Biometrika (2005), **92,** 3, *pp.* 724–731

Printed in Great Britain

Minimum distance estimation for the logistic regression model

By HOWARD D. BONDELL

Department of Statistics, Rutgers University, Piscataway, New Jersey 08854, U.S.A.

bondell@stat.rutgers.edu

Summary

It is well known that the maximum likelihood fit of the logistic regression parameters can be greatly affected by atypical observations. Several robust alternatives have been proposed. However, if we consider the model from the case-control viewpoint, it is clear that current techniques can exhibit poor behaviour in many common situations. A new robust class of estimation procedures is introduced. The estimators are constructed via a minimum distance approach after identifying the model with a semiparametric biased sampling model. The approach is developed under the case-control sampling scheme, yet is shown to be applicable under prospective sampling as well. A weighted Cramér–von Mises distance is used as an illustrative example of the methodology.

Some key words: Biased sampling problem; Bounded influence; Case-control data; Logistic regression; Minimum distance; Robust regression.

1. Introduction

Given a binary variable Y and a $p \times 1$ vector X of covariates, the logistic regression model states that

$$\mathrm{pr}(Y=1|X=x) = \frac{\exp(\alpha^* + x'\beta)}{1+\exp(\alpha^* + x'\beta)} \equiv Q_{\alpha^*,\beta}(x), \tag{1·1}$$

say, where α^* is a scalar parameter and β is a $p \times 1$ vector of parameters. Under this model, the marginal distribution of X is left completely unspecified.

It is well known that atypical observations can have a dramatic impact on the maximum likelihood fit. Several robust alternatives have been proposed in the literature, all of which may be expressed as methods that downweight observations which are, in some sense, far from the centre of the data. Downweighting can occur in terms of elliptical contours (Künsch et al., 1989; Carroll & Pederson, 1993) or in terms of extreme predicted probabilities (Copas, 1988; Carroll & Pederson, 1993; Bianco & Yohai, 1996). The majority of the current procedures were borrowed from the standard regression setting and do not take the unique structure of binary regression into account.

If we consider the model via the case-control viewpoint, it is clear that some techniques of downweighting may be problematic. Since the marginal distribution of X is a mixture of two distributions, the use of elliptical contours to measure distance and downweight observations is not appropriate. Estimators based on this methodology tend to be robust against highly extreme outliers, but can be greatly influenced by moderate outliers, and can also lose a great deal of efficiency for some common configurations of the covariates. In the extreme case of one group much smaller than the other, with sufficiently separated centres, the entire smaller group would be systematically considered as outliers and receive little or no weight. In fact, although estimators of this form are regularly implemented in standard software, they often behave poorly.

The purpose of this paper is to introduce a new general methodology for constructing robust and highly efficient estimators for the logistic regression model. This construction is developed based on case-control data, but remains applicable under prospective sampling. The logistic regression model under case-control sampling is equivalent to a two-sample semiparametric biased sampling problem. Exploiting this equivalence allows a new methodology to be defined in terms of a minimum distance criterion under the case-control formulation. The resulting family of procedures can yield estimators that are highly efficient if the model is true, and yet are robust to small deviations in the model. The proposed procedures do not depend on elliptical contours, and therefore perform well even when the marginal covariate distribution is far from elliptical. The minimum distance idea also yields a natural class of goodness-of-fit statistics for testing the validity of the logistic regression model.

2. A minimum distance approach

One approach to conducting a study would be an observational study, or prospective sampling, where the sampling is a simple random sample from the joint distribution, F_{XY}. This is the setting in which all current robust procedures are constructed. However, the work of Carroll et al. (1995) has shown that, for most of these procedures, point estimators and the corresponding asymptotic standard errors for the slope parameters, β, derived from this sampling scheme are still valid if the sampling is actually of the case-control form. This result had been previously derived for the maximum likelihood estimator by Prentice & Pyke (1979). Perhaps this is why the construction of new robust procedures has been limited to the prospective formulation.

Case-control sampling, also known as retrospective sampling, refers to independently sampling from F_0, the conditional distribution of $X|(Y=0)$, and F_1, defined similarly. As in Qin & Zhang (1997), if we let f_0 and f_1 denote the corresponding conditional densities and use (1·1), an application of Bayes' rule yields the equivalent two-sample semiparametric biased sampling model,

$u_1, \ldots, u_{n_0}$ is a random sample with density $f_0(x)$,
$z_1, \ldots, z_{n_1}$ is a random sample with density $f_1(x) = \exp(\gamma + x'\beta) f_0(x)$,

where $\gamma = \alpha^* + \log\{(1-\pi)/\pi\}$ and $\pi = \mathrm{pr}(Y=1)$.

Under the above model, for a fixed (γ, β), F_0 can be estimated in the following two ways:

$$\tilde{F}_0(t; \gamma, \beta, \rho) = \sum_{i=1}^{n} \tilde{p}_i(\gamma, \beta, \rho) I(x_i \leqslant t), \tag{2·1}$$

$$\hat{F}_0(t) = \sum_{i=1}^{n_0} \hat{p}_i I(u_i \leqslant t), \tag{2·2}$$

where $\hat{p}_i = n_0^{-1}$ and $\tilde{p}_i(\gamma, \beta, \rho) = n_0^{-1}\{1 + \rho \exp(\gamma + x_i'\beta)\}^{-1}$, with $\rho \equiv n_1/n_0$. The combined sample is denoted by $(x_1, \ldots, x_n) = (u_1, \ldots, u_{n_0}, z_1, \ldots, z_{n_1})$.

Qin & Zhang (1997) proposed a Kolmogorov–Smirnoff statistic as a measure of goodness-of-fit with the parameters estimated via maximum likelihood. This idea is further extended to the estimation procedure by minimising a discrepancy to construct the estimator. In particular, a class of weighted Cramér–von Mises distance measures will be used. However, it is stressed that other choices are possible and deserve further investigation.

3. The weighted Cramér–von Mises procedure

3·1. *Defining the estimator*

Denote the difference process between the two distribution functions in (2·1) and (2·2) by

$$\begin{aligned} d(t; \alpha, \beta) \equiv d(t; \gamma, \beta, \rho) &\equiv \tilde{F}_0(t; \gamma, \beta, \rho) - \hat{F}_0(t) \\ &= n_0^{-1} \sum_{i=1}^{n} \{y_i - Q_{\alpha,\beta}(x_i)\} I(x_i \leqslant t), \end{aligned} \tag{3·1}$$

where $\alpha \equiv \gamma + \log \rho$, and Q has the functional form as in (1·1). In the context of goodness-of-fit, Su & Wei (1991) arrived at this same process as a cumulative residual process from the prospective viewpoint.

For a given weight function, W_{F_X}, which as written may depend on the distribution of X, define the weighted Cramér–von Mises distance as

$$D(\alpha, \beta) \equiv \int \{d(t; \alpha, \beta)\}^2 dW_{F_X}(t) \tag{3·2}$$

with $dW_{F_X}(t) \geqslant 0$ for all t and $\int dW_{F_X}(t) < \infty$. The choice of weight function used in this paper will be the following family of weight functions, indexed by the constant $c \geqslant 0$:

$$dW_{F_X}(t) \equiv [F_X(t)\{1 - F_X(t)\}]^c dF_X(t). \tag{3·3}$$

Note that $c \geqslant 0$ implies that $\int dW_{F_X}(t) < \infty$ regardless of F_X, which in practice will be estimated from the data via the empirical marginal distribution. More judicious choices of the weight function may be possible, and deserve further investigation.

The parameter estimators will then be defined as the solution to the following minimisation problem:

$$(\tilde{\alpha}, \tilde{\beta}) \equiv \arg\min_{(\alpha,\beta)} D(\alpha, \beta), \tag{3·4}$$

subject to $d(\infty; \tilde{\alpha}, \tilde{\beta}) = 0$. The side condition ensures that $\tilde{F}_0$ is a cumulative distribution function.

The functional form of the optimisation problem for the weighted Cramér–von Mises distance can be written as follows. Define $\alpha \equiv \alpha(\beta, F_{XY})$ as a functional of β and the distribution F_{XY}, through the relationship

$$\int \{y - Q_{\alpha,\beta}(x)\} dF_{XY}(x, y) = 0. \tag{3·5}$$

Now define β_0 as the solution to the minimisation problem

$$\beta_0 \equiv \arg\min_{\beta} \int \left[\int \{y - Q_{\alpha,\beta}(x)\} I(x \leqslant t) dF_{XY}(x, y) \right]^2 dW_{F_X}(t). \tag{3·6}$$

Then $\alpha_0 \equiv \alpha(\beta_0, F_{XY})$.

The designation of case or control group is arbitrary, so that any estimator in logistic regression should be equivariant under changing of the labels. Clearly this is the case for the minimum weighted Cramér–von Mises estimator since the objective function depends on y only through $y - Q_{\alpha,\beta}(x)$. Proposition 1 shows equivariance under location-scale changes of the covariate vector.

PROPOSITION 1. *Let A be a $p \times p$ diagonal matrix with strictly positive entries and let b be a $p \times 1$ vector. Under the transformation given by $T : x \to (Ax + b)$, $(\tilde{\alpha}, \tilde{\beta})$ is transformed to $(\tilde{\alpha} - b'A^{-1}\tilde{\beta}, A^{-1}\tilde{\beta})$.*

Remark 1. It appears not possible to achieve full affine equivariance directly in the multivariate setting regardless of the choice of weight function. This is because of the inherent partial ordering imposed by the cumulative distribution function.

Now further define

$$w_1 \equiv w_1(x, \alpha, \beta, F_X) = \int C(x, z) Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(z)\}(z - K) dF_X(z), \tag{3·7}$$

with

$$C(x, z) = \int I(t \geqslant x) I(t \geqslant z) dW_{F_X}(t), \quad K \equiv K(\alpha, \beta, F_X) = \int x dH_{\alpha,\beta,F_X}(x),$$

$$dH_{\alpha,\beta,F_X}(x) \equiv \left[\int Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(x)\} dF_X(x) \right]^{-1} Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(x)\} dF_X(x).$$

Proposition 2. *Let* (α_0, β_0) *be a solution to the functional version of the minimisation problem given by* (3·5) *and* (3·6), *and suppose that the following conditions hold*:

(i) $\int dW_{F_X}(t) < \infty$;

(ii) *the vector* $K(\alpha, \beta, F_X)$ *as defined above has all components finite at* (α_0, β_0);

(iii) *there exists a solution such that* $\|\beta_0\| < \infty$.

Then (α_0, β_0) *is a generalised M-functional satisfying the following first-order condition*:

$$\int w(x, \alpha_0, \beta_0, F_X)\{y - Q_{\alpha_0,\beta_0}(x)\}dF_{XY}(x, y) = 0,$$

where $w(x, \alpha_0, \beta_0, F_X) = (1, w_1')'$, *and* $w_1 = w_1(x, \alpha_0, \beta_0, F_X)$ *is as defined in* (3·7).

3·2. *Asymptotics*

The asymptotic distribution will now be derived within the framework of prospective sampling as opposed to case-control, but, as a consequence of Theorem 1, inference for the slope parameters will remain valid under case-control sampling. A standard tool in asymptotic theory, and particularly that of robustness, is the influence function (Hampel et al., 1986). The influence function of a functional of a distribution heuristically measures the change in the functional upon the addition or deletion of a single point at a particular location. For robustness, a bounded influence function is desirable. The influence function is the linear term in the asymptotic expansion of a statistic so that, under standard regularity conditions (Hampel et al., 1986, Ch. 4), the asymptotic variance is given by $E(\text{IF IF}')$, where IF denotes the vector-valued influence function.

From Proposition 2, under the prospective model, standard theory yields

$$\text{IF}\{(x, y), (\alpha, \beta), F_X\} = A^{-1}w(x, \alpha, \beta, F_X)\{y - Q_{\alpha,\beta}(x)\}, \quad (3·8)$$

where the $(p+1) \times (p+1)$ matrix A is given by

$$A \equiv A(\alpha, \beta, F_X) = E_X[w(x, \alpha, \beta, F_X)(1, x')Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(x)\}].$$

Conditions (i) and (ii) of Proposition 2 ensure that the influence function is bounded for all (x, y).

From the influence function, the asymptotic distribution of the minimum weighted Cramér–von Mises estimator, $(\tilde{\alpha}, \tilde{\beta})$, under the prospective logistic regression model (1·1) with true parameters (α_0, β_0), is given by

$$n^{1/2}\{\tilde{\alpha} - \alpha_0, (\tilde{\beta} - \beta_0)'\}' \to N(0, V),$$

in distribution, with $V \equiv V(\alpha_0, \beta_0, F_X) = A^{-1}B(A^{-1})'$, in which the $(p+1) \times (p+1)$ matrix B is

$$B \equiv B(\alpha, \beta, F_X) = E_X[w(x, \alpha, \beta, F_X)w(x, \alpha, \beta, F_X)'Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(x)\}].$$

To conduct inference, $V(\alpha_0, \beta_0, F_X)$ will be consistently estimated by $V(\tilde{\alpha}, \tilde{\beta}, F_n)$, where F_n denotes the empirical marginal distribution of X. Based on the results of Carroll et al. (1995), asymptotically, it remains valid to use $\tilde{\beta}$ and $V(\tilde{\alpha}, \tilde{\beta}, F_n)$ for inference about the vector of slope parameters, β, under case-control sampling.

Theorem 1. *Under the logistic regression model,* $\tilde{\beta}$ *will consistently estimate* β, *and* $V(\tilde{\alpha}, \tilde{\beta}, F_n)$ *will consistently estimate the variance matrix of* $\tilde{\beta}$, *regardless of whether the sampling was done prospectively or retrospectively.*

4. Examples

4·1. *Some existing robust estimators*

The minimum weighted Cramér–von Mises proposal of this paper is compared to the maximum likelihood estimator as well as three benchmark robust proposals. The first two are the conditionally unbiased bounded influence estimator of Künsch et al. (1989) and the Mallows-type estimator as discussed by Carroll & Pederson (1993) with weights defined as functions of a robust Mahalanobis distance. These proposals were computed as implemented in the standard Robust package of S-Plus. The third proposal is the estimator of Bianco & Yohai (1996) with choice of objective function and procedure for implementation as in Croux & Haesbroeck (2003).

All of these estimators, except for the maximum likelihood estimator, are indexed by a tuning constant, variation of which yields a trade-off between robustness and efficiency. In the univariate case, all three existing robust proposals, along with the proposal of this paper, have strictly bounded influence functions, but this is not true in the multivariate case. The Bianco–Yohai estimator has unbounded influence for contamination points with $\|x\| \to \infty$, but $\alpha + x'\beta$ remaining bounded, i.e. points which remain close to the 'separating hyperplane', or all points if $\beta = 0$.

4·2. *Simulation results*

Two normal distributions with common scale or two gamma distributions with common shape are possible ways of generating the univariate logistic regression model. Using these two different models for the covariate, we generated 2000 samples of size 100 with true $\beta = 2{\cdot}5$. The effect of contamination was examined by adding three cases ($y = 1$) at various values of x. The value $\beta = 2{\cdot}5$ is chosen so that the mixture distribution is just slightly bimodal under normality with 50 controls and 50 cases. The tuning constants for the three existing robust estimators were matched to yield similar robustness and efficiency properties in this set-up and then fixed for the remainder of the simulation study. Two choices, $c = 0, 1$, are used for the minimum weighted Cramér–von Mises estimator.

Table 1 gives the bias and root mean squared error for each of the estimators under normal distributions using 75 controls and 25 cases. As expected, because of the asymmetry of the group proportions, there is a loss of efficiency for the Künsch et al. and Mallows estimators, while they also exhibit worse behaviour under contamination, particularly under moderate contamination.

Table 1: *Simulation study. Root mean squared error, and bias in parentheses, of slope parameter estimator for three contaminating cases at various locations x, under normal distributions for 75 controls and 25 cases*

	Uncontaminated	$x = -1$	$x = -2$	$x = -3$	$x = -5$
MLE	0·756 (0·262)	1·048 (1·036)	1·277 (1·271)	1·453 (1·449)	1·716 (1·715)
CUBIF $c = 0{\cdot}5$	0·816 (0·283)	1·116 (1.103)	0·999 (0·975)	0·988 (0·958)	0·972 (0·938)
Mallows $c = 5$	0·789 (0·258)	1.061 (1·048)	1·150 (1·140)	1·107 (1·088)	0·741 (0·475)
BY $c = 3$	0·783 (0·266)	0·886 (0·837)	0·897 (0·823)	0·877 (0·751)	0·829 (0·544)
MCVM $c = 0$	0·784 (0·258)	0·987 (0·966)	1·000 (0·978)	0·998 (0·976)	0·996 (0·974)
MCVM $c = 1$	0·792 (0·262)	1·026 (1·008)	1·036 (1·017)	1·035 (1·016)	1·035 (1·016)

MLE, maximum likelihood estimator; CUBIF, estimator of Künsch et al.; Mallows, estimator of Carroll & Pederson; BY, Bianco & Yohai estimator; MCVM, minimum weighted Cramér–von Mises estimator.

Table 2 gives the results under exponential distributions, with 50 controls and 50 cases. As anticipated, the efficiency loss for the Künsch and Mallows estimators is severe, as evidenced by the large root mean squared errors. The Mallows estimator performed particularly poorly in terms of root mean squared error in this set-up, and was extremely unstable. Surprisingly, the Bianco–Yohai estimator was extremely nonrobust for this configuration. The efficiency loss for the Mallows and Künsch et al. estimators becomes even greater when the proportions are made to be unequal. Overall, in the simulations, the proposed minimum weighted Cramér–von Mises estimator exhibited the most stable behaviour.

Table 2: *Simulation study. Root mean squared error, and bias in parentheses, of slope parameter estimator for three contaminating cases at various locations x, under exponential distributions for 50 controls and 50 cases*

	Uncontaminated	$x=0$	$x=-0{\cdot}5$	$x=-1$	$x=-3$
MLE	0·731 (0·133)	0·650 (0·267)	0·845 (0·729)	1·139 (1·093)	1·899 (1·893)
CUBIF $c=0{\cdot}5$	0·955 (0·178)	0·875 (0·181)	0·848 (0·296)	0·846 (0·295)	0·843 (0·293)
Mallows $c=5$	0·980 (0·210)	2·140 (0·583)	1·969 (0·301)	1·959 (0·298)	1·959 (0·254)
BY $c=3$	0·738 (0·130)	0·651 (0·276)	0·852 (0·735)	1·141 (1·088)	1·877 (1·830)
MCVM $c=0$	0·783 (0·118)	0·719 (0·273)	0·755 (0·480)	0·783 (0·520)	0·800 (0·536)
MCVM $c=1$	0·784 (0·116)	1·721 (0·255)	0·746 (0·438)	0·770 (0·478)	0·792 (0·500)

MLE, maximum likelihood estimator; CUBIF, estimator of Künsch et al.; Mallows, estimator of Carroll & Pederson; BY, Bianco & Yohai estimator; MCVM, minimum weighted Cramér–von Mises estimator.

4·3. *Leukaemia data*

The leukaemia data from Cook & Weisberg (1982, p. 193) were analysed in Carroll & Pederson (1993), among others. There are 33 observations, with the response being survival for 52 weeks and two predictors, white blood cell count and the indicator of acute granuloma. It is well documented that a possible response outlier with $Y=1$ lies in a cluster of five points with identical values of the white blood cell count. Including only the continuous covariate in the model yields a situation with which it is more difficult to deal with as the outlier is now contained in a large group with equal leverage. Table 3 gives the estimated coefficients and asymptotic standard errors for the various procedures. Also shown for comparison is MLE_{32}, the maximum likelihood estimate after removing the possible response outlier. The Mallows estimator is highly unstable in this situation because of the extreme skewness and bimodality of the distribution of white blood cell counts. An alternative choice of tuning constant, $c=8$, the default in the current Robust package of S-Plus, is also shown. The Bianco–Yohai estimator with the previously used tuning constant behaves similarly to the maximum likelihood estimator, but with a large standard error and hence a very small t-statistic. An alternative choice of tuning constant for the Bianco–Yohai estimator is shown as well. The estimator is now more robust, but its efficiency is much lower, as evidenced by its large standard error. The minimum weighted Cramér–von Mises estimator and that of Künsch et al. perform best here, dampening the effect of the single point as seen by the fact that they move towards the estimate MLE_{32}.

Table 3: *Leukaemia data. Estimates of parameters, with estimated standard errors in parentheses*

	Intercept	White blood count
MLE	−0·028 (0·498)	−2·933 (1·844)
CUBIF $c=0{\cdot}5$	1·075 (0·902)	−9·570 (5·472)
Mallows $c=5$	3·264 (4·122)	−75·260 (84·380)
Mallows $c=8$	2·053 (1·967)	−43·593 (40·709)
BY $c=3$	−0·027 (0·818)	−2·945 (5·789)
BY $c=0{\cdot}5$	0·890 (0·898)	−11·970 (8·542)
MCVM $c=0$	0·781 (0·683)	−9·578 (5·283)
MLE_{32}	1·377 (0·820)	−19·387 (9·348)

MLE, maximum likelihood estimator; CUBIF, estimator of Künsch et al.; Mallows, estimator of Carroll & Pederson; BY, Bianco & Yohai estimator; MCVM, minimum weighted Cramér–von Mises estimator; MLE_{32}, maximum likelihood estimator based on 32 observations.

Acknowledgement

The author would like to thank the editor and two referees for their help in improving the manuscript. The author is also grateful to David E. Tyler for his comments and guidance during the preparation of this manuscript. This research was supported by a grant from the U.S. National Science Foundation.

Appendix

Proofs

Proof of Proposition 1. First note that $Q_{\alpha-b'A^{-1}\beta,A^{-1}\beta}(Ax+b)=Q_{\alpha,\beta}(x)$ for all x. Hence (3·5) remains unchanged and, since A is diagonal with strictly positive entries, (3·6) becomes

$$\beta_0 \equiv \arg\min_{\beta} \int [\int \{y - Q_{\alpha,\beta}(x)\} I\{x \leqslant A^{-1}(t-b)\} dF_{XY}(x,y)]^2 dW_{F_{AX+b}}(t).$$

However, $dW_{F_{AX+b}}(t) \propto [F_X\{A^{-1}(t-b)\}]^c [1 - F_X\{A^{-1}(t-b)\}]^c dF_X\{A^{-1}(t-b)\}$. The optimisation problem then remains unchanged, up to proportionality, after a change of variable. □

Proof of Proposition 2. Differentiating (3·5) with respect to β yields the equation

$$\int Q_{\alpha,\beta}(x)\{1 - Q_{\alpha,\beta}(x)\} \left\{ x + \frac{\partial}{\partial \beta} \alpha(\beta, F_{XY}) \right\} dF_X(x) = 0.$$

Hence

$$\frac{\partial}{\partial \beta} \alpha(\beta, F_{XY}) = -K(\alpha, \beta, F_X).$$

After a little algebra, the objective function (3·6) may be written as

$$\iint C(x,z)\{y - Q_{\alpha,\beta}(x)\}\{u - Q_{\alpha,\beta}(x)\} dF_{XY}(x,y) dF_{XY}(z,u).$$

The integrability of the weight function allows for the interchange of the order of integration in the above derivation. Also, by condition (iii), (α_0, β_0) must be a critical point of (3·6). The result is then obtained after differentiation with respect to β. □

Proof of Theorem 1. It is straightforward to see that the functional defined by Proposition 2 is not only prospectively unbiased, but also prospectively conditionally unbiased, that is

$$\int w(x, \alpha_0, \beta_0, F_X)\{y - Q_{\alpha_0,\beta_0}(x)\} dF_{Y|X=x}(y) = 0,$$

since $w(x, \alpha_0, \beta_0, F_X)$ is independent of y, and $E_{Y|X=x}(Y) = Q_{\alpha_0,\beta_0}(x)$. Hence the theorem follows directly from the lemma of Carroll et al. (1995, § 5). □

References

Bianco, A. M. & Yohai, V. J. (1996). Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods, Lecture Notes in Statistics* **109**, Ed. H. Rieder, pp. 17–34. New York: Springer-Verlag.

Carroll, R. J. & Pederson, S. (1993). On robust estimation in the logistic regression model. *J. R. Statist. Soc.* B **55**, 693–706.

Carroll, R. J., Wang, S. & Wang, C. Y. (1995). Prospective analysis of logistic case-control studies. *J. Am. Statist. Assoc.* **90**, 157–69.

Cook, R. D. & Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Chapman and Hall.

Copas, J. B. (1988). Binary regression models for contaminated data (with Discussion). *J. R. Statist. Soc.* B **50**, 225–65.

Croux, C. & Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.* **44**, 273–95.

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

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.

Prentice, R. L. & Pyke, R. (1979). Logistic disease incidence models and case-control studies *Biometrika* **66**, 403–11.

Qin, J. & Zhang, B. (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika* **84**, 609–18.

Su, J. Q. & Wei, L. J. (1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Am. Statist. Assoc.* **86**, 420–6.

[*Received October* 2004. *Revised March* 2005]