The Annals of Statistics
1976, Vol. 4, No. 4, 746–754

ASYMPTOTIC BEHAVIOR OF MINQUE-TYPE ESTIMATORS OF VARIANCE COMPONENTS

By K. G. Brown

University of Maine, Orono

The limiting distributions are obtained for two estimators of variance components: C. R. Rao's MINQUE, and an estimator produced by an iterative procedure, referred to as I-MINQUE. Limits are taken as the number of independent and identically distributed vector observations on the model assumed gets large. This approach provides the asymptotics of interest when an experiment with a large number of observations can be thought of as independent replications of a smaller experiment, a condition applying to some common experimental designs. The main result, from which the limiting distributions are obtained, is essentially an extension of a theorem due to T. W. Anderson (1973), who provides an application in time series.

Both estimators considered here are consistent, and require only modest assumptions on the sampled distribution. The I-MINQUE has a limiting distribution which is functionally independent of the choice of norm; when it is further assumed that the sampled distribution is normal, the estimator is asymptotically equivalent to the m.l.e. and asymptotically efficient. The MINQUE itself is less robust in the sense that these two properties do not always apply, the conditions being dependent on the choice of design.

1. Introduction and summary. To be considered is the estimation of variance components in a general analysis of variance model of the form

$$Y = X\beta + \varepsilon \tag{1.1}$$

where Y is a random n-vector, X is a given $n \times m$ matrix, β is an unknown m-vector of parameters. Further, the error term may have the linear structure

$$\varepsilon = U_1\xi_1 + \cdots + U_k\xi_k .$$

where U_i is a given $n \times k_i$ matrix, ξ_i is a k_i-vector of uncorrelated variables, each with mean 0, variance σ_i^2, and finite fourth moment; the components of ξ_i and ξ_j are uncorrelated, $i \neq j$. Then $V(\varepsilon) = \Sigma = \sum_{i=1}^k V_i\sigma_i^2$, where $V_i = U_iU_i'$. It is assumed that Σ is positive definite (p.d.) and that the V_i's are linearly independent matrices.

For Y normally distributed, T. W. Anderson (1969, 1970) has given the maximum likelihood estimator (m.l.e.) of σ_i^2 in a closely related model,[1] and obtained

Received January 1975; revised December 1975.

[1] In Anderson's formulation, the parameters corresponding to the variance components above are unrestricted in sign.

AMS 1970 *subject classifications*. Primary 62E20; Secondary 62J10.

Key words and phrases. Variance components, quadratic unbiased estimation, MINQUE, asymptotic distribution.

the limiting distribution as N, the number of independent and identically distributed vector observations on Y, approaches infinity; the estimator is consistent and asymptotically efficient. The asymptotic results also apply to the m.l.e. in the model given here; however, computation of the m.l.e. is further complicated in this model by the restricted parameter space, $\sigma_i^2 \geqq 0$. In a later paper, Anderson (1973) has proposed a simpler estimator which has the property of being equivalent (as $N \to \infty$) to the m.l.e. of the variance components.

In this paper is determined the limiting distribution (as $N \to \infty$) of the MINQUE (MInimum Norm Quadratic Unbiased Estimator[2]) of components due to C. R. Rao (1970, 1971 a, 1972). This approach provides the asymptotics of interest for experimental designs in which a large number of observations is conceptually equivalent to observing several independent observations from a reduced design. Applications include some common linear models, as illustrated in Section 6.

The MINQUE requires no distributional assumptions (beyond the existence of the first four moments) and the computations are tractable. Somewhat on the negative side, however, the estimates of the variance components may be negative and can depend on the choice of norm being minimized. One question to be considered is whether these two negative characteristics may vanish in the limit.

In the notation to follow, H is a matrix specified by the user which determines the specific Euclidean norm to be minimized. As will be apparent from its limiting distribution, the MINQUE is consistent for any choice of H, from which it follows that the probability of a negative estimate approaches zero under any norm. However, the limiting distribution of the estimator is not always functionally independent of the choice of norm; its limiting covariance matrix may depend on H.

The asymptotic dependence on the choice of norm motivates a related estimator with stronger asymtotic properties. To be referred to as the Iterated-MINQUE, it is obtained by repeating the estimation procedure using the MINQUE of the unknown covariance matrix as a weight matrix. This estimator, also consistent, has a limiting distribution which is functionally independent of the choice of H in *all* cases; it is identical to the MINQUE in those particular cases where the latter is invariant of the norm chosen; when the variables of the model are assumed to be normally distributed, it is asymptotically equivalent to the maximum likelihood estimator and asymptotically efficient.

In the next section, the MINQUE equations are derived and simplified for $N > 1$. A general theorem is then stated and proved, from which the limiting distributions of the MINQUE and I-MINQUE are obtained

2. The fundamental equations of MINQUE. The matrix determining the norm to be minimized in the MINQUE is $H = \sum_{i=1}^k V_i$ or $H = \sum_{i=1}^k V_i \alpha_i^2$,

[2] The E may also stand for Estimate or Estimation depending on the context.

chosen to be p.d., where $\alpha_1^2, \cdots, \alpha_k^2$ are a priori ratios of the unknown components. Let $P = X(X'H^{-1}X)^{-}X'H^{-1}$, $R = H^{-1}(I - P)$ and $\sigma = (\sigma_1^2, \cdots, \sigma_k^2)'$, the vector of unknown variance components. The fundamental equations for the estimation of σ (Rao, 1972) are

$$S\hat{\sigma}_M = u\,, \tag{2.1}$$

where $s_{ij} = \text{tr}\, RV_iRV_j$ (tr means trace), $u_i = \text{tr}\, RV_iRee'$, and $e = (I - P)Y$, $i, j = 1, \cdots, k$. (The form of u_i is equivalent to that of the reference.) The variance components are simultaneously estimable by MINQUE if and only if the equations have a unique solution.

To consider the asymptotic behavior of the estimate of σ, it will be necessary to develop the equations when there is more than one observation on the random vector Y. Let $Y_1, \cdots, Y_N$ be independent and identically distributed (i.i.d.) as Y of model (1.1). The model, based on N vectors, becomes

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon} \tag{2.2}$$

where $\tilde{Y} = (Y_1', \cdots, Y_N')'$, $X = (X' \vdots \cdots \vdots X')'$, $\tilde{\varepsilon} = (\varepsilon_1', \cdots, \varepsilon_N')'$, and $\varepsilon_i = Y_i - X\beta$. Then $V(\tilde{Y}) = \tilde{\Sigma}$, which is block diagonal with N blocks of Σ. Similarly, let $\tilde{H}$ be a block diagonal matrix with N blocks of H, and $\tilde{R} = \tilde{H}^{-1}(I - \tilde{P})$, where $\tilde{P} = \tilde{X}(\tilde{X}'\tilde{H}^{-1}\tilde{X})^{-}\tilde{X}'\tilde{H}^{-1}$. Further, let $\tilde{V}_i$ be block diagonal with N blocks of V_i, $i = 1, \cdots, k$. The equations for estimating σ based on $Y_1, \cdots, Y_N$ are

$$\tilde{S}\hat{\sigma}_M = \tilde{u}\,, \tag{2.3}$$

where $\tilde{s}_{ij} = \text{tr}\, \tilde{R}\tilde{V}_i\tilde{R}\tilde{V}_j$, $\tilde{u}_i = \text{tr}\, \tilde{R}\tilde{V}_i\tilde{R}\tilde{e}\tilde{e}'$, and $\tilde{e} = (I - \tilde{P})\tilde{Y}$, $i, j = 1, \cdots, k$.

The terms of (2.3) involve matrices with dimensions dependent on N; this can be remedied. For convenience, multiply both sides of (2.3) by N^{-1}, redefining $\tilde{s}_{ij}$ and $\tilde{u}_i$. A simple (but lengthy) reduction of terms gives

$$\tilde{s}_{ij} = \text{tr}\, H^{-1}V_iH^{-1}V_j + N^{-1}\,\text{tr}\,(RV_iRV_j - H^{-1}V_iH^{-1}V_j) \tag{2.4}$$

and $\tilde{u}_i = \text{tr}\, H^{-1}V_iH^{-1}N^{-1}\sum_{i=1}^{N} e_ie_i' + \text{tr}\,(RV_iR - H^{-1}V_iH^{-1})\bar{e}\bar{e}'$, where e_i is the ith n-subvector of $\tilde{e}$ (i.e., $\tilde{e} = (e_1', \cdots, e_N')'$) and $\bar{e} = N^{-1}\sum_{i=1}^{N} e_i$.

In the next two sections the main result of the paper is stated and proved. We then return to use (2.3) in applying the result to the MINQUE and I-MINQUE of σ.

3. Main result. A class of equations somewhat more general than those of (2.3) will be considered. Of interest is the asymptotic behavior of the solution of an arbitrary member of the class. First, however, some notational preliminaries will be dismissed.

If A is a finite dimensional matrix dependent on $Y_1, \cdots, Y_N$, then $A = o_p(N^{-k})$ will mean $N^ka_{ij} \to_p 0$ as $N \to \infty$ for all i, j; $A = O_p(N^{-k})$ will indicate that N^ka_{ij} is bounded in probability as $N \to \infty$. Throughout, limits will be taken as $N \to \infty$. The notation $A \to_p B$, where B is a matrix (which will always be fixed), will mean $a_{ij} \to_p b_{ij}$, all i, j. All matrices are finite dimensional.

DEFINITION OF TERMS.

(i) W_0 is any fixed linear combination of $V_1, \cdots, V_k$ which is p.d. (e.g., $W_0 = H$ or $W_0 = \Sigma$).

(ii) W is any linear combination of $V_1, \cdots, V_k$, with coefficients possibly dependent on $Y_1, \cdots, Y_N$, which is p.d. and a consistent estimator of W_0.

(iii) $C = N^{-1} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i'$ and D is an $n \times n$ symmetric matrix, possibly dependent on $Y_1, \cdots, Y_N$, such that $C - D = o_p(N^{-\frac{1}{2}})$.

(iv) $R(W)$, $P(W)$, $e_i(W)$, and $\bar{e}(W)$ are defined as R, P, e_i, and $\bar{e}$, respectively, except with H replaced by W.

To be considered is an arbitrary equation of the form

$$[G(W) + K(W)]\hat{\sigma} = a(W, D) + b(W) , \tag{3.1}$$

where $g_{ij} = \operatorname{tr} W^{-1}V_iW^{-1}V_j$, $k_{ij} = o_p(N^{-\frac{1}{2}})$, $a_i = \operatorname{tr} W^{-1}V_iW^{-1}D$, and $b_i = o_p(N^{-\frac{1}{2}})$, $i, j = 1, \cdots, k$. As will be seen, asymptotic covariance matrix of the solution to (3.1) (suitably normalized) depends on the covariance matrix of the squares and cross-products of the elements of ε in (1.1); the form of the matrix will be briefly explained.

Let $C = (c_{ij})$ and define c to be the $n(n+1)/2$ vector of C's components, $c = (c_{11}, c_{22}, \cdots, c_{nn}, c_{12}, c_{13}, \cdots, c_{n-1,n})'$. We'll say that c is the *vector form* of the matrix C. Denote the covariance of c for the case $N = 1$ by $\Psi = (\psi_{ij,kl})$, where $\psi_{ij,kl} = \operatorname{Cov}(c_{ij}, c_{kl})$, $i \leqq j$, $k \leqq l$. To find the value of $\psi_{ij,kl}$, let $U = (U_1 \vdots \cdots \vdots U_k) = (u_{ij})$ and $\xi = (\xi_1', \cdots, \xi_k')' = (\xi_{(i)})$. The component s of ε in (1.1) is $\varepsilon_{(s)} = \sum_{t=1}^{p} u_{st}\xi_{(t)}$ (where $p = \sum_{i=1}^{k} k_i$), which can be substituted into $\psi_{ij,kl} = E(\varepsilon_{(i)}\varepsilon_{(j)}\varepsilon_{(k)}\varepsilon_{(l)}) - E(\varepsilon_{(i)}\varepsilon_{(j)})E(\varepsilon_{(k)}\varepsilon_{(l)})$. Imposing the condition[3] $E(\varepsilon_{(i)}^a \varepsilon_{(j)}^b \varepsilon_{(k)}^c \varepsilon_{(l)}^d) = E\varepsilon_{(i)}^a E\varepsilon_{(j)}^b E\varepsilon_{(k)}^c E\varepsilon_{(l)}^d$, where a, b, c, d are any nonnegative integers which sum to four, and simplifying gives $\psi_{ij,kl} = \phi_{ij,kl} + \lambda_{ij,kl}$, where $\phi_{ij,kl} = \sigma_{il}\sigma_{jk} + \sigma_{ik}\sigma_{jl}$ $(\Sigma = (\sigma_{ij}))$ and $\lambda_{ij,kl} = \sum_{t=1}^{p} u_{it}u_{kt}u_{lt}u_{jt}(\mu_{4(t)} - 3\mu_{2(t)}^2)$, with $\mu_{2(t)}$ and $\mu_{4(t)}$ being the second and fourth moments of $\xi_{(t)}$. In matrix notation, defining $\Phi(\Sigma) = (\phi_{ij,kl})$ and $\Lambda = (\lambda_{ij,kl})$ gives $\Psi = \Phi(\Sigma) + \Lambda$. The term $\mu_{4(t)} - 3\mu_{2(t)}^2$ is the numerator of the kurtosis of $\xi_{(t)}$ (kur $\xi_{(t)}$). Note that if kur $\xi_{(t)} = 0$ for all t, as when $\xi_{(t)}$'s are normally distributed, then $\Psi = \Phi(\Sigma)$.

THEOREM. *Let $Y_1, \cdots, Y_N$ be random vectors* i.i.d. *as Y of model* (1.1). *Let $\hat{\sigma}$ be a solution to* (3.1) *when such exists, and an arbitrary k-vector otherwise. Then $N^{\frac{1}{2}}(\hat{\sigma} - \sigma)$ has a limiting normal distribution with mean vector* 0 *and covariance matrix $M(W_0)\Psi M(W_0)'$, where $M(W_0) = [L'\Phi^{-1}(W_0)L]^{-1}L'\Phi^{-1}(W_0)$, and L is the matrix with* jth *column equal to the vector form of V_j, $j = 1, \cdots, k$.*

In some cases $M(W_0)$ can be expressed more simply. As is well known, when all the V_i commute in pairs, $(V_iV_j = V_jV_i)$, which includes most experimental design models with *equal* numbers in the subclasses (Graybill and Hultquist,

[3] The asymptotic results to follow only require the left-hand side of the equality to be finite. The condition is used here because it provides a very substantial simplification to the covariance matrix.

1961), there exists an orthogonal matrix P such that PV_iP' is diagonal, $i = 1, \cdots, k$. Even when the V_i do not commute, sometimes there exists a nonsingular matrix R for which each RV_iR' is diagonal (Hultquist and Atzinger, 1972). The MINQUE, to which the theorem is to be applied, is invariant under nonsingular transformations (Rao, 1971 a).

Premultiplying model (1.1) by a diagonalizing matrix makes the resultant V_i and W_0 matrices diagonal, simplifying $M(W_0)$. The new matrix L has all zero elements beyond the nth row, and the new $\Phi(W_0)$ is diagonal, making inversion easy. The diagonal elements of $\Phi(W_0)$ are $2w_1^2, 2w_2^2, \cdots, 2w_n^2, w_1w_2, \cdots, w_{n-1}w_n$, where $W_0 = \text{diag}(w_1, \cdots, w_n)$. Letting L_1 be the matrix consisting of the first n rows of L, its ith column is the vector of elements on the diagonal of V_i. The matrix $L'\Phi^{-1}(W_0)L$ has ijth element $\frac{1}{2}\sum_{s=1}^n v_{is}w_s^{-2}v_{js}$, where v_{is} is element s on the diagonal of V_i, and $L'\Phi^{-1}(W_0)$ has ijth element $\frac{1}{2}v_{ij}w_j^{-2}$, $j = 1, \cdots, n$ and $0, j > n$.

4. Proof of theorem. A lemma will be established first, regarding the estimation of β in model (2.2) (assume for now that X of (1.1) is of full column rank). An estimate using W as a weight matrix is the solution to

$$X'W^{-1}X\hat{\beta}(W) = X'W^{-1}\bar{Y}\,, \tag{4.1}$$

where $\bar{Y} = N^{-1}\sum_{i=1}^N Y_i$. If Σ were known, using it as the weight matrix would give the Gauss–Markov estimate of β.

LEMMA 1. *Let $Y_1, \cdots, Y_N$ be* i.i.d. *as Y of model* (1.1). *Then $N^{\frac{1}{2}}(\hat{\beta}(W) - \beta) \to_d N(0, J(W_0)\Sigma J(W_0)')$, where $J(W_0) = (X'W_0^{-1}X)^{-1}X'W_0^{-1}$.*

The proof is given in T. W. Anderson (1973) for the case $W_0 = \Sigma$; the extension follows in the same manner.

When Y of (1.1) is normally distributed (or, more weakly, when kur $\xi_{(t)} = 0$, all t), then $\text{Cov}(c) = N^{-1}\Phi(\Sigma)$, and it can be shown (T. W. Anderson (1969)), that

$$L'\Phi^{-1}(\Sigma)c \equiv \tfrac{1}{2}a(\Sigma, C)\,. \tag{4.2}$$

Replacing Σ by W gives

$$L'\Phi^{-1}(W)c \equiv \tfrac{1}{2}a(W, C)\,. \tag{4.3}$$

From $EC = \Sigma = \sigma_1^2V_1 + \cdots + \sigma_k^2V_k$, it follows that

$$Ec = L\sigma\,. \tag{4.4}$$

A weighted estimate of σ, using $\Phi(W)$ as the weight matrix with model (4.4), is given by the solution to

$$L'\Phi^{-1}(W)L\hat{\sigma}_1 = L'\Phi^{-1}(W)c\,. \tag{4.5}$$

Equation (4.5) is like (4.1) with X, W, and $\bar{Y}$ replaced by L, $\Phi(W)$, and c. Since $\Phi(W) \to_p \Phi(W_0)$, Lemma 1 applies giving

$$N^{\frac{1}{2}}(\hat{\sigma}_1 - \sigma) \to_d N(0, M(W_0)\Psi M(W_0)')\,. \tag{4.6}$$

The jth column of L is the vector form of V_j, so the jth column of $L'\Phi^{-1}(W)L$ is, applying identity (4.3), just $\frac{1}{2}a(W, V_j)$. Observing that the jth column of $G(W)$ is $a(W, V_j)$ and applying identity (4.3) to the r.h.s. of (4.5), equation (4.5) can be written $G(W)\hat{\sigma}_1 = a(W, C)$. This proves the theorem for the case $K(W) = 0$, $D = C$, and $b(W) = 0$ (all a.s.). These restrictions will be lifted in turn.

Dropping the restriction that $K(W) = 0$ gives $[G(W) + K(W)]\hat{\sigma}_2 = a(W, C)$. Since $G(W) = L'\Phi^{-1}(W)L$ is p.d. (a.s.) for all N, and $K(W) = o_p(N^{-\frac{1}{2}})$, the probability that the matrix of coefficients is p.d. approaches 1 as $N \to \infty$. The difference $N^{\frac{1}{2}}(\hat{\sigma}_2 - \sigma) - N^{\frac{1}{2}}(\hat{\sigma}_1 - \sigma)$ is $N^{\frac{1}{2}}([G(W) + K(W)]^{-1}G(W) - I)\hat{\sigma}_1 = -[G(W) + K(W)]^{-1}N^{\frac{1}{2}}K(W)\hat{\sigma}_1 \to_p 0$, since $[G(W) + K(W)]^{-1} \to_p G^{-1}(W_0)$, $K(W) = o_p(N^{-\frac{1}{2}})$, and, by implication of (4.6), $\hat{\sigma}_1 \to_p \sigma$. Hence, $N^{\frac{1}{2}}(\hat{\sigma}_2 - \sigma)$ has the same limiting distribution as $N^{\frac{1}{2}}(\hat{\sigma}_1 - \sigma)$ given in (4.6).

The next step of generality gives $[G(W) + K(W)]\hat{\sigma}_3 = a(W, D)$. The difference $N^{\frac{1}{2}}(\hat{\sigma}_3 - \sigma) - N^{\frac{1}{2}}(\hat{\sigma}_2 - \sigma)$ is $[G(W) + K(W)]^{-1}N^{\frac{1}{2}}[a(W, D) - a(W, C)]$. $[G(W) + K(W)]^{-1} \to_p G^{-1}(W_0)$. For the ith term of the remaining expression, $\operatorname{tr} W^{-1}V_iW^{-1}N^{\frac{1}{2}}[D - C] \to_p 0$, since $C - D = o_p(N^{-\frac{1}{2}})$ and $W^{-1}V_iW^{-1} = O_p(1)$. $N^{\frac{1}{2}}(\hat{\sigma}_3 - \sigma)$ has the same limiting distribution as $N^{\frac{1}{2}}(\hat{\sigma}_2 - \sigma)$ given in (4.6).

Adding the term $b(W)$ gives the full equation of (3.1). Letting $\hat{\sigma}_4$ be a solution, $N^{\frac{1}{2}}(\hat{\sigma}_4 - \sigma) - N^{\frac{1}{2}}(\hat{\sigma}_3 - \sigma) = [G(W) + K(W)]^{-1}N^{\frac{1}{2}}b(W) \to_p 0$, since $[G(W) + K(W)]^{-1} \to_p G^{-1}(W_0)$ and $b(W) = o_p(N^{-\frac{1}{2}})$. Thus, $N^{\frac{1}{2}}(\hat{\sigma}_4 - \sigma)$ has the same limiting distribution as $N^{\frac{1}{2}}(\hat{\sigma}_3 - \sigma)$, which completes the proof. □

5. Application to the method of MINQUE. To show that the theorem applies to MINQUE, the following result is established.

LEMMA 2. (i) $\bar{e}(W)\bar{e}(W)' = o_p(N^{-\frac{1}{2}})$
(ii) $N^{-1}\sum_{i=1}^{N} e_i(W)e_i(W)' - C = o_p(N^{-\frac{1}{2}})$.

PROOF. The term $e_i(W)$ reduces to

$$e_i(W) = \varepsilon_i - P(W)\bar{\varepsilon}\,, \tag{5.1}$$

where $\bar{\varepsilon} = N^{-1}\sum_{i=1}^{N}\varepsilon_i$. Then $N^{\frac{1}{4}}\bar{e}(W) = (I - P(W))N^{\frac{1}{4}}\bar{\varepsilon}$. $N^{\frac{1}{2}}\bar{\varepsilon}$ has a limiting distribution by the C.L.T., so $N^{\frac{1}{4}}\bar{\varepsilon} \to_p 0$. Observing that $I - P(W) = O_p(1)$ gives $\bar{e}(W) = o_p(N^{-\frac{1}{4}})$, which proves (i). Multiplying the expression in part (ii) by $N^{\frac{1}{2}}$ and applying (5.1) gives

$$-P(W)N^{\frac{1}{4}}\bar{\varepsilon}N^{\frac{1}{4}}\bar{\varepsilon}' - N^{\frac{1}{4}}\bar{\varepsilon}N^{\frac{1}{4}}\bar{\varepsilon}'P(W)' + P(W)N^{\frac{1}{4}}\bar{\varepsilon}N^{\frac{1}{4}}\bar{\varepsilon}'P(W)' \to_p 0\,,$$

since $N^{\frac{1}{4}}\bar{\varepsilon} \to_p 0$ and $P(W) = O_p(1)$. □

The lemma shows that equation (2.3), with terms as given in (2.4), is of the form (3.1) with $W = H$ (a.s.), which leads to a corollary to the theorem.

COROLLARY 1. *Let* $\hat{\sigma}_M$ *be the* MINQUE *of variance components for model* (2.2). *Then* $N^{\frac{1}{2}}(\hat{\sigma}_M - \sigma)$ *has a limiting normal distribution with mean vector* 0 *and covariance matrix* $M(H)\Psi M(H)'$.

Observe that the estimator is consistent, independent of the choice of H, but the limiting distribution may depend on H through the covariance matrix. For those cases where the MINQUE is invariant of H, it follows that the limiting covariance also must be invariant. Unfortunately, the condition for invariance of H (Rao, 1971 b, Section 6) is very difficult to check, and does not hold in all cases. This difficulty is resolved by using the MINQUE of σ in an interative procedure, a notion which is heuristically appealing from another viewpoint as well.

Since the matrix H is essentially being used as an initial, or a priori, weight matrix in place of the unknown Σ, it seems reasonable to replace H by the estimated Σ (obtained by substituting for the unknown parameters their MINQUE), and re-estimate. In terms of the previous notation, if $\hat{\sigma}_M = (\hat{\sigma}_1^2, \cdots, \hat{\sigma}_k^2)'$ is the MINQUE based on $Y_1, \cdots, Y_N$, then set $W = \sum_{i=1}^k V_i \hat{\sigma}_i^2$ (if the r.h.s. is not p.d., then set $W = H$, which is equivalent to not iterating in that case). Since $\sum_{i=1}^k V_i \hat{\sigma}_i^2 \to_p \Sigma$ which is p.d., the probability $\sum_{i=1}^k V_i \hat{\sigma}_i^2$ is not p.d. approaches zero; the theorem can be applied with $W_0 = \Sigma$. The resultant estimator is the Iterated-MINQUE referred to previously.

COROLLARY 2. *Let $\hat{\sigma}_I$ be the* I-MINQUE *of variance components for model* (2.2). *Then $N^{\frac{1}{2}}(\hat{\sigma}_I - \sigma)$ has a limiting normal distribution with mean vector* 0 *and covariance matrix* $[L'\Phi^{-1}(\Sigma)L]^{-1} + M(\Sigma)\Lambda M(\Sigma)'$.

Observe that the initial starting matrix H is always immaterial to the limiting distribution. If Y is normally distributed, $\Lambda = 0$ and the covariance matrix reduces to $[L'\Phi^{-1}(\Sigma)L]^{-1}$; using (4.2) it may be written $[\frac{1}{2} \operatorname{tr} \Sigma^{-1}V_i\Sigma^{-1}V_j]^{-1}$. This matrix is the asymptotic covariance of the m.l.e. of σ, which is asymptotically efficient (T. W. Anderson (1973)). Hence, when Y of model (1.1) is normally distributed, *without further restrictions*, the I-MINQUE is asymptotically efficient.

While the case of normality is of particular interest, the reader is reminded that the method of MINQUE is "distribution-free," beyond requiring finite first four moments.

In practice it may be helpful to iterate an estimate of σ more than once, possibly until either the components change only slightly or a fixed number of iterations is reached; the asymptotic-distribution will remain as given in Corollary 2.

It is usually of interest to estimate β of (1.1), or a parametric function of β. If W is set equal to either the MINQUE or I-MINQUE of Σ (or another consistent estimator of Σ), then the probability the matrix of coefficients in expression (4.1) is p.d. approaches 1 as $N \to \infty$, when β is estimable. The limiting distribution is given by Lemma 1; the estimator is seen to be asymptotically equivalent to the Gauss–Markov estimator of β formed when Σ is known.

6. Applications. In the notation of this paper, the estimator of T. W. Anderson (1973) is found by initially letting the matrix W be an arbitrary p.d. matrix (e.g., $W = H$) and $D = N^{-1} \sum_{i=1}^N e_i(W)e_i(W)'$. The initial estimator, say $\hat{\sigma}_0$, is the solution to (3.1) with $K(W) \equiv 0$ and $b(W) \equiv 0$. Applying the theorem, $\hat{\sigma}_0 \to_p \sigma$. Resetting W to the estimator of Σ based on $\hat{\sigma}_0$, it follows that $W \to_p \Sigma$.

Solving (3.1) with the new value of W gives the estimator sought, with limiting distribution given by the theorem. Anderson's approach was to work with the maximum likelihood equations for a normal distribution, making $\Lambda = 0$ in the limiting covariance matrix. The results are applicable to estimation in the moving average stationary stochastic process of finite order.

A frequently occurring variance components model of interest is the nested classification. The two-fold nested design can be written $y_{ijk} = \mu + a_i + b_{ij} + c_{ijk}$, for $i = 1, \cdots, r$, $j = 1, \cdots, s$, $k = 1, \cdots, t$, where μ is fixed and the remaining terms on the right are independent variables with zero means and variances $\sigma_a^2, \sigma_b^2, \sigma_c^2$. For example, imagine a survey of a given type of deciduous tree where the observation of interest, y_{ijk}, is a measure of the nitrogen content of the kth leaf from the jth branch of the ith tree being sampled. The sample size could reasonably be made large by holding s and t fixed while increasing r. To apply the results of this paper, the model is put into the form of (1.1) with $r = 1$. Then for r an arbitrary positive integer, the model is conceptually equivalent to $N = r$ replications of model (1.1).

It can be verified directly that the ANOVA estimator, obtained by partitioning the sum of squares and equating to their expectation, is the MINQUE in the two-fold design with $r = 2$ (the smallest value for which the components are estimable) and $H = I$. The large sample approximations given here do not depend on the usual assumptions of normality.

To illustrate the computations, consider the simplest one-way classification $y_{ij} = \mu + a_i + b_{ij}$, $i = 1, \cdots, r$, $j = 1, 2$, with assumptions as above and variance components σ_a^2 and σ_b^2. The asymptotics will be as $r \to \infty$. The terms of model (1.1), constructed for $r = 1$, are $X = (1, 1)$, $U_1 = (1\ \ 1)'$, $U_2 = I_2$. Then V_1 is a 2×2 matrix of 1's, $V_2 = I_2$, and $\Sigma = V_1\sigma_a^2 + V_2\sigma_b^2$. Denoting the ith column of L by l_i, $l_1 = (1, 1, 1)'$ and $l_2 = (1, 1, 0)'$. Choosing $H = V_1 \cdot 0 + V_2 \cdot 1 = I_2$, $\Phi(H) = \operatorname{diag}(2, 2, 1)$. The limiting distribution of $r^{\frac{1}{2}}(\hat{\sigma}_M - \sigma)$ as $r \to \infty$ is normal with mean vector 0 and covariance matrix

$$M(I)\Phi(\Sigma)M(I)' + M(I)\Lambda M(I)' = \begin{pmatrix} 2\sigma_a^4 + \sigma_b^4 + 2\sigma_a^2\sigma_b^2 & -\sigma_b^4 \\ -\sigma_b^4 & 2\sigma_b^4 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2\lambda_a & 0 \\ 0 & \lambda_b \end{pmatrix},$$

where $\lambda_a = Ea_i^4 - 3\sigma_a^4$, $\lambda_b = Eb_{ij}^4 - 3\sigma_b^4$. The limiting distribution of the I-MINQUE is identical in this case.

An example of a mixed effects model is the balanced two-way classification with interaction given by $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \gamma_{ijk}$, for $i = 1, \cdots, r$, $j = 1, \cdots, s$, $k = 1, \cdots, t$, where τ_i is fixed and the remaining terms on the right are variables with zero means and variances σ_β^2, $\sigma_{\tau\beta}^2$, σ_γ^2, respectively. If the τ_i represent fixed treatment differences and the β_j are random block differences, the asymptotics of interest could be as the number of blocks (s) increases. (If a block consists of a single observation ($t = 1$), then the term γ_{ijk} cannot be distinguished from $(\tau\beta)_{ij}$ and is dropped from the model.) Model (1.1) is formed with $s = 1$, r and t arbitrary. The preceding results apply in which $N = s \to \infty$.

The method of MINQUE, with or without iteration, does not require that subclasses have equal numbers of observations. To apply the asymptotic results, however, it is necessary to be able to view a large experiment as replications of a smaller one. For unbalanced data, this is not apt to be the case if the imbalance is due to sporadic missing observations, but possibly applies in some situations of stratified sampling.

7. Concluding remarks. The limiting distribution of both the iterated and uniterated estimators depend on the numerator of kur $(\xi_{(t)})$, $t = 1, \cdots, p$. For a random variable Z with $E(Z) = 0$, kur $(Z) + 2 = \text{Var}(Z^2)/(\text{Var}(Z))^2$, indicating the kurtosis measures the ratio of the variance of the squared variable to the square of the variance of the variable. Recently Ali (1974) concluded that for a symmetrical density, kurtosis can be interpreted as the degree of "tailedness" relative to the normal distribution, but warns that it can be misleading as a measure of departure from normality.

Acknowledgment. The author would like to express his appreciation to the referee and editors for their helpful suggestions.

REFERENCES

[1] Ali, Mukhtar M. (1974). Stochastic ordering and kurtosis measure. *J. Amer. Statist. Assoc.* **69** 543–545.

[2] Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. In *Proceedings of the Second International Symposium on Multivariate Analysis*, 55–66 (P. R. Krishnaiah, ed.). Academic Press, New York.

[3] Anderson, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear conbinations of given matrices. In *Essays in Probability and Statistics*, 1–24 (R. C. Bose, et al., eds.). Univ. of North Carolina Press, Chapel Hill.

[4] Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

[5] Graybill, F. A. and Hultquist, R. A. (1961). Theorems concerning Eisenhart's model II. *Ann. Math. Statist.* **32** 261–269.

[6] Hultquist, R. A. and Atzinger, E. M. (1972). The mixed effects model and simultaneous diagonalization of symmetric matrices. *Ann. Math. Statist.* **43** 2024–2030.

[7] Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **65** 161–172.

[8] Rao, C. R. (1971a). Estimation of variance and covariance components—MINQUE theory. *J. Multivariate Anal.* **1** 257–275.

[9] Rao, C. R. (1971b). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Anal.* **1** 445–456.

[10] Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115.

Department of Mathematics
English-Math Bldg.
University of Maine
Orono, Maine 04473