Further Remarks on Robustness in the Logistic Regression Model

by

Raymond J. Carroll* and Shane Pederson
Texas A & M University
Los Alamos National Lab

Technical Report # 90–43C

Department of Statistics
Purdue University

August, 1990

*The research of Professor Raymond J. Carroll is supported in part by the National Science Foundation Grants DMS–8923071 and DMS–8717799 at Purdue University.

△

FURTHER REMARKS ON ROBUSTNESS IN THE LOGISTIC REGRESSION MODEL

R. J. Carroll
Department of Statistics
Texas A&M University
College Station, TX 77843

Shane Pederson
Statistics Group
Los Alamos National Laboratory
Los Alamos, NM 87545

SUMMARY

We investigate robustness in the logistic regression model. Copas (1987) studied two forms of robust estimators, a robust/resistant estimate of Pregibon (1982) and an estimate based on a misclassification model. He concluded that robust/resistant estimates are much more biased in small samples than the usual logistic estimate, and recommends a bias–corrected version of the misclassification estimate. We show that there are other versions of robust/resistant estimates which have bias approximately the same as and sometimes even less than the logistic estimate; these estimates belong to the Mallows class. In addition, the corrected misclassification estimate is inconsistent at the logistic model; we develop two ways to modify it to obtain the desired consistency. The first method, based on correcting the score function, is a member of the Mallows class. The second method adjusts the misclassification estimate directly; an asymptotic theory for this new estimate is developed. The results are illustrated on data sets featuring different kinds of outliers.

Keywords: ASYMPTOTICS; BINARY REGRESSION; CONSISTENT ESTIMATION; LEVERAGE; MALLOWS ESTIMATORS; MISCLASSIFICATION; SMALL SAMPLE BIAS EXPANSIONS; OUTLIERS

1. Introduction

We consider robust estimation in the logistic regression model

$$\Pr(Y = 1|x) = F(x^T \beta_T), \qquad F(v) = \{1 + \exp(-v)\}^{-1}. \tag{1.1}$$

Along with the usual logistic model, we will be concerned with a misclassification model in which each response is misclassified with probability γ, so that

$$\Pr(Y = 1|x) = F(x^T \beta_T) + \gamma \left\{1 - 2F(x^T \beta_T)\right\} = G(x_i^T \beta_T, \gamma). \tag{1.2}$$

In an important paper, Copas (1988) contrasts two forms of robust estimates:

- Robust/resistant estimates due to Pregibon (1982);
- The misclassification maximum likelihood estimate (mle) for model (1.2), corrected to be approximately consistent for the logistic model (1.1).

Copas concludes that the latter is a preferable method of estimation at the logistic model, as it seems to have the robustness properties of the former method while at the same time having less small to moderate sample size bias. Model (1.2) also has uses when response misclassification is of interest (Copas, 1988); these will not be discussed in detail here. Copas focused on small values of γ and the use of (1.2) in generating robust estimators and diagnostics.

Despite this work, a number of important problems remain to be investigated. Among these are the following:

- Pregibon's estimate is inconsistent at the logistic model. Thus, it might be the case that the bias observed in Copas' study is really not so much a function of the entire class of robust/resistant estimates as it is a function of the particular method employed. For example, the "Mallows" or "Schweppe" classes (Künsch, et al., 1989) are consistent. Our results will indicate that robust/resistant estimates can have much less small sample bias than previously thought.
- The corrected misclassification mle is only approximately consistent, with no large sample theory available. We will provide the required theory for the corrected estimate.
- In addition, we will propose and study two estimates which are closely related to the misclassification mle but which are consistent at the logistic model. The first method modifies the score

function of the misclassification mle to make it unbiased at the logistic model; this estimate is a member of the Mallows class. The second method starts from the misclassification mle, and adjusts it to be consistent at the logistic model.

- We will investigate modifications of the usual logistic estimate which partially and/or fully correct for misclassification, along with studying the large sample properties of these estimates.

The outline of this paper is as follows. In section 2, we will give an overview of the robust estimates previously proposed in the literature, making a connection between robust/resistant estimates and the misclassification estimate. In section 3, we discuss corrected estimates and their asymptotic theory. In section 4, we introduce our misclassification–based estimate consistent at the logistic model, and our logistic–based estimate consistent at the misclassification model. In section 5, we illustrate the behavior of these estimates on three sets of data. In section 6, we investigate the small sample bias properties of the robust/resistant estimates, and compute these biases numerically in the two examples.

2. Robust Estimates

2.1. Introduction

The primary robust estimates for the logistic model (1.1) solve equations of the form

$$0 = \sum_{i=1}^{n} w_i x_i \left\{ Y_i - F(x_i^T \beta) - c(x_i, \beta) \right\}, \tag{2.1}$$

where the $\{w_i\}$ are weights, which may depend on the response. If $w_i \equiv 1$ and $c(x_i, \beta) \equiv 0$, then (2.1) yields the usual logistic regression estimate. If $w_i = w(x_i, x_i^T\beta)$ and $c(x_i, x_i^T\beta) \equiv 0$, then the weights depend only on the design and we are in the so–caled "Mallows" class. Finally, if $w_i = w(x_i, Y_i)$, then we are in the "Schweppe" class.

2.2. The Mallows Class

In the Mallows–type formulation, the weights do not depend on the response directly, but are instead functions only of the design $\{x_i\}$ and the parameter β. Basically, the idea is that points which have high leverage are "dangerous", and should be downweighted. Estimators $\widehat{\beta}_0$ in this class are necessarily consistent, since in this case (2.1) is an unbiased estimating equation. While less efficient than the usual logistic mle, the Mallows class has an easily computed small–sample $O(1/n)$ bias which can be smaller than that of the mle, especially if there are unusual design points. If superscript (j) denotes the jth derivative, and if we define

$$V_{nj}(\beta) = -n^{-1} \sum_{i=1}^{n} w_i^j x_i x_i^T F^{(1)}(x_i^T \beta);$$

$$T_{nj}(\beta) = V_{n1}^{-1}(\beta) n^{-1} \sum_{i=1}^{n} w_i^{(2-j)} x_i x_i^T V_{n1}^{-1}(\beta) V_{n2}(\beta) V_{n1}^{-1}(\beta) x_i F^{(j)}(x_i^T \beta),$$

then by an analysis similar to that of Copas (1988), at the logistic model the first order bias is

$$E(\widehat{\beta}_0 - \beta_T) \approx -(2n)^{-1} \left\{ T_{n2}(\beta_T) + 2T_{n1}(\beta_T) \right\}.$$

There are two important subcases: (i) the weights depend only on the design; and (ii) the weights depend only on the probabilities.

CASE I: Leverage Downweighting Suppose that $w(x_i, x_i^T \beta) = w(x_i)$, so that we downweight strictly on the basis of leverage. In this case, $w^{(1)} = T_{n1} = 0$. For regression through the origin (Copas, 1988, p. 230), the bias is

$$E(\widehat{\beta}_0 - \beta_T) \approx -(2n)^{-1} \frac{n^{-1} \sum_{i=1}^{n} w_i x_i^3 F^{(2)}(x_i \beta_T) n^{-1} \sum_{i=1}^{n} w_i^2 x_i^2 F^{(1)}(x_i \beta_T)}{\left\{ n^{-1} \sum_{i=1}^{n} w_i x_i^2 F^{(1)}(x_i \beta_T) \right\}^3}.$$

For example, if $n = 50$ and the x's consist of 10 each at ± 2, ± 1, 0, then the bias for the mle is approximately 0.019 at $\beta_T = 0.5$ and approximately 0.048 at $\beta_T = 1$. If instead we give weight of only 0.5 to those observations with $x = \pm 2$, then the corresponding biases for the Mallows estimate are 0.027 and 0.062, respectively. This is not by any means a design with any unusual leverage. In practice, the Mallows weight assigned to the design points ± 2 will be very close to 1.0, and there will be little difference in the bias behavior of the two methods.

While this might at first suggest that the Mallows estimators are generally more biased, this is not the case in general. Of particular concern is the case of bias when there are unusual design points, where the classic robustness theory (Hampel, et al., 1986) suggests that robust methods should have smaller biases. We illustrate this numerically. Replace the previous design with the same one except that 2 points are placed at ± 5, and give weight 0.5 to these unusual points, all other weights being 1.0. The mle has approximate biases 0.036 and 0.060 at $\beta_T = 0.5$ and 0.1, while the Mallows estimate has biases 0.029 and 0.056. A Mallows estimate which gives zero weight to the four most extreme design points has biases 0.021 and 0.053.

The main point of this illustration is that in a design with extreme design points, selective downweighting can lead to less biased estimates when compared to the usual mle. The decrease in bias, coupled with the robustness protection, can be worth the loss of efficiency. In the previous example, a weight of 0.25 assigned to the unusal design points leads to a MSE efficiency of over 80% at the logistic model.

In his North Carolina Ph.D. thesis, L. Stefanski suggests downweighting on the basis of a robust Mahalanobis distance. The basic idea is that those points which are not near the center of the design space should be downweighted. Any of these robust methods depend on how well one can measure the leverage of an individual point; this can be a nontrivial task. In the leukemia data listed by Cook & Weisberg (1982), the most influential point occurs in a situation where there are three identical design points. As a group, these points are highly leveraged, but automatic methods for measuring leverage, such as that given above, may have problems identifying that each of the points individually has leverage.

The appendix defines the Mallows leverage downweighting method used in our calculations.

CASE II: Prediction Downweighting An alternative is to set $w(x_i, x_i^T\beta) = w(x_i^T\beta)$, so that extreme fitted probabilities are downweighted. An example of this kind is

$$w(x^T\beta) = \left[F(x^T\beta)\left\{1 - F(x^T\beta)\right\}\right]^c \left[F(x^T\beta)^\lambda + \left\{1 - F(x^T\beta)\right\}^\lambda\right],$$

for constants (c, λ). The choices (c, λ) = (0,0) or (1,-1) yields the usual logistic estimate. The choice (c, λ) = (1,0) weights according to the variance of Y given x, and will handle outliers at very low or very high predicted probabilities.

2.3. Schweppe Class

Pregibon (1982), Stefanski, et al. (1985) and Künsch, et al. (1989) have defined estimates in which the weights depend both on the leverage and on the response. Here the correction terms $c(x_i, \beta)$ in (2.1) are necessary in order that (2.1) be an unbiased estimating equation given x.

Pregibon's method chooses $c(x,\beta) \equiv 0$ and is thus inconsistent. The versions of the method of Künsch, et al. used here are discussed in the Appendix.

2.4. Misclassification Estimate

For the misclassification model (1.2), the mle $\hat{\beta}_{Mc}$ is an M–estimator solving

$$0 = \sum_{i=1}^{n} w_i(x_i^T\beta,\gamma)x_i \left\{Y_i - G(x_i^T\beta,\gamma)\right\}, \qquad \text{where} \tag{2.2}$$

$$w_i(x_i^T\beta,\gamma) = (1-2\gamma)F(x_i^T\beta)\left\{1 - F(x_i^T\beta)\right\}\left[G(x_i^T\beta,\gamma)\left\{1 - G(x_i^T\beta,\gamma)\right\}\right]^{-1}.$$

These equations lie in the class defined by (2.1). Generally, (2.2) has more than one solution if there is an influential outlier. Computation of the solution which maximizes the likelihood can be tricky; use of GLIM often fails.

Recognizing that $\hat{\beta}_{Mc}$ is inconsistent for β_T at the logistic model Copas (1988) suggested a bias–corrected version appropriate for small γ, see section 3.3. This estimate remains inconsistent at the logistic model.

2.5. Misclassification and Robust/Resistant Estimates

There are two easily identified methods for modifying the misclassification estimate $\hat{\beta}_{Mc}$ to be consistent at the logistic model:

- Directly adjust it for its bias, see section 4.1;
- Adjust the score function (2.2) to make it unbiased at the logistic model.

We pursue the second approach here.

The typical way to adjust an estimating equation to make it unbiased at a model is to subtract the former's expectation at that model. In (2.2), this simply reduces to solving

$$0 = \sum_{i=1}^{n} w_i(x_i^T\beta,\gamma)x_i \left\{Y_i - F(x_i^T\beta)\right\}. \tag{2.3}$$

Note that the solution to (2.3) is a member of the prediction downweighting Mallows class. Thus, Copas' methods can be looked upon as robust resistant estimates, and his criticism is not of robust estimates in general but of the Schweppe class in particular.

The form (2.3) is also instructive as to the behavior of the resulting estimate. We see that the weights become small when $|x_i^T\beta|$ is large, with γ controlling how extreme the fitted probabilities have to be before significant downweighting occurs. For those γ *not* near 0, the weights become nearly proportional to $F(x^T\beta)\left\{1 - F(x^T\beta)\right\}$, an option discussed in section 2.2.

2.6. Influence Functions

It can be shown that, in general, the Mallows leverage downweighted and Schweppe estimates have bounded influence functions. The misclassification mle has a bounded influence function only when $\|x\|\ F(x^T\beta_T)\ \{1-F(x^T\beta_T)\}$ is bounded, which occurs, in general, only for a scalar predictor x; otherwise a ridge can be formed. The Mallows prediction downweighted estimates have a similar behavior: their influence function is bounded when $\|x\|w(x^T\beta)$ is bounded.

Section 3: Estimates Corrected for Misclassification

In this section, we suggest three different estimates motivated by consideration of small misclassification rates. In section 3.1, we will define an estimate which corrects for the bias of the ordinary weighted logistic regression estimate due to misclassification. In section 3.2, we suggest a refinement of this estimate. In section 3.3, we study the behavior of the maximum likelihood estimate under a small error rate misclassification model, computing its bias at the ordinary logistic model. This leads to a third corrected estimate, which upon further modification becomes (27) of Copas (1988). In section 3.4, we compute the asymptotic limit distributions of all these estimates in a unified way, using techniques due to Stefanski (1985).

3.1: First Order Correction

Consider the Mallows class of solutions to (2.1), so that the weights $\{w_i\}$ do not depend on the response. In the case that we have misclassification so that the model is given by (1.2), $\widehat{\beta}_0$ estimates not β_T, but rather β_{0n}, the solution to

$$0 = \sum_{i=1}^{n} w_i x_i E\left\{Y_i - F(x_i\beta_{0n})|x_i\right\} = \sum_{i=1}^{n} w_i x_i \left\{G(x_i^T\beta_T,\gamma) - F(x_i\beta_{0n})\right\}. \tag{3.1}$$

Via a Taylor series expansion about β_T, one can show that to terms of order $O(\gamma^2)$, $\beta_T = \beta_{0n} + \mathcal{H}_{n,1}^{-1}(\beta_T)\mathcal{C}_{n,1}(\beta_T)$, where

$$\mathcal{C}_{n,1}(\beta) = \gamma\sum_{i=1}^{n} w_i x_i\left\{2F(x_i^T\beta) - 1\right\}; \qquad \mathcal{H}_{n,1}(\beta) = \sum_{i=1}^{n} w_i x_i x_i^T F(x_i^T\beta)\left\{1 - F(x_i^T\beta)\right\}. \tag{3.2}$$

This leads to the estimate

$$\widehat{\beta}_{c,1} = \widehat{\beta}_0 + \mathcal{H}_{n,1}^{-1}(\widehat{\beta}_0)\mathcal{C}_{n,1}(\widehat{\beta}_0) \tag{3.3}$$

3.2. A Modified Bias Correction

One potential problem with (3.3) is that the substitution leading to (3.3) replaces β_T by an inconsistent estimate, $\widehat{\beta}_0$. Possible improvement can be made by noting that, by (3.1),

$$0 = \sum_{i=1}^{n} w_i x_i \left\{ \frac{F(x_i^T \beta_{0n}) - \gamma}{1 - 2\gamma} - F(x_i \beta_T) \Big| x_i \right\}.$$

This suggests modifying the definition of $\mathcal{C}_{n,1}$ in (3.2), leading to the estimate

$$\widehat{\beta}_{c,2} = \widehat{\beta}_0 + \mathcal{H}_{n,1}^{-1}(\widehat{\beta}_0)\mathcal{C}_{n,2}(\widehat{\beta}_0), \qquad \text{where} \tag{3.4}$$

$$\mathcal{C}_{n,2}(\beta) = \frac{\gamma}{1 - 2\gamma} \sum_{i=1}^{n} w_i x_i \left\{ 2F(x_i^T \beta) - 1 \right\}.$$

3.3. Correcting the Misclassification MLE

The estimates defined by (3.3) and (3.4) are modifications of the usual logistic regression estimate to take into account possible misclassification. An alternative approach is to compute the misclassification mle for a fixed γ, and then correct it to be approximately consistent at the logistic model. This is the approach followed by Copas (1988), who does not compute the limit distribution of his estimate. Such an estimate should inherit the robustness properties of the misclassification mle, and, it would be hoped, advantageous bias behavior.

Under the logistic model (1.1), the solution $\widehat{\beta}_{Mc}$ to (2.2) estimates β_{Mn}, where

$$0 = \sum_{i=1}^{n} \mathcal{K}_i(\beta_T, \gamma, \beta_{Mn}) = \sum_{i=1}^{n} w_i(x_i^T \beta_{Mn}, \gamma) x_i \left\{ F(x_i^T \beta_T) - G(x_i^T \beta_{Mn}, \gamma) \right\}. \tag{3.5}$$

Expanding β_{Mn} about β_T in (3.5) yields $\beta_T = \beta_{Mn} + \mathcal{H}_{n3}^{-1}(\beta_T, \gamma)\mathcal{C}_{n3}(\beta_T, \gamma) + o(\gamma)$, where

$$\mathcal{H}_{n3}(\beta, \gamma) = n^{-1} \sum_{i=1}^{n} w_i(x_i^T \beta, \gamma) x_i x_i^T F(x_i^T \beta) \left\{ 1 - F(x_i^T \beta) \right\};$$

$$\mathcal{C}_{n3}(\beta, \gamma) = \frac{\gamma}{1 - 2\gamma} n^{-1} \sum_{i=1}^{n} w_i(x_i^T \beta, \gamma) x_i \left\{ 1 - 2F(x_i^T \beta) \right\}. \tag{3.6}$$

Since the ordinary logistic estimate $\widehat{\beta}_L$ is a consistent estimate of β_T, this leads to the estimate

$$\widehat{\beta}_{c,3} = \widehat{\beta}_{Mc} + \mathcal{H}_{n3}^{-1}(\widehat{\beta}_L, \gamma)\mathcal{C}_{n3}(\widehat{\beta}_L, \gamma). \tag{3.7}$$

The bias corrected estimate developed by Copas (1988, equation 28) has the form (3.7), but with (3.6) multiplied by $1-2\gamma$.

An alternative estimator may be obtained by expanding β_T about β_{Mn} in (3.5), leading to an estimate like (3.7) but with the logistic estimate $\widehat{\beta}_L$ replaced by the misclassification mle $\widehat{\beta}_{Mc}$.

3.4. Asymptotic Distribution Theory

The estimates (3.3), (3.4) and (3.7) are of the same form, requiring initial M–estimates followed by a correction. For fixed $\{x_i\}$, the limit theory follows from an easy application of Stefanski's (1985) techniques. Here we merely present the results.

Consider initial estimates $\widehat{\beta}_j$ for $j=0,1$ solving the equations

$$0=\sum_{i=1}^{n}\psi_{ij}(\beta).$$

Define β_{jn} as the solution to $0=\sum_{i=1}^{n}E\{\psi_{ij}(\beta)|x_i\}$. A corrected estimate is given as

$$\widehat{\beta}=\widehat{\beta}_0+\mathcal{H}_N^{-1}(\widehat{\beta}_1)\mathcal{C}_n(\widehat{\beta}_1),\qquad \text{where}$$
$$\mathcal{H}_n(\beta,\gamma)=n^{-1}\sum_{i=1}^{n}H_i(x_i^T\beta);\qquad \mathcal{C}_n(\beta,\gamma)=n^{-1}\sum_{i=1}^{n}C_i(x_i^T\beta).$$

Of course, $\widehat{\beta}$ estimates $\beta_{0n}+\mathcal{H}_N^{-1}(\beta_{1n})\mathcal{C}_n(\beta_{1n})$. If we define

$$V_{nj}(\beta,\gamma)=n^{-1}\sum_{i=1}^{n}E\left\{\frac{\partial}{\partial\beta^T}\psi_{ij}(\beta)\Big|x_i\right\};\qquad L_n(\beta)=\frac{\partial}{\partial\beta^T}\left\{\mathcal{H}_n^{-1}(\beta)\mathcal{C}_n(\beta)\right\},$$
$$Z_in(\beta_{0n},\beta_{1n})=n^{-1}\sum_{i=1}^{n}\left\{V_{n0}^{-1}(\beta_{0n})\psi_{i0}(\beta_{0n})+L_n(\beta_{1n})V_{n1}^{-1}(\beta_{1n})\psi_{i1}(\beta_{1n})\right\},$$

then we clearly have that

$$\begin{aligned}
n^{1/2}(\widehat{\beta}-\beta_n)&\approx n^{1/2}(\widehat{\beta}_0-\beta_{0n})+L_n(\beta_{1n})n^{1/2}(\widehat{\beta}_1-\beta_{1n})\\
&\approx -V_{n0}^{-1}(\beta_{0n})n^{-1/2}\sum_{i=1}^{n}\psi_{i0}(\beta_{0n})-L_n(\beta_{1n})V_{n1}^{-1}(\beta_{1n})n^{-1/2}\sum_{i=1}^{n}\psi_{i1}(\beta_{1n})\\
&\approx -n^{-1/2}\sum_{i=1}^{n}Z_{in}(\beta_{0n},\beta_{1n}).\\
&\approx \text{Normal}\left\{0,\Omega_n(\beta_{0n},\beta_{1n},\beta_T,\gamma)\right\}\qquad\text{where}
\end{aligned}$$
$$\Omega_n(\beta_{0n},\beta_{1n},\beta_T,\gamma)=n^{-1}\sum_{i=1}^{n}E\left\{Z_{in}(\beta_{0n},\beta_{1n})Z_{in}^T(\beta_{0n},\beta_{1n})\right\}.$$

A consistent estimate of the covariance of $\widehat{\beta}$ is

$$n^{-2} \sum_{i=1}^{n} \left\{ Z_{in}(\widehat{\beta}_0, \widehat{\beta}_1) Z_{in}^T(\widehat{\beta}_0, \widehat{\beta}_1) \right\},$$

although in certain cases improved covariance estimates can be constructed, see below.

We illustrate the general result with Copas' estimate, defined as (3.7) with $w_i \equiv 1$. Here $\widehat{\beta}_1 = \widehat{\beta}_L$, the ordinary logistic regression estimate, $\widehat{\beta}_0 = \widehat{\beta}_{Mc}$, $\beta_{1n} = \beta_T$, $\beta_{0n} = \beta_{Mc}$, $V_{n1}(\beta_T) = -\mathcal{H}_n(\beta_T)$, and

$$\psi_{i0}(\beta_{Mn}) = w_i(x_i^T \beta_{Mn}, \gamma) x_i \left\{ Y_i - G(x_i^T \beta_{Mn}, \gamma) \right\}; \qquad \psi_{i1}(\beta_T) = x_i \left\{ Y_i - F(x_i^T \beta_T) \right\};$$

$$\mathcal{H}_n(\beta_T) = n^{-1} \sum_{i=1}^{n} x_i x_i^T F^{(1)}(x_i^T \beta_T); \qquad \mathcal{C}_n(\beta_T, \gamma) = \frac{\gamma}{1-2\gamma} n^{-1} \sum_{i=1}^{n} x_i \left\{ 1 - 2F(x_i^T \beta_T) \right\}$$

$$V_{n0}(\beta_{Mn}) = -(1-2\gamma) n^{-1} \sum_{i=1}^{n} w_i(x_i^T \beta_{Mn}, \gamma) x_i x_i^T F^{(1)}(x_i^T \beta_{Mn});$$

$$L_n(\beta_T, \gamma) = -\frac{2\gamma}{1-2\gamma} I - \mathcal{H}_n^{-1}(\beta_T, \gamma) n^{-1} \sum_{i=1}^{n} x_i x_i^T F^{(2)}(x_i^T \beta_T) x_i^T \mathcal{H}_n^{-1}(\beta_T) \mathcal{C}_n(\beta_T, \gamma).$$

This means that

$$\begin{aligned} Z_{in}(\beta_{Mn}, \beta_T) &= -\left\{ V_{n0}^{-1}(\beta_{Mn}) w_i(x_i^T \beta_{Mn}, \gamma) + L_n(\beta_T, \gamma) V_{n1}^{-1}(\beta_T) \right\} x_i Y_i + \kappa \\ &= \mathcal{B}_i(\beta_T, \beta_{Mn}, \gamma) x_i Y_i + \kappa, \end{aligned}$$

where κ depends only on the design but not Y_i. Now Z_{in} has mean zero, so that under the logistic model, by taking covariance, we see that

$$\Omega_n(\beta_{Mn}, \beta_T, \gamma) = n^{-1} \sum_{i=1}^{n} F^{(1)}(x_i^T \beta_T) \mathcal{B}_i(\beta_T, \beta_{Mn}, \gamma) x_i x_i^T \mathcal{B}_i^T(\beta_T, \beta_{Mn}, \gamma).$$

A consistent covariance estimate for $\widehat{\beta}_{c,3}$ is $n^{-1}\ \Omega_n(\widehat{\beta}_{Mc}, \widehat{\beta}_L, \gamma)$.

4. Consistent Estimates

4.1. Logistic Model

Under the logistic model, the misclassification mle and its corrected versions are inconsistent. We obtain consistency by using a method of moments argument. By starting at the misclassification mle, we hope to retain its robustness properties, especially to non–isolated outliers.

The solution $\widehat{\beta}_{Mc}$ to (2.2) estimates β_{Mn}, the solution to (3.5). Thus, if $\widehat{\beta}$ solves

$$0 = \sum_{i=1}^{n} \mathcal{K}_i(\widehat{\beta}, \gamma, \widehat{\beta}_{Mc}), \tag{4.1}$$

then $\widehat{\beta}$ is consistent for β, see (3.5). Solutions to (4.1) usually exist because they are logistic regression estimates with "response" $G(x_i^T \widehat{\beta}_{Mc}, \gamma)$ and weights $w_i(x_i^T \widehat{\beta}_{Mc}, \gamma)$. By a Taylor series,

$$n^{1/2}(\widehat{\beta} - \beta_T) \approx -A_{2n}^{-1}(\beta_T, \gamma, \beta_{Mn}) A_{1n}(\beta_T, \gamma, \beta_{Mn}) n^{1/2}(\widehat{\beta}_{Mc} - \beta_{Mn}), \qquad \text{where} \tag{4.2}$$

$$A_{1n}(\beta_T, \gamma, \beta) = n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \beta^T} \mathcal{K}_i(\beta_T, \gamma, \beta);$$

$$A_{2n}(\beta_T, \gamma, \beta) = n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \beta_T^T} \mathcal{K}_i(\beta_T, \gamma, \beta).$$

Since $\widehat{\beta}_{Mc}$ is an M–estimate, it has a standard asymptotic theory, so that

$$n^{1/2}(\widehat{\beta} - \beta_T) \Longrightarrow \text{Normal}\left(0, A_{2n}^{-1} A_{1n} \Lambda_n A_{1n}^T A_{2n}^{-T}\right), \qquad \text{where}$$

$$\Lambda_n = V_n^{-1}(\beta_T, \gamma, \beta_{Mn}) R_n(\beta_T, \gamma, \beta_{Mn}) V_n^{-1}(\beta_T, \gamma, \beta_{Mn})$$

$$V_n(\beta_T, \gamma, \beta) = n^{-1} \sum_{i=1}^{n} x_i \left\{ \frac{\partial}{\partial \beta^T} \left[w_i(x_i^T \beta, \gamma) \right\} \left\{ F(x_i^T \beta_T) - G(x_i^T \beta, \gamma) \right\} \right]$$

$$R_n(\beta_T, \gamma, \beta_{Mn}) = n^{-1} \sum_{i=1}^{n} w_i^2(\beta, \gamma) x_i x_i^T F^{(1)}(x_i^T \beta_T)$$

A consistent covariance estimate may be obtained by evaluating these terms at $(\widehat{\beta}, \gamma, \widehat{\beta}_{Mc})$.

4.2. Misclassification–Consistent Estimate

In this subsection, we consider method of moments type corrections to the usual estimates, corrections which yield estimates which are consistent for the misclassification model. The underlying probabilities satisfy (1.2), while $\widehat{\beta}_0$ solves (2.1) with $w_i \equiv c_i \equiv 0$. As $n \longrightarrow \infty$, $\widehat{\beta}_0$ converges to the solution to $\sum w_i x_i \left\{ G(x_i^T \beta_T, \gamma) - F(x_i^T \beta) \right\}$. This suggests a correction to obtain an unbiased estimating equation. Let $\widehat{\beta}$ solve

$$0 = \sum_{i=1}^{n} w_i x_i \left\{ F(x_i^T \widehat{\beta}_0) - \gamma - (1 - 2\gamma) F(x_i^T \widehat{\beta}) \right\}. \tag{4.3}$$

If we define $z_i = \left\{ F(x_i^T \widehat{\beta}_0) - \gamma \right\} / (1 - 2\gamma)$, then the solution to (4.3) is unique in general, as it is the same as a weighted logistic regression applied to the "responses" z_i.

It is easy to show that $\widehat{\beta}$ is consistent for β_T. As in the previous subsection, $\widehat{\beta}$ is asymptotically normally distributed with mean β_T and covariance matrix $n^{-1}\Omega$, where, *under the misclassification model (4.1)*, Ω is consistently estimated by

$$\widehat{\Omega} = \widehat{A}_{3n}^{-1}\widehat{B}_n\widehat{A}_{3n}^{-1}, \qquad \text{where}$$

$$\widehat{A}_{3n} = -(1-2\gamma)n^{-1}\sum_{i=1}^{n} w_i x_i x_i^T F(x_i^T\widehat{\beta})\left\{1 - F(x_i^T\widehat{\beta})\right\};$$

$$\widehat{B}_n = n^{-1}\sum_{i=1}^{n} w_i^2 x_i x_i^T G(x_i^T\widehat{\beta},\gamma)\left\{1 - G(x_i^T\widehat{\beta},\gamma)\right\}$$

The solution to (4.1) has an interesting connection to the bias–corrected estimates of the previous section. The first step of a Newton–Raphson solution to (4.3) has the form

$$\widehat{\beta}_1 = \widehat{\beta}_0 - \left[\sum_{i=1}^{n} w_i x_i x_i^T F(x_i^T\widehat{\beta}_0)\left\{1 - F(x_i^T\widehat{\beta}_0)\right\}\right]^{-1}\sum_{i=1}^{n} w_i x_i\left\{1 - 2F(x_i^T\widehat{\beta}_0)\right\}\frac{\gamma}{1-2\gamma}.$$

This is the same as the bias–corrected estimate (3.4), and its asymptotic distribution theory under the logistic model is discussed in the previous section.

5. Examples

5.1 Introduction

We investigate three examples, one of which involves an extreme leverage and prediction outlier, i.e., an obvious leverage point with extreme predicted probabilities. The second involves an observation which is only an extreme prediction outlier, while the third involves points which are moderate prediction outliers but have considerable leverage. We shall see that the methods behave differently in these data sets.

For the Mallows leverage downweighted and modified Künsch estimates defined in the appendix, we used the tuning constant $b = 8$. For the estimates of Künsch, et al. (1989), we used the tuning constants $b = 18$; Künsch, et al. use $b = 16$. For the misclassification mle and its modifications, we used $\gamma = 0.03$ and 0.01.

5.2. Food Stamp Data

These data are discussed in Künsch, et al. (1989, p. 465). There is a single case, #5, which is isolated in the design space, and appears to be a response outlier. The interesting coefficient is the fourth component β_4, which changes considerably in numerical value and significance level with the deletion of case #5.

The last component of x is at least 4.4, except for case #5 where this component was 0. The fitted logistic coefficient without case #5 was (-5.3, 1.8, -.7, 1.1). The fitted probabilities were at least .33, except for case #5 which had fitted probability .005. In sum, case #5 is both a leverage and a prediction outlier, and so is amenable for analysis by all the methods.

We see from Table 1 that each of the robust/resistant methods which involve leverage down-weight case #5 sufficiently to obtain the desired change from the logistic mle. In this case, we must choose γ rather large, 0.04, in order for Copas' correction to the misclassification mle to to show the desired effect. By looking at posterior probabilities of misclassification, given that data are misclassified with rate γ, it can be seen that several other cases were more likely to be misclassified than case #5, and hence γ needed to be as large as 0.04 for case #5 to be fit as misclassified. For comparison, we show the results for $\gamma = 0.03$, which do not pick up the outlier.

5.3. Leukemia Data

These data are listed in Cook and Weisberg (1982, page 193). One observation, #17, appears to be a response outlier, but it is in a group of three points with identical extreme design values. Without #17, the logistic coefficients were (.2, -23.5, 2.6). Observation #17 had prediction from this fit of less than .001, but the response was $Y = 1$. This is an extreme prediction outlier. However, #17 is not an extreme leverage outlier by itself, and quantifying its leverage is a difficult task. The method of Künsch, et al. and the corrected version of the misclassification mle both show the desired large change from the ordinary logistic mle, although former seems to have elevated standard errors. Case #17 is a likely misclassified point, under model (1.2), and the new parameter values reflect this. The Mallows and modified Künsch methods perform poorly here, as they had difficulty assessing the leverage of the unusual observation. There are other ways of assessing leverage, of course, and they might have led to better performance of the these estimates.

5.4. Rare Event Data Set

The final data set consists of 300 observations with two predictors (X_1, X_2), generated as follows. First, 296 observations were selected from an ongoing study of diet and breast cancer. The 37 cases with $Y = 1$ were included along with a non–random subset of the cases $Y = 0$. Here, X_j

ranges from $-j$ to j, for $j = 1, 2$. In these data, the logistic regression coefficients were (-2.6, 1.5, -.9). The 10^{th} percentile of the probabilities from this fit was .05, while the 90^{th} was .24.

We then added 4 artificial outliers to the data, all of which were extreme in the design space. These were (Y, X_1, X_2) = (1, -.25, 4.5), (1, -.25, 5.5), (0, -.25, -4.5) and (0, -.25, -5.5). The probabilities from the original logistic fit were .0011. .0005, .85, .94. Thus, the first two observations are extreme in the prediction space. The latter two are not, although they are extreme compared to the rest of the data.

Our hypothesis was that all the methods would easily handle the first two generated points, but only the methods which account for leverage can handle the last two points. This is borne out in Table 3, which lists the fitted values and their standard errors.

5.5. Summary

The key feature of these examples that they indicate that leverage and prediction each have individual roles to play in the performance of fitting methods. Methods which only handle leverage can get fooled by a leverage group. Methods which only handle extreme predictions can get fooled by leverage points which have unusual predictions not near 0 or 1. The method of Künsch, et al. (1989) works well on all these examples. However, there surely are examples which can fool this method, and we by no means recommend it uniquely above the others.

6. Further Bias Expansions

6.1. Theoretical Developments

By means of formal asymptotic expansions, one can compute the bias of the robust/resistant estimates discussed in this paper to order $o\left(n^{-1}\right)$. We have already shown how to do this at the logistic model for the Mallows estimates, see section 2.2, where we found that unusual design points can make the bias of the Mallows leverage downweighted estimate smaller than that of the mle. Bias expansions for the Künsch and modified Künsch estimates can be constructed as well, but analytic expressions seem infeasible. Both estimates are M–estimates, and in the appendix we provide formulae for constructing their approximate biases, as well as the bias of the consistent estimates derived in section 4.1. In the appendix, we also compute the bias of the consistent modification of the misclassification mle discussed in section 4.1. In our examples, we computed the necessary derivatives for the Künsch type methods numerically.

6.2. Biases in the Foodstamp Data

In Table 4, we present the approximate biases of the various estimates at the parameter value β^T = (-6.88, 2.02, -.76, 1.33) obtained by deleting cases #5 and #66 and refitting by logistic regression. We see in this table that the Mallows and modified Künsch estimates, which performed very well on these data, have biases approximately as large as the usual logistic regression estimates, while the Künsch method has only a moderately higher bias, at least when compared to the standard errors in Table 1. The consistent misclassification estimate of section 4.1 also has reasonable bias behavior. The score adjusted estimate (2.3) has worse behavior.

Copas' corrected estimate is asymptotically inconsistent, so that one can speak of an asymptotic bias to describe the difference between the correct logistic parameter and the value estimated. With $\gamma = 0.01$, these asymptotic biases are $(.58, -.07, .02, -.10)^T$, while with $\gamma = 0.04$, the asymptotic biases are $(5.3, -.92, .21, -.94)^T$. Thus in this particular example, γ has to remain fairly small in order that the asymptotic bias not become too large. Note that we had to use a value of $\gamma = 0.04$ in order to detect the isolated outlier in this example.

6.3. Biases in the Leukemia Data

In Table 5, we present the approximate biases of the various estimates at the parameter value β^T = (-1.3, -3.2, 2.26) obtained at the logistic regression estimate. It seemed to us nonsensical to evaluate bias at the mle having deleted the unusual point, as Table 2 indicates that the standard errors of any of the estimates having deleted this point are extremely large.

The biases of the Mallows, modified Künsch and consistent misclassification estimates are similar in this example. The score adjusted methods had larger but still reasonable biases. The Künsch method has a much larger bias in the second component. There may be some numerical instability here in our numerical calculation of the necessary derivatives.

6.4. Rare Event Data Set

For this data set we used $\beta_T^T = (-2.55, 1.4, 0.88)$. The bias behavior was much like in the food stamp data, although note that in the third component of β_T, the robust/resistant methods which account for leverage have much smaller bias than does the ordinary logistic mle.

6.5. Summary

While the Künsch method has the best performance in the examples, it has higher $O(1/n)$ bias. The biases were not unacceptably large for the food stamp and rare event data sets, but were unusually large for the leukemia data.

ACKNOWLEDGEMENT

Carroll's research was supported by the Air Force Office of Scientific Research, and partially completed during a visit to the Center for Decision Sciences at Purdue University, which is supported by the National Science Foundation.

APPENDIX

A.1. M-estimates of Location and Scale

Stefanski's idea was as follows. Let $p = \dim(x)$ and set $w_i = w(x_i, \beta)$, where with $m(x,\beta) = \max\{F(x^T\beta), 1 - F(x^T\beta)\}$, for a tuning constant b,

$$w(x,\beta) = \min\left[1, \frac{bp^{1/2}}{m(x,\beta)(x^T M^{-1} x)^{1/2}}\right];$$

$$M = n^{-1}\sum_{i=1}^{n} x_i x_i^T w^2(x_i,\beta) F(x_i^T\beta)\left\{1 - F(x_i^T\beta)\right\}.$$

Instead of this formulation, we have found that the following somewhat simpler method works reasonably well in practice. Write $x_i^T = (1,\ z_i^T)^T$, and let (μ, M) be a "robust" estimate of the center and covariance matrix of the $\{z_i\}$. Let ψ_{1b} be any odd function, and define $\psi_{2b}(v) = \psi_{1b}^2(v)/\xi$, where $\xi = E\psi_{1b}^2(\|Z_{p-1}\|)$ and Z_{p-1} is a $(p-1)$–dimensional normal random variable with zero mean and identity covariance. Define $u_{ib}(v) = \psi_{ib}(v)/v$. The estimates (μ, M) are the solutions to:

$$n^{-1}\sum_{i=1}^{n} u_{1b}\left[\left\{(z_i-\mu)^T M^{-1}(z_i-\mu)\right\}^{1/2}\right](z_i-\mu) = 0;$$

$$n^{-1}\sum_{i=1}^{n} u_{2b}\left\{(z_i-\mu)^T M^{-1}(z_i-\mu)\right\}(z_i-\mu)(z_i-\mu)^T = M.$$

In the calculations, we used the trisquared redescending function

$$\psi_{1b}(v) = v\left\{1-(v/b)^2\right\}^3 I(|v| \le b).$$

A.2. Mallows Leverage Downweighted Estimates

If $d_i = (z-\mu)^T M^{-1}(z-\mu)$, then the Mallows weights to be used in (2.1) are

$$w(x_i,\beta) = w(x_i) = u_{1b}\left\{d_i/(p-1)^{1/2}\right\}.$$

Of course, in (2.1) we set $c(x_i,\beta) \equiv 0$. Note that these weights can redescend to zero, so that points which are extremely outlying in the design space receive zero weight. Note too that different estimates of center and location can cause the weights to change.

A.3. Schweppe Estimates

For a given matrix M, the estimates defined by Künsch, et al (1989) follow (2.1), with

$$w_i = u_{1b}\left\{|Y_i - F(x_i^T\beta) - c(x_i^T\beta, x_i^T M^{-1}x_i)|(x_i^T M^{-1}x_i)^{1/2}\right\},$$

where the function c(a,b) is chosen so that the right hand side of (2.1) has mean zero when evaluated at β_T and the logistic model. The matrix M can be estimated using equations (2.8) and (2.9) of Künsch, et al (1989).

Alternatively, one might estimate the center and scatter matrices (μ, M) as before and define

$$w_i = u_{1b}\left\{|Y_i - F(x_i^T\beta) - c(x_i^T\beta, d_i)|d_i^{1/2}\right\}.$$

In the text, we call this the modified Künsch estimate.

A.4. Bias Expansions for M–estimates

We assume that all expectations are conditional on $x_i, \ldots, x_n$, and that $E\Psi_i(\beta) = 0$. The M–estimate is defined as the solution to $0 = \sum_1^n \Psi_i(\widehat{\beta})$. Write the jth component of Ψ_i as Ψ_{ij}. Dropping the argument β and letting subscript β's denote derivatives, define

$$A_n = n^{-1}\sum_{i=1}^n E\Psi_{i\beta}; \qquad B_n = n^{-1}\sum_{i=1}^n E\Psi_i\Psi_i^T; \qquad C_n = n^{-1}\sum_{i=1}^n E\Psi_{i\beta}A_n^{-1}\Psi_i;$$

$$D_n = \text{vec}\left[n^{-1}\sum_{i=1}^n \text{trace}\left\{(E\Psi_{ij\beta\beta})\,A_n^{-1}B_nA_n^{-1}\right\}\right].$$

We claim that

$$E\left(\widehat{\beta}-\beta\right) = -n^{-1}A_n^{-1}\left\{(1/2)D_n - C_n\right\} + o(1). \qquad (A.1)$$

Proving (A.1) involves two simple steps. First recall from standard theory that

$$\left(\widehat{\beta}-\beta\right) = -A_n^{-1}n^{-1}\sum_{i=1}^n \Psi_i + O_p\left(n^{-1}\right), \qquad (A.2)$$

so that $\left(\widehat{\beta}-\beta\right)$ has covariance $n^{-1}A_n^{-1}B_nA_n^{-1}$. By a simple Taylor series,

$$0 = n^{-1}\sum_{i=1}^{n}\left[\Psi_{ij} + \Psi_{ij\beta}^T\left(\widehat{\beta}-\beta\right) + (1/2)\text{trace}\left\{\Psi_{ij\beta\beta}\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^T\right\}\right] + O_p(n^{-3/2}).$$

If a_{nj}^T is the jth row of A_n, using (A.2) we thus have

$$0 = n^{-1}\sum_{i=1}^{n}\Psi_{ij} + a_{nj}^T\left(\widehat{\beta}-\beta\right) - n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left\{\Psi_{ij\beta} - E\left(\Psi_{ij\beta}\right)\right\}^T A_n^{-1}\Psi_i\right]$$
$$+ (1/2)n^{-1}\sum_{i=1}^{n}\text{trace}\left\{E\left(\Psi_{ij\beta\beta}\right)\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^T\right\} + O_p(n^{-3/2}).$$

Taking expectations yields the result.

It is possible to eliminate the second partial derivatives, as follows. Write $\Psi_{ij} = \Psi_{ij}(Y_i|x_i,\beta)$ and $h(y|x\beta) = F(x^T\beta)^y\left\{1-F(x^T\beta)\right\}^{1-y}$. By definition, for *every* β,

$$0 = \sum_{y=0}^{1}\Psi_{ij}(y|x,\beta)h(y|x,\beta).$$

Taking two derivatives, we obtain

$$E\Psi_{ij\beta\beta} = -\sum_{y=0}^{1}\left\{\Psi_{ij\beta}(y)h_\beta^T(y) + h_\beta(y)\Psi_{ij\beta}^T(y) + \Psi_{ij}(y)h_{\beta\beta}(y)\right\}.$$

A.5. Bias Expansion for the Consistent Misclassification Estimate

In this section we compute a bias expansion for the consistent misclassification estimate defined by (4.1). If we write $w_i = w_i(x_i^T\beta_{Mn},\gamma)$, $G_i = G(x_i^T\beta_{Mn},\gamma)$, $U_i = w_iG_i$ and $F_i = F(x_i^T\beta_T)$, then by a Taylor series of (4.1) we obtain

$$0 = n^{-1}\sum_{i=1}^{n}x_ix_i^T\left\{\left(w_i^{(1)}F_i - U_i^{(1)}\right)\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right) + w_iF_i^{(1)}\left(\widehat{\beta}-\beta_T\right)\right\}$$
$$+ (1/2)n^{-1}\sum_{i=1}^{n}x_i\left[\left(w_i^{(2)}F_i - U_i^{(2)}\right)\left\{x_i^T\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right)\right\}^2 + w_iF_i^{(2)}\left\{x_i^T\left(\widehat{\beta}-\beta_T\right)\right\}^2\right]$$
$$+ (1/2)n^{-1}\sum_{i=1}^{n}x_iw_i^{(1)}F_i^{(1)}\left\{x_i^T\left(\widehat{\beta}-\beta_T\right)\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right)^Tx_i + x_i^T\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right)\left(\widehat{\beta}-\beta_T\right)^Tx_i\right\}.$$

From (4.2), replace $\left(\widehat{\beta}-\beta_T\right)$ by $A_{2n}^{-1}A_{1n}\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right)$ in the last two sums, and recall from standard M–estimator theory that the covariance of $\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right)$ under the logistic model is $n^{-1}V$, where $V = S_1^{-1}S_4S_1^{-1}$, where

$$S_1 = n^{-1}\sum_{i=1}^{n}x_ix_i^T\left(w_i^{(1)}F_i - U_i^{(1)}\right); \qquad S_4 = n^{-1}\sum_{i=1}^{n}x_ix_i^Tw_i^2F_i^{(1)}.$$

Taking expectations, we obtain

$$E\left(\widehat{\beta}-\beta_T\right) = -L_1^{-1}L_2E\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right) - (2n)^{-1}L_1^{-1}(L_3+L_4+L_5)+o(1), \text{ where} \quad (A.3)$$

$$L_1 = n^{-1}\sum_{i=1}^{n} x_i x_i^T w_i F_i^{(1)}; \qquad L_2 = n^{-1}\sum_{i=1}^{n} x_i x_i^T (w_i^{(1)}F_i - U_i^{(1)});$$

$$L_3 = n^{-1}\sum_{i=1}^{n} x_i w_i F_i^{(2)} a_{3i}; \qquad L_4 = n^{-1}\sum_{i=1}^{n} x_i a_{4i}(w_i^{(2)}F_i - U_i^{(2)});$$

$$L_5 = n^{-1}\sum_{i=1}^{n} x_i w_i^{(1)} F_i^{(1)} a_{5i}; \qquad a_{3i} = x_i^T A_{2n}^{-1} A_{1n} V A_{1n} A_{2n}^{-1} x_i;$$

$$a_{4i} = x_i^T V x_i; \qquad a_{5i} = -x_i^T A_{2n}^{-1} A_{1n} V x_i - x_i^T V A_{1n} A_{2n}^{-1} x_i.$$

Thus, from (A.3), we have to compute the first order bias for the misclassification estimate. This is just an M–estimate, and the previous section can be used to compute this expansion. Detailed calculations show that

$$E\left(\widehat{\beta}_{Mc}-\beta_{Mn}\right) = -(2n)^{-1}S_1^{-1}S_2 + n^{-1}S_1^{-1}S_3, \qquad \text{where}$$

$$S_2 = n^{-1}\sum_{i=1}^{n} x_i (x_i^T V x_i \left(w_i^{(2)}F_i - U_i^{(2)}\right)$$

$$S_3 = n^{-1}\sum_{i=1}^{n} x_i w_i w_i^{(1)} F_i^{(1)} x_i^T S_1^{-1} x_i.$$

REFERENCES

Copas, J. B. (1988). Binary regression models for contaminated data (with discussion). *Journal of the Royal Statistical Society, Series B*, 50, 225–265.

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

Künsch, H. R., Stefanski, L. A. and Carroll, R. J. (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.

Pregibon, D. (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, 38, 485-498.

Stefanski, L. A., Carroll, R. J. and Ruppert, D. (1986). Optimally bounded score functions for generalized linear models, with applications to logistic regression. *Biometrika*, 73, 413–425.

TABLE 1 – FOODSTAMP DATA SET

For the various methods, parameter estimates are listed, along with standard errors in parentheses. A "*" indicates lack of convergence.

Method	β_1	β_2	β_3	β_4
Logistic mle	-.93	1.85	-.90	.33
	(1.62)	(.53)	(.50)	(.27)
Mallows Leverage, b = 8	-5.22	1.80	-.66	1.05
Sections 2.2 and A.2	(2.63)	(.54)	(.52)	(.44)
Künsch, b = 18	-5.22	1.67	-.69	1.04
Sections 2.3 and A.3	(2.46)	(.50)	(.50)	(.42)
Modified Künsch, b = 8	-5.77	1.83	-.69	1.15
Section A.3	(2.68)	(.54)	(.52)	(.46)
Copas, $\gamma = 0.01$	-.88	1.83	-.89	.32
	(1.56)	(.53)	(.50)	(.27)
Copas, $\gamma = 0.03$	-.71	1.83	-.84	.29
	(1.51)	(.58)	(.50)	(.26)
Copas, $\gamma = 0.04$	-7.90	1.85	-.62	1.51
	(4.03)	(.63)	(.55)	(.72)
Consistent, $\gamma = 0.03$	-.81	1.79	-.86	.31
Equation (4.1)	(1.47)	(.40)	(.46)	(.25)
Consistent, $\gamma = 0.04$	-6.43	1.86	-.71	1.26
	(2.54)	(.41)	(.48)	(.42)
Score Adjusted, $\gamma = 0.03$	-.87	1.83	-.88	.32
Equation (2.3)	(1.62)	(.53)	(.50)	(.27)

TABLE 2 – LEUKEMIA DATA SET

For the various methods, parameter estimates are listed, along with standard errors in parentheses.

Method	β_1	β_2	β_3
Logistic mle	-1.31 (.81)	-3.18 (1.86)	2.26 (.95)
Mallows Leverage, b = 8	-1.20 (.83)	-4.05 (2.35)	2.24 (.97)
Künsch, b = 18	.02 (1.08)	-18.55 (13.25)	2.38 (1.17)
Modified Künsch, b = 8	-1.22 (.84)	-4.20 (2.46)	2.30 (.98)
Copas, $\gamma = 0.01$	.22 (1.34)	-22.97 (26.50)	2.51 (1.62)
Consistent, $\gamma = 0.01$	.19 (1.05)	-20.27 (12.40)	2.42 (1.18)
Score Adjusted $\gamma = 0.01$	-1.27 (.81)	-3.25 (1.91)	2.22 (.98)
Score Adjusted $\gamma = 0.03$	.15 (1.08)	-21.41 (13.58)	2.47 (1.22)

TABLE 3 – SMALL RATE DATA SET

For the various methods, parameter estimates are listed, along with standard errors in parentheses.

Method	β_1	β_2	β_3
Logistic mle	-2.34	1.05	.07
	(.30)	(.54)	(.23)
Mallows Leverage, b = 8	-2.61	1.49	-.96
	(.34)	(.58)	(.33)
Künsch, b = 18	-2.44	1.22	-.85
	(.33)	(.58)	(.32)
Modified Künsch, b = 8	-2.68	1.49	-.96
	(.34)	(.58)	(.33)
Copas, $\gamma = 0.01$	-2.33	1.04	.05
	(.30)	(.54)	(.23)
Copas, $\gamma = 0.03$	-2.58	1.41	-.35
	(.41)	(.66)	(.15)
Consistent, $\gamma = 0.03$	-2.47	1.27	-.27
	(.23)	(.46)	(.20)
Score Adjusted	-2.33	1.04	-.01
$\gamma = 0.03$	(.30)	(.54)	(.23)

TABLE 4 – FOODSTAMP DATA SET

For the various methods, approximate biases of the parameter estimates are listed. A "*" indicates lack of convergence.

Method	β_1	β_2	β_3	β_4
Logistic mle	-.487	.127	-.027	.091
Mallows Leverage, b = 8	-.463	.123	-.027	.086
Künsch, b = 18	-.634	.197	-.043	.119
Modified Künsch, b = 8	-.518	.135	-.026	.097
Consistent Misclass, γ = .01	-.529	.140	-.033	.098
Consistent Misclass, γ = .03	-.638	.179	-.045	.119
Score Adjusted, $\gamma = 0.03$	-.794	.246	-.058	.156

TABLE 5 – LEUKEMIA DATA SET

For the various methods, approximate biases of the parameter estimates are listed. A "*" indicates numerical instability.

Method	β_1	β_2	β_3
Logistic mle	-.103	-.976	.270
Mallows Leverage, b = 8	-.107	-.934	.269
Künsch, b = 18	*	*	*
Modified Künsch, b = 8	-.119	-.823	.270
Consistent Misclass, γ = .01	-.121	-1.00	.279
Consistent Misclass, γ = .03	-.109	-1.07	.292
Score Adjusted, γ = 0.01	-.120	-1.22	.326
Score Adjusted, γ = 0.03	-.145	-1.51	.393

TABLE 6– SMALL RATE DATA SET

For the various methods, approximate biases of the parameter estimates are listed.

Method	β_1	β_2	β_3
Logistic mle	-.052	.041	-.045
Mallows Leverage, b = 8	-.052	.042	-.013
Künsch, b = 18	-.070	.059	-.015
Modified Künsch, b = 8	-.056	.045	-.002
Consistent Misclass, γ = .01	-.053	.042	-.046
Consistent Misclass, γ = .03	-.064	.054	-.051
Score Adjusted, γ = 0.01	-.067	.057	-.056
Score Adjusted, γ = 0.03	-.088	.075	-.071