The Annals of Statistics
1984, Vol. 12, No. 2, 497–509

ESTIMATION FOR THE MULTIVARIATE ERRORS-IN-VARIABLES MODEL WITH ESTIMATED ERROR COVARIANCE MATRIX[1]

By Yasuo Amemiya and Wayne A. Fuller

Iowa State University

The errors-in-variables model in which the unobserved true values satisfy multiple linear restrictions is considered. Under the assumptions that the unobservable true values are normally distributed and that an estimator of the covariance matrix of the measurement error is available, the maximum likelihood estimators are derived. The limiting properties of the estimators are obtained for a wide range of assumptions, including the assumption of fixed true values.

1. Introduction. In the errors-in-variables model, the true values of a set of variables satisfy exact relationships. Inference is based on the observed values which are the sums of the true values and errors of measurement. To define the model, let a set of r-dimensional row vectors $\mathbf{y}_t$ and a set of k-dimensional row vectors $\mathbf{x}_t$ satisfy

$$\mathbf{y}_t = \beta_0 + \mathbf{x}_t\beta, \quad t = 1, 2, \cdots, n, \tag{1.1}$$

where β_0 is a $1 \times r$ vector of parameters and β is a $k \times r$ matrix of parameters. We observe $\mathbf{Y}_t$ and $\mathbf{X}_t$, which satisfy

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{y}_t + \mathbf{e}_t, \\ \mathbf{X}_t &= \mathbf{x}_t + \mathbf{u}_t, \quad t = 1, 2, \cdots, n, \end{aligned} \tag{1.2}$$

where $\mathbf{e}_t$ and $\mathbf{u}_t$ are unobservable error vectors of dimensions r and k, respectively. Let $\varepsilon_t = (\mathbf{e}_t, \mathbf{u}_t)$, and let $p = r + k$. The ε_t are assumed to be independently and identically distributed with mean zero and covariance matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon}$. Assume that the ε_t are independent of the $\mathbf{x}_j$ for all t and j. Equations (1.1) and (1.2) and the associated assumptions define the multivariate errors-in-variables model.

If the $\mathbf{x}_t$ are constant vectors, the model is called a functional model. If the $\mathbf{x}_t$ are independently and identically distributed random vectors, the model is called a structural model. In the functional model, the $\mathbf{x}_t$, $t = 1, 2, \cdots, n$, are incidental parameters, which enter the distribution of only finitely many observations as $n \to \infty$.

Most research on the multivariate structural model has concentrated on the factor analysis model which assumes the error covariance matrix $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ to be diagonal.

Received September 1982; revised August 1983.

[1] This research was partly supported by the U.S. Bureau of the Census through Joint Statistical Agreement J.S.A. 82-6.

AMS 1980 *subject classifications*. Primary 62J99, 62H12; secondary 62H25, 62F10, 62F12.

Key words and phrases. Errors-in-variables, structural relationship, functional relationship, multivariate regression model, maximum likelihood estimator, asymptotic distribution, measurement errors.

The maximum likelihood estimators for the factor model are discussed in Lawley (1940, 1941, 1943, 1967, 1976), Anderson and Rubin (1956), Jöreskog (1967), and Jennrich and Thayer (1973). Other work on the multivariate structural model includes Jöreskog (1970), Jöreskog and Goldberger (1972), Browne (1974), Theobald (1975), and Anderson (1984).

The estimation for the multivariate functional model was first discussed by Tintner (1945) and Geary (1948). Anderson (1951a) derived the maximum likelihood estimators for the replicated functional model with unknown $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ in the context of estimating linear restrictions on the coefficients of the multivariate regression model. Gleser (1981) gave the limiting distribution of the maximum likelihood estimators for the functional model with $\mathbf{\Sigma}_{\varepsilon\varepsilon} = \sigma^2 \mathbf{I}$. Gleser also showed that the estimators could be obtained by minimizing certain norms. Dahm and Fuller (1981) applied the generalized least squares method to the functional model. Other literature on the multivariate functional model includes Anderson (1951b), Whittle (1952), Anderson and Rubin (1956), Gleser and Watson (1973), Theobald (1975), Höschel (1978), Nussbaum (1979), Chan (1980), Healy (1980), Villegas (1982), Chan and Mak (1983), and Anderson (1984).

We consider the estimation of the multivariate model (1.1) and (1.2) for both the functional and structural cases, when there is available an estimator $\mathbf{S}_{\varepsilon\varepsilon}$ of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$. The model with intercept term β_0 is chosen because of its wider use in practice. We assume that $\mathbf{S}_{\varepsilon\varepsilon}$ is an estimator of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ rather than of a multiple of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ because we consider this to be the usual case. There are two common sources for $\mathbf{S}_{\varepsilon\varepsilon}$. Independent experiments in the past often provide such estimators. Also, when replicated observations are measured at some of the true values $(\mathbf{y}_t, \mathbf{x}_t)$, the within replicates mean squares matrix can be used as an estimator of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$. With normality of the errors, the means over replicates used as the data points $(\mathbf{Y}_t, \mathbf{X}_t)$ are independent of the estimator of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ based on the within sum of squares.

We derive the maximum likelihood estimators for the multivariate structural model. The large sample properties of the estimators are obtained under weak assumptions on ε_t for a wide range of assumptions on $\mathbf{x}_t$.

2. The maximum likelihood estimator. Let the model (1.1) and (1.2) hold and assume that there is available an estimator $\mathbf{S}_{\varepsilon\varepsilon}$ of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$. In this section we derive the maximum likelihood estimators of the parameters in the model for the normal structural case.

First, we introduce some notation. Let $\mathbf{z}_t = (\mathbf{y}_t, \mathbf{x}_t)$, and $\mathbf{Z}_t = (\mathbf{Y}_t, \mathbf{X}_t)$, and define the statistics

$$\overline{\mathbf{Z}} = n^{-1} \textstyle\sum_{t=1}^{n} \mathbf{Z}_t = (\overline{\mathbf{Y}}, \overline{\mathbf{X}}),$$

$$\mathbf{m}_{ZZ} = (n-1)^{-1} \textstyle\sum_{t=1}^{n} (\mathbf{Z}_t - \overline{\mathbf{Z}})'(\mathbf{Z}_t - \overline{\mathbf{Z}}) = \begin{pmatrix} \mathbf{m}_{YY} & \mathbf{m}_{YX} \\ \mathbf{m}_{XY} & \mathbf{m}_{XX} \end{pmatrix},$$

with analogous definitions holding for $\bar{\mathbf{z}}$ and $\mathbf{m}_{zz}$. Throughout this paper, $\mathbf{m}_{ab}$ denotes the corrected cross product matrix with divisor $(n-1)$ for any sequences of vectors $\mathbf{a}_t$ and $\mathbf{b}_t$, $t = 1, 2, \cdots, n$.

We assume that $(\mathbf{x}_t, \boldsymbol{\varepsilon}_t)'$ are independently and identically distributed as normal vectors with mean $(\boldsymbol{\mu}_x, \mathbf{0})'$ and covariance matrix block diag$\{\boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon}\}$, that $\mathbf{S}_{\varepsilon\varepsilon}$ is independent of $\mathbf{Z}_t$, and that $d\ \mathbf{S}_{\varepsilon\varepsilon}$ is distributed as a Wishart matrix with parameter $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and degrees of freedom d. Because it is more common to work with the unbiased sample covariance matrix $\mathbf{m}_{ZZ}$ constructed with a divisor $(n-1)$, we define a slightly modified log likelihood function

$$\begin{aligned} &\log L \\ (2.1)\qquad &= C_0 - 2^{-1}(n-1)[\log|\boldsymbol{\Sigma}_{ZZ}| + \operatorname{tr}(\mathbf{m}_{ZZ}\boldsymbol{\Sigma}_{ZZ}^{-1}) + (\overline{\mathbf{Z}} - \boldsymbol{\mu}_Z)\boldsymbol{\Sigma}_{ZZ}^{-1}(\overline{\mathbf{Z}} - \boldsymbol{\mu}_Z)'] \\ &\quad - 2^{-1}d[\log|\boldsymbol{\Sigma}_{\varepsilon\varepsilon}| + \operatorname{tr}(\mathbf{S}_{\varepsilon\varepsilon}\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1})], \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{ZZ} = (\boldsymbol{\beta}, \mathbf{I})'\boldsymbol{\Sigma}_{xx}(\boldsymbol{\beta}, \mathbf{I}) + \boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \boldsymbol{\Sigma}_{zz} + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}, \quad \boldsymbol{\mu}_Z \doteq (\boldsymbol{\beta}_0, \mathbf{0}) + \boldsymbol{\mu}_x(\boldsymbol{\beta}, \mathbf{I}),$$

and C_0 is a constant. We call the function (2.1) the log likelihood adjusted for degrees of freedom. Theorem 1 presents the maximum likelihood estimators adjusted for degrees of freedom, that is, the values which maximize (2.1). The maximum likelihood estimators can be obtained from the results of Theorem 1 by replacing $\mathbf{m}_{ZZ}$ with $n^{-1}(n-1)\mathbf{m}_{ZZ}$.

THEOREM 1. *Let the model* (1.1) *and* (1.2) *hold. Assume that*

$$\begin{pmatrix} \mathbf{x}_t' \\ \boldsymbol{\varepsilon}_t' \end{pmatrix} \sim NI\left(\begin{pmatrix} \boldsymbol{\mu}_x' \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \end{pmatrix}\right),$$

where $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ are positive definite. Also, assume that $\mathbf{S}_{\varepsilon\varepsilon}$ is independent of $\mathbf{Z}_t$ for all t and that the distribution of d $\mathbf{S}_{\varepsilon\varepsilon}$ is the Wishart distribution with parameter $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ and degrees of freedom d. Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$ be the eigenvalues of $\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}$ and let $\mathbf{Q} = (\mathbf{Q}_1, \mathbf{Q}_2)$ be the matrix of the corresponding orthonormal eigenvectors such that

$$\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{Q}_j = \mathbf{Q}_j\hat{\boldsymbol{\Lambda}}_j, \quad j = 1, 2,$$

where

$$\hat{\boldsymbol{\Lambda}}_1 = \operatorname{diag}\{\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_k\}, \quad \hat{\boldsymbol{\Lambda}}_2 = \operatorname{diag}\{\hat{\lambda}_{k+1}, \hat{\lambda}_{k+2}, \cdots, \hat{\lambda}_p\}.$$

If $\hat{\lambda}_k \leq 1$, the maximum likelihood estimators adjusted for degrees of freedom do not exist. If $\hat{\lambda}_k > 1$, the maximum likelihood estimators adjusted for degrees of freedom are

$$\begin{aligned} (\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\mu}}_x) &= (\overline{\mathbf{Y}} - \overline{\mathbf{X}}\,\hat{\boldsymbol{\beta}}, \overline{\mathbf{X}}), \quad \hat{\boldsymbol{\beta}} = (\mathbf{P}_{kk}')^{-1}\mathbf{P}_{rk}' = -\mathbf{T}_{kr}\mathbf{T}_{rr}^{-1}, \\ \hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} &= (n-1+d)^{-1}[(n-1)(\mathbf{m}_{ZZ} - \hat{\boldsymbol{\Sigma}}_{zz}) + d\ \mathbf{S}_{\varepsilon\varepsilon}], \\ \hat{\boldsymbol{\Sigma}}_{xx} &= \mathbf{P}_{kk}(\hat{\boldsymbol{\Lambda}}_1 - \mathbf{I})\mathbf{P}_{kk}', \end{aligned}$$

where

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{zz} &= (\hat{\boldsymbol{\beta}}, \mathbf{I})'\ \hat{\boldsymbol{\Sigma}}_{xx}(\hat{\boldsymbol{\beta}}, \mathbf{I}) = \mathbf{P}_1(\hat{\boldsymbol{\Lambda}}_1 - \mathbf{I})\mathbf{P}_1', \\ \mathbf{P}_1 &= \mathbf{S}_{\varepsilon\varepsilon}^{1/2}\mathbf{Q}_1 = (\mathbf{P}_{rk}', \mathbf{P}_{kk}')', \quad \mathbf{T}_2 = \mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{Q}_2 = (\mathbf{T}_{rr}', \mathbf{T}_{kr}')'. \end{aligned}$$

PROOF. Because $\hat{\mu}_Z = \bar{\mathbf{Z}}$ maximizes (2.1) with respect to μ_Z, the results for β_0 and μ_x are immediate. Using the matrix $\mathbf{T} = \mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{Q}$, we define new parameters

$$\mathbf{\Sigma}_{T\varepsilon\varepsilon} = \mathbf{T}'\mathbf{\Sigma}_{\varepsilon\varepsilon}\mathbf{T}, \quad \mathbf{\Sigma}_{TZZ} = \mathbf{T}'\mathbf{\Sigma}_{ZZ}\mathbf{T} = \mathbf{A}\,\mathbf{A}' + \mathbf{\Sigma}_{T\varepsilon\varepsilon},$$

where $\mathbf{A}$ is a $p \times k$ matrix satisfying

$$\mathbf{A}\,\mathbf{A}' = \mathbf{\Sigma}_{Tzz} = \mathbf{T}'\mathbf{\Sigma}_{zz}\mathbf{T}.$$

Since only $[kr + 2^{-1}k(k+1)]$ elements in $\mathbf{A}$ are free, we impose a restriction

$$\mathbf{A}'\hat{\mathbf{\Lambda}}^{-1}\mathbf{A} = \mathbf{D}, \tag{2.2}$$

where $\hat{\mathbf{\Lambda}} = \text{block diag}\{\hat{\mathbf{\Lambda}}_1, \hat{\mathbf{\Lambda}}_2\}$, and $\mathbf{D}$ is a diagonal matrix with free diagonal elements. If the log likelihood (2.1) is expressed in terms of the tranformed variables $\mathbf{Z}_t\mathbf{T}$ and the new parameters, maximizing (2.1) is equivalent to minimizing

$$f(\theta) = (n-1)[\log|\mathbf{\Sigma}_{TZZ}| + \text{tr}(\hat{\mathbf{\Lambda}}\,\mathbf{\Sigma}_{TZZ}^{-1})] + d[\log|\mathbf{\Sigma}_{T\varepsilon\varepsilon}| + \text{tr}(\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1})], \tag{2.3}$$

where θ contains the elements of $\mathbf{A}$ and the distinct elements of $\mathbf{\Sigma}_{T\varepsilon\varepsilon}$. Setting the derivatives of $f(\theta)$ with respect to θ equal to zero, we obtain the necessary conditions for critical points of $f(\theta)$:

$$(\mathbf{\Sigma}_{TZZ} - \hat{\mathbf{\Lambda}})\mathbf{\Sigma}_{TZZ}^{-1}\mathbf{A} = \mathbf{0}, \tag{2.4}$$

$$(n-1)\mathbf{\Sigma}_{TZZ}^{-1}(\mathbf{\Sigma}_{TZZ} - \hat{\mathbf{\Lambda}})\mathbf{\Sigma}_{TZZ}^{-1} + d\,\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1}(\mathbf{\Sigma}_{T\varepsilon\varepsilon} - \mathbf{I})\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1} = \mathbf{0}. \tag{2.5}$$

By (2.2) and (2.4),

$$\mathbf{A} = \mathbf{\Sigma}_{TZZ}\hat{\mathbf{\Lambda}}^{-1}\mathbf{A} = \mathbf{A}\mathbf{D} + \mathbf{\Sigma}_{T\varepsilon\varepsilon}\hat{\mathbf{\Lambda}}^{-1}\mathbf{A}$$

and

$$\hat{\mathbf{\Lambda}}^{1/2}\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1}\hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{\Lambda}}^{-1/2}\mathbf{A} = \hat{\mathbf{\Lambda}}^{-1/2}\mathbf{A}\mathbf{\Gamma}_1,$$

where $\mathbf{\Gamma}_1 = (\mathbf{I} - \mathbf{D})^{-1} = \text{diag}\{\gamma_1, \gamma_2, \cdots, \gamma_k\}$ is a diagonal matrix. Therefore, $\hat{\mathbf{\Lambda}}^{-1/2}\mathbf{A}_{\cdot i}$, where $\mathbf{A}_{\cdot i}$ is the ith column of $\mathbf{A}$, is the eigenvector of $\hat{\mathbf{\Lambda}}^{1/2}\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1}\hat{\mathbf{\Lambda}}^{1/2}$ corresponding to the eigenvalue γ_i, for $i = 1, 2, \cdots, k$. Let $\mathbf{\Gamma} = \text{block diag}\{\mathbf{\Gamma}_1, \mathbf{\Gamma}_2\}$ be the diagonal matrix of the eigenvalues of $\hat{\mathbf{\Lambda}}^{1/2}\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1}\hat{\mathbf{\Lambda}}^{1/2}$, where $\mathbf{\Gamma}_1$ contains the k roots corresponding to $\hat{\mathbf{\Lambda}}^{-1/2}\mathbf{A}$. Let $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ be the corresponding matrix of orthonormal eigenvectors of $\hat{\mathbf{\Lambda}}^{1/2}\mathbf{\Sigma}_{T\varepsilon\varepsilon}^{-1}\hat{\mathbf{\Lambda}}^{1/2}$. Then, by (2.2),

$$\mathbf{A} = \hat{\mathbf{\Lambda}}^{1/2}\mathbf{H}_1(\mathbf{I} - \mathbf{\Gamma}_1^{-1})^{1/2}, \tag{2.6}$$

provided the k roots chosen for $\mathbf{\Gamma}_1$ are all greater than unity. Substituting (2.6) into (2.5), we have

$$\mathbf{H}'\hat{\mathbf{\Lambda}}\mathbf{H} = \text{block diag}\{\mathbf{\Gamma}_1, d[(n-1+d)\mathbf{\Gamma}_2^{-1} - (n-1)\mathbf{I}]^{-1}\} = \mathbf{\Lambda}^*. \tag{2.7}$$

Since $\mathbf{\Lambda}^*$ and $\hat{\mathbf{\Lambda}}$ are diagonal matrices and $\mathbf{H}$ is orthogonal, it follows that $\mathbf{H}$ is a permutation matrix and that diagonals of Λ^* and $\hat{\Lambda}$ are composed of the same p elements. Let $\lambda_1^0, \lambda_2^0, \cdots, \lambda_k^0$ be the roots $\hat{\lambda}_i$ chosen for $\mathbf{\Gamma}_1$, where $\lambda_i^0 > 1$, $i = 1, 2, \cdots, k$, and let $\hat{\mathbf{\Lambda}}_{10} = \text{diag}\{\lambda_1^0, \lambda_2^0, \cdots, \lambda_k^0\}$. Then, it follows from (2.6) and

(2.7) that the critical points of $f(\theta)$ satisfying (2.2) have the form:

$$\hat{\mathbf{A}} = \hat{\mathbf{\Lambda}}^{1/2}\mathbf{H}_1(\mathbf{I} - \hat{\mathbf{\Lambda}}_{10}^{-1})^{1/2}, \quad \hat{\mathbf{\Sigma}}_{T\varepsilon\varepsilon} = \operatorname{diag}\{\delta_1, \delta_2, \cdots, \delta_p\}, \tag{2.8}$$

where $\delta_i = 1$ if $\hat{\lambda}_i$ is in the chosen set $\{\lambda_i^0\}_{i=1}^k$, and $\delta_i = (n - 1 + d)^{-1}[d + (n-1)\hat{\lambda}_i]$ if $\hat{\lambda}_i$ is not in the set. If $\hat{\lambda}_k \le 1$, the function $f(\theta)$ has no critical point on the open parameter space for θ and the maximum likelihood estimators adjusted for degrees of freedom do not exist. Now, assume $\hat{\lambda}_k > 1$ so that $\lambda_i^0 > 1$ for $i = 1, 2, \cdots, k$. It can be shown that the function (2.3) evaluated at (2.8) is

$$C_1 + \textstyle\sum_{i=1}^k g(\lambda_i^0),$$

where C_1 is a constant free of k and of the choice of λ_i^0, and

$$g(w) = (n-1)\log w - (n - 1 + d)\log\{(n - 1 + d)^{-1}[(n-1)w + d]\}. \tag{2.9}$$

The function $g(w)$ is monotone decreasing for $w > 1$ and $g(1) = 0$. If we let $\hat{\theta}$ be the critical point (2.8) with the k largest $\hat{\lambda}_i$ chosen for the set $\{\lambda_i^0\}_{i=1}^k$, then $\hat{\theta}$ gives the minimum of $f(\theta)$ among all critical points (2.8). The parameter space Ω for θ consists of all symmetric positive definite matrices $\mathbf{\Sigma}_{T\varepsilon\varepsilon}$ and all $p \times k$ matrices $\mathbf{A}$ of rank k. We note that $f(\theta) \to \infty$ if $\mathbf{\Sigma}_{T\varepsilon\varepsilon}$ approaches a singular matrix or if one or more elements of $(\mathbf{\Sigma}_{T\varepsilon\varepsilon}, \mathbf{A})$ approaches infinity in absolute value. See Anderson (1958, page 47). The boundary of Ω associated with an $\mathbf{A}$ of rank less than k is the union of k disjoint sets Ω_j, $0 \le j \le k-1$, where on Ω_j, $f(\theta)$ is a function of a symmetric positive definite $\mathbf{\Sigma}_{T\varepsilon\varepsilon}$ and a $p \times j$ matrix $\mathbf{A}$ of rank j. By applying the above argument to the parameter space Ω_j, we find that $[C_1 + \sum_{i=1}^j g(\hat{\lambda}_i)]$ is the minimum of $f(\theta)$ among all critical points on Ω_j, $1 \le j \le k-1$. On Ω_j, $f(\theta) \to \infty$ if $\mathbf{\Sigma}_{T\varepsilon\varepsilon}$ approaches a singular matrix or if one or more elements of $(\mathbf{\Sigma}_{T\varepsilon\varepsilon}, \mathbf{A}_j)$ approaches infinity in absolute value, and the boundary of Ω_j associated with an $\mathbf{A}_j$ of rank less than j is the union of Ω_ℓ, $0 \le \ell \le j-1$. On Ω_0, the absolute minimum of $f(\theta)$ is C_1. Since $g(w) < 0$ for $w > 1$, the absolute minimum of $f(\theta)$ on the union of Ω_j, $0 \le j \le k-1$ is $[C_1 + \sum_{i=1}^{k-1} g(\hat{\lambda}_i)]$. Hence, $f(\hat{\theta}) = [C_1 + \sum_{i=1}^k g(\hat{\lambda}_i)]$ is the absolute minimum of $f(\theta)$ on Ω. The expressions for the estimators follow by transforming $\hat{\theta}$ to the original parameterization. □

Anderson (1946, 1984) gave the maximum likelihood estimators for $\mathbf{\Sigma}_{zz}$ and $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ in the structural model where the rank of $\mathbf{\Sigma}_{zz}$ is at most k. For the functional model, the maximum likelihood estimators of the parameters were derived by Anderson (1951a), and alternative derivations have been given by Healy (1980) and Villegas (1982).

The maximum likelihood estimators of β_0 and β derived in Theorem 1 for the structural model are the same as the maximum likelihood estimators for the functional model. On the other hand, the maximum likelihood estimator of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$ for the structural model differs from that for the functional model.

Under the assumption of Theorem 1, we can perform a goodness of fit test for the model. An alternative model is the unrestricted model where $\mathbf{Z}_t' \sim NI(\boldsymbol{\mu}_Z', \mathbf{\Sigma}_{ZZ})$, $\boldsymbol{\mu}_Z$ is an unrestricted p-dimensional vector, $\mathbf{\Sigma}_{ZZ}$ is an unrestricted $p \times p$ positive definite matrix, and $\mathbf{S}_{\varepsilon\varepsilon}$ is a multiple of a Wishart matrix that is independent of $\mathbf{Z}_t$.

COROLLARY 1.1. *Let the null model be defined by* (1.1), (1.2), *and the assumptions of Theorem* 1. *Let the alternative model be the unrestricted model. Then, the likelihood ratio statistic adjusted for degrees of freedom is*

$$\chi^2 = (n - 1 + d) \textstyle\sum_{i=\ell+1}^{p} \log\{(n - 1 + d)^{-1}[(n - 1)\hat{\lambda}_i + d]\} - (n - 1) \sum_{i=\ell+1}^{p} \log \hat{\lambda}_i,$$

where $\ell = \min\{k, q\}$, *and* q *is the number of* $\hat{\lambda}_i$ *which are greater than one. Under the null model, the limiting distribution of the test statistic is that of a chi-squared random variable with* $2^{-1}r(r + 1)$ *degrees of freedom.*

PROOF. Using the notation in the proof of Theorem 1, the maximum likelihood estimators of the transformed parameters for the unrestricted model are

$$\tilde{\boldsymbol{\Sigma}}_{TZZ} = \hat{\boldsymbol{\Lambda}}, \quad \tilde{\boldsymbol{\Sigma}}_{T\varepsilon\varepsilon} = \mathbf{I}.$$

The value of $f(\theta)$ in (2.3) evaluated at the unrestricted maximum likelihood estimators $\tilde{\boldsymbol{\Sigma}}_{TZZ}$ and $\tilde{\boldsymbol{\Sigma}}_{T\varepsilon\varepsilon}$ is $[C_1 + \sum_{i=1}^{p} g(\hat{\lambda}_i)]$, where C_1 and $g(w)$ are given in the proof of Theorem 1. By the argument used in the proof of Theorem 1, the infimum of $f(\theta)$ under the null model is $[C_1 + \sum_{i=1}^{\ell} g(\hat{\lambda}_i)]$. Hence, if we let τ be the likelihood ratio test statistic, then $\chi^2 = -2 \log \tau = \sum_{i=\ell+1}^{p} g(\hat{\lambda}_i)$. For the unrestricted parameter space, $\boldsymbol{\Sigma}_{ZZ}$ and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ are any symmetric positive definite matrices. We transform the parameters defining $\boldsymbol{\Sigma}_{ZZ}$ into β, $\mathbf{H}_{xx}$, and $\mathbf{H}_{yy\cdot x}$ by the one-one transformation, where

$$\boldsymbol{\Sigma}_{ZZ} = \mathbf{H}_{zz} + \boldsymbol{\Sigma}_{\varepsilon\varepsilon}, \quad \mathbf{H}_{zz} = (\beta, \mathbf{I})'\mathbf{H}_{xx}(\beta, \mathbf{I}) + (\mathbf{I}, \mathbf{0})'\mathbf{H}_{yy\cdot x}(\mathbf{I}, \mathbf{0}),$$

β is a $k \times r$ matrix, $\mathbf{H}_{xx}$ is a $k \times k$ symmetric matrix, and $\mathbf{H}_{yy\cdot x}$ is an $r \times r$ symmetric matrix such that $\mathbf{H}_{zz} + \boldsymbol{\Sigma}_{zz}$ is positive definite. Under the unrestricted model $\mathbf{H}_{xx}$ and $\mathbf{H}_{yy\cdot x}$ are not necessarily nonnegative definite matrices. Under the null model, $\mathbf{H}_{xx} = \boldsymbol{\Sigma}_{xx}$ is positive definite and $\mathbf{H}_{yy\cdot x} = \mathbf{0}$. Since there are $2^{-1}r(r + 1)$ distinct elements in $\mathbf{H}_{yy\cdot x}$, the result follows from the standard likelihood theory. □

The test in Corollary 1.1 provides a test of the model specification (1.1) and (1.2) against the unrestricted alternative. The test can be used to check the goodness-of-fit of the model (1.1) and (1.2). Anderson (1946, 1984) discusses the likelihood ratio test of the null hypothesis that the rank of $\boldsymbol{\Sigma}_{zz}$ is less than or equal to k against the alternative hypothesis that the rank of $\boldsymbol{\Sigma}_{zz}$ is greater than k. Anderson's test statistic differs from that of Corollary 1.1 and does not have an asymptotic chi-squared distribution.

3. Strong consistency. In this section, we discuss the strong consistency of the estimators $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}$, and $\hat{\boldsymbol{\Sigma}}_{xx}$ defined in Theorem 1. In the derivations we use weaker assumptions than those of Theorem 1. We assume that $\mathbf{S}_{\varepsilon\varepsilon}$ is based on d degrees of freedom and that $\mathbf{S}_{\varepsilon\varepsilon} \to \boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, a.s., as $d \to \infty$. Let

$$c = \lim_{n\to\infty} d^{-1}n. \tag{3.1}$$

For the functional model with $0 \le c < \infty$, Healy (1980) showed that $\hat{\beta}_0$ and $\hat{\beta}$ are consistent and that the maximum likelihood estimator of $\Sigma_{\varepsilon\varepsilon}$ constructed under the functional model is not consistent. The following theorem shows that $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\Sigma}_{\varepsilon\varepsilon}$, and $\hat{\Sigma}_{xx}$ in Theorem 1 are strongly consistent under both the functional and structural models if both n and d tend to infinity.

THEOREM 2. *Let the model* (1.1) *and* (1.2) *hold, and let* $\mathbf{S}_{\varepsilon\varepsilon}$ *be an estimator of* $\Sigma_{\varepsilon\varepsilon}$. *Assume that the* ε_t *are independently and identically distributed with mean zero and positive definite covariance matrix* $\Sigma_{\varepsilon\varepsilon}$. *Assume that* $\mathbf{S}_{\varepsilon\varepsilon} \to \Sigma_{\varepsilon\varepsilon}$, a.s., *as* $d \to \infty$, *and that* $\lim_{n\to\infty} d = \infty$. *Also, assume either* (a) $\mathbf{x}_t$ *are independently and identically distributed with mean* μ_x, *positive definite covariance matrix* Σ_{xx}, *and* $\mathbf{x}_t$ *and* ε_j *are independent for all* t *and* j, *or* (b) $\mathbf{x}_t$ *are fixed and satisfy*

$$\lim_{n\to\infty} \bar{\mathbf{x}} = \mu_x, \quad \lim_{n\to\infty} \mathbf{m}_{xx} = \Sigma_{xx},$$

where Σ_{xx} *is a positive definite matrix. Then, as* $n \to \infty$, *with probability one,* $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\Sigma}_{\varepsilon\varepsilon}$, *and* $\hat{\Sigma}_{xx}$ *converge to* β_0, β, $\Sigma_{\varepsilon\varepsilon}$, *and* Σ_{xx}, *respectively.*

PROOF. By the strong law of large numbers for case (a) and by Lemma 3.1 of Gleser (1981) for case (b), $\mathbf{m}_{ZZ} \to \Sigma_{ZZ}$, a.s., and

$$\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2} \to \Sigma_{\varepsilon\varepsilon}^{-1/2}\Sigma_{ZZ}\Sigma_{\varepsilon\varepsilon}^{-1/2}, \quad \text{a.s.,} \quad \text{as} \quad n \to \infty.$$

Let

$$\Sigma_{\rho\rho} = [(\beta, \mathbf{I})\Sigma_{\varepsilon\varepsilon}^{-1}(\beta, \mathbf{I})']^{-1} = \Sigma_{uu} - \Sigma_{uv}\Sigma_{vv}^{-1}\Sigma_{vu},$$

where

$$\Sigma_{vv} = (\mathbf{I}, -\beta')\Sigma_{\varepsilon\varepsilon}(\mathbf{I}, -\beta')', \quad \Sigma_{uv} = \Sigma_{ue} - \Sigma_{uu}\beta.$$

Also, let $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_k > 0$ be the eigenvalues of $\Sigma_{\rho\rho}^{-1/2}\Sigma_{xx}\Sigma_{\rho\rho}^{-1/2}$, and let $\mathbf{R}$ be the matrix of the corresponding orthonormal eigenvectors. Then, the eigenvalues of $\Sigma_{\varepsilon\varepsilon}^{-1/2}\Sigma_{ZZ}\Sigma_{\varepsilon\varepsilon}^{-1/2}$ are $\lambda_i = 1 + \nu_i$ for $i = 1, 2, \cdots, k$, and $\lambda_i = 1$ for $i = k + 1, k + 2, \cdots, p$. The matrix of the corresponding orthonormal eigenvectors has the form

$$(3.2) \qquad (\mathbf{Q}_1^0\mathbf{G}_1, \mathbf{Q}_1^0\mathbf{G}_2) = (\mathbf{Q}_1^0 \text{ block diag}\{\mathbf{G}_{11}, \mathbf{G}_{12}, \cdots, \mathbf{G}_{1s}\}, \mathbf{Q}_2^0\mathbf{G}_2),$$

where

$$\mathbf{Q}_1^0 = \Sigma_{\varepsilon\varepsilon}^{-1/2}(\beta, \mathbf{I})'\Sigma_{\rho\rho}^{1/2}\mathbf{R}, \quad \mathbf{Q}_2^0 = \Sigma_{\varepsilon\varepsilon}^{1/2}(\mathbf{I}, -\beta')'\Sigma_{vv}^{-1/2},$$

$\mathbf{G}_{11}, \mathbf{G}_{12}, \cdots, \mathbf{G}_{1s}$ and $\mathbf{G}_2$ are orthogonal matrices, and s is the number of distinct roots ν_i. The eigenvalues $\hat{\lambda}_i$ are locally continuous functions of the elements in $\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}$. (See Franklin, 1968, page 191.) Thus, $\hat{\lambda}_i \to \lambda_i$, a.s., for $i = 1, 2, \cdots, p$. Let ω be a point in the probability space of all sequences of observations. The following arguments are similar to those used in the proof of Lemma 3.3 by Gleser (1981). Fix ω such that $\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2} \to \Sigma_{\varepsilon\varepsilon}^{-1/2}\Sigma_{ZZ}\Sigma_{\varepsilon\varepsilon}^{-1/2}$, and $\hat{\lambda}_i \to \lambda_i$, $i = 1, 2, \cdots, p$. The set of such ω has probability one. Since $\mathbf{Q}(\omega)$ is orthogonal for all n, each element of $\mathbf{Q}(\omega)$ is bounded. Thus, for every subsequence of $\{\mathbf{Q}(\omega)\}$, there exists a convergent subsubsequence. The limit of such a convergent subsubsequence has the form (3.2) for some orthogonal matrices $\mathbf{G}_{11}, \cdots, \mathbf{G}_{1s}$, and $\mathbf{G}_2$. Hence, the limit of $\mathbf{Q}_1(\omega)\hat{\mathbf{\Lambda}}_1(\omega)\mathbf{Q}_1'(\omega)$ over such a subsubsequence is

$$\mathbf{Q}_1^0\mathbf{G}_1 \operatorname{diag}\{\lambda_1, \cdots, \lambda_k\}\mathbf{G}_1'\mathbf{Q}_1^0 = \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}[(\boldsymbol{\beta}, \mathbf{I})'\mathbf{\Sigma}_{\rho\rho}(\boldsymbol{\beta}, \mathbf{I}) + \mathbf{\Sigma}_{zz}]\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2} = \mathbf{\Sigma}_{11}.$$

Since the limit $\mathbf{\Sigma}_{11}$ does not depend on the subsubsequence,

$$\mathbf{Q}_1(\omega)\hat{\mathbf{\Lambda}}_1(\omega)\mathbf{Q}_1'(\omega) \to \mathbf{\Sigma}_{11}.$$

Therefore,

$$\mathbf{Q}_1\hat{\mathbf{\Lambda}}_1\mathbf{Q}_1' \to \mathbf{\Sigma}_{11}, \quad \text{a.s.} \tag{3.3}$$

Similarly,

$$\mathbf{Q}_1\mathbf{Q}_1' \to \mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}(\boldsymbol{\beta}, \mathbf{I})'\mathbf{\Sigma}_{\rho\rho}(\boldsymbol{\beta}, \mathbf{I})\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1/2}, \quad \text{a.s.}$$

Thus,

$$(\hat{\boldsymbol{\beta}}, \mathbf{I})'\hat{\mathbf{\Sigma}}_{xx}(\hat{\boldsymbol{\beta}}, \mathbf{I}) = \mathbf{S}_{\varepsilon\varepsilon}^{1/2}\mathbf{Q}_1(\hat{\mathbf{\Lambda}}_1 - \mathbf{I})\mathbf{Q}_1'\mathbf{S}_{\varepsilon\varepsilon}^{1/2}$$

$$\to \mathbf{\Sigma}_{zz} = (\boldsymbol{\beta}, \mathbf{I})'\mathbf{\Sigma}_{xx}(\boldsymbol{\beta}, \mathbf{I}), \quad \text{a.s.},$$

and the strong consistency of $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{\Sigma}}_{\varepsilon\varepsilon}$, and $\hat{\mathbf{\Sigma}}_{xx}$ follows. For either case (a) or (b), $\bar{\mathbf{X}} \to \boldsymbol{\mu}_x$, a.s., $\bar{\mathbf{Y}} \to \boldsymbol{\beta}_0 + \boldsymbol{\mu}_x\boldsymbol{\beta}$, a.s., and $\hat{\boldsymbol{\beta}}_0 \to \boldsymbol{\beta}_0$, a.s. □

4. Limiting distribution. In this section, the limiting distribution of the estimators $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{\Sigma}}_{\varepsilon\varepsilon}$, and $\hat{\mathbf{\Sigma}}_{xx}$ is derived. Under the assumption that the degrees of freedom for $\mathbf{S}_{\varepsilon\varepsilon}$ increases at the same rate as the number of observations n, the limiting covariance matrix contains a contribution from $\mathbf{S}_{\varepsilon\varepsilon}$ as an estimator of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$. We show that the limiting distribution of $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}$, and $\hat{\mathbf{\Sigma}}_{\varepsilon\varepsilon}$ can be derived under relatively weak assumptions on $\mathbf{x}_t$, and that the limiting covariance matrix has a common form for a wide class of $\mathbf{x}_t$. A stronger assumption on $\mathbf{x}_t$ is necessary to obtain the limiting distribution of $\hat{\mathbf{\Sigma}}_{xx}$ for the structural case. For the functional model, Anderson (1951b) discussed the limiting distribution of the maximum likelihood estimators under the assumptions that the $\mathbf{x}_t$ are fixed, that n is fixed and that the error variances tend to zero. For the univariate case, Fuller (1980) gave the limiting distribution for fixed error variances and identified the contribution to the limiting covariance matrix of $\hat{\boldsymbol{\beta}}$ associated with the estimation of $\mathbf{\Sigma}_{\varepsilon\varepsilon}$.

For an $m \times n$ matrix $\mathbf{A}$, let vec $\mathbf{A}$ be the $mn \times 1$ vector obtained by listing the columns one beneath the other beginning with the first column. For an $m \times m$ symmetric matrix $\mathbf{B}$, let vech $\mathbf{B}$ be the $2^{-1}m(m+1) \times 1$ vector obtained by listing the elements that are on or below the diagonal beginning with the first column. Let $\mathbf{\Phi}_m$ be the $m^2 \times 2^{-1}m(m+1)$ matrix and $\boldsymbol{\psi}_m$ the $2^{-1}m(m+1) \times m^2$ matrix such that

$$\operatorname{vec} \mathbf{B} = \mathbf{\Phi}_m \operatorname{vech} \mathbf{B}, \quad \boldsymbol{\psi}_m = (\mathbf{\Phi}_m'\mathbf{\Phi}_m)^{-1}\mathbf{\Phi}_m'.$$

See Henderson and Searle (1979) for a discussion of the vec and vech operators.

The following lemma will be used in the derivation of the limiting distribution of the estimators. A proof is given in Amemiya (1982).

LEMMA 1. *Assume that* $\varepsilon_t' \sim NI(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$. *Assume either* (a′) $\mathbf{x}_t$ *and* ε_j *are independent for all t and j, and* $\mathbf{x}_t$ *are independently and identically distributed with covariance matrix* $\boldsymbol{\Sigma}_{xx}$, *or* (b′) $\mathbf{x}_t$ *are fixed and satisfy* $\lim_{n\to\infty}\mathbf{m}_{xx} = \boldsymbol{\Sigma}_{xx}$. *Then, as* $n \to \infty$,

$$n^{1/2}\begin{pmatrix}\text{vec } \mathbf{m}_{x\varepsilon} \\ \text{vech}(\mathbf{m}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\varepsilon})\end{pmatrix} \to_L N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \otimes \boldsymbol{\Sigma}_{xx} & \mathbf{0} \\ \mathbf{0} & 2\,\psi_p(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon})\psi_p'\end{pmatrix}\right).$$

The following theorem presents our principal results on the limiting distribution of the estimators $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon}$, and $\hat{\boldsymbol{\Sigma}}_{xx}$.

THEOREM 3. *Let the model* (1.1) *and* (1.2) *hold, and let* $\mathbf{S}_{\varepsilon\varepsilon}$ *be an unbiased estimator of* $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$. *Assume that* ε_t *are independently distributed* $N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$ *random variables with positive definite* $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$. *Assume that* $\mathbf{S}_{\varepsilon\varepsilon}$ *is independent of* $\mathbf{Z}_t$ *for all t, and that d* $\mathbf{S}_{\varepsilon\varepsilon}$ *is distributed as Wishart with parameter* $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ *and degrees of freedom d. Also, assume that* $0 < c < \infty$, *where c is defined in* (3.1).

(i) *Assume that the* $\mathbf{x}_t$ *satisfy either assumption* (a) *or assumption* (b) *of Theorem* 2. *Then,*

$$n^{1/2}\begin{pmatrix}(\hat{\beta}_0 - \beta_0)' \\ \text{vec}(\hat{\beta} - \beta) \\ \text{vech}(\hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\varepsilon})\end{pmatrix} \to_L N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{V}_{00} & \mathbf{V}_{0\beta} & \mathbf{V}_{0\varepsilon} \\ \mathbf{V}_{0\beta}' & \mathbf{V}_{\beta\beta} & \mathbf{V}_{\beta\varepsilon} \\ \mathbf{V}_{0\varepsilon}' & \mathbf{V}_{\beta\varepsilon}' & \mathbf{V}_{\varepsilon\varepsilon}\end{pmatrix}\right),$$

where

$$\mathbf{V}_{00} = \boldsymbol{\Sigma}_{vv} + (\mathbf{I}_{rxr} \otimes \boldsymbol{\mu}_x)\mathbf{V}_{\beta\beta}(\mathbf{I}_{rxr} \otimes \boldsymbol{\mu}_x'), \quad \mathbf{V}_{0\beta} = -(\mathbf{I}_{rxr} \otimes \boldsymbol{\mu}_x)\mathbf{V}_{\beta\beta},$$

$$\mathbf{V}_{0\varepsilon} = -(\mathbf{I}_{rxr} \otimes \boldsymbol{\mu}_x)\mathbf{V}_{\beta\varepsilon}, \quad \mathbf{V}_{\beta\beta} = \boldsymbol{\Sigma}_{vv} \otimes \{\boldsymbol{\Sigma}_{xx}^{-1}[\boldsymbol{\Sigma}_{xx} + (1 + c)\boldsymbol{\Sigma}_{\rho\rho}]\boldsymbol{\Sigma}_{xx}^{-1}\},$$

$$\mathbf{V}_{\beta\varepsilon} = -2c\{\boldsymbol{\Sigma}_{v\varepsilon} \otimes [\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{\rho\rho}(\beta, \mathbf{I})]\}\psi_p',$$

$$\mathbf{V}_{\varepsilon\varepsilon} = 2c\psi_p[(\boldsymbol{\Sigma}_{\varepsilon\varepsilon} \otimes \boldsymbol{\Sigma}_{\varepsilon\varepsilon}) - (1 + c^{-1})^{-1}(\boldsymbol{\Sigma}_{\varepsilon v}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{v\varepsilon}) \otimes (\boldsymbol{\Sigma}_{\varepsilon v}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{v\varepsilon})]\psi_p',$$

$$\boldsymbol{\Sigma}_{v\varepsilon} = \boldsymbol{\Sigma}_{v\varepsilon}' = (\mathbf{I}, -\beta')\boldsymbol{\Sigma}_{\varepsilon\varepsilon}.$$

(ii) *Assume that* $\mathbf{x}_t$ *are independently distributed* $N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ *random variables and that* $\mathbf{x}_t$ *and* ε_j *are independent for all t and j. Then,*

$$n^{1/2}\begin{bmatrix}(\hat{\beta}_0 - \beta_0)' \\ \text{vec}(\hat{\beta} - \beta) \\ \text{vech}(\hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\varepsilon}) \\ \text{vech}(\hat{\boldsymbol{\Sigma}}_{xx} - \boldsymbol{\Sigma}_{xx})\end{bmatrix} \to_L N\left[\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\mathbf{V}_{00} & \mathbf{V}_{0\beta} & \mathbf{V}_{0\varepsilon} & \mathbf{V}_{0x} \\ \mathbf{V}_{0\beta}' & \mathbf{V}_{\beta\beta} & \mathbf{V}_{\beta\varepsilon} & \mathbf{V}_{\beta x} \\ \mathbf{V}_{0\varepsilon}' & \mathbf{V}_{\beta\varepsilon}' & \mathbf{V}_{\varepsilon\varepsilon} & \mathbf{V}_{\varepsilon x} \\ \mathbf{V}_{0x}' & \mathbf{V}_{\beta x}' & \mathbf{V}_{\varepsilon x}' & \mathbf{V}_{xx}\end{bmatrix}\right],$$

where

$$\mathbf{V}_{0x} = -(\mathbf{I}_{rxr} \otimes \boldsymbol{\mu}_x)\mathbf{V}_{\beta x}, \quad \mathbf{V}_{\beta x} = 2\{\boldsymbol{\Sigma}_{vu} \otimes \boldsymbol{\Sigma}_{xx}^{-1}[\boldsymbol{\Sigma}_{xx} + (1 + c)\boldsymbol{\Sigma}_{\rho\rho}]\}\psi_k',$$

$$\mathbf{V}_{\varepsilon x} = 2c\,\psi_p[(\boldsymbol{\Sigma}_{\varepsilon v}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu} \otimes \boldsymbol{\Sigma}_{\varepsilon v}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu}) - (\boldsymbol{\Sigma}_{\varepsilon u} \otimes \boldsymbol{\Sigma}_{\varepsilon u})]\psi_k'$$

$$\mathbf{V}_{xx} = 2\,\psi_k[(\boldsymbol{\Sigma}_{XX} \otimes \boldsymbol{\Sigma}_{XX}) + c(\boldsymbol{\Sigma}_{uu} \otimes \boldsymbol{\Sigma}_{uu}) - (1 + c)(\boldsymbol{\Sigma}_{uv}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu} \otimes \boldsymbol{\Sigma}_{uv}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu})]\psi_k',$$

$$\boldsymbol{\Sigma}_{XX} = \boldsymbol{\Sigma}_{xx} + \boldsymbol{\Sigma}_{uu}.$$

PROOF OF (i). Observe that

$$
\begin{aligned}
&\mathrm{vec}[(\beta, \mathbf{I})\mathbf{S}_{\varepsilon\varepsilon}^{-1}\mathbf{m}_{ZZ}(\mathbf{I}, -\beta')'] \\
&\qquad = \mathrm{vec}[(\beta, \mathbf{I})\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}(\mathbf{Q}_1\hat{\mathbf{\Lambda}}_1\mathbf{Q}_1' + \mathbf{Q}_2\hat{\mathbf{\Lambda}}_2\mathbf{Q}_2')\mathbf{S}_{\varepsilon\varepsilon}^{1/2}(\mathbf{I}, -\beta')'] \\
&\qquad = \mathrm{vec}[\mathbf{F}_{kk}(\hat{\beta} - \beta) + (\hat{\beta} - \beta)\mathbf{F}_{rr}] \\
&\qquad = (\mathbf{I}_{r\times r} \otimes \mathbf{F}_{kk} + \mathbf{F}_{rr}' \otimes \mathbf{I}_{k\times k})\mathrm{vec}(\hat{\beta} - \beta),
\end{aligned} \tag{4.1}
$$

where $\mathbf{I}_{r\times r}$ and $\mathbf{I}_{k\times k}$ are identity matrices of dimensions $r \times r$ and $k \times k$, respectively, and

$$\mathbf{F}_{kk} = (\beta, \mathbf{I})\mathbf{S}_{\varepsilon\varepsilon}^{-1}(\hat{\beta}, \mathbf{I})\mathbf{P}_{kk}\hat{\mathbf{\Lambda}}_1\mathbf{P}_{kk}', \quad \mathbf{F}_{rr} = -\mathbf{T}_{rr}\hat{\mathbf{\Lambda}}_2\mathbf{T}_{rr}'(\mathbf{I}, -\hat{\beta}')\mathbf{S}_{\varepsilon\varepsilon}(\mathbf{I}, -\beta')'.$$

The argument used to obtain (3.3) can be used to show that, as $n \to \infty$,

$$
\begin{aligned}
&\mathbf{P}_{kk}\hat{\mathbf{\Lambda}}_1\mathbf{P}_{kk}' \to \mathbf{\Sigma}_{xx} + \mathbf{\Sigma}_{\rho\rho}, \quad \text{a.s.,} \quad \mathbf{T}_{rr}\hat{\mathbf{\Lambda}}_2\mathbf{T}_{rr}' \to \mathbf{\Sigma}_{vv}^{-1}, \quad \text{a.s.,} \\
&\mathbf{T}_{rr}\mathbf{T}_{rr}' \to \mathbf{\Sigma}_{vv}^{-1}, \quad \text{a.s.}
\end{aligned} \tag{4.2}
$$

Thus,

$$\mathbf{I}_{r\times r} \otimes \mathbf{F}_{kk} + \mathbf{F}_{rr}' \otimes \mathbf{I}_{k\times k} \to \mathbf{I}_{k\times k} \otimes (\mathbf{\Sigma}_{\rho\rho}^{-1}\mathbf{\Sigma}_{xx}), \quad \text{a.s.,} \tag{4.3}$$

and the limit is nonsingular. Also,

$$
\begin{aligned}
&(\beta, \mathbf{I})\mathbf{S}_{\varepsilon\varepsilon}^{-1}\mathbf{m}_{ZZ}(\mathbf{I}, -\beta')' \\
&\quad = (\beta, \mathbf{I})\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}[(\beta, \mathbf{I})'\mathbf{m}_{x\varepsilon} + \mathbf{m}_{\varepsilon\varepsilon} - \mathbf{\Sigma}_{\varepsilon\varepsilon}](\mathbf{I}, -\beta')' \\
&\quad\quad - (\beta, \mathbf{I})\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}(\mathbf{S}_{\varepsilon\varepsilon} - \mathbf{\Sigma}_{\varepsilon\varepsilon})(\mathbf{I}, -\beta')' + O_p(n^{-1}) \\
&\quad = \mathbf{\Sigma}_{\rho\rho}^{-1}(\mathbf{m}_{xv} + \mathbf{m}_{\xi v} - \mathbf{S}_{\xi v}) + O_p(n^{-1}) = O_p(n^{-1/2}),
\end{aligned} \tag{4.4}
$$

where

$$
\begin{aligned}
&\mathbf{m}_{xv} = \mathbf{m}_{x\varepsilon}(\mathbf{I}, -\beta')', \quad \mathbf{m}_{\xi v} = \mathbf{\Sigma}_{\rho\rho}(\beta, \mathbf{I})\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{m}_{\varepsilon\varepsilon}(\mathbf{I}, -\beta')', \\
&\mathbf{S}_{\xi v} = \mathbf{\Sigma}_{\rho\rho}(\beta, \mathbf{I})\mathbf{\Sigma}_{\varepsilon\varepsilon}^{-1}\mathbf{S}_{\varepsilon\varepsilon}(\mathbf{I}, -\beta')'.
\end{aligned}
$$

By (4.1), (4.3), and (4.4),

$$\mathrm{vec}(\hat{\beta} - \beta) = (\mathbf{I}_{r\times r} \otimes \mathbf{\Sigma}_{xx}^{-1})\mathrm{vec}(\mathbf{m}_{xv} + \mathbf{m}_{\xi v} - \mathbf{S}_{\xi v}) + o_p(n^{-1/2}). \tag{4.5}$$

Also,

$$\hat{\beta}_0 - \beta_0 = \bar{\mathbf{v}} - [\mathrm{vec}(\hat{\beta} - \beta)]'(\mathbf{I}_{r\times r} \otimes \mu_x') + o_p(n^{-1/2}). \tag{4.6}$$

Observe that

$$
\begin{aligned}
(\mathbf{I}, -\beta')\mathbf{m}_{ZZ}(\mathbf{I}, -\hat{\beta}')'\mathbf{T}_{rr} &= (\mathbf{I}, -\beta')\mathbf{m}_{ZZ}\mathbf{S}_{\varepsilon\varepsilon}^{-1/2}\mathbf{Q}_2 \\
&= (\mathbf{I}, -\beta')\mathbf{S}_{\varepsilon\varepsilon}(\mathbf{I}, -\hat{\beta}')'\mathbf{T}_{rr}\hat{\mathbf{\Lambda}}_2.
\end{aligned}
$$

Hence, by (4.2) and (4.5),

$$\mathbf{T}_{rr}(\hat{\mathbf{\Lambda}}_2 - \mathbf{I})\mathbf{T}_{rr}' = \mathbf{\Sigma}_{vv}^{-1}(\mathbf{m}_{vv} - \mathbf{S}_{vv})\mathbf{\Sigma}_{vv}^{-1} + o_p(n^{-1/2}) = O_p(n^{-1/2}). \tag{4.7}$$

We also have

$$\text{(4.8)} \quad \mathbf{m}_{ZZ} = (\hat{\beta}, \mathbf{I})'\hat{\boldsymbol{\Sigma}}_{xx}(\hat{\beta}, \mathbf{I}) + \mathbf{S}_{\varepsilon\varepsilon} + \mathbf{S}_{\varepsilon\varepsilon}(\mathbf{I}, -\hat{\beta}')'\mathbf{T}_{rr}(\hat{\boldsymbol{\Lambda}}_2 - \mathbf{I})\mathbf{T}_{rr}'(\mathbf{I}, -\hat{\beta}')\mathbf{S}_{\varepsilon\varepsilon}.$$

By (4.5), (4.7), and (4.8),

$$\text{(4.9)} \quad \hat{\boldsymbol{\Sigma}}_{\varepsilon\varepsilon} = \mathbf{S}_{\varepsilon\varepsilon} + (n-1+d)^{-1}(n-1)\boldsymbol{\Sigma}_{\varepsilon v}\boldsymbol{\Sigma}_{vv}^{-1}(\mathbf{m}_{vv} - \mathbf{S}_{vv})\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{v\varepsilon} + o_p(n^{-1/2}).$$

By the Wishart assumption on $\mathbf{S}_{\varepsilon\varepsilon}$, the joint distribution of $n^{1/2}\text{vech}(\mathbf{S}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$, $n^{1/2}\text{vec}\, \mathbf{S}_{\xi v}$, and $n^{1/2}\text{vech}(\mathbf{S}_{vv} - \boldsymbol{\Sigma}_{vv})$ tends to a normal distribution. By Lemma 1, the joint distribution of $n^{1/2}\bar{\mathbf{v}}'$, $n^{1/2}\text{vec}\, \mathbf{m}_{xv}$, $n^{1/2}\text{vec}\, \mathbf{m}_{\xi v}$, and $n^{1/2}\text{vech}(\mathbf{m}_{vv} - \boldsymbol{\Sigma}_{vv})$ tends to a normal distribution. Using Lemma 1, (4.5), (4.6), and (4.9), we obtain the result by evaluating the limiting covariance matrix.

PROOF OF (ii). Observe that

$$\text{(4.10)} \quad \mathbf{S}_{\varepsilon\varepsilon}(\mathbf{I} - \hat{\beta}')' = \boldsymbol{\Sigma}_{\varepsilon v} + O_p(n^{-1/2}).$$

Using (4.7) and (4.10), we write the lower right $k \times k$ corner of (4.8) as

$$\text{(4.11)} \quad \hat{\boldsymbol{\Sigma}}_{xx} = \mathbf{m}_{XX} - \mathbf{S}_{uu} - \boldsymbol{\Sigma}_{uv}\boldsymbol{\Sigma}_{vv}^{-1}(\mathbf{m}_{vv} - \mathbf{S}_{vv})\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{vu} + o_p(n^{-1/2}).$$

Hence, the result follows from Lemma 1, (4.5), (4.6), (4.9), (4.11), and the argument used in the proof of part (i). □

In Theorem 3, if the normality of ε_t in part (i) and part (ii) and the normality of $\mathbf{x}_t$ in part (ii) are replaced by the existence of fourth moments, and if the Wishart distribution of $d\, \mathbf{S}_{\varepsilon\varepsilon}$ is replaced by the condition that the distribution of $d^{1/2}(\mathbf{S}_{\varepsilon\varepsilon} - \boldsymbol{\Sigma}_{\varepsilon\varepsilon})$ tends to normal as $d \to \infty$, then the estimators have a limiting normal distribution, but with different parameters than those given in Theorem 3. It can be shown that the results of Theorem 3 hold when $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ is singular, provided $\boldsymbol{\Sigma}_{vv}$ is nonsingular.

In the literature on the structural equation model, the limiting distribution of the estimators is often derived by the standard normal likelihood theory. Such an argument is not directly applicable under the assumptions of part (i) of Theorem 3.

Under the assumptions for the functional model stated in part (i) of Theorem 3, $n^{1/2}\text{vech}(\hat{\boldsymbol{\Sigma}}_{xx} - \mathbf{m}_{xx})$ has a limiting normal distribution. The limiting covariance matrix takes the form of that given in part (ii) with $(\boldsymbol{\Sigma}_{XX} \otimes \boldsymbol{\Sigma}_{XX})$ in $\mathbf{V}_{xx}$ replaced by $(2\, \boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{\Sigma}_{uu} + \boldsymbol{\Sigma}_{uu} \otimes \boldsymbol{\Sigma}_{uu})$.

It can be shown that the results for $\hat{\beta}_0$ and $\hat{\beta}$ in Theorem 3 with $c = 0$ provide the limiting distribution of the maximum likelihood estimators $\tilde{\beta}_0$ and $\tilde{\beta}$ for the model (1.1) and (1.2) with known $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$, where $\tilde{\beta}_0$ and $\tilde{\beta}$ are obtained from $\hat{\beta}_0$ and $\hat{\beta}$ by replacing $\mathbf{S}_{\varepsilon\varepsilon}$ by the known $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$. Gleser (1981) obtained the limiting distribution of $\tilde{\beta}_0$ and $\tilde{\beta}$ for the functional model with $\boldsymbol{\Sigma}_{\varepsilon\varepsilon} = \sigma^2\mathbf{I}$, where σ^2 is unknown. If $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{\varepsilon\varepsilon}$ are replaced by $\lim_{n\to\infty} n^{-1}\Sigma_{t=1}^{n}\mathbf{x}_t'\mathbf{x}_t$ and $\sigma^2\mathbf{I}$, respectively, the limiting covariance matrix for $\tilde{\beta}$ is the same as that obtained by Gleser for the no-intercept model. Also, the limiting covariance matrix for $\hat{\beta}$ reduces to Fuller's (1980) result for the univariate model.

The limiting covariance matrices in Theorem 3 are simple functions of the unknown parameters β, $\Sigma_{\varepsilon\varepsilon}$, and Σ_{xx}. Thus, consistent estimators of the limiting covariance matrices can be obtained by replacing β, $\Sigma_{\varepsilon\varepsilon}$ and Σ_{xx} with $\hat{\beta}$, $\hat{\Sigma}_{\varepsilon\varepsilon}$, and $\hat{\Sigma}_{xx}$. The expression for the limiting covariance matrix in Theorem 3 is of practical importance because the only matrices to be inverted in the evaluation of the covariance matrix in Theorem 3 are $\hat{\Sigma}_{xx}$ and $\hat{\Sigma}_{vv}$. These matrices are of dimension $k \times k$ and $r \times r$, respectively, while the total number of parameters in the model is $r + kr + \frac{1}{2} k(k + 1) + \frac{1}{2} (k + r)(k + r + 1)$.

Acknowledgement. We thank the referee for thoughtful comments and T. W. Anderson, Leon Gleser and C. M. Theobald for useful discussions.

REFERENCES

AMEMIYA, Y. (1982). Estimators for the errors-in-variables model. Ph.D. Thesis. Iowa State University, Ames, Iowa.

ANDERSON, T. W. (1946). Analysis of multivariate variance. Unpublished.

ANDERSON, T. W. (1951a). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22** 327–351.

ANDERSON, T. W. (1951b). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Statist. Probab.* Univ. of Calif. Press, Berkeley.

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Statist. Probab.* Univ. of Calif. Press, Berkeley.

BROWNE, M. W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South African Statist. J.* **8** 1–24.

CHAN, N. N. (1980). Estimating linear functional relationships. *Recent Developments in Statistical Inference and Data Analysis.* K. Matusita, ed. North Holland, Amsterdam.

CHAN, N. N. and MAK, T. K. (1983). Estimation of multivariate linear functional relationships. *Biometrika* **70** 263–267.

DAHM, P. F. and FULLER, W. A. (1981). Generalized least squares estimation of the functional multivariate linear errors in variables model. Report to the U.S. Census Bureau on J.S.A. 81-14.

FRANKLIN, J. N. (1968). *Matrix Theory.* Prentice Hall, New Jersey.

FULLER, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8** 407–422.

GEARY, R. C. (1948). Studies in relations between economic time series. *J. Roy. Statist. Soc. B* **10** 140–158.

GLESER, L. J. (1981). Estimation in a multivariate "errors-in-variables" regression model: large sample results. *Ann. Statist.* **9** 24–44.

GLESER, L. J. and WATSON, G. S. (1973). Estimation of a linear transformation. *Biometrika* **60** 525–534.

HEALY, J. D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. *J. Multivariate Anal.* **10** 243–251.

HENDERSON, H. V. and SEARLE, S. R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canadian J. Statist.* **7** 65–81.

HÖSCHEL, H. P. (1978). Generalized least squares estimators of linear functional relations with known error covariance. *Math. Operationsforsch. Statist. Ser. Statist.* **9** 9–26.

JENNRICH, R. I. and THAYER, D. T. (1973). A note on Lawley's formulas for standard errors in maximum likelihood factor analysis. *Psychometrika* **38** 571–580.

JÖRESKOG, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32** 443–482.

JÖRESKOG, K. G. (1970). A general method for analysis of covariance structures. *Biometrika* **57** 239–251.

JÖRESKOG, K. G. and GOLDBERGER, A. S. (1972). Factor analysis by generalized least squares. *Psychometrika* **37** 243–260.

LAWLEY, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinburgh A* **60** 64–82.

LAWLEY, D. N. (1941). Further investigations in factor estimation. *Proc. Roy. Soc. Edinburgh A* **61** 176–185.

LAWLEY, D. N. (1943). The application of the maximum likelihood method to factor analysis. *Brit. J. Psychol.* **33** 172–175.

LAWLEY, D. N. (1967). Some new results in maximum likelihood factor analysis. *Proc. Roy. Soc. Edinburgh A* **67** 256–264.

LAWLEY, D. N. (1976). The inversion of an augmented information matrix occurring in factor analysis. *Proc. Roy. Soc. Edinburgh A* **75** 171–178.

NUSSBAUM, M. (1979). Asymptotic efficiency of estimators of a multivariate linear functional relation. *Math. Operationsforsch. Statist. Ser. Statist.* **10** 505–527.

THEOBALD, C. M. (1975). An inequality with application to multivariate analysis. *Biometrika* **62** 461–466.

TINTNER, G. (1945). A note on rank, multicollinearlity, and multiple regression. *Ann. Math. Statist.* **16** 304–308.

VILLEGAS, C. (1982). Maximum likelihood and least squares estimation in linear and affine functional models. *Ann. Statist.* **10** 256–265.

WHITTLE, P. (1952). On principal components and least squares methods of factor analysis. *Skand. Aktuaretidskr.* **35** 223–239.

IOWA STATE UNIVERSITY
STATISTICAL LABORATORY AND
DEPARTMENT OF STATISTICS
102 SNEDECOR HALL
AMES, IOWA 50011