BIBLIOTECA DI SCIENZE STATISTICHE
SERVIZIO ... NAZIONALE
BID ______ BID ______
ACQ. ____/____ INV. 79377
COLL. SCOLL WP, CLASS. ______
2000-16

ROBUST INFERENCE FOR GENERALIZED LINEAR MODELS WITH APPLICATION TO LOGISTIC REGRESSION

G. Adimari, L. Ventura

2000.16

Dipartimento di Scienze Statistiche
Università degli Studi
Via S. Francesco, 33
35121 Padova

Dicembre 2000

Robust inference for generalized linear models with application to logistic regression

Gianfranco Adimari and Laura Ventura
Dipartimento di Scienze Statistiche, Università di Padova
Via S. Francesco, 33, 35121 Padova,
e-mail: `adimari@stat.unipd.it`, `ventura@stat.unipd.it`

Abstract

In this paper we consider a suitable scale adjustment of the estimating function which defines a class of robust M-estimators for generalized linear models. This leads to a robust version of the quasi-profile loglikelihood which allows the derivation of robust likelihood ratio-type tests for inference and model selection, with a standard asymptotic behaviour. An application to logistic regression is discussed.

Keywords: Likelihood ratio test, Logistic regression, M-estimator, Quasi-likelihood, Robustness.

1 Introduction

Generalized linear models (GLM) (McCullagh and Nelder, 1989) are a technique for modeling the relationship between $p+1$ predictors $\boldsymbol{x}_i = (1, x_{1i}, \ldots, x_{pi})^T$ and a function of the means μ_i of continuous or discrete response variables y_i, for $i = 1, \ldots, n$. More precisely, a GLM assumes that

$$\eta_i = g(\mu_i) = \boldsymbol{\beta}^T \boldsymbol{x}_i,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is a vector of unknown parameters belonging to $\mathbb{R}^{p+1}$ and $g(\cdot)$ is the link function. Usually, the response variables are supposed to come from a distribution belonging to the exponential family.

It is well known that the maximum likelihood estimator for $\boldsymbol{\beta}$ is not robust as well as the usual quasi-likelihood estimator (see e.g. McCullagh and Nelder, 1989, Heyde, 1997), both defined as solutions of estimating equations of the form

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)}{V(\mu_i)} \boldsymbol{\mu}_i^{'} = \mathbf{0} \,, \tag{1}$$

where $\mu_i = g^{-1}(\boldsymbol{\beta}^T \boldsymbol{x}_i)$, $\boldsymbol{\mu}_i^{'} = \partial\mu_i/\partial\boldsymbol{\beta}^T$ and $V(\mu_i) = Var(y_i)$ is the variance function, which is assumed to be known in the quasi-likelihood approach. In fact, these estimators are M-estimators with unbounded influence function (see e.g. Hampel *et al.*, 1986). Therefore, large deviations of the response from its mean or outlying points in the explanatory variables may have a large influence on the estimators. In view of this, estimators with better robustness properties are needed and several robust alternatives have been proposed in the literature (see e.g. Pregibon, 1982, Stefanski *et al.*, 1986, Künsch *et al.*, 1989, Carroll and Pederson, 1993 and Preisser and Qaqish, 1999). In particular, Preisser and Qaqish (1999) propose a class of robust estimators in the more general setup of generalized estimating equations.

Starting from the class of robust estimators discussed by Preisser and Qaqish (1999), Cantoni and Ronchetti (2000) propose robust tools for inference for the whole class of GLM, based on a natural generalization of the quasi-likelihood approach. In particular, Cantoni and Ronchetti define robust deviances and present related tests for model selection playing the same role as the classical tests based on quasi-deviances. However, unlike their classical counterpart, these tests do not have a standard asymptotic distribution. This is because the estimating function which defines the class of M-estimators considered by Cantoni and Ronchetti does not satisfy a condition equivalent to the second Bartlett identity.

The aim of this paper is to discuss a suitable scale adjustment of the estimating function defining the class of robust M-estimators considered by Cantoni and Ronchetti (2000). Such an adjustment allows the derivation of a robust quasi-profile loglikelihood function which may be used as an ordinary profile loglikelihood to make inference about a scalar parameter of interest, in the presence of nuisance parameters. In particular, the related likelihood ratio-type tests for inference and model selection present a standard asymptotic behaviour, as in the classical framework.

2 M-estimators and robust inference

Cantoni and Ronchetti (2000) consider a class of Mallows type robust estimators for GLM, where the influence of deviations on the response and on the predictors are bounded separately. A Mallows quasi-likelihood estimator for $\boldsymbol{\beta}$ is defined as a solution $\hat{\boldsymbol{\beta}}$ of the estimating equation

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\psi}(y_i, \mu_i) = \mathbf{0} \ , \tag{2}$$

with

$$\boldsymbol{\psi}(y_i, \mu_i) = \left[\psi_k(r_i) w(\boldsymbol{x}_i) \frac{1}{\sqrt{V(\mu_i)}} \boldsymbol{\mu}_i' - \boldsymbol{a}(\boldsymbol{\beta}) \right] ,$$

where $\psi_k(u) = \min\{k, \max\{u, -k\}\}$ is the Huber ψ-function, for fixed $k > 0$, and $r_i = (y_i - \mu_i)/\sqrt{V(\mu_i)}$ are the Pearson residuals. The correction term $\boldsymbol{a}(\boldsymbol{\beta}) = (1/n)\sum_{i=1}^{n} E[\psi_k(r_i)] w(\boldsymbol{x}_i) \boldsymbol{\mu}_i'/\sqrt{V(\mu_i)}$ ensures the Fisher consistency of the estimator $\hat{\boldsymbol{\beta}}$ and can be computed explicitly for the binomial, the Poisson and the logistic models; here the expectation is taken with respect to the conditional distribution of $y|\boldsymbol{x}$. The estimating equation (2) for GLM has a structure suggested by the classical quasi-likelihood equation (1) and is a special case of an estimating equation given in Preisser and Qaqish (1999) in a more general setup.

The influence function for the M-estimator defined by (2) is $M^{-1}\boldsymbol{\psi}(y, \mu)$, where $M = X^T B X/n$, with $B = \mathrm{diag}(b_1, \ldots, b_n)$,

$$b_i = E\left[\psi_k(r_i) \left(\frac{\partial}{\partial \mu_i} \log h(y_i, \mu_i) \right) \right] \frac{w(\mathbf{x}_i)}{\sqrt{V(\mu_i)}} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

and $h(\cdot,\cdot)$ is the conditional density or probability of $y_i|\boldsymbol{x}_i$. The shape of the function $\psi(\cdot,\cdot)$ ensures robustness by putting a bound on the influence function. In particular, the function $\psi_k(\cdot)$ controls deviations in the y-space and leverage points are down-weighted by the weights $w(\boldsymbol{x}_i)$. A simple choice for $w(\boldsymbol{x}_i)$ is $\sqrt{1-h_i}$, where h_i is the i-th diagonal element of $H = X(X^TX)^{-1}X^T$ and X denotes the design matrix. This choice is suggested by the classical linear models theory (see Staudte and Sheather, 1990, Sec. 7). More sophisticated choices for $w(\cdot)$ are also available (see Cantoni and Ronchetti, 2000, for a discussion and some references). Subject to some regularity conditions, $\hat{\boldsymbol{\beta}}$ admits an asymptotic normal distribution with mean $\boldsymbol{\beta}$ and variance $M^{-1}QM^{-1}$, where $Q = X^TAX/n - \boldsymbol{a}(\boldsymbol{\beta})\boldsymbol{a}(\boldsymbol{\beta})^T$, with $A = \text{diag}(a_1, \ldots, a_n)$ and

$$a_i = E[\psi_k(r_i)^2]\frac{w^2(\mathbf{x}_i)}{V(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 .$$

Equation (2) corresponds to the minimization with respect to $\boldsymbol{\beta}$ of the quantity

$$l_M(\boldsymbol{\beta}) = \int^{\boldsymbol{\beta}} \boldsymbol{\Psi}(\boldsymbol{b})d\boldsymbol{b} ,$$

which can be seen as the robust counterpart of the classical quasi-loglikelihood function. Cantoni and Ronchetti (2000) use $l_M(\boldsymbol{\beta})$ to define robust quasi-deviances and obtain robust tests for model selection. Such tests are generalisations of the quasi-deviance tests for GLM. In particular, in order to test $H_0 : \boldsymbol{\beta}_{(1)} = \mathbf{0}$ against $H_1 : \boldsymbol{\beta}_{(1)} \neq \mathbf{0}$, where $\boldsymbol{\beta}_{(1)}$ is a set of $q \leq p$ components of $\boldsymbol{\beta}$, an adequate robust statistic based on $l_M(\boldsymbol{\beta})$ is

$$\Lambda_M = 2(l_M(\hat{\boldsymbol{\beta}}) - l_M(\dot{\boldsymbol{\beta}})) ,$$

where $\dot{\boldsymbol{\beta}}$ is the estimate of $\boldsymbol{\beta}$ under H_0. In view of the structure of the function $\psi(\cdot,\cdot)$, $l_M(\cdot)$ takes the form

$$l_M(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int^{\mu_i} \left\{ \psi_k\left(\frac{y_i - t}{\sqrt{V(t)}}\right) - E\left[\psi_k\left(\frac{y_i - t}{\sqrt{V(t)}}\right)\right]\right\} \frac{w(\boldsymbol{x}_i)}{\sqrt{V(t)}}\, dt,$$

where $\mu_i = g^{-1}(\boldsymbol{\beta}^T\boldsymbol{x}_i)$. Consequently, the computation of the statistic Λ_M involves n one-dimensional integrations, which can be performed numerically.

Observe that $\Lambda_M = 2[l_{MP}(\hat{\boldsymbol{\beta}}_{(1)}) - l_{MP}(\mathbf{0})]$, where $l_{MP}(\cdot)$ is the quasi-profile loglikelihood for $\boldsymbol{\beta}_{(1)}$.

A difficulty in using tests based on $l_M(\boldsymbol{\beta})$ is that, unlike their classical counterpart, they do not have a standard asymptotic χ^2 distribution. Indeed, Λ_M is asymptotically distributed as a linear combination of q independent χ_1^2 variables, whose coefficients are the eigenvalues of a suitable matrix. In general, these coefficients depend on the unknown parameter $\boldsymbol{\beta}$; see Proposition 1 in Cantoni and Ronchetti (2000) and Hèritier and Ronchetti (1994).

The discrepancy between the asymptotic behaviour of quasi-likelihood ratio-type tests in the classical and in the robust framework occurs because $l_M(\boldsymbol{\beta})$ does not verify the relation

$$Var\left(\boldsymbol{\Psi}(\boldsymbol{\beta})\right) = -E\left(\frac{\partial}{\partial \boldsymbol{\beta}^T}\boldsymbol{\Psi}(\boldsymbol{\beta})\right),$$

that is known as the second Bartlett identity when $\boldsymbol{\Psi}(\boldsymbol{\beta})$ is the usual score function. However, following Adimari and Ventura (2000), for inference on a scalar component of $\boldsymbol{\beta}$, it is possible to modify the estimating function in (2) yielding to quasi-profile loglikelihood functions with the standard asymptotic behaviour.

Let β_j, a scalar component of $\boldsymbol{\beta}$, be the parameter of interest. Let $\Psi_{\beta_j}(\beta_j, \boldsymbol{\lambda})$ be the estimating function corresponding to β_j. Here $\boldsymbol{\lambda}$ indicates the vector $\boldsymbol{\beta}$ without its j-th element. The adjusted quasi-profile loglikelihood function for β_j (Adimari and Ventura, 2000) can be written as

$$l_{QP}(\beta_j) = \int^{\beta_j} \omega(b, \hat{\boldsymbol{\lambda}}_b)\Psi_{\beta_j}(b, \hat{\boldsymbol{\lambda}}_b)db \ , \tag{3}$$

with

$$\omega(\beta_j, \boldsymbol{\lambda}) = \frac{M_{\beta_j,\beta_j} - \boldsymbol{\xi}_{\beta_j}^T M_{(-j)}^{-1}\boldsymbol{\xi}_{\beta_j}}{Q_{\beta_j,\beta_j} - 2\boldsymbol{\xi}_{\beta_j}^T M_{(-j)}^{-1}\boldsymbol{\zeta}_{\beta_j} + \boldsymbol{\xi}_{\beta_j}^T M_{(-j)}^{-1} Q_{(-j)} M_{(-j)}^{-T}\boldsymbol{\xi}_{\beta_j}} \ ,$$

where M_{β_j,β_j} is the j-th diagonal element of the matrix M, $\boldsymbol{\xi}_{\beta_j}$ is the j-th column of M without its j-th element, $M_{(-j)}$ denotes M without the j-th column and the j-th row, $\boldsymbol{\zeta}_{\beta_j}$ is the j-th column of Q without its j-th element and $\hat{\boldsymbol{\lambda}}_{\beta_j}$ is the estimate of $\boldsymbol{\lambda}$ for β_j fixed. Function (3) is obtained by a scale adjustment which corrects the quasi-profile score $\Psi_{\beta_j}(\beta_j, \hat{\lambda}_{\beta_j})$ to have information bias at the proper order $O(1)$ (see Adimari and Ventura, 2000, and McCullagh and Tibshirani, 1990). As a consequence, $l_{QP}(\beta_j)$ has similar properties to the ordinary profile loglikelihood. In particular, for setting quasi-likelihood confidence regions or for testing hypotheses, the adjusted quasi-likelihood ratio statistic

$$W_{QP}(\beta_j) = 2\left(l_{QP}(\hat{\beta}_j) - l_{QP}(\beta_j)\right) = 2\int_{\beta_j}^{\hat{\beta}_j} \omega(b, \hat{\lambda}_b)\Psi_{\beta_j}(b, \hat{\lambda}_b)db \qquad (4)$$

may be used. Unlike Λ_M, under $H_0 : \beta_j = 0$ and usual regularity conditions, $W_{QP}(0)$ is approximately χ_1^2 distributed (see Adimari and Ventura, 2000) and, for instance, asymptotic confidence regions with nominal coverage $1 - \alpha$ for β_j can be constructed as $\{\beta_j : W_{QP}(\beta_j) \leq \chi^2_{1;1-\alpha}\}$, where $\chi^2_{1;1-\alpha}$ is the $(1-\alpha)$−quantile of the χ_1^2 distribution. Alternatively, the quasi-directed likelihood $r_{QP}(\beta_j) = \text{sgn}(\hat{\beta}_j - \beta_j)\sqrt{W_{QP}(\beta_j)}$, which is approximately standard normal, may be used. Observe that statistic (4) suffices to check the significance of one variable in a selection model procedure.

It must be noted that the adjustment of $\Psi_{\beta_j}(\beta_j, \hat{\lambda}_{\beta_j})$ by the factor $\omega(\cdot, \cdot)$ leaves the M-estimator for β_j unchanged. Consequently, the robustness properties are maintained. According to the results obtained in Cantoni and Ronchetti (2000) and Heritier and Ronchetti (1994), such robustness properties will carry over to quasi-likelihood based inferential procedures.

3 Application: Logistic Regression

In this section we consider the U.S. Food Stamp data previously analyzed by Stefanski *et al.* (1986), Künsch *et al.* (1989), Carroll and Pederson (1993), in the framework of robust estimation, and Hèritier and Ronchetti (1994). For these data, the response (y) indicates partecipation in the Federal Food Stamp program and the covariates employed for study include two dichotomous variables, i.e. tenancy (x_1) and supplemental income (x_2), and a logarithmic tranformation of monthly income [log(monthly income + 1)] (x_3). The data consists on observations on 150 individuals. Previous analyses show that data contain at least a leverage point (case 5); some authors also suggest that case 66 is somewhat outlying.

Consider the logit model

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \ .$$

We have $\mu_i = pr(y_i = 1 \mid \boldsymbol{x}_i) = e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}/(1+e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})$, with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and $\boldsymbol{x}_i = (1, x_{1i}, x_{2i}, x_{3i})^T$, $V(\mu_i) = \mu_i(1-\mu_i)$ and $\boldsymbol{\mu}_i' = \mu_i(1-\mu_i)\boldsymbol{x}_i$. The Mallows quasi-likelihood estimator is defined by the estimating function

$$\boldsymbol{\Psi}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[\psi_k\left(\frac{y_i-\mu_i}{\sqrt{\mu_i(1-\mu_i)}}\right) w(\boldsymbol{x}_i)\sqrt{\mu_i(1-\mu_i)}\ \boldsymbol{x}_i - \boldsymbol{a}(\boldsymbol{\beta})\right] \ , \qquad (5)$$

where

$$\boldsymbol{a}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\left[\psi_k\left(\sqrt{e^{-\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right)\mu_i - \psi_k\left(\sqrt{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right)(1-\mu_i)\right] w(\boldsymbol{x}_i)\sqrt{\mu_i(1-\mu_i)}\ \boldsymbol{x}_i \ .$$

The diagonal elements of the matrices A and B are given by

$$a_i = \left[\psi_k^2\left(\sqrt{e^{-\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right)\mu_i - \psi_k^2\left(\sqrt{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right)(1-\mu_i)\right] w^2(\boldsymbol{x}_i)\mu_i(1-\mu_i)$$

and

$$b_i = \left[\psi_k\left(\sqrt{e^{-\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right) + \psi_k\left(\sqrt{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}\right)\right] w(\boldsymbol{x}_i)[\mu_i(1-\mu_i)]^{3/2} \ .$$

The parameter of interest is the fourth component β_3, which corresponds to monthly income. In fact, previous analyses show that this coefficient changes considerably its numerical value and significance level if case 5 is deleted. For this data set, the Mallows estimator defined by (5) with $k = 1.55$ and $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$ produces very similar results to the conditional estimators proposed in Künsch *et al.* (1989) and appears to be sufficiently robust. In particular, for the component β_3 the Mallows estimator $\hat{\beta}_3$ yields the fit -1.18 with estimated standard error 0.50. Conversely, the classical logistic maximum likelihood estimator yields the fit -0.33 with estimated standard error 0.27.

To illustrate the use of the adjusted quasi-profile loglikelihood (3), we test the hypothesis

$$H_0 : \beta_3 = 0 \qquad (6)$$

against the alternative $H_1 : \beta_3 \neq 0$. By the results given above, the Wald-type test based on the Mallows estimator $\hat{\beta}_3$ with $k = 1.55$ and $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$ rejects the hypothesis (6) at the usual 5% significance level. In contrast, the related test based on the adjusted quasi-likelihood ratio statistic (4) performs similarly to the logistic Wald test and gives a p-value of 0.135. The discrepancy between these results can be explained by looking at Figure 1, which gives the quasi-profile score $\Psi_{\beta_3}(\beta_3, \hat{\boldsymbol{\lambda}}_{\beta_3})$ and the adjusted quasi-profile score (dashed line) $\omega(\beta_3, \hat{\boldsymbol{\lambda}}_{\beta_3})\Psi_{\beta_3}(\beta_3, \hat{\boldsymbol{\lambda}}_{\beta_3})$ for β_3, when $k = 1.55$ and:

(a) $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$;

(b) $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$ and case 5 is removed;

(c) $w(\boldsymbol{x}_i) = (1-h_i)^2$.

In fact, Figure 1 shows the influence of observation 5 on the estimate $\hat{\beta}_3$ and on the shape of the adjusted quasi-profile score. It may be noted that, when $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$, the shape of the adjusted quasi-profile score is still greatly influenced by the observation 5, unlike the numerical value of the estimate $\hat{\beta}_3$. Since the adjusted quasi-likelihood ratio statistic is twice the area under the dashed line between $\hat{\beta}_3$ and 0, it is clear why the likelihood ratio-type test fails to reject the null hypothesis when $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$. This result suggests that more care is necessary in choosing the weight function and confirms some limits of automatic methods for measuring leverage. In particular, in this case it seems that the function $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$ does not down-weight observation 5 enough. For comparison, the following table shows the weights given to cases 5 and 66 by three different functions, namely $\sqrt{1-h_i}$, $(1-h_i)^2$ and $1-\sqrt{h_i}$.

case	$\sqrt{1-h_i}$	$(1-h_i)^2$	$1-\sqrt{h_i}$
5	0.828	0.471	0.440
66	0.992	0.969	0.875

If we use the adjusted quasi-likelihood ratio statistic based on (5) with $k = 1.55$ and $w(\boldsymbol{x}_i) = (1-h_i)^2$ we obtain a p-value of 0.022, which leads to the rejection of the hypothesis (6).

The adjusted quasi-likelihood approach can also be used to construct confidence intervals. As an example, Figure 2 gives the plot of the adjusted quasi-profile loglikelihood ratio function $W_{QP}(\beta_3)$ computed from (5) with $k = 1.55$

and $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$. The horizontal line shown is at the asymptotically justified 95% level. Confidence intervals obtained by this method do not present the predetermined symmetry which appears when one uses the classical technique based on the asymptotic distribution of $\hat{\beta}_3$. A partial simulation experiment (based on 5000 Monte Carlo trials) has also been made to evaluate coverage probabilities of the nominal 1-α confidence intervals for β_3 obtained by the adjusted quasi-profile loglikelihood. For this experiment, responses y_i have been generated according to the model $pr(y_i = 1 \mid \boldsymbol{x}_i) = e^{\boldsymbol{\beta}_*^T \boldsymbol{x}_i}/(1 + e^{\boldsymbol{\beta}_*^T \boldsymbol{x}_i})$, where $\boldsymbol{\beta}_* = (6, -1.8, 0.7, -1.2)^T$. For nominal 0.90, 0.95 and 0.99 coverage probabilities we obtained empirical coverages probabilities 0.898, 0.953 and 0.988, respectively.

4 Conclusion

The adjusted quasi-profile loglikelihood obtained from an estimating function defining Mallows type robust estimators for GLM (Cantoni and Ronchetti, 2000) is appealing. It represents a robust version of the profile loglikelihood and allows robust inference and model selection in a standard way. In particular, it allows the derivation of robust likelihood ratio-type tests and confidence regions by using a standard χ^2_1 approximation. In addition, the application example discussed in Section 3 shows that this tool is also useful to evaluate the resistance properties of the underlying estimating function.

References

Adimari, G., Ventura, L. (2000), Quasi-profile loglikelihoods for unbiased estimating equations, *Working Paper* n. 2000.4, Dipartimento di Scienze Statistiche, Università di Padova. Revised for *Ann. Inst. Statist. Math.*.

Cantoni, E., Ronchetti, E. (2000), Robust inference for generalized linear models, *Working Paper*, Department of Econometrics, University of Geneva. Revised for *Jour. Amer. Statist. Assoc.*.

Carroll, R.J., Pederson, S. (1993), On robustness in the logistic regression model, *J. R. Statist. Soc.* B, **55**, 693-706.

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

Héritier, S., Ronchetti, E. (1994), Robust bounded-influence tests in general parametric models, *Jour. Amer. Statist. Assoc.*, **89**, 897-904.

Heyde, C.C. (1997), *Quasi-Likelihood and its Application*, Springer-Verlag, New York.

Künsch, H.R., Stefanski, L.A., Carroll, R.J. (1989), Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models, *Jour. Amer. Statist. Assoc.*, **84**, 460-466.

McCullagh, P., Nelder, J.A. (1989), *Generalized Linear Models*, 2nd Ed., Chapman and Hall, London.

McCullagh, P. and Tibshirani, R. (1990), A simple method for the adjustment of profile likelihoods, *J. R. Statist. Soc.* B, **52**, 325-344.

Pregibon, D. (1982), Resistant fits for some commonly used logistic models with medical applications, *Biometrics*, **55**, 574-579.

Preisser, J.S., Qaqish, B.F. (1999), Robust regression for clustered data with applications to binary regression, *Biometrics*, **55**, 574-579.

Staudte, R.G., Sheather, S.J. (1990), *Robust Estimation and Testing*, Wiley, New York.

Stefanski, L.A., Carroll, R.J., Ruppert, D. (1986), Optimally bounded score functions for generalized linear models with applications to logistic regression, *Biometrika*, **73**, 413-424.

Figure 1: *Quasi-profile score and adjusted quasi-profile score (dashed line) for* β_3 *when* $k = 1.55$ *and* (a) $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$, (b) $w(\boldsymbol{x}_i) = \sqrt{1-h_i}$ *and case 5 is removed,* (c) $w(\boldsymbol{x}_i) = (1-h_i)^2$.

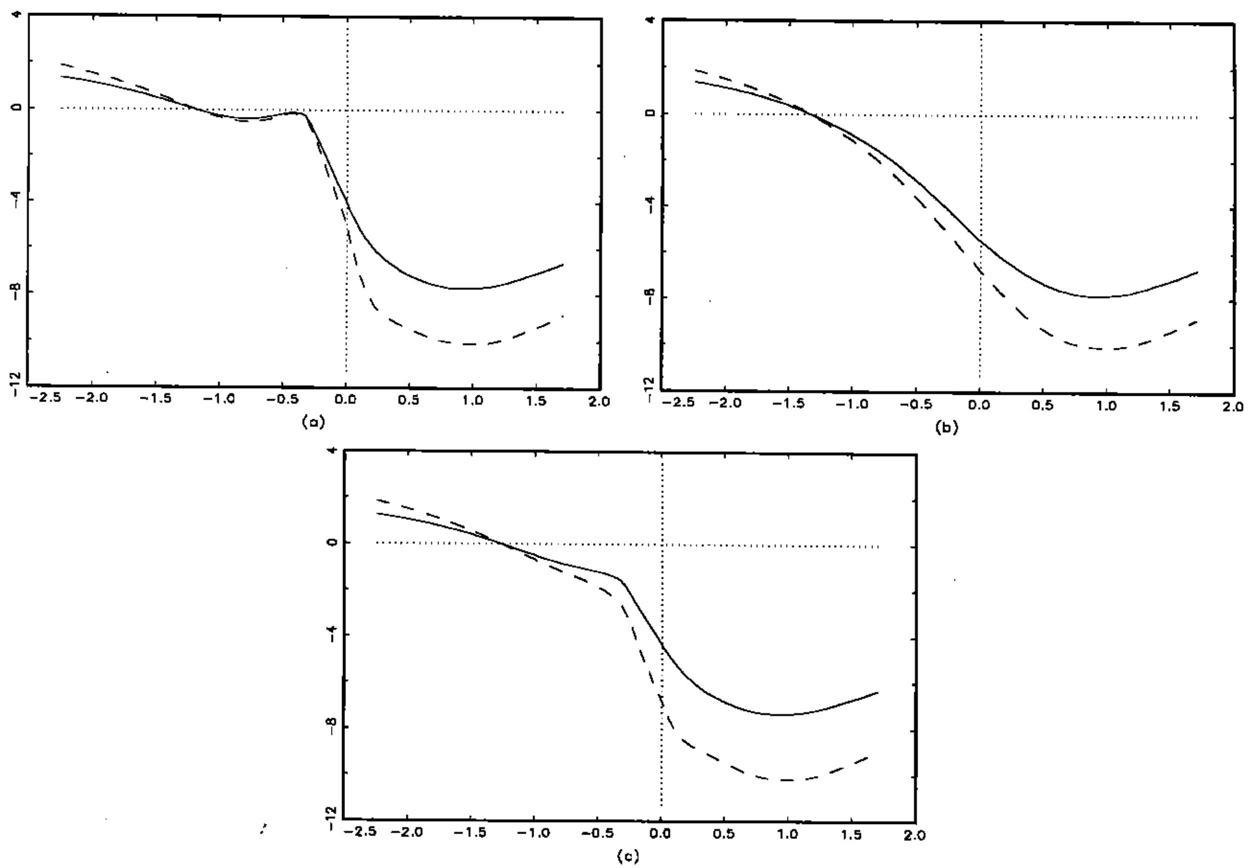

Figure 2: *Adjusted quasi-profile loglikelihood ratio function* W_{QP} *for the parameter* β_3 *when* $k = 1.55$ *and* $w(\boldsymbol{x}_i) = \sqrt{1 - h_i}$. *The horizontal line is at the asymptotically justified 95% level.*