence function

Missing data

Outliers

Robust estimation

Robust testing

ABSTRACT

In many situations, data follow a generalized linear model in which the mean of the responses is modelled, through a link function, linearly on the covariates. Robust estimators for the regression parameter in order to build test statistics for this parameter, when missing data occur in the responses, are considered. The asymptotic behaviour of the robust estimators for the regression parameter is obtained, under the null hypothesis and under contiguous alternatives. This allows us to derive the asymptotic distribution of the robust Wald-type test statistics constructed from the proposed estimators. The influence function of the test statistics is also studied. A simulation study allows us to compare the behaviour of the classical and robust tests, under different contamination schemes. Applications to real data sets enable to investigate the sensitivity of the p -value to the missing scheme and to the presence of outliers.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The generalized linear model (McCullagh and Nelder, 1989), GLM, is a popular technique for modelling a wide variety of data and assumes that the observations (y_i, \mathbf{x}_i^T) , $1 \leq i \leq n$, $\mathbf{x}_i \in \mathbb{R}^k$, are independent with the same distribution as $(y, \mathbf{x}^T) \in \mathbb{R}^{k+1}$ such that the conditional distribution of $y|\mathbf{x}$ belongs to the canonical exponential family

$$\exp \{ [y\theta(\mathbf{x}) - B(\theta(\mathbf{x}))] / A(\tau) + C(y, \tau) \}, \quad (1)$$

for known functions A , B and C . In this situation, if we denote by B' the derivative of B , the mean $\mu(\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = B'(\theta(\mathbf{x}))$ is modelled linearly through a known link function, g , i.e., $g(\mu(\mathbf{x})) = \theta(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$. Robust procedures for generalized linear models have been considered, among others, by Stefanski et al. (1986), Künsch et al. (1989), Bianco and Yohai (1996), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2003) and Bianco et al. (2005); see also, Maronna et al. (2006). Recently, robust tests for the regression parameter under a logistic model were considered by Bianco and Martínez (2009).

In practice, some response variables may be missing by design (as in two-stage studies) or by happenstance. As is well known, the methods proposed by the above mentioned authors are designed for complete data sets and problems arise when missing observations are present. Even if there are many situations in which both the response and the explanatory variables are missing, we will focus our attention on those cases in which missing data occur only in the responses. Actually, missingness of responses is very common in opinion polls, market research surveys, mail enquiries, social-economic investigations, medical studies and other scientific experiments, when the explanatory variables can be controlled. This pattern appears, for example, in the scheme of double sampling proposed by Neyman (1938), where first a complete sample

* Correspondence to: Instituto de Cálculo, FCEyN, UBA Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina. Tel.: +54 11 45763375; fax: +54 11 45763375.

E-mail addresses: abianco@dm.uba.ar (A.M. Bianco), gboente@dm.uba.ar, gboente@fibertel.com.ar (G. Boente), irodrig@math.ist.utl.pt (I.M. Rodrigues).

is obtained and then, some additional covariate values are computed since, perhaps, this is less expensive than to obtain more response values. Hence, we will focus our attention on robust inference when the response variable may have missing observations, but the covariate \mathbf{x} is totally observed.

In this paper, we consider the robust estimators for the regression parameter β introduced by Bianco et al. (2011a), under a GLM model. When there are no missing data, these estimators include the family of estimators previously studied by several authors such as Bianco and Yohai (1996), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2003) and Bianco et al. (2005). It is shown that the robust estimators of β are asymptotically normally distributed which allows us to construct a robust procedure to test the hypothesis $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$. The paper is organized as follows. The robust proposal is given in Section 2, the asymptotic distribution of the regression estimators and a robust Wald-type test for the regression parameter are provided in Section 3. The results of a Monte Carlo study are summarized in Section 4, while in Section 5 we investigate the empirical breakdown point of the different procedures. The proposed procedure is illustrated over two real data examples in Section 6 where we carried out a sensitivity study for the p -value. An expression for the influence function of the test is obtained in Section 7. Proofs are relegated to the Appendix.

2. Robust inference

2.1. Framework and the robust estimators

Suppose we obtain a random sample of incomplete data $(y_i, \mathbf{x}_i^T, \delta_i)$, $1 \leq i \leq n$, of a generalized linear model where $\delta_i = 1$ if y_i is observed, $\delta_i = 0$ if y_i is missing and $(y_i, \mathbf{x}_i^T) \in \mathbb{R}^{k+1}$ are such that the conditional distribution $F(\cdot, \mu_i, \tau)$ of $y_i|\mathbf{x}_i$ belongs to the canonical exponential family given in (1), with $\mu_i = H(\mathbf{x}_i^T \beta)$ and $\text{VAR}(y_i|\mathbf{x}_i) = A^2(\tau)V(\mu_i) = A^2(\tau)B''(\theta(\mathbf{x}_i))$ with B'' the second derivative of B . Let (β, τ) denote the true parameter values and \mathbb{E}_F the expectation under the true model; thus $\mathbb{E}_F(y|\mathbf{x}) = H(\mathbf{x}^T \beta)$. In a more general situation, we will think of τ as a nuisance parameter such as an additional scale or dispersion parameter or even, the tuning constant for the score function to be considered below. For instance, under a Gamma regression model τ is related to the shape parameter, while for Poisson and logistic regression, $\tau = 1$.

Let $(y, \mathbf{x}^T, \delta)$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i^T, \delta_i)$. Bianco et al. (2011a) defined robust estimators of the regression parameter when missing responses occur under an ignorable missing mechanism. To be more precise, they assumed that y is missing at random (MAR), that is, δ and y are conditionally independent given \mathbf{x} , i.e.,

$$P(\delta = 1|y, \mathbf{x}) = P(\delta = 1|\mathbf{x}) = p(\mathbf{x}). \tag{2}$$

A common assumption in the literature states that $\inf_{\mathbf{x}} p(\mathbf{x}) > 0$, meaning that at any value of the covariate response variables are observed.

For the sake of completeness, we remind the definition of the simplified estimators considered in Bianco et al. (2011a) where also a propensity score approach is considered. Through a heuristic argument based on the influence function, Bianco et al. (2011a) showed that in some situations, such as the Gamma model to be considered below, the asymptotic variance of the robust simplified estimators is smaller than that of the propensity score ones. For that reason, we will focus here on test statistics based on the robust simplified estimators.

Let $w_1 : \mathbb{R}^k \rightarrow \mathbb{R}$ be a weight function to control leverage points on the carriers \mathbf{x} and $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ a loss function. For any $\mathbf{b} \in \mathbb{R}^k$, $t \in \mathbb{R}$, let us define

$$S_n(\mathbf{b}, t) = \frac{1}{n} \sum_{i=1}^n \delta_i \rho(y_i, \mathbf{x}_i^T \mathbf{b}, t) w_1(\mathbf{x}_i), \tag{3}$$

$$S(\mathbf{b}, t) = \mathbb{E}_F[\delta \rho(y, \mathbf{x}^T \mathbf{b}, t) w_1(\mathbf{x})] = \mathbb{E}_F[p(\mathbf{x}) \rho(y, \mathbf{x}^T \mathbf{b}, t) w_1(\mathbf{x})]. \tag{4}$$

In order to define Fisher-consistent estimators, Bianco et al. (2011a) assumed that $w_1(\cdot)$ and $\rho(\cdot)$ are such that $S(\beta, \tau) = \min_{\mathbf{b}} S(\mathbf{b}, \tau)$. As mentioned above, the parameter t in $S(\mathbf{b}, t)$ plays the role of a nuisance parameter.

Let $\hat{\tau} = \hat{\tau}_n$ be robust consistent estimators of τ , the robust simplified estimator $\hat{\beta}$ of the regression parameter is defined as

$$\hat{\beta} = \underset{\mathbf{b}}{\operatorname{argmin}} S_n(\mathbf{b}, \hat{\tau}). \tag{5}$$

Under mild conditions the consistency of $\hat{\beta}$ is derived in Bianco et al. (2011a).

When ρ is continuously differentiable, if we denote by $\Psi(y, u, t) = \partial \rho(y, u, t) / \partial u$, then β and $\hat{\beta}$ satisfy the differentiated equations $S^{(1)}(\beta, \tau) = \mathbf{0}_k$ and $S_n^{(1)}(\mathbf{b}, \hat{\tau}) = \mathbf{0}_k$, respectively, where $S^{(1)}(\mathbf{b}, t) = \mathbb{E}_F(\Psi(y, \mathbf{x}^T \mathbf{b}, t) w_1(\mathbf{x}) p(\mathbf{x}) \mathbf{x})$ and $S_n^{(1)}(\mathbf{b}, t) = (1/n) \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \mathbf{b}, t) w_1(\mathbf{x}_i) \mathbf{x}_i$.

Remark 2.1.1. Two classes of loss functions ρ have been considered in the literature. One of them aims to bound the deviances, while the other one introduced by Cantoni and Ronchetti (2001) bounds the Pearson residuals. In both cases, the correction term needed to ensure Fisher-consistency is included in the function ρ . For a complete description,

see Bianco et al. (2011a). In particular, the Poisson and log-Gamma model were considered therein. We refer to Bianco et al. (2005) for a description on robust estimators based on deviances for regression models with asymmetric errors where all the responses are observed and also to Heritier et al. (2009) for a description on M -type estimators for the log-Gamma model. One of the main advantages of robust estimators based on deviances is that they do not need a correction term when the response distribution belongs to a continuous family of distributions with strongly unimodal density function (Bianco et al., 2005). As pointed in Heritier et al. (2009), the nature of possible deviations in generalized linear models is close to those appearing in regression models, in the sense that atypical observations may correspond to outliers in the responses (producing large vertical residuals) and/or leverage points in the design space. The deviation from the model that affects the responses can be measured either through the Pearson residuals or through the deviances, while the weights $w_1(\mathbf{x}_i)$ control covariates with high leverage. However, when considering a logistic model, deviations in the responses correspond to misclassification (a zero instead of a one, or vice versa), which is naturally measured through the deviance, so robust estimators bounding deviances should be considered. One of the advantages of estimators bounding the Pearson residuals is their direct interpretation. In fact, the estimating equations in this last situation can be interpreted as the classical estimating equations weighted (both with respect to the Pearson residuals and to \mathbf{x}_i) and re-centred via the correction term to ensure consistency. In this way, the estimators can be computed through iterative reweighted least squares algorithms; see Heritier et al. (2009) for a discussion. The general proposal based on Pearson residuals considered in Cantoni and Ronchetti (2001) and in particular, for Gamma models in Cantoni and Ronchetti (2006), use a Huber's score function to bound large values of the residuals. We suspect that using a bounded loss function increases the breakdown point of the final estimators, even when considering Pearson residuals. In our simulation study and in the examples analysed, we choose a Tukey's score function to bound large values of the deviance.

It is worth noting that when defining the estimators in (5), a preliminary estimator of the nuisance parameter is needed. This parameter includes the situation of exponential families with overdispersion, such as the negative binomial $y_i|\mathbf{x}_i \sim \mathcal{NB}(\tau, \mu_i)$, $\tau > 0$, where $\mathbb{E}(y_i|\mathbf{x}_i) = \mu_i$ and $\text{VAR}(y_i|\mathbf{x}_i) = \mu_i + \tau\mu_i^2$ or the Gamma regression model $y_i|\mathbf{x}_i \sim \Gamma(\tau, \mu_i)$ where $\mathbb{E}(y_i|\mathbf{x}_i) = \mu_i$ and $\text{VAR}(y_i|\mathbf{x}_i) = \mu_i^2/\tau$, when $\tau \leq 1$ and $\mu_i > 1$. A simple way to define a preliminary estimator $\hat{\tau}$ of τ is to consider joint estimators of $(\boldsymbol{\beta}, \tau)$ as the solution of $(\hat{\boldsymbol{\beta}}, \hat{\tau}) = \text{argmin}_{\mathbf{b}, t} S_n(\mathbf{b}, t)$, for some choice of the loss function $\rho = \rho_0$, leading to estimators with lower efficiency but higher breakdown point. Then, to improve the efficiency of the regression estimators, a different loss function may be used to plug-in the value of $\hat{\tau}$ in (5). However, even when dealing with no missing responses, in some situations, such as location-scale models or regression models simultaneous estimators lead to lower breakdown point estimators of the target parameter than regression estimators computed using a preliminary S -estimator with a high breakdown point (see, for instance Maronna et al., 2006, for a discussion). For that reason, and since Gamma models can be transformed to a regression model with asymmetric errors, the proposal described in Remark 3.2 is considered.

2.2. Test statistics

When the responses are missing at random, the results in Section 3 show that under mild conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(\mathbf{0}_k, \boldsymbol{\Sigma}_\beta)$ where $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ and the symmetric matrices \mathbf{A} and \mathbf{B} are defined as

$$\mathbf{A} = \mathbb{E}_F(\chi(y, \mathbf{x}^T\boldsymbol{\beta}, \tau) w_1(\mathbf{x})p(\mathbf{x})\mathbf{x}\mathbf{x}^T) \quad (6)$$

$$\mathbf{B} = \mathbb{E}_F(\Psi^2(y, \mathbf{x}^T\boldsymbol{\beta}, \tau) w_1^2(\mathbf{x})p(\mathbf{x})\mathbf{x}\mathbf{x}^T) \quad (7)$$

with $\chi(y, u, \tau) = \partial\Psi(y, u, \tau)/\partial u$.

Estimators of \mathbf{A} and \mathbf{B} can easily be obtained through their sample versions. So, let us define $\hat{\mathbf{A}} = \sum_{i=1}^n \delta_i \chi(y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}}, \hat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T / n$, $\hat{\mathbf{B}} = \sum_{i=1}^n \delta_i \Psi^2(y_i, \mathbf{x}_i^T \hat{\boldsymbol{\beta}}, \hat{\tau}) w_1^2(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T / n$. A Wald-type test statistic to test the hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ can thus be defined as $\hat{\mathbf{W}}_n = n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, with $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$. This test statistics will be asymptotically χ_k^2 distributed under the null hypothesis; hence H_0 is rejected if $\hat{\mathbf{W}}_n > \chi_{k, \alpha}^2$, where $\chi_{k, \alpha}^2$ denotes the $1 - \alpha$ quantile of the χ_k^2 distribution, i.e., $P(\chi_k^2 > \chi_{k, \alpha}^2) = \alpha$.

The asymptotic behaviour of $\hat{\mathbf{W}}_n$ under contiguous alternatives is derived in Section 3. When the tested parameter is one-dimensional, the quasi-profile log-likelihood approach considered by Adimari and Ventura (2002) may be modified to include missing responses. Also, a score type test as defined in Heritier and Ronchetti (1994) when there are no missing responses in the sample, can be adapted to the present situation. As in the complete sample case, we expect these tests to be asymptotically equivalent. As is well known, for data sets with no missings, the advantage of these two procedures over robust likelihood ratio-type tests based on a loss function ρ , also called dispersion tests, is that the robust likelihood-ratio ones behave asymptotically as a linear combination of independent χ_1^2 whose coefficients needs to be estimated, since they depend on the unknown parameters. On the other hand, Wald tests can suffer from the Hauck–Donner effect in a logistic regression whereas this is not the case for the likelihood ratio or score tests. From this point of view the latter should be preferred to the former. For missing responses, the robust likelihood ratio-type tests can be written as $2\{S_n(\hat{\boldsymbol{\beta}}, \hat{\tau}) - S_n(\boldsymbol{\beta}_0, \hat{\tau})\}$ where $S_n(\mathbf{b}, t)$ is defined in (3). It is worth noting that the asymptotic behaviour of robust likelihood ratio-type tests is, up

to our knowledge, not known in this situation. This interesting topic may be the subject of future research. For complete data sets, an interesting discussion on robust likelihood ratio-type tests is given in [Heritier and Ronchetti \(1994\)](#). Besides, as pointed out in [Amaral Turkman and Silva \(2000\)](#), classical Wald tests are preferable to classical likelihood ratio-type tests when overdispersion is present. For that reason, we have followed an approach based on the Wald-type test.

3. Asymptotic behaviour of the test statistics

In this section, we derive the asymptotic distribution of the test statistics under the null hypothesis and under contiguous alternatives. We will consider the following set of assumptions.

- N1. The functions $w_1(\mathbf{x})$ and $w_1(\mathbf{x})\|\mathbf{x}\|$ are bounded.
- N2. $\mathbb{E}_F(p(\mathbf{x})w_1(\mathbf{x})\|\mathbf{x}\|^2) < \infty$.
- N3. $\Psi(y, u, v)$ and $\chi(y, u, v) = \partial\Psi(y, u, v)/\partial u$ are bounded continuous functions.
- N4. The matrix \mathbf{A} defined in (6) is non-singular.
- N5. The class of functions $\mathcal{F} = \{f_\tau(y, \mathbf{x}, \delta) = \delta\Psi(y, \mathbf{x}^\top\boldsymbol{\beta}, \tau)w_1(\mathbf{x}), \tau \in \mathcal{K}\}$, where \mathcal{K} is a compact neighbourhood of τ , has finite entropy.
- N6. $\mathbb{E}_F(\Psi(y, \mathbf{x}^\top\boldsymbol{\beta}, \tau) | \mathbf{x}) = \mathbf{0}_k$ for any fixed $\tau \in \mathcal{K}$.

Remark 3.1. Assumptions N1 and N3 are standard requirements since they state that the weight function controls large values of the covariates and that the score function bounds large residuals, respectively. N2 is fulfilled for instance, for a 0–1 weight function and more generally, if $w_1(\mathbf{x})\|\mathbf{x}\|^2$ is bounded. Assumption N4 is a standard condition in the robustness literature to guarantee that the regression estimators will be root $-n$ consistent. On the other hand, N5 is fulfilled for the family of functions studied in [Bianco et al. \(2005\)](#), when τ plays the role of the tuning constant, $\rho(y, \mathbf{x}^\top\boldsymbol{\beta}, \tau) = \rho(\sqrt{d^*}(y, \mathbf{x}, \boldsymbol{\beta})/\tau)$ with $d^*(y, \mathbf{x}, \boldsymbol{\beta}) = -1 - (\log(y) - \mathbf{x}^\top\boldsymbol{\beta}) + y \exp(-\mathbf{x}^\top\boldsymbol{\beta})$, if ρ is twice continuously differentiable and there exists M such that $|u| > M$ implies that $\rho(u) = \sup_v \rho(v)$. Finally, note that N6 holds for the usual functions considered in robustness, it is the conditional Fisher-consistency defined by [Künsch et al. \(1989\)](#). For instance, for regression models, $y_i = \mathbf{x}_i^\top\boldsymbol{\beta} + u_i$ with symmetric errors u_i , the usual choice is to select an even loss function bounding the residuals to guarantee N6. On the other hand, as mentioned in [Remark 2.1.1](#), for generalized regression models a correction term is usually included in the score function to guarantee the conditional Fisher-consistency (see, for instance, [Cantoni and Ronchetti, 2001](#), for estimators bounding the quasi-likelihood, [Bianco and Yohai, 1996](#); [Croux and Haesbroeck, 2003](#), for estimators bounding the deviance under a logistic model). [Bianco et al. \(2005\)](#) note that, when bounding the deviance for continuous responses, the correction term equals 0.

The following theorem states the behaviour of the $\widehat{\boldsymbol{\beta}}$ under the null hypothesis and under contiguous alternatives. Its proof is given in the [Appendix](#).

Theorem 3.1. Assume that $(y_i, \mathbf{x}_i^\top, \delta_i), 1 \leq i \leq n$ are independent random vectors satisfying (2) and such that $y_i | \mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ where $\mu_i = H(\mathbf{x}_i^\top\boldsymbol{\beta}_n)$, with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}$. Assume that N1–N6 hold and that $\widehat{\tau}_n \xrightarrow{p} \tau$. Let $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where the symmetric matrices \mathbf{A} and \mathbf{B} are defined in (6) and (7), respectively. Then, we have that

- (a) Under $H_0 : \mathbf{c} = \mathbf{0}_k$, i.e., under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}_k, \boldsymbol{\Sigma})$.
- (b) Under $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$, i.e., under $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$, if $\mathbb{E}(|H'(\mathbf{x}^\top\boldsymbol{\beta}_0)|\|\mathbf{x}\|^2) < \infty$, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(-\mathbf{c}, \boldsymbol{\Sigma})$. Moreover, $\widehat{\mathbf{B}}^{-1/2}\widehat{\mathbf{A}}\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(-\mathbf{B}^{-1/2}\mathbf{A}\mathbf{c}, \mathbf{I}_k)$ where $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ are defined in Section 2.2.

The following theorem states the asymptotic behaviour of the proposed Wald-type test statistics and its proof follows easily applying [Theorem 3.1](#) and [Lemma A.1](#) in the [Appendix](#).

Theorem 3.2. Assume that $(y_i, \mathbf{x}_i^\top, \delta_i), 1 \leq i \leq n$ are independent random vectors satisfying (2) and such that $y_i | \mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ where $\mu_i = H(\mathbf{x}_i^\top\boldsymbol{\beta}_n)$, with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-1/2}$. Let $\boldsymbol{\Sigma} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$, where the symmetric matrices \mathbf{A} and \mathbf{B} are defined in (6) and (7), respectively. Assume that N1–N6 hold and that $\widehat{\tau}_n \xrightarrow{p} \tau$, then we have that

- (a) Under $H_0 : \mathbf{c} = \mathbf{0}_k$, i.e., under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\widehat{\mathcal{W}}_n \xrightarrow{d} \chi_k^2$.
- (b) Under $H_{1,n} : \mathbf{c} \neq \mathbf{0}_k$, i.e., under $H_{1,n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$, if $\mathbb{E}(|H'(\mathbf{x}^\top\boldsymbol{\beta}_0)|\|\mathbf{x}\|^2) < \infty$, $\widehat{\mathcal{W}}_n \xrightarrow{d} \chi_k^2(\theta)$, where $\theta = \mathbf{c}^\top\boldsymbol{\Sigma}^{-1}\mathbf{c}$.

From [Theorem 3.2](#), as mentioned in Section 2.2, to test the null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ at a given asymptotic level α , the decision rule rejecting H_0 if $\widehat{\mathcal{W}}_n > \chi_{k,\alpha}^2$ leads to a consistent test.

In regression, one of the most frequent hypothesis testing problems involves only a subset of the regression parameter. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}^\top, \boldsymbol{\beta}_{(2)}^\top)^\top$, $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_{(1)}^\top, \widehat{\boldsymbol{\beta}}_{(2)}^\top)^\top$ and $\mathbf{x} = (\mathbf{x}_{(1)}^\top, \mathbf{x}_{(2)}^\top)^\top$, where $\boldsymbol{\beta}_{(1)} \in \mathbb{R}^{k_1}$ with $k_1 < k$. In order to test $H_{0\boldsymbol{\beta}_{(1)}} : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\boldsymbol{\beta}_{(2)}$ unspecified, one may use the statistic $\widehat{\mathcal{W}}_{1,n} = n(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})^\top \widehat{\boldsymbol{\Sigma}}_{11}^{-1}(\widehat{\boldsymbol{\beta}}_{(1)} - \boldsymbol{\beta}_{(1),0})$, where $\widehat{\boldsymbol{\Sigma}}_{11}$ denotes the $k_1 \times k_1$ sub-matrix of $\widehat{\boldsymbol{\Sigma}}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$.

Theorem 3.3. Assume that $(y_i, \mathbf{x}_i^T, \delta_i)$, $1 \leq i \leq n$ are independent random vectors satisfying (2) and such that $y_i | \mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ where $\mu_i = H(\mathbf{x}_i^T \boldsymbol{\beta}_n)$, with $\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + \mathbf{c}n^{-\frac{1}{2}}$ with $\mathbf{c} = (\mathbf{c}_{(1)}^T, \mathbf{0}_{k-k_1}^T)^T$. Assume that N1–N6 hold and that $\widehat{\tau}_n \xrightarrow{p} \tau$. Denote by $\boldsymbol{\Sigma}_{11}$ the $k_1 \times k_1$ submatrix of $\boldsymbol{\Sigma} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$, corresponding to the coordinates of $\boldsymbol{\beta}_{(1)}$ where the symmetric matrices \mathbf{A} and \mathbf{B} are defined in (6) and (7), respectively. Then, we have that

- (a) Under $H_0 \boldsymbol{\beta}_{(1)} : \mathbf{c}_{(1)} = \mathbf{0}_{k_1}$, i.e., under $H_0 : \boldsymbol{\beta}_{(1)} = \boldsymbol{\beta}_{(1),0}$, $\widehat{\mathbf{W}}_{1,n} \xrightarrow{\mathcal{D}} \chi_{k_1}^2$.
 (b) Under $H_1 \boldsymbol{\beta}_{(1),n} : \mathbf{c}_{(1)} \neq \mathbf{0}_{k_1}$, i.e., under $H_1 \boldsymbol{\beta}_{(1),n} : \boldsymbol{\beta} = \boldsymbol{\beta}_n$, if $\mathbb{E}(|H'(\mathbf{x}^T \boldsymbol{\beta}_0)| \|\mathbf{x}\|^2) < \infty$, $\widehat{\mathbf{W}}_{1,n} \xrightarrow{\mathcal{D}} \chi_{k_1}^2(\theta)$, where $\theta = \mathbf{c}_{(1)}^T \boldsymbol{\Sigma}_{11}^{-1} \mathbf{c}_{(1)}$.

Remark 3.2. For any $\tau > 0$ and $\mu > 0$, we denote by $\Gamma(\tau, \mu)$ the parametrization of the Gamma distribution given by the density $f(v, \tau, \mu) = \tau^\tau v^{\tau-1} \exp(-(\tau/\mu)v) \{\mu^\tau \Gamma(\tau)\}^{-1} I_{v \geq 0}$. Note that, if $v \sim \Gamma(\tau, \mu)$, we have that $\mathbb{E}(v) = \mu$ and $\text{VAR}(v) = \mu^2/\tau$, where τ is a shape parameter.

Assume that $(y_i, \mathbf{x}_i^T, \delta_i)$, $1 \leq i \leq n$, is a sample from a generalized linear model such that $(y_i, \mathbf{x}_i^T, \delta_i) \sim (y, \mathbf{x}^T, \delta)$ where $y | \mathbf{x} \sim \Gamma(\tau, \mu(\mathbf{x}))$ with link function $\log(\mu(\mathbf{x})) = \mathbf{x}^T \boldsymbol{\beta}$. It is worth noticing that, in this case, the responses can be transformed so that they are modelled through a linear regression model with asymmetric errors. Indeed, let $z = \log(y)$, then $z = \mathbf{x}^T \boldsymbol{\beta} + u$, where $u \sim \log(\Gamma(\tau, 1))$ and u and \mathbf{x} are independent. Besides, the density of u is $g(u, \tau)$ where

$$g(u, \tau) = \frac{\tau^\tau}{\Gamma(\tau)} \exp(\tau(u - \exp(u))), \quad (8)$$

is asymmetric and unimodal with maximum at $u_0 = 0$. Since we are dealing with the situation in which some of the responses y , the transformed responses $z = \log(y)$, may be missing according to the MAR model (2), δ and z are conditionally independent given \mathbf{x} , so δ and u are independent.

When no missing responses arise, the estimators defined in Bianco et al. (2005) bound the deviances using a loss function ρ . We describe briefly how to adapt the estimators defined therein to allow missing responses and a weight function to control leverage points which will be effective when testing hypothesis on the regression parameter.

Denote by $d_i(\mathbf{b}, \tau)$ the deviance component of the i -th observation, i.e., $d_i(\mathbf{b}, \tau) = 2\tau d^*(y_i, \mathbf{x}_i, \mathbf{b})$ where $d^*(y, \mathbf{x}, \mathbf{b}) = -1 - (\log(y) - \mathbf{x}^T \mathbf{b}) + y \exp(-\mathbf{x}^T \mathbf{b})$.

First note that, since the tuning constant of the loss function depends on the unknown parameter τ , Bianco et al. (2005) introduce an adaptive sequence of tuning constants $\widehat{c}_{M,n}$ to define a sequence of M -estimators, $\widehat{\boldsymbol{\beta}}_{M,n}$, for data sets with no missing observations. These estimators, which satisfy $\widehat{\boldsymbol{\beta}}_{M,n} = \text{argmin}_{\mathbf{b}} \sum_{i=1}^n \phi(\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}/\widehat{c}_{M,n})$, with $\widehat{c}_{M,n} \xrightarrow{p} c_0$, have an asymptotic covariance matrix equal to $(B(\phi, \tau, c_0)/A^2(\phi, \tau, c_0)) \mathbb{E}(\mathbf{xx}^T)^{-1}$. The constants $B(\phi, \tau, c_0)$ and $A^2(\phi, \tau, c_0)$ depend only on the derivative of the score function ϕ and the shape parameter τ , but not on the covariates. Hence, the estimators can be calibrated to attain a given efficiency. From now on, denote $C_e(\tau)$ as the value of the tuning constant c_0 such that the M -estimator has efficiency e with respect to the maximum likelihood estimator.

In our modification, we consider the following three step algorithm to compute a generalized MM -estimator.

- *Step 1.* We first compute an initial S -estimate $\widetilde{\boldsymbol{\beta}}_n$ of the regression parameter and the corresponding scale estimate $\widehat{\sigma}_n$, taking $b = \frac{1}{2} \sup \phi$ with the complete data set. To be more precise, for each value of \mathbf{b} let $\sigma_n(\mathbf{b})$ be the M -scale estimate of $\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}$ given by

$$\sum_{i=1}^n \delta_i \phi \left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})}}{\sigma_n(\mathbf{b})} \right) = b \sum_{i=1}^n \delta_i,$$

where ϕ is Tukey's bisquare function,

$$\phi(u) = \min \left\{ \frac{u^2}{2} \min \left(1 - u^2 + \frac{u^4}{3} \right), 1 \right\}.$$

The S -estimate of $\boldsymbol{\beta}$ for the considered model is defined as $\widetilde{\boldsymbol{\beta}}_n = \text{argmin}_{\mathbf{b}} \sigma_n(\mathbf{b})$ and the corresponding scale estimate as $\widehat{\sigma}_n = \min_{\mathbf{b}} \sigma_n(\mathbf{b})$.

Let u be a random variable with density (8), $h(u) = 1 - u - \exp(u)$ and G the distribution of the errors u . Write $\sigma^*(\tau)$ the solution of $\mathbb{E}_G[\phi(\sqrt{h(u)}/\sigma^*(\tau))] = b$. Under mild conditions, $\widetilde{\boldsymbol{\beta}}_n \xrightarrow{\text{a.s.}} \boldsymbol{\beta}$ and $\widehat{\sigma}_n \xrightarrow{\text{a.s.}} \sigma^*(\tau)$. Besides, $\sigma^*(\tau)$ is a continuous and strictly decreasing function so, an estimator of τ can be defined as $\widehat{\tau}_n = \sigma^{*-1}(\widehat{\sigma}_n)$ leading to a strongly consistent estimator for τ .

- *Step 2.* As mentioned above, let $C_e(\tau)$ be the constant given efficiency e for data sets with no missing responses. In the second step, we compute $\widehat{\tau}_n = \sigma^{*-1}(\widehat{\sigma}_n)$ and $\widehat{c}_n = \max(\widehat{\sigma}_n, C_e(\widehat{\tau}_n)) = \max(\widehat{\sigma}_n, C_e(\sigma^{*-1}(\widehat{\sigma}_n)))$. Then, we have that $\widehat{c}_n \xrightarrow{p} c_0 = \max\{\sigma^*(\tau), C_e(\tau)\}$.

• Step 3. Let $\widehat{\beta}_n$ be the adaptive estimator of β defined by

$$\widehat{\beta}_n = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i=1}^n \delta_i \phi \left(\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})} / \widehat{c}_n \right) w_1(\mathbf{x}_i). \tag{9}$$

Note that in this case, $\rho(y_i, \mathbf{x}_i^T \mathbf{b}, t) = \phi \left(\sqrt{d^*(y_i, \mathbf{x}_i, \mathbf{b})} / c(t) \right)$, where $c(t) = \max\{\sigma^*(t), C_c(t)\}$, so that $\rho(y, v, t) = \phi \left(\sqrt{-1 - \log(y) + v + y \exp(-v)} / c(t) \right)$. Hence, for differentiable loss functions ϕ , we have that

$$\psi(y, v, t) = \phi' \left(\sqrt{-1 - \log(y) + v + y \exp(-v)} / c(t) \right) \frac{(1 - y \exp(-v))}{2c(t)\sqrt{-1 - \log(y) + v + y \exp(-v)}}.$$

Taking conditional expectation and using Lemma 1 in Bianco et al. (2005), we have that the conditional Fisher-consistency required in N6 holds for differentiable loss functions ϕ such that

- (i) $\phi(0) = 0$.
- (ii) $0 < \|\phi\|_\infty = \sup_u \phi(u) < 1$.
- (iii) If $0 \leq u < v$, then $\phi(u) \leq \phi(v)$.
- (iv) If $\phi(u) < \|\phi\|_\infty$ and $0 \leq u < v$, then $\phi(u) < \phi(v)$.
- (v) $\psi = \phi'$ is twice continuously differentiable, $\psi(0) = 0$, $\psi'(0) > 0$ and $\psi''(0) = 0$.

More generally, let us consider the regression model $z_i = \mathbf{x}_i^T \beta + u_i$ where the errors u_i have a (symmetric or not) continuous distribution with strictly positive and strictly unimodal density, f_0 . Denote $d(y, v) = 2(\log f_0(u_0) - \log f_0(y - v))$, $u_0 = \operatorname{argmax}_f f_0(u)$ and assume that ϕ satisfies (i) to (v). Then, Lemma 1 in Bianco et al. (2005) also ensures that N6 holds for the estimators defined as $\widehat{\beta}_n = \underset{\mathbf{b}}{\operatorname{argmin}} \sum_{i=1}^n \delta_i \phi \left(\sqrt{d(y_i, \mathbf{x}_i^T \mathbf{b})} / c \right) w_1(\mathbf{x}_i)$.

As a consequence of Theorem 3.1 and since $\widehat{c}_n \xrightarrow{p} c_0$, we have that

$$\sqrt{n}(\widehat{\beta}_n - \beta) \xrightarrow{d} N(\mathbf{0}_k, (B(\phi, \tau, c_0) / A^2(\phi, \tau, c_0)) \mathbf{D})$$

where $\mathbf{D} = \mathbf{C}^{-1} \mathbb{E} (p(\mathbf{x}) w_1^2(\mathbf{x}) \mathbf{x} \mathbf{x}^T) \mathbf{C}^{-1}$ with $\mathbf{C} = \mathbb{E} (p(\mathbf{x}) w_1(\mathbf{x}) \mathbf{x} \mathbf{x}^T)$. Note that when a 0–1 weight function is considered, the asymptotic matrix \mathbf{D} reduces to $\mathbf{D} = \mathbb{E} (p(\mathbf{x}) w_1(\mathbf{x}) \mathbf{x} \mathbf{x}^T)^{-1}$. Besides, when $w_1 \equiv 1$, we get the same efficiency e as the MM-estimators defined in Bianco et al. (2005) for data sets with no missing responses.

When dealing with missing responses, the asymptotic relative efficiency of $\widehat{\beta}_n$ depends on the asymptotic efficiency for the complete data set and on the matrix \mathbf{D} . For the sake of simplicity, in the simulation study we have calibrated the estimators to attain an efficiency e when $w_1 \equiv 1$ so, an extra loss of efficiency should be expected for the GM-estimators, leading to a loss of power.

4. Monte Carlo study for Gamma models

We have performed a simulation study for the following Gamma regression model

$$y_i | \mathbf{x}_i \sim \Gamma(\tau, \mu(\mathbf{x}_i)) \quad \text{with } \log(\mu(\mathbf{x}_i)) = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3, \quad 1 \leq i \leq n, \tag{10}$$

with $\tau = 3$, $\beta_1 = \beta_2 = \beta_3 = 0$, $\mathbf{x}_i^T = (x_{1i}, x_{2i}, 1)$ and $(x_{1i}, x_{2i}) \sim N(\mathbf{0}_2, \mathbf{I})$. The sample size was $n = 100$ and the number of Monte Carlo replications was $K = 1000$. The results of a simulation study when $\tau = 1$ are reported in Bianco et al. (2011b).

We studied the behaviour of the test statistics for samples without outliers and samples contaminated with 5% outliers. In the contaminated samples, the outliers were all equal, say (y_0, \mathbf{x}_0^T) . Since the magnitude of the effect of these outliers depends on x_{10} and x_{20} only throughout $x_{10}^2 + x_{20}^2$, without loss of generality they were taken of the form (y_0, \mathbf{x}_0^T) with $\mathbf{x}_0^T = (x_0, 0, 1)$ and $y_0 = \exp(m x_0)$. The value m represents the slope of the outlying observations. We chose two values of x_0 corresponding to moderate ($x_0 = 3$) and high ($x_0 = 10$) leverage outliers and also, the value $x_0 = 1$ to take into account the effect of having only vertical outliers when m is large. As values for m we considered $m = 0.5$ and 2.5 . These contaminations are denoted by C_{m, x_0} . In Bianco et al. (2011b) other contamination schemes are considered, however those reported here are representative of most of them.

We considered Wald test statistics related to the maximum likelihood estimators, $\widehat{\beta}_{ML}$, the robust estimators based on the deviance with $w_1 \equiv 1$, $\widehat{\beta}_M$, and two robust weighted estimators, $\widehat{\beta}_{GM}$ and $\widehat{\beta}_{HR}$. The robust estimators were computed as described in Remark 3.2 choosing as score function ϕ the Tukey's function. Besides, for all robust estimators, the tuning constant was chosen equal and such that $\widehat{\beta}_M$ attains an efficiency $e = 0.90$ with respect to the maximum likelihood estimator. For both weighted robust estimators, the weights were computed over the robust Mahalanobis distances based on an S-estimator, $(\widehat{\mu}_x, \widehat{\Sigma}_x)$, with breakdown point 0.5 using 1000 sub-samples. For $\widehat{\beta}_{GM}$, we used the Tukey's bisquare weight function with tuning constant $c = \chi_{k, 0.95}^2$, while for $\widehat{\beta}_{HR}$ a hard rejection weight function was considered, that is, $w_1(\mathbf{x}) = 1$ when $(\mathbf{x} - \widehat{\mu}_x)^T \widehat{\Sigma}_x^{-1} (\mathbf{x} - \widehat{\mu}_x) \leq \chi_{k, 0.95}^2$ and $w_1(\mathbf{x}) = 0$ elsewhere.

From now on, we denote by \widehat{W}_{ML} , \widehat{W}_M , \widehat{W}_{GM} and \widehat{W}_{HR} the tests based on $\widehat{\beta}_{ML}$, $\widehat{\beta}_M$, $\widehat{\beta}_{GM}$ and $\widehat{\beta}_{HR}$, respectively.

Table 1

Observed frequencies of rejection in the non-contaminated case C_0 with $p \equiv 1$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\mathbf{x}^T\boldsymbol{\lambda} + 0.4))^2$ and $p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T\boldsymbol{\lambda} - 2))$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$.

n	$p \equiv 1$				$p(\mathbf{x}) = 0.4 + 0.5(\cos(\mathbf{x}^T\boldsymbol{\lambda} + 0.4))^2$				$p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T\boldsymbol{\lambda} - 2))$			
	50	100	250	500	50	100	250	500	50	100	250	500
$\widehat{\mathcal{W}}_{ML}$	0.114	0.097	0.062	0.051	0.174	0.114	0.081	0.061	0.163	0.111	0.068	0.047
$\widehat{\mathcal{W}}_M$	0.136	0.101	0.066	0.061	0.253	0.142	0.074	0.063	0.233	0.121	0.053	0.051
$\widehat{\mathcal{W}}_{GM}$	0.121	0.084	0.055	0.053	0.207	0.110	0.063	0.049	0.192	0.108	0.062	0.046
$\widehat{\mathcal{W}}_{HR}$	0.147	0.091	0.076	0.052	0.236	0.141	0.077	0.052	0.203	0.104	0.060	0.053

We considered three models for the missing probability $p \equiv 1$, $p(\mathbf{x}) = 0.4 + 0.5(\cos(\mathbf{x}^T\boldsymbol{\lambda} + 0.4))^2$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$ and $p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T\boldsymbol{\lambda} - 2))$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$, i.e., a logistic model for the missing probability. In order to study the convergence speed to the χ^2 distribution, Table 1 reports the observed frequencies of rejection in the non-contaminated case C_0 , for different sample sizes $n = 50, 100, 250$ and 500 for a nominal level 5%. In all cases the convergence is quite slow. Note that the hard rejection test $\widehat{\mathcal{W}}_{HR}$ has a similar performance to the situation in which Tukey's weights are used to downweight high leverage points. For that reason, only the results corresponding to $\widehat{\mathcal{W}}_{GM}$ are reported for contaminated samples.

Figs. 1 and 2 allow us to study the power performance of the tests when $n = 100$ for $p(\mathbf{x}) \equiv 1$ and for the logistic missingness model $p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T\boldsymbol{\lambda} - 2))$. Therein, we have plotted the observed frequencies of rejection under the null hypothesis $H_0 : \boldsymbol{\beta} = \mathbf{0}_3$ and the alternatives $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + \Delta n^{-1/2}(1, 0, 0)^T$ with $\Delta = \pm 6, \pm 4.8, \pm 3.6, \pm 2.4, \pm 1.2, \pm 0.8, \pm 0.4$ and ± 0.2 . The solid (black), broken (green) and dashed (red) lines correspond to the test statistics $\widehat{\mathcal{W}}_{ML}$, $\widehat{\mathcal{W}}_M$, $\widehat{\mathcal{W}}_{GM}$, respectively, while the horizontal broken (- -) line corresponds to the nominal level 0.05. Fig. S.1 in the supplementary file reports the observed frequency of rejection when $p(\mathbf{x}) = 0.4 + 0.5(\cos(\mathbf{x}^T\boldsymbol{\lambda} + 0.4))^2$.

As expected, under C_0 the test procedures based on classical or robust estimators perform quite similarly, under all the missing schemes. For large values of m and/or x_0 , the classical procedure is non-informative. When considering moderate vertical outliers, we observe some loss of power and level for the $\widehat{\mathcal{W}}_{GM}$, in particular, when missing responses arise, while the level of the classical procedure $\widehat{\mathcal{W}}_{ML}$ is sensitive to large values of the responses ($m = 2.5$) combined with $x_0 = 1$. On the other hand, the test based on $\widehat{\mathcal{W}}_M$ shows its sensitivity to moderate outliers ($m = 0.5$ and $x_0 = 3$) for all the missingness models. This test is also sensitive to extreme outliers ($x_0 = 10$) when $p \equiv 1$, while for a logistic missingness model, it shows a poor performance under $C_{0.5,10}$ probably due to the effect of high leverage points on the estimation of the asymptotic covariance matrix. Their weighted versions are more stable with respect to all the contaminations considered. It is worth noting that the situation with vertical outliers is not as harmful as when high leverage points arise.

5. Empirical breakdown point

A popular measure of robustness is the finite sample breakdown point (BDP). For test statistics the level and power BDP should be distinguished (see, for instance He et al., 1990). Indeed, it is desirable to keep a stable level and a good power under small departures from specified alternative values. As expected and according to the results reported in Section 4, the classical test statistics are sensitive to atypical observations, so they will breakdown both in level and power with just few atypical observation. It is well known that in some situations, estimators with the maximum possible breakdown point (0.5) may lead to test statistics that may breakdown with smaller amounts of contaminations due to the performance of the estimated asymptotic covariance matrix. This phenomenon was illustrated in Bianco and Martínez (2009) and also in the results obtained in Section 4 where the Wald tests, $\widehat{\mathcal{W}}_M$, based on the unweighted robust estimators show their sensitivity when a 5% of moderate outliers ($m = 0.5$ and $x_0 = 3$) are included in the sample.

To have a deeper comprehension of the test performance, we begin by investigating in the complete data case, i.e. $p \equiv 1$, the sensitivity to outliers of both robust Wald tests, $\widehat{\mathcal{W}}_M$ and $\widehat{\mathcal{W}}_{GM}$, when different amounts/sizes of contamination are included in the sample. This allows us to get some insight into its finite sample breakdown point. To be more precise, we conducted a simulation study with a similar scheme of that used in Section 4, but with different amounts of contamination included in the sample. As before, the contaminated samples, include n_0 outliers all equal, say (y_0, \mathbf{x}_0^T) where $\mathbf{x}_0^T = (x_0, 0, 1)$ and $y_0 = \exp(m x_0)$. We choose $n_0 = 5, 10, 15, 20$ corresponding to 5%, 10%, 15% and 20% of contamination. Figs. S.2 and S.3 in the supplementary file report the observed frequencies of rejection of the Wald test statistics $\widehat{\mathcal{W}}_{ML}$ and $\widehat{\mathcal{W}}_M$, while Fig. 3 reports the same quantities for $\widehat{\mathcal{W}}_{GM}$ when $p \equiv 1$. Besides, Figs. 4 and S.4 in the supplementary file correspond to $\widehat{\mathcal{W}}_{GM}$ when the same missing probabilities studied in Section 4 are included in the sample. In all figures, the solid lines in black correspond to the uncontaminated samples, while the solid line with triangles (blue), the broken line (green), the dashed line (red) and the solid one with squares (purple) correspond to samples containing 5, 10, 15 and 20 outliers, respectively.

Fig. S.2 in the supplementary file shows the severe effect of different contaminations on the behaviour of the classical test, perhaps, with the exception of scheme $C_{0.5,1}$ where contamination percentages smaller than 15% are less harmful. It is also apparent from these plots that in some contamination schemes, the symmetric pattern of the power function of the classical test is completely altered even with 5 outliers. On the other hand, from Fig. S.3 in the supplementary file and Fig. 3, we can conclude that the statistic $\widehat{\mathcal{W}}_{GM}$ leads to a much more stable power function than $\widehat{\mathcal{W}}_M$. Only under contamination

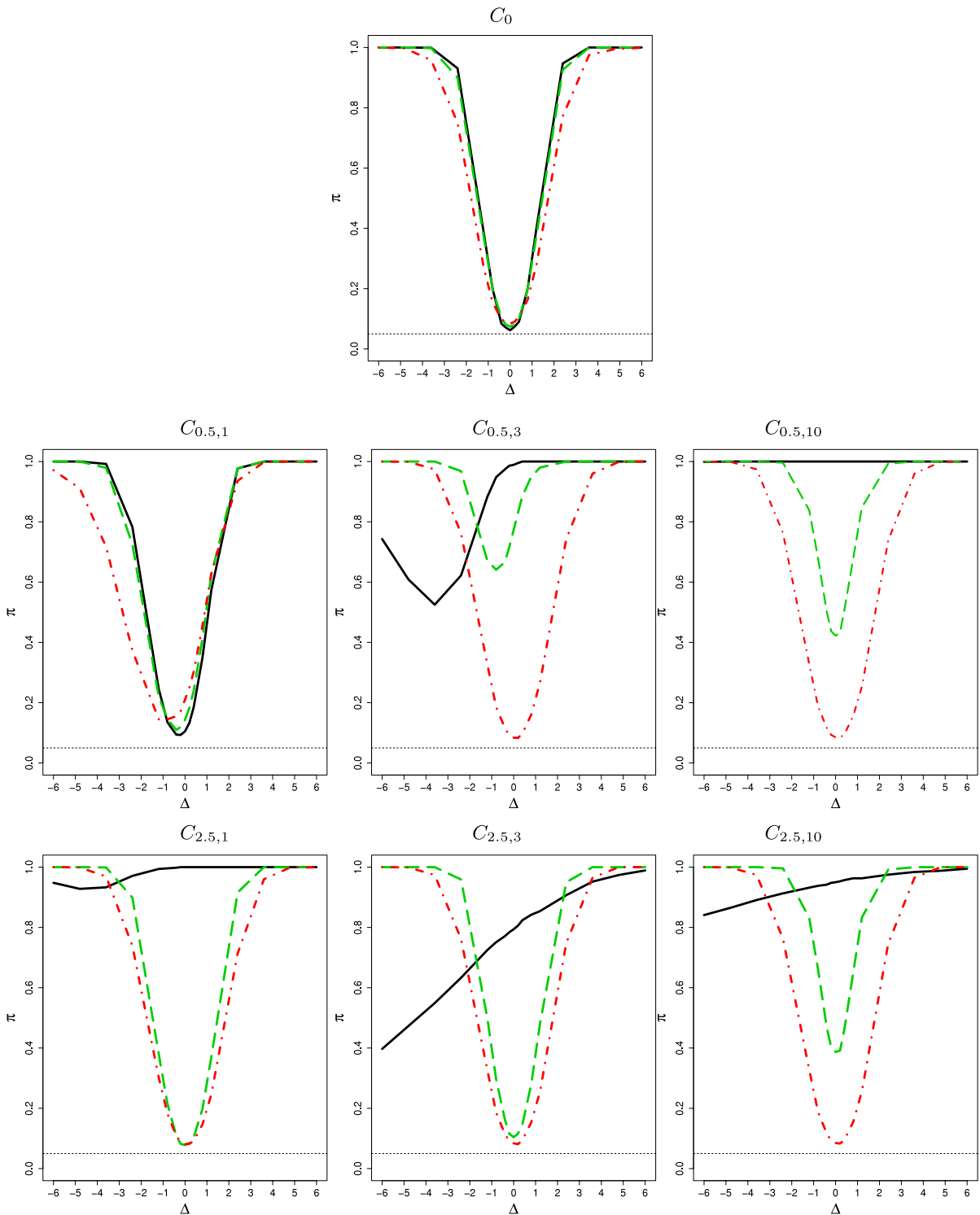


Fig. 1. Observed frequencies of rejection under the Gamma model when $p(\mathbf{x}) = 1$. The solid, broken and dashed lines correspond to the tests based on \widehat{W}_{ML} , \widehat{W}_M and \widehat{W}_{GM} , respectively.

schemes $C_{0.5,1}$ and $C_{0.5,3}$ the power of \widehat{W}_{GM} shows some instability for large percentages of contamination, mainly due to the fact that, in these two contaminations, the mild vertical outliers combined with low or moderate leverage points are hard to be controlled. It is also evident, from these two figures, the different performance of the robust tests for the considered percentages of contamination: one can explain these differences in terms of their covariance matrices since in the case of

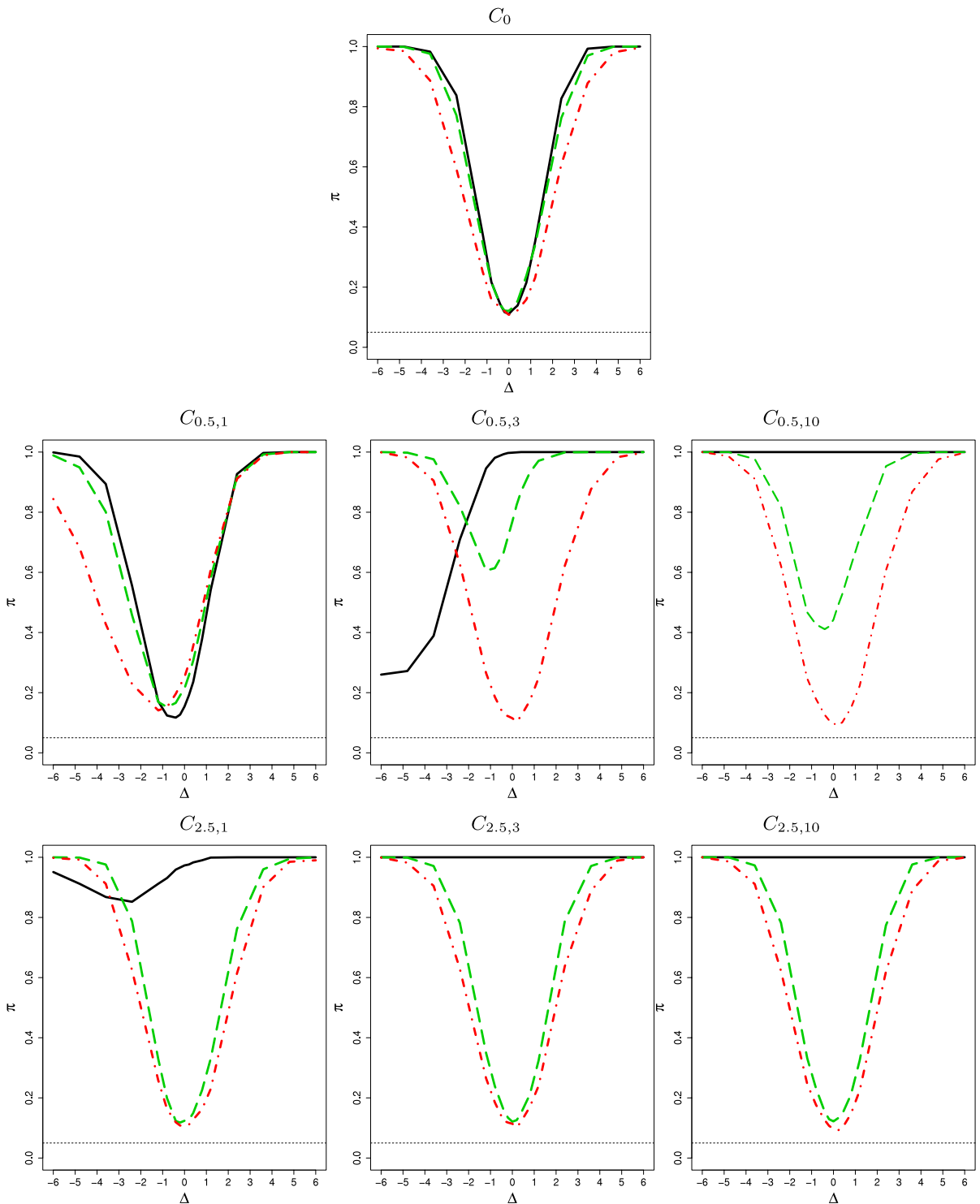


Fig. 2. Observed frequencies of rejection under the Gamma model when $p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T \boldsymbol{\lambda} - 2))$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$. The solid, broken and dashed lines correspond to the tests based on $\widehat{\mathcal{W}}_{ML}$, $\widehat{\mathcal{W}}_M$ and $\widehat{\mathcal{W}}_{GM}$, respectively.

the $\widehat{\mathcal{W}}_{GM}$ statistic the leverage is downweighted by the weights and hence, the covariance matrix is controlled. Even if the performance of $\widehat{\mathcal{W}}_M$ looks satisfactory in power under $C_{0.5,1}$, except for 20 outliers, its performance is very poor both in level and power when 5 outliers are introduced in the sample under $C_{0.5,3}$, $C_{0.5,10}$ and $C_{2.5,10}$. In this sense, the plots suggest that $\widehat{\mathcal{W}}_{GM}$ may manage without any problems 5 outliers, while some level sensitivity is observed for 10 outliers under $C_{0.5,1}$

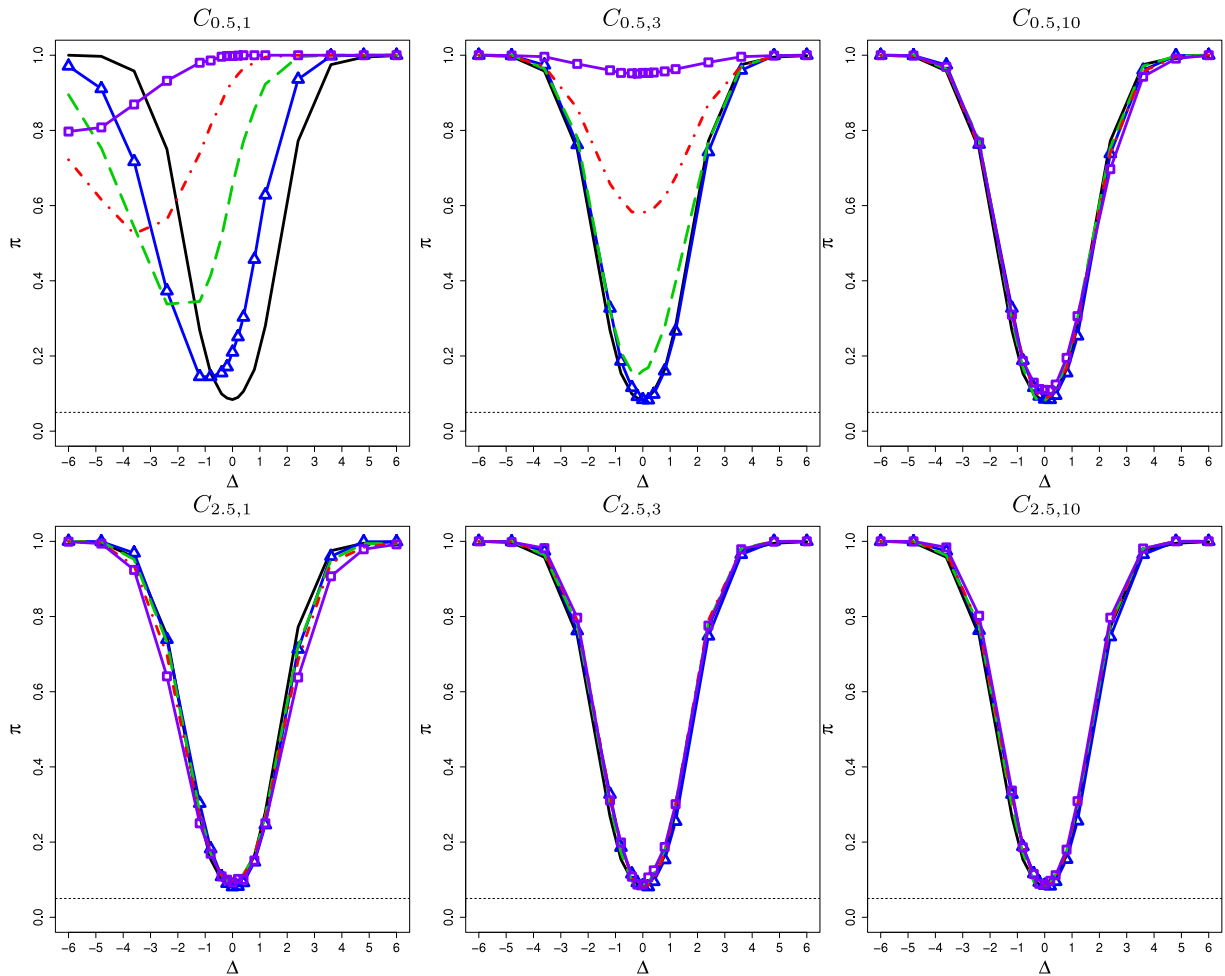


Fig. 3. Observed frequencies of rejection under the Gamma model for the test based on \widehat{W}_{GM} when $p(\mathbf{x}) = 1$. The solid lines (black) correspond to the uncontaminated samples, while the solid line with triangles (blue), the broken (green), dashed (red) and solid line with squares (purple) correspond to samples containing 5, 10, 15 and 20 outliers, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and $C_{0.5,3}$. The test handles without any trouble, up to 20 outliers, with large vertical residuals or large leverage points (see the plots for $C_{0.5,10}$ and those $C_{2.5,x_0}$ with $x_0 = 1, 3$ and 10).

Figs. 4 and S.4 in the supplementary file summarize the performance of the Wald test statistic \widehat{W}_{GM} under the logistic and cosine missing probability models. As it can be seen the main features of these plots resemble those obtained for $p \equiv 1$, except for samples with 20% of contamination, where some loss in the level is observed in some situations.

6. Examples

6.1. Hospital cost data

Marazzi and Yohai (2004) introduced a data set that corresponds to the costs of 100 patients in a Swiss hospital in 1999 for *medical back problems*. They concerned on the relationship between the hospital cost of stay, y , (Cost, in Swiss francs) and the following administrative explanatory variables:

- *LOS*: length of stay in days
- *ADM*: admission type (0 = planned; 1 = emergency)
- *INS*: insurance type (0 = regular; 1 = private)
- *AGE*: years
- *SEX*: (0 = female; 1 = male)
- *DEST*: discharge destination (1 = home; 0 = other).

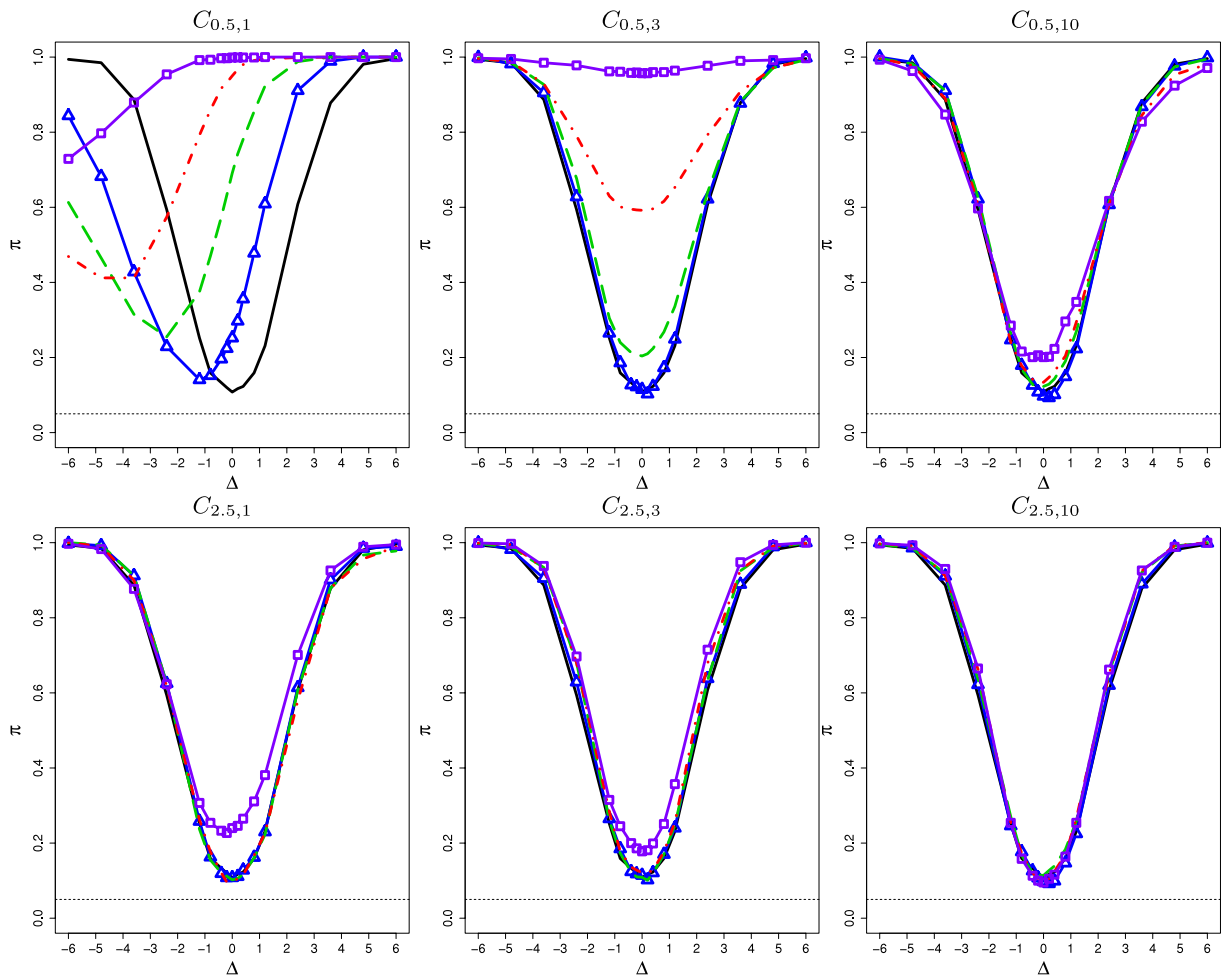


Fig. 4. Observed frequencies of rejection under the Gamma model for the test based on \widehat{w}_{GM} when $p(\mathbf{x}) = 1/(1 + \exp(-\mathbf{x}^T \boldsymbol{\lambda} - 2))$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$. The solid lines (black) correspond to the uncontaminated samples, while those solid line with triangles (blue), the broken (green), dashed (red) and solid line with squares (purple) correspond to samples containing 5, 10, 15 and 20 outliers, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Cantoni and Ronchetti (2006) fitted to the complete data set the model $\log(\mathbb{E}(y_i|\mathbf{x}_i)) = \boldsymbol{\beta}^T \mathbf{x}_i$ which for Gamma responses is equivalent to

$$\log(y_i) = \boldsymbol{\beta}^T \mathbf{x}_i + u_i,$$

where u_i has $\log \Gamma(\tau, 1)$ and $\mathbf{x} = (\log \text{LOS}, \text{ADM}, \text{INS}, \text{AGE}, \text{SEX}, \text{DEST}, 1)$. Using their robust proposal, they identified 5 outliers corresponding to observations labelled as 14, 21, 28, 44 and 63, whose weights are less or equal than 0.5. They realized that the atypical points affected the classical estimates of the coefficient of variable *INS* and the shape parameter. In particular, the effect of the outliers on the classical estimate of the shape parameter is remarkable since it achieved almost half the value obtained with the robust method.

As in the simulation study, we compute $\widehat{\boldsymbol{\beta}}_{ML}$, $\widehat{\boldsymbol{\beta}}_M$ and $\widehat{\boldsymbol{\beta}}_{GM}$ for the complete data set and also the maximum likelihood estimator without the 5 outlying observations, $\widehat{\boldsymbol{\beta}}_{ML}^{(-5)}$, and the corresponding estimators of the shape parameter τ . Table 2 summarized the obtained estimators including the p -values for testing the significance of the covariates *SEX* and *DEST*. The obtained estimators are analogous to those obtained by Cantoni and Ronchetti (2006). Moreover, the value of $\widehat{\boldsymbol{\beta}}_{ML}^{(-5)}$ and the related estimator of τ are very similar to those obtained using $\widehat{\boldsymbol{\beta}}_{GM}$, showing its good performance in the presence of outliers. With respect to the test, it is worth noticing that the classical Wald test \widehat{w}_{ML} does not reject the null hypothesis of no effect of *SEX* or *DEST* at 5% significance level, while from $\widehat{w}_{ML}^{(-5)}$, the classical test statistic obtained after removing the 5 outliers, the conclusion is reversed. On the other hand, as the classical procedure, the unweighted procedure \widehat{w}_M does not reject the null hypothesis $H_0 : \beta_5 = 0$, while the Wald tests based on the weighted robust estimators reject the null hypothesis as the classical method does after removing the outliers. For the variable *DEST*, both the weighted and unweighted procedures lead to the same conclusions.

Table 2
Analysis of hospital cost data.

	Estimated coeff.				$H_0 : \beta_j = 0$ p-value, 5% level			
	$\hat{\beta}_{ML}$	$\hat{\beta}_{ML}^{[5]}$	$\hat{\beta}_M$	$\hat{\beta}_{GM}$	\hat{W}_{ML}	$\hat{W}_{ML}^{[5]}$	\hat{W}_M	\hat{W}_{GM}
log LOS	0.8218	0.8473	0.8640	0.8892				
ADM	0.2132	0.2151	0.2576	0.2375				
INS	0.0960	-0.0235	-0.0523	-0.0437				
AGE	-0.0005	-0.0015	-0.0009	-0.0010				
SEX	0.0954	0.0706	0.0489	0.0739	0.0561	0.0410	0.1620	0.0222
DEST	-0.1040	-0.1413	-0.1024	-0.1225	0.1331	0.0026	0.0344	0.0064
Intercept	7.2331	7.2764	7.1796	7.1268				
$\hat{\tau}$	20.1876	44.2838	48.9791	41.1086				

Table 3
Analysis of hospital cost data with missing responses.

	$p(\mathbf{x}) = 1/(1 + \exp(0.2 \log LOS - 2))$			$H_0 : \beta_j = 0$ p-value, 5% level		
	Estimated coeff.			\hat{W}_{ML}	\hat{W}_M	\hat{W}_{GM}
	$\hat{\beta}_{ML}$	$\hat{\beta}_M$	$\hat{\beta}_{GM}$			
log LOS	0.8048	0.8194	0.8500			
ADM	0.1891	0.2087	0.1852			
INS	0.1487	-0.0021	-0.0106			
AGE	-0.0010	-0.0014	-0.0012			
SEX	0.0905	0.0711	0.0906	0.0990	0.0562	0.0099
DEST	-0.0998	-0.1214	-0.1518	0.0961	0.0243	0.0024
Intercept	7.3009	7.3259	7.2672			
$\hat{\tau}$	17.3636	35.2136	41.8678			

To evaluate the performance of the proposed estimators and test statistics for incomplete data sets, we introduced artificially missing data to this example and we took the above analysis as a natural counterpart. Missing responses were introduced at random among the non-outlying points reported by [Cantoni and Ronchetti \(2006\)](#) according to the logistic scheme $p(\mathbf{x}) = 1/(1 + \exp(0.2 \log LOS - 2))$ resulting in about 20% of missing responses. The estimators and test statistics were computed over this new sample which includes missing responses. [Table 3](#) reports the obtained results. Comparing with the counterpart estimates based on $\hat{\beta}_{ML}^{[5]}$ reported in [Table 2](#), we observe that, for the new sample, all the estimates based on $\hat{\beta}_{GM}$ remain very stable and close to these values, while the maximum likelihood estimates and the estimates based on $\hat{\beta}_M$ are farther away; in particular, the estimators of the coefficient of *INS* and τ . Moreover, for the new sample, according to the robust test statistic \hat{W}_{GM} the variable *SEX* is significant at 5% level, while both the classical test and \hat{W}_M lead to the opposite conclusion. With respect to *DEST*, both robust Wald statistics lead to the same conclusion rejecting the null hypothesis, while the classical statistic does not reject it, as when no missing responses arise. Besides, for this new sample with missing responses, we recompute the Wald test based on the ML-estimator after removing the 5 outliers identified by [Cantoni and Ronchetti \(2006\)](#). The results obtained for the p-values are now consistent with those given by \hat{W}_{GM} . Indeed, the tests lead now to the same conclusion obtained when no missing responses arise. As with the situation with no missing responses, we can see the benefits of introducing weights in order to avoid the effect of high leverage outlying data in the presence of missing responses.

As in [Croux et al. \(2008\)](#) and in [Bianco and Martínez \(2009\)](#), in order to assess the effect of contamination on the decision, one observation is added to the data at position (\mathbf{x}, y) with $\mathbf{x} = (s, 0, 0, 56, 1, 0, 1)^T$ and $y = \exp(sm)$, for values of (m, s) in $[-3, 3] \times [-10, 0]$ by steps of 0.1. [Fig. 5](#) shows the p-values of the classical and the robust tests for the null hypothesis $H_0 : \beta_5 = 0$, for the different values of s and m together with its contour plot beneath the surface. The function $\xi(m) = -7.5|m + 3|^3/3.2^3 - 1.7$ is plotted also in the different surface plots as a red line with diamonds. This function was fitted to provide an approximation of the region where the shape of the p-values change. The p-values corresponding to the classical test \hat{W}_{ML} and also to the test based on the robust estimators with $w_1 \equiv 1$, \hat{W}_M , increase rapidly along a curve. Moreover, for large or moderate negative values of s , the Wald test based on the maximum likelihood estimators changes the decision from non rejection to rejection when m moves from positive to negative values. On the other hand, the p-values of the test statistic \hat{W}_M have a lower bound which corresponds to the effect of the 5 outliers as noted in [Table 2](#). On the contrary, \hat{W}_{GM} is very stable and only shows a very small sensitivity along a curve without changing the decision. To explain the sensitivity along this curve which is close to $(m, \xi(m))$, [Fig. S.5](#) in the supplementary file shows surface plots for the robust test p-values together with the values of $\log(\sqrt{d^*(y, \mathbf{x}, \hat{\beta}_{GM})})$ and those of the estimators of the *SEX* component, $\hat{\beta}_{5,GM}$ when $\mathbf{x} = (s, 0, 0, 56, 1, 0, 1)^T$ and $y = \exp(sm)$. The flat green surface in the plot corresponding to $\log(\sqrt{d^*(y, \mathbf{x}, \hat{\beta}_{GM})})$

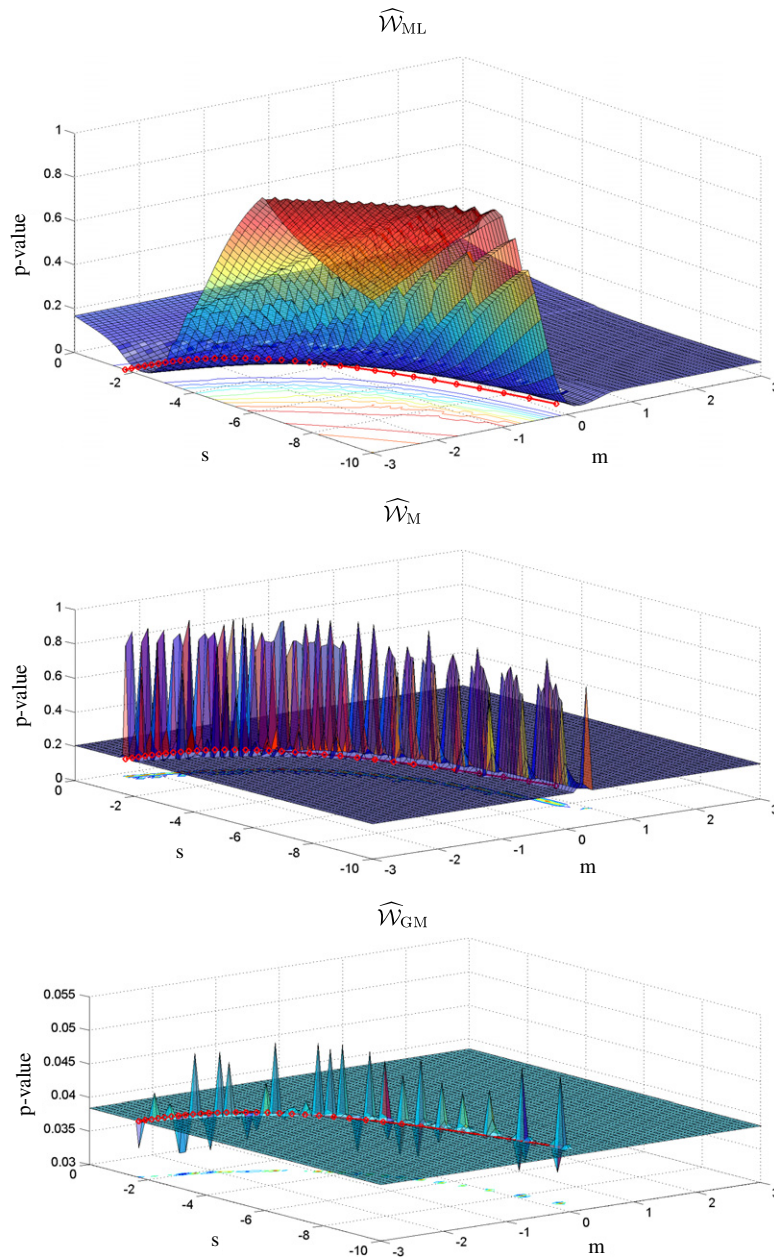


Fig. 5. Sensitivity of the p -value to an outlier located at (\mathbf{x}, y) , with $\mathbf{x} = (s, 0, 0, 56, 1, 0, 1)^T$ and $y = \exp(sm)$ when testing the significance of the variable SEX for the Hospital data set.

represents the logarithm of the tuning constant, $\log(\widehat{c}_n)$, used in Step 3, which is almost constant and close to $\log(0.3722)$, when (s, m) varies. Note that the robust estimators truncate the observations such that $\sqrt{d^*(y, \mathbf{x}, \widehat{\beta}_{GM})} > \widehat{c}_n$. Thus, the region where the surface $\log(\sqrt{d^*(y, \mathbf{x}, \widehat{\beta}_{GM})})$ crosses the plane $\{(m, s, v) : v = \log(\widehat{c}_n)\}$ correspond to mild outliers observations (\mathbf{x}, y) with $\mathbf{x} = (s, 0, 0, 56, 1, 0, 1)^T$ and $y = \exp(sm)$ that can be suspected as atypical by the bounding deviance procedure.

6.2. Leukaemia data

The data of Feigl and Zelen (1965) represent the survivorship of 33 patients of acute myelogenous leukaemia divided in two groups, that correspond to a factor variable AG which classifies the patients as positive or negative depending on the presence or absence of a morphological characteristic in the white cells. The original data are time at death and also the

Table 4
Analysis of Feigl and Zelen data. Complete data set.

	Estimated coefficients				$H_0 : \beta_1 = 0$ p-value			
	$\hat{\beta}_{ML}$	$\hat{\beta}_{ML}^{(-4)}$	$\hat{\beta}_M$	$\hat{\beta}_{GM}$	\hat{w}_{ML}	$\hat{w}_{ML}^{(-4)}$	\hat{w}_M	\hat{w}_{GM}
$\frac{WBC}{1000}$	-0.007	-0.051	-0.051	-0.089	0.2151	0	0	0.0001
AG	-1.101	-1.574	-1.802	-1.510				
Intercept	4.227	4.795	4.849	5.101				

Table 5
Analysis of Feigl and Zelen data with two missing probabilities.

	$p(\mathbf{x}) = 0.9$			$H_0 : \beta_1 = 0$ p-value		
	$\hat{\beta}_{ML}$	$\hat{\beta}_M$	$\hat{\beta}_{GM}$	\hat{w}_{ML}	\hat{w}_M	\hat{w}_{GM}
$\frac{WBC}{1000}$	-0.008	-0.050	-0.084	0.1477	0	0.0001
AG	-0.974	-1.469	-1.364			
Intercept	4.333	4.841	5.055			
	$p(\mathbf{x}) = 1/(1 + \exp(0.2 WBC - 4))$			$H_0 : \beta_1 = 0$ p-value		
$\frac{WBC}{1000}$	-0.0012	-0.0004	-0.1210	0.8000	0.9375	0.0000
AG	-1.3718	-1.4371	-1.4315			
Intercept	4.4432	4.5055	5.2617			

white blood cells count WBC , which is a useful tool for diagnosing the initial condition of the patient; indeed higher counts seem to be associated with more severe conditions. Bianco et al. (2005) fit, to the complete data set, the model

$$\log(y_i) = \beta_1 WBC_i + \beta_2 AG_i + \beta_3 + u_i,$$

where u_i has $\log \Gamma(\tau_0, 1)$ distribution using an m -estimator based on deviance. The QQ-plot of the residuals of the m -estimate computed by Bianco et al. (2005) reveals 4 clear outliers corresponding to patients with very high values of WBC who survived more than expected.

In this example, since AG is a factor variable, when computing $\hat{\beta}_{GM}$ the weights $w_1(\mathbf{x})$ were based only on the variable WBC and the tuning constant was chosen as $c = \chi_{1,0.95}^2$. The robust Mahalanobis distance of WBC equals in this case $|WBC_i - \text{median}_i(WBC_i)|/\text{MAD}(WBC_i)$.

Table 4 reports the values of $\hat{\beta}_{ML}$, $\hat{\beta}_M$, $\hat{\beta}_{GM}$ and the classical estimate without the 4 outliers, $\hat{\beta}_{ML}^{(-4)}$, together with the p -values of the related Wald-type statistics to check $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$. It is worth noticing that according to the results given by $\hat{\beta}_{ML}$, the coefficient of the variable WBC is non-significant at a 5% significance level. On the other hand, the ML -estimator applied to the sample without the 4 outliers and the robust estimators lead to the opposite conclusion.

As in Ibrahim (1990), in order to evaluate the performance of the proposed tests for incomplete data sets, we introduced artificially missing data to this example and we took the above analysis as a natural counterpart. Missing responses among the non-outlying points were introduced at random according to two missing schemes, a completely at random situation with $p(\mathbf{x}) = 0.9$ and a missing at random case with logistic probability of missing $p(\mathbf{x}) = 1/(1 + \exp(0.2 WBC - 4))$. In this way, for the logistic case, 8 responses (almost 25% of the data) result in missing observations. The analysis was repeated for each of the obtained samples. In Table 5 we summarize the corresponding results. Different conclusions are derived depending on the missing scheme. As expected, when missing responses occur completely at random, analogous results to those obtained with the complete data set are obtained. In this sense, when $p(\mathbf{x}) = 0.9$, according to the computed robust tests the variable WBC is significant, while from the ML -estimator we conclude otherwise. On the other hand, for the incomplete sample obtained through a logistic missing probability, the estimators $\hat{\beta}_{ML}$ and $\hat{\beta}_M$ take similar values. Moreover, according to the robust test statistic \hat{w}_{GM} , based on Tukey’s weights, the variable WBC is significant, while both the classical test and \hat{w}_M lead to the opposite conclusion. Besides, for this sample with missing responses, we recompute the Wald test based on the ML -estimator after removing the 4 outliers. The results obtained for the p -values are now consistent with those given by \hat{w}_{GM} . Indeed, the tests lead again to the same conclusion obtained when no missing responses arise.

As in Section 6.1 a sensitivity analysis is conducted, but this time to measure the influence of the missing probability on the resulting p -value. We considered a logistic probability of missing $p(\mathbf{x}) = 1/(1 + \exp(a WBC - b))$ where $(a, b) \in [0, 0.20] \times [4, 10]$ among the non-outlying responses. The grid step was 0.01 for a and 0.5 for b . In this way, at most 25% of the data result in missing observations. We compute the p -values of the classical and the robust tests for the null hypothesis $H_0 : \beta_1 = 0$, for the different values of a and b . The surface plots in Fig. 6 show clearly that the p -values corresponding to the robust test \hat{w}_{GM} are more stable leading to p -values almost 0 except when the probability of missing results in more than 20% of missing responses in which the p -value rapidly grows to 0.00026 without changing the decision. As in the previous example, the p -values corresponding to \hat{w}_{ML} and \hat{w}_M increase rapidly and attain values close to 1. Moreover, for the Wald

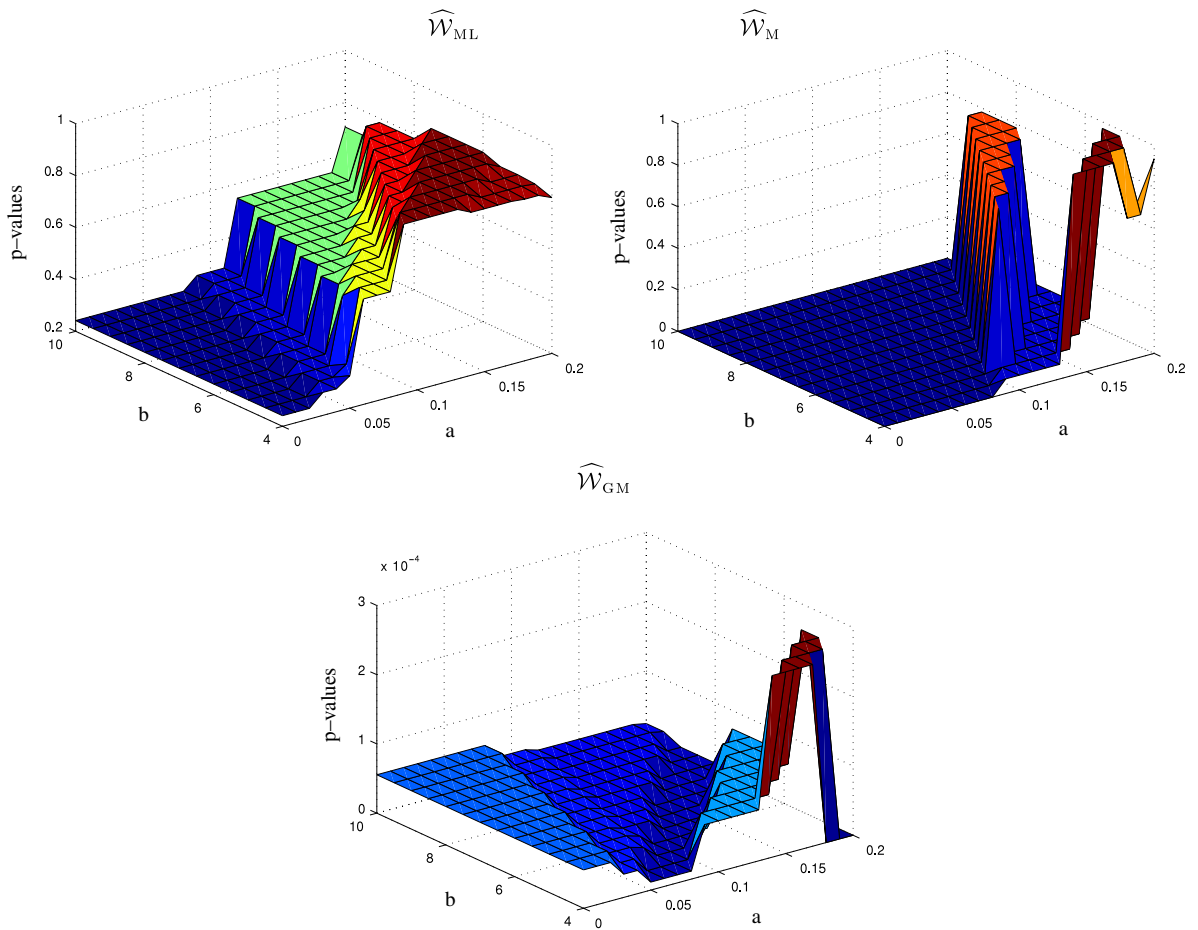


Fig. 6. Sensitivity of the p -value under the logistic missing model $p(\mathbf{x}) = 1/(1 + \exp(aWBC - b))$, when testing the significance of the variable WBC for the Leukaemia data set.

test based on the maximum likelihood estimators has p -values always larger than 0.2413 and smaller than 0.8529, while for the Wald test based on $\hat{\beta}_M$ the p -values move from 0 to 0.9688 changing the decision from rejection to no rejection.

These results show the advantage of introducing weights as a useful tool to prevent from outlying points under the different missing schemes considered.

7. Influence functions of the test functionals

Influence functions are measures of robustness with respect to single outliers. The influence function allows us to study the local robustness and the asymptotic efficiency of the estimators, providing a rationale for choosing appropriate weight functions and tuning parameters. It can be thought as the first derivative of the functional version of the estimator. The influence function of a functional $T(F)$ is defined as $IF(\mathbf{z}_0, T, F) = \lim_{\epsilon \rightarrow 0} (T(F_{\mathbf{z}_0, \epsilon}) - T(F))/\epsilon$, where $F_{\mathbf{z}_0, \epsilon} = (1 - \epsilon)F + \epsilon \Delta_{\mathbf{z}_0}$ and $\Delta_{\mathbf{z}_0}$ denotes the probability measure which puts mass 1 at the point $\mathbf{z}_0 = (y_0, \mathbf{x}_0^T, \delta_0)$ and represents the contaminated model.

The examples given in Section 6 show the importance of measuring the effect of a single outlier on the p -value. Rieder (1978) studied the effect of contamination on level and power, while a link between the influence function of a test statistic and its power and level influence functions is given in Hampel et al. (1986).

For that reason, in this section, we derive the influence function of the Wald test statistic. We will also give an appropriate definition that takes into account the missing responses.

For any distribution F_1 , let $\mathbf{V}(F_1)$ be a Fisher-consistent scatter functional at F , i.e., such that $\mathbf{V}(F) = \Sigma$.

Denote by $\beta(F_1)$ and $\tau(F_1)$ the functionals related to the estimators $\hat{\beta}$ and $\hat{\tau}$, respectively, and assume that $\beta(F_1)$, the solution of $S^{(1)}(\beta(F_1), \tau(F_1)) = \mathbb{E}_{F_1} (\delta\Psi(y, \mathbf{x}^T\beta(F_1), \tau(F_1)) w_1(\mathbf{x})\mathbf{x}) = \mathbf{0}_k$, is a Fisher-consistent functional at F , i.e., $\beta(F) = \beta$. Bianco et al. (2011a) derived the influence function of the regression functional $\beta(F_1)$ at F and showed that, under N3, N4 and N6, $IF(\mathbf{z}_0, \beta, F)$ exists and, when $\tau(F) = \tau$,

$$IF(\mathbf{z}_0, \beta, F) = -\Psi(y_0, \mathbf{x}_0^T\beta, \tau) w_1(\mathbf{x}_0)\delta_0\mathbf{A}^{-1}\mathbf{x}_0. \tag{11}$$

To obtain the influence function of the test, define the functionals

$$\begin{aligned} \mathbf{A}(F_1) &= \mathbb{E}_{F_1} (\delta \chi (y, \mathbf{x}^T \boldsymbol{\beta}(F_1), \tau(F_1)) w_1(\mathbf{x}) \mathbf{x} \mathbf{x}^T) \\ \mathbf{B}(F_1) &= \mathbb{E}_{F_1} (\delta \Psi^2 (y, \mathbf{x}^T \boldsymbol{\beta}(F_1), \tau(F_1)) w_1^2(\mathbf{x}) \mathbf{x} \mathbf{x}^T). \end{aligned}$$

The Wald-type test functional related to the statistic used to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is given by $\mathcal{W}(F_1) = (\boldsymbol{\beta}(F_1) - \boldsymbol{\beta}_0)^T \mathbf{V}(F_1)^{-1} (\boldsymbol{\beta}(F_1) - \boldsymbol{\beta}_0)$. It is easy to see that, under H_0 , $\text{IF}(\mathbf{x}, \mathcal{W}, F) = 0$.

As in Hampel et al. (1986), in order to obtain a non-null influence function related to the level and power influence, we consider the square root of the test statistics, $\mathcal{T}(F_1) = \mathcal{W}(F_1)^{1/2}$. As for the linear model, using that, $\mathcal{W}(F) = 0$ under H_0 , we have that $\text{IF}(\mathbf{x}, \mathcal{T}, F) = \left\{ (1/2) \partial^2 \mathcal{W}(F_{\mathbf{x}, \epsilon}) / \partial \epsilon^2 |_{\epsilon=0} \right\}^{1/2}$. The following theorem gives the value of the influence function of the test functional $\mathcal{T}(F)$.

Theorem 7.1. Let $\mathbf{V}(F_1)$ be a scatter functional such that $\mathbf{V}(F) = \boldsymbol{\Sigma} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$, where $\mathbf{A} = \mathbf{A}(F)$ and $\mathbf{B} = \mathbf{B}(F)$ are the symmetric matrices defined in (6) and (7). Assume that the influence function $\text{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F)$ and that $\partial^2 \boldsymbol{\beta}(F_{\mathbf{x}, \epsilon}) / \partial \epsilon^2 |_{\epsilon=0}$ exist. Then, the influence function at F of the functional $\mathcal{T}(F_1)$ to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is given by

$$\text{IF}(\mathbf{z}_0, \mathcal{T}, F)^2 = \text{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F)^T \boldsymbol{\Sigma}^{-1} \text{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F). \tag{12}$$

Besides, under N3, N4 and N6, we have

$$\text{IF}(\mathbf{z}_0, \mathcal{T}, F)^2 = \Psi^2 (y_0, \mathbf{x}_0^T \boldsymbol{\beta}(F), \tau(F)) w_1^2(\mathbf{x}_0) \delta_0 \mathbf{x}_0^T \mathbf{B}^{-1} \mathbf{x}_0. \tag{13}$$

Remark 7.1. It is worth noticing that the influence function depends on the indicator of the missing response δ_0 so, it will be 0 if no responses arise. A more reliable function to measure the sensitivity to outliers of a given functional $T(F_1)$ under a missing scheme may be to consider the *expected influence function* of an observed data $\mathbf{z}_0^* = (y_0, \mathbf{x}_0^T)^T$, denoted as $\text{EIF}(\mathbf{z}_0^*, T, F)$, i.e., $\text{EIF}(\mathbf{z}_0^*, T, F) = \mathbb{E}(\text{IF}(\mathbf{z}_0, T, F) | (y_0, \mathbf{x}_0))$. For the functionals under study, we have that

$$\begin{aligned} \text{EIF}(\mathbf{z}_0^*, \boldsymbol{\beta}, F) &= -\Psi (y_0, \mathbf{x}_0^T \boldsymbol{\beta}, \tau) w_1(\mathbf{x}_0) p(\mathbf{x}_0) \mathbf{A}^{-1} \mathbf{x}_0 \\ \text{EIF}(\mathbf{z}_0^*, \mathcal{T}, F)^2 &= \Psi^2 (y_0, \mathbf{x}_0^T \boldsymbol{\beta}(F), \tau(F)) w_1^2(\mathbf{x}_0) p^2(\mathbf{x}_0) \mathbf{x}_0^T \mathbf{B}^{-1} \mathbf{x}_0. \end{aligned}$$

When considering a test, a different measure may be to consider the *expected squared influence function* $\text{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F) = \mathbb{E}(\text{IF}(\mathbf{z}_0, \mathcal{T}, F)^2 | (y_0, \mathbf{x}_0))$. In our case, we obtain

$$\text{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F) = \Psi^2 (y_0, \mathbf{x}_0^T \boldsymbol{\beta}(F), \tau(F)) w_1^2(\mathbf{x}_0) p(\mathbf{x}_0) \mathbf{x}_0^T \mathbf{B}^{-1} \mathbf{x}_0.$$

Note that $\text{EIF}(\mathbf{z}_0^*, \mathcal{T}, F)^2 = \text{EIF}_2(\mathbf{z}_0^*, \mathcal{T}, F) p(\mathbf{x})$ so, the difference between both measures is the role played by the missing probability.

7.1. Expected influence functions for the Gamma model

We will compute the influence function for the Gamma regression model considered in the simulation study, that is, the response y given the covariates \mathbf{x} has distribution $\Gamma(\tau, \mu(\mathbf{x}))$ with $\tau = 3$ and $\mu(\mathbf{x}) = \beta_1 x_1 + \beta_2 x_2 + \beta_3$, where $\beta_1 = \beta_2 = \beta_3 = 0$, $\mathbf{x}^T = (x_1, x_2, 1)$ and $(x_1, x_2) \sim N(\mathbf{0}_2, \mathbf{I})$.

Figures S.6 and S.7 in the supplementary file show the square of the *expected influence functions* (EIF^2) and the *expected squared influence* (EIF_2) at $y = \exp(1)$ of the Wald tests corresponding to the maximum likelihood estimators, the estimators related to those introduced in Bianco et al. (2005) based on the deviance, i.e., with $w_1 \equiv 1$, and the weighted estimators computed with the Tukey’s bisquare weight function, denoted by \widehat{w}_{ML} , \widehat{w}_{M} and \widehat{w}_{GM} , respectively, as in Sections 4 and 6. In the latter, the weights were computed over the Mahalanobis distances with tuning constant $c = \chi_{k, 0.95}^2$. The missing probabilities correspond to the complete case situation, i.e., $p \equiv 1$, the logistic model $p(\mathbf{x}) = 1 / (1 + \exp(-\mathbf{x}^T \boldsymbol{\lambda} - 2))$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$ and $p(\mathbf{x}) = 0.4 + 0.5(\cos(\mathbf{x}^T \boldsymbol{\lambda} + 0.4))^2$ with $\boldsymbol{\lambda} = (2, 2, 0)^T$, respectively. To plot the influence functions a grid of values for each component $x_j, j = 1, 2$, was taken between -4 and 4 with step 0.1.

When $p \equiv 1$, a reduced range was also considered to compare the behaviour near the origin by using a grid of points between -1 and 1 with step 0.025. These plots can be seen in Bianco et al. (2011b). As expected, the shape of both EIF^2 and EIF_2 for the test functionals \mathcal{W}_{M} and \mathcal{W}_{GM} are comparable to that of their classical relatives at the centre of the distribution of the covariates. Note that EIF^2 and EIF_2 are unbounded due to leverage points for the classical test and also for that based on the Bianco et al. (2005) estimators. This feature makes us suspect that even when the latter are based on a robust procedure to estimate the regression parameter, the test statistic may be sensitive to outliers as observed in the examples studied. On the other hand, when using a weight function to downweight carriers with large Mahalanobis distances, the expected influence at points further away is downweighted.

8. Concluding remarks

We have introduced a resistant procedure to test hypotheses on the regression parameter under a generalized linear regression model, when there are missing observations in the responses and it can be suspected that anomalous observations

are present in the sample. The estimators turn out to be asymptotically normally distributed. The test statistics are robust versions of the classical Wald-type statistic. Even when the tests statistics have a limiting χ^2 -distribution under the null hypothesis and under contiguous alternatives, the Monte Carlo study illustrates the convergence to the asymptotic distribution.

The simulation also confirms the expected inadequate behaviour of the classical Wald-type test in the presence of outliers and of the unweighted robust estimators under some contamination schemes. The proposed weighted robust procedure for the regression parameter is stable both in level and power when 5% of contamination is included, while some sensitivity is observed at 10% for some of the considered situations. The effect of high leverage points is also observed through the influence function.

Finally, through real data sets, we confirm the stability of the decision rule induced by \widehat{W}_{GM} , which is based on weights that control high leverage points, when outlying observations are present and under different missing probability patterns.

Acknowledgements

This work began while Isabel Rodrigues was visiting the Instituto de Cálculo of the Universidad de Buenos Aires and received financial support from Portuguese National Funds through FCT (Fundao para a Cincia e a Tecnologia) under the scope of project PEst-OE/MAT/UI0822/2011 and from *Fundação Calouste Gulbenkian* (FCG). This research was also partially supported by Grants PIP 112-200801-00216 from CONICET, PICT 0821 from ANPCYT and W276 from the Universidad de Buenos Aires at Buenos Aires, Argentina and by the joint cooperation program ANPCYT-FCT PO/09/05-Proc. 44100.

Appendix A

The following lemma will be useful when deriving the asymptotic distribution of the robust estimators defined in Section 2.1. We omit its proof since it follows using analogous arguments to those considered in Lemma 1 of Bianco and Boente (2002).

Lemma A.1. Assume that $(y_i, \mathbf{x}_i^T, \delta_i), 1 \leq i \leq n$ are independent random vectors satisfying (2) and such that $y_i|\mathbf{x}_i \sim F(\cdot, \mu_i, \tau)$ where $\mu_i = H(\mathbf{x}_i^T \boldsymbol{\beta})$. Let $\varphi(y, u, v)$ be a continuous function and assume that $\tilde{\tau} \xrightarrow{p} \tau$ and $\tilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Define $\mathbf{V} = \mathbb{E}_F(\varphi(y, \mathbf{x}^T \boldsymbol{\beta}, \tau) w_1(\mathbf{x}) p(\mathbf{x}) \mathbf{x} \mathbf{x}^T)$ and $\widehat{\mathbf{V}} = (1/n) \sum_{i=1}^n \delta_i \varphi(y_i, \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}, \tilde{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T$. Then, under N1 and N2, $\widehat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$.

Proof of Theorem 3.1. (a) Since $S_n^{(1)}(\widehat{\boldsymbol{\beta}}, \widehat{\tau}) = 0$, a Taylor’s expansion of order one leads to

$$\frac{1}{n} \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i + \frac{1}{n} \sum_{i=1}^n \delta_i \chi(y_i, \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{0}_k,$$

so, we have that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}_n^{-1} \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i / \sqrt{n}$. Note that Lemma A.1 entails that $\mathbf{A}_n = \sum_{i=1}^n \delta_i \chi(y_i, \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T / n \xrightarrow{p} \mathbf{A}$. Using that $\widehat{\tau} \xrightarrow{p} \tau$, N5 and the standard empirical processes arguments, we easily get that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \tau) w_1(\mathbf{x}_i) \mathbf{x}_i \xrightarrow{p} \mathbf{0}.$$

Hence, $\sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \widehat{\tau}) w_1(\mathbf{x}_i) \mathbf{x}_i / \sqrt{n} \xrightarrow{\mathcal{D}} N(\mathbf{0}_k, \mathbf{B})$ entailing that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \tau) w_1(\mathbf{x}_i) \mathbf{x}_i + o_p(1) \tag{A.1}$$

so, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}_k, \boldsymbol{\Sigma})$.

(b) Let T_n stand for $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T$ or $\widehat{\mathbf{B}}^{-1/2} \widehat{\mathbf{A}} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T$. In order to prove (b), we will use Le Cam’s third Lemma (see Van der Vaart, 2000, p. 90). Therefore, we need to obtain the asymptotic distribution of $(T_n, \ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})))$, under $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})$ is the joint density under the null hypothesis and $q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})$ is the corresponding one under the alternative, $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$.

Let $\sigma^2 = \mathbf{c}^T \mathbb{E}(H'(\mathbf{x}_i^T \boldsymbol{\beta}_0)) \mathbf{x}_i \mathbf{x}_i^T \mathbf{c} / A(\tau)$. Using standard arguments (see Bianco et al., 2011b, for details), we have that the following expansion holds

$$A(\tau) \ln \frac{q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})}{p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})} = \frac{1}{\sqrt{n}} \sum_{i=1}^n [y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}_0)] \mathbf{x}_i^T \mathbf{c} - \frac{1}{2n} \sum_{i=1}^n H'(\mathbf{x}_i^T \boldsymbol{\beta}_0) (\mathbf{x}_i^T \mathbf{c})^2 + o_p(1),$$

implying that $\ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})) \xrightarrow{\mathcal{D}} N(-\sigma^2/2, \sigma^2)$.

On the other hand, from (A.1), we have that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{A}^{-1} \mathbf{C}_n + o_p(1)$, where \mathbf{A} is defined in (6) and $\mathbf{C}_n = \sum_{i=1}^n \delta_i \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \tau) w_1(\mathbf{x}_i) \mathbf{x}_i / \sqrt{n}$. Moreover, Lemma A.1 entails that $\widehat{\mathbf{A}} \xrightarrow{p} \mathbf{A}$ and $\widehat{\mathbf{B}} \xrightarrow{p} \mathbf{B}$, thus, using that \mathbf{C}_n is asymptotically normally distributed, we get that $\widehat{\mathbf{B}}^{-1/2} \widehat{\mathbf{A}} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\mathbf{B}^{-1/2} \mathbf{C}_n + o_p(1)$. Hence, to derive the joint asymptotic distribution of $(T_n, \ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})))^T$, it is enough to compute the covariance between \mathbf{C}_n and $R_1 = (1/A(\tau)) \sum_{i=1}^n [y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}_0)] \mathbf{x}_i^T \mathbf{c} / \sqrt{n}$. Using that $\mathbb{E}_F(y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}_0)) = 0$, we get that

$$\text{Cov}(\mathbf{C}_n, R_1) = \frac{1}{A(\tau)} \mathbb{E}_F[(y_i - H(\mathbf{x}_i^T \boldsymbol{\beta}_0)) \Psi(y_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \tau) p(\mathbf{x}_i) w_1(\mathbf{x}_i) \mathbf{x}_i^T] \mathbf{c}.$$

Using N6 we get $\mathbb{E}_F[(y_1 - H(\mathbf{x}_1^T \boldsymbol{\beta}_0)) \Psi(y_1, \mathbf{x}_1^T \boldsymbol{\beta}_0, \tau) | \mathbf{x}_1] = -A(\tau) \mathbb{E}_F[\chi(y_1, \mathbf{x}_1^T \boldsymbol{\beta}_0, \tau) | \mathbf{x}_1]$, so, $\text{Cov}(\mathbf{C}_n, R_1) = -\mathbf{A} \mathbf{c}$, which implies that

$$(\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T, \ln(q_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})/p_n(\mathbf{y}, \mathbf{X}, \boldsymbol{\delta})))^T \xrightarrow{d} N\left(\begin{pmatrix} \mathbf{0}_k \\ 1 \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & -\mathbf{c} \\ -\mathbf{c} & \sigma^2 \end{pmatrix}\right)$$

so, $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(-\mathbf{c}, \boldsymbol{\Sigma})$ under $H_{1,n}$, concluding the proof. \square

Proof of Theorem 7.1. We need to compute $\partial^2 \mathcal{W}(F_{z_0, \epsilon}) / \partial \epsilon^2 |_{\epsilon=0}$. Using that, under the null hypothesis, $\boldsymbol{\beta}(F) = \boldsymbol{\beta}_0$, straightforward calculations lead to

$$\partial^2 \mathcal{W}(F_{z_0, \epsilon}) / \partial \epsilon^2 |_{\epsilon=0} = 2 (\partial(\boldsymbol{\beta}(F_{z_0, \epsilon}) - \boldsymbol{\beta}_0) / \partial \epsilon)^T \mathbf{V}^{-1}(F_{z_0, \epsilon}) (\partial(\boldsymbol{\beta}(F_{z_0, \epsilon}) - \boldsymbol{\beta}_0) / \partial \epsilon) |_{\epsilon=0}$$

so,

$$\text{IF}(\mathbf{z}_0, \mathcal{I}, F)^2 = \text{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F)^T \boldsymbol{\Sigma}^{-1} \text{IF}(\mathbf{z}_0, \boldsymbol{\beta}, F). \tag{A.2}$$

The proof follows now applying the expression for the influence function of $\boldsymbol{\beta}(F)$ given in (11). \square

Appendix B. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.csda.2012.05.008>.

References

Adimari, G., Ventura, L., 2002. Quasi-profile loglikelihoods for unbiased estimating functions. *Ann. Inst. Statist. Math.* 54, 235–244.
 Amaral Turkman, M.A., Silva, G., 2000. Modelos Lineares Generalizados—da Teoria à Prática. Edições SPE, Lisboa.
 Bianco, A., Boente, G., 2002. On the asymptotic behavior of one-step estimates in heteroscedastic regression models. *Statist. Probab. Lett.* 60, 33–47.
 Bianco, A., Boente, G., Rodrigues, I.M., 2011a. Resistant estimators in Poisson and gamma models with applications to outlier detection. Available on: www.ic.fcen.uba.ar/preprints/biancoenterodrigues_2011.pdf.
 Bianco, A., Boente, G., Rodrigues, I.M., 2011b. Robust test in generalized linear models with missing responses. Available on: www.ic.fcen.uba.ar/preprints/test_GLM_TR.pdf.
 Bianco, A., García Ben, M., Yohai, V., 2005. Robust estimation for linear regression with asymmetric errors. *Canad. J. Statist.* 33, 511–528.
 Bianco, A., Martínez, E., 2009. Robust testing in the logistic regression model. *Comput. Statist. Data Anal.* 53, 4095–4105.
 Bianco, A., Yohai, V., 1996. Robust estimation in the logistic regression model. In: *Lecture Notes in Statistics*, vol. 109. Springer-Verlag, New York, pp. 17–34.
 Cantoni, E., Ronchetti, E., 2001. Robust inference for generalized linear models. *J. Amer. Statist. Assoc.* 96, 1022–1030.
 Cantoni, E., Ronchetti, E., 2006. A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. *J. Health Econ.* 25, 198–213.
 Croux, C., Haesbroeck, G., 2003. Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Statist. Data Anal.* 44, 273–295.
 Croux, C., Haesbroeck, G., Joossens, K., 2008. Logistic discrimination using robust estimators: An influence function approach. *Canad. J. Statist.* 36 (2), 157–174.
 Feigl, P., Zelen, M., 1965. Estimation of exponential survival probabilities with concomitant information. *Biometrics* 21, 826–838.
 Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
 Heritier, S., Cantoni, E., Copt, S., Victoria-Feser, M.P., 2009. *Robust Methods in Biostatistics*. In: *Wiley Series in Probability and Statistics*, Wiley.
 Heritier, S., Ronchetti, E., 1994. Robust bounded-influence tests in general parametric models. *J. Amer. Statist. Assoc.* 89, 897–904.
 He, X., Simpson, D., Portnoy, S., 1990. Breakdown robustness of test. *J. Amer. Statist. Assoc.* 85, 446–452.
 Ibrahim, J., 1990. Incomplete data in generalized linear models. *J. Amer. Statist. Assoc.* 85, 765–769.
 Künsch, H., Stefanski, L., Carroll, R., 1989. Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. *J. Amer. Statist. Assoc.* 84, 460–466.
 Marazzi, A., Yohai, V., 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. *J. Statist. Plann. Inference* 122, 271–291.
 Maronna, R., Martin, D., Yohai, V., 2006. *Robust Statistics: Theory and Methods*. Wiley, New York.
 McCullagh, P., Nelder, J., 1989. *Generalized Linear Models*, second ed. Chapman and Hall, London.
 Neyman, J., 1938. Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.* 33, 101–116.
 Rieder, H., 1978. A robust asymptotic testing model. *Ann. Statist.* 6, 1080–1094.
 Stefanski, L., Carroll, R., Ruppert, D., 1986. Bounded score functions for generalized linear models. *Biometrika* 73, 413–424.
 Van der Vaart, A., 2000. *Asymptotic Statistics*. In: *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press.