Biometrika (1989), **76**, 3, *pp.* 622-3
Printed in Great Britain

Restricted unbiased iterative generalized least-squares estimation

BY HARVEY GOLDSTEIN
Department of Mathematics, Statistics and Computing, Institute of Education, University of London, London WC1H 0AL, U.K.

SUMMARY

It is shown that the iterative least-squares procedure for estimating the parameters in a general multilevel random coefficients linear model can be modified to produce unbiased estimates of the random parameters. In the multivariate normal case these are equivalent to restricted maximum likelihood estimates.

Some key words: Iterative generalized least-squares; Maximum likelihood; Multilevel model; Restricted generalized least-squares; Restricted maximum likelihood.

Goldstein (1986, Appen. 1) showed that the iterative generalized least-squares estimates used in the general multilevel random coefficient linear model are equivalent to maximum likelihood estimates under multivariate normality. Like the maximum likelihood estimates, the iterative generalized least-squares estimates are biased and the purpose of the present paper is to show how unbiased estimates can be obtained. Under the assumption of multivariate normality these are shown to be equivalent to restricted maximum likelihood estimates, so called because the likelihood is based upon a set of $n-p$ linearly independent error contrasts, where p is the number of coefficients in the linear model and n is the number of observations. By analogy the unbiased iterative generalized least-squares estimates will be referred to as restricted iterative generalized least-squares estimates. The notation is that used by Goldstein (1986).

The general model can be written

$$Y = X\beta + Ze, \quad E(\,) = 0, \quad E\{(Ze)(Ze)^{\mathrm{T}}\} = V,$$

where β is a vector of fixed coefficients and e is a vector of variables random at any level of the hierarchy. The matrices X and Z are the design matrices for the fixed and random variables in the model respectively, and V is the covariance matrix of the response vector Y. In the estimation of the random parameters of the multilevel model, that is the variances and covariances of the random variables, at each iteration we regress vech $\{(Y-X\hat{\beta})(Y-X\hat{\beta})^{\mathrm{T}}\}$ on the corresponding design matrix for the unknown random parameters (Goldstein, 1986).

When β is estimated using generalized-least squares with V known

$$E\{(Y-X\hat{\beta})(Y-X\hat{\beta})^{\mathrm{T}}\} = V - X(X^{\mathrm{T}}V^{-1}X)^{-1}X^{\mathrm{T}}, \tag{1}$$

where X is the design matrix for the fixed effects in the model, and is assumed to have full rank. In the general case where V is to be estimated, at each iteration we use

$$(Y-X\hat{\beta})(Y-X\hat{\beta})^{\mathrm{T}} + X(X^{\mathrm{T}}\hat{V}^{-1}X)^{-1}X^{\mathrm{T}} \tag{2}$$

to provide an updated estimate of V which is based upon the current value $\hat{V}$; the term $X(X^{\mathrm{T}}\hat{V}^{-1}X)^{-1}X^{\mathrm{T}}$ can be regarded as a bias correction term.

Thus we work with (2) rather than just the first term of (2) when estimating the random parameters. This involves adding the correction term to the calculated cross-product matrix for every highest level unit; all the relevant matrices are already available from the computations. Note that the matrix $(X^{\mathrm{T}}\hat{V}^{-1}X)^{-1}$ can be written in the form $U^{-1}(U^{-1})^{\mathrm{T}}$, where $(X^{\mathrm{T}}\hat{V}^{-1}X) = U^{\mathrm{T}}U$ and U is upper triangular. With $Z = XU^{-1}$, expression (2) becomes $\tilde{Y}\tilde{Y}^{\mathrm{T}} + ZZ^{\mathrm{T}}$, where $\tilde{Y} = Y - X\hat{\beta}$.

Hence, with minor modifications, we can apply computational procedures similar to those for ordinary iterative generalized least-squares.

We now show that the restricted iterative generalized least-squares estimates are equivalent to restricted maximum likelihood estimates when e has a multivariate normal distribution.

The log likelihood function which is minimized in restricted maximum likelihood is the full likelihood function plus half the term $-\log(|X^T V^{-1} X|)$, where X is assumed to have full rank (Harville, 1977). On differentiating the log likelihood this gives the extra term

$$\operatorname{tr}\left\{(X^T V^{-1} X)\frac{\partial (X^T V^{-1} X)^{-1}}{\partial \beta^*}\right\} = \operatorname{tr}\left\{-X^T \frac{\partial V^{-1}}{\partial \beta^*} X (X^T V^{-1} X)^{-1}\right\},$$

where β^* is the vector of random parameters. Following Appendix 1 of Goldstein (1986), in the generalized least-squares minimization we replace S by S^*, where

$$S^* = S + X(X^T \Omega^{-1} X)^{-1} X^T, \tag{3}$$

with $\Omega = E\{(Y - X\beta)(Y - X\beta)^T\}$. Thus, on differentiating and setting $V = \Omega$, the second term in (3) leads to an additional term in the generalized least-squares estimation equations

$$-\operatorname{tr}\left\{\frac{\partial V^{-1}}{\partial \beta} X (X^T V^{-1} X)^{-1} X^T\right\}.$$

Since $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, we have

$$-\log(|X^T V^{-1} X|) = -\operatorname{tr}\left\{\frac{\partial V^{-1}}{\partial \beta} X (X^T V^{-1} X)^{-1} X^T\right\}$$

and the estimation procedures are equivalent.

Finally, in the ordinary least-squares model, with $V = \sigma^2 I$, we see from (1) that

$$E\{\operatorname{tr}(\tilde{Y}\tilde{Y}^T)\} = \sigma^2[n - \operatorname{tr}\{X(X^T X)^{-1} X^T\}] = \sigma^2(n - p),$$

where $(X^T X)$ is of order p, and hence an unbiased estimate of σ^2 is given by the usual formula, $\hat{\sigma}^2 = \operatorname{tr}(\tilde{Y}\tilde{Y}^T)/(n - p)$. In the special case where X is the unit vector we obtain $\hat{\sigma}^2 = s^2/(n - 1)$, which is the usual formula for an unbiased estimate of a variance.

Acknowledgements

I am most grateful to Tony Bryk, Bob Prosser and John Rasbash for their helpful comments. This work was carried out as part of a research project supported by the Economic and Social Research Council (U.K.).

References

GOLDSTEIN, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* **73**, 43-56.

HARVILLE, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Assoc.* **72**, 320-37.

[*Received December* 1988. *Revised February* 1989]