Functional, Structural and Ultrastructural Errors-in-Variables Models (Preliminary Draft)

by

Leon Jay Gleser
Purdue University

Technical Report #83-16

Purdue University
West Lafayette, Indiana

May 1983

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

Functional, Structural and Ultrastructural Errors-in-Variables Models

(Preliminary Draft)

by

Leon Jay Gleser

1. INTRODUCTION

Consider the model

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \binom{0}{\alpha} \underset{\sim}{1}'_N + \binom{I_r}{B} U + E \qquad (1.1)$$

where Y_1: r×N and Y_2: s×N are observed matrices of random variables,

$$\underset{\sim}{1}_N = (1,1,\ldots,1)': N\times 1, \quad I_a = \text{the } a\times a \text{ Identity matrix},$$

and the columns $\underset{\sim}{e}_i$, $1 \le i \le N$, of E: (r+s)×N are i.i.d. random vectors with mean vector $\underset{\sim}{0}$ and covariance matrix Σ_e. This model is known as a <u>multivariate errors-in-variables regression model</u>.

Two types of errors-in-variables model are identified in the literature:

a. <u>The Structural Type.</u> Here, the columns u_i, $1 \le i \le N$, of U are i.i.d. r×1 vectors with common mean vector μ and common covariance matrix Σ_u. The random matrices U and E are assumed to be independent.

b. <u>The Functional Type.</u> Here, the matrix $U = E(Y_1)$ is an unknown matrix of constants. For purposes of comparison with the structural errors-in-variables model, define

$$\mu^{(N)} = N^{-1} U \underset{\sim}{1}_N, \qquad \Sigma_u^{(N)} = N^{-1} U (I_N - N^{-1} \underset{\sim}{1}_N \underset{\sim}{1}'_N) U'.$$

The errors-in-variables models (structural and functional) as defined above are overparameterized (non-identifiable). One way of attacking this problem is to put restrictions on the parameters -- e.g., set some of the paraemters equal to known values. Such an approach implicitly defines the parameters α, B, μ (or $u^{(N)}$), Σ_u (or $\Sigma_u^{(N)}$), and Σ_e as functions of a vector-valued parameter θ taking values in an open subset Θ of, say, t-dimensional Euclidean space R^t. Another method is to obtain an independent estimator S of Σ_e, usually from replications. A third method (of which the method of replications is a special case) is to find instrumental variables v_i related to the columns u_i of U, but independent of E. After some algebra, this method reduces to expressing U in the form

$$U = \Xi V + U^* \tag{1.2}$$

where the matrix V has columns v_i and is assumed known, Ξ is an unknown matrix of slopes, and U^* is either a matrix whose columns are i.i.d. with mean vector $\underset{\sim}{0}$ and covariance matrix Σ_u^* (structural case), or $U^* = \underset{\sim}{0}$ (functional case). The model defined by (1.1) and (1.2) in the functional case is discussed by Healy (1980), who does not, however, mention the connection between his model and the method of instrumental variables. When the columns of Y are normally distributed, Healy shows that the model defined by (1.1) and (1.2) can be reduced to a canonical form of the type (1.1) in which an independent estimator S of Σ_e is available. Although Healy only proved this result in the functional case, his arguments also apply to the structural case.

There is an extensive literature on errors-in-variables models of both functional and structural type. T. W. Anderson's 1982 Wald Lectures (Anderson, 1982) provide a clear and thorough review of this literature, with special emphasis on the connection between errors-in-variables models, the models of factor analysis, and econometric simultaneous equations models. Other useful surveys of this literature are given by Madansky (1959), Moran (1971) and Kendall and Stuart (1979). Most past research has dealt totally with one or the other of the two types (structural and functional) of errors-in-variables models. However, there have been basic papers, such as Anderson and Rubin (1956) and Nussbaum (1977), which have made use of results known for one type of errors-in-variables model to infer results for the other type of errors-in-variables model. In Section 2, we attempt to formalize some of the relationships between structural and functional errors-in-variables models that are useful in such efforts. In particular, we show that identifiability results for structural models can be used to determine whether or not consistent estimators exist for the parameters of corresponding functional errors-in-variables models, while on the other hand large-sample consistency and distributional results for a functional model imply similar large-sample consistency and distributional results for the corresponding structural model. Because structural errors-in-variables models, with Y having normally distributed columns, permit efficiency calculations of the Fisher-Cramér type, this last relationship enables us to define efficiencies for estimators and tests for functional errors-in-variables models, despite the presence of incidental parameters in such models.

Another attempt to connect structural and functional models is Dolby's (1976) ultrastructural model. Both Dolby and Cox (1976), who independently treated the model of Dolby's paper, point out that this model can be viewed as a replicated structural model. In Dolby's version of the model (which is defined for $r = s = 1$), we observe

$$y_{ij} = \binom{0}{\alpha} + \binom{1}{b} u_{ij} + e_{ij}: 2\times 1,\ 1 \le i \le m,\ 1 \le j \le n, \qquad (1.3)$$

where

the e_{ij}'s are i.i.d. $N(\underset{\sim}{0}, \begin{pmatrix} \sigma_{e1}^2 & 0 \\ 0 & \sigma_{e2}^2 \end{pmatrix})$,

the u_{ij}'s are independent, $u_{ij} \sim N(\mu_i, \sigma_u^2)$,

the μ_i's, α, b, σ_{e1}^2, σ_{e2}^2, and σ_u^2 are unknown paraemters, and the e_{ij}'s and u_{ij}'s are mutually independent. This model is called ultrastructural because it reduces to a structural model (1.1) with $Y = (y_{11}, \ldots, y_{1n}, \ldots, y_{m1}, \ldots, y_{mn})$ when $\mu_1 = \mu_2 = \ldots = \mu_m$, and to a functional model (replicated) when $\sigma_u^2 = 0$. (Allowing $\sigma_u^2 = 0$ is actually an extension of the usual parametric assumptions for structural models, which require $\sigma_u^2 > 0$.)

Although Dolby and Cox are correct in describing (1.3) as a replicated (when $n > 1$) structural model, it can also be described as a replicated functional model. Indeed, this model is a special case of Anderson's (1951) replicated functional model (specialized in that there are inequality constraints imposed on the parameters). This fact is demonstrated in Section 3. It is shown that viewing the ultrastructural model as a replicated functional model ($n > 1$) leads immediately to formulas for the maximum likelihood estimators of the parameters, and to consistency and asymptotic normality results for such estimators.

2. RELATIONSHIPS BETWEEN FUNCTIONAL AND STRUCTURAL MODELS

We will confine discussion in this section to errors-in-variables models (1.1) in which the parameters α, B, μ (or $u^{(N)}$), Σ_u (or $\Sigma_u^{(N)}$) and Σ_e are expressed as functions of a vector-valued parameter θ.

The key to the results which we prove in this section is the following fact.

Theorem 2.1. If Y obeys the errors-in-variables model (1.1) of structural type, then Y conditional upon the value of U obeys an errors-in-variables model of functional type with the same parameters α, B, Σ_e as in the structural model.

Proof. Immediate, once we note that in the structural model, U and E are assumed to be independent. □

Consequently, we can define a sample space and set of probability measures on which we can define correspondingly parameterized structural and functional models for arbitrary sample size N. Let

$$\mathscr{E} = \text{space of all sequences } \underset{\sim}{e} = \{e_i\} \text{ of } (r+s)\times 1 \text{ vectors } e_i,$$

$$\mathscr{U} = \text{space of all sequences } \underset{\sim}{u} = \{u_i\} \text{ of } r\times 1 \text{ vectors } u_i.$$

For each parameter value θ, define the probability measure $P_{\theta e}$ on $\mathscr{E}$ following the assumption (see (1.1)) that the e_is are i.i.d. with a common distribution (which may or may not be specified) having mean vector $\underset{\sim}{0}$ and covariance matrix $\Sigma_e(\theta)$. On $\mathscr{U}$, define the probability measure $P_{\theta u}$ following the structural assumption that the u_is are i.i.d. with a common distribution having mean vector $\mu(\theta)$ and covariance matrix $\Sigma_u(\theta)$. Of course, on $\mathscr{E}$ and $\mathscr{U}$ define sigma fields compatible with the above measures. Finally, define the product space $\mathscr{E}\times\mathscr{U}$ and the product probability measure $P_{\theta e}\times P_{\theta u}$ on the appropriate product sigma field.

On the probability space so constructed, columns y_i of Y under either a structural or a functional model are defined by $y_i = u_i + e_i$. In the structural model, the probability measure is defined by $P_{\theta e}\times P_{\theta u}$. In the functional model, the probability measure is defined by $P_{\theta e}$ on $\mathscr{E}$, and a fixed element $\underset{\sim}{u}$ of $\mathscr{U}$; that is, by $P_{\theta e}\times\underset{\sim}{u}$.

Let {N = 1,2,...} be a sequence of nonnegative integers. Define

$$Y_N = (y_1,\dots,y_N) = (u_1+e_1,\dots,u_N+e_N) = U_N + E_N \quad .$$

Thus, Y_N is a measurable cylinder function on our product space.

<u>Theorem 2.2.</u> Suppose that for every $\underset{\sim}{u}$ in a measurable set B with $P_{\theta u}(B) = 1$, $\{\hat{g}_N(Y_N), : N \geq 1\}$ is a strongly (weakly) consistent sequence of estimators of the matrix-valued function $g(\theta)$ of θ under the functional model, and that $P_{\theta u}(B) = 1$. Then $\{\hat{g}_N(Y_N: N \geq 1\}$ is also strongly (weakly) consistent for $g(\theta)$ under the structural model.

Proof. To prove strong consistency, let $\varepsilon > 0$ be an arbitrary positive constant and

$$A_N = \bigcup_{\ell=N}^{\infty} \{||g_\ell(Y_\ell) - g(\theta)|| > \varepsilon\},$$

where $||\cdot||$ is the Euclidean norm. Since $\{\hat{g}_N\}$ is strongly consistent for all $\underset{\sim}{u} \in B$ in the functional case,

$$\lim_{N\to\infty} [P_{\theta e} \times \underset{\sim}{u}](A_N) = 0, \text{ all } \underset{\sim}{u} \in B.$$

Consequently, by the Lebesgue dominated convergence theorem,

$$\lim_{N\to\infty} [P_{\theta e} \times P_{\theta u}](A_N) = \lim_{N\to\infty} \int_{\mathcal{U}} [P_{\theta e} \times \underset{\sim}{u}](A_N)\, dP_{\theta u}(\underset{\sim}{u}) = 0,$$

proving strong consistency in the structural case. A similar proof establishes the weak consistency result with A_N replaced by

$$A_N^* = \{||\hat{g}_N(Y_N) - g(\theta)|| > \varepsilon\}. \quad \square$$

Theorem 2.3. Suppose that there exist a sequence of constants $\{c(N)\}$ and a random matrix Z with c.d.f.

$$F(z) = P\{Z \le z, \text{ elementwise}\}$$

such that in the functional case

$$c(N)(\hat{g}_N(Y_N) - g(\theta)) \xrightarrow{\mathscr{L}} Z \qquad (2.1)$$

for all $\underset{\sim}{u} \in B$, $P_B(\underset{\sim}{u}) = 1$. Then (2.1) also holds in the structural case.

Proof. Let

$$Q_N(z) = \{c(N)(\hat{g}_N(Y_N) - g(\theta)) \leq z, \text{ elementwise}\}.$$

By the given, for all $\underset{\sim}{u} \in B$, all continuity points z of F(z),

$$\lim_{t\to\infty} [P_{\theta e} \times \underset{\sim}{u}](Q_N(z)) = F(z).$$

Now apply the Lebesgue dominated convergence theorem to show that

$$\lim_{N\to\infty} [P_{\theta e} \times P_{\theta u}](Q_N(z)) = \lim_{N\to\infty} \int_{\mathscr{U}} [P_{\theta e} \times \underset{\sim}{u}](Q_N(z))dP_{\theta u}(\underset{\sim}{u})$$

$$= \int_{\mathscr{U}} F(z)dP_{\theta u}(\underset{\sim}{u}) = F(z)$$

at all points of continuity z of F(z). $\square$

Remark 1. In most applications of Theorems 2.2 and 2.3, the set B will be

$$B = \{\underset{\sim}{u}: \lim_{N\to\infty} \mu^{(N)} = \mu(\theta),\ \lim_{N\to\infty} \Sigma_u^{(N)} = \Sigma_u(\theta)\}.$$

The assertion that $P_{\theta u}(B) = 1$ is a direct consequence of the SLLN.

As an example of the applicability of these theorems, consider the functional errors-in-variables model

$$Y = \binom{0}{\alpha}\underset{\sim}{1}_N + \binom{I_r}{B}U + E, \qquad Y: (r+s)\times N,$$

with $\Sigma_e = \sigma_e^2 I_{r+s}$. Gleser (1981) shows that the maximum likelihood estimators (MLEs) of α, $\mu = \lim_{N\to\infty} \mu^{(N)}$ and B are strongly consistent, and that

$$\tilde{\sigma}_e^2 = r^{-1}(r+s)\hat{\sigma}_e^2, \text{ where } \hat{\sigma}_e^2 = \text{MLE of } \sigma_e^2,$$

is strongly consistent for σ_e^2. He also finds a strongly consistent estimator $\tilde{\Sigma}_u$ of $\Sigma_u = \lim_{N\to\infty} \Sigma_u^{(N)}$. To do so, he assumes that the limits μ and Σ_u exist, and that $\Sigma_u > 0$. Applying Theorem 2.2. (see also Remark 1 above), it follows that these same estimators are strongly consistent in the corresponding structural model. It is worth mentioning that the MLEs of α, μ, B, and σ_e^2 are the same in both the structural and functional models (Anderson, 1982).

Gleser (1981) also finds the limiting joint distribution of the MLEs $\hat{\alpha}$, $\hat{B}$ of α, B, and of $\tilde{\sigma}_e^2$, in the functional case. Because these limiting distributions depend upon the sequence $\underset{\sim}{u} = \{u_i: i \geq 1\}$ only through the limits μ and Σ_u of $\mu^{(N)}$ and $\Sigma_u^{(N)}$, Theorem 2.3 and Remark 1 show that $(\hat{\alpha}, \hat{B}, \tilde{\sigma}_e^2)$ have the identical limiting joint distribution in the structural case.

As another example, consider replicated functional and structural models, where θ is equivalent to the parameters α, B, etc. in (1.1). Anderson (1982) shows that the MLEs of corresponding parameters in the functional and structural models are the same. Consequently, consistency and asymptotic joint normality of the MLEs in the functional case imply consistency and asymptotic joint normality in the structural case (Theorems 2.2 and 2.3). Amemiya and Fuller (1982) give separate proofs for the functional- and structural-case MLEs. Our approach, using Theorems 2.2 and 2.3, shows that only the results for the functional case require proof.

We now demonstrate some additional theoretical consequences of Theorems 2.2 and 2.3.

2.1 Consistency and Identifiability

Consider a (possibly) matrix-valued function $g(\theta)$ of $\theta \in \Theta$. We might wish to know whether in the functional errors-in-variables model, there exists a consistent (strong or weak) sequence of estimators for $g(\theta)$. Because the probability measures in the functional model depend upon the sequence $\underset{\sim}{u}$ of columns u_i, $i=1,2,\ldots$, of U, this question is often difficult to answer. However, the corresponding structural model has a finite-dimensional parameter space, and is consequently easier to handle.

Indeed, Nussbaum (1977) has pointed out that (in the case $r = s = 1$) there is a relationship between the existence of a consistent sequence of estimators for $g(\theta)$ in the functional model, and the identifiability of $g(\theta)$ in the corresponding structural model. We generalize his approach in Theorem 2.4, below.

Definition 2.1. The function $g(\theta)$ is identifiable in the structural model if $g(\theta_1) \neq g(\theta_2)$ only for identifiable $\theta_1, \theta_2 \in \Theta$.

Theorem 2.4. In order for there to exist a consistent sequence of estimators for $g(\theta)$ in a functional errors-in-variables model it is necessary that $g(\theta)$ be identifiable in the corresponding structural model.

Proof. If $\{\hat{g}_N\}$ is consistent (weakly or strongly) for $g(\theta)$ in the functional model, then since this consistency holds for all $\underset{\sim}{\mu} \in \mathcal{U}$, this sequence is consistent for $\underset{\sim}{\mu}$'s satisfying

$$\lim_{N\to\infty} \mu^{(N)} = \mu(\theta), \ \lim_{N\to\infty} \Sigma_u^{(N)} = \Sigma_u(\theta), \ \theta \in \Theta .$$

Suppose that $g(\theta)$ is not identifiable in the corresponding structural model. Then there exist θ_1, θ_2, $\in \Theta$ which are not identifiable, but for which $g(\theta_1) \neq g(\theta_2)$. Thus, by Theorem 2.2 and Remark 1, under the structural model

$$\hat{g}_N \to g(\theta_i) \text{ under } \theta_i, \ i = 1,2,$$

where the convergence "$\to$" is either in probability (weak convergence) or almost surely (strong convergence) as $N\to\infty$. However, θ_1 and θ_2 are not identifiable, and thus describe the same distribution $P_e \times P_u$ on $\mathcal{E} \times \mathcal{U}$. Consequently, the probability (or almost sure) limits of $\{\hat{g}_N\}$ must be the same under θ_1 and θ_2; that is, $g(\theta_1) = g(\theta_2)$. Since $g(\theta_1) \neq g(\theta_2)$, this is a contradiction, and the result is established. □

Nussbaum (1977) applies the argument of Theorem 2.4 in the model (1.1) where $r = s = 1$, the columns of Y are normally distributed, and $\Sigma_e = \text{diag}(\sigma_{11},\sigma_{22})$. He shows that when σ_{11}/σ_{22} is unknown, no consistent estimator of B exists in the functional model. Gleser (1981, Section 5) remarks that this argument also applies to the general model (1.1) with $r \geq 1$, $s \geq 1$, where the form of Σ_e is unspecified. To see this, note that when the columns of Y are normally distributed, the distribution of y_i, $i = 1,2,\ldots$, is determined (identified) in the structural case by α, μ and

$$\Sigma_{\underset{\sim}{y}} = \Sigma_e + \binom{I_r}{B}\Sigma_u\binom{I_r}{B}'. \tag{2.2}$$

Let

$$\Sigma_y = \begin{pmatrix} \Sigma_{y11} & \Sigma_{y12} \\ \Sigma_{y12}' & \Sigma_{y22} \end{pmatrix}, \quad \Sigma_e = \begin{pmatrix} \Sigma_{e11} & \Sigma_{e12} \\ \Sigma_{e12}' & \Sigma_{e22} \end{pmatrix}$$

where Σ_{y11} and Σ_{e11} are $r\times r$. There exists $\varepsilon > 0$ such that

$$\Sigma_{e11} = \Sigma_{y11} - \varepsilon I_r > 0,\ \Sigma_{e22} = \Sigma_{y22} - \varepsilon^3 LL' > 0$$

and

$$\Sigma_e = \begin{pmatrix} \Sigma_{y11}-\varepsilon I_r & \Sigma_{y12}-\varepsilon^2 L' \\ (\Sigma_{y12}-\varepsilon^2 L')' & \Sigma_{y22}-\varepsilon^3 LL' \end{pmatrix} > 0,$$

where $L = \underset{\sim}{1}_s \underset{\sim}{1}_r'$. Thus, Σ_y can be obtained by letting

$$\Sigma_e = \begin{pmatrix} \Sigma_{y11}-\varepsilon I_r & \Sigma_{y12}-\varepsilon^2 L' \\ (\Sigma_{y12}-\varepsilon^2 L')' & \Sigma_{y22}-\varepsilon^3 LL' \end{pmatrix},\ B = \varepsilon L,\ \Sigma_u = \varepsilon I_r.$$

On the other hand, there exists $\delta > 0$ such that

$$\Sigma_{y11} - \delta I_r > 0,\ \Sigma_{y22} - \Sigma'_{y12}(\Sigma_{y11}-\delta I_r)^{-1}\Sigma_{y12} > 0.$$

Thus, Σ_y can also be obtained by letting

$$\Sigma_e = \begin{pmatrix} \Sigma_{y11}-\delta I_r & \Sigma_{y12} \\ \Sigma'_{y12} & \delta I_s \end{pmatrix} \quad , B = (\Sigma_{y11}-\delta I_r)^{-1}\Sigma'_{y12},\ \Sigma_u = \Sigma_{y11} - \delta I_r.$$

Since εL cannot equal $(\Sigma_{y11}-\delta I_r)^{-1}\Sigma'_{y12}$ for all Σ_y, this proves that B is not identifiable in the structural model, and hence no consistent sequence of estimators for B can exist in the corresponding functional model.

Theorem 2.4 provides a way to determine if inference for $g(\theta)$ in the functional model is possible, but only in the negative sense. That is, we can sometimes show that $g(\theta)$ cannot be consistently estimated in a functional model by an appeal to the corresponding structural model. However, if $g(\theta)$ is identifiable in a structural model, Theorem 2.4 gives no guarantee that $g(\theta)$ can be consistently estimated in the corresponding functional model. (That is, identifiability may not be necessary and sufficient for consistency.)

If $g(\theta)$ is identifiable in a structural model, then <u>in this structural model</u> we can usually find a consistent sequence of estimators $\{\hat{g}_N\}$ for $g(\theta)$. Looking at the proof of Theorem 2.2, with $B = \mathcal{U}$, we see that for each $\theta \in \Theta$, consistency of $\{g_N\}$ in the structural model implies consistency of $\{g_N\}$ in the functional model, except possibly for a set $H(\theta)$ of sequences $\underset{\sim}{u} = \{u_i: i=1,2,\ldots\}$ having $P_{\theta u}$-probability equal to 0. Unfortunately, each such $H(\theta)$ need not be empty, and it is even possible that the union (over Θ) of such sets is equal to $\mathcal{U}$. Consequently, our methods do not allow us to show that $\{\hat{g}_N\}$ is consistent for $g(\theta)$ in the functional model.

Nevertheless, if we can find a reasonably regular, consistent sequence of estimators $\{\hat{g}_N\}$ for $g(\theta)$ in the structural model, the above discussion suggests that we try to directly prove that $\{\hat{g}_N\}$ is also consistent for $g(\theta)$ in the functional case. Frequently, there is additional analytic structure in the structural errors-in-variables model, such as rotational invariance of the columns of Y in the normal case, that could enable us to show that $H(\theta) = \phi$, all θ. Even so, once we see that $\{\hat{g}_N\}$ may be consistent for $g(\theta)$ in the functional model, it is usually easier to prove this fact directly.

2.2 Asymptotic Efficiency in Functional Models

Consider a structural model (1.1) for Y. Suppose that the parameterization through $\theta \in \Theta$ of this model is identifiable. It is now possible to calculate the information matrix $I(\theta)$ of the model, and to find (or establish) best asymptotic normal (BAN) estimators $\{\hat{g}_N\}$ of $g(\theta)$.

For the corresponding functional model, no such theory exists because of the incidental or nuisance parameter U, whose dimension increases with the sample size N. Indeed, no definition of an information matrix $I(\theta,\underset{\sim}{u})$ of infinite dimension seems to have been given in the literature.

However, note that

$$U = \mu^{(N)}\underset{\sim}{1}_N + (\Sigma_u^{(N)})^{\frac{1}{2}}\Gamma^{(N)} \quad , \tag{2.3}$$

where $(\Sigma_u^{(N)})^{\frac{1}{2}}$ is the symmetric square root of $\Sigma_u^{(N)}$ and

$$\Gamma^{(N)} = (\Sigma_u^{(N)})^{-\frac{1}{2}} U(I_N - N^{-1} \underset{\sim}{1}_N \underset{\sim}{1}_N') \tag{2.4}$$

is a row-orthogonal r×N matrix satisfying $\Gamma^{(N)}\underset{\sim}{1}_N = 0$. (Although $(\Sigma_u^{(N)})^{-\frac{1}{2}}$ exists only when $\Sigma_u^{(N)} > 0$, $\Gamma^{(N)}$ satisfying (2.3) and $\Gamma^{(N)}\underset{\sim}{1}_N = 0$ always exists.)

Let $\underset{\sim}{\gamma} = \{\Gamma^{(N)}: N = 1,2,\ldots\}$. Note that $(\theta,\underset{\sim}{\gamma})$ parameterizes the functional model. We restrict the parameterization and the parameter space by the following assumptions.

Assumption 1. $\lim_{N\to\infty} \mu^{(N)}(\theta) = \mu(\theta)$, $\lim_{N\to\infty} \Sigma_u^{(N)}(\theta) = \Sigma_u(\theta) > 0$, all $\theta \in \Theta$.

Assumption 2. $\underset{\sim}{\gamma}$ is functionally independent of θ.

In this case, we can "borrow" the BAN theory for estimators in structural models and construct a BAN theory for the functional model.

Theorem 2.5. Suppose that $\{\hat{g}_N\}$ is any asymptotically normal sequence of estimators for $g(\theta)$ in the functional model; that is,

$$\sqrt{N}(\hat{g}_N - g(\theta)) \xrightarrow{L} N(\underset{\sim}{0}, M(\theta)), \tag{2.5}$$

for all $(\theta,\underset{\sim}{\gamma})$. Then

$$M(\theta) - I^{-1}(\theta) \geq 0 \quad \text{in the sense of positive definiteness.}$$

Proof. By Theorem 2.3, Assumption 1, and Remark 2, (2.5) also holds for the corresponding functional model. The conclusion of the theorem now follows from BAN theory for the structural model. □

In the structural model, there are well known methods (e.g., maximum likelihood) to find BAN estimators of $g(\theta)$. Unfortunately, just as in Section 2.1, we cannot reverse (the proof of) Theorem 2.3 to assert that BAN estimators $\{\hat{g}_N\}$ for the structural model are also BAN for the corresponding functional model (under Assumptions 1 and 2). If, however, we can show that in the functional model (2.5) holds for $\{\hat{g}_N\}$ with $M(\theta) = I(\theta)$, then by Theorem 2.5 this sequence of estimators is BAN in the sense that no other asymptotically normal sequence of estimators of $g(\theta)$ can have a smaller asymptotic covariance matrix.

For example, in the model (1.1) with $\Sigma_e = \sigma^2 I_{r+s}$, normally distributed data, the MLE of the slope B is the same function of the data Y under both the functional and structural forms of the model. Since Gleser (1981) shows that the MLE of B is asymptotically normally distributed in the functional model (under Assumption 1), and this estimator is BAN in the structural model, Theorem 2.5 shows that the MLE of B is BAN in the functional model.

2.3. Summary and Discussion

The points made in this section can be simply summarized as follows:

(1) Corresponding structural and functional forms of errors-in-variables should be analyzed together.

(2) The parametric regularity of structural models make such models preferable for considering identifiability questions, problems of showing nonexistence of consistent sequences of estimators, and

asymptotic efficiency theory. It is also frequently easier to identify good classes of estimators in the context of such models.

(3) Proofs of asymptotic consistency and asymptotic distribution for sequences of estimators $\{\hat{g}_N\}$ need be given for the functional models only, since results on consistency and asymptotic distributions in functional models imply similar results in structural models (Theorems 2.2 and 2.3, Remark 1).

In connection with point (3), it should be noted that in functional models the matrix U, treated as a matrix of known constants, plays a role similar to the design matrix X in classical regression models. Consequently (Gallo, 1982a, b), existing asymptotic theory for classical regression models can be utilized in proving asymptotic consistency and asymptotic normality of estimators in functional models.

In connection with the last sentence of point (2), it is worth noting that MLEs for a functional model may not exist even when the corresponding structural model is fully identified (θ is identifiable). In this case, MLEs for the structural model (which do typically exist) can serve as good estimators in the functional model -- "good" since, as we have shown, if these estimators are asymptotically normal in the functional case, they must be BAN. This approach is utilized in the context of factor analysis by Anderson and Rubin (1956).

Finally, it is worth noting the parallelism of this discussion to both the fixed factor (Model I)-random factor (Model II) categorizations of ANOVA, and to the compound decision-empirical Bayes statistical decision-theoretic formulations of Robbins (1951, 1955). The structural form of (1.1) is an empirical Bayes reformulation of the corresponding functional model. However, by imposing Assumption 1 on the parameters of the functional model, we remove the need for a random mechanism to generate the u_i's, requiring only (in the spirit of compound decision models) that the sequence $\underset{\sim}{u} = \{u_i: i=1,2,\ldots\}$ of columns of U have some of the properties of i.i.d. sequences of random $r\times 1$ vectors.

3. ULTRASTRUCTURAL MODELS

Dolby's ultrastructural model (1.4) can be rewritten as follows:

$$y_{ij} = \binom{0}{\alpha} + \binom{1}{b}\mu_i + f_{ij}, \quad 1 \leq i \leq m, \; 1 \leq j \leq n, \tag{3.1}$$

where

$$f_{ij} = \binom{1}{b}(u_{ij}-\mu_i) + e_{ij} \quad \text{are i.i.d. } N(\underset{\sim}{0},\Sigma_f)$$

and

$$\Sigma_f = \begin{pmatrix} \sigma_{f11} & \sigma_{f12} \\ \sigma_{f12} & \sigma_{f22} \end{pmatrix} = \begin{pmatrix} \sigma_{e1}^2+\sigma_u^2 & b\sigma_u^2 \\ b\sigma_u^2 & \sigma_{e2}^2+b^2\sigma_u^2 \end{pmatrix} . \tag{3.2}$$

As σ_{e1}^2, σ_{e2}^2, σ_u^2 and b range over their possible values, Σ_f can equal any 2×2 positive definite matrix. However, the requirement that σ_{e1}^2, σ_{e2}^2 and σ_u^2 be nonnegative places the following restrictions on Σ_f and b:

$$|\sigma_{f11}^{-1}\sigma_{f12}| \leq |b| \leq (|\sigma_{f22}^{-1}\sigma_{f12}|)^{-1}, \quad b\sigma_{f12} \geq 0. \tag{3.3}$$

Except for (3.3), when $n > 1$ the model (3.1) is recognizable as an equally-replicated functional model. The requirement (3.3) doesn't change the parameterization of the model, but does limit the parameter values to a subspace (of the same dimension). Consequently, if the MLEs for b and Σ_f in this model satisfy (3.3), they are also MLEs for the parameters of Dolby's ultrastructural model (1.4). (This is essentially the approach of Cox (1976).)

A model equivalent to (3.1) is treated by Anderson (1951). Anderson's model is (in our notation)

$$y_{ij} = \gamma + \xi_i + f_{ij}, \qquad (3.4)$$

where $\sum_{i=1}^{m} \xi_i = 0$, and there exists $\lambda = (\lambda_1,\lambda_2)'$ such that

$$\lambda'\xi_i = 0,\ 1 \le i \le m. \qquad (3.5)$$

It is easy to see that the following correspondences hold:

$$\gamma = \binom{0}{\alpha} + \bar{\mu}. \quad , \quad \xi_i = \mu_i - \bar{\mu}. \quad , \quad \lambda = c\binom{-b}{1}, \qquad (3.6)$$

where $\bar{\mu}. = m^{-1} \sum_{i=1}^{m} \mu_i$, and c is an arbitrary non-zero constant. The relationship of the models (1.4) and (3.1) to the model (3.4) appears to have been overlooked by both Dolby (1976) and Cox (1976).

Let

$$W = m^{-1} \sum_{i=1}^{m} (\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{i.} - \bar{y}_{..})', \quad S = (nm)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij}-\bar{y}_{i.})(y_{ij}-\bar{y}_{i.})',$$

where $\bar{y}_{i.} = n^{-1} \sum_{j=1}^{n} y_{ij}$, $\bar{y}_{..} = (nm)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}$. Let $t_1 > t_2$ be the ordered roots of the determinantal equation

$$|W - tS| = 0$$

and let g_1, g_2: 2×1 satisfy

$$Wg_i = t_i S g_i, \quad g_i' W g_i = 1, \qquad i = 1,2.$$

Then Anderson (1951, 1982) shows that the MLEs of the parameters of the model (3.4) are

$$\hat{\lambda} = \begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = g_2, \quad \hat{\xi}_i = S g_1 g_1' (\bar{y}_{i.} - \bar{y}_{..}), \quad 1 \le i \le m,$$
$$\hat{\gamma} = S g_1 g_1' \bar{y}_{..}, \quad \hat{\Sigma}_f = S + t_2 S g_2 g_2' S. \tag{3.7}$$

Consequently, the MLE of the unrestricted model (3.1) are $\hat{\Sigma}_f$ and

$$\hat{b} = -\frac{\hat{\lambda}_1}{\hat{\lambda}_2}, \ \hat{\alpha} = \binom{-\hat{b}}{1}\hat{\gamma}, \ \mu_i = S g_1 g_1' (\bar{y}_{i.}) - \binom{0}{\hat{\alpha}}, \ 1 \le i \le m. \tag{3.8}$$

Let

$$\hat{\Sigma}_f = \begin{pmatrix} \hat{\sigma}_{f11} & \hat{\sigma}_{f12} \\ \hat{\sigma}_{f12} & \hat{\sigma}_{f22} \end{pmatrix} .$$

If

$$|\hat{\sigma}_{f11}^{-1} \hat{\sigma}_{f12}| \le |\hat{b}| \le (|\hat{\sigma}_{f22}^{-1} \hat{\sigma}_{f12}|)^{-1}, \quad \hat{b}\hat{\sigma}_{f12} \ge 0, \tag{3.9}$$

then the MLEs of α, b, μ_i in Dolby's model are given by (3.7), (3.8), and

$$\hat{\sigma}_u^2 = \hat{b}^{-1}\hat{\sigma}_{f12},\ \hat{\sigma}_{e1}^2 = \hat{\sigma}_{f11} - \hat{b}^{-1}\hat{\sigma}_{f12},\ \hat{\sigma}_{e2}^2 = \hat{\sigma}_{f22} - \hat{b}\hat{\sigma}_{f12}. \qquad (3.10)$$

are the MLEs of σ_u^2, σ_{e1}^2, σ_{e2}^2, respectively.

If (3.9) fails to hold, then the supremum of the likelihood occurs at one of the three boundaries of the parameter space (defined by $\sigma_u^2 = 0$, by $\sigma_{e1}^2 = 0$, and by $\sigma_{e2}^2 = 0$). The boundary defined by $\sigma_u^2 = 0$ is part of the parameter space of Dolby's model. Cox (1976) shows how to obtain the maximum of the likelihood on this boundary (and also on the boundaries $\sigma_{e1}^2 = 0$, $\sigma_{e2}^2 = 0$), but recommends ignoring the $\sigma_u^2 = 0$ boundary as an approximation to the correct maximum likelihood procedure. (On the $\sigma_u^2 = 0$ boundary, $\hat{b}$ is the solution of a quartic equation.) This approximation makes sense in Cox's approach (where the model is regarded as a replicated structural model, and $\sigma_u^2 > 0$ is assumed), but is not acceptable from the ultrastructural viewpoint of Dolby (where $\sigma_u^2 = 0$ means that there is no structural component to the model).

When $\sigma_u^2 > 0$ (and $\sigma_{e1}^2 > 0$, $\sigma_{e2}^2 > 0$), (3.9) holds with probability tending to one as $n \to \infty$ (m fixed). This assertion follows from the consistency of the MLEs of b and Σ_f in the unrestricted model (3.1), as shown by Anderson (1951) and also Healy (1980). Consequently, the MLEs for b, σ_u^2, σ_{e1}^2, σ_{e2}^2 in the ultrastructural model are consistent estimators of these parameters. The MLEs of α and μ_i are also consistent, by the same argument. The joint asymptotic normality of the MLEs in the unrestricted model (3.1) has been shown by Amemiya and Fuller (1982) for both normal and nonnormal data; for normal data, the asymptotic covariance matrix of these estimators has been calculated by several authors. Because (3.9) holds with probability tending to one as $n \to \infty$, these large sample results can be used to show that $\hat{\alpha}$, $\hat{\mu}_i$, $1 \leq i \leq m$, $\hat{b}$, $\hat{\sigma}_u^2$, $\hat{\sigma}_{e1}^2$, $\hat{\sigma}_{e2}^2$ are jointly asymptotically normal as $n \to \infty$, even for nonnormal data, and the asymptotic covariance matrix of these estimators can be calculated. The asymptotic variances are obtained by Cox (1976). As Cox (1976) notes, the accuracy of the large-sample distributional approximation depends upon (3.9) holding with high probability for the given replication size n (and number of groups m).

It should be noted that the magnitude of m relative to that of n cannot be ignored in determining whether n is large enough to permit asymptotic distributions to be used as approximations to the exact distributions of the MLEs. Healy (1980) considers Anderson's (1951) replicated functional models, but allows the number of groups m to go to infinity. In the present context (Healy's results are more general), if $m \to \infty$ and n stays fixed, $n > 1$. Healy's results show that $\hat{b}$ is consistent but $\hat{\Sigma}_f$ is <u>not</u> consistent. Whether, (3.9) continues to hold with probability tending to one as $m \to \infty$ (n fixed) is thus an open question. If the probability that (3.9) holds fails to converge to one, then the MLEs in the models (1.4) and (3.1) are not necessarily asymptotically equivalent for distributional purposes.

The method of approach used here is one that is frequently useful: embed a given errors-in-variables model in a similarly parameterized model whose parameter space is less restricted, and whose MLEs are either known, or easily determined. If this approach is successful, one can also apply consistency and asymptotic distribution results for the MLEs of the broader model to obtain similar large-sample results for the MLEs of the given model. However, in using this approach care must be taken to make sure that the two parameter spaces have the same dimension, differing only in that the parameter space of the given model is a open subset of the parameter space of the broader model. Although this approach is mentioned in popular mathematical statistics textbooks [e.g., Bickel and Doksum (1977)], it is frequently overlooked, with a resulting expenditure of unnecessary effort.

Dolby (1976) also considers the unreplicated case (n=1) of the ultra-structural model (1.4), noting that without additional restrictions on the parameters, the formal solutions of the likelihood equaitons will be unsatisfactory as estimators. This fact is not surprising since as Dolby notes (quoting Solari (1969)), these formal solutions define saddlepoints of the likelihood. Indeed, using the general approach outlined in Section 2.1, it is easy to show that the structural model corresponding to (3.1), where the μ_i's are i.i.d. $N(\mu,\sigma^2)$, is not identified. More specifically, in this structural model none of the parameters σ_u^2, σ_{e1}^2, σ_{e2}^2 and b are identifiable. Consequently, in the model (1.4) with n = 1, <u>no consistent</u> (as $m \to \infty$) <u>estimators of</u> σ_u^2, σ_{e1}^2, σ_{e2}^2, and b <u>can exist.</u> In this sense, the unreplicated ultrastructural model is nonidentifiable.

Dolby imposes the conditions

$$\frac{\sigma_{e2}^2}{\sigma_{e1}^2} = k_1, \quad \frac{\sigma_u^2}{\sigma_{e1}^2} = k_2,$$

where k_1 and k_2 are known nonnegative constants. In this case, Σ_f in the corresponding model (3.1) becomes

$$\Sigma_f = \sigma_{e1}^2 \begin{pmatrix} 1+k_2 & bk_2 \\ bk_2 & k_1+b^2k_2 \end{pmatrix} \quad , \tag{3.11}$$

and the restrictions (3.3) are no longer needed. It is interesting to note that the model (3.1) with Σ_f defined by (3.11) generalizes the model in Gleser (1981). Gleser (1981) treated the case

$$\Sigma_f = \sigma^2 \Sigma_0, \ \Sigma_0 \text{ known}.$$

In contrast, (3.10) has the form

$$\Sigma_f = \sigma_{e1}^2 [\begin{pmatrix} 1 & 0 \\ 0 & k_1 \end{pmatrix} + k_2 \binom{1}{b} \binom{1}{b}'].$$

Dolby's MLEs for α, b, and σ_{e1}^2 in his model, yield MLEs for α, b, and Σ_f in the model (3.1), (3.10), and vice versa. Unfortunately, Dolby's MLE for b involves solution of a quintic polynomial equation. It is possible that Gleser's (1981) algebraic approach to finding MLEs in functional models may either yield simpler formulas, or show that only one of the roots of the quintic polynomial is an acceptable solution. Work on this question is in progress.

REFERENCES

Amemiya, Y. and Fuller, W. (1982). Estimation for the multivariate errors-in-variables model with estimated error covariance matrix. Unpublished.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. Ann. Math. Statist. 22, 327-351. (Correction, Ann. Statist. 8 (1980), 1400.)

Anderson, T. W. (1982). Estimating linear statistical relationships (1982 Abraham Wald Memorial Lectures). Technical Report No. 389, Instit. for Studies in the Soc. Sci., Stanford Univ. To appear in Ann. Statist.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. Proc. Third Berkeley Symp. Math. Statist. and Prob. (Jerzy Neyman, ed.), University of California Press 5, 111-150.

Bickel, Peter J. and Doksum, Kjell A. (1977). Mathematical Statistics. Holden-Day: San Francisco.

Cox, N. R. (1976). The linear structural relation for several groups of data. Biometrika 63, 231-237.

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

Gallo, Paul P. (1982a). Properties of estimators in errors-in-variables regression models. Ph.D. dissertation, University of North Carolina at Chapel Hill.

Gallo, Paul P. (1982b). Consistency of regression estimates when some variables are subject to error. Comm. in Statist., Ser. A 11, 973-983.

Gleser, Leon Jay (1981). Estimation in a multivariate "errors in variables" regression model: Large sample results. Ann. Statist. 9, 24-44.

Healy, John D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. J. Multivar. Anal. 10, 243-251.

Kendall, M.G. and Stuart, A. (1979). The Advanced Theory of Statistics (vol. 2, 4th ed.). Macmillan: New York.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. Ann. Math. Statist. 27, 887-906.

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. J. Amer. Statist. Assoc. 54, 173-205.

Moran, P. A. P. (1971). Estimating structural and functional relationshps. J. Multivar. Anal. 1, 232-255.

Nussbaum, N. (1977). Asymptotic optimality of estimators of a linear functional relation if the ratio of the error variances is known. Math. Op. Statist. Ser. Statist. 8, 173-198.

Robbins, H. (1951). Asymptotic subminimax solutions of compound statistical decision problems. Proc. Second Berkeley Symp. Math. Statist. and Prob. Univ. California Press, 131-148.

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. Third Berkeley Symp. Math. Statist. and Prob. U. California Press 5, 157-163.