A Note on Dolby's Ultrastructural Model

by

Leon Jay Gleser
Purdue University

Technical Report #83-30

Purdue University
West Lafayette, Indiana

July, 1983

This research supported by the National Science Foundation under Grant No. MCS 81-21948.

A Note on Dolby's Ultrastructural Model

by

Leon Jay Gleser
Purdue University

SUMMARY

It is shown that in the unreplicated case of Dolby's (1976) ultrastructural model where the ratio k_1 of error variances is known, maximum likelihood estimates exist for the intercept, slope, and unknown error variance, even when the ratio k_2 of the variability of the means to the variability of the errors is unknown. This corrects an incorrect assertion in Dolby's paper. The resulting maximum likelihood estimators are shown to be consistent and asymptotically normal, with consistently estimable covariance matrix. On the other hand, the corresponding estimators of Dolby, which require knowledge of both k_1 and k_2, are shown to be inconsistent.

Some key words: Errors in variables, functional relation, maximum likelihood, consistency, asymptotic normality, asymptotic confidence intervals.

A Note on Dolby's Ultrastructural Model

by

Leon Jay Gleser
Purdue University

1. INTRODUCTION

Dolby (1976) has proposed a model in which independent pairs $(x_{ij}, y_{ij})'$ of random variables are observed.

$$\begin{pmatrix} x_{ij} \\ y_{ij} \end{pmatrix} = \begin{pmatrix} u_{ij} \\ \alpha+\beta u_{ij} \end{pmatrix} + \begin{pmatrix} e_{ij} \\ f_{ij} \end{pmatrix}, \tag{1}$$

where

$$\begin{pmatrix} u_{ij} - \mu_i \\ e_{ij} \\ f_{ij} \end{pmatrix} \text{ are i.i.d. } N(\underset{\sim}{0}, \begin{pmatrix} \sigma_u^2 & & 0 \\ & \sigma_e^2 & \\ 0 & & \sigma_f^2 \end{pmatrix}), \tag{2}$$

$1 \leq i \leq m$, $1 \leq j \leq n$. The parameters of this model are α, β, $\underset{\sim}{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)'$; σ_e^2, σ_f^2 and σ_u^2. The model (1), (2) is called <u>ultrastructural</u> by Dolby because the u_{ij}'s are allowed to have different unknown means (the μ_i's). When $\mu_1 = \mu_2 = \ldots = \mu_m$, the model reduces to the usual structural errors-in-variables model based on mn observations. On the other hand, when $\sigma_u^2 = 0$, the model reduces to a (replicated) functional errors-in-variables model.

The replicated (n=1) case of the ultrastructural model has been studied by Dolby (1976) and by Cox (1976). Each author independently finds maximum likelihood estimators (MLEs) for the parameters of the model in this case, and determines asymptotic (m fixed, $n \to \infty$) properties of these estimators. Gleser (1983) shows that Dolby's replicated ultrastructural model can also be described as a replicated functional errors-in-variables model with inequality constraints on the parameters, and uses results of Anderson (1951) to provide a brief alternative derivation of the MLEs and their asymptotic properties.

The present note concerns the unreplicated (n=1) case of Dolby's ultrastructural model. Here, MLEs of the parameters do not exist unless restrictions are imposed on the parameter space. Dolby (1976) asserts that for MLEs to exist it is sufficient for the ratios

$$k_1 = \frac{\sigma_f^2}{\sigma_e^2} \quad , \qquad k_2 = \frac{\sigma_u^2}{\sigma_e^2} \tag{3}$$

to be known, and determines MLEs of the remaining parameters α, β, σ_e^2 and $\underset{\sim}{\mu}$ under this assumption. Unfortunately, the MLE $\hat{\beta}$ of β is not found explicitly, but only as a root of a certain quintic polynomial.

Dolby (1976) also claims to obtain the asymptotic covariance matrix of the MLEs $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_e^2$ of α, β, and σ_e^2 when k_1 and k_2 are known. However, it is not clear what he means by "asymptotic," since the Fisher information matrix which he uses for his derivations is well defined only when m is fixed. In fact, it is shown (Theorem 3) in Section 2 that the MLEs $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}_e^2$ are <u>inconsistent</u> ($m \to \infty$) estimators of α, β, σ_e^2, respectively, when $k_2 > 0$. Consequently, these estimators are not asymptotically unbiased, and Dolby's Fisher information calculations cannot yield the asymptotic covariance matrix of these estimators. (See also Patefield (1978).)

Dolby (1976) briefly considered the unreplicated ultrastructural model when k_1 is known, but $k_2 \geq 0$ is unknown, and asserted that MLEs of the unknown parameters (α, β, σ_e^2, $\underset{\sim}{\mu}$, k_2) do not exist in this case. In Section 2 (Theorem 1) it is shown that this assertion is incorrect. Indeed, the MLEs α^*, β^*, $(\sigma_e^*)^2$, $\underset{\sim}{\mu}^*$ of α, β, σ_e^2

and $\underset{\sim}{\mu}$ are identical to the MLEs of these parameters for the usual functional errors-in-variables model with known error variance ratio k_1(Gleser, 1981), corresponding to the special case of the ultrastructural model in which k_1 is known and $k_2 = 0$. [Not surprisingly, the MLE k_2^* of k_2 is $k_2^* = 0$.] Theorem 2 of Section 2 then shows that α^*, β^*, and $2(\sigma_e^*)^2$ are consistent $(m \to \infty)$ estimators of α, β, and σ_e^2, respectively, regardless of the value of k_2.

Finally, the asymptotic $(m \to \infty)$ joint distribution of α^*, β^*,and $2(\sigma_e^*)^2$ is given in Section 3.

The results of this note overlap to some extent with those of Patefield (1978). Patefield demonstrates the inconsistency of Dolby's MLE β (but not of the MLEs of the other parameters). His argument, however, fails to show that the inconsistency is not a trivial one, such as is the inconsistency of $(\sigma^*)^2$ for σ_e^2. Patefield also suggests estimators for α, β, and σ_e^2 based on α^*, β^*, and $(\sigma^*)^2$, but derives these estimators by an ad hoc adjustment of the likelihood equations for Dolby's model (k_1 and k_2 known). He asserts consistency of these estimators under somewhat more general sequences of the incidental parameters μ_i than those given by (13), but only for normally distributed errors. Finally, he gives a formula for the covariance matrix of his estimators in large samples. Although the leading (order m^{-1}) term of this formula agrees with the results given in Theorem 4 of this note, Patefield's formula is actually obtained from an asymptotic expansion of the finite sample covariances of his estimators in powers of m^{-1}. For this expansion to be rigorous, it is necessary that the estimators in question have finite second moments. However, it is well known that $E|\beta^*| = \infty$ for all finite sample sizes m. Theorem 4 of the present note, on the other hand, shows asymptotic joint normality of the estimators, and gives the covariance matrix of this asymptotic normal distribution.

2. MAXIMUM LIKELIHOOD ESTIMATORS

Since n = 1, we drop the subscript j in (1) and (2). Thus, the model is

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} u_i \\ \alpha+\beta u_i \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \end{pmatrix} \tag{4}$$

$$\begin{pmatrix} u_i - \mu_i \\ e_i \\ f_i \end{pmatrix} \text{ are i.i.d. } N(\underset{\sim}{0}, \begin{pmatrix} \sigma_u^2 & & 0 \\ & \sigma_e^2 & \\ 0 & & \sigma_f^2 \end{pmatrix}),$$

$1 \leq i \leq m$. Alternatively,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} \mu_i \\ \alpha+\beta\mu_i \end{pmatrix} + \underset{\sim}{\epsilon}_i \quad , \tag{5}$$

where

$$\underset{\sim}{\epsilon}_i = \begin{pmatrix} e_i \\ f_i \end{pmatrix} + \begin{pmatrix} 1 \\ \beta \end{pmatrix} (u_i-\mu_i) \text{ are i.i.d. } N(0,\Sigma),$$

and

$$\Sigma = \begin{pmatrix} \sigma_e^2 + \sigma_u^2 & \beta\sigma_u^2 \\ \beta\sigma_u^2 & \sigma_f^2 + \beta^2\sigma_u^2 \end{pmatrix} = \sigma_e^2 \left[\begin{pmatrix} 1 & 0 \\ 0 & k_i \end{pmatrix} + k_2 \begin{pmatrix} 1 \\ \beta \end{pmatrix}\begin{pmatrix} 1 \\ \beta \end{pmatrix}'\right]. \tag{6}$$

We assume that $k_1 = \sigma_e^{-2}\sigma_f^2$ is known, $k_1 \geq 0$.

For notational convenience, we drop the subscript on σ_e^2; thus, $\sigma^2 = \sigma_e^2$. Note from (6) that

$$|\Sigma| = \sigma^4 (k_1+k_1k_2+k_2\beta^2),$$

and that

$$\Sigma^{-1} = \sigma^2 |\Sigma|^{-1}\left\{\begin{pmatrix} k_1 & 0 \\ 0 & 1 \end{pmatrix} + k_2 \begin{pmatrix} -\beta \\ 1 \end{pmatrix}\begin{pmatrix} -\beta \\ 1 \end{pmatrix}'\right\}.$$

Consequently, the likelihood of the data $\underset{\sim}{x} = (x_1, \ldots, x_m)'$, $y = (y_1, \ldots, y_m)'$ is

$$L(\underset{\sim}{x},\underset{\sim}{y}|\alpha,\beta,\sigma^2,\underset{\sim}{\mu},k_2) = \frac{\exp - \frac{1}{2\sigma^2} Q(\alpha,\beta,\underset{\sim}{\mu},k_2)}{(2\pi\sigma^2)^m \; (k_1+k_1k_2+k_2\beta^2)^{m/2}}, \tag{7}$$

where

$$Q(\alpha,\beta,\underset{\sim}{\mu},k_2) = \frac{k_1 \sum_{i=1}^{m}(x_i-\mu_i)^2 + \sum_{i=1}^{m}(y_i-\beta\mu_i-\alpha)^2 + k_2 \sum_{i=1}^{m}(y_i-\beta x_i-\alpha)^2}{(k_1+k_1k_2+k_2\beta^2)}.$$

Let $\underset{\sim}{1}_m = (1,1,\ldots,1)': m \times 1$, $\overline{x} = m^{-1} \sum_{i=1}^{m} x_i$, $\overline{y} = m^{-1} \sum_{i=1}^{m} y_i$, and

$$S = m^{-1}\sum_{i=1}^{m} \begin{pmatrix} x_i - \overline{x} \\ y_i - \overline{y} \end{pmatrix} \begin{pmatrix} x_i - \overline{x} \\ y_i - \overline{y} \end{pmatrix}' = \begin{pmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{pmatrix}.$$

Lemma 1. For each fixed β, k_2, $0 \leq k_2 < \infty$, and all α, σ^2, $\underset{\sim}{\mu}$,

$$L(\underset{\sim}{x},\underset{\sim}{y})|\alpha,\beta,\sigma^2,\underset{\sim}{\mu},k_2) \leq L(\underset{\sim}{x},\underset{\sim}{y}| \hat{\alpha}(\beta),\beta,\hat{\sigma}^2(\beta),\hat{\underset{\sim}{\mu}}(\beta),k_2), \tag{8}$$

where

$$\hat{\alpha}(\beta) = \overline{y} - \beta\overline{x}, \quad \hat{\underset{\sim}{\mu}}(\beta) = (k_1+\beta^2)^{-1} \; [k_1\underset{\sim}{x} + \beta(\underset{\sim}{y} - \hat{\alpha}(\beta)\underset{\sim}{1}_m)],$$

$$\hat{\sigma}^2(\beta) = (2(k_1+\beta^2))^{-1} \begin{pmatrix} -\beta \\ 1 \end{pmatrix}' S \begin{pmatrix} -\beta \\ 1 \end{pmatrix}, \tag{9}$$

and

$$L(\underset{\sim}{x}, \underset{\sim}{y}|\hat{\alpha}(\beta), \beta, \hat{\sigma}^2(\beta), \hat{\underset{\sim}{\mu}}(\beta), k_2)$$

$$= \frac{\exp(-m)}{[2\pi\hat{\sigma}^2(\beta)]^m \; (k_1+k_1k_2+k_2\beta^2)^{m/2}}. \tag{10}$$

Proof. Fix β, σ^2 and k_2. To maximize the likelihood over $\underset{\sim}{\mu}$ and α, we see from (7) that we need to minimize $Q(\alpha, \beta, \underset{\sim}{\mu}, k_2)$. Minimizing first over $\underset{\sim}{\mu}$ (α fixed), it is not hard to show that this minimum is attained for $\underset{\sim}{\mu} = \hat{\underset{\sim}{\mu}}(\alpha,\beta)$ where

$$\hat{\underset{\sim}{\mu}}(\alpha,\beta) = (k_1+\beta^2)^{-1} \; [k_1 \underset{\sim}{x} + \beta(\underset{\sim}{y} - \alpha \underset{\sim}{1}_m)].$$

Substituting $\hat{\underset{\sim}{\mu}}(\alpha,\beta)$ for $\underset{\sim}{\mu}$ in $Q(\alpha, \beta, \underset{\sim}{\mu}, k_2)$ and simplifying yields

$$Q(\alpha,\beta,\hat{\underset{\sim}{\mu}}(\alpha,\beta),k_2) = \frac{\sum_{i=1}^{m} (y_i - \alpha - \beta x_i)^2}{k_1 + \beta^2} \quad . \tag{11}$$

Minimizing this expression in turn over α, we see that the minimum is attained for $\alpha = \hat{\alpha}(\beta) = \bar{y} - \beta\bar{x}$. Plugging in $\hat{\alpha}(\beta)$ for α in $\hat{\underset{\sim}{\mu}}(\alpha,\beta)$ and (11) yields

$$\hat{\underset{\sim}{\mu}}(\hat{\alpha}(\beta),\beta) = \hat{\underset{\sim}{\mu}}(\beta)$$

and

$$Q(\hat{\alpha}(\beta),\beta,\hat{\underset{\sim}{\mu}}(\beta)) = \frac{\sum_{i=1}^{m}(y_i-\bar{y}-\beta(x_i-\bar{x}))^2}{k_1 + \beta^2} = \frac{\begin{pmatrix}-\beta\\1\end{pmatrix}' S \begin{pmatrix}-\beta\\1\end{pmatrix}}{k_1 + \beta^2} \quad .$$

Finally, we maximize $L(\underset{\sim}{x},\underset{\sim}{y}|\hat{\alpha}(\beta),\beta,\sigma^2,\hat{\underset{\sim}{\mu}}(\beta),k_2)$ over σ^2 (β,k_2 fixed), and arrive at the conclusion given in (8), (9) and (10). □

2.1 The Case k_1 Known, k_2 Unknown

Using Lemma 1, we can find the MLEs of α, β, $\sigma^2 = \sigma_e^2$, $\underset{\sim}{\mu}$ and k_2 for the model (4) when k_1 is known.

Theorem 1. For the model (4) with k_1 known, the MLEs of α, β, σ^2, $\underset{\sim}{\mu}$ and k_2 are:

$$\alpha^* = \bar{y} - \beta^* \bar{x}, \; \underset{\sim}{\mu}^* = (k_1+(\beta^*)^2)^{-1} \; (k_1\underset{\sim}{x} + \beta^*(\underset{\sim}{y} - \alpha^* \underset{\sim}{1}_m)),$$

$$(\sigma^*)^2 = \frac{\begin{pmatrix}-\beta^*\\1\end{pmatrix}' S \begin{pmatrix}-\beta^*\\1\end{pmatrix}}{2\;(k_1+(\beta^*)^2)} \quad , \quad k_2^* = 0,$$

and

$$\beta^* = \frac{s_{yy} - k_1 s_{xx} + [(k_1 s_{xx} - s_{yy})^2 + 4k_1 s_{xy}^2]^{1/2}}{2s_{xy}} .$$

Proof It follows from Lemma 1, that to find MLEs we must maximize $L(\underset{\sim}{x},\underset{\sim}{y}|\hat{\alpha}(\beta), \beta, \hat{\sigma}^2(\beta), \hat{\underset{\sim}{\mu}}(\beta), k_2)$ over β and k_2, or equivalently (see (10)) minimize

$$G(\beta,k_2) = \left[\frac{\begin{pmatrix}-\beta\\1\end{pmatrix}' S \begin{pmatrix}-\beta\\1\end{pmatrix}}{k_1 + \beta^2}\right]^2 (k_1+k_1k_2+k_2\beta^2) \qquad (12)$$

over β and k_2. For fixed β, it is apparent that $G(\beta,k_2)$ is minimized over $k_2 \geq 0$ when $k_2 = 0$. Minimizing $G(\beta,0)$ over β then yields the result that the minimum occurs for $\beta = \beta^*$. The formulas for α^*, $(\sigma^*)^2$ and $\underset{\sim}{\mu}^*$ result from plugging β^* in for β in (9). □

It may be useful for future hypothesis testing problems to note that the maximum likelihood in Theorem 1 is

$$L\,(\underset{\sim}{x},\underset{\sim}{y}|\alpha^*,\beta^*,(\sigma^*)^2,\underset{\sim}{\mu}^*,k_2^*) = \frac{\exp\,(-m)}{[2\pi(\sigma^*)^2]^m} .$$

This is equal to the maximum likelihood for the special case of the model (4) in which k_1 is known and $k_2 = 0$ (which, as already noted, is equivalent to the classical functional errors-in-variables model with known ratio k_1 of error variances). Consequently, it is not possible to test the hypothesis $H{:}k_2 = 0$ (equivalently $\sigma_u^2 = 0$) in the unreplicated ultrastructural model (3), at least by likelihood ratio methods.

In fact, using the methods of Section 2 of Gleser (1983), it can be shown that the parameter σ_u^2 cannot be consistently estimated ($m \to \infty$), since this parameter is confounded with the variation of the unknown means $\mu_1, \mu_2, \ldots, \mu_m$. On the other hand, σ_u^2 can be consistently estimated (m, $n \to \infty$) in the replicated ultrastructural model (1), (2). To summarize: The unreplicated ultrastructural model (4) with k_1 known, $k_2 = \sigma_e^{-2}\sigma_u^2$ unknown, cannot be distinguished statistically from the unreplicated functional errors-in-variables model with known ratio k_1 of error variances, but the replicated cases of these two models can be distinguished. This fact is of interest because one of Dolby's motivations in introducing the ultrastructural model was that "it may be specialized to the functional and structural relations, thereby facilitating a unified approach to both."

Next, we show that α^* and β^* are consistent ($m \to \infty$) estimators of α and β, respectively, while $2(\sigma^*)^2$ is a consistent estimator of $\sigma^2 = \sigma_e^2$.

<u>Theorem 2</u>. Suppose that as $m \to \infty$,

$$\lim_{m\to\infty} \bar{\mu} = \lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} \mu_i = \mu, \quad \lim_{m\to\infty} m^{-1}\sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 = \Delta, \tag{13}$$

exist. Then, provided that Δ and k_2 are not both 0,

$$\lim_{m\to\infty} \alpha^* = \alpha, \ \lim_{m\to\infty} \beta^* = \beta, \ \lim_{m\to\infty} 2(\sigma^*)^2 = \sigma^2 = \sigma_e^2 , \tag{14}$$

with probability one for all α, β, σ^2 and k_2.

<u>Proof</u>. Using arguments similar to those used to prove Lemma 3.1 of Gleser (1981), when (13) holds it can be shown that

$$\lim_{m\to\infty}\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix}=\begin{pmatrix}\mu\\ \alpha+\beta\mu\end{pmatrix}, \quad \lim_{m\to\infty} S = \sigma^2\begin{pmatrix}1 & 0\\ 0 & k_1\end{pmatrix} + (k_2\sigma^2+\Delta)\begin{pmatrix}1\\ \beta\end{pmatrix}\begin{pmatrix}1\\ \beta\end{pmatrix}' , \tag{15}$$

with probability one for all α, β, σ^2, and k_2. Since α^*, β^* and $2(\sigma^*)^2$ are continuous functions of $\bar{x}$, $\bar{y}$ and S (except when $s_{xy} = 0$, an event of zero probability), the assertion (14) follows directly from (15) and the formulas for α^*, β^*, $(\sigma^*)^2$ given in Theorem 1. □

Remark 1. The results in Theorem 2 do not require that $(u_i - \mu_i, e_i, f_i)$ have, for each i, a trivariate normal distribution. For Theorem 2 to hold it is sufficient that $(x_i, y_i)'$ satisfies the model (5) with the $\underset{\sim}{\varepsilon}_i$'s i.i.d. , with common mean $\underset{\sim}{0}$ and common covariance matrix Σ. The common distribution of the $\underset{\sim}{\varepsilon}_i$'s need not be normal.

2.2. Case when k_1 and k_2 are known

Now consider the model (4), when both k_1 and k_2 are known, $k_2 > 0$. Since we have additional information about the parameters, we would intuitively expect that the MLEs for this model would be more efficient (accurate) than the MLEs for the model where k_2 is unknown. However, the MLEs $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$ are not even consistent estimators of their respective parameters. (Nor is $2\hat{\sigma}^2$ consistent for σ^2.)

Lemma 2. The MLEs of α, $\underset{\sim}{\mu}$ and $\sigma^2 = \sigma_e^2$ for the model (4) with k_1 and k_2 known are

$$\alpha = \bar{y} - \hat{\beta}\bar{x}, \qquad \hat{\underset{\sim}{\mu}} = (k_1 + \hat{\beta}^2)^{-1} (k_1 \underset{\sim}{x} + \hat{\beta}(\underset{\approx}{y} - \hat{\alpha}\underset{\sim}{1}_m)),$$

$$\hat{\sigma}^2 = \frac{\begin{pmatrix} -\hat{\beta} \\ 1 \end{pmatrix}' S \begin{pmatrix} -\hat{\beta} \\ 1 \end{pmatrix}}{2(k_1 + \hat{\beta}^2)} ,$$

where the MLE $\hat{\beta}$ of β minimizes $G(\beta, k_2)$ defined by (12).

Proof. Follows directly from Lemma 1. □

Theorem 3. If as $m \to \infty$ the limits in (13) exist, and if $k_2 > 0$, then with probability one, all α, β, σ^2,

$$\lim_{m\to\infty} \hat{\alpha} = \alpha + (1-\zeta)\beta, \qquad \lim_{m\to\infty} \hat{\beta} = \zeta\beta$$

and

$$\lim_{m\to\infty} 2\hat{\sigma}^2 = \sigma^2 + (\zeta-1)^2 \frac{\beta^2(\sigma^2 k_2+\Delta)}{k_1+\zeta^2\beta^2} \quad ,$$

where $\zeta, \zeta \neq 1$, is the unique value of z minimizing

$$h(z) = \left[\frac{(z-1)^2 \ \beta^2(\sigma^2 k_2+\Delta)}{k_1+\beta^2 z^2}\right] + \sigma^2(k_1+k_1k_2+k_2\beta^2 z^2). \tag{16}$$

Proof. Since $\hat{\beta}$ minimizes $G(\beta, k_2)$, $(d/d\beta)\ G(\beta, k_2)$ is jointly continuous in β and the elements of S, and $(d/d\beta)^2\ G(\beta, k_2)$ is not 0 when $\beta = \hat{\beta}$, all S, it follows from the implicit function theorem that $\hat{\beta}$ is a continuous function of the elements of S. Consequently, (15) implies that $\lim_{m\to\infty} \hat{\beta}$ exists with probability one. Let

$$\lim_{m\to\infty} \hat{\beta} = \zeta\beta. \tag{17}$$

Then

$$\lim_{m\to\infty} G(\hat{\beta},k_2) = \left[\frac{\begin{pmatrix}-\beta\zeta \\ 1\end{pmatrix}\left(\lim_{m\to\infty} S\right)\begin{pmatrix}-\beta\zeta \\ 1\end{pmatrix}}{k_1 + \beta^2\zeta^2}\right]^2 (k_1+k_1k_2+k_2\beta^2\zeta^2)$$

$$= h(\zeta)$$

with probability one, where $h(z)$ is defined by (16). It is straightforward to show that ζ must uniquely minimize $h(z)$. [Consider $z\beta^*$ as an alternative estimator. It can be shown that $G(z\beta^*,k_2)$ converges with probability 1 to $h(z)$ as $m \to \infty$. Hence, if ζ does not uniquely minimize $h(z)$, then for large enough m, $\hat{\beta}$ does not minimize $G(\beta,k_2)$, contradicting Lemma 2.] Finally, when $k_2 > 0$, $\beta \neq 0$,

$$\frac{d}{dz}\ h(z)\Big|_{z=1} = 2k_2\sigma^4\ \beta^2 > 0,$$

and $(d/dz)\ h(z)$ is continuous in z in a neighborhood of z=1. Consequently, z=1 cannot minimize $h(z)$, and hence $\zeta \neq 1$. The limiting values for $\hat{\alpha}$ and $2\hat{\sigma}^2$ follow as a direct consequence of (15), the formulas for $\hat{\alpha}$ and $\hat{\sigma}^2$ in terms of $\hat{\beta}$ in Lemma 2, and (17). □

What has gone wrong with the usually reliable maximum likelihood method? One possibility is that we have been given too much information. When k_2 was unknown, the maximum likelihood estimation procedure simply ignored the fact that the variables u_i could be nondegenerate random variables, and variation in either the u_i's or in their means μ_i was assigned to the variation of the μ_i's. However, if we are told the value of k_2, and k_2 is not 0, the maximum likelihood procedure tries to separate the variance of the u_i's from the variability of the μ_i's. Because $\sigma_u^2 = k_2\sigma_e^2$, the estimation procedure borrows information in the data about σ_e^2 to help identify σ_u^2. This biases the estimate of σ_e^2, and consequently the estimate of β. The moral here is that "a little knowledge can sometimes be a dangerous thing."

Remark 2. Dolby (1976) asserts that when $k_1=0$, $k_2>0$, the MLE $\hat{\beta}$ of β is Teissier's (1948) estimator $(s_{yy}/s_{xx})^{\frac{1}{2}}$. This is clearly incorrect since β can be negative while $(s_{yy}/s_{xx})^{\frac{1}{2}}$ is nonnegative, and s_{xy} contains information about the sign of β. In fact, in this case

$$\hat{\beta} = (\text{sign of } s_{xy}) \left(\frac{s_{yy}}{s_{xx}}\right)^{\frac{1}{2}}.$$

Although with this modification Teissier's estimator may have the theoretical status of an MLE, this status is somewhat an empty one since, as Theorem 3 shows, β is inconsistent for β as $m \to \infty$.

Remark 3. Theorem 3 does not require normality of the random vectors $(u_i-\mu_i, e_i, f_i)$. See Remark 1.

2.3 Discussion

As estimators of the basic parameters α, β, σ_e^2 of the unreplicated structural model, the estimators α^*, β^*, $2(\sigma^*)^2$ are clearly preferable to $\hat{\alpha}$, $\hat{\beta}$, $2\hat{\sigma}^2$(or $\hat{\sigma}^2$) even when k_2 is assumed known. Not only are α^*, β^*, $2(\sigma^*)^2$ consistent estimators for α, β, σ_e^2, respectively (while $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ are not consistent unless $k_2=0$, in which case $\hat{\alpha}=\alpha^*$, $\hat{\beta}=\beta^*$, $\hat{\sigma}^2=(\sigma^*)^2$), but their calculation does not require knowledge of the value of k_2 (robustness with respect to misspecification of k_2). Finally (Theorem 1), these estimators are easily computed from the data.

In Section 3, it is shown that α^*, β^*, and $2(\sigma^*)^2$ have a large-sample ($m \to \infty$) trivariate normal distribution, and that the covariance matrix of this large-sample distribution can be consistently estimated. Hence, it is possible to construct large-sample confidence intervals for α, β, and σ_e^2.

3. ASYMPTOTIC NORMALITY

Assume that the limits defined in (13) exist, and that

$$\tau^2 = k_2\sigma^2 + \Delta > 0. \tag{18}$$

Using arguments similar to those on pp. 38-9 of Gleser (1981), replacing Lemma 4.4 (which is false) by a direct proof of Corollary 4.1, it can be shown that

$$\sqrt{m}\begin{pmatrix}\bar{x}-\mu \\ \bar{y}-\alpha-\beta\mu\end{pmatrix} \xrightarrow{L} N(\underset{\sim}{0},\Sigma)$$

and

$$\sqrt{m}\left[\begin{pmatrix}s_{xx} \\ s_{xy} \\ s_{yy}\end{pmatrix}-\begin{pmatrix}\sigma^2+\tau^2 \\ \tau^2\beta \\ k_1\sigma^2+\tau^2\beta^2\end{pmatrix}\right] \xrightarrow{L} N(\underset{\sim}{0},\Lambda),$$

where

$$\Lambda = \sigma^4\begin{pmatrix}2 & 0 & 0 \\ 0 & k_1 & 0 \\ 0 & 0 & 2k_1^2\end{pmatrix}+\sigma^2\tau^2\begin{pmatrix}4 & 2\beta & 0 \\ 2\beta & \beta^2+k_1 & 2k_1\beta \\ 0 & 2k_1\beta & 4k_1\beta^2\end{pmatrix}+(2\tau^4-\Delta^2)\begin{pmatrix}1 \\ \beta \\ \beta^2\end{pmatrix}\begin{pmatrix}1 \\ \beta \\ \beta^2\end{pmatrix}'.$$

Further, $(\bar{x},\bar{y})$ is independent of (s_{xx}, s_{xy}, s_{yy}).

Standard Taylor series expansions (the "delta method") and Slutzky's theorem can be used on the formulas for α^*, β^* and $2(\sigma^*)^2$ given in Theorem 1 to show that

$$\sqrt{m}\,(\alpha^*-\alpha) = \begin{pmatrix}-\beta\\ 1\end{pmatrix}' \sqrt{m}\left[\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix} - \begin{pmatrix}\mu\\ \alpha+\beta\mu\end{pmatrix}\right] - \mu\sqrt{m}\,(\beta^*-\beta) + o_p(1),$$

$$\sqrt{m}(\beta^*-\beta) = \frac{1}{\tau^2(k_1+\beta^2)} \begin{pmatrix}-k_1\beta\\ -\beta^2+k_1\\ \beta\end{pmatrix}' \sqrt{m}\left[\begin{pmatrix}s_{xx}\\ s_{xy}\\ s_{yy}\end{pmatrix} - \begin{pmatrix}\sigma^2+\tau^2\\ \tau^2\beta\\ k_1\sigma^2+\tau^2\beta^2\end{pmatrix}\right] + o_p(1),$$

$$\sqrt{m}(2(\sigma^*)^2-\sigma^2) = \frac{1}{k_1+\beta^2} \begin{pmatrix}\beta^2\\ -2\beta\\ 1\end{pmatrix}' \sqrt{m}\left[\begin{pmatrix}s_{xx}\\ s_{xy}\\ s_{yy}\end{pmatrix} - \begin{pmatrix}\sigma^2+\tau^2\\ \tau^2\beta\\ k_1\sigma^2+\tau^2\beta^2\end{pmatrix}\right] + o_p(1)\,.$$

Consequently, the following result can be demonstrated.

Theorem 4. If the limits defined in (13) exist and (18) holds,

$$\sqrt{m}\left[\begin{pmatrix}\alpha^*\\ \beta^*\\ 2(\sigma^*)^2\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\\ \sigma^2\end{pmatrix}\right] \xrightarrow{L} N\left(\underset{\sim}{0},\ \begin{pmatrix}\sigma^2(k_1+\beta^2)+\mu^2\psi & -\mu\psi & 0\\ -\mu\psi & \psi & 0\\ 0 & 0 & 2\sigma^4\end{pmatrix}\right),$$

where

$$\psi = \text{asymp. var of } \beta^* = \sigma^2\left(\frac{k_1\sigma^2}{\tau^4} + \frac{(k_1+\beta^2)}{\tau^2}\right) .$$

Remark 4. The above results continue to hold if $(x_i,y_i)'$, $i = 1,2,\ldots$, satisfy the model (5), with the $\underset{\sim}{\varepsilon}_i$'s i.i.d., but not necessarily normally distributed. However, the common distribution of the $\underset{\sim}{\varepsilon}_i$'s must have mean vector $\underset{\sim}{0}$, covariance matrix Σ, and third and fourth moments and cross-moments identical to those of a bivariate normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix Σ. (See Gleser, 1981, Section 4.)

To obtain large-sample $100(1-\alpha)\%$ confidence intervals (or joint confidence regions) for α, β, and $\sigma^2 = \sigma_e^2$, we need a consistent estimator of the asymptotic covariance matrix of $(\alpha^*, \beta^*, 2(\sigma^*)^2)'$. This clearly can be constructed, estimating μ by $\bar{x}$, σ^2 by $2(\sigma^*)^2$ and β by β^*, provided that we can find a consistent estimator of τ^2. One such consistent estimator is provided by

$$\hat{\tau}^2 = \frac{\begin{pmatrix} k_1 \\ \beta^* \end{pmatrix}' S \begin{pmatrix} k_1 \\ \beta^* \end{pmatrix}}{(k_1+(\beta^*)^2)^2} - \frac{k_1\, 2(\sigma^*)^2}{k_1+(\beta^*)^2} .$$

It can be shown that $\hat{\tau}^2$ is positive with probability one. Note that although $\sigma_\mu^2 = k_2\sigma^2$ cannot be consistently estimated (see Section 2), and Δ also cannot be consistently estimated, their sum τ^2 can be consistently estimated. Fortunately, this is all that is needed to estimate the asymptotic covariance matrix of $(\alpha^*, \beta^*, 2(\sigma^*)^2)'$.

Using the methods outlined in Gleser (1983, Section 2.3), it can be shown that when k_2 is unknown (and $\tau^2 > 0$), and the limits (13) exist, the estimators $\alpha^*, \beta^*, 2(\sigma^*)^2$ are BAN within the class of all asymptotically unbiased and asymptotically normal estimators of α, β, $\sigma^2=\sigma_e^2$ whose asymptotic covariance matrix depends on the sequence $\{\mu_i : i=1,2,\ldots\}$ of unknown means only through the limits μ and Δ. Since this class includes all estimators which depend upon the data $\underset{\sim}{x}$, $\underset{\sim}{y}$ only through $\bar{x}$, $\bar{y}$ and S (and are regular enough to permit Taylor series expansions in $\bar{x}$, $\bar{y}$, s_{xx}, s_{xy}, and s_{yy}), this provides justification for the use of $\alpha^*, \beta^*, 2(\sigma^*)^2$ in practice.

Research on this paper was supported by the National Science Foundation under Grant No. MCS 81-21948.

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Stat 17, 409-431. (Correction, Ann. Math. Stat. 35 (1964), 923-4).

ANDERSON, T. W. (1982). Estimating linear statistical relationships. Institute for Mathematical Studies in the Social Sciences, Stanford University, Tech. Report No. 389.

COX, N. R. (1976). The linear structural relation for several groups of data. Biometrika 63, 231-7.

DOLBY, G. R. (1976). The ultrastructural relation: A synthesis of the functional and structural relations. Biometrika 63, 39-50.

GLESER, L. J. (1981). Estimation in a multivariate 'errors in variables' regression model: Large sample results. Ann. Stat. 9, 24-44.

GLESER, L. J. (1983). Functional, structural and ultrastructural errors-in-variables models. Purdue University, Department of Statistics, Tech. Report #83-16.

PATEFIELD, W. M. (1978). The unreplicated ultrastructural relation: Large sample properties. Biometrika 65, 535-40.

TEISSIER, G. (1948). La relation d'allométrie: sa signification statistique et biologique (with discussion). Biometrics 4, 14-53.