Estimating variance components from unbalanced data: a review

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N. Y.

BU-453-M March, 1973

Abstract

This paper is mostly just Part III of the 1971 Biometrics review paper "Topics in variance component estimation" (revised to be self-contained), as invited for publication by the journal Economics and Mathematical Methods, Academy of Sciences of the U.S.S.R. The revision consists only of changing equation, section and table numbers so as to start at 1, and of deleting references to Parts I and II. A final section, added to update certain topics, is also available in mimeo form.

12. SOME RECENT WORK

The preceding review is adapted from Searle [1971b], which also reviews variance component estimation from balanced data. Additional details, including a catalogue of computing formulae for unbalanced data, are available in chapters 9, 10 and 11 of Searle [1971a]. Documentation of computer programs (in FORTRAN IV) for calculating these formulae is available in Searle and Corbeil [1973]. A brief description now follows of some of the developments that have occurred since these 1971 publications.

Insofar as established methods are concerned, one important development is that Henderson's Method 2 (see section 3 above) has been shown by Henderson et al [1973] to be uniquely defined. Thus the assertion of non-uniqueness made by Searle [1968, 1971a, 1971b] has been found in error. However, this does not affect the fact that, as established in Searle [1968], the method can be used only when the model contains no interactions between fixed and random effects. Indeed, it is this fact which provides the basis for establishing the uniqueness of the method. The importance of the uniqueness result is that in some instances it provides for mixed models having no interactions between fixed and random effects an estimation procedure that is easier to compute than the hitherto only available method, namely the fitting constants method. A general description of how to use Method 2 is given in Henderson et al [1973] and specific details for the 2-way crossed classification are given by Searle [1973].

On the topic of maximum likelihood estimation (section 8), Rudan and Searle [1971] give explicit formulae for deriving sampling variances of large-sample maximum likelihood estimators for the 3-way nested classification. They also suggest, in Rudan and Searle [1973], that for the 2-way crossed classification the inverse $\underline{V}^{-1}$ of the variance-covariance matrix cannot be obtained for use in equation (43).

Covariance models have also been considered. First, Searle and Rounsaville [1973] show that for the general class of quadratic estimators, covariance components can be estimated by applying the well-known formula $2\sigma_{xy} = \sigma^2_{x+y} - \sigma^2_x - \sigma^2_y$. This holds true for any estimation procedure that involves quadratic forms of the observations, including all cases of unbalanced data, mixed and random models. Variance component estimation in models that include covariates has also been considered. Nerlove [1971] suggests an empirical estimator which Mount and Searle [1972] show is biased. They also derive, using Henderson's Method 3, specific estimators for the 1-way classification model with unbalanced data, and for the 2-way crossed classification with one observation per cell, in both cases with multiple covariates included in the model. It is interesting that their results reduce to calculating sums of sums of squares of residuals.

Best quadratic estimators developed in Townsend [1968] have been summarized in Townsend and Searle [1971]. Generalizations are made by Rao [1970, 1971a,b and 1972] in procedures known as MINQUE and MIVQUE. In considering the same model as in section 8.2

$$\underline{y} = \underline{X}\underline{\beta} + \sum_{\theta=A}^{K+1} \underline{Z}_\theta \underline{u}_\theta$$

only with $\underline{u}_{K+1} = \underline{e}$ and $\underline{Z}_{K+1} = \underline{I}$, we write as do Swallow and Searle [1973],

$$\underline{V}_\theta = \underline{Z}_\theta \underline{Z}'_\theta$$

$$\underline{V}_u = \sum_{\theta=A}^{K+1} \underline{V}_\theta$$

and

$$\underline{V} = \mathrm{var}(\underline{y}) = \sum_{\theta=A}^{K+1} \sigma^2_\theta \underline{V}_\theta \quad .$$

[Rao uses the symbol $\underline{V}$ for $\underline{V}_u$ and $\underline{V}^*$ for $\underline{V}$, but the above notation is more compatible with the common usage of $\underline{V}$ for var($\underline{y}$).]. Rao then estimates a variance component by $\underline{y}'\underline{A}\underline{y}$ chosing $\underline{A}$, symmetric, so that the estimator is both unbiased and invariant to changes in $\underline{\beta}$. He suggests two different estimators. One minimizes the Euclidian norm $tr(\underline{V}_u\underline{A})^2$ and is called the MInimum Norm Quadratic Unbiased Estimator, or MINQUE. (See e.g. Rao [1971a, p. 268 and 1972 pp. 112-3].) Swallow and Searle [1973] have named this Basic MINQUE to distinguish it from what they call Alternative MINQUE, which minimizes $tr(\underline{V}\underline{A})^2$. This too is suggested by Rao [1971a, p. 268 and 1972, p. 113] although he does not use distinguishing names. The second estimator suggested by Rao [1972, pp. 447, 453] is one which has minimum variance, the MInimum Variance Quadratic Unbiased Estimator, MIVQUE, which is derived as $\underline{y}'\underline{A}\underline{y}$ by minimizing

$$var(\underline{y}'\underline{A}\underline{y}) = 2tr(\underline{V}\underline{A})^2 + \text{a term in } \underline{A} \text{ and kurtosis parameters.}$$

Under normality assumptions kurtosis parameters are zero and MIVQUE is then equivalent to alternative MINQUE. Rao [<u>loc cit</u>] shows that for

$$\underline{R} = \underline{V}^{-1}[\underline{I} - \underline{X}(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}]$$

$$\underline{S} = \{s_{\theta\theta'}\} = \{tr(\underline{V}_\theta\underline{R}\underline{V}_{\theta'}\underline{R})\} \qquad \text{for } \theta,\theta' = A,\ldots,K,K+1$$

and

$$\underline{u} = \{u_\theta\} = \{\underline{y}'\underline{R}\underline{V}_\theta\underline{R}\underline{y}\}$$

the vector of MIVQUE's under normality (alternative MINQUE's)

$$\hat{\underline{\sigma}} = \underline{S}^{-1}\underline{u} \quad .$$

The same procedure used with $\underset{\sim}{V}$ of $\underset{\sim}{R}$ replaced by $\underset{\sim}{V}_u$ gives the basic MINQUE's. This summary is given in Swallow and Searle [1973], whose main results are explicit expressions for these estimators and their sampling variances for unbalanced data in the 1-way classification, one model with non-zero μ and the other with $\mu \equiv 0$. In the latter case the MIVQUE procedure under normality gives, as it should, the BQUE (best quadratic unbiased estimation) results of Townsend and Searle [1971].

ACKNOWLEDGEMENT

Preparation of Searle [1971b] was partially supported by Research Grant GM-05900 from the National Institute of General Medical Science, Bethesda, Maryland, and preparing this supplement was supported by Grant GJ 31746 from the National Science Foundation, Washington, D. C.

Additional References

Henderson, C. R., Searle, S. R. and Schaeffer, L. R. [1973] The uniqueness of Method 2 of estimating variance components. Paper BU-437-M, Biometrics Unit, Cornell University, Ithaca, N. Y.

Mount, T. D. and Searle, S. R. [1972] Estimating variance components in co-variance models. Paper BU-403-M, Biometrics Unit, Cornell University, Ithaca, N. Y.

Nerlove, M. [1971] Further evidence on the estimation of dynamic economic relations from a time series of cross sections. Econometrica, 39, 359-382.

Rao, C. R. [1970] Estimation of heteroscedastic variances. J. Am. Stat. Ass., 65, 161-172.

Rao, C. R. [1971a] Estimation of variance and covariance components — MINQUE theory. J. Multivariate Analysis, 1, 257-275.

Rao, C. R. [1971b] Minimum variance quadratic unbiased estimation of variance components. J. Multivariate Analysis, 1, 445-456.

Rao, C. R. [1972] Estimation of variance and covariance components in linear models. J. Am. Stat. Ass., 67, 112-115

Rudan, J. W. and Searle, S. R. [1971] Large sample variances of maximum likelihood estimators of variance components in the 3-way nested classification. Biometrics, 27, 1087-1091.

Rudan, J. W. and Searle, S. R. [1973] Wanted: an inverse matrix. Communications in Statistics, II No. 2 (in press).

Searle, S. R. [1971a] Linear Models, John Wiley and Sons, Inc., New York.

Searle, S. R. [1971b] Topics in variance component estimation. Biometrics, 27, 1-77.

Searle, S. R. [1973] Computing procedures for estimating variance components from unbalanced data in the 2-way crossed classification, no interaction mixed model. BU-450-M, Biometrics Unit, Cornell University, Ithaca, N. Y.

Searle, S. R. and Corbeil, R. R. [1973] Estimation of variance components and their sampling variances: documentation of computer routines. BU-442-M, Biometrics Unit, Cornell University, Ithaca, N. Y.

Searle, S. R. and Rounsaville, T. R. [1973] A remark on estimating covariance components. Biometrics (in press).

Swallow, W. H. and Searle, S. R. [1973] Minimum norm and minimum variance quadratic unbiased estimators of variance components from unbalanced data in the 1-way classification. Paper BU-447-M, Biometrics Unit, Cornell University, Ithaca, N. Y.

Townsend, E. C. and Searle, S. R. [1971] Best quadratic unbiased estimation of variance components from unbalanced data in the 1-way classification. Biometrics, 27, 643-657.