Biometrika (1992), **79**, 4, pp. 747-54
Printed in Great Britain

Least-absolute-deviations fits for generalized linear models

BY STEPHAN MORGENTHALER

Department of Mathematics, Swiss Federal Institute of Technology,
1015 Lausanne, Switzerland

SUMMARY

The fitting by quasi-likelihoods is based on Euclidean distance and thereby related to the least-squares norm. This paper examines the consequences of replacing the L_2-norm by the L_1-norm in the derivation of quasi-likelihoods. Since the least-absolute-deviations centre of a distribution is its median rather than its mean, the natural models for the L_1-fitting involve medians. However, even if we model the mean response rather than the median response, an L_1-type criterion is applicable and leads to alternatives for maximum likelihood fits.

Some key words: Generalized linear model; L_1-fit; quasi-likelihood.

1. INTRODUCTION

Quasi-likelihoods are used in fitting mean responses as a function of explanatory variables. They go beyond the weighted least-squares method in allowing for a dependence of the variances of the responses on the unknown means. Denote by $y=(y_1,\ldots,y_n)^{\mathrm{T}}$ the observed responses and let $x_1,\ldots,x_n$ be the p-vectors of the values of the corresponding explanatory variables. Let $\mu=(\mu_1,\ldots,\mu_n)^{\mathrm{T}}$ be the mean of y. A quasi-likelihood function is defined by way of its gradient

$$\frac{\partial}{\partial\mu}K(\mu;y)=V(\mu)^{-1}(y-\mu), \tag{1·1}$$

where $V(\mu)$ denotes the variance matrix of the vector of responses. If V is constant, this leads to the usual weighted least-squares solution. If we postulate a structural model of the form

$$g\{\mu_i(\beta)\}=x_i^{\mathrm{T}}\beta \quad (i=1,\ldots,n)$$

for some monotone and continuously differentiable link function $g(.)$, the quasi-likelihood approach consists in estimating the unknown β by the solution of the system of equations

$$U(\beta)=D^{\mathrm{T}}V(\mu)^{-1}(y-\mu)=0. \tag{1·2}$$

The matrix D contains the partial derivatives $\partial\mu_i/\partial\beta_j=D_{ij}$. It is clear that this estimator is consistent if the mean μ is correctly specified, since

$$E\{D^{\mathrm{T}}V(\mu)^{-1}(Y-\mu)\}=0,$$

for all random vectors $Y=(Y_1,\ldots,Y_n)^{\mathrm{T}}$ with $E(Y)=\mu$. Linearizing (1·2) shows that assuming certain regularity conditions

$$0=U(\hat\beta)=U(\beta)-H(\hat\beta-\beta)+o_p(n^{-\frac{1}{2}}),$$

where $H=E_\beta(-\partial U/\partial\beta)$. It follows that

$$\operatorname{var}_\beta(\hat\beta)=H^{-1}\operatorname{var}\{U(\beta)\}(H^{-1})^{\mathrm{T}}. \tag{1·3}$$

In the case of a correctly specified quasi-likelihood model, i.e. if indeed the variance of the responses is equal to V, this formula leads to

$$\text{var}_\beta\,(\hat{\beta}) = \{D^{\mathrm{T}} V(\mu)^{-1} D\}^{-1}. \qquad (1\cdot4)$$

See Wedderburn (1974) and McCullagh & Nelder (1989, p. 327) for more details.

Historically, the least-absolute-deviations approach to combining observations emerged at about the same time as the least-squares approach. It is therefore quite natural to ask in what sense the quasi-likelihood approach, as we have briefly described it, can be put to use in conjunction with least-absolute-deviations instead of least-squares. This paper shows that replacing the L_2-norm in the definition of the quasi-likelihood by an arbitrarily chosen L_q-norm for $q \geqslant 1$ leads to biased estimating equations. This bias can be eliminated only in the case where the underlying distributions are assumed known. The corrected equations lead then to alternative estimates for parameters of generalized linear models. When $q = 1$ these are in some sense analogues of the L_1-estimate for normal models.

A quasi quasi-likelihood, i.e. an approach based on absolute-deviations without assuming complete knowledge of the underlying distributions, exists if we do not eliminate the bias. Such an approach can make sense for continuous error distributions. It essentially consists in modelling the median response rather than the mean response.

The paper is organized as follows. We begin by introducing alternative estimates for models with completely specified response distributions and then discuss quasi-likelihood fitting based on an absolute-deviation likelihood.

2. Robust fits for generalized linear models

2·1. *General theory*

Pregibon (1982), Stefanski, Carroll & Ruppert (1986) and Künsch, Stefanski & Carroll (1989) consider the robust estimation of generalized linear models parameters with particular emphasis on logistic regression. Since the original justification of the least-absolute-deviations principle is its resistance to gross errors, it is of interest to see first how this principle can be applied with traditional generalized linear models and to compare it to these existing robust procedures.

With a constant diagonal scatter matrix, $V = \text{diag}\,(V_1, \ldots, V_n)$, fitting by least-$L_q$-norm ($q \geqslant 1$) corresponds to minimizing

$$\sum_{i=1}^{n} \left| \frac{y_i - \mu_i}{V_i^{\frac{1}{2}}} \right|^q .$$

The corresponding gradient is

$$\{\text{diag}\,(V_1, \ldots, V_n)\}^{-q/2}\{|y-\mu|^{q-1}\,\text{sgn}\,(y-\mu)\}, \qquad (2\cdot1)$$

where sgn(.) denotes the sign-function that takes values ±1. The product $|y-\mu|^{q-1}\,\text{sgn}\,(y-\mu)$ in (2·1) is to be taken componentwise. With $q = 2$, this formula leads back to the quasi-likelihood (1·1) discussed in the introduction. But as a straightforward generalization of (1·1), (2·1) is a failure, since it does not lead to consistent estimates unless $q = 2$ or unless the responses are symmetrically distributed around their means. In general, one must correct for the asymmetry and that demands detailed knowledge of the underlying distribution. Simply specifying means and variances, as was possible with quasi-likelihoods, is not sufficient.

To obtain consistent estimates we calculate the correcting quantities

$$c_i = c_i(\mu_i) = E\{|Y_i - \mu_i|^{q-1}\,\text{sgn}\,(Y_i - \mu_i)\},$$

where the random variables Y_i follow the true underlying distribution with mean μ_i and variance V_i. Let $c=(c_1,\dots,c_n)^{\mathrm{T}}$. The corrected gradient is

$$\{\operatorname{diag}(V_1,\dots,V_n)\}^{-q/2}\{|y-\mu|^{q-1}\operatorname{sgn}(y-\mu)-c\},\tag{2·2}$$

and the estimating equation for β is

$$U_q(\beta)=D^{\mathrm{T}}\{\operatorname{diag}(V_1,\dots,V_n)\}^{-q/2}\{|y-\mu|^{q-1}\operatorname{sgn}(y-\mu)-c\}.\tag{2·3}$$

In models where the complete underlying distribution is specified, one can use this approach to derive alternatives to the usual maximum likelihood estimates. As $q\rightarrow 1$, the solution of (2·3) tends to a corrected L_1-fit of the model.

Of course, the estimation of standard errors via (1·3) leads to a formula that can be considerably more complicated than (1·4), namely

$$\operatorname{var}_\beta(\hat\beta)=(D^{\mathrm{T}}V^{-q/2}QD)^{-1}(D^{\mathrm{T}}V^{-q/2}RV^{-q/2}D)(D^{\mathrm{T}}V^{-q/2}QD)^{-1}.\tag{2·4}$$

Here R denotes the diagonal matrix with diagonal elements

$$R_{ii}=E(|Y_i-\mu_i|^{2q-2})-c_i^2,$$

and Q is also diagonal with elements

$$Q_{ii}=(q-1)E(|Y_i-\mu_i|^{q-2})-E\{|Y_i-\mu_i|^{q-1}\,\partial\operatorname{sgn}(Y_i-\mu_i)/\partial\mu_i\}+c_i'(\mu_i).$$

Since

$$\partial\operatorname{sgn}(y-\mu_i)/\partial\mu_i=-\partial\operatorname{sgn}(y-\mu_i)/\partial y=-\partial\{2h(y-\mu_i)-1\}/\partial y,$$

where $h(u)$ is the distribution function for a point mass at zero, the second term in this last formula can be interpreted as an integral with respect to a Dirac measure. One obtains zero except for $q=1$ where the result of the expectation is (-2) times the value of the underlying density evaluated at μ_i.

To illustrate these formulae, we next consider two examples and discuss computational aspects.

2·2. *Logistic regression*

If the ith response follows a Bernoulli distribution with probability of success μ_i, the correction for consistency discussed above is

$$c_i=E\{|Y_i-\mu_i|^{q-1}\operatorname{sgn}(Y_i-\mu_i)\}=(1-\mu_i)^{q-1}\mu_i-\mu_i^{q-1}(1-\mu_i).$$

The generalized quasi-likelihood gradient (2·2) is equal to

$$\{\operatorname{diag}(V_1,\dots,V_n)\}^{-q/2}[(y-\mu)\{(1-\mu)^{q-1}+\mu^{q-1}\}],$$

where we assume that the responses are uncorrelated. The variance is $V_i=\mu_i(1-\mu_i)$. If we model the means as

$$\mu_i=\frac{\exp(x_i^{\mathrm{T}}\beta)}{1+\exp(x_i^{\mathrm{T}}\beta)},$$

this leads to estimating the unknown β by the root of $U_q(\hat\beta)=0$, where

$$U_q(\beta)=D^{\mathrm{T}}V^{-q/2}[(y-\mu)\{(1-\mu)^{q-1}+\mu^{q-1}\}].$$

Since $D=VX$, the least-absolute-deviations estimate for this particular model therefore satisfies

$$X^{\mathrm{T}}V^{\frac{1}{2}}(y-\mu)=\sum_{i=1}^{n}\{\hat\mu_i(1-\hat\mu_i)\}^{\frac{1}{2}}(y_i-\hat\mu_i)x_i=0.$$

This estimating equation differs from the maximum likelihood equation only through the inclusion of the weight

$$\{\hat{\mu}_i(1-\hat{\mu}_i)\}^{\frac{1}{2}},$$

which tends to downweight observations at the fringe of the data set. Equation (2·4) leads to

$$\operatorname{var}_\beta(\tilde{\beta}) = (X^{\mathrm{T}}V^{3/2}X)^{-1}(X^{\mathrm{T}}V^2X)(X^{\mathrm{T}}V^{3/2}X)^{-1}.$$

This particular form of the least-absolute-deviations estimate may come as a surprise since we are used to such estimates being determined by equations involving the balance of residual signs. The Bernoulli case is obviously rather special. The likelihood equation itself can in fact be interpreted as providing balanced signs, since it is equivalent to

$$\textstyle\sum_{\{y_i=1\}}(1-\hat{\mu}_i)x_i = \sum_{\{y_i=0\}}\hat{\mu}_i x_i.$$

On the left-hand side of this equation are all observations with positive residuals, counted with weight $1-\hat{\mu}_i = 1-\hat{\mathrm{p}}\mathrm{r}\,(Y_i=1)$, whereas on the right-hand side are all the cases with negative residuals counted with weight $\hat{\mu}_i = 1-\hat{\mathrm{p}}\mathrm{r}\,(Y_i=0)$. The least-absolute-deviations criteria leads in this case simply to a different scheme for weighting. It is a bit surprising that the additional weights depend only on μ_i and do not take into account the observations, as is the case in the estimates considered in the papers cited at the beginning of § 2·1. Stefanski et al. (1986, p. 418, in particular equation (3·1)) consider simple robust estimates that are vaguely similar to ours in the sense that they make use of an additional weight depending only on x_i and β.

Example. As a practical example, consider the data on vaso-constriction of the skin given by Finney (1947, p. 322). This data set consists of $n=39$ observations of a Bernoulli variable in the presence of two explanatory variables, V, the volume of air inspired and R, the inspiration rate. The logistic model using variables log V and log R leads to the linear predictor

$$-2{\cdot}873 + 5{\cdot}177 \log V + 4{\cdot}559 \log R$$

with standard errors of 1·32, 1·86 and 1·83 for the fitted coefficients. The corrected L_1 criterion on the other hand gives a very different linear predictor, namely

$$-21{\cdot}6 + 34{\cdot}9 \log V + 28{\cdot}1 \log R$$

with standard errors of 13·5, 22·5 and 17·1. This drastic change is due to the fact that the negative responses are very nearly separated from the positive responses by a straight line in the log V, log R plane. The 4th and the 18th observations contain practically all the information about the rise of the logistic surface. In the scoring algorithm for the L_1-fit, they receive a huge final weight. For an alternative robust analysis of this data, see Pregibon (1982, p. 496), whose solution is close to the one obtained with $q=1{\cdot}5$.

2·3. *Gamma regression*

Let $y_1, \ldots, y_n$ be positive observations following densities of the form

$$f_i(y_i) = \Gamma(\nu_i)^{-1}(\nu_i y_i/\mu_i)^{\nu_i}\exp(-\nu_i y_i/\mu_i)(1/y_i). \qquad (2{\cdot}5)$$

The correcting quantity for $q=1$ is

$$\begin{aligned} c_i &= \int_{\mu_i}^{\infty} f_i(y)\,dy - \int_0^{\mu_i} f_i(y)\,dy \\ &= 1 - 2\int_0^{\nu_i} y^{\nu_i - 1}\Gamma(\nu_i)^{-1}\exp(-y)\,dy, \end{aligned}$$

which depends only on ν_i. If $\nu_i \equiv \nu$ is constant for all cases, we have $c_1 = c_2 = \ldots = c_n$ and the least-absolute-deviations estimating equation is

$$U_1(\hat\beta) = D^{\mathrm{T}}\mathrm{diag}\,(\hat\mu_i^2/\nu)^{-\frac{1}{2}}\{\mathrm{sgn}\,(y-\hat\mu) - c\} = 0, \tag{2·6}$$

where $D_{ij} = \partial\mu_i/\partial\beta_j$ and $c = (c_1, \ldots, c_1)^{\mathrm{T}}$. This estimation equation must be interpreted as a limit for $q \to 1$. At $q=1$ it may not have a solution due to the discontinuity of the sign-function. The scoring algorithm based on (2·6) will work, however, if special care is taken. To use this estimator, we need a simultaneous estimate of ν, since the correcting quantity c_1 is a function thereof. Any $n^{\frac{1}{2}}$-consistent estimate of ν will do this job without affecting the asymptotic characteristics of the estimate of β. The matrices Q and R from (2·4) are diagonal with elements $R_{ii} = 1 - c_1^2$ and $Q_{ii} = 2f_i(\mu_i)$, for $i = 1, \ldots, n$.

For example, when $\mu_i = \exp(x_i^{\mathrm{T}}\beta)$, we have $D = \mathrm{diag}\,(\mu_1, \ldots, \mu_n)X$ and the estimating equation is

$$X^{\mathrm{T}}\{\mathrm{sgn}\,(y-\hat\mu) - c\} = 0, \tag{2·7}$$

which is a regression quantile equation (Koenker & Bassett, 1978). The positive residuals obtain a weight of $(1-c_1)$, whereas the negative residuals get weight $(-1-c_1)$. The term $-c$ in (2·7) corrects for the fact that μ_i is not the median of the density f_i but rather the $((1-c_1)/2)$-quantile.

It is clear that, in this model, the least-absolute-deviations estimator will curb the influence of outlying observations as long as they do not occur at positions of high leverage.

2·4. *Computation*

Suppose the estimating equation is of the form

$$U(\hat\beta) = \sum_{i=1}^{n} w(\hat\mu_i)x_i(y_i - \hat\mu_i) = 0, \tag{2·8}$$

where $w(.)$ is some weight function and x_i is the vector of explanatory variables for the ith case. The Newton-Raphson algorithm is based on the derivative

$$\frac{\partial U_k}{\partial\beta_l} = \sum_{i=1}^{n}\{w'(\mu_i)h'(x_i^{\mathrm{T}}\beta)x_{il}x_{ik}(y_i - \mu_i) - w(\mu_i)x_{il}x_{ik}h'(x_i^{\mathrm{T}}\beta)\},$$

where we assume that $\mu_i = h(x_i^{\mathrm{T}}\beta) = g^{-1}(x_i^{\mathrm{T}}\beta)$. This leads to the linear expansion

$$U(\beta_{\mathrm{old}}) - (X^{\mathrm{T}}MX)(\beta_{\mathrm{new}} - \beta_{\mathrm{old}}) = 0,$$

where M is diagonal with elements

$$M_{ii} = h'(x_i^{\mathrm{T}}\beta)\{w(\mu_i) - w'(\mu_i)(y_i - \mu_i)\} \quad (i = 1, \ldots, n). \tag{2·9}$$

The Newton-Raphson update is

$$\begin{aligned} \beta_{\mathrm{new}} &= \beta_{\mathrm{old}} + (X^{\mathrm{T}}MX)^{-1}U(\beta_{\mathrm{old}}) \\ &= (X^{\mathrm{T}}MX)^{-1}\{X^{\mathrm{T}}MX\beta_{\mathrm{old}} + X^{\mathrm{T}}W(y-\mu)\} \\ &= (X^{\mathrm{T}}MX)^{-1}(X^{\mathrm{T}}M)\{X\beta_{\mathrm{old}} + M^{-1}W(y-\mu)\}, \end{aligned}$$

where W is the diagonal matrix with the weights $w(\mu_i)$ on the diagonal. In the last equation, the updating step is written as the solution of a weighted least-squares problem with weight matrix M and response vector $X\beta_{\text{old}}+M^{-1}W(y-\mu)$. A simpler algorithm is obtained if we ignore the term involving the derivatives of the weights $w'(\mu_i)$ in the matrix M.

3. Absolute-deviations quasi-likelihoods

It was evident from the start that (2·1), which is the natural equation for a quasi-likelihood based on the L_q-norm, does not lead to consistent estimates of the usual generalized linear model parameters. Besides correcting the estimating equation as we did in the last section, a second sense of the term 'absolute-deviations estimate' is obtained if we accept (2·1) for $q=1$ as is. This evidently means that we model the median rather than the mean of the response as a function of the explanatory variables x. The question is whether specifying the median as a smooth function of $x_i^{\mathrm{T}}\beta$ and indicating the scatter S as a function of the median is sufficient to compute estimates and their standard errors.

In discrete models, the median is clearly not a good functional to use. In the Bernoulli case, for example, the median is either 0, 1 or indeterminate. No smooth function of $x_i^{\mathrm{T}}\beta$ can model this behaviour. Absolute-deviations quasi-likelihoods are, therefore, available only for responses with continuous distributions. In those cases, however, modelling the median rather than the mean can be attractive, because the resulting fitted surface is more directly interpretable.

Denote by f_i the density of the distribution of the ith observation and suppose that the median m_i of this distribution is such that $f_i(m_i)>0$. Suppose further, that

$$g(m_i)=x_i^{\mathrm{T}}\beta,$$

for some link function $g(.)$. Let $m=(m_1,\ldots,m_n)^{\mathrm{T}}$. Substitution of $q=1$ in (2·1) gives the gradient of our quasi-likelihood

$$\frac{\partial}{\partial m}K(m;y)=S(m)^{-\frac{1}{2}}\operatorname{sgn}(y-m),$$

where $S(m)=\operatorname{diag}(S(m_1),\ldots,S(m_n))$ and $S_i=S(m_i)$ is a user-supplied function that models the scatter of the responses as a function of the median. Formally, the least-absolute-deviations estimate $\hat{\beta}$ is associated with the estimating equation

$$D^{\mathrm{T}}\{\operatorname{diag}(S_1,\ldots,S_n)\}^{-\frac{1}{2}}\{\operatorname{sgn}(y-m)\}=0, \qquad (3\cdot1)$$

where $D_{ij}=\partial\mu_i/\partial\beta_j$. For any random variable Y_i with density f_i we have

$$E\{\operatorname{sgn}(Y_i-m_i)\}=0,$$

so that (3·1) produces consistent estimates of the parameter β. Under additional regularity conditions we have

$$\operatorname{var}_\beta(\hat{\beta})=\{D^{\mathrm{T}}S(m)^{-\frac{1}{2}}FD\}^{-1}\{D^{\mathrm{T}}S(m)^{-1}D\}\{D^{\mathrm{T}}S(m)^{-\frac{1}{2}}FD\}^{-1},$$

where $F=\operatorname{diag}\{2f_1(m_1),\ldots,2f_n(m_n)\}$. The underlying distributions make their appearance in this formula unless we assume that $S_i^{\frac{1}{2}}=2f_i(m_i)$, in which case we have the formula

$$\operatorname{var}_\beta(\hat{\beta})=\{D^{\mathrm{T}}S(m)^{-1}D\}^{-1}. \qquad (3\cdot2)$$

To specify a quasi-likelihood model in the least-absolute-deviations framework, we must therefore assume that the scatter function $S(m)$ is properly matched to the choice of the median as our location functional. With the right choice, the formula for computing standard errors (3·2) is the same as in the case of models for the mean (1·4).

In a classical quasi-likelihood model, it is often assumed that the variance is only determined up to a proportionality constant. That constant must then also be estimated. One possible such estimator can be based on the sum of squares of normalized residuals. In the case of L_1-estimates, the same task is more involved, since we must estimate the value of a density. But for many models, this is entirely feasible and can be based on one of the methods of density estimation.

Example. McCullagh & Nelder (1989, p. 300) discuss an example involving clotting times of blood where the clotting was induced by two different lots of an agent, and was measured for nine different dilutions X. They fit a Gamma regression model with the inverse link and find the linear predictors:

$-0{\cdot}01655(0{\cdot}00086)+0{\cdot}01534(0{\cdot}00038)\log X$ for the first lot,
$-0{\cdot}02391(0{\cdot}00143)+0{\cdot}02360(0{\cdot}00062)\log X$ for the second lot.

The numbers in parentheses are standard errors. Taking the estimating equation (2·6) for the Gamma regression without correction

$$D^{\mathrm{T}}\operatorname{diag}(\hat{m}_i^2)^{-\frac{1}{2}}\{\operatorname{sgn}(y-\hat{m})\}=0,$$

and rewriting it as

$$D^{\mathrm{T}}\operatorname{diag}(\hat{m}_i^2)^{-\frac{1}{2}}\left\{\frac{\operatorname{sgn}(y-\hat{m})}{y-\hat{m}}\right\}(y-\hat{m})=0$$

to match (2·8), we can compute the estimates with the algorithm discussed in § 2·4. Ignoring the term involving the derivative of the weight, one obtains

$$M_{ii}=m_i^2/|y_i-m_i|$$

for the estimation weights (2·9). The adjusted responses to be used in the updating algorithm are

$$x_i^{\mathrm{T}}\beta-(y_i-m_i)/m_i^2 \quad (i=1,\ldots,n).$$

All of this is valid for the inverse link. The fitted linear predictors are:

$-0{\cdot}0168(0{\cdot}0016)+0{\cdot}0157(0{\cdot}0007)\log X$ for the first lot,
$-0{\cdot}0226(0{\cdot}0026)+0{\cdot}0231(0{\cdot}0011)\log X$ for the second lot.

This fit is easy to describe since it is characterized by the exact fitting of the observations $\{1, 9\}$ in the first lot and $\{10, 15\}$ in the second lot. Care has to be taken in running the algorithm, since the weights for these four observations tend to infinity as the iterations converge.

To yield a correct inference for the Gamma, we could proceed as follows. From (2·5) it is clear that the median m_i and the mean μ_i of the ith observation are linked by a proportionality relation $m_i=m_0\mu_i$. Evaluating the density at its median we find

$$f_i(m_i)=\Gamma(\nu)^{-1}\nu^{\nu}m_0^{\nu}\exp(-\nu m_0)(1/m_i)\propto 1/m_i.$$

We were, therefore, quite justified in choosing $S_i\propto m_i^2$. The proportionality constant between $2f_i(m_i)$ and $1/m_i$,

$$2\Gamma(\nu)^{-1}\nu^{\nu}m_0^{\nu}\exp(-\nu m_0),$$

is nothing else but twice the value of the density of $z_i = y_i/m_i$ evaluated at 1. This can be estimated by the value of a density estimate based on $y_1/\hat{m}_1, \ldots, y_n/\hat{m}_n$ evaluated at 1. A kernel density estimate gives for our example a value of about 6·0. The variance matrix for the estimates is therefore equal to

$$\tfrac{1}{144}\{X^{\mathrm{T}} \operatorname{diag}(m_i^2) X\}^{-1},$$

where $144 = (2 \times 6)^2$. This leads to the standard errors indicated above.

Acknowledgements

I would like to thank the two referees for their careful work and their suggestions which led to a better manuscript. The work on this paper was supported by the Swiss National Science Foundation.

References

FINNEY, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34**, 320–34.

KOENKER, R. W. & BASSETT, G. W. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

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. *J. Am. Statist. Assoc.* **84**, 460-6.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

PREGIBON, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics* **38**, 485-98.

STEFANSKI, L. A., CARROLL, R. J. & RUPPERT, D. (1986). Optimally bounded score functions for generalized linear models with applications to logistic regression. *Biometrika* **73**, 413-25.

WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439-47.

[*Received October* 1991. *Revised June* 1992]