The Annals of Statistics
1978, Vol. 6, No. 5, 1149–1158

ESTIMATION FOR A LINEAR REGRESSION MODEL WITH UNKNOWN DIAGONAL COVARIANCE MATRIX

BY WAYNE A. FULLER AND J. N. K. RAO

Iowa State University and Carleton University

A method of estimating the parameters of a linear regression model when the covariance matrix is an unknown diagonal matrix is investigated. It is assumed that the observations fall into k groups with constant error variance for a group. The estimation is carried out in two steps, the first step being an ordinary least squares regression. The least squares residuals are used to estimate the covariance matrix and the second step is the calculation of the generalized least squares estimator using the estimated covariance matrix. The large sample properties of the estimator are derived for increasing k, assuming the numbers in the groups form a fixed sequence.

1. Introduction. We consider the problem of estimating the parameter, $\boldsymbol{\beta}$, in the linear regression model with heteroscedastic error variances:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}\,, \tag{1}$$

where $\mathbf{y}$ is an n-vector of observations $y_{ij}(j = 1, \cdots, n_i; i = 1, \cdots, k; \sum_i n_i = n)$, $\mathbf{X}$ is an $n \times r$ full rank matrix of known constants $(n > r)$, $\boldsymbol{\beta}$ is an r-vector of regression parameters and $\mathbf{e}$ is an n-vector of random variables e_{ij} with mean zero and dispersion matrix

$$\mathbf{V} = \text{block diag.}\ \{\sigma_1^2\mathbf{I}_{n_1}, \cdots, \sigma_k^2\mathbf{I}_{n_k}\}\,. \tag{2}$$

In (2), σ_i^2 is the unknown error variance associated with e_{ij} and $\mathbf{I}_{n_i}$ is the $n_i \times n_i$ identity matrix.

The ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y\,.$$

Let the two step weighted least squares (WLS) estimator of $\boldsymbol{\beta}$ be defined by

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y})\,, \tag{3}$$

where

$$\hat{\mathbf{V}} = \text{block diag.}\ \{\hat{\sigma}_1^2\mathbf{I}_{n_1}, \cdots, \hat{\sigma}_k^2\mathbf{I}_{n_k}\} \tag{4}$$

and $\hat{\sigma}_i^2$ is an estimator of σ_i^2. The estimator $\tilde{\boldsymbol{\beta}}$ is obtained in two steps. The first step is to obtain an estimator $\hat{\mathbf{V}}$. The second step is the calculation outlined in (3). Bement and Williams [2] and Jacquez et al. [5], among others, have investigated the efficiency of $\tilde{\boldsymbol{\beta}}$, relative to the OLS estimator, given that $\hat{\sigma}_i^2 = s_i^2 = \sum_j (y_{ij} - \bar{y}_i)^2/(n_i - 1)$, where $\bar{y}_i = \sum_j y_{ij}/n_i$, is used in the estimated covariance matrix $\hat{\mathbf{V}}$. Jacquez et al. argued that results for small n_i are important

Received March 1977; revised October 1977.
AMS 1970 *subject classification*. Primary 62J05.
Key words and phrases. Two step estimators, linear regression, unknown variances, weighted least squares.

because it is often impractical to obtain more than 4 or 5 replicates at a point and 2 or 3 replicates at a point is a common situation.

Rao and Subrahmaniam [9] used the OLS residuals $\hat{e} = y - X\hat{\beta}$ and the method of MINQUE (minimum norm quadratic unbiased estimation), to estimate the σ_i^2. They empirically investigated the efficiency of the resulting WLS estimator in the equireplicated case $n_i = m$ for all i. Their WLS estimator was found to be considerably more efficient than that based on the s_i^2 when m is small ($\leqq 5$) and k is relatively large.

Williams [10] reviewed some recent developments in WLS estimation, giving a number of references. Williams obtained lower bounds on convergence rates of WLS estimators to best linear unbiased estimators when the σ_i^2 are a function of a finite number of parameters.

In the present context the approach discussed by Williams corresponds to fixing the number of groups k and letting the number of elements n_i within each group increase. A notable exception to this type of approach is the pioneering work of Neyman and Scott [7], where the estimation difficulties associated with an increasing number of groups are discussed. Assuming that the errors e_{ij} are normally distributed, they studied estimators of μ for the model

$$\mathbf{y} = \mu\mathbf{1} + \mathbf{e} ,$$

where $\mathbf{1}$ is a column of ones. Neyman and Scott constructed an estimator which is asymptotically (as $k \to \infty$) more efficient than the maximum likelihood estimator of μ, provided the n_i differ.

We shall study the class of two step estimators of $\boldsymbol{\beta}$ given by

$$(5) \qquad \tilde{\boldsymbol{\beta}}_w = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{y}) ,$$

where $\hat{\mathbf{e}}' = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' = (\hat{e}_{11}, \cdots, \hat{e}_{1n_1}; \cdots; \hat{e}_{k1}, \cdots, \hat{e}_{kn_k})$,

$$(6) \qquad \hat{\sigma}_i^2 = n_i^{-1} \textstyle\sum_{j=1}^{n_i} \hat{e}_{ij}^2$$

$$(7) \qquad \mathbf{W} = \text{block diag. } \{w_1\mathbf{I}_{n_1}, \cdots, w_k\mathbf{I}_{n_k}\}$$

and $w_i = g(n_i)$ for some $g(\cdot)$ such that $0 < \gamma_1 < w_i < \gamma_2 < \infty$ for all i (the w_i generate the class). For the equireplicated case, $\tilde{\boldsymbol{\beta}}_w$ reduces to $\tilde{\boldsymbol{\beta}}$ given by (3).

The asymptotic distribution of $\tilde{\boldsymbol{\beta}}_w$ is derived in Section 2. Under the postulated model the individual $\hat{\sigma}_i^2$ do not converge in probability to the true σ_i^2 as the number of groups increases. Consequently, the distribution of the two step estimator of $\boldsymbol{\beta}$ is a function of the distribution of the differences $\hat{\sigma}_i^2 - \sigma_i^2$. Because the $\hat{\sigma}_i^2$ are functions of $\hat{\boldsymbol{\beta}}$, it follows that replacing $\hat{\boldsymbol{\beta}}$ in equation (6) by an alternative estimator of $\boldsymbol{\beta}$ will produce a different limiting distribution for the estimator (5). One might use the two step estimator $\tilde{\boldsymbol{\beta}}_w$ to construct new estimators of σ_i^2 and insert these estimated variances into (5) to obtain a three step estimator of $\boldsymbol{\beta}$. One can view the maximum likelihood estimator as the limit of such an iterative process with $\mathbf{W} \equiv \mathbf{I}$. Hence, the distribution of the maximum likelihood estimator differs from that of the two step estimator.

The special case of the two step estimator of a common mean is investigated in Section 3. The two step estimator is demonstrated to be superior to the maximum likelihood estimator for a considerable range of parameter values.

2. Asymptotic distribution of $\tilde{\boldsymbol{\beta}}_w$. In deriving the large sample properties of $\tilde{\boldsymbol{\beta}}_w$ we shall use the following assumptions:

(a) The sequences $\{\sigma_i^2\}$ and $\{n_i\}$ satisfy $0 < \sigma_L^2 \leqq \sigma_i^2 \leqq \sigma_U^2 < \infty$ and $3 \leqq n_i < n^* < \infty$ for all i.

(b) The rows of $\mathbf{X}$, $(x_{ij1}, \cdots, x_{ijr})$, form a fixed sequence with $\sum_{t=1}^{r} x_{ijt}^2 < \delta < \infty$ for all (i, j).

(c) The limits (as $k \to \infty$) of the matrices $n^{-1}\mathbf{X}'\mathbf{X}$, $n^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}$, $n^{-1}\mathbf{X}'\mathbf{W}\mathbf{X}$, $n^{-1}\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}$, $n^{-1}\mathbf{X}'\mathbf{W}\mathbf{V}^{-1}\mathbf{L}\mathbf{W}\mathbf{X}$ and $n^{-1}\mathbf{X}'\mathbf{V}_w^{-1}\mathbf{X}$ exist and are positive definite, where

$$\mathbf{G} = \text{block diag.}\left\{\frac{1}{n_1 - 2}\mathbf{I}_{n_1}, \cdots, \frac{1}{n_k - 2}\mathbf{I}_{n_k}\right\} \tag{8}$$

$$\mathbf{L} = \text{block diag.}\left\{\frac{n_1}{n_1 - 2}\mathbf{I}_{n_1}, \cdots, \frac{n_k}{n_k - 2}\mathbf{I}_{n_k}\right\} \tag{9}$$

and

$$\mathbf{V}_w^{-1} = \text{block diag.}\left\{\frac{n_1 w_1}{(n_1 - 2)\sigma_1^2}\mathbf{I}_{n_1}, \cdots, \frac{n_k w_k}{(n_k - 2)\sigma_k^2}\mathbf{I}_{n_k}\right\}. \tag{10}$$

Strictly, an additional subscript k should be introduced, e.g., $n_k^{-1}X_k'X_k$, when discussing the asymptotic behavior as $k \to \infty$. Because such notation becomes cumbersome we have omitted the subscript.

2.1. *The lemmas.* We present seven lemmas prior to our principal result. Lemma 1 is proven by Chung [4], page 111, and Lemma 7 is a straightforward integration result. Proofs of Lemmas 2 through 6 are given in the Appendix.

LEMMA 1. *Let $\{Z_i\}$ be a sequence of independent random variables with distribution functions $\{F_i\}$ and let $S_k = \sum_i Z_i$. Let $\{c_k\}$ be an increasing sequence of positive numbers. Suppose*

(i) $\sum_{i=1}^{k} \int_{|z|>c_k} dF_i(z) = o(1)$,

(ii) $c_k^{-2} \sum_{i=1}^{k} \int_{|z|\leqq c_k} z^2\, dF_i(z) = o(1)$

and

(iii) $a_k = \sum_{i=1}^{k} \int_{|z|\leqq c_k} z\, dF_i(z)$.

Then

$$c_k^{-1}(S_k - a_k) \to_p 0\,.$$

LEMMA 2. *Let assumption* (a) *hold, let $\{b_i\}$ be a sequence such that $|b_i| \leqq b^* < \infty$ for all i, and let*

$$\lim_{k\to\infty} k^{-1} \sum_{i=1}^{k} b_i n_i (n_i - 2)^{-1} \sigma_i^{-2} = A$$

where A is a real number. Then

$$k^{-1} \sum_{i=1}^{k} \tilde{\sigma}_i^{-2} b_i \to_p A\,,$$

where

$$\tilde{\sigma}_i^2 = n_i^{-1} \sum_{j=1}^{n_i} e_{ij}^2 . \tag{11}$$

LEMMA 3. *Let assumption* (a) *hold. Then, for* $3^{-1} < \alpha < 2^{-1}$,

$$k^{-3\alpha} \sum_{i=1}^{k} (\sum_{j=1}^{n_i} e_{ij}^2)^{-2} |e_{i1}| \to_p 0 .$$

LEMMA 4. *Let assumptions* (a) *and* (b) *hold. Then*

$$k^{-1} \sum_{i=1}^{k} |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| \to_p 0 .$$

LEMMA 5. *Let assumption* (a) *hold. Then, for* $3^{-1} < \alpha < 2^{-1}$,

$$k^{-3\alpha} \sum_{i=1}^{k} (e_{i1}^2 + e_{i2}^2 + e_{i3}^2)^{-3} e_{i1}^2 |e_{i2}| \to_p 0 .$$

LEMMA 6. *Let assumptions* (a) *and* (b) *hold. Then*

$$k^{-\frac{1}{2}}\{\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e} - \mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e} - 2\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\} \to_p 0 ,$$

where

$$\tilde{\mathbf{V}} = \text{block diag.}\ \{\tilde{\sigma}_1^2 \mathbf{I}_{n_1}, \cdots, \tilde{\sigma}_k^2 \mathbf{I}_{n_k}\} \tag{12}$$

and $\hat{\mathbf{V}}$ *is given by* (4).

LEMMA 7. *Let* U_1 *and* U_2 *be independent chi-square random variables with* 1 *and* 2 *degrees of freedom respectively. Then*

$$E\left\{\left(\frac{U_1^{\frac{1}{2}}}{U_1 + U_2}\right)^{2+\delta}\right\} = \frac{\Gamma\left(\frac{3+\delta}{2}\right)\Gamma\left(\frac{1+\delta}{2}\right)}{\Gamma(\frac{1}{2})\Gamma\left(\frac{5+\delta}{2}\right)} 2^{-\frac{1}{2}(2+\delta)}$$

for a positive $\delta < 1$, *where* $\Gamma(\cdot)$ *is the gamma function.*

2.2. *The main result.* The asymptotic distribution of $\tilde{\boldsymbol{\beta}}_w$ is given by the following:

THEOREM. *Let* $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ *where the* e_{ij} *are independently and normally distributed, and* e_{ij} *has mean zero and variance* σ_i^2. *Let* $\tilde{\boldsymbol{\beta}}_w$ *be defined by* (5) *and suppose the assumption* (a)—(c) *hold. Then* $n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_w - \boldsymbol{\beta})$ *has a limiting normal distribution with mean* **0** *and covariance matrix* **H** *given by*

$$\mathbf{H} = \lim_{k\to\infty} n(\mathbf{X}'\mathbf{V}_w^{-1}\mathbf{X})^{-1}\mathbf{D}(\mathbf{X}'\mathbf{V}_w^{-1}\mathbf{X})^{-1} \tag{13}$$

where

$$\mathbf{D} = \mathbf{X}'\mathbf{W}\mathbf{V}^{-1}\mathbf{L}\mathbf{W}\mathbf{X} + 2(\mathbf{M} + \mathbf{M}') + 4(\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}) \tag{14}$$

and

$$\mathbf{M} = \mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{X} . \tag{15}$$

PROOF. We have, from (1) and (5),

$$n^{\frac{1}{2}}(\tilde{\boldsymbol{\beta}}_w - \boldsymbol{\beta}) = (n^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{X})^{-1} n^{-\frac{1}{2}}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e}) . \tag{16}$$

We now follow the steps used in obtaining the asymptotic theory for standard regression models (e.g., Anderson [1], page 23):

STEP 1. Demonstrate that

$$\text{plim}_{n\to\infty} n^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{X}) = \lim_{n\to\infty} n^{-1}(\mathbf{X}'\mathbf{V}_w^{-1}\mathbf{X}) .$$

STEP 2. Show that $n^{-\frac{1}{2}}\boldsymbol{\lambda}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e})$ has a limiting normal distribution with mean zero and variance $\boldsymbol{\lambda}'\mathbf{D}\boldsymbol{\lambda}$ for every arbitrary $\boldsymbol{\lambda}$ ($\neq \mathbf{0}$).

PROOF OF STEP 1. Consider the (t, s)th element of $n^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{X})$ given by

$$\frac{1}{n} \sum_{i=1}^{k} w_i \hat{\sigma}_i^{-2} \sum_{j=1}^{n_i} x_{ijt} x_{ijs} \qquad (t, s = 1, \cdots, r) .$$

By assumptions (a) and (b),

$$|n^{-1} \sum_{i=1}^{k} w_i(\hat{\sigma}_i^{-2} - \tilde{\sigma}_i^{-2}) \sum_{j=1}^{n_i} x_{ijt}x_{ijs}| \leqq ck^{-1} \sum_{i=1}^{k} |\hat{\sigma}_i^{-2} - \tilde{\sigma}_i^{-2}| ,$$

where c is a positive constant. The quantity on the right converges to zero in probability by Lemma 4. Furthermore

$$\text{plim} \frac{1}{n} \sum_{i=1}^{k} \frac{w_i}{\tilde{\sigma}_i^2} \sum_{j=1}^{n_i} x_{ijt}x_{ijs} = \lim \frac{1}{n} \sum_{i=1}^{k} \frac{n_i w_i}{(n_i - 2)\sigma_i^2} \sum_{j=1}^{n_i} x_{ijt}x_{ijs}$$

by Lemma 2 with $b_i = w_i \sum_j x_{ijt}x_{ijs}$. The proof of step 1 is complete.

PROOF OF STEP 2. Consider

$$\hat{\theta} = n^{-\frac{1}{2}}\boldsymbol{\lambda}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e})$$

which by Lemma 6 can be written as

$$\hat{\theta} = \tilde{\theta} + o_p(1)$$

where

$$\begin{aligned} \tilde{\theta} &= n^{-\frac{1}{2}}\boldsymbol{\lambda}'[\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e} + 2\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}] \\ &= n^{-\frac{1}{2}} \sum_{i=1}^{k} \xi_i , \\ \xi_i &= \sum_{j=1}^{n_i} \sum_{s=1}^{r} (\lambda_s w_i \tilde{\sigma}_i^{-2} x_{ijs} e_{ij} + \gamma_s x_{ijs} e_{ij}) , \end{aligned}$$

and

$$(\gamma_1, \cdots, \gamma_r) = \boldsymbol{\gamma}' = 2\boldsymbol{\lambda}'\mathbf{X}'\mathbf{W}\mathbf{G}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} .$$

The ξ_i's are independent with uniformly bounded variances. By Lemma 7 and the fact that x_{ijt} and σ_i^2 are uniformly bounded, it follows that the ξ_i have bounded $2 + \delta$ moments for some $\delta(0 < \delta < 1)$. Therefore the conditions of Liapounov's central limit theorem are met and $\hat{\theta}$ converges in distribution to a normal random variable.

To evaluate the asymptotic variance of $\hat{\theta}$, we first consider the expectation of the (t, s)th element of $\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e}\mathbf{e}'\mathbf{W}\tilde{\mathbf{V}}^{-1}\mathbf{X}$, say δ_{ts}. We have

$$\delta_{ts} = E\{\sum_{i=1}^{k} w_i^2\tilde{\sigma}_i^{-4}(\sum_{j=1}^{n_i} x_{ijt}e_{ij})(\sum_{j=1}^{n_i} x_{ijs}e_{ij})\} .$$

Hence, following the proof of Lemma 7, we obtain

$$E\{\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e}\mathbf{e}'\mathbf{W}\tilde{\mathbf{V}}^{-1}\mathbf{X}\} = \mathbf{X}'\mathbf{W}\mathbf{V}^{-1}\mathbf{L}\mathbf{W}\mathbf{X} . \tag{17}$$

Next consider the expectation of the (t, s)th element of $\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e}\mathbf{e}'\mathbf{X}$, say ω_{ts}. We have

$$E\{\omega_{ts}\} = E\{\sum_{i=1}^{k} w_i \tilde{\sigma}_i^{-2}(\sum_{j=1}^{n_i} x_{ijt} e_{ij})(\sum_{j=1}^{n_i} x_{ijs} e_{ij})\}$$
$$= \sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} x_{ijt} x_{ijs} E\{e_{ij}^2 \tilde{\sigma}_i^{-2}\} = \sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} x_{ijt} x_{ijs} .$$

Hence

(18) $$E\{\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{W}\mathbf{e}\mathbf{e}'\mathbf{X}\} = \mathbf{X}'\mathbf{W}\mathbf{X} .$$

Finally

(19) $$E\{\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}\} = \mathbf{X}'\mathbf{V}\mathbf{X} .$$

Using (17)—(19), we obtain the expression $\boldsymbol{\lambda}'\mathbf{D}\boldsymbol{\lambda}$ as the asymptotic variance of $\hat{\theta}$. The proof of step 2 is complete. □

The asymptotic covariance matrix of $\tilde{\boldsymbol{\beta}}_w$ simplifies considerably in the important special case of all $n_i = m \geqq 3$:

(20) $$V(\tilde{\boldsymbol{\beta}}_w) = V(\tilde{\boldsymbol{\beta}}) \doteq (1 + 2m^{-1} - 8m^{-2})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} + 4m^{-2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} .$$

3. Combining independent estimators. The problem of combining independent estimators of a common mean μ has long been a concern of statisticians. The model for this problem is $y_{ij} = \mu + e_{ij} (j = 1, \cdots, n_i; i = 1, \cdots, k; \sum_i n_i = n)$ which is a special case of the regression model (1). Here k denotes the number of independent estimators $\bar{y}_1, \cdots, \bar{y}_k$ of μ and $\bar{y}_i = n_i^{-1} \sum_j y_{ij}$. The two step estimator (3) for the special case $n_i = m$ for all i reduces to

(21) $$\tilde{\mu} = (\sum_{i=1}^{k} \hat{\sigma}_i^{-2} n_i)^{-1} \sum_{i=1}^{k} \hat{\sigma}_i^{-2} n_i \bar{y}_i ,$$

where $\hat{\sigma}_i^2 = n_i^{-1} \sum_j (y_{ij} - \bar{y})^2$ and $\bar{y} = n^{-1} \sum_i n_i \bar{y}_i$. For $n_i = m$, the variances of the limiting distributions for $\tilde{\mu}$, $\bar{y}$ and the maximum likelihood estimator μ^* are given by

(22) $$V(\tilde{\mu}) \doteq m^{-2}(m - 2)^2(m + 4)V(\mu^*) + 4m^{-2}V(\bar{y})$$

(23) $$V(\bar{y}) = k^{-2}m^{-1} \sum_{i=1}^{k} \sigma_i^2$$

and

(24) $$V(\mu^*) \doteq (m - 2)^{-1}(\sum_{i=1}^{k} \sigma_i^{-2})^{-1} .$$

The formula (22) is obtained from (20) whereas (24) follows Neyman and Scott [7]. Note that the Neyman–Scott estimator is identical to μ^* in the special case $n_i = m$ for all i.

It is clear from (22) that $\tilde{\mu}$ is a compromise between the sample mean $\bar{y}$ and the maximum likelihood estimator μ^*. For a range of variability in the σ_i^2 it is superior to both $\bar{y}$ and μ^*. This is demonstrated by Table 1 which contains selected values of asymptotic relative efficiencies for the Cochran and Carroll [3] variance model. In this model it is assumed that one-third of the k groups

TABLE 1
Asymptotic relative efficiency of estimators of the mean for variances (α, 1, α^{-1}) *and* $n_i = m$

Variance Ratio	α	m = 3	4	5	6	10
$V(\mu^*)/V(\tilde{\mu})$	1	2.45	1.59	1.34	1.23	1.08
	2	2.17	1.49	1.35	1.19	1.06
	3	1.76	1.31	1.18	1.12	1.04
	4	1.40	1.09	1.05	1.03	1.01
	5	1.12	0.97	0.95	0.95	0.97
	6	0.91	0.82	0.84	0.86	0.93
$V(\bar{y})/V(\tilde{\mu})$	1	0.82	0.80	0.81	0.82	0.86
	2	0.98	1.02	1.05	1.08	1.16
	3	1.22	1.37	1.48	1.55	1.73
	4	1.43	1.74	1.95	2.11	2.47
	5	1.60	2.07	2.42	2.69	3.31
	6	1.72	2.35	2.86	3.27	4.23

have variance $1/\alpha$, one-third variance 1 and the remaining one-third variance α. For small $m\,(\leqq 6)$ the sample mean is most efficient until α approaches 2 (note that α^2 is the ratio of the largest σ_i to the smallest). For α in the range 3 to 4 and $m \leqq 10$, $\tilde{\mu}$ is most efficient and the gain in efficiency over μ^* and $\bar{y}$ is often large for $m \leqq 5$. For larger α, μ^* is most efficient but the gain in efficiency over $\tilde{\mu}$ is modest. As m increases the relative efficiency of $\tilde{\mu}$ to μ^* approaches one.

APPENDIX

PROOFS OF LEMMAS.

A.1. *Proof of Lemma* 2. Let $Z_i = b_i/\tilde{\sigma}_i^2 = (b_i n_i)/(\sigma_i^2 U_{(i)})$ where $U_{(i)}$ is a chi-squared variable with $n_i\,(\geqq 3)$ degrees of freedom. It can be verified that $\lim_{k\to\infty} k^{-(1+\delta)} \sum_i E\{|Z_i|^{1+\delta}\} = 0$ for a positive $\delta \leqq 1$ and the conclusion follows. (See Rao [8], page 118, exercise 4.5.) □

A.2. *Proof of Lemma* 3. Let $Z_i = |v_{i1}|(v_{i1}^2 + u_2)^{-2}$, where $u_2 = v_{i2}^2 + v_{i3}^2$ and the v_{ij} are independent $N(0, 1)$ variables. We first demonstrate that $T_k = k^{-3\alpha} \sum_i Z_i \to_p 0$ by verifying that the sequence $\{Z_i\}$ of independent random variables satisfies the conditions (i) and (ii) of Lemma 1 with $c_k = k^{3\alpha}$. Now

$$\begin{aligned} P[|Z_i| > k^{3\alpha}] &= P[|v_{i1}| > k^{3\alpha}(v_{i1}^2 + u_2)^2] \\ &\leqq P[|v_{i1}| > k^{3\alpha}(v_{i1}^4 + u_2^2)] \\ &\leqq P[|v_{i1}| < k^{-\alpha};\, u_2 < k^{-2\alpha}] \\ &\leqq P[v_{i1}^2 + u_2 < 2k^{-2\alpha}] < 2^{\frac{3}{2}}k^{-3\alpha} \end{aligned}$$

since $P(U_3 < d) < d^{\frac{3}{2}}$, where $U_3 = v_{i1}^2 + u_2$ is a chi-square variable with 3 degrees of freedom. Hence, condition (i) of Lemma 1 is established, noting that

$1 - 3\alpha < 0$. Let $u_1 = v_{i1}^2$, then for any i

$$\begin{aligned}\int_{|z|<k^{3\alpha}} z^2\, dF_i(z) &\leqq \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{1}{2})} \int_{u_1+u_2>k^{-2\alpha}} \frac{u_1^{\frac{1}{2}}}{(u_1+u_2)^4} e^{-\frac{1}{2}(u_1+u_2)}\, du_1\, du_2 \\ &\leqq c \int_{y>k^{-2\alpha}} y^{-4} y^{\frac{3}{2}} e^{-y/2}\, dy \\ &< ck^{3\alpha} \int_{y>k^{-2\alpha}} y^{-1} e^{-y/2}\, dy = O(k^{3\alpha}),\end{aligned}$$

where c is a positive constant. Hence, condition (ii) of Lemma 1 is also established. Similarly,

$$k^{-3\alpha} \sum_{i=1}^k \int_{|z|<k^{3\alpha}} z\, dF_i(z) \leqq ck^{1-3\alpha} \int_{y>k^{-2\alpha}} y^{-1} e^{-y/2}\, dy = o(1).$$

We conclude that $T_k \to_p 0$. If the v_{ij}'s are independent $N(0, \sigma_i^2)$, the integrals for the ith random variable Z_i change by a constant which is bounded for all i since $\sigma_L^2 \leqq \sigma_i^2 \leqq \sigma_U^2$. In a similar manner if $u_3 = u_2 + u_4$, where u_4 is a chi-square random variable, then $|v_{i1}|(v_{i1}^2 + u_2)^{-2} > |v_{i1}|(v_{i1}^2 + u_3)^{-2}$. □

A.3. *Proof of Lemma* 4. We may write

$$\frac{1}{n_i}(\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}) = -\frac{\sum_j (e_{ij} + e_{ij} - g_{ij})g_{ij}}{[\sum_j e_{ij}^2][\sum_j (e_{ij} - g_{ij})^2]},$$

where $g_{ij} = \sum_s (\hat{\beta}_s - \beta_s)x_{ijs}$ and $\hat{e}_{ij} = e_{ij} - g_{ij}$. Let $\delta > 0$ and $\varepsilon > 0$ be given and α be as defined in Lemma 3. Let A_{1k} be the event that, for at least one i, $\max_j |e_{ij}| < k^{-\alpha}$ and let A_{2k} be the event that $\max_{ij} |g_{ij}| > (2k)^{-\alpha}$. Let $\bar{A}_k = \bar{A}_{1k} \cap \bar{A}_{2k}$, where $\bar{A}_{hk}$ denotes the complement of $A_{hk} (h = 1, 2)$. Then

$$\begin{aligned}P[k^{-1} \textstyle\sum_{i=1}^k |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| > \delta] &\leqq P(A_k) + P[k^{-1} \textstyle\sum_{i=1}^k |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| > \delta \mid \bar{A}_k] \\ &\leqq P(A_k) + P[k^{-1} \textstyle\sum_{i=1}^k |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| > B_2 k^{-\frac{3}{2}} \sum_{i=1}^k D_i \mid \bar{A}_k] \\ &\quad + P[B_2 k^{-\frac{3}{2}} \textstyle\sum_{i=1}^k D_i > \delta]\end{aligned}$$

where B_2 is some positive constant, $D_i = \sum_j |e_{ij}|(\sum_h e_{ih}^2)^{-1}$ and A_k is the complement of $\bar{A}_k$.

By the property $P(U_3 > d) < d^{\frac{3}{2}}$ for a chi-square variable U_3 with 3 degrees of freedom we have $P(\bar{A}_{1k}) \geqq 1 - ck^{1-3\alpha}$, where c is a positive constant. Therefore, noting that $\tilde{\beta} - \beta = O_p(k^{-\frac{1}{2}})$, there exists a K_1 such that $P(A_k) > \varepsilon/3$ for $k > K_1$. If $|g_{ij}| < (2k)^{-\alpha}$ for $j = 1, \cdots, n_i$ and $|e_{it}| = \max_j |e_{ij}| > k^{-\alpha}$, then there is a positive constant B_1 such that

$$d_i = \frac{|\sum_j (e_{ij} + e_{ij} - g_{ij})|}{[\sum_j e_{ij}^2][\sum_j (e_{ij} - g_{ij})^2]} < B_1 \frac{\sum_j |e_{ij}|}{(\sum_j e_{ij}^2)^2} = B_1 D_i.$$

This is true since

$$\textstyle\sum_j (e_{ij} - g_{ij})^2 \geqq (4n_i)^{-1} \sum_j e_{ij}^2.$$

Because the x_{ijs} are uniformly bounded and $(\hat{\beta} - \beta) = O_p(k^{-\frac{1}{2}})$, there is a B_2 such that $P(k^{-1} \sum |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| > B_2 k^{-\frac{3}{2}} \sum_i D_i) < \varepsilon/3$ for all k. By Lemma 3 there is a K_2 such that $P[B_2 k^{-\frac{3}{2}} \sum_i D_i > \delta] < \varepsilon/3$ for $k > K_2$. Letting $K^* = \max(K_1, K_2)$, we have $P[k^{-1} \sum_i |\tilde{\sigma}_i^{-2} - \hat{\sigma}_i^{-2}| > \delta] < 3(\varepsilon/3) = \varepsilon$ for $k > K^*$. □

A.4. *Proof of Lemma* 5. The proof parallels that of Lemma 3. We verify that the sequence $\{Z_i\}$ of independent random variables satisfies conditions (i) and (ii) of Lemma 1 with $c_k = k^{3\alpha}$, where $Z_i = v_{i1}^2|v_{i2}|(v_{i1}^2 + v_{i2}^2 + v_{i3}^2)^{-3}$ and the v_{ij} are independent $N(0, 1)$ variables. Now for any i,

$$\begin{aligned} P[Z_i > k^{3\alpha}] &= P[v_{i1}^2|v_{i2}| > k^{3\alpha}(v_{i1}^2 + v_{i2}^2 + v_{i3}^2)^3] \\ &\leqq P[(v_{i1}^2 + v_{i2}^2 + v_{i3}^2)^{\frac{3}{2}} > k^{3\alpha}(v_{i1}^2 + v_{i2}^2 + v_{i3}^2)^3] \\ &= P[(v_{i1}^2 + v_{i2}^2 + v_{i3}^2)^{-\frac{3}{2}} > k^{3\alpha}] = P[U_3 > k^{-2\alpha}] < k^{-3\alpha} . \end{aligned}$$

Hence, condition (i) of Lemma 1 is satisfied. Likewise

$$\begin{aligned} P\{Z_i > z\} &= 1 - F_i(z) \leqq P\{v_{i3}^2 > z^{-\frac{2}{3}}\} \\ &= [\Gamma(\tfrac{3}{2})2^{\frac{3}{2}}]^{-1} \textstyle\int_0^{z^{-\frac{2}{3}}} x^{\frac{1}{2}} e^{-x/2}\, dx . \end{aligned}$$

Let $G(q) = 0$ for $q \leqq 1$; $= 1 - [3\Gamma(\frac{3}{2})2^{\frac{1}{2}}q]^{-1}$ for $q > 1$, and $g(q) = q^2$ for $q < k^{3\alpha}$; $= k^{6\alpha}$ for $q \geqq k^{3\alpha}$. Then $F_i(q) > G(g)$ and

$$\int_{q<k^{3\alpha}} q^2\, dF_i(g) < \int_0^\infty g(q)\, dG(q) = 1 - 2[3(\tfrac{3}{2})2^{\frac{1}{2}}]^{-1}(1 - k^{3\alpha}) ,$$

where we have used Problem 11, Chapter 3 of Lehmann [6]. Therefore, condition (ii) of Lemma 1 is established. □

A.5. *Proof of Lemma* 6. The tth element of $n^{-\frac{1}{2}}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{We}$ is given by

$$\begin{aligned} &k^{-\frac{1}{2}} \textstyle\sum_{i=1}^k \hat{\sigma}_i^{-2} w_i \sum_{j=1}^{n_i} x_{ijt} e_{ij} \\ &\quad = k^{-\frac{1}{2}} \textstyle\sum_{i=1}^k w_i \sum_{j=1}^{n_i} x_{ijt} e_{ij} \{\tilde{\sigma}_i^{-2} + 2\tilde{\sigma}_i^{-4}[\sum_{s=1}^r (\hat{\beta}_s - \beta_s) n_i^{-1} \sum_{h=1}^{n_i} x_{ihs} e_{ih}]\} \\ &\qquad - S_t - R_t , \end{aligned}$$

where

$$\begin{aligned} S_t &= k^{-\frac{1}{2}} \textstyle\sum_{i=1}^k n_i^{-1} \tilde{\sigma}_i^{-4} w_i \sum_{j=1}^{n_i} g_{ij}^2 \sum_{h=1}^{n_i} x_{iht} e_{ih} , \\ R_t &= k^{-\frac{1}{2}} \textstyle\sum_{i=1}^k n_i^{-2} \tilde{\sigma}_i^{-4} \hat{\sigma}_i^{-2} w_i \sum_{h=1}^{n_i} x_{iht} e_{ih} [\sum_{j=1}^{n_i} (2e_{ij} - g_{ij}) g_{ij}]^2 , \\ \hat{e}_{ij} &= e_{ij} - g_{ij} = e_{ij} - \textstyle\sum_{s=1}^r (\hat{\beta}_s - \beta_s) x_{ijs} , \end{aligned}$$

and we have used

$$\hat{\sigma}_i^{-2} = \tilde{\sigma}_i^{-2} - \tilde{\sigma}_i^{-4}(\hat{\sigma}_i^2 - \tilde{\sigma}_i^2) + \tilde{\sigma}_i^{-4}\hat{\sigma}_i^{-2}(\hat{\sigma}_i^2 - \tilde{\sigma}_i^2)^2 .$$

Noting that $\hat{\beta}_s - \beta_s = O_p(k^{-\frac{1}{2}})$ and using Lemma 3 we have $S_t \to_p 0$ as $k \to \infty$. Using Lemma 5 and the result $d_i < B_1 D_i$ established in Lemma 4 it follows that $R_t \to_p 0$.

By the arguments of Lemma 2,

$$\begin{aligned} &k^{-1} \textstyle\sum_{i=1}^k n_i^{-1} \tilde{\sigma}_i^{-4} w_i (\sum_{j=1}^{n_i} x_{ijt} e_{ij})(\sum_{h=1}^{n_i} x_{ihs} e_{ih}) \\ &\qquad = k^{-1} \textstyle\sum_{i=1}^k n_i^{-1} w_i \sum_{j=1}^{n_i} x_{ijt} x_{ijs} E\{\tilde{\sigma}_i^{-4} e_{ij}^2\} + o_p(1) \\ &\qquad = k^{-1} \textstyle\sum_{i=1}^k \sigma_i^{-2} (n_i - 2)^{-1} w_i \sum_{j=1}^{n_i} x_{ijt} x_{ijs} + o_p(1) , \end{aligned}$$

noting that $E\{e_{ij}^2(\sum_h e_{ih}^2)^{-2}\} = [n_i(n_i - 2)\sigma_i^2]^{-1}$. Hence

$$k^{-\frac{1}{2}}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{We} = k^{-\frac{1}{2}}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{We} + 2k^{-\frac{1}{2}}\mathbf{X}'\mathbf{WGV}^{-1}\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1) . \qquad □$$

Acknowledgment. This research was partially supported by a grant from the National Research Council of Canada. We thank David Hasza, Tom O'Connor, the Associate Editor and the referee for useful comments and constructive suggestions.

REFERENCES

[1] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

[2] BEMENT, T. R. and WILLIAMS, J. S. (1969). Variance of weighted regression estimators when sampling errors are independent and heteroscedastic. *J. Amer. Statist. Assoc.* **64** 1369–1382.

[3] COCHRAN, W. G. and CARROLL, S. P. (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. *Biometrics* **9** 447–459.

[4] CHUNG, K. L. (1974). *A Course in Probability Theory*. Academic Press, New York.

[5] JACQUEZ, J. A., MATHER, F. J. and CRAWFORD, C. R. (1968). Linear regression with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24** 607–626.

[6] LEHMANN, E. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[7] NEYMAN, J. and SCOTT, E. (1948). Consistent estimators based on partially consistent observations. *Econometrica* **16** 1–32.

[8] RAO, C. R. (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.

[9] RAO, J. N. K. and SUBRAHMANIAM, K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics* **27** 971–990.

[10] WILLIAMS, J. S. (1975). Lower bounds on convergence rates of weighted least squares to best linear unbiased estimators. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 555–569. Academic Press, New York.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50010

DEPARTMENT OF MATHEMATICS
CARLETON UNIVERSITY
OTTAWA, CANADA K1S 5B6