STATISTICS IN MEDICINE, VOL. 2, 71–78 (1983)

BIAS CORRECTION IN MAXIMUM LIKELIHOOD LOGISTIC REGRESSION

ROBERT L. SCHAEFER

Department of Mathematics and Statistics, Miami University, Oxford, Ohio, U.S.A.

SUMMARY

This paper provides an expression for bias of the maximum likelihood logistic regression estimates for use with small sample sizes. This bias correction is based on an expansion of the maximum likelihood equation. Simulation results show these corrections to be highly effective in small samples.

KEY WORDS Bias Maximum Likelihood Logistic Regression Simulation

1. INTRODUCTION

Logistic regression using Maximum Likelihood (ML) estimation has found widespread use in the medical field in modelling the probability of survival and the assessment of risk factors.[1,2] The assessment of risk factors is critical in the care of the medical patient. Comparing different treatment regimens as well as assessing the effects of demographic characteristics on survival are two important components of risk factor assessment. In this paper we address only this risk factor assessment aspect of logistic regression.

It is well known that ML estimates are asymptotically unbiased and their asymptotic variances approximate the variances of these estimates.[3] Since most 'risk factor assessment' analyses entail large sample sizes (from 500 to 14,000), use of these asymptotic results appears appropriate. The use of such approximations, however, is questionable when the sample size is small. For small samples, the ML estimates may have substantial bias and, if no account is made of the bias, could lead to incorrect conclusions concerning the effects of the risk factors.

Assuming the risk factors to have a joint multivariate normal distribution, Anderson and Richardson[4] considered adjustment of the logistic discriminant estimates using a bias correction factor; their simulation results indicated the correction factor was effective. In this paper we consider a bias correction for the ML logistic regression estimate. In Section 2 we derive the correction and investigate its order. Section 3 contains the results of the simulation and shows that the bias correction substantially improves the estimates in certain situations.

2. DERIVATION OF THE BIAS CORRECTION

Letting y_i denote the 0/1 binary dependent variable, its probability density function (pdf) is

$$pdf(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i};\ y_i = 0, 1;\ i = 1, 2, \ldots, n \tag{1}$$

where $\pi_i = (1+\exp(-\beta_0 - \beta_1 x_{i1} - \ldots - \beta_{p-1}x_{ip-1}))^{-1}$, the βs are the unknown regression parameters, and the xs are the observable risk factors or independent variables. Letting $\mathbf{X}$ be the

0277–6715/83/010071–08$01.00

Received October 1981
Revised July 1982

$(n \times p)$ matrix of risk factors, the conditional (on $\mathbf{X}$) ML estimate, $\hat{\beta}$, is the solution of

$$\mathbf{X}^{\mathrm{T}}(\mathbf{y} - \hat{\pi}) = \mathbf{0} \tag{2}$$

where $\mathbf{y}$ is the $(n \times 1)$ vector of the y_i and $\hat{\pi}$ is the $(n \times 1)$ vector of the $\hat{\pi}_i$, the ML estimate of π_i. (Alternatively, one could use iterative weighted least squares on the model

$$\mathbf{y} = \pi + \varepsilon$$

where π is the $(n \times 1)$ vector of the π_i and ε is an $(n \times 1)$ vector of independent Bernoulli random errors, ε_i, with $E(\varepsilon) = \mathbf{0}$ and $\mathrm{Var}(\varepsilon) = \mathbf{V} = \mathrm{diag}\{v_i\} = \mathrm{diag}\{\pi_i(1-\pi_i)\}$ to arrive at the same estimate.[5])

We assume that the independent variables are bounded so that $\lim(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})/n$ is finite and positive definite as $n \to \infty$. Then, under certain regularity conditions, it is well known[3] that $(\hat{\beta} - \beta)$ converges in distribution to a p-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}$. Thus if the sample size is large, $\hat{\beta}$ is approximately unbiased and has approximate covariance matrix given by $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}$. If the sample size is small, however, $\hat{\beta}$ might be biased.

To obtain an approximation for the bias of $\hat{\beta}$, we expand $\hat{\pi}$ in (2) in a Taylor series expansion in $\hat{\beta}$ about β to obtain

$$\mathbf{X}^{\mathrm{T}}\left[\mathbf{y} - \pi - \mathbf{V}\mathbf{X}(\hat{\beta} - \beta) - (1/2)\begin{bmatrix} (\hat{\beta} - \beta)^{\mathrm{T}} E^1 (\hat{\beta} - \beta) \\ \vdots \\ (\hat{\beta} - \beta)^{\mathrm{T}} E^n (\hat{\beta} - \beta) \end{bmatrix}\right] = \mathbf{0} \tag{3}$$

where

$$E^k = v_k(1 - 2\pi_k)\mathbf{x}_k\mathbf{x}^{\mathrm{T}}{}_k$$

and $\mathbf{x}^{\mathrm{T}}{}_k$ is the kth row of $\mathbf{X}$. We obtain an approximate solution by rewriting (3) as

$$(\hat{\beta} - \beta) = (\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\left\{\mathbf{X}^{\mathrm{T}}\varepsilon - (1/2)\mathbf{X}^{\mathrm{T}}\begin{bmatrix} (\hat{\beta} - \beta)^{\mathrm{T}} E^1 (\hat{\beta} - \beta) \\ \vdots \\ (\hat{\beta} - \beta)^{\mathrm{T}} E^n (\hat{\beta} - \beta) \end{bmatrix}\right\} \tag{4}$$

finding an initial solution for $(\hat{\beta} - \beta)$, say $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\varepsilon$, and substituting for $(\hat{\beta} - \beta)$ in the RHS of (4) to obtain

$$(\hat{\beta} - \beta) \approx (\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\varepsilon - (1/2)(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\begin{bmatrix} \varepsilon^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}E^1(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\varepsilon \\ \vdots \\ \varepsilon^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}E^n(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\varepsilon \end{bmatrix} \tag{5}$$

From (5) we can find an approximation for the bias of $\hat{\beta}$ since we have assumed $E(\varepsilon) = \mathbf{0}$ and $\mathrm{Var}(\varepsilon) = \mathbf{V}$.

The bias is

$$\begin{aligned} \mathrm{bias}(\hat{\beta}) \approx\ & (-1/2)(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\{\mathrm{trace}(\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}E^k(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\,\mathrm{Var}(\varepsilon))\} \\ & (-1/2)(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}\{(1 - 2\pi_k)\mathbf{x}^{\mathrm{T}}{}_k(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_k\} \end{aligned} \tag{6}$$

where $\{a_k\}$ represents a vector whose kth element is a_k.

By writing a typical element of (6) as a sum and noting that, since we assumed $\lim (\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})/n$ finite, the elements of $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}$ are $O(1/n)^6$, one can show that the elements in (6) are $O(1/n)$. To ensure accounting for all terms of $O(1/n)$, we carried the Taylor series expansion in (3) to the third order. The resulting additional terms were all $O(1/n^2)$, hence (6) represents the bias of $\hat{\boldsymbol{\beta}}$ to order $O(1/n)$.

One might argue that the bias should be corrected to order $O(1/n^2)$ since these additional terms might have a substantial effect on the bias if n is small. We found, however, that these additional terms were computationally impractical. The number of $(n \times n)$ *matrix computations* required to compute these terms increases by a factor of n^2 beyond the computations required for the $O(1/n)$ terms.

Further, as shown in the next section, one can obtain substantial improvements by using only the $O(1/n)$ terms in the bias when the sample size is small. In fact, one attains the greatest improvements for small sample sizes. In the light of such improvements with small sample size, the utility of these additional terms is questionable.

Since the bias is a function of $\boldsymbol{\beta}$, through $\mathbf{V}$, (6) has limited practical usefulness. One can, however, estimate the bias by obtaining $\hat{\boldsymbol{\beta}}$, computing $\hat{\boldsymbol{\pi}}$ and $\hat{\mathbf{V}}$, then estimating the bias by

$$\widehat{\text{bias}}\,(\hat{\boldsymbol{\beta}}) \approx (-1/2)\,(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}\{(1-2\pi_k)\mathbf{x}^{\mathrm{T}}{}_k(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})^{-1}\mathbf{x}_k\}$$

The 'bias corrected' or, simply, corrected ML estimate of $\hat{\boldsymbol{\beta}}$ is then

$$\hat{\boldsymbol{\beta}} - \widehat{\text{bias}}\,(\hat{\boldsymbol{\beta}}).$$

3. SIMULATION

In designing the simulation we relied heavily on the strong similarities between multiple regression and logistic regression, namely, in regard to models, estimation techniques, and goals of analysis. A view of logistic regression as a non-linear regression problem and the use of iterative weighted least squares estimation make logistic regression similar in many respects to regression. Hence, we designed the simulation, in part, from findings of simulations to investigate biased estimation in regression analysis.[7-11] These multiple regression studies found that the number of independent variables $(p-1)$ and the angle between $\boldsymbol{\beta}$ and the eigenvector associated with the smallest eigenvalue of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ affected the performance of the Ridge and other biased estimators. Owing to the similarities between multiple and logistic regression, we investigated two values of $p-1$ (2 and 5) and two angles (0° and 90°). We determined the angle by setting $\boldsymbol{\beta}$ to be the eigenvector associated with the smallest (largest) eigenvalue of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ thus obtaining an orientation of 0°(90°). (Ideally, we should have set $\boldsymbol{\beta}$ to be the eigenvectors of $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})$ but since $\mathbf{V}$ is a function of $\boldsymbol{\beta}$ we could not obtain $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})$ without first obtaining $\boldsymbol{\beta}$; hence, we used $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$.)

We also felt that collinearity among the independent variables would affect the bias, since $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})$ would be near singular and would make the bias unstable. Thus we investigated three degrees of collinearity, none, moderate, and severe. These degrees of collinearity are represented by the maximum R^2 values (the maximum coefficient of determination among the independent variables) of 0, 90 and 99 per cent. We did not investigate more severe collinearity since the collinearity problem is magnified fourfold in logistic regression. This would have caused severe problems with non-singular $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})$ matrices and would have made simulation nearly impossible.

Further, since the bias is $O(1/n)$, we also investigated the effect of sample size on the bias. The values of n selected were 25, 50, and 200. We felt that samples below 25 represented an unrealistic use of logistic regression and that sample sizes above 200 would yield an insignificant bias correction term.

Finally, owing to the cost of running such simulations, we did not investigate every combination of these four factors. Rather, to obtain the general pattern of performance we ran only the situations indicated by an asterisk in Table I. Hence there were a total of $2 \times 2 \times 6 = 24$ different simulations.

Table I. The combinations of sample size and degree of collinearity investigated in the simulation for *all* combinations of $(p-1)$ and angle.

Sample size	Degree of collinearity: 0%	90%	99%
25	*	*	
50		*	*
200	*		*

The model used in the simulation was

$$\Pr(y_i = 1) = \pi_i = (1 + \exp(-\beta_0 - \beta_1 x_{i1} - \ldots - \beta_{p-1} x_{ip-1}))^{-1}.$$

For each replication of each simulation setting, we generated the **X** matrix using the uniform (0, 1) psuedo-random number generator in the statistical analysis system (SAS).[12] First, we set all values of the independent variables in the **X** matrix as independently generated psuedo-random numbers. We then introduced the degree of collinearity. For those simulations with no collinearity, we left the **X** matrix unchanged. For the other two categories, we introduced a single collinearity by replacing the last variable, x_{ip-1}, with a linear combination of other variables as follows:

$$\begin{aligned} &\text{for } p-1 = 2: \quad x_{i2} = x_{i1} + k^* u(0, 1),\ k = 0{\cdot}3333 \text{ and } 0{\cdot}1005 \\ &\text{for } p-1 = 5: \quad x_{i5} = x_{i3} + x_{i4} + k^* u(0, 1),\ k = 0{\cdot}45 \text{ and } 0{\cdot}142 \end{aligned} \tag{7}$$

where $u(0, 1)$ was another independently generated uniform (0, 1). The resulting theoretical correlation matrices of the independent variables were:

$$\text{(i)} \begin{bmatrix} 1 & p \\ p & 1 \end{bmatrix} \qquad \text{(ii)} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & p \\ 0 & 0 & 0 & 1 & p \\ 0 & 0 & p & p & 1 \end{bmatrix}$$

$p = 0{\cdot}9487$ and $0{\cdot}9950$ $\qquad$ $p = 0{\cdot}6738$ and $0{\cdot}7036$

corresponding respectively to the ks in (7). The resulting degree of collinearity was then max $R^2 = 0$, 90 and 99 per cent.

Once we calculated **X**, we set $\boldsymbol{\beta}$ to be the eigenvector associated with the smallest (largest) eigenvalue of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$. We then computed the π_i using $\boldsymbol{\beta}$ and the rows of **X**. We generated the y_i by

$$y_i = \begin{cases} 1, & \text{if } \pi_i \leqslant u(0, 1) \\ 0, & \text{otherwise} \end{cases}$$

where $u(0, 1)$ was another independently generated uniform (0, 1). We replicated $\mathbf{X}$ and $\mathbf{y}$, the vector of y_i, approximately 500 times for each of the 24 simulation settings. (On occasions, the randomly generated data yielded a singular $\mathbf{X}$ matrix and hence we could not compute the ML estimate. We removed such occurrences from the simulation results.)

Once we generated $\mathbf{X}$ and $\mathbf{y}$, we computed $\hat{\beta}$ using the SAS logistic regression program, LOGIST.[13] We used the PROC MATRIX command in SAS to compute the corrected ML estimate. We then computed the squared error (SQE), $(\hat{\beta}-\beta)^{\mathrm{T}}(\hat{\beta}-\beta)$, for each estimate and averaged these over all replications to yield an estimate of the mean squared error (MSE).

Table II contains selected results from the simulation, the mean and standard deviation of the SQEs for both estimates. These findings yield two conclusions. First is the effect of sample size on the magnitude of SQEs. For every cell in Table II, the estimated MSE decreased as the sample size increased. Further, the estimated MSE of the corrected ML estimate approached that of the ML estimate, verifying that, as expected, the bias is $O(1/n)$ and hence should be negligible for large samples. As a result it would seem that one need not correct estimates based on samples of size 200. One should certainly consider, however, bias correction for sample sizes of 25 and 50. Based on the simulation performed, one cannot make recommendations for sample sizes between 50 and 200. The minimal cost and effort of computing the corrected ML estimate, however, suggest use of the correction.

The second conclusion from the results in Table II is the magnitude and instability in the squared errors induced by the collinearity. The collinearity affected not only the size of the SQEs (examine any row in Table II), but the variability of the SQEs also increased as the degree of collinearity increased. One would expect this result since the collinearity affects the stability of the $(\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X})$ matrix and, hence, also makes the ML estimate very unstable. Although the corrected ML estimate was also affected, it was not affected to the same extent.

Table III presents the main results of the simulation and contains the ratio of the estimated MSE of the corrected ML estimate to the estimated MSE of the ML estimate. The number of replications of each simulation setting appears in parentheses. We compared simulation results after 150, 300, 450, and 600 replications and, in general, found that results stabilized at about 300 replications; results varied little beyond 300. Hence the one 'small' replication size (number of replications = 378) should not pose any problem.

The results in Table III yield four major conclusions. First is the effect of sample size. As expected, the most substantial improvements occurred with small sample sizes, with decreasing gains observed as n increased. One might offset the rather small gains realized when the sample size was 200 by the additional computations necessary to compute the corrected ML estimate; hence, we do not recommend bias correction in these circumstances. For small sample sizes, however, which realized reductions from 30 per cent up to 75 per cent, we recommend correction of the ML estimate for bias. We deem the extra computations as certainly warranted in this situation.

Second, no clear cut pattern emerged in regard to the Degree of collinearity. There appear to be minor improvements as the degree of collinearity increased, but they were so minor, one is tempted to ignore degree of collinearity and its effect on the bias correction.

Third, in regard to the number of risk factors, the corrected ML estimate was much better than the ML estimate when the number of risk factors was large. Fourth, and in a similar fashion, the angle between β and the eigenvector associated with smallest eigenvalue of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ also played an important role. Substantial gains accrued as β became orthogonal to this eigenvector. This was especially true in small sample sizes.

Overall, the results indicate that one should use the corrected ML estimate when (1) the sample size is small, (2) the number of independent variables is large and/or (3) the parameter vector is orthogonal to the eigenvector associated with the smallest eigenvalue of $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$.

Table II. Mean of the squared errors for the ML estimate with the standard deviation in parentheses followed by the mean of the squared errors for the corrected ML estimate with the standard deviation in parentheses.

Angle	$p-1$	Sample size	Degree of collinearity 0% ML	0% Corrected ML	90% ML	90% Corrected ML	99% ML	99% Corrected ML
0°	2	25	10·38 (12·22)	7·46 (7·94)	69·30 (115·13)	45·67 (70·22)		
		50			28·80 (41·79)	24·55 (35·21)	244·38 (375·95)	205·97 (313·18)
		200	4·79 (2·37)	4·68 (2·30)			54·53 (72·85)	52·51 (70·10)
0°	5	25	39·75 (74·36)	13·03 (9·99)	106·31 (150·71)	40·81 (50·41)		
		50			33·17 (49·01)	24·09 (34·08)	172·71 (248·30)	125·09 (177·58)
		200	5·63 (2·63)	5·38 (2·48)			45·14 (64·20)	42·21 (59·00)
90°	2	25	16·73 (33·54)	8·91 (7·04)	93·46 (152·07)	52·70 (76·89)		
		50			38·02 (78·65)	30·58 (55·82)	265·67 (390·81)	215·73 (315·62)
		200	4·73 (1·05)	4·63 (1·01)			68·44 (89·98)	65·44 (85·88)
90°	5	25	50·22 (66·25)	12·48 (8·10)	134·44 (211·22)	33·87 (39·95)		
		50			45·00 (65·43)	29·04 (32·67)	151·31 (217·76)	106·30 (150·60)
		200	5·94 (1·53)	5·65 (1·41)			59·27 (74·81)	54·87 (68·95)

Table III. Ratio of the estimated MSE of the corrected ML estimate to the estimated MSE of the ML estimate and the number of replications in the simulation in parentheses.

Angle	$p-1$	Sample size	Degree of collinearity 0%	90%	99%
0°	2	25	71·87% (515)	65·90% (600)	
		50		85·24% (600)	84·28% (600)
		200	97·70% (487)		96·30% (600)
0°	5	25	32·78% (516)	38·39% (585)	
		50		72·63% (520)	72·43% (542)
		200	95·56% (600)		93·51% (600)
90°	2	25	53·26% (596)	56·39% (596)	
		50		80·43% (600)	81·20% (423)
		200	97·86% (524)		95·62% (542)
90°	5	25	24·85% (418)	25·19% (424)	
		50		64·53% (566)	70·25% (378)
		200	95·12% (600)		92·58% (600)

4. CONCLUSION AND SUMMARY

The bias correction for ML logistic regression estimates has utility particularly with small samples. Of the four components addressed in the simulation, only collinearity did not substantially affect the bias correction. With small sample sizes the bias correction can achieve large reductions in MSE; with larger sample sizes the reduction is less. As the number of risk factors increases, the correction appears to reduce the MSE by a factor of 10 per cent to 50 per cent. Furthermore, when the parameter and collinearity vectors are orthogonal, the bias correction can result in smaller MSE. (Very large reductions occurred in the small sample size—no collinearity setting with orthogonal parameter and collinearity vectors, i.e. the eigenvector associated with the smallest eigenvalue of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$.) Finally, the effect of collinearity increases the magnitude and variability of the two estimates. Hence the bias correction would seem to be affected by the degree of collinearity in a similar manner as the ML estimate so that one observes no improvement in MSE as the degree of collinearity increases.

In summary, the particular appeals of this bias correction are its relative ease of computation and implementation and its relatively low cost for small samples. We strongly recommend bias correction of ML estimates when the sample size is small. In the small sample size logistic regression, bias correction should substantially improve the assessment of risk factors.

ACKNOWLEDGEMENTS

This research was funded by a 1981 Summer Research Grant from Miami University. The helpful comments of two referees are also acknowledged.

REFERENCES

1. D'Agostino, R. B., Pozen, M. W., Mitchell, J., Teebagg, N. C., Guglielmino, J. T., Bielawski, L. I. and Hood, W. B. 'Comparison of logistic regression and discriminant analysis as emergency room decision models for the diagnosis of acute coronary disease', *Research Report #2–78*, Department of Mathematics, Boston University, Boston, Massachusetts, 1978.
2. Cornell, R. G. and Feller, I. 'Evaluation of emergency medical services with a national burn registry', *Technical Report*, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, 1979.
3. Cox, D. R. *The Analysis of Binary Data*, Metheun, London, 1970.
4. Anderson, J. A. and Richardson, S. C. 'Logistic discrimination bias correction in maximum likelihood estimation', *Technometrics*, **21**, 71–78 (1979).
5. Walker, S. H. and Duncan, D. B. 'Estimation of the probability of an event as a function of several independent variables', *Biometrika*, **54**, 167–179 (1967).
6. Bishop, Y. M. M., Feinberg, S. E., and Holland, P. W. *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, 1975.
7. Gunst, R. F. and Mason, R. L. 'Biased estimation in regression: an evaluation using mean squared error', *Journal of the American Statistical Association*, **72**, 616–628 (1977).
8. Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. 'Ridge regression: some simulations', *Communications in Statistics—Theory and Methodology*, **4**, 105–123 (1975).
9. Lawless, J. F. and Wang, P. 'A simulation study of ridge and other regression estimators', *Communications in Statistics—Theory and Methodology*, **5**, 307–323 (1976).
10. McDonald, G. C. and Galarneau, D. I. 'A Monte Carlo evaluation of some ridge-type estimators', *Journal of the American Statistical Association*, **70**, 407–416 (1975).
11. Wichern, D. W. and Churchill, G. A. 'A Comparison of ridge estimators', *Technometrics*, **20**, 301–311 (1978).
12. *SAS User's Guide*, SAS Institute, North Carolina, 1979.
13. *SAS Supplemental Library User's Guide*, SAS Institute, North California, 1980.