*The Annals of Statistics*
1986, Vol. 14, No. 4, 1572–1582

# ASYMPTOTIC NORMALITY OF THE ANOVA ESTIMATES OF COMPONENTS OF VARIANCE IN THE NONNORMAL, UNBALANCED HIERARCHAL MIXED MODEL

BY PETER H. WESTFALL

*Texas Tech University*

Despite their lack of optimality in unbalanced normally distributed models, the ANOVA estimates of components of variance are convenient and widely used. The hierarchal (nested) design is well suited to this estimation scheme. In this paper the nonnormal, unbalanced hierarchal design is considered and mild conditions for a sequence of such designs are specified so that the vector of normalized ANOVA estimates converges to a multivariate normal distribution. The nested structure permits an expression of the estimates in terms of a sum of independent quadratic forms in mean zero random variables plus smaller order remainder, and a theorem of Whittle (1960) establishes the Liapounov criterion. Distinguishing features of this paper are the limit theory of nonnormal unbalanced models and the allowance that some variances other than the error variance may be null.

**1. Introduction and results.** The ANOVA estimates of components of variance are widely used. Their popularity is attributable to the simplicity of their formulation and computation, and to their wide exposure as estimates routinely computed by packaged computer programs. The ANOVA estimates are particularly appealing in the hierarchal (or nested) mixed model because the computations are simplified considerably and the estimates are uniquely defined.

Despite their nonoptimality in unbalanced normally distributed models (Olson, Seely, and Birkes (1976)), the ANOVA estimates have fared reasonably well in simulation studies when compared to estimates such as maximum likelihood, MINQUE, and I-MINQUE (Corbeil and Searle (1976), Swallow and Monahan (1984)). These simulations compare the estimators under normally distributed models for which the maximum likelihood and MINQUE-type estimates are best suited.

Asymptotic theory for variance component estimates has been limited largely to normally distributed or balanced models. Hartley and Rao (1967) and Miller (1977) have considered asymptotic theory for maximum likelihood estimates in the normally distributed case. While Miller's results encompass a wide variety of unbalanced mixed models, his theory is restricted to normally distributed models whose variance parameters lie in the interior of the parameter space. Brown (1976) has proven asymptotic distribution results for MINQUE and I-MINQUE estimates of variance components in nonnormal models; however, his theory only applies to models having a special balanced structure.

---

Received September 1984; revised February 1986.
*AMS* 1980 *subject classifications*. Primary 62E20; secondary 62J10.
*Key words and phrases*. Variance components, ANOVA estimates, hierarchal mixed model, asymptotic normality.

By way of motivation, consider the following example, which is similar to one given in Scheffé ((1959), pages 221–223): an experiment is tried in a factory with $t$ workers and a machine run by a single worker. Let $y_{ij}$ be the output of worker $i$ on day $j$, $1 \le i \le t$, $1 \le j \le n_i$. In the design phase of the experiment the $n_i$ may be bounded by five, since there are five working days in a week; however, due to scheduling problems, it is not possible to observe all workers exactly five days. Thus, the experiment is designed to be unbalanced.

Assume the model $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ with $\mu$ fixed and $\{\alpha_1, \dots, \alpha_t, \varepsilon_{11}, \dots, \varepsilon_{tn_t}\}$ independent mean zero random variables with $\mathrm{Var}(\alpha_i) \equiv \sigma_a^2$ and $\mathrm{Var}(\varepsilon_{ij}) \equiv \sigma_e^2$. The analysis of variance (ANOVA) technique may be used to obtain estimates for $\sigma_a^2$ and $\sigma_e^2$; it is natural to question whether these estimates have a limiting bivariate normal distribution under some conditions as the number of workers becomes large. The fact that the estimate are quadratic forms in an increasing number of mean zero random variables it not enough to guarantee normal convergence, since there are instances for which such quadratic forms have nonnormal limiting distributions (e.g., Fox and Taqqu (1985) and Varberg (1966)).

In this paper asymptotic distribution theory is given for the ANOVA estimates in unbalanced, nonnormal hierarchal models. The conditions for convergence are quite mild, not requiring that the data vector be composed of i.i.d. subvectors. The nonerror components of variance are allowed to be null, and it is seen that the error random components play a crucial role in the asymptotic theory. In order to prove the central limit theorem, arbitrary linear combinations of variance component estimates are represented approximately as summations of independent quadratic forms in mean zero random variables, and a theorem on the expectation of absolute moments of quadratic forms due to Whittle (1960) is employed to establish the Liapounov criterion.

**2. The model.** The model under consideration is

$$Y = U_0\alpha_0 + U_1\alpha_1 + \cdots + U_c\alpha_c, \tag{1}$$

where $Y$ has dimension $n \times 1$, $U_i$ is $n \times m(i)$, and $\alpha_i$ is $m(i) \times 1$. For $i = 0, \dots, c$ let $\alpha_i = (\alpha_{i1}, \dots, \alpha_{im(i)})'$. It is assumed throughout that

(1a) (fixed effects) $\alpha_0$ is a vector of fixed unknown constants;

(1b) (nonerror random effects) for $i = 1, \dots, c-1$, $\alpha_i$ is a vector of i.i.d. mean zero random variables such that $E(\alpha_{i1}^2) = \sigma_i^2 \ge 0$ and $E|\alpha_{i1}|^{4+2\delta} < \infty$ for some $\delta > 0$;

(1c) (error random effects) $\alpha_c$ is an $n \times 1$ vector of i.i.d. mean zero random variables such that $E(\alpha_{c1}^2) = \sigma_c^2 > 0$, $E|\alpha_{c1}|^{4+2\delta} < \infty$, and $\mathrm{Var}(\alpha_{c1}^2) > 0$;

(1d) (independence of factors) $\alpha_1, \dots, \alpha_c$ are independent random vectors;

(1e) (ANOVA design matrices) $U_i$ is a matrix having exactly 1 one and $m(i) - 1$ zeros in every row and no columns containing all zeros, $U_c$ is the identity matrix of dimension $n \times n$; and

(1f) (hierarchal property) every column in $U_i$ is the sum of some of the columns in $U_{i+1}$, for all $i < c$. Formally, letting $u(i, s)$ denote the $s$th column of $U_i$, we assume for every $(i, j)$ such that $0 \le i \le j \le c$ and every $s \in \{1, \dots, m(i)\}$ there exists a set of indices $T(s, i, j) = \{t_1, t_2, \dots\} \subset \{1, \dots, m(j)\}$ such that

$$u(i, s) = \sum_{t \in T(s, i, j)} u(j, t). \tag{2}$$

Note that (1e) and (1f) imply that for $0 \le i < j \le c$ the sets $T(s, i, j)$, for $1 \le s \le m(i)$, partition the set $\{1, \dots, m(j)\}$; in particular, this entails $m(i) \le m(j)$. Equality $m(i) = m(j)$ can obtain only if the design matrices $U_i$ and $U_j$ have identical (up to permutations) columns; this would mean that the effects $\alpha_i$ and $\alpha_j$ cannot be distinguished and that the ANOVA estimation scheme described below cannot yield a unique set of estimates. Hence, throughout the remainder of the paper it is assumed that $m(i) < m(j)$; i.e., for every $i < c$ there exists some $s \in \{1, \dots, m(i)\}$ such that $|T(s, i, i+1)| > 1$. Assumption (i) (to follow) is an asymptotic version of this requirement.

For example, the preceding one-way random effects model may be written as $Y = U_0\alpha_0 + U_1\alpha_1 + U_2\alpha_2$, where $Y = (y_{11}, \dots, y_{1n_1}, \dots, y_{m1}, \dots, y_{mn_m})'$, $n = \sum_{i=1}^m n_i$, $U_0 = (1, \dots, 1)' = e_n'$, $U_1$ is block diagonal with blocks $e_{n_i}$, and $U_2$ is the identity matrix of order $n$. Letting $i = 1$, $j = 2$ in (2) we have for $s = 1, \dots, m$, $u(1, s) = \sum_{t \in T(s,1,2)} u(2, t)$, where $u(1, s)$ is the $s$th column of $U_1$, $u(2, t)$ is the $t$th column of the identity matrix, and $T(s, 1, 2)$ is the set of indices

$$\left\{\sum_{i=1}^{s-1} n_i + 1, \sum_{i=1}^{s-1} n_i + 2, \dots, \sum_{i=1}^{s} n_i\right\} \quad \left(\text{define } \sum_{i=1}^{0} n_i = 0\right).$$

**3. Main results.** The model specified by (1a)–(1f) defines an experiment at any point in a conceptual sequence of designs. In what follows, a sequence of models satisfying (1a)–(1f) is defined for $n$ increasing to infinity. Thus, all model quantities are regarded as functions of $n$; however, for the sake of simplicity the dependence on $n$ may at times be omitted.

Define $\nu(i, s) = u'(i, s)u(i, s)$; i.e., $\nu(i, s)$ is the number of times factor $i$ is observed at level $s$. The sequence of hierarchal models is specified as follows:

ASSUMPTION (i). $n^{-1}m(i) \to a_i$ for $0 \le i \le c$, where $0 = a_0 < a_1 < \cdots < a_c = 1$.

ASSUMPTION (ii). There are universal constants $k$, $K$, with $0 < k < K < \infty$, such that $k \le n^{-1}\nu(0, s) \le K$, for all $n$ and $s = 1, \dots, m(0)$.

(Note that $\nu(0, s) \le \sum_{t=1}^{m(0)} \nu(0, t) = n$ and $\nu(0, s) \ge 1$; hence, all that is really required here is that $n^{-1}\nu(0, s) \ge k$ for all $n$.)

ASSUMPTION (iii). There is a universal constant $M$ such that $\nu(i, t) \le M$ for all $n$, $1 \le i \le c$ and $1 \le t \le m(i)$.

These three assumptions are quite mild: In many nested designs it is easy to imagine the model comes from a sequence satisfying Assumptions (i)–(iii); a particular application is considered in Section 5.

To define the estimates, let $P_i = U_i(U_i'U_i)^{-1}U_i'$ for $i = 0, \ldots, c$ and let $q_i = Y'(P_i - P_{i-1})Y$ for $i = 1, \ldots, c$. Then (e.g., Searle (1971), pages 443–445) the ANOVA estimates may be obtained by equating the $q_i$ to their expectations and solving for $\sigma_1^2, \ldots, \sigma_c^2$. Because $U_j'(P_i - P_{i-1}) = 0$ for $j < i$, we have

$$q_i = \sum_{j=1}^{c} \sum_{k=i}^{c} \alpha_j' U_j'(P_i - P_{i-1}) U_k \alpha_k \tag{3}$$

and

$$E(q_i) = \sigma_i^2(n - \operatorname{tr}(U_i'P_{i-1}U_i)) + \sum_{j=i+1}^{c} \sigma_j^2 \operatorname{tr}\big(U_j'(P_i - P_{i-1})U_j\big).$$

Writing $Q_n = (q_1, q_2, \ldots, q_c)'$ and $\Sigma = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2)'$, we have $E(n^{-1}Q_n) = F_n\Sigma$, where $F_{ii}(n) = 1 - n^{-1}\operatorname{tr}(U_i'P_{i-1}U_i)$; $F_{ij}(n) = n^{-1}\operatorname{tr}(U_j'(P_i - P_{i-1})U_j)$ for $i < j$, and $F_{ij}(n) = 0$ for $j < i$. If $F_n$ is invertible, then

$$\hat{\Sigma}_n = n^{-1}F_n^{-1}Q_n \tag{4}$$

is the vector of ANOVA estimates. A sufficient condition for invertibility of $F_n$ is that $C(U_0) \subsetneq C(U_1) \subsetneq \cdots \subsetneq C(U_c)$, where $C(\cdot)$ denotes vector space spanned by column vectors. This condition is implied by our assumption that $m(i) < m(i + 1)$ for $i = 0, \ldots, c - 1$, and is also guaranteed as $n \to \infty$ by Assumption (i) and the conditions (1e) and (1f).

We will study the asymptotic behavior of the vector $Q_n$, from which the results for $\hat{\Sigma}_n$ will readily follow. In particular, we have the following:

MAIN LEMMA. *Given the model definitions* (1a)–(1f) *and Assumptions* (i)–(iii), *and for any* $d \in \mathbb{R}^c - \{0\}$, *then* $[\operatorname{Var}^{-1/2}(d'Q_n)](d'Q_n - E(d'Q_n))$ *converges weakly to* $N(0,1)$.

Assumptions (i)–(iii) are not sufficient to insure that the covariance matrix of the normalized estimates $n^{1/2}(\hat{\Sigma}_n - \Sigma)$ converges. In order to assert convergence of this covariance matrix we assume the following:

ASSUMPTION (iv). $F_n \to F$, where $F$ is invertible.

ASSUMPTION (v). $n^{-1}\operatorname{Cov}(Q_n) \to \Delta$, where $\Delta$ is positive definite.

These assumptions may seem somewhat stringent at first glance; however, they are actually only mild strengthenings of Assumptions (i)–(iii). That these assumptions are consistent with Assumptions (i)–(iii) is demonstrated in Section 4, Lemmas 1 and 2, and is further illustrated in the application of Section 5.

Thus we have:

MAIN THEOREM. *Given the premises of the main lemma, and adding Assumptions* (iv) *and* (v), *we have that* $n^{1/2}(\hat{\Sigma}_n - \Sigma)$ *has a limiting c-variate normal distribution with mean* 0 *and covariance matrix* $(F^{-1})\Delta(F^{-1})'$.

**4. Proofs.** Asymptotic notation and conventions are as follows: If $\{a_n\}$ is a real sequence and $f(n)$ is a positive real function of $n$, then $a_n = O(f(n))$ if $f(n)^{-1}|a_n|$ is bounded away from $\infty$, and $a_n = o(f(n))$ if $f(n)^{-1}a_n \to 0$. If $\{x_n\}$ is a sequence of random variables, then $x_n = o_p(f(n))$ if $f(n)^{-1}x_n$ converges to 0 in probability. The following convenient notation is less familiar: $a_n = B(f(n))$ if $f(n)^{-1}a_n$ is bounded away from both 0 and $\infty$ for all $n$ large.

In order to establish consistency of Assumption (iv) with (i)–(iii), it is necessary that the elements of $F_n$ are bounded and that $F_n$ does not approach singularity for increasing $n$. These results are established in Lemma 1.

LEMMA 1. *Suppose Assumptions* (i)–(iii) *hold. Letting* $F_n$ *be defined as* $E(n^{-1}Q_n) = F_n\Sigma$, *we have*

(a) $F_{ij}(n) = O(1)$ *for all* $1 \le i,\ j \le c$, *and*
(b) $\det(F_n) = B(1)$.

PROOF. We first establish the order of magnitude of $|T(s, i, j)|$, where $|\cdot|$ denotes cardinality.

For $0 \le i < j \le c$ note that $\nu(i, s) = \sum_{t \in T(s, i, j)} \nu(j, t)$; since $\nu(0, s) = B(n)$ and $\nu(j, t) \le M$ for $j = 1, \dots, c$ this establishes

$$
\begin{aligned}
|T(s, i, j)| &= B(n), \qquad 1 \le s \le m(0), 0 = 1 < j \le c, \\
&\le M, \qquad 1 \le s \le m(i), 1 \le i < j \le c.
\end{aligned} \tag{5}
$$

We now establish order of magnitude for $n^{-1}\mathrm{tr}(U_j'P_iU_j)$. For $0 \le i < j \le c$ note that

$$
\mathrm{tr}(U_j'P_iU_j) = \mathrm{tr}\left\{(U_i'U_i)^{-1}U_i'U_jU_j'U_i\right\} = \sum_{s=1}^{m(i)} \nu^{-1}(i, s) \sum_{t \in T(s, i, j)} \nu^2(j, t).
$$

Thus we have

$$
\begin{aligned}
n^{-1}\mathrm{tr}(U_j'P_iU_j) &= O(1), \qquad 0 = i < j \le c, \\
&= B(1), \qquad 1 \le i < j \le c.
\end{aligned} \tag{6}
$$

Equation (6) establishes (a).

Because $F_n$ is upper triangular, we need only show $F_{ii}(n) = B(1)$ for $i = 1, \dots, c$ to prove (b). Note that

$$
F_{ii}(n) = 1 - n^{-1} \sum_{s=1}^{m(i-1)} \nu^{-1}(i - 1, s) \sum_{t \in T(s, i-1, i)} \nu^2(i, t).
$$

Since $\nu(i, t) \geq 1$ and

$$\sum_{t \in T(s, i-1, i)} \nu(i, t) = \nu(i-1, s),$$

we have

$$\begin{aligned}\nu^2(i-1, s) &= \sum_{t \in T(s, i-1, i)} \sum_{t' \in T(s, i-1, i)} \nu(i, t)\nu(i, t') \\ &\geq \sum_{t \in T(s, i-1, i)} \nu^2(i, t) + (|T(s, i-1, i)| - 1)\nu(i-1, s),\end{aligned}$$

implying that

$$\nu^{-1}(i-1, s) \sum_{t \in T(s, i-1, i)} \nu^2(i, t) \leq \nu(i-1, s) - |T(s, i-1, i)| + 1.$$

Hence,

$$\begin{aligned}F_{ii}(n) &\geq 1 - n^{-1} \sum_{s=1}^{m(i-1)} (\nu(i-1, s) - |T(s, i-1, i)| + 1) \\ &= 1 - n^{-1}(n - m(i) + m(i-1)) = n^{-1}m(i) - n^{-1}m(i-1).\end{aligned}$$

Part (b) of Lemma 1 now follows from Assumption (i). □

The consistency of Assumption (v) with (i)–(iii) is demonstrated in Lemma 2.

LEMMA 2. *Suppose Assumptions* (i)–(iii) *hold. Then any* $d \in \mathbb{R}^c - \{0\}$, $\mathrm{Var}(d'Q_n) = B(n)$.

PROOF. We first note the intimate relationship between the variance of a quadratic form and the sum of squares of the defining matrix in a special case relevant to this paper: Let $\alpha = (\alpha_1' : \alpha_2' : \cdots : \alpha_c')'$ and let $A$ be an $m \times m$ symmetric matrix composed of $m(i) \times m(j)$ submatrices $A_{ij}$ (here $m = \Sigma m(i)$). Letting $E(\alpha_{i1}^4) = \mu_i$, we have

$$\mathrm{Var}(\alpha' A \alpha) = 2\sum_{i=1}^{c} \sum_{j=1}^{c} \sigma_i^2 \sigma_j^2 \mathrm{tr}(A_{ij}A_{ji}) + \sum_{i=1}^{c} (\mu_i - 3\sigma_i^4)\mathrm{tr}(A_{ii}\mathrm{diag}(A_{ii})) \tag{7}$$

and

$$k\,\mathrm{tr}(A'A) \leq \mathrm{Var}(\alpha' A \alpha) \leq K\,\mathrm{tr}(A'A), \tag{8}$$

where $k = \min_{1 \leq i \leq c}\{2\sigma_i^4 \wedge (\mu_i - \sigma_i^4)\}$ and $K = \max_{1 \leq i \leq c}\{2\sigma_i^4 \vee (\mu_i - \sigma_i^4)\}$, and the notations $j \wedge k$ and $j \vee k$ refer to $\min(j, k)$ and $\max(j, k)$, respectively. Equation (7) is verified by direct computation, and (8) follows from (7). For the remainder we let $S(A) = \mathrm{tr}(A'A)$ for any (not necessarily square) matrix $A$.

Using (3) and interchanging the order of summation we have

$$d'Q_n = \sum_{j=1}^{c} \sum_{k=1}^{c} \alpha_j' U_j' \sum_{i=1}^{j \wedge k} d_i(P_i - P_{i-1})U_k \alpha_k. \tag{9}$$

To demonstrate that $\mathrm{Var}(d'Q_n) = O(n)$ it will suffice to show that $S(U_j'P_iU_k) = O(n)$ for $0 \le i \le j \wedge k$.

For $i = j \wedge k$ and $1 \le j \le k \le c$ we have

$$S(U_j'P_iU_k) = S(U_j'U_k) = \sum_{s=1}^{m(j)} \sum_{t \in T(s,j,k)} \nu^2(k,t) = O(n).$$

By symmetry the same holds for $1 \le k \le j \le c$.

Suppose that $i < j \wedge k$ and $1 \le j \le k \le c$. Manipulation of the trace operator and the Cauchy–Schwarz inequality gives

$$S(U_j'P_iU_k) \le S^{1/2}(U_j'P_iU_j)S^{1/2}(U_k'P_iU_k). \tag{10}$$

Now,

$$S(U_j'P_iU_j) = \sum_{s=1}^{m(i)} \nu^{-2}(i,s)\left\{\sum_{t \in T(s,i,j)} \nu^2(j,t)\right\}^2;$$

hence,

$$\begin{aligned} S(U_j'P_iU_j) &= O(1), \qquad 0 = i < j \le c, \\ &= B(n), \qquad 1 \le i < j \le c. \end{aligned} \tag{11}$$

By symmetry we have the same result for $k \le j$; combining (10) and (11) yields $\mathrm{Var}(d'Q_n) = O(n)$.

To complete the proof that $\mathrm{Var}(d'Q_n) = B(n)$ we need only show that $n^{-1}\mathrm{Var}(d'Q_n)$ is bounded away from 0 for $n$ large. From (9) obtain $d'Q_n = \alpha'G\alpha$, where $G$ is composed of submatrices $G_{jk} = U_j'\sum_{i=1}^{j\wedge k} d_i(P_i - P_{i-1})U_k$. Using an argument similar to that used to prove (8), it may be shown that $\mathrm{Var}(\alpha'G\alpha)$ is bounded below by a quantity that depends only on the lower right submatrix $G_{cc} = \sum_{i=1}^{c} d_i(P_i - P_{i-1})$. It then follows that

$$\begin{aligned} \mathrm{Var}(d'Q_n) &\ge \left(2\sigma_c^4 \wedge \left(\mu_c - \sigma_c^4\right)\right)\mathrm{tr}(G_{cc}G_{cc}) \\ &= \left(2\sigma_c^4 \wedge \left(\mu_c - \sigma_c^4\right)\right)\sum_{i=1}^{c} d_i^2\,\mathrm{tr}(P_i - P_{i-1}) \\ &= \left(2\sigma_c^4 \wedge \left(\mu_c - \sigma_c^4\right)\right)\sum_{i=1}^{c} d_i^2(m(i) - m(i-1)) = B(n); \end{aligned}$$

since at least one $d_i$ is nonzero, $\sigma_i^2 > 0$, $\mathrm{Var}(\sigma_c^2) > 0$, and $m(i) - m(i-1) = B(n)$. □

The crucial role of the error random effects in the asymptotic theory has just been established: Regardless of whether $\sigma_i^2 = 0$ (or for that matter, $\sigma_i^2 > 0$ and $\mathrm{Var}(\sigma_{i1}^2) = 0$ for $i = 1, \dots, c-1$), $\mathrm{Var}(d'Q_n) = B(n)$ due to the error effects. This fact is essential in the application of the central limit theorem.

PROOF OF MAIN LEMMA. For arbitrary $d \in \mathbb{R}^c - \{0\}$, Equation (3) gives

$$d'Q_n = \left[\sum_{i=2}^{c} d_i \sum_{j=1}^{c}\sum_{k=i}^{c} \alpha_j' U_j'(P_i - P_{i-1})U_k\alpha_k + d_1 \sum_{j=1}^{c}\sum_{k=1}^{c} \alpha_j' U_j' P_1 U_k \alpha_k\right]$$

$$-d_1 \sum_{j=1}^{c}\sum_{k=1}^{c} \alpha_j' U_j' P_0 U_k \alpha_k = \theta(n) - R(n) \quad \text{(say)}.$$

By (10) and (11) the variance of $R(n)$ is $O(1)$, while $\mathrm{Var}(d'Q_n)$ is $B(n)$ by Lemma 2. It follows that if $[\mathrm{Var}^{-1/2}\{\theta(n)\}]\{\theta(n) - E(\theta(n))\}$ converges weakly to $N(0,1)$ then so does $[\mathrm{Var}^{-1/2}\{d'Q_n\}]\{d'Q_n - E(d'Q_n)\}$. We now show that $\theta(n)$ may be expressed as a sum of independent random variables.

For $1 \le i \le j \wedge k$, note that

$$\alpha_j' U_j' P_i U_k \alpha_k = \sum_{s=1}^{m(i)} \sum_{t \in T(s,i,j)} \sum_{t' \in T(s,i,k)} \alpha_{jt}\alpha_{kt'}\nu^{-1}(i,s)\nu(j,t)\nu(k,t')$$

$$(12) \qquad = \sum_{r=1}^{m(1)} \sum_{s \in T(r,1,i)} \sum_{t \in T(s,i,j)} \sum_{t' \in T(s,i,k)} \alpha_{jt}\alpha_{kt'}\nu^{-1}(i,s)\nu(j,t)\nu(k,t').$$

Substituting (12) in the expression for $\theta(n)$ and interchanging the order of summation we have

$$\theta(n) = \sum_{r=1}^{m(1)} \Bigg\{ \sum_{i=2}^{c} d_i \sum_{j=i}^{c}\sum_{k=i}^{c} \sum_{s \in T(r,1,i)} \sum_{t \in T(s,i,j)} \times \sum_{t' \in T(s,i,k)} \alpha_{jt}\alpha_{kt'}\nu^{-1}(i,s)\nu(j,t)\nu(k,t')$$

$$- \sum_{i=2}^{c} d_i \sum_{j=i}^{c}\sum_{k=i}^{c} \sum_{s \in T(r,1,i-1)} \sum_{t \in T(s,i-1,j)} \times \sum_{t' \in T(s,i-1,k)} \alpha_{jt}\alpha_{kt'}\nu^{-1}(i-1,s)\nu(j,t)\nu(k,t')$$

$$+ d_1 \sum_{j=1}^{c}\sum_{k=1}^{c} \sum_{s \in T(r,1,1)} \sum_{t \in T(s,1,j)} \times \sum_{t' \in T(s,1,k)} \alpha_{jt}\alpha_{kt'}\nu^{-1}(1,s)\nu(j,t)\nu(k,t') \Bigg\}$$

$$= \sum_{r=1}^{m(1)} X(r).$$

Let $H(r) = \{\alpha_{kt} | \alpha_{kt}$ appears in the expression for $X(r)\}$, and let $H = \bigcup_{r=1}^{m(1)} H(r)$. Since $T(r,1,i)$ for $r = 1, \dots, m(1)$ partitions $(1, \dots, m(i))$ for all $i = 1, \dots, c$, it follows that $\{H(r)\}$ partitions $H$; hence, $X(1), \dots, X(m(1))$ are independent.

Since for $1 \le i \le j \le c$, $|T(r, i, j)| \le M$ and $1 \le \nu(i, j) \le M$, $X(r)$ is a quadratic form in a uniformly bounded number of independent mean zero random variables, with all coefficients uniformly bounded. Letting $X(r) = x_r' A(r) x_r$, it follows from Theorem 2 of Whittle (1960) that

$$E\left(|x_r' A(r) x_r - E(x_r' A(r) x_r)|^{2+\delta}\right) \le L(2+\delta) D[S(A(r))]^{1+\delta/2},$$

where $L$ is a function of $(2+\delta)$ only, and $D$ is a constant such that $E(|\alpha_{i1}|^{4+2\delta}) \le D$, $i = 1, \dots, c$. Since the dimension and coefficient of $A(r)$ are bounded uniformly in $r = 1, 2, \dots, m(1)$,

$$\sum_{r=1}^{m(1)} E\left(|X(r) - E(X(r))|^{2+\delta}\right) = O(n).$$

Since $\operatorname{Var}(\theta(n)) = B(n)$, the Liapounov criterion as applied to $X(1), \dots, X(m(1))$ follows, and $[\operatorname{Var}^{-1/2}(d'Q_n)](d'Q_n - E(d'Q_n))$ converges weakly to $N(0,1)$ for all $d \in \mathbb{R}^c - \{0\}$. □

PROOF OF MAIN THEOREM. Results to this point have not depended on Assumptions (iv) and (v). Invoking these assumptions and Slutsky's theorem yields that $\{d'(F^{-1})\Delta(F^{-1})'d\}^{-1/2}\{d'n^{1/2}(\hat{\Sigma}_n - \Sigma)\}$ converges weakly to $N(0,1)$ for all $d \in \mathbb{R}^c - \{0\}$, implying the results. □

**5. An application.** A simple example of an unbalanced hierarchal model is the one-way classification $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, for $i = 1, \dots, t$, $j = 1, \dots, n_i$, where $\mu$ is fixed and the remaining variables are independent with mean zero and variances $\operatorname{Var}(\alpha_i) = \sigma_a^2$ and $\operatorname{Var}(\varepsilon_{ij}) = \sigma_e^2$. We require that $\alpha_1 \dots, \alpha_t$ are i.i.d. with $E|\alpha_i|^{4+2\delta} < \infty$, $\varepsilon_{11}, \dots, \varepsilon_{tn_t}$ are i.i.d. with $E|\varepsilon_{ij}|^{4+2\delta} < \infty$, and $\operatorname{Var}(\varepsilon_{i1}^2) > 0$.

In order to satisfy Assumptions (i)–(v), it is sufficient that $n = \sum n_i \to \infty$ in such a way that $n_i \le M$, $n^{-1}t \to c \in (0,1)$, $n^{-1}\sum n_i^2 \to \gamma$ and $n^{-1}\sum n_i^{-1} \to \xi$. Assumptions (i)–(iii) are satisfied by $n_i \le M$ and $n^{-1}t \to c \in (0,1)$; the remaining assumptions insure that $F_n$ and $n^{-1}\operatorname{Cov}(Q_n)$ converge. Once convergence is established, invertibility follows from Lemmas 1 and 2.

For such a sequence of models $n^{1/2}(\hat{\sigma}_a^2 - \sigma_a^2, \hat{\sigma}_e^2 - \sigma_e^2)'$ has a limiting bivariate normal distribution with mean 0 and covariance matrix

$$(F^{-1})\Delta(F^{-1})' = \begin{bmatrix} 2\gamma\sigma_a^4 + 4\sigma_a^2\sigma_e^2 + 2c(1-c)^{-1}\sigma_e^4 & -2c(1-c)^{-1}\sigma_e^4 \\ -2c(1-c)^{-1}\sigma_e^4 & 2(1-c)^{-1}\sigma_e^4 \end{bmatrix} + \begin{bmatrix} \gamma k_a + (\xi - c^2)(1-c)^{-2}k_e & (c^2 - \xi)(1-c)^{-2}k_e \\ (c^2-\xi)(1-c)^{-2}k_e & (1 - 2c + \xi)(1-c)^{-2}k_e \end{bmatrix}, \tag{13}$$

where $k_a = E(\alpha_1^4) - 3\sigma_a^4$ and $k_e = E(\varepsilon_{i1}^4) - 3\sigma_e^4$.

In the balanced case $n_i \equiv m$ (fixed), $c = m^{-1}$, $\gamma = m$, and $\xi = m^{-2}$. In this case

$$(14)\quad (F^{-1})\Delta(F^{-1})' = \begin{bmatrix} 2m\sigma_a^4 + 4\sigma_a^2\sigma_e^2 + 2(m-1)^{-1}\sigma_e^4 & -2(m-1)^{-1}\sigma_e^4 \\ -2(m-1)^{-1}\sigma_e^4 & 2(m-1)^{-1}m\sigma_e^4 \end{bmatrix} + \begin{bmatrix} mk_a & 0 \\ 0 & k_e \end{bmatrix}.$$

From this expression it is readily apparent that asymptotically level-robust procedures are available for testing $H_0$: $\sigma_a^2 = 0$ in the nonnormal balanced case (as noted by Scheffé (1959), page 344). In the unbalanced case it is clear from (13) that the same test is asymptotically conservative for $k_e < 0$ and asymptotically liberal for $k_e > 0$. The extent to which the test is conservative or liberal depends largely on the quantity $(\xi - c^2)$.

Apart from a constant of proportionality, the covariance matrix (14) is identical to that obtained by Brown (1976) for the MINQUE estimators when $m = 2$. (Brown uses $t$ as a normalizing constant. It is well known that the ANOVA and MINQUE estimates coincide in the standard balanced models.)

**6. Concluding remarks.** The asymptotic normality of the ANOVA estimates establishes the asymptotic validity of inferential procedures when the random effects have normal distributions (or more generally, when kurtosis parameters are null). It is seen that in some balanced cases, asymptotically valid procedures for testing whether a component of variance is null are obtained when the kurtosis parameters are nonzero. In cases where the design is unbalanced and the kurtosis parameters are nonzero, the results give a means of assessing the degree of nonlevel robustness of normal-theory based inferential procedures when a priori information on the "tailedness" of the random effect distributions is available.

**Acknowledgments.** This paper is a part of the author's doctoral dissertation at the University of California at Davis. I wish to acknowledge the motivation and direction I received from my advisor, Professor Alan P. Fenech, and support from the University of California at Davis throughout the course of preparing my dissertation. I would also like to thank a referee and an Associate Editor for some corrections and many helpful suggestions that have improved the manuscript.

## REFERENCES

Brown, K. G. (1976). Asymptotic behaviour of MINQUE-type estimators of variance components. *Ann. Statist.* **4** 746–754.

Corbeil, R. R. and Searle, S. R. (1976). A comparison of variance component estimators. *Biometrics* **32** 779–791.

Fox, R. and Taqqu, M. S. (1985). Non-central limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Probab.* **13** 428–446.

HARTLEY, H. O. and RAO, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93–108.

MILLER, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

OLSEN, A., SEELY, J. and BIRKES, D. (1976). Invariant quadratic unbiased estimation for two variance components. *Ann. Statist.* **4** 878–890.

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

SEARLE, S. R. (1971). *Linear Models*. Wiley, New York.

SWALLOW, W. H. and MONAHAN, J. F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. *Technometrics* **26** 47–57.

VARBERG, D. E. (1966). Convergence of quadratic forms in independent random variables. *Ann. Math. Statist.* **37** 567–576.

WHITTLE, P. (1960). Bounds for moments of linear and quadratic forms in independent random variables. *Theory Probab. Appl.* **5** 302–305.

TEXAS TECH UNIVERSITY
COLLEGE OF BUSINESS ADMINISTRATION
INFORMATION SYSTEMS AND
QUANTITATIVE SCIENCES
P.O. BOX 4320
LUBBOCK, TEXAS 79409