Biometrika (1990), 77, 1, *pp.* 23–31
Printed in Great Britain

Principal component estimation for generalized linear regression

BY BRIAN D. MARX
Department of Experimental Statistics, Louisiana State University, Baton Rouge, Louisiana 70803-5606, U.S.A.

AND ERIC P. SMITH
Department of Statistics, Virginia Polytechnic Institute & State University, Blacksburg, Virginia 24061, U.S.A.

SUMMARY

The generalized linear model (Nelder & Wedderburn, 1972) has become an elegant and practical option to classical least-squares linear model building. We consider the specific problem of generalized linear regression utilizing a set of continuous explanatory variables to model an exponential family response. It is the objective of this paper to develop and present an asymptotically biased principal component parameter estimation technique, as an option to traditional maximum likelihood estimation for generalized linear regression. Both iterative and one-step principal component estimators are developed, directly compared, and can be particularly useful with the presence of an ill-conditioned information matrix. The bias, variance and mean squared error of principal component estimation will be quantified. Generalizations for rules of deletion of components will be examined. Lastly, an example employs principal component estimation for Poisson response data.

Some key words: Generalized linear model; Maximum likelihood estimator; Mean squared error; Principal component estimator.

1. BACKGROUND OF GENERALIZED LINEAR REGRESSION

Let $X^* = (x_1^*, \ldots, x_p^*)$ be a $N \times p$ matrix of continuous regression explanatory variables. Consider centring and scaling X^* so that $X'X$ is in correlation form. Define

$$x_{ij} = q_j^{*-1}(x_{ij}^* - \bar{x}_j^*), \tag{1}$$

where

$$q_j^* = \left\{ \sum_{i=1}^{N} (x_{ij}^* - \bar{x}_j^*)^2 \right\}^{\frac{1}{2}}.$$

Augment the matrix X with the constant vector of ones. Let Y be a $N \times 1$ random response vector for which each entry in Y follows the same distribution in the exponential family. Consider a single entry in the response vector, Y_i. The probability density function for Y_i can be expressed as

$$f(y; \theta, \phi) = \exp\left[\{yb(\theta) + c(\theta)\}/q(\phi) + d(y, \phi)\right],$$

where b, c, d, q are known functions. If $b(\theta) = \theta$, then denote θ as the natural parameter. Let the nuisance parameter ϕ be constant for all Y_i.

Given a set of p continuous explanatory variables and a constant, generalized linear regression utilizes the relationship,

$$g(\mu_i) = x_i'\beta = \eta_i, \tag{2}$$

satisfying:

(i) $\mu_i = E(Y_i)$;
(ii) g is a monotone, twice differentiable link function with $g^{-1} = h$;
(iii) x_i' is a $1\times(p+1)$ vector of continuous regressors including the constant;
(iv) β is the unknown parameter vector;
(v) the estimation of β does not depend on having an estimate of ϕ.

Standard theory of the generalized linear model is given by McCullagh & Nelder (1989) and Aitkin et al. (1989).

The method of scoring for maximum likelihood estimation can be expressed as

$$\begin{aligned}\hat\beta_t &= \hat\beta_{t-1} + (X'\hat K_{t-1}^{-1}X)^{-1}\left(\sum_{i=1}^{N} x_i \hat k_{ii}^{-1}\hat e_i \frac{\partial\eta_i}{\partial\mu_i}\right)_{t-1}\\ &= (X'\hat K_{t-1}^{-1}X)^{-1}X'\hat K_{t-1}^{-1}y_{t-1}^*,\end{aligned} \tag{3}$$

where the residual $\hat e_i = y_i - \hat\mu_i$, and $y_i^* = \eta_i + \hat e_i(\partial\eta_i/\partial\mu_i)$ are evaluated at $\hat\beta_{t-1}$ and

$$\hat K^{-1} = \mathrm{diag}\,[\{h'(\hat\eta_i)\}^2/\mathrm{var}\,(Y_i)].$$

Note that in the most general setting, $\hat K^{-1}$ and the working variable $\hat y^*$ must be updated at each iteration step of the parameter estimate, since they are a function of the iterated $\hat\eta_{t-1}$. Further, $\hat\beta$ is the least-squares estimate for normal data with $g(\mu) = \mu$.

2. Generalizing principal component regression

In a great deal of the literature addressing normal response data and the identity link function, principal component regression often utilizes explanatory variables that are standardized so that $X'X$ is proportional to the correlation matrix as given in (1). Certainly similar derivations to the methods presented in this paper can be made for uncentred or unscaled explanatory variables. In addition, if by design the explanatory variables are measured in the same units, then standardization may not be needed. It is not our purpose to discuss the controversy of alternate standardizations. For generalized linear regression, define the principal components for each observation as $Z = XM$, where the (i, j)th element of Z is the score of the jth principal component for the ith observation (Jolliffe, 1986, Ch. 8). Define M as the $(p+1)\times(p+1)$ matrix whose jth column is the jth eigenvector of the information matrix, Φ. Hence M is an orthogonal matrix and $M'\Phi M = \mathrm{diag}\,(\lambda_i) = \Lambda$, where λ_i are the eigenvalues of information.

Since $X\beta = XMM'\beta$, define $\alpha = M'\beta$. Hence (2) can be rewritten as

$$\eta_i = z_i'\alpha, \tag{4}$$

where z_i' is a row vector of the Z matrix. An orthogonally transformed full principal component model is given in (4). The reduced principal component model can also be useful and is defined as

$$\eta_{i,s} = z_{i,s}'\alpha_s,$$

where $z_{i,s}'$ is a subset row vector of Z_s, a $(p+1)\times s$ subset matrix of Z. The subset vector α_s is associated with high information.

3. Advantages of principal component regression

In the least-squares multiple regression framework for normal data with identity link function, the options for model building are well documented. Multicollinearity usually leads to an investigation of $X'X$. Variable deletion, Stein (1960) estimation, ridge regression (Hoerl & Kennard, 1970), and principal component regression (Webster, Gunst & Mason, 1974) are all tractable alternatives depending on the researcher's interests. However, more detrimental to the generalized linear model than multicollinearity among the columns of X is multicollinearity among the columns of $W = K^{-\frac{1}{2}}X$. Multicollinearity among the columns of W leads to ill-conditioning of the information matrix. Notice that multicollinearity among columns of X and W are equivalent if and only if $K^{-\frac{1}{2}} \simeq cI$, for a constant $c > 0$.

Define an ill-conditioned information matrix as one which has a small eigenvalue relative to the largest eigenvalue (Hartree, 1952, Ch. 8; Belsley, Kuh & Welsch, 1980). Let $\lambda_0, \ldots, \lambda_p$ be the eigenvalues of information in decreasing order. Define a condition index $CI_j = (\lambda_{\max}/\lambda_j)^{\frac{1}{2}}$, for $j = 0, \ldots, p$. Notice that information can be expressed as

$$\Phi = \sum_{j=0}^{p} \lambda_j m_j m_j'.$$

Hence at each iteration step of the method of scoring maximum likelihood

$$\hat{\beta}_t - \hat{\beta}_{t-1} = \sum_{j=0}^{p} \hat{\lambda}_j^{-1} \hat{m}_j \hat{m}_j' \sum_{i=1}^{N} x_i \hat{k}_{ii}^{-1} \hat{e}_i \frac{\partial \eta_i}{\partial \mu_i}, \tag{5}$$

where the right-hand side of (5) is evaluated at $t-1$. An iteration step can be highly unstable in the presence of multicollinearity among the columns of W. The deletion of a proper subset of principal components can help stabilize coefficients.

In addition, utilizing asymptotic properties, we have:

(i) as $\lambda_j \to 0$,

$$\sum_{j=0}^{p} \operatorname{var}(\hat{\beta}_j) = \operatorname{tr}(\Phi^{-1}) \to \infty;$$

(ii) for predictions outside the mainstream of the columns of W when combined with a small λ_j,

$$\operatorname{var}\{\hat{\mu}(x_0)\} \simeq \left(\frac{\partial \mu_0}{\partial \eta_0}\right)^2 \sum_{j=0}^{p} z_{j,0}^2 \lambda_j^{-1} \to \infty; \tag{6}$$

(iii) for the test

$$H_0: \beta = \beta_C, \quad H_1: \beta = \beta_F,$$

define the test statistic,

$$\chi^2 = \sum_{j=0}^{p} (\hat{\alpha}_{j,C} - \hat{\alpha}_{j,F})^2 \lambda_j,$$

where C and F denote the current and full model respectively such that $C \subset F$. Hence $\chi^2 \to 0$ if $(\hat{\alpha}_{j,C} - \hat{\alpha}_{j,F})^2 \lambda_j \to 0$ for some j.

One procedure to alleviate the above detriments is to delete the terms in the sum corresponding to small λ_j. In doing so, an iterative principal component estimator can be defined of the form

$$\hat{\beta}^{pc}_{t,s} = \hat{\beta}^{pc}_{t-1,s} + \sum_{j=0}^{s-1} \tilde{\lambda}_j^{-1} \tilde{m}_j \tilde{m}_j' \left(\sum_{i=1}^{N} x_i \tilde{k}_{ii}^{-1} \hat{e}_{i,s} \frac{\partial \eta_i}{\partial \mu_i} \right)_{t-1}, \tag{7}$$

where $\tilde{\lambda}_j$ and $\tilde{m}_j$ correspond to the eigenvalues and eigenvectors of the converged maximum likelihood estimate of information $\tilde{\Phi} = X'\tilde{K}^{-1}X$, if they exist. Define $\hat{e}_{i,s} = y_i - h(\hat{\eta}_{i,s})$. Note that $\lambda_s, \ldots, \lambda_p$ are usually the $r = p+1-s$ very small eigenvalues. Alternatively, (7) can be expressed in terms of $\hat{\alpha}^{pc}$

$$\begin{aligned} \hat{\alpha}^{pc}_{t,s} &= \hat{\alpha}^{pc}_{t-1,s} + \tilde{\Lambda}_s^{-1} \left(\sum_{i=1}^{N} z_{i,s} \tilde{k}_{ii}^{-1} \hat{e}_{i,s} \frac{\partial \eta_i}{\partial \mu_i} \right)_{t-1} \\ &= \tilde{\Lambda}_s^{-1} Z_s' \tilde{K}^{-1} y^{*pc}_{t-1}, \end{aligned} \tag{8}$$

where $z_{i,s}$ are the columns of Z_s' and $y_i^{*pc} = \hat{\eta}_{i,s} + \hat{e}_{i,s}\, \partial\eta_i/\partial\mu_i$ evaluated at iteration $t-1$ until specified convergence. Naturally, the principal components estimators can be constructed upon convergence of $\hat{\alpha}^{pc}_s$ as

$$\hat{\beta}^{pc}_s = \tilde{M}_s \hat{\alpha}^{pc}_s. \tag{9}$$

Using the full set of components, $\hat{\beta}^{pc}_s = \hat{\beta}$ as in (3).

4. Justification for maximum likelihood estimate in principal component estimators

A justification for using maximum likelihood $\tilde{K}$ as an estimate of K_t in principal component estimation follows from using a variance argument. Consider a first-order Taylor series expansion of an estimate of a diagonal element in $\tilde{K}^{-1}$, denoted by $\tilde{k}^{-1}$. Let var $(Y) = t(\eta)$. The subscript i is suppressed in (10)-(12). We have

$$\tilde{k}^{-1} \simeq \frac{\{h'(\eta)\}^2}{t(\eta)} + \frac{2t(\eta)h''(\eta)h'(\eta) - \{h'(\eta)\}^2 t'(\eta)}{\{t(\eta)\}^2} (\hat{\eta} - \eta). \tag{10}$$

Therefore

$$\operatorname{var}(\tilde{k}^{-1}) \simeq \frac{[2t(\eta)h''(\eta)h'(\eta) - \{h'(\eta)\}^2 t'(\eta)]^2}{\{t(\eta)\}^4} \sum_{j=0}^{p} z_j^2 \tilde{\lambda}_j^{-1}, \tag{11}$$

from equations (10), (3) and (4). Thus

$$\operatorname{var}(\tilde{k}^{-1}) = C^*(\eta) \sum_{j=0}^{p} z_j^2 \tilde{\lambda}_j^{-1}. \tag{12}$$

The quantity $C^*(\eta)$ is a constant. The variance given in (12) will not be as affected for observations in the original data set as for new observations outside the mainstream of collinearity in W. In general, original data cannot deviate much in the z_j direction corresponding to relatively small λ_j; hence $\tilde{K}^{-1}$ will estimate K^{-1} relatively well.

5. Generalization of Schaefer's logistic estimator

Schaefer (1986) presented a principal component estimator for logistic regression with $r=1$. The logistic principal component estimator is a one-step adjustment to the maximum likelihood estimator. Let $\hat{\beta}$ denote the maximum likelihood estimator upon convergence of method of scoring, if it exists. In experimental situations where the maximum likelihood parameter estimates are highly unstable and do not converge, then it is perverse to apply a one-step principal component estimator. In such a setting, perhaps the iterative principal component approach can be used as a resort to obtain converged parameter estimates.

Define the Moore–Penrose generalized inverse

$$(X'\tilde{K}^{-1}X)_s^+ = \sum_{j=0}^{s-1} \tilde{\lambda}_j^{-1} \tilde{m}_j \tilde{m}_j'.$$

In following an approach similar to that of Schaefer's, a one-step principal component estimator can be expressed as

$$b_s^{pc} = (X'\tilde{K}^{-1}X)_s^+ X'\tilde{K}^{-1}X\hat{\beta}. \tag{13}$$

Notice that if the iterative principal component technique, given in (7)–(9), uses starting values of zero, then we can rewrite the iterative scheme as follows

$$\hat{\beta}_s^{pc} = \sum_{t=1}^{T} \left(\sum_{j=0}^{s-1} \tilde{\lambda}_j^{-1} \tilde{m}_j \tilde{m}_j' \sum_{i=1}^{N} x_i \tilde{k}_{ii}^{-1} \hat{e}_{i,s} \frac{\partial \eta_i}{\partial \mu_i} \right)_t, \tag{14}$$

where T denotes the iteration of convergence. Furthermore, (13) and (14) are identical except that the maximum likelihood estimate $\hat{\beta}$ in (13) utilizes the residual $\hat{e}_i$ rather than $\hat{e}_{i,s}$. However, based on the variance result in (6), the residual is well behaved for the original data even with multicollinearity in W. Given a principal component estimator in (13) or (14), standard approaches can be taken to obtain uncentred and unscaled estimated principal component regression coefficients.

6. Properties of principal component estimators

Consider expressing the principal component generalized linear regression as follows

$$g(\mu) = (Z_s Z_r) \begin{pmatrix} \alpha_s \\ \alpha_r \end{pmatrix},$$

where

$$(Z_s Z_r)'(Z_s Z_r) = \begin{pmatrix} \Lambda_s & 0 \\ 0 & \Lambda_r \end{pmatrix}$$

and $s+r=p+1$. Note that $M=(M_s M_r)$ are the full set of eigenvectors of information. Hence an asymptotic reduction of variance in the principal component estimates can be expressed as

$$\text{var}\,(\hat{\beta}_s^{pc}) = \text{var}\,(b_s^{pc}) = \text{var}\,(\hat{\beta}) - M_r \Lambda_r^{-1} M_r'. \tag{15}$$

For an ill-conditioned information matrix, var $(\hat{\beta}_s^{pc})$ or var (b_s^{pc}) can be greatly reduced by deletion of principal components associated with small eigenvalues. The estimated asymptotic covariance matrix for the centred and scaled principal component estimates,

$\hat{\beta}_s^{pc}$, is $\tilde{M}_s\tilde{\Lambda}_s^{-1}\tilde{M}_s'$. Using (1), the corresponding uncentred and unscaled estimated standard errors can be expressed as

$$\text{SE}(\hat{\beta}_{s,j}^{*pc}) = q_j^{*-1}\,\text{SE}(\hat{\beta}_{s,j}^{pc}) \quad (j=1,\ldots,p),$$

$$\text{SE}(\hat{\beta}_{s,0}^{*pc}) = \left\{\text{var}(\hat{\beta}_{s,0}^{pc}) + \sum_{j=1}^{p} (q_j^{*-1}\bar{x}_j^*)^2\,\text{var}(\hat{\beta}_{s,j}^{pc}) + 2\sum_{i<j}\sum_{j\neq 0} q_i^{*-1}q_j^{*-1}\bar{x}_i^*\bar{x}_j^*\,\text{cov}(\hat{\beta}_{s,i}^{pc}, \hat{\beta}_{s,j}^{pc}) - 2\sum_{i=1}^{p} q_j^{*-1}\bar{x}_j^*\,\text{cov}(\hat{\beta}_{s,0}^{pc}, \hat{\beta}_{s,j}^{pc})\right\}^{\frac{1}{2}}.$$

Furthermore, the asymptotic bias associated with b_s^{pc} can be quantified as

$$E(b_s^{pc}) = \beta - M_r\alpha_r,$$

which is minimal if $\alpha_r \doteq 0$. Hence an asymptotic mean squared error criterion for b_s^{pc} is written as

$$\text{tr}\{\text{MSE}(b_s^{pc})\} = \sum_{j=0}^{p}\sum_{a=0}^{s-1} m_{aj}^2\lambda_a^{-1} + \sum_{j=0}^{p}\left(\sum_{k=s}^{p}\alpha_k m_{jk}\right)^2.$$

Asymptotic arguments yield an approximate distribution for b_s^{pc},

$$b_s^{pc} \sim N(M_s\alpha_s, M_s\Lambda_s^{-1}M_s'),$$

neglecting the stochastic nature in which components are deleted.

7. Deletion of principal components

For normal response data with the identity link function, there is an assortment of rules for deleting the proper principal components, if any at all. Development of rules of deletion in the framework of the generalized linear model will parallel standard principal component regression results. Jolliffe (1986, Ch. 8) provides an excellent summary. The full set of components $\hat{\alpha} \sim N(\alpha, \Lambda^{-1})$. Consider the test H_0: $C\alpha = 0$, where C is a $q\times(p+1)$ matrix of constants. The corresponding test statistic can be constructed $\hat{\alpha}'C'(C\Lambda^{-1}C')^{-1}C\hat{\alpha} \sim \chi_q^2$. Hence the test statistic criterion for a single component simplifies to $\hat{\alpha}_j^2\lambda_j \sim \chi_1^2$. The above tests statistics are compared to the appropriate percentage point of the asymptotic chi-squared distribution. In practice λ_j is usually unknown; therefore, the test statistic

$$t_j^* = \hat{\alpha}_j\hat{\lambda}_j^{\frac{1}{2}} \tag{16}$$

can be applied to test a single component using a t distribution on $N-p-1$ degrees of freedom.

Jolliffe develops several strategies for the selection of principal component standard multiple regression. One such strategy is to simply delete all the components associated with small eigenvalues below a specified cut-off, perhaps 0·01. A different approach from deleting small eigenvalues is one which incorporates the t test given in (16). Hence a procedure could be used which deletes components based on their contribution to the regression via a t test. However, Jolliffe warns, for standard principal component regression, that usually more components will be retained than are really necessary if components are deleted in succession until a significant t-statistic is reached.

Hill, Fomby & Johnson (1977) consider a more sophisticated approach to deletion of components which we can extend into the generalized linear model. A weak criterion is one where the objective is to get b_s^{pc} close to β. That is b_s^{pc} is preferred over $\hat{\beta}$ if

$$\operatorname{tr}\{\text{MSE}(b_s^{pc})\} \leqslant \operatorname{tr}\{\text{MSE}(\hat{\beta})\}. \tag{17}$$

Notice that (17) is equivalent to

$$\left(\sum_{j=0}^{p}\sum_{a=s}^{p} m_{aj}^2 \lambda_j^{-1}\right) - \left\{\sum_{j=0}^{p}\left(\sum_{k=s}^{p} \alpha_k m_{jk}\right)^2\right\} \geqslant 0,$$

A stronger criterion from (17) (Hill et al., 1977) is more oriented toward prediction rather than estimation of the coefficients. The requirement, which is more difficult to apply in practice, is now $\text{MSE}(c'b_s^{pc}) \leqslant \text{MSE}(c'\hat{\beta})$, for all nonnull c of proper dimension.

8. Illustrative example

Beauchamp & Gehrs (1978) present data from an extensive study on the population of *Diaptomus Clavipes*, a Zooplankton. Effects of independent variables including water temperature, adult density and female size were analysed for roles regulating the populations. A weighted least-squares approach was taken by Beauchamp & Gehrs fitting a full second-order model with the identity link function. In this paper, a generalized linear model is used assuming a Poisson response and the natural log link function. The square root link and identity link were also observed and dismissed based on deviance.

Perhaps if Beauchamp & Gehrs (1977) took a generalized linear regression approach rather than a weighted least-squares approach, then the following model would have been chosen:

$$\log \lambda_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i}, \tag{18}$$

where λ_i is the expected average number of eggs on ith sampling date; X_{1i} is the average water temperate over a four weak period prior to ith date, X_{2i} is the log of the adult density on ith sampling date, and X_{3i} is the female size observed on ith sampling date.

The model in (18) is adequate for prediction based on analysis of deviance and will yield parameter estimates of the explanatory variables of interest. Principal component estimates are provided along with the maximum likelihood estimates in Table 1. Standard approaches have been taken to uncentre and unscale the parameter estimates. The estimates are quite stable in sign and magnitude across techniques. Table 1 also confirms that the residuals $\hat{e}_i$ and $\hat{e}_{i,s}$, in (13) and (14), are behaving similarly. The iterative and one-step principal component estimates are identical, except for a slight difference in the intercept. All methods in this example converged in 6 iterations. The corresponding

Table 1. *Maximum likelihood,* ML, *and principal component parameter,* PC, *estimation for model* (18). *Link,* log; *error, Poisson*; $N=25$, $p+1=4$

		Iterative		One-step	
	ML	PC (−1)	PC (−2)	PC (−1)	PC (−2)
Dev.	22·9913	23·0117	23·0138	23·0117	23·0138
$\hat{\beta}_0$	0·9588	0·7264	0·7517	0·7266	0·7517
$\hat{\beta}_1$	−0·0225	−0·0210	−0·0209	−0·0210	−0·0209
$\hat{\beta}_2$	−0·1217	−0·1241	−0·1265	−0·1241	−0·1265
$\hat{\beta}_3$	0·0261	0·0282	0·0280	0·0282	0·0280

Table 2. *Estimated asymptotic standard errors for model* (18)

CI		ML	PC (−1)	PC (−2)
1·0000	SE ($\hat{\beta}_0$)	1·7522	0·6523	0·3373
14·1439	SE ($\hat{\beta}_1$)	0·0110	0·0036	0·0030
136·8008	SE ($\hat{\beta}_2$)	0·0594	0·0569	0·0181
4029·7880	SE ($\hat{\beta}_3$)	0·0164	0·0067	0·0040

CI, condition indices.

uncentred and unscaled estimated asymptotic standard errors are given in Table 2 which also provides condition indices, demonstrating that model (18) has ill-conditioned information. This example provides a considerable reduction in asymptotic standard errors for the principal component techniques. An asymptotic reduction is guaranteed, with deletion of components, by (15).

9. Discussion

The iterative principal component estimator given in (7)–(9) and the one-step adjustment estimator given in (13) are completely consistent with Webster et al. (1974) and Jolliffe (1986). In the special case of normal data and the identity link function, we have

$$Y = X\beta + \varepsilon \sim N(\mu, \sigma^2 I), \quad \hat{\beta}_s^{pc} = b_s^{pc} = M_s(Z_s'Z_s)^{-1}Z_s'y.$$

As in the classical linear regression model, principal component estimation for generalized linear regression is not always the best choice for model building. Given specific theoretical models oriented toward parameter estimation, principal component estimation can yield desirable variance properties with minimal bias. In models aimed toward prediction, perhaps wary variable deletion to reduce the multicollinearity among the columns of W can satisfy the researcher's needs.

Acknowledgement

The authors would like to thank the referee, Sir David Cox and Dr Luis A. Escobar for their helpful suggestions.

References

AITKIN, M., ANDERSON, D., FRANCIS, B. & HINDE, J. (1989). *Statistical Modelling in GLIM.* Oxford: Clarendon Press.

BEAUCHAMP, J. J. & GEHRS, C. W. (1978). Use of regression analysis to evaluate environmental effects: exploring methods of analysis. In *Proceedings of the 1977 Department of Energy Statistical Symposium.* Compiled and edited by D. A. Gardiner and T. Truett. Oak Ridge, Tennessee: Oak Ridge National Laboratory Report CONF-771042.

BELSLEY, D. A., KUH, E. & WELSCH, R. E. (1980). *Regression Diagnostics: Influential Data and Sources of Collinearity.* New York: Wiley.

HARTREE, D. R. (1952). *Numerical Analysis.* Oxford University Press.

HILL, R. C., FOMBY, T. B. & JOHNSON, S. R. (1977). Component selection norms for principal component regression. *Comm. Statist.* A **6**, 309–34.

HOERL, A. E. & KENNARD, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.

JOLLIFFE, I. T. (1986). *Principal Component Analysis.* New York: Springer-Verlag.

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

NELDER, J. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. R. Statist. Soc.* A **135**, 370–84.

SCHAEFER, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. *J. Statist. Comput. Simul.* **25**, 75–91.

STEIN, C. M. (1960). Multiple regression. In *Contributions to Probability and Statistics*, Essays in Honor of Harold Hotelling, Ed. I. Olkin, pp. 424–43. Stanford Univ. Press.

WEBSTER, J. T., GUNST, R. F. & MASON, R. L. (1974). Latent root regression analysis. *Technometrics* **16**, 513–22.

[*Received March* 1989. *Revised September* 1989]