Robust and Accurate Inference for Generalized Linear Models

by

Serigne N. Lô

The George Institute

University of Sydney, Australia

and

Elvezio Ronchetti

Department of Econometrics

University of Geneva, Switzerland

February 2008 / Revised: May 2009

Abstract

In the framework of generalized linear models, the nonrobustness of classical estimators and tests for the parameters is a well known problem and alternative methods have been proposed in the literature. These methods are robust and can cope with deviations from the assumed distribution. However, they are based on first order asymptotic theory and their accuracy in moderate to small samples is still an open question. In this paper we propose a test statistic which combines robustness and good accuracy for moderate to small sample sizes. We combine results from Cantoni and Ronchetti (2001) and Robinson, Ronchetti and Young (2003) to obtain a robust test statistic for hypothesis testing and variable selection which is asymptotically χ^2-distributed as the three classical tests but with a *relative error* of order $O(n^{-1})$. This leads to reliable inference in the presence of small deviations from the assumed model distribution and to accurate testing and variable selection even in moderate to small samples.

Keywords: M-estimators, Monte Carlo, Robust inference, Robust variable selection, Saddlepoint techniques, Saddlepoint Test.

1 Introduction

Generalized linear models (GLM) (McCullagh and Nelder, 1989) have become the most commonly used class of models in the analysis of a large variety of data. In particular, GLM can be used to model the relationship between predictors and a function of the mean of a continuous or discrete response variable. Let $Y_1, ..., Y_n$ be n independent observations of a response variable. Assume that the distribution of Y_i belongs to the exponential family with $E[Y_i] = \mu_i$ and $Var[Y_i] = V(\mu_i)$, and

$$g(\mu_i) = \eta_i = x_i^T \beta, \qquad i = 1, ..., n, \tag{1}$$

where $\beta \in \mathbb{R}^q$ is a vector of unknown parameters, $x_i \in \mathbb{R}^q$, and $g(.)$ is the link function.

The estimation of β can be carried out by maximum likelihood or quasi-likelihood methods, which are equivalent if $g(.)$ is the canonical link, such as the logit function for logistic regression or the *log* for Poisson regression. Standard asymptotic inference based on likelihood ratio, Wald, and score test is then readily available for these models.

However, two main problems can potentially invalidate p-values and confidence intervals based on standard classical techniques. First of all, the models are ideal approximations to reality and deviations from the assumed distribution can have important effects on classical estimators and tests for these models (nonrobustness). Secondly, even when the model is exact, standard classical inference is based on approximations to the distribution of the test statistics provided by (first order) asymptotic theory. This can lead to inaccurate p-values and confidence intervals when the sample size is moderate to small or when probabilities in the far tails are required (and in some cases both are required). Since these tests are typically used for model comparison and variable selection, these problems can have important implications in the final choice of the explanatory variables. As an illustration, consider for instance the data set discussed in section 5, where a Poisson regression is used to model adverse events of a drug on 117 patients affected by Crohn's disease (a chronic inflammatory disease of the intestine) by means of 7 explanatory variables describing the characteristics of each patient. In this case a classical variable selection is affected by the presence of outlying observations, while a deviance analysis obtained using our new test is more reliable; see section 5.

The nonrobustness of classical estimators and tests for β is a well known problem and alternative methods have been proposed in the literature; see, for instance Pregibon (1982), Stefanski, Carroll, and Ruppert (1986), Künsch, Stefanski, and Carroll (1989), Morgenthaler (1992), Bianco and Yohai (1996), Ruckstuhl and Welsh (2001), Victoria-Feser (2002), and Croux and Haesbroeck (2003) for robust estimators and Cantoni and Ronchetti (2001) for robust inference. Although these methods are devised to cope with deviations from the assumed model distribution, their statistical properties are based on first order asymptotic theory and the accuracy of the asymptotic approximation of their distributions in moderate to small samples is still an open question.

In this paper we propose a test statistic which combines robustness and good accuracy for small sample sizes. As a first step we apply the results in Robinson, Ronchetti, and Young (2003) to the GLM case and obtain the new test statistic in this case. We then combine the results of Cantoni and Ronchetti (2001) and Robinson, Ronchetti, and Young (2003) to obtain a robust test statistic for hypothesis testing and variable selection in GLM which is asymptotically χ^2-distributed as the three classical tests but with a *relative error* of order $O(n^{-1})$, i.e. the difference between the exact tail probability and that obtained by the χ^2 distribution divided by the exact is of order $O(n^{-1})$. This is in contrast with the absolute error of order $O(n^{-\frac{1}{2}})$ for the classical tests, where the difference between the exact tail probability and that obtained by the χ^2 distribution is of order $O(n^{-\frac{1}{2}})$. For a more detailed discussion of these properties we refer to Robinson, Ronchetti, and Young (2003), p.1155-1156. The accuracy of the new robust test statistic is stable in a neighborhood of the model distribution and this leads to robust inference even in moderate to small samples. The new test statistic is easily computed. Given a robust estimator for β, it has an explicit form in the case of a simple hypothesis and it requires an additional minimization in the case of a composite hypothesis. S-PLUS code is available from the authors upon request.

The paper is organized as follows. Section 2 reviews the classical and robust estimators for GLM. In section 3.1 we review the saddlepoint test statistic in a general setup and in section 3.2 we give its explicit form in the case of GLM. Three important special cases (Normal, Poisson, Binomial) are treated in detail. In section 3.3 we present the robustified version of the saddlepoint test which is obtained by replacing the classical score function by its robust version in the saddlepoint test statistic. Section 4 presents a simulation study in the case of Poisson regression which shows the advantage of robust saddlepoint tests with respect to standard classical tests. As an illustration, the new procedure is applied to a real data example in section 5. Finally, section 6 concludes the article with some potential research directions.

2 Classical and Robust Inference for Generalized Linear Models

Let $Y_1, ..., Y_n$ be n of independent random variables with density (or probability function) belonging to the exponential family

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + d(y; \phi)\right\}, \tag{2}$$

for some specific functions $a(\cdot)$, $b(\cdot)$ and $d(\cdot; \cdot)$. Then $E[Y_i] = \mu_i = b'(\theta_i)$ and $Var[Y_i] = b''(\theta_i)a(\phi)$. Given n observations $x_1, ..., x_n$ of a set of q explanatory variables ($x_i \in \mathbb{R}^q$), (1) defines the relationship between a linear predictor of the x_i's and a function $g(\mu_i)$ of the mean response μ_i. When $g(\mu_i)$ is the canonical link, $g(\mu_i) = \theta_i$, the maximum likelihood estimator and the quasi-likelihood estimator of β are the solution of the system of equations

$$\sum_{i=1}^{n} (y_i - \mu_i) \cdot x_{ij} = 0, \qquad j = 1, ..., q, \tag{3}$$

where $\mu_i = g^{-1}(x_i^T \beta)$.

The maximum likelihood and the quasi-likelihood estimator defined by (3) can be viewed as an M-estimator (Huber, 1981) with score function

$$\psi(y_i; \beta) = (y_i - \mu_i) \cdot x_i, \tag{4}$$

where $x_i = (x_{i1}, ..., x_{iq})^T$.

Since $\psi(y; \beta)$ is in general unbounded in x and y, the influence function of the estimator defined by (3) is unbounded and the estimator is not robust; see Hampel, Ronchetti, Rousseeuw, and Stahel (1986). Several alternatives have been proposed. One of these methods is the class of M-estimators of Mallows's type (Cantoni and Ronchetti 2001) defined by the score function:

$$\psi(y_i;\beta) = \nu(y_i,\mu_i)w(x_i)\mu_i' - \tilde{a}(\beta), \tag{5}$$

where $\tilde{a}(\beta) = \frac{1}{n}\sum_{i=1}^{n} E[\nu(y_i,\mu_i)]w(x_i)\mu_i', \quad \mu_i' = \frac{\partial\mu_i}{\partial\beta}$,
$\nu(y_i,\mu_i) = \psi_c(r_i)\frac{1}{V^{1/2}(\mu_i)}$, $r_i = \frac{y_i-\mu_i}{V^{1/2}(\mu_i)}$ are the Pearson residuals, $V^{1/2}(.)$ the square root of the variance function, and ψ_c is the Huber function defined by

$$\begin{aligned} \psi_c(r) &= r & |r| \le c \\ &= c \cdot sign(r) & |r| > c. \end{aligned}$$

When $w(x_i) = 1$, we obtain the so-called Huber quasi-likelihood estimator.

The tuning constant c is typically chosen to ensure a given level of asymptotic efficiency and $\tilde{a}(\beta)$ is a correction term to ensure Fisher consistency at the model. that can be computed explicitly for binomial and Poisson models and does not require numerical integration. The choice of this estimator is due to the fact that standard (first order aymptotic) inference based on robust quasi-deviances is available; see Cantoni and Ronchetti (2001). This will allow us to compare our new robust test with classical and robust tests based on first order asymptotic theory.

3 Small Sample Accuracy and Robustness

3.1 Saddlepoint Test Statistic

Let $Y_1, ..., Y_n$ be an independent, identically distributed sample of random vectors from a distribution F on some sample space $\mathcal{Y}$. Define the M-functional $\beta(F)$ to satisfy

$$E[\psi(Y;\beta)] = 0, \tag{6}$$

where ψ is assumed to be a smooth function from $\mathcal{Y} \times \mathbb{R}^q \longrightarrow \mathbb{R}^q$ with $q = dim(\beta)$ and the expectation is taken with respect to F. Suppose we wish to test the hypothesis $u(\beta) = \eta_0$, where $u : \mathbb{R}^q \rightarrow \mathbb{R}^{q_1}$, $q_1 \leq q$ and consider test statistics based on $u(T_n)$, where T_n is the M-estimate of β given by the solution of

$$\sum_{i=1}^{n} \psi(Y_i; T_n) = 0. \tag{7}$$

When $q_1 = 1$, saddlepoint approximations with relative error of order $O(n^{-1})$ for the p-value $P[u(T_n) > u(t_n)]$, where t_n is the observed value of T_n, are available; see for instance DiCiccio, Field, and Fraser (1990), Tingley and Field (1990), Daniels and Young (1991), Wang (1993), Jing and Robinson (1994), Fan and Field (1995), Davison, Hinkley, and Worton (1995), Gatto and Ronchetti (1996), and Butler (2007) for a recent general overview on saddlepoint methods. In the multidimensional case ($q_1 > 1$), Robinson, Ronchetti, and Young (2003) proposed the one dimensional test statistic $h(u(T_n))$, where

$$h(y) = \inf_{\{\beta: u(\beta)=y\}} \sup_{\lambda} \{-K_\psi(\lambda; \beta)\} \tag{8}$$

and

$$K_\psi(\lambda; \beta) = logE[e^{\lambda^T \psi(Y;\beta)}] \tag{9}$$

is the cumulant generating function of the score function $\psi(Y; \beta)$ and the expectation is taken with respect to F under the null hypothesis.

Using the saddlepoint approximation of the density of the M-estimator T_n, they proved that under the null hypothesis, $2nh(u(T_n))$ is asymptotically $\chi^2_{q_1}$ with a *relative error of order* $O(n^{-1})$. Therefore, although this test is asymptotically (first order) equivalent to the three standard tests, it has better small sample properties, the classical tests being asymptotically $\chi^2_{q_1}$ with only an absolute error of order $O(n^{-\frac{1}{2}})$.

Notice that (8) can be rewritten as

$$h(y) = \inf_{\{\beta: u(\beta)=y\}} \{-K_\psi(\lambda(\beta);\beta)\}, \tag{10}$$

where K_ψ is defined by (9) and $\lambda(\beta)$ is the so-called saddlepoint satisfying

$$K'_\psi(\lambda;\beta) \equiv \frac{\partial}{\partial\lambda} K_\psi(\lambda;\beta) = 0. \tag{11}$$

Moreover, in the case of a simple hypothesis, i.e. $u(\beta) = \beta$, (10) simply becomes $h(\beta) = -K_\psi(\lambda(\beta);\beta)$.

In order to apply the saddlepoint test statistic to GLM, we first adapt this result to the case when the observations $Y_1, ..., Y_n$ are independent but not identically distributed. In this case the formulas given above still hold with the cumulant generating function (9) replaced by

$$K_\psi(\lambda;\beta) = \frac{1}{n}\sum_{i=1}^{n} K^i_\psi(\lambda;\beta), \tag{12}$$

where $K^i_\psi(\lambda;\beta) = logE_{F^i}[e^{\lambda^T\psi(Y_i;\beta)}]$ and F^i is the distribution of Y_i.

This follows from the fact that the proof about the accuracy of the test requires the saddlepoint approximation of the density of the M-estimator T_n, which in the case of independent but not identically distributed observations is given in section 4.5c of Field and Ronchetti (1990) or in section 4 of Ronchetti and Welsh (1994) and is based on the cumulant generating function (12).

The saddlepoint test statistic can now be applied to GLM with different score functions ψ, such as those defined by (4) and (5). In the next section, we will exploit the structure of GLM to provide explicit formulas for the new test statistic.

3.2 Saddlepoint Test Statistic with Classical Score Function

In this section we first consider the classical situation. Robust versions of the test will be derived in section 3.3. The quasi-likelihood and the maximum likelihood estimators of β are defined by the same score function. The solution of (3) is an M-estimator defined by the score function (4). We now derive the explicit form of the saddlepoint test statistic (8) with the classical score function (4). The complete computations are provided in Appendix A, B, C, D in the document "Robust and Accurate Inference for Generalized Linear Models: Complete Computations" available at *http://www.unige.ch/ses/metri/ronchetti/ERpapers01.html*.

We consider first the case of a simple hypothesis $\beta = \beta_0$. Let $K_\psi(\lambda;\beta) = \frac{1}{n}\sum_{i=1}^n K_\psi^i(\lambda;\beta)$, where $K_\psi^i(\lambda;\beta) = logE_{F_0^i}[e^{\lambda^T\psi(Y_i;\beta)}]$ and F_0^i is the distribution of Y_i defined by the exponential family (2) with $\theta = \theta_{0i}$ and $b'(\theta_{0i}) = \mu_{0i} = g^{-1}(x_i^T\beta_0)$. Then by (4) we can write

$$K_\psi^i(\lambda;\beta) = log\int e^{\lambda^T\psi(y;\beta)} f_{Y_i}(y;\theta_{0i},\phi)\cdot dy$$

$$\begin{aligned}
&= log\int e^{\lambda^T(y-\mu_i)x_i}\cdot e^{\frac{y\theta_{0i}-b(\theta_{0i})}{a(\phi)}}\cdot e^{d(y;\phi)}\cdot dy \\
&= log\int e^{-\mu_i\lambda^T x_i}\cdot e^{\frac{-b(\theta_{0i})}{a(\phi)}}\cdot e^{\frac{y(\theta_{0i}+\lambda^T x_i a(\phi))}{a(\phi)}}\cdot e^{d(y;\phi)}\cdot dy \\
&= log\int e^{-\mu_i\lambda^T x_i}\cdot e^{\frac{-b(\theta_{0i})}{a(\phi)}}\cdot e^{\frac{b(\theta_{0i}+\lambda^T x_i a(\phi))}{a(\phi)}}\cdot e^{\frac{y(\theta_{0i}+\lambda^T x_i a(\phi))-b(\theta_{0i}+\lambda^T x_i a(\phi))}{a(\phi)}}\cdot e^{d(y;\phi)}\cdot dy \\
&= log\left[e^{-[\mu_i\lambda^T x_i+\frac{b(\theta_{0i})}{a(\phi)}]}\cdot e^{\frac{b(\theta_{0i}+\lambda^T x_i a(\phi))}{a(\phi)}}\cdot\int e^{\frac{y(\theta_{0i}+\lambda^T x_i a(\phi))-b(\theta_{0i}+\lambda^T x_i a(\phi))}{a(\phi)}}\cdot e^{d(y;\phi)}\cdot dy\right] \\
&= \frac{b(\theta_{0i}+\lambda^T x_i a(\phi))-b(\theta_{0i})}{a(\phi)} - \mu_i\lambda^T x_i\ . \qquad (13)
\end{aligned}$$

By taking into account the fact that $\mu_i = b'(\theta_i)$, and that $b'(.)$ is injective, the solution $\lambda(\beta)$ of (11) with K_ψ defined by (12) and (13) is unique and given by (see

Appendix A):

$$\lambda(\beta) = \frac{\beta - \beta_0}{a(\phi)}.$$

Therefore,

$$h(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{b'(x_i^T\beta)x_i^T(\beta-\beta_0) - (b(x_i^T\beta) - b(x_i^T\beta_0))}{a(\phi)} \ . \tag{14}$$

The test statistic $2nh(\hat{\beta})$ given by (14) where $\hat{\beta}$ is MLE (the solution of (3)) is asymptotically χ^2_q under the simple null hypothesis $\beta = \beta_0$ and can be used to test this null hypothesis.

Notice that in this case (*simple hypothesis* and *canonical link*), the classical saddlepoint test statistic $2nh(\hat{\beta})$ defined by (14) is the log-likelihood ratio test statistic. Therefore, in this case the latter is asymptotically χ^2_q with a relative error of order $O(n^{-1})$.

To test the more general hypothesis $u(\beta) = \eta_0$, where $u : \mathbb{R}^q \to \mathbb{R}^{q_1}$, $q_1 \leq q$, the test statistic is given by $2nh(u(\hat{\beta}))$, where $h(y)$ is defined by (10) and from (13), (14)

$$-K_\psi(\lambda(\beta);\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{b'(x_i^T\beta)x_i^T(\beta-\beta_0) - (b(x_i^T\beta) - b(x_i^T\beta_0))}{a(\phi)} \ , \tag{15}$$

and β_0 such that $u(\beta_0) = \eta_0$, i.e. β_0 is the estimator of β under the null hypothesis.

Three special cases

(i) $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$

$b(\theta) = \frac{\theta^2}{2}$, $\quad a(\phi) = \sigma^2$

Then,

$$h(\beta) \ = \ \frac{1}{2n\sigma^2}(\beta-\beta_0)^T\Big[\sum_{i=1}^{n} x_i x_i^T\Big](\beta-\beta_0).$$

(ii) $Y_i \sim \mathcal{P}(\mu_i)$

$b(\theta) = e^{\theta}\,, \quad a(\phi) = 1$

Then,

$$h(\beta) = \frac{1}{n}\Big[\sum_{i=1}^{n} e^{x_i^T\beta} x_i^T(\beta - \beta_0) - \sum_{i=1}^{n}(e^{x_i^T\beta} - e^{x_i^T\beta_0})\Big]\,.$$

(iii) $Y_i \sim \mathcal{B}in(m, \pi_i)$

$b(\theta) = m\log(1 + e^{\theta})\,, \quad a(\phi) = 1$

Then,

$$h(\beta) \;=\; \frac{m}{n}\Big[\sum_{i=1}^{n} \frac{e^{x_i^T\beta}}{1 + e^{x_i^T\beta}} x_i^T(\beta - \beta_0) - \sum_{i=1}^{n}\big[log(1 + e^{x_i^T\beta}) - log(1 + e^{x_i^T\beta_0})\big]\Big]\,.$$

When the model is exact and for composite hypotheses, the saddlepoint test will be more accurate than the standard classical likelihood ratio test. However, both are based on the (unbounded) classical score function (4) and will be inaccurate (even for large n) in the presence of deviations from the model. In the next section, we construct a robustified version of the saddlepoint test.

3.3 Saddlepoint Test Statistic with Robust Score Function

From (5), the robust score function is defined by $\tilde{\psi}_R(y;\beta) = \psi_c(r)w(x)\frac{1}{V^{1/2}(\mu)}\mu' - \tilde{a}(\beta)$ and the cumulant generating function of the robust score function by

$$K_{\tilde{\psi}_R}(\lambda;\beta) = \frac{1}{n}\sum_{i=1}^{n} K^{i}_{\tilde{\psi}_R}(\lambda;\beta), \qquad (16)$$

where

$$K^{i}_{\tilde{\psi}_R}(\lambda;\beta) = logE_{F^i}\big[e^{\lambda^T\tilde{\psi}_R(Y_i;\beta)}\big]\,.$$

As in the classical case, the robust cumulant generating function $K^i_{\tilde{\psi}_R}(.)$ for each observation i can be written as

$$
\begin{aligned}
K^i_{\tilde{\psi}_R}(\lambda;\beta) &= \log \int e^{\lambda^T \tilde{\psi}_R(y;\beta)} f_{Y_i}(y;\theta_{0i},\phi)\cdot dy \\
&= log \int e^{\lambda^T \psi_c(r_i)\frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot e^{\frac{y\theta_{0i}-b(\theta_{0i})}{a(\phi)}} \cdot e^{d(y;\phi)} \cdot dy \\
&= log \Big[\int_{r_i<-c} e^{-\lambda^T c \frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot e^{\frac{y\theta_{0i}-b(\theta_{0i})}{a(\phi)}} \cdot e^{d(y;\phi)} \cdot dy \qquad (I_{i1}) \\
&+ \int_{-c<r_i<c} e^{\lambda^T \frac{y-\mu_i}{V^{1/2}(\mu_i)} \frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot e^{\frac{y\theta_{0i}-b(\theta_{0i})}{a(\phi)}} \cdot e^{d(y;\phi)} \cdot dy \qquad (I_{i2}) \\
&+ \int_{r_i>c} e^{\lambda^T c \frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot e^{\frac{y\theta_{0i}-b(\theta_{0i})}{a(\phi)}} \cdot e^{d(y;\phi)} \cdot dy \qquad (I_{i3}) \Big] \\
&= log[I_{i1}+I_{i2}+I_{i3}],
\end{aligned}
$$

where $r_i = \frac{y-\mu_i}{V^{1/2}(\mu_i)}$.

For the explicit calculations of I_{ij} for $j = 1,2,3$, we refer to Appendix B. Finally, the cumulant generating function can be written as

$$
\begin{aligned}
K^i_{\tilde{\psi}_R}(\lambda;\beta) &= log[I_{i1}+I_{i2}+I_{i3}] \\
&= log\Big[e^{-\lambda^T c \frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot P(Z^i \le -cV^{1/2}(\mu_i)+\mu_i) \\
&+ e^{\frac{-\lambda^T \mu_i \mu_i' w(x_i)}{V(\mu_i)}} \cdot e^{-\lambda^T \tilde{a}(\beta)} \cdot e^{\frac{b(\theta_{0i}+\frac{\lambda^T \mu_i' w(x_i) a(\phi)}{V(\mu_i)})-b(\theta_{0i})}{a(\phi)}} \\
&. \quad P(-cV^{1/2}(\mu_i)+\mu_i < Z^i_\lambda < cV^{1/2}(\mu_i)+\mu_i) \\
&+ e^{\lambda^T c \frac{w(x_i)}{V^{1/2}(\mu_i)}\mu_i' - \lambda^T \tilde{a}(\beta)} \cdot P(Z^i \ge cV^{1/2}(\mu_i)+\mu_i)\Big],
\end{aligned}
$$

where Z^i is a random variable with distribution (2) with $\theta = \theta_{0i}$ and Z^i_λ is a random variable with distribution (2) with $\theta = \theta_{0i} + \frac{\lambda^T \mu_i' w(x_i) a(\phi)}{V(\mu_i)}$.

To obtain $h_R(\beta)$, we have to solve the equation

$$
\frac{\partial K_{\tilde{\psi}_R}(\lambda;\beta)}{\partial \lambda} = \frac{1}{n}\sum_{i=1}^n \frac{\partial K^i_{\tilde{\psi}_R}(\lambda;\beta)}{\partial \lambda} = 0, \qquad (17)
$$

with respect to λ, i.e.

$$s(\lambda;\beta) = \sum_{i=1}^{n} \frac{\partial K^i_{\tilde{\psi}_R}(\lambda;\beta)}{\partial\lambda} = \sum_{i=1}^{n} \frac{\partial log(I_{i1}+I_{i2}+I_{i3})}{\partial\lambda} = \sum_{i=1}^{n} \frac{\frac{\partial I_{i1}}{\partial\lambda}+\frac{\partial I_{i2}}{\partial\lambda}+\frac{\partial I_{i3}}{\partial\lambda}}{I_{i1}+I_{i2}+I_{i3}} = 0\,. \qquad (18)$$

(18) can be easily solved numerically. Alternatively, we can approximate the solution of (18) by a one-step Newton's algorithm, i.e.

$$\tilde{\lambda}(\beta) \cong \lambda^0 - \big[\frac{\partial s(\lambda;\beta)}{\partial\lambda}\,|_{\lambda^0}\big]^{-1}\cdot s(\lambda^0;\beta)\,, \qquad (19)$$

where $\lambda^0 = \frac{\hat{\beta}_R-\beta_0}{a(\phi)}$ and $\hat{\beta}_R$ is the robust estimator defined by (7) and (5). The explicit computations of $s(\lambda;\beta)$ and $\frac{\partial s(\lambda;\beta)}{\partial\lambda}$ are provided in Appendix C.

For a given distribution of Y_i this leads to the following expression for the robust saddlepoint test statistic $h_R(.)$:

$$h_R(\beta) = \frac{1}{n}\sum_{i=1}^{n} K^i_{\tilde{\psi}_R}(\tilde{\lambda}(\beta);\beta)\,, \qquad (20)$$

where $\tilde{\lambda}(\beta) \cong \frac{\hat{\beta}_R-\beta_0}{a(\phi)} - \big[\sum_{i=1}^{n} x_i x_i^T A_i(\frac{\hat{\beta}_R-\beta_0}{a(\phi)})\big]^{-1}\cdot s(\frac{\hat{\beta}_R-\beta_0}{a(\phi)};\beta)$

and $A_i(.)$ a scalar function defined by the distribution of Y_i. For the important cases of Normal, Poisson and Binomial distributions, we refer to the corresponding expressions in Appendix D.

The test statistic $2nh_R(\hat{\beta}_R)$ given by (20) where $\hat{\beta}_R$ is the robust M-estimator defined by (7) with the score function given by (5) is asymptotically χ^2_q under the simple null hypothesis $\beta = \beta_0$ and can be used to test this null hypothesis.

To test the more general hypothesis $u(\beta) = \eta_0$, where $u : \mathbb{R}^q \to \mathbb{R}^{q_1}$, $q_1 \leq q$, the robust test statistic is given by $2nh_R(u(\hat{\beta}_R))$, where $h_R(y)$ is defined by

$$h_R(y) = \inf_{\{\beta: u(\beta)=y\}} \{-K_{\tilde{\psi}_R}(\tilde{\lambda}(\beta); \beta)\}. \tag{21}$$

4 A Simulation Study

To illustrate and compare the different tests, we consider a Poisson regression model with canonical link $g(\mu) = log(\mu)$ and 3 explanatory variables plus intercept ($q = 4$), i.e.

$$log(\mu_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4}\,,$$

where $x_{ij} \sim U[0,\, 1]$, $j = 2, 3, 4$. The $Y_i's$ are generated according to the Poisson distribution $P(\mu_i)$ and a perturbed distribution of the form $(1-\epsilon)P(\mu_i)+\epsilon P(\nu\mu_i)$, where $\epsilon = 0.05, 0.10$ and $\nu = 2$. The latter represents situations where the distribution of the data is not exactly the model but possibly lies in a small neighborhood of the model. The null hypothesis is $\beta_2 = \beta_3 = \beta_4 = 0$ ($q_1 = 3$) and we choose two sample sizes $n = 30, 100$. To simulate the data, the parameter β_1 was set equal to 1.

We consider four tests: the classical test, the robust quasi-deviance test developed in Cantoni and Ronchetti (2001), and the two saddlepoint tests derived from them in sections 3.2 and 3.3. The latter are defined by the new test statistics $2nh(\hat{\beta})$ and $2nh_R(\hat{\beta}_R)$ respectively. The tuning constant c in the robust score function (5) is set to 1.345. Since the x-design is balanced and uniform on $[0, 1]$, it is not necessary to use weights on the covariates $x_i's$ and we set $w(x_i) \equiv 1\ \forall i$.

The computation of the new saddlepoint test statistics involves explicit expressions ((14) and (20)) in the case of a simple hypothesis and an additional minimization in the case of a composite hypothesis. In our simulations and in the real data example of section 5, we computed (10)-(11) by direct numerical optimization without using (20). In higher dimensional problems the latter would certainly be useful. S-PLUS code is available from the authors upon request. The evaluation of the robust version of the saddlepoint test requires the computation of $\hat{\beta}_R$, the robust estimator defined by (7) and (5). Code is available in **R** (function `glmrob` in the `robustbase` package) and S-PLUS (at *http://www.unige.ch/ses/metri/cantoni/*).

The results of the simulations are represented by PP-plots of p-values against $U[0, 1]$ probabilities. In Figures 1 to 3, PP-plots for the classical test (left) and the saddlepoint test based on the classical score function (right) are given in Panel (a). Panel (b) shows the corresponding PP-plots for their robust versions. The first row reports the simulation results for sample size $n = 30$ and the second one for $n = 100$.

Figure 1 shows the results when there are no deviations from the model. Even in this case the asymptotic χ^2 approximation of the classical test statistic is inaccurate (deviation from the 45^o line) both for $n = 30$ and 100 while the χ^2 approximation of the distribution of the new test statistic is clearly better. The robust quasi-deviance test is already doing better than its classical counterpart and the χ^2 approximation to the new robust saddlepoint test statistic provides a very high degree of accuracy. In the presence of small deviations from the model (Figure 2), the χ^2 approximation of the classical test is extremely inaccurate (even for $n = 100$), its saddlepoint version and robust quasi-deviance version are better but still inaccurate, while the robust saddlepoint test is very accurate even down to $n = 30$.

Finally, in the presence of larger deviations from the model (Figure 3), the robust saddlepoint test is not as accurate as in the previous cases but it is still useful. Notice however that this is an extreme scenario especially for $n = 30$.

To summarize: The new saddlepoint statistic clearly improves the accuracy of the test. When it is used with a robust score function, it can control the bias due to deviations from the model and the resulting test is very accurate in the presence of small deviations from the model and even down to small sample sizes.

5 A Real Data Example

To illustrate the use of the new test for variable selection, we consider a data set issued from a study of the adverse events of a drug on 117 patients affected by Crohn's disease (a chronic inflammatory disease of the intestines).

In addition to the response variable AE (number of adverse events), 7 explanatory variables were recorded for each patient: BMI (body mass index), HEIGHT, COUNTRY (one of the two countries where the patient lives), SEX, AGE, WEIGHT, and TREAT (the drug taken by the patient in factor form: placebo, Dose 1, Dose 2). We consider a Poisson regression model.

Table 1 presents the $p-$values of an analysis of deviance based on the classical test, the (first order) robust quasi-deviance test developed in Cantoni and Ronchetti (2001), and the new robust saddlepoint test respectively. The three analyses agree on the selection of the variables Dose 1, BMI, HEIGHT, SEX and the non-selection of Dose 2. The variable COUNTRY is also essentially significant everywhere. Finally, AGE is supported by the two robust analyses, while WEIGHT is not selected by the classical and the robust saddlepoint analysis, which seems to be reasonable in view of the inclusion of BMI and HEIGHT. Additional information on the influential points is provided by Figure 4 which shows the robust weights obtained by the robust analysis. Points with small weights can have a big influence on the classical analysis and can lead to a wrong variable selection when using the classical test. In view of the robustness *and* better finite sample behavior of the new test, we recommend the result obtained by the third analysis.

6 Conclusion

We derived a robust test for GLM with good small sample accuracy. It keeps its level in the presence of small deviations from the assumed model and the χ^2 approximation of its distribution is accurate even down to small sample sizes. Since this test requires only a robust score function, similar test procedures can be developed for other models where such score functions are available.

Acknowledgments Part of the first author's work was done while he was with the Dept. of Econometrics of the University of Geneva. The second author would like to thank the Dept. of Operations Research and Financial Engineering of Princeton University for the hospitality during part of this work and the Swiss National Science Foundation Pro*Doc Program 114533 for financial support. Both authors would like to thank the Editor, two referees, E. Cantoni, M.-P. Victoria-Feser, J. Krishnakumar, C. Field, J. Jureckova, M. Huskova, F. Peracchi, C. Croux, S. Heritier, V. Czellar, M. Mächler, W. Stahel, Y. Ma, M. Genton, E. Parzen, N. Wang, and seminar participants at the Universities of Sydney and Macquarie (Australia), the Francqui Foundation Workshop (Brussels), Dalhousie University (Canada), Texas A&M University, ETH Zürich, Università Cattolica (Milano) for useful remarks and suggestions on an early draft of the paper. They also thank Dr. Enrico Chavez (Genexion SA, Geneva) who kindly provided the data set analyzed in section 5.

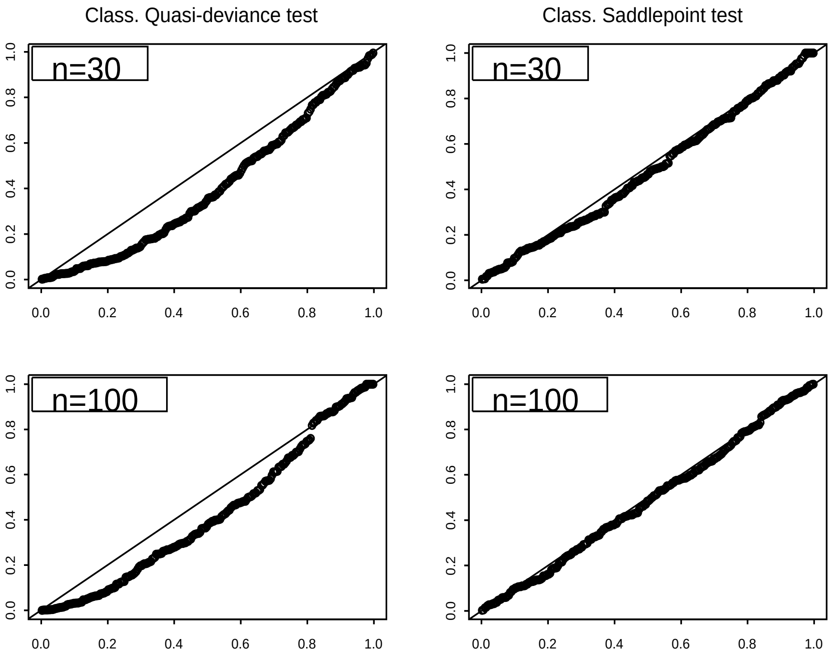

Figure 1 (a): *PP-plots of* **classical** $p-$*values vs.* $U[0,1]$ *when the data are generated from* $P(\mu_i)$.

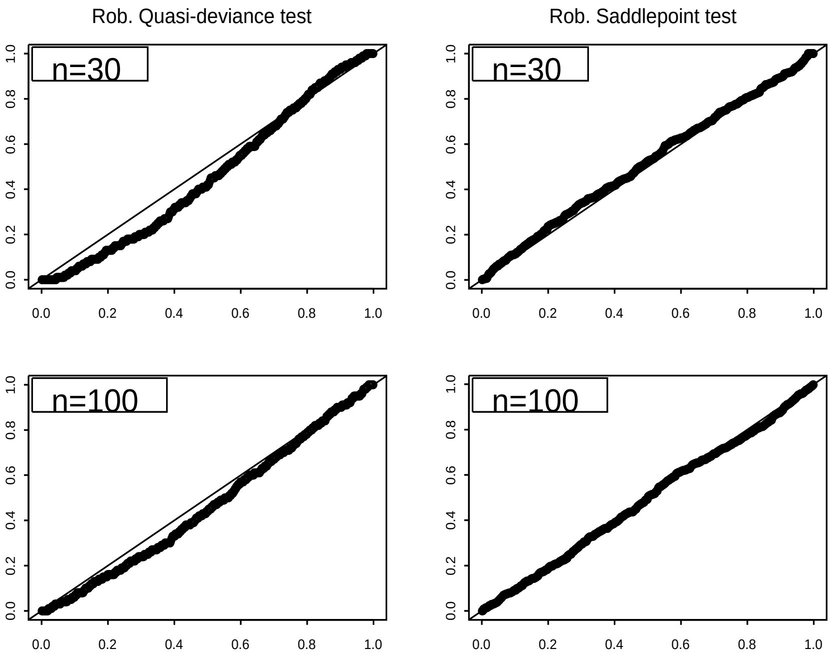

Figure 1 (b): *PP-plots of* **robust** $p-$*values vs.* $U[0,1]$ *when the data are generated from* $P(\mu_i)$.

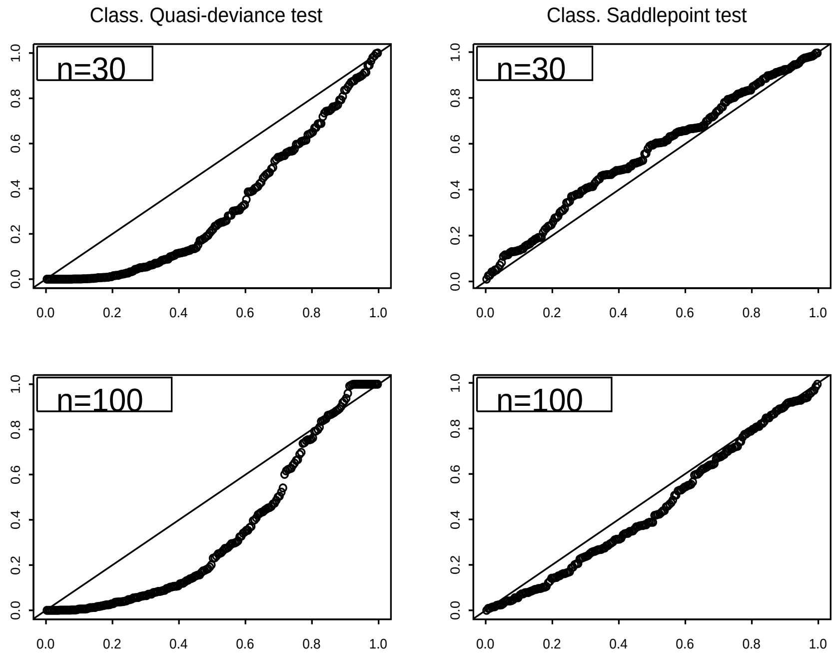

Figure 2 (a): *PP-plots of* **classical** $p-$*values vs.* $U[0,1]$ *when the data are generated from a contaminated Poisson model with* $\epsilon = 0.05$, $\nu = 2$

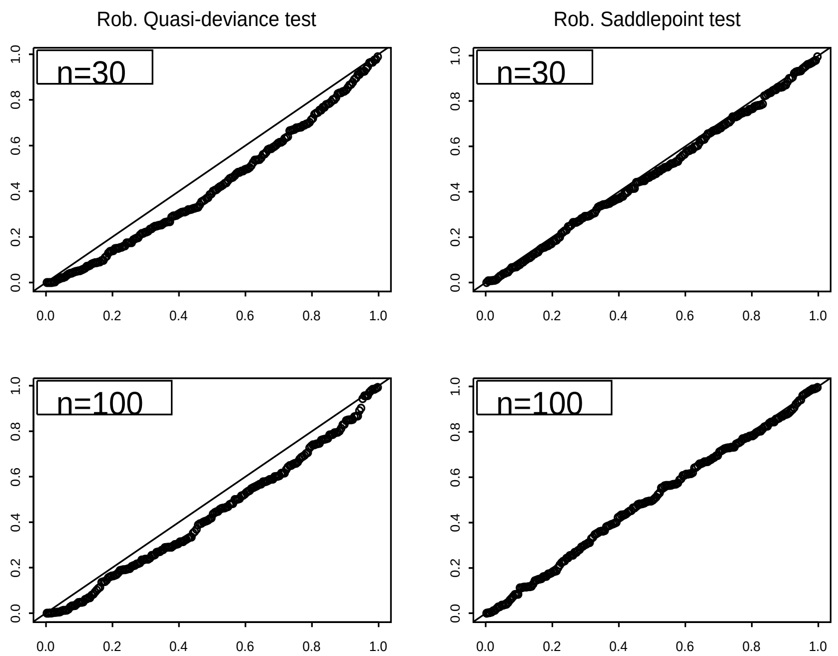

Figure 2 (b): *PP-plots of* **robust** $p-$*values vs.* $U[0,1]$ *when the data are generated from a contaminated Poisson model with* $\epsilon = 0.05$, $\nu = 2$

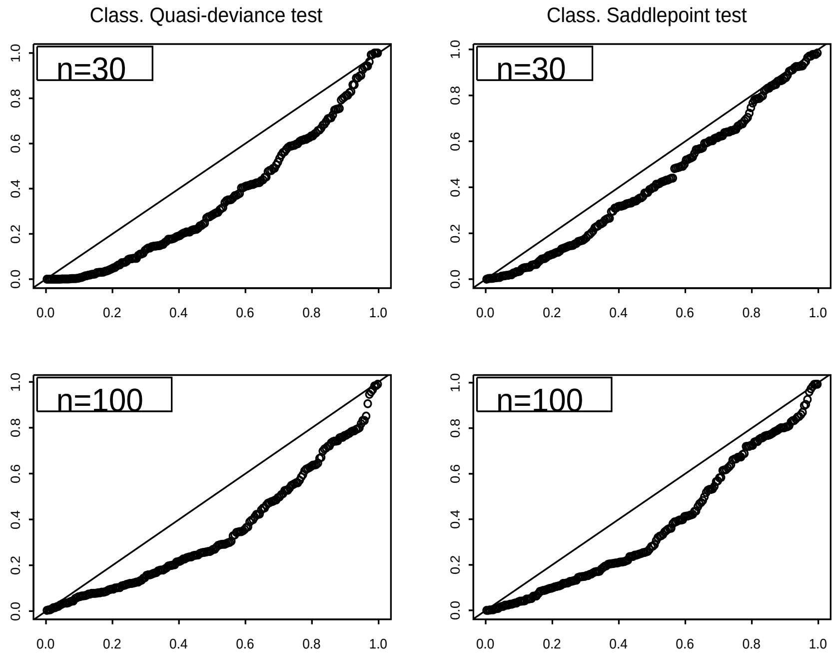

Figure 3 (a): *PP-plots of* **classical** *$p-$values vs. $U[0,1]$ when the data are generated from a contaminated Poisson model with $\epsilon = 0.10$, $\nu = 2$*

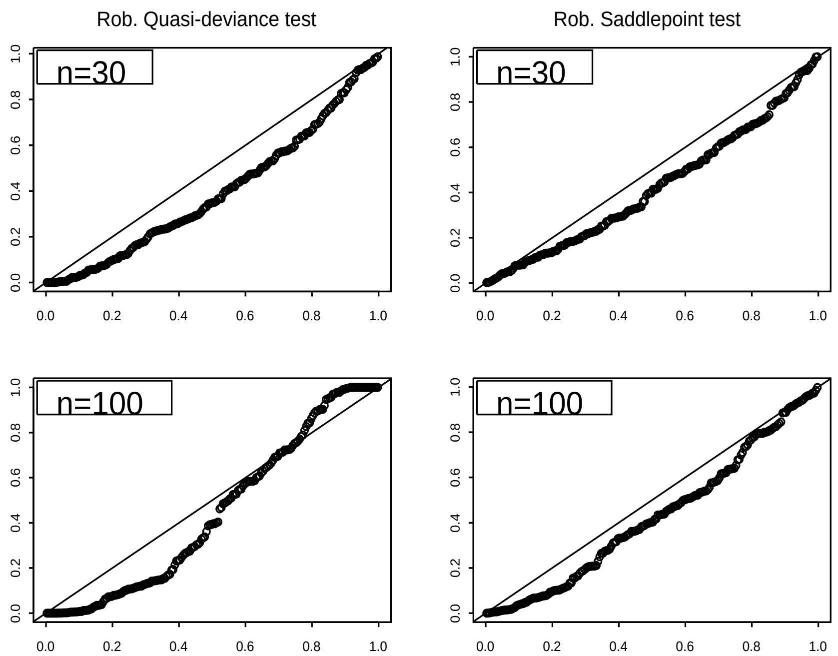

Figure 3 (b): *PP-plots of* **robust** *$p-$values vs. $U[0,1]$ when the data are generated from a contaminated Poisson model with $\epsilon = 0.10$, $\nu = 2$*

Table 1: Analysis of deviance for Crohn's disease data

Variable	P.val (class.)	P.val (rob. as.)	P.val (rob. sad.)
NULL	-	-	-
Dose 1	**0.010**	**0.007**	**0.019**
Dose 2	0.408	0.798	0.730
BMI	$<$ **0.0001**	**0.007**	**0.0001**
HEIGHT	$<$ **0.0001**	**0.0008**	**0.0003**
COUNTRY	**0.003**	0.06	**0.009**
SEX	**0.001**	**0.0004**	$<$ **0.0001**
AGE	0.079	**0.045**	**0.043**
WEIGHT	0.401	**0.027**	0.291

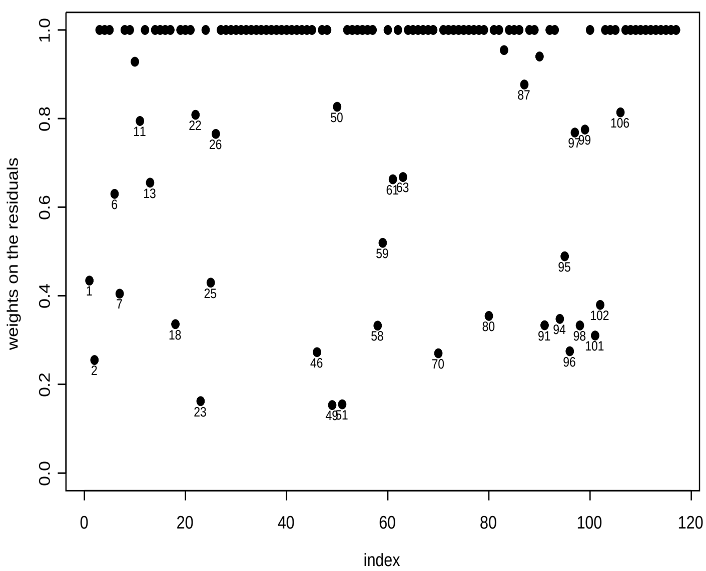

Figure 4: *Plot of the robust weights for each observation*

References

Bianco, A. and Yohai, V. (1996), "Robust Estimation in the Logistic Regression Model," *In Robust Statistics, Data Analysis, and Computer Intensive Methods*, H. Rieder ed., Springer, Lecture Notes in Statistics 109, 17–34.

Butler, R. (2007), *Saddlepoint Approximations with Applications*, Cambridge: Cambridge University Press.

Cantoni, E. and Ronchetti, E. (2001), "Robust Inference for Generalized Linear Models," *Journal of the American Statistical Association*, 96, 1022–1030.

Croux, C. and Haesbroeck, G. (2003), "Implementing the Bianco and Yohai Estimator for Logistic Regression," *Computational Statistics and Data Analysis*, 44, 273–295.

Daniels, H. and Young, G. (1991), "Saddlepoint Approximation for the Studentized Mean, with an Application to the Bootstrap," *Biometrika*, 78, 169–179.

Davison, A., Hinkley, D., and Worton, B. (1995), "Accurate and Efficient Construction of Bootstrap Likelihoods," *Statistics and Computing*, 5, 257–264.

DiCiccio, T., Field, C., and Fraser, D. (1990), "Approximations for Marginal Tail Probabilities and Inference for Scalar Parameters," *Biometrika*, 77, 77–95.

Fan, J. and Field, C. (1995), "Marginal Densities of M-estimators," *Canadian Journal of Statistics*, 23, 185–197.

Field, C. and Ronchetti, E. (1990), *Small Sample Asymptotics*, IMS Monograph Series: Hayward, California.

Gatto, R. and Ronchetti, E. (1996), "General Saddlepoint Approximations of Marginal Densities and Tail Probabilities," *Journal of the American Statistical Association*, 91, 666–673.

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986), *Robust Statistics: The Approach on Influence Functions*, New York: Wiley.

Huber, P. (1981), *Robust Statistics*, New York: Wiley.

Jing, B.-Y. and Robinson, J. (1994), "Saddlepoint Approximations for Marginal and Conditional Probability of Transformed Variables," *The Annals of Statistics*, 22, 1115–1132.

Künsch, H., Stefanski, L., and Carroll, R. (1989), "Conditionally Unbiased Bounded Influence Estimation in General Regression Models, with Applications to Generalized Linear Models," *Journal of the American Statistical Association*, 84, 460–466.

McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, London: Chapman and Hall.

Morgenthaler, S. (1992), "Least-absolute-deviations Fits for Generalized Linear Models," *Biometrika*, 79, 747–754.

Pregibon, D. (1982), "Resistant Fits for Some Commonly Used Logistic Models with Medical Applications," *Biometrics*, 38, 485–498.

Robinson, J., Ronchetti, E., and Young, G. (2003), "Saddlepoint Approximations and Tests Based on Multivariate M-estimators," *The Annals of Statistics*, 31, 1154–1169.

Ronchetti, E. and Welsh, A. (1994), "Empirical Saddlepoint Approximations for Multivariate M-estimators," *Journal of the Royal Statistical Society*, Series B, 56, 313–326.

Ruckstuhl, A. and Welsh, A. (2001), "Robust Fitting of the Binomial Model," *The Annals of Statistics*, 29, 1117–1136.

Stefanski, L., Carroll, R., and Ruppert, D. (1986), "Optimally Bounded Score Functions for Generalized Linear Models, with Applications to Logistic Regression," *Biometrika*, 73, 413–425.

Tingley, M. and Field, C. (1990), "Small-sample Confidence Intervals," *Journal of American Statistical Association*, 85, 427–434.

Victoria-Feser, M.-P. (2002), "Robust Inference with Binary Data," *Psychometrica*, 67, 21–32.

Wang, S. (1993), "Saddlepoint Approximations in Conditional Inference," *Journal of Applied Probability*, 30, 397–404.