Biometrika (1975), 62, 2, *p*. 383
Printed in Great Britain

The effect of errors in the independent variables in linear regression

By R. B. DAVIES and B. HUTTON*
Applied Mathematics Division, D.S.I.R., Wellington

SUMMARY

Suppose that the independent variables in a linear regression are subject to error. This paper is concerned with the bias introduced into the least squares estimators by these errors, first when they are regarded as fixed and second when they are regarded as random. Simple criteria are introduced for deciding whether the bias is likely to be serious. The paper also considers the effect of the occasional large error in either the dependent or independent variables.

Some key words: Errors in data matrix; Least squares; Linear regression; Singularity of matrix.

1. INTRODUCTION

The usual multiple linear regression model in Scheffé's (1959, § 1·3) notation is

$$Y = X'\beta + \epsilon, \tag{1·1}$$

where Y is a vector of n observations, X is a known $m \times n$ matrix, β is a vector of m unknown parameters and ϵ is a vector of independent random variables with zero mean and variance, σ^2. In practice it may happen that X cannot be measured exactly, rather $W = X + \Delta$ is known, where Δ is the measurement error. Then the usual least squares estimators for β and σ^2 with W in place of X, namely

$$b = (WW')^{-1} WY, \tag{1.2}$$

$$\hat{\sigma}^2 = Y'\{I - W'(WW')^{-1} W\} Y/(n-m), \tag{1·3}$$

will, in general, be subject to bias and one will want to know if this bias is likely to be serious.

This paper is particularly concerned with the criteria for deciding, when the general order of magnitude of the components of Δ is known, whether significant bias is possible. One might expect that, if the data matrix, X, is close to one of less than full rank, then measurement errors would have a serious effect. Section 2 introduces a distance from singularity of the matrix W and suggests that if this is small the estimates will be meaningless. In § 3 a simple bound for the maximum bias that can be introduced into the estimate b in relation to its standard deviation is found. In § 4 it is supposed that the columns of Δ are themselves independent random vectors and the asymptotic distribution of the estimators found as the number of observations, n, tends to infinity. Again a simple formula is found for deciding whether the bias is significant when compared with the standard error. Some of the results obtained in this section are similar to the small-sample approximations obtained by Hodges & Moore (1972). In § 5 these results are compared with a generalization of some

* Now at Mathematics Institute, University of Warwick, Coventry.

results of Brownlee (1965, pp. 414–7), where the components of X are also random variables. In this situation exact results similar to the asymptotic ones of §4 may be obtained. Section 6 is concerned with the sensitivity of the estimates to an occasional large error in X or Y, in particular with deciding whether such an error can have a serious effect on the estimates and yet escape detection when the residuals are examined. Section 7 contains a suggested practical procedure and an example.

Throughout this paper the symbol $\|.\|$ will denote the L_2 norm of a vector and the Hilbert norm, $\sup \|Ax\|/\|x\|$, of a matrix, A. Also $|x|$ will denote the sum of the absolute values of the elements of a vector x. A limit in probability will be denoted by plim.

2. Distance from singularity

We assume here as elsewhere that $n > m$. If X has rank less than m then we will say that X is singular. In this case there is insufficient information for estimating β. One might ask whether it is possible for X to be singular even though the observed matrix W is nonsingular. In this instance the estimates would be meaningless. However even if the matrix X was only nearly singular one would expect the estimates to be unreliable.

We suppose that in each element of the ith row of X a measurement error with an absolute value of about r_i might be expected. In this section we will not define r_i any more precisely. Let $R = \operatorname{diag}(r_1, \ldots, r_m)$. Then we will say that the distance of the matrix W from singularity is given by

$$\rho(W) = \inf[\{\operatorname{tr}(DD')\}^{\frac{1}{2}}: W + RD \text{ is singular}], \tag{2·1}$$

where D is an $m \times n$ matrix. If $\rho(W)$ is large compared with $(m_1 n)^{\frac{1}{2}}$, where m_1 is the number of rows of X subject to error, then X cannot be singular. However if $\rho(W)$ is small compared with $(m_1 n)^{\frac{1}{2}}$ one might expect the estimates to be very inaccurate

An alternative expression for $\rho(W)$ is given by the following theorem.

Theorem 2·1. *If $\rho(W)$ is defined as in* (2·1), *WW' is nonsingular and R is diagonal, then*

$$\rho(W) = \|R(WW')^{-1}R\|^{-\frac{1}{2}}. \tag{2·2}$$

Further if R is nonsingular, $\rho(W)$ is equal to the square root of the minimum eigenvalue of $R^{-1}WW'R^{-1}$.

Proof. Suppose that R is nonsingular. Apply a singular value decomposition to $R^{-1}W$, namely

$$R^{-1}W = Q[\operatorname{diag}(\lambda_1, \ldots, \lambda_m) : O]P,$$

where Q and P are each orthogonal matrices and λ_1 has minimum absolute value and is supposed to be positive. Then $R^{-1}W + D$ is singular with $\{\operatorname{tr}(DD')\}^{\frac{1}{2}} = \lambda_1$ if

$$D = -Q\left[\begin{array}{c:c} \lambda_1 & O \\ \hdashline O & O \end{array}\right]P.$$

Thus $\rho(W) \leqslant \lambda_1$, the square root of the minimum eigenvalue of $R^{-1}WW'R^{-1}$. If $W+RD$ is singular $(R^{-1}W+D)'x = 0$ for some $x \neq O$ and so $\|D'x\| = \|(R^{-1}W)'x\| \geqslant \lambda_1\|x\|$. Thus $\{\operatorname{tr}(DD')\}^{\frac{1}{2}} \geqslant \|D\| \geqslant \lambda_1$ and the result follows.

If R is singular, the result may be obtained either by an argument involving the taking of limits or by applying the theorem to $W_1\{I - W_2'(W_2W_2')^{-1}W_2\}$ and R_1 in place of W and R, where W_1 denotes the rows of W subject to error, W_2 the remaining rows, and R_1 the diagonal matrix formed from the nonzero diagonal elements of R.

Since the variances of the b_i are generally found in any analysis it is simpler to calculate

$$\rho_1(W) = [\operatorname{tr}\{R(WW')^{-1}R\}]^{-\frac{1}{2}} = \hat{\sigma}\{\Sigma r_i^2 \operatorname{var}(b_i)\}^{-\frac{1}{2}} \tag{2·3}$$

in place of $\rho(W)$, where var (b_i) denotes an estimated variance; this should be sufficiently accurate since $\rho_1(W) \leqslant \rho(W) \leqslant m_1^{\frac{1}{2}}\rho_1(W)$.

A generalization of Theorem 2·1 is that

$$\inf\{\operatorname{tr}(DD'):\ W+RD \text{ of rank } m-k\} = \sum_{i=1}^{k} 1/\mu_i, \tag{2·4}$$

where the μ_i are the eigenvalues in decreasing order of $R(WW')^{-1}R$. This result is included for completeness; however it does indicate that the situation in which $\rho_1 = (\Sigma\mu_i)^{-\frac{1}{2}}$ is substantially smaller than $\rho = \mu_1^{-\frac{1}{2}}$ corresponds to one in which W may be 'near' to a matrix of much less than full rank.

3. Maximum bias

In § 2 we let r_i $(i = 1, \ldots, m)$ denote the order of magnitude of the measurement errors in the elements of the data matrix. In this section we will make the more precise assumption that r_i is the square root of the average of the squares of the errors in the ith row of the data matrix, that is the square root of the ith diagonal element of $\Delta\Delta'/n$, and as before we let $R = \operatorname{diag}(r_1, \ldots, r_m)$. In practice, since the r_i will not be known exactly, they will be replaced by the experimenter's estimate of likely error.

Regard $a'b$ as an estimator of the linear combination $a'\beta$, with b as in (1·2) and a a vector; then the bias of $a'b$ might be considered important if it is not markedly less than the estimated standard deviation of $a'b$, namely $\hat{\sigma}\{a'(WW')^{-1}a\}^{\frac{1}{2}}$. Thus we may regard the bias introduced by the measurement error in X as unimportant if

$$c_1 = \sup_a |a'\{E(b)-\beta\}|/[\hat{\sigma}\{a'(WW')^{-1}a\}^{\frac{1}{2}}] \tag{3·1}$$

is somewhat less than one. Since

$$\sup_a\{(a'x)^2/a'Qa\} = x'Q^{-1}x, \tag{3·2}$$

when x is a vector and Q a positive-definite matrix, it may be shown that

$$\begin{aligned} c_1 &= \{\beta'\Delta W'(WW')^{-1}W\Delta'\beta\}^{\frac{1}{2}}/\hat{\sigma} \\ &\leqslant \|\Delta'\beta\|/\hat{\sigma} & (3·3) \\ &\leqslant n^{\frac{1}{2}}\,|R\beta|/\hat{\sigma}, & (3·4) \end{aligned}$$

where R is as defined at the beginning of this section. Expression (3·3) has been given by Swindel & Bower (1972), but since Δ is unknown cannot be applied directly. However (3·4) can be calculated if R is based on an estimate of the likely error. It will also be necessary to replace β by its estimate b. To guard against the random errors in b one might increase the magnitudes of its elements by their standard deviations when carrying out this substitution.

It is still possible that the measurement errors in X have rendered meaningless b and hence the estimate of (3·4). However

$$\begin{aligned} \big|\,|R\beta| - |R\,E(b)|\,\big| &\leqslant m_1^{\frac{1}{2}} \|R\{\beta - E(b)\}\| \\ &\leqslant m_1^{\frac{1}{2}} \|\Delta'\beta\|/\rho(W) \\ &\leqslant (m_1 n)^{\frac{1}{2}} |R\beta|/\rho(W), \end{aligned}$$

with $\rho(W)$ as defined by (2·1). Thus the bias in b will not significantly change the estimated value of (3·4) if $\rho(W)$ is at least several times the value of $(m_1 n)^{\frac{1}{2}}$. A further inequality leads to

$$c_1 \leqslant (m_1 n \beta' W W' \beta)^{\frac{1}{2}} / \{\hat{\sigma} \rho(W)\}. \tag{3·5}$$

We note that if a constant term is being fitted in the regression then $\beta' W W' \beta$ may be calculated after the row means have been subtracted from the rows of W and in this case this term is equal to the treatment sum of squares. Formula (3·5) may be rewritten

$$\|(W W')^{\frac{1}{2}} \{E(b) - \beta\}\| \leqslant \|(W W')^{\frac{1}{2}} \beta\| \, (m_1 n)^{\frac{1}{2}} / \rho(W); \tag{3·6}$$

that is the bias in the estimate of the 'normalized' form $(WW')^{\frac{1}{2}}\beta$ is less than the length $\|(WW')^{\frac{1}{2}}\beta\|$ multiplied by $(m_1 n)^{\frac{1}{2}}/\rho(W)$. Both (3·5) and (3·6) indicate the desirability of avoiding small values of $\rho(W)$.

The covariance matrix of b is $\sigma^2(WW')^{-1}$ and so bias in the estimation of this matrix enters only via the estimate of σ^2. In fact

$$E(\hat{\sigma}^2) = \sigma^2 + \beta' \Delta \{I - W'(W W')^{-1} W\} \Delta' \beta / (n - m), \tag{3·7}$$

and thus $0 < E(\hat{\sigma}^2) - \sigma^2 < \|\Delta'\beta\|^2/(n-m)$. Therefore, not unexpectedly, $E(\hat{\sigma}^2) > \sigma^2$ but the difference will be very small compared with the standard deviation of $\hat{\sigma}^2$ if (3·3) is small.

4. Effect of random errors in the data matrix

The condition considered in the last section may be unnecessarily stringent; it assumes that errors have been arranged in the worst possible way. In this section it will be supposed that the columns of Δ are independently and identically distributed with zero mean and covariance matrix S and are independent of ϵ. There is little point in allowing for a nonzero mean in Δ since most linear regressions involve a constant term and a constant error in a row of X will merely result in a different constant term. In the present context it is convenient to let r_i denote the standard deviation of the elements of the ith row of Δ, that is the square root of the ith diagonal element of S. If S is diagonal, that is the elements in each column of Δ are uncorrelated, then $R = S^{\frac{1}{2}}$.

The results obtained in this section are very similar to those obtained by Hodges & Moore (1972). However rather than letting $S \to 0$ with fixed n we will consider the asymptotic performance of b and $\hat{\sigma}$ as the number of observations, n, tends to infinity. We will replace the symbols Y, ϵ, X, Δ, W by Y_n, ϵ_n, X_n, Δ_n, W_n to denote the vectors and matrices with n observations. Applying the law of large numbers to $W_n W_n'$ and $W_n Y_n$ we find the following Theorem.

THEOREM 4·1. *Suppose that $X_n X_n'/n \to A$ as $n \to \infty$, where A is a nonsingular matrix with finite elements. Let $b_n = (W_n W_n')^{-1} W_n Y_n$. Then $b_n \to \beta - (A+S)^{-1} S\beta$ in probability as $n \to \infty$.* Thus the estimate will tend to be $(A+S)^{-1}S\beta$ in error if n is large. Note, however, that if S

is diagonal and the design balanced, that is A diagonal, the major effect of the measurement error will be multiplicative errors in the estimates of the components of β; the contamination of one component of β by the other components will be of a lower order of magnitude.

If a norm, or pseudo-norm if S is singular, is defined by $\|\beta\|_S = \|S^{\frac{1}{2}}\beta\|$ and S is diagonal, then the asymptotic error, $(A+S)^{-1}S\beta$, may be related to the 'distance from singularity' defined in § 2 as follows:

$$\begin{aligned}\sup \|(A+S)^{-1}S\beta\|_S/\|\beta\|_S &= \|S^{\frac{1}{2}}(A+S)^{-1}S^{\frac{1}{2}}\| \\ &= \operatorname*{plim}_{n\to\infty} n/\rho_n^2,\end{aligned} \tag{4·1}$$

where $\rho_n = \|R(W_n W_n')^{-1}R\|^{-\frac{1}{2}}$, $R = S^{\frac{1}{2}}$. Thus substantial errors may be present in the estimates if (4·1) is not small compared with one.

A more precise statement of the asymptotic distribution of b is given by the following theorem.

THEOREM 4·2. *Suppose that the fourth moments of the elements of Δ are finite. Then under the conditions of Theorem 4·1, with $\mathscr{L}$ denoting probability distribution,*

$$\mathscr{L}[n^{\frac{1}{2}}\{b_n-(X_nX_n'+nS)^{-1}X_nX_n'\beta\}] \to N\{0, \sigma^2(A+S)^{-1}+(A+S)^{-1}T(A+S)^{-1}\},$$

where T is the limit as $n \to \infty$ of the covariance matrix of

$$n^{-\frac{1}{2}}(X_n\Delta_n'+\Delta_n\Delta_n'-nS-\Delta_nX_n'A^{-1}S)(A+S)^{-1}A\beta.$$

If the elements of Δ_n are jointly normally distributed then

$$\begin{aligned}T = (A+S)\beta'A(A+S)^{-1}S(A+S)^{-1}A\beta \\ +S\beta'A(A+S)^{-1}SA^{-1}S(A+S)^{-1}A\beta \\ -S(A+S)^{-1}A\beta\beta'A(A+S)^{-1}S.\end{aligned}$$

Proof. We have, after some manipulation, that

$$\begin{aligned}b_n-(X_nX_n'+nS)^{-1}X_nX_n'\beta &= (A+S)^{-1}(X_n+\Delta_n)\epsilon_n/n \\ &\quad -(A+S)^{-1}(X_n\Delta_n'+\Delta_n\Delta_n'-nS-\Delta_nX_nA^{-1}S)(A+S)^{-1} \\ &\quad \times A\beta/n+o(n^{-\frac{1}{2}}),\end{aligned}$$

where the final term indicates small order in probability. The first two terms have null correlation matrix and null mean vectors. The first term has covariance matrix $\sigma^2(A+S)^{-1}/n$ and normality results from the Central Limit Theorem. The covariance matrix of the second term, under the hypothesis of normality of Δ, may be obtained by the application of the following expressions:

$$E(Z\xi\xi'Z') = \|\xi\|^2 I,$$
$$E(Z\xi\eta'Z) = \eta\xi',$$
$$E\{(Z'Z-nI)\xi\xi'(Z'Z-nI)\} = n(\|\xi\|^2 I+\xi\xi'),$$

where Z denotes an $n \times m$ matrix of independent standard normal variables, ξ an m dimensional vector and η an n dimensional vector. This completes the proof.

We note that for S close to the null matrix the expression for the asymptotic covariance matrix of b_n can be written

$$\{(A+S)^{-1}(\sigma^2+\beta'S\beta)+O(S^2)\}/n. \tag{4·2}$$

The estimate (1·3) of σ^2 is examined by the following theorem which may be proved using the law of large numbers.

THEOREM 4·3. *As $n \to \infty$, we have that*

$$\operatorname{plim} \hat{\sigma}_n^2 = \sigma^2 + \beta' S \beta - \beta' S (A+S)^{-1} S \beta.$$

Note that for S close to the null matrix the $O(S)$ term is the same as that in the coefficient of $(A+S)^{-1}/n$ in (4·2) and so (Hodges & Moore, 1972) the estimate of the covariance matrix of b_n will tend not to be unduly unbiased.

For large n the bias of (1·2) will be close to $(A+S)^{-1} S\beta$. However the standard deviations of its components will be of order $n^{-\frac{1}{2}}$ and so for large n the bias overwhelms the random error. Hence when considering the bias of the components of b_n in relation to their standard deviations it is appropriate to let S tend to zero as n tends to infinity, in particular to suppose that the covariance matrix of the columns of Δ_n is $S_n = n^{-\frac{1}{2}} S_1$.

COROLLARY 4·4. *If $S_n = n^{-\frac{1}{2}} S_1$, then*

$$\mathscr{L}\{n^{\frac{1}{2}}(b_n - \beta)\} \to N(A^{-1} S_1 \beta, \sigma^2 A^{-1}). \tag{4·3}$$

Following the approach of § 3 we consider the bias of b_n as important if the ratio of the asymptotic expectation of $n^{\frac{1}{2}} a'(b_n - \beta)$ to the asymptotic standard deviation of $n^{\frac{1}{2}} a' b_n$ is not markedly less than one for every nonnull m dimensional vector a. The maximum value of this ratio is

$$c_2 = (\beta' S_1 A^{-1} S_1 \beta)^{\frac{1}{2}} / \sigma. \tag{4·4}$$

If S_n is diagonal and $R_n = S_n^{\frac{1}{2}}$ and $\rho_n = \|R_n (W_n W_n')^{-1} R_n\|^{-\frac{1}{2}}$ then

$$c_2 \leqslant \lim n \|R_n \beta\| / (\rho_n \sigma). \tag{4·5}$$

On the other hand, if S_n is not diagonal and R_n is the diagonal matrix composed of the square roots of the diagonal elements of S_n, we find

$$c_2 \leqslant \lim m_1^{\frac{1}{2}} n |R_n \beta| / (\rho_n \sigma).$$

Thus, on dropping the subscript n, if n is large and the errors in the data matrix can be considered to be independent random variables, then the bias in b is negligible if

$$n \|R\beta\| / \{\rho(W)\sigma\} \tag{4·6}$$

is small compared with one. If only the columns of Δ are considered to be independent, (4·6) should be replaced with

$$m_1^{\frac{1}{2}} n |R\beta| / \{\rho(W)\sigma\}. \tag{4·7}$$

As in § 3, it will be necessary to replace β by b when calculating (4·6) or (4·7). If $R_n = S_n^{\frac{1}{2}}$ it follows from (4·1) that as $n \to \infty$

$$\operatorname{plim} \|R_n(\beta - b_n)\| / \|R_n \beta\| \leqslant \operatorname{plim} (n/\rho_n^2),$$

and so the substitution of β by b in (4·6) will be acceptable if $\rho(W)$ is large compared with $n^{\frac{1}{2}}$. In the general case one can show that

$$\operatorname{plim} |R_n(\beta - b_n)| / |R_n \beta| \leqslant \operatorname{plim} (m_1 n/\rho_n^2),$$

and so the substitution in (4·7) will be acceptable if $\rho(W)$ is large compared with $(m_1 n)^{\frac{1}{2}}$.

We note that if $S = O(n^{-\frac{1}{2}})$ then $\hat{\sigma}_n^2$ will be subject to little relative error for large n. However smallness of c_2 in (4·4) does not guarantee that the bias of $\hat{\sigma}_n^2$ will be small compared with its standard deviation.

An alternative asymptotic approach which supposes that β is of order $n^{-\frac{1}{2}}$ with S remaining fixed leads to essentially the same results as far as checking for significant bias in b is concerned. However with this approach to the limit bias will be present in the estimate of σ^2.

5. Random X matrix

In this section we consider the model (1·1) but where the columns of X are supposed to be independent $N(0, \Sigma)$ random variables. The results found here represent a generalization of those given by Brownlee (1965, § 12·8) and are included here because exact results similar to the asymptotic ones of the last section can be found. For simplicity we have deleted the constant term from Brownlee's formulation.

We suppose that the ϵ and Δ are as in the last section, but in addition we suppose that they are normally distributed independently of the X matrix. Then the vector $(w', y)'$, where w denotes a column of W and y the corresponding element of Y, has a multivariate normal distribution with zero mean and covariance matrix

$$\begin{bmatrix} \Sigma+S & \Sigma\beta \\ \beta'\Sigma & \sigma^2+\beta'\Sigma\beta \end{bmatrix},$$

so that

$$E(y|W) = w'\beta - w'(\Sigma+S)^{-1}S\beta,$$

$$\operatorname{var}(y|W) = \sigma^2+\beta'S\beta-\beta'S(\Sigma+S)^{-1}S\beta.$$

It follows that the problem is equivalent to the one without error in the measurement of the independent variables with β, Σ and σ^2 replaced by

$$\beta_S = \beta-(\Sigma+S)^{-1}S\beta, \quad \Sigma_S = \Sigma+S,$$

$$\sigma_S^2 = \sigma^2+\beta'S\beta-\beta'S(\Sigma+S)^{-1}S\beta.$$

Hence the estimate (1·2) of β will have bias, conditional on W, of $-(\Sigma+S)^{-1}S\beta$, and variance, conditional on W, of $(WW')^{-1}\{\sigma^2+\beta'S\beta-\beta'S(\Sigma+S)^{-1}S\beta\}$. The estimate of the residual sum of squares will have expectation $\sigma^2+\beta'S\beta-\beta'S(\Sigma+S)^{-1}S\beta$. These results are to be compared with the corresponding results of § 4.

6. Sensitivity to occasional large errors

So far we have been looking at the type of errors arising from measurement; that is, in each element in a row of the W matrix one might expect the same order of magnitude of error. Unfortunately one might also expect occasional large errors and it seems important that the results of an analysis do not depend too much on any one of the observations. Let W_i denote the ith column of W, the observed data matrix, and $s_i = W_i'(WW')^{-1}W_i$, that is the ith diagonal element of $W'(WW')^{-1}W$.

Suppose either that $W = X$ and an error ξ is introduced into just one, say the ith, element of Y or that an error $-\xi/\beta_j$ is introduced into the (j, i)th element of X, with all the other elements of W being equal to the corresponding elements of X. In either case an error ξ is introduced into the ith regression equation and this, in turn, introduces an error $\xi(WW')^{-1}W_i$ into the estimate b. Following the approach of §§ 3 and 4 this may be considered important if $c_3 = s_i^{\frac{1}{2}}|\xi|/\sigma$ is not small compared with one.

Of course, if ξ is large enough the error will be detected when the residuals,

$$\{I - W'(WW')^{-1}W\}\,Y$$

are examined. We ask if it is possible for the error ξ to have a significant effect on the estimates and yet escape detection when the residuals are checked. We suppose that there are a moderate number of observations, so that σ can be estimated fairly accurately in spite of the occasional large error. The ith residual will be contaminated by an amount $(1-s_i)\,\xi$. Also this residual has variance $(1-s_i)\,\sigma^2$ (Behnken & Draper, 1972). Thus we may regard ξ as detectable if, say, $(1-s_i)\,\xi > 4(1-s_i)^{\frac{1}{2}}\,\sigma$. If a bound, $c_3 \leqslant 1$, is acceptable, that is an error of one standard deviation permissible in some linear combination of the estimates, we find s_i must be less than 0·06 for the error, ξ, to be either detectable, $\xi/\sigma > 4(1-s_i)^{-\frac{1}{2}}$, or negligible, $\xi/\sigma < s_i^{-\frac{1}{2}}$. This is a rather stringent condition since the diagonal elements $\{s_i\}$ sum to m, the number of unknown parameters, and hence have average value m/n which itself may be well over 0·06. If a more generous bound, $c_3 < 2$, is acceptable the bound on s_i rises to a more reasonable value of 0·2.

The work involved in finding the values of the s_i is comparable with that required for the estimates. However, it does seem worthwhile including the computation of the s_i as an option in a regression or analysis of variance program, so that the user can be warned if any of his observations have an excessive influence on the results.

7. Practical procedure and example

We have supposed that the coefficients x_{ij} in the linear regression model

$$Y_j = \sum_{i=1}^{m} \beta_i x_{ij} + \epsilon_j \quad (j = 1, \ldots, n)$$

are subject to measurement error; an error with a magnitude of about r_i being expected in the measurement of x_{ij} with m_1 of the r_i being nonzero. We suggested that the quantity introduced in § 2, $\rho_1 = \hat{\sigma}(\Sigma r_i^2 \operatorname{var} b_i)^{-\frac{1}{2}}$ be calculated. If ρ_1 is not somewhat larger $(m_1 n)^{\frac{1}{2}}$, or possibly $n^{\frac{1}{2}}$ if n is large, then at least some of the estimates b_i are likely to have little meaning. $n^{\frac{1}{2}}$ If this test is passed then $n\Sigma r_i|b_i|/\hat{\sigma}$ introduced in § 3 should be evaluated. If this quantity is markedly less than one then the errors in the data matrix can be ignored. If this test is failed and the measurement errors can be regarded as random and independent then one should calculate $n(\Sigma r_i^2 b_i^2)^{\frac{1}{2}}/(\rho_1\hat{\sigma})$ from § 4. If this is markedly less than one, then the effects of the errors are probably negligible, particularly if n is large. On the other hand if this term is larger than one the bias is likely to constitute a major part in the error in at least some of the estimates. One might estimate it from Theorem 4·1 in order to see which estimates are affected and how serious the bias is.

We also suggested in § 6 that the diagonal elements of $X'(XX')^{-1}X$ should be calculated in order to check whether any single observation has an undue effect on the estimates. In particular if any diagonal element is greater than 0·2 it is possible for a moderate error in the corresponding observation to affect the estimates significantly and yet go undetected when the residuals are checked.

As an example we apply formulae (2·3), (3·4), (3·5) and (4·6) to the regression problem of Hodges & Moore (1972). In that example $n = 30$, $m = 3$, $m_1 = 2$ and $\hat{\sigma} = 37$, the treatment sum of squares = 390 000, and the values of the estimates, their standard deviations and the magnitude of the errors in the coefficients are

b_i	−4·5	0·98	28·6
s.d. (b_i)	10·9	0·11	7·2 .
r_i	0	7·0	0·16

Thus, from (2·3), $\rho_1 = 27$, which is somewhat larger than $(nm_1)^{\frac{1}{2}} = 7{\cdot}7$. From (3·4) we find $c_1 \leqslant 1{\cdot}7$ which suggests that with the worst arrangement of errors significant bias may be present. With ρ replaced by ρ_1, (3·5) gives the less sharp bound $c_1 \leqslant 4{\cdot}9$. From (4·6), again with ρ replaced by ρ_1, we find $c_2 \leqslant 0{\cdot}25$ indicating a nearly negligible bias.

The authors wish to thank J.H. Maindonald of Victoria University of Wellington for suggesting this project.

References

Behnken, D. W. & Draper, N. R. (1972). Residuals and their variance patterns. *Technometrics* **14**, 101–11.

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*. New York: Wiley.

Hodges, S. D. & Moore, P. G. (1972). Data uncertainties and least squares regression. *Appl. Statist.* **21**, 185–95.

Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.

Swindel, B. F. & Bower, D. R. (1972). Rounding errors in the independent variables in a general linear model. *Technometrics* **14**, 215–8.

[*Received April* 1974. *Revised October* 1974]