1979

Estimation of the parameters of the multivariate linear errors in variables model

Paul Frederick Dahm
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
Dahm, Paul Frederick, "Estimation of the parameters of the multivariate linear errors in variables model" (1979). Retrospective Theses and Dissertations. 7199.
https://lib.dr.iastate.edu/rtd/7199
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction:

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University Microfilms International
300 N. ZEEB ROAD, ANN ARBOR. MI 48106
18 BEDFORD ROW. LONDON WC1R 4EJ, ENGLAND
DAHM, PAUL FREDERICK
ESTIMATION OF THE PARAMETERS OF THE
MULTIVARIATE LINEAR ERRORS IN VARIABLES
MODEL.
IOWA STATE UNIVERSITY, PH.D., 1979
Estimation of the parameters of the multivariate linear errors in variables model

by

Paul Frederick Dahm

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1979
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>I. INTRODUCTION AND LITERATURE REVIEW</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>B. Definition of the Multivariate Linear Errors in Variables Model</td>
<td>2</td>
</tr>
<tr>
<td>C. Review of the Multivariate Linear Errors in Variables Model</td>
<td>6</td>
</tr>
<tr>
<td>1. Structural relationships</td>
<td>6</td>
</tr>
<tr>
<td>2. Functional relationships</td>
<td>13</td>
</tr>
<tr>
<td>3. Factor analysis</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. DEFINITIONS AND PRELIMINARY RESULTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. ESTIMATION OF THE MULTIVARIATE LINEAR STRUCTURAL MODEL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Preliminary Notation</td>
<td>38</td>
</tr>
<tr>
<td>B. Generalized Least Squares Estimation of the Multivariate Linear Structural Model</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IV. ESTIMATION OF THE MULTIVARIATE LINEAR FUNCTIONAL MODEL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Preliminary Notation</td>
<td>62</td>
</tr>
<tr>
<td>B. Generalized Least Squares Estimation of the Multivariate Linear Functional Model</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. ESTIMATION OF NONNORMAL MULTIVARIATE LINEAR ERRORS IN VARIABLES MODELS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Introduction</td>
<td>92</td>
</tr>
<tr>
<td>B. Estimation of the Structural Model when $\varepsilon_t$ and $x_{t'}$, $t = 1,2,\ldots$ are Nonnormal</td>
<td>93</td>
</tr>
<tr>
<td>C. Estimation of the Functional Model when the $x_{t'}$, $t = 1,2,\ldots$ are Nonnormal</td>
<td>105</td>
</tr>
</tbody>
</table>
### VI. EXAMPLES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. An Internal Estimate of Measurement Error in Available Soil Moisture</td>
<td>115</td>
</tr>
<tr>
<td>B. Estimation of Heritability in Beef Cattle</td>
<td>133</td>
</tr>
</tbody>
</table>

### VII. BIBLIOGRAPHY

Page 159

### VIII. ACKNOWLEDGMENTS

Page 165
I. INTRODUCTION AND LITERATURE REVIEW

A. Introduction

The statistical consideration of models containing measurement errors began as early as 1877. Most research of the errors in variables problem has been of univariate models, that is, models with one dependent variable (cf. p. 3). The multivariate errors in variables model generalizes the univariate errors in variables model by allowing more than one dependent variable. This thesis considers the estimation of the multivariate linear errors in variables model.

The presence of measurement error in the independent variables in the univariate model usually necessitates the use of additional information to estimate the parameters of the model. Three general types of additional information have been used.

1. Distributional knowledge
2. Knowledge about the error variances and covariances
3. Instrumental variables

Diverse methods of estimation have evolved depending upon the nature of the additional information available. Reviews of various estimation procedures for linear univariate errors in variables models are given in Madansky (1959), Cochran (1968), Wolter (1974), and Carter (1976). The nonlinear univariate model is reviewed by Kendall and Stuart (1967) and Wolter (1974).
A more limited variety of estimation techniques have been developed for the multivariate linear errors in variables model. This thesis considers a generalized least squares (G.L.S.) approach to the analysis of the covariance structure of the observed variables. Before discussing the G.L.S. approach, we define the multivariate linear errors in variables model and review the literature on the estimation of these models.

B. Definition of the Multivariate Linear Errors in Variables Model

Let \( \{ \mathbf{y}_t \}_{t=1}^{\infty} \) be a sequence of \( p \)-dimensional random row vectors satisfying

\[
\mathbf{y}_t = f(\mathbf{x}_t, \mathbf{B}), \quad t = 1, 2, \ldots, \tag{1.1}
\]

where \( \{ \mathbf{x}_t \}_{t=1}^{\infty} \) is a sequence of \( k \)-dimensional random row vectors, \( \mathbf{B} \in \mathbb{R}^{kp} \) is a \( kp \times 1 \) vector of parameters, and the components of the \( p \)-vector \( f \) are real valued Borel measurable functions mapping \( \mathbb{R}^k \times \emptyset \) into \( \mathbb{R}^1 \). We define the multivariate errors in variables model by

\[
\mathbf{y}_t = \mathbf{y}_t + \mathbf{e}_t, \quad t = 1, 2, \ldots, \tag{1.2}
\]

\[
\mathbf{x}_t = \mathbf{x}_t + \mathbf{u}_t, \quad t = 1, 2, \ldots. \tag{1.3}
\]
where $y_t$ and $x_t$ are observable random row vectors of dimension $p(\geq 2)$ and $k$, respectively, $e_t$ and $u_t$ are unobservable error vectors of dimension $p$ and $k$, and $\xi_t = (e_t, u_t)$ satisfies

$$E(\xi_{ti}) = E(\xi_{tj}x_{ti}) = 0,$$  \hspace{0.5cm} (1.4)

for all $t = 1, 2, \ldots, n$, $t' = 1, 2, \ldots, n$, $i = 1, 2, \ldots, p+k$, and $j = 1, 2, \ldots, k$. If $p=1$, model (1.1-1.4) is called the univariate errors in variables model. Suppose the elements of $B$ are not functionally related and

$$f(x_t, B) = x_t \delta, \hspace{0.5cm} t = 1, 2, \ldots,$$  \hspace{0.5cm} (1.5)

where $\delta$ is a $k \times p$ matrix of parameters formed from the $kp \times 1$ vector $B$. The model (1.1-1.5) defines the multivariate linear errors in variables model.

The distinction between $x_t$ being fixed or random is important in the estimation of the parameters of the multivariate errors in variables model. If the $x_t, t = 1, 2, \ldots$, are constant vectors, the multivariate errors in variables model is termed a functional relationship model, while if the $x_t, t = 1, 2, \ldots$, are nonconstant random vectors the model is called a structural relationship model. This terminology is due to Kendall (1951). Maximum likelihood (M.L.) estimation of the functional multivariate linear errors in variables model would include the $x_t, t = 1, 2, \ldots$, as
unknown parameters to be estimated. Because of the indefinitely increasing number of unknown parameters in this model, the classical asymptotic theory of M.L. estimation is inapplicable. Neyman and Scott (1948) referred to parameters, such as \( x_t, \ t = 1, 2, \ldots \), which enter the distribution of the observed random variables for finitely many \( t \) as incidental parameters and those entering for infinitely many \( t \) as structural parameters. Generally one is interested in estimating only structural parameters.

For the multivariate model the standard assumption appearing in the literature is that the \( \varepsilon_t, \ t = 1, 2, \ldots \), are independent identically distributed \((p+k)\)-variate normal random vectors with mean zero and positive definite covariance matrix which we will denote by \( \Xi \). If the model is a structural multivariate linear model, the usual assumption is that the \( \tilde{x}_t, \ t = 1, 2, \ldots \), are independent identically distributed \( k \)-variate normal random vectors with mean \( \mu_x \) and positive definite covariance matrix \( \Xi_x (\sigma_x^2 \text{ if } k=1) \). Unless otherwise stated, the above distributional assumptions are understood to hold throughout the remainder of this dissertation.

It will be convenient at times to write the model defined by (1.1-1.5) as

\[
\tilde{z}_t = \mu + \varepsilon_t, \ t = 1, 2, \ldots, n,
\]
where

\[ Z_t = (y_t, x_t), \]

\[ \mu_t = (\beta_t, x_t), \]

and, as previously defined,

\[ \xi_t = (e_t, u_t). \] (1.6)

Thus, \( Z_t \) is a \((p+k)\)-dimensional row vector of observable random variables, \( \xi_t \) is a \((p+k)\)-dimensional row vector of unobservable random variables, and \( \mu_t \) is a \((p+k)\)-dimensional row vector either of unknown constants or unobservable random variables. The same definitions and assumptions apply to the model given in the form (1.6) as to the model when written as (1.1-1.5). Hence, we may unambiguously refer to either the structural or functional form of model (1.6). We denote by \( Z_{it} \), \( \mu_{it} \), and \( \xi_{it} \) the \( i \)th elements of \( Z_t \), \( \mu_t \), and \( \xi_t \), respectively.

The concept of identifiability is important in the estimation of models such as (1.6). Generally speaking, we say the parameters of a model are identifiable if they can be determined uniquely from knowledge of the distribution function of the observable random variables. In multivariate errors in variables models it is often the case that parameters can be determined from knowledge of the joint
distribution function of the observable random variables in more than one way. The parameters are then said to be over-identified. We shall say more about identifiability in later sections.

C. Review of the Multivariate Linear Errors in Variables Model

We classify the multivariate errors in variables literature into three categories - structural relationship models, functional relationship models, and factor analysis models. We first review the literature on the multivariate linear structural model.

1. Structural relationships

Grubbs (1948) considered the multivariate linear structural model with \( k = 1, \) \( \xi \in \mathbb{E} = \text{diag}(\sigma^2_1, \ldots, \sigma^2_{p+1}), \) and \( \xi = \mathbf{1}_p, \) where \( \mathbf{1}_p \) is a \( 1 \times p \) vector of ones. This model arose in an experiment to determine the precisions, \( \sigma^2_i, \) \( i = 1, \ldots, p+1, \) of different measuring instruments and product variability, \( \sigma^2_x. \) Using the notation of model (1.6), let \( Z \) denote the \( (p+1) \times (p+1) \) covariance matrix of the observable random variables, \( Z_t, \) \( t = 1, 2, \ldots, n. \) The \((i,j)\)th element of \( Z \) is \( Z_{ij}, \) where
The matrix $V_n = n^{-1} \sum_{t=1}^n (Z_t - \bar{Z})(Z_t - \bar{Z})'$, where

$$\bar{Z} = n^{-1} \sum_{t=1}^n Z_t,$$

is the maximum likelihood estimator of $\bar{Z}$. Let $V_{ij}$ be the $(i,j)$th element of $V_n$. Grubbs obtained consistent estimators of $\sigma_x^2, \sigma_i^2, i = 1, \ldots, p+1$, by solving the equations $\bar{Z}_{ij} = V_{ij}, i = 1, \ldots, p+1, j = i, i+1, \ldots, p+1$. If $p = 1$, Grubbs' estimators are maximum likelihood estimators, since the transformation from $\bar{Z}_{11}, \bar{Z}_{12}, \bar{Z}_{22}$ to $\sigma_x^2, \sigma_1^2, \sigma_2^2$ is one to one. Three objections to Grubbs' estimators are: (1) the estimates of variance can be negative; (2) if $p \geq 2$, the procedure does not yield unique estimates; and (3) requiring $\beta \mathbin{1 \!\! \sim} p$ is overly restrictive in some situations.

Although not explicitly stated by Grubbs, the above procedure assumes that any systematic difference between measuring instruments is constant throughout the range of observations. Smith (1950) considered the special case of Grubbs' model where $p = 1$ and the scales of the two measuring instruments are linearly related by $y_t = \beta_0 + \beta_1 x_t, t = 1, 2, \ldots$. Smith suggested using Wald's method [Wald (1940)] to estimate $\beta_1$. Grubbs' procedure could then be employed with the true value $\beta_1$ replaced by its esti-
mate to obtain consistent estimates of instrument precisions.

The relationship of Grubbs' procedure to analysis of variance techniques was pointed out by Gaylor (1956) and Russell and Bradley (1958). Consider the two-way classification model

\[ y_{tj} = \mu + \alpha_t + \beta_j + e_{tj}, \quad t = 1,2,\ldots,n, \]
\[ j = 1,2,\ldots,p+1. \quad (1.7) \]

Gaylor viewed (1.7) as a random effects model where \( \alpha_t, \beta_j, \) and \( e_{tj} \) are all assumed to be independently distributed with expected values equal to zero, \( E(\alpha_t^2) = \sigma^2, \quad E(\beta_j^2) = \sigma^2, \) and \( E(e_{tj}^2) = \sigma^2. \) If we write \( \mathbf{z}_t = (y_{t1}, y_{t2}, \ldots, y_{tp})', \quad \mathbf{X}_t = (y_{t,p+1})', \quad \mathbf{z}_t = (\mu+\alpha_t)', \) and \( \mathbf{\varepsilon}_t = (\beta_1 + e_{t1}, \beta_2 + e_{t2}, \ldots, \beta_{p+1} + e_{t,p+1})' \) we see that model (1.7) is a special case of Grubbs' model with \( \chi_{\varepsilon\varepsilon} = \sigma^2 I_{(p+1) \times (p+1)}. \)

Gaylor proved that a component of variance estimate of \( \sigma^2 \) is equivalent to Grubbs' estimate of product variance, \( \sigma^2. \)

Russell and Bradley viewed (1.7) as a fixed effects model, but assumed \( e_{tj} \) to be a normal variate, independent of any other \( e_{t',j'}, \) with zero mean and variance \( \text{Var}(e_{tj}) = \sigma^2, \quad t = 1,2,\ldots,n, \quad j = 1,2,\ldots,p+1. \) An estimate, \( Q_s, \) of \( \sigma^2, \) \( s = 1,2,\ldots,p+1, \) which is a quadratic form in the original observations \( y_{tj}, \) has the general form
Russell and Bradley imposed the following conditions on $Q_s$:

(i) $Q_s$ must be invariant under interchange of order of items.

(ii) $Q_s$ must be independent of the parameters $\mu$, $a_t$, and $b_j$.

(iii) $Q_s$ must be an unbiased estimator of $\sigma_s^2$.

With these conditions, the estimators (1.8) are identical to Grubbs' estimators for instrument precision. For the case $p = 2$, Russell and Bradley also demonstrated the equivalence of restricted maximum likelihood estimators of $\sigma_s^2$, $s = 1, 2, 3$ and their quadratic form estimators.

For the two instrument case ($p = 1$), Thompson (1962) used restricted maximum likelihood estimation to derive non-negative estimators of instrument precision. While removing the first objection to Grubbs' procedure, Thompson did not consider the problems of nonunique estimates when $p \geq 2$, or the restrictiveness of the assumption that $b_j = \lambda$. Thompson (1963), for the two instrument case, presented exact tests of significance for the relative precisions of the instruments. The relative precision of an instrument is the ratio of the instrument precision to the product variance.

In addition, Thompson tabled values of parameters used to form simultaneous confidence regions for $\sigma_x^2$, $\sigma_l^2$, and $\sigma_2^2$.

Prompted by an example from the medical field, Barnett
(1969) generalized Grubbs' model by allowing arbitrary coefficients on the true value for all but one "reference" instrument. Estimators for the parameters of this model were obtained by analogy to Grubbs' estimators, by equating the covariance matrix, $\mathbf{Z}$, of the observed variables to the sample covariance matrix $\mathbf{V}_n$. Through the use of Taylor's series expansions, Barnett derived the asymptotic variances of his estimators.

Several significance tests for making inferences about instrument precision have been suggested. These tests are based upon Grubbs' estimators. For the case where the data consist of one reading with one device and two additional independent readings with a second device on each of $n$ units, Hahn and Nelson (1970) demonstrated how to form exact one and two sided tests of the equality of instrument precisions, and confidence intervals for the ratio of instrument precisions. Maloney and Rastogi (1970) presented an exact test of equality of instrument precision based upon the test of Morgan (1939) and Pitman (1939) for comparing the marginal variances in a bivariate distribution. For testing the hypothesis that one of the two instruments is precise, i.e., that instrument's precision variance is zero, Maloney and Rastogi proposed an approximate likelihood ratio test.

Jaech (1971) extended the results of Maloney and Rastogi
using approximate likelihood ratio tests of the hypotheses that: (1) one of the two precisions is equal to some specified value; (2) the two precisions are jointly equal to specified values; (3) the ratio of the two precisions is some specified constant. Grubbs (1973) and Jaech (1976) extended the results of Maloney and Rastogi to the case of more than two instruments. Shukla (1973) derived the exact distribution of Maloney and Rastogi's statistic which tests that instrument precision is zero. Shukla's result can be used to obtain exact confidence limits for the relative precision.

A special case of the multivariate linear structural model occurs when observations on two variables are sampled from two or more distinct populations. Cox (1976) proposed the following model for such data

\[ Y_{it} = \beta_0 + \beta_1 x_{it} + e_{it} \]

\[ X_{it} = x_{it} + u_{it}, \quad i = 1,2,\ldots,p, \]

\[ t = 1,2,\ldots,n_i, \quad (1.9) \]

where the underlying independent variables, $x_{it}$, are assumed to have independent normal distributions with means $\mu_i$ and a common variance, $\sigma^2_x$. The $u_{it}$ are assumed to have independent normal distributions with means zero and variance $\sigma^2_u$, and are distributed independently of the $e_{it}$ which also have independent normal distributions with means...
zero and variance $\sigma_e^2$. The observed variables $(Y_{it}, X_{it})$
thus have independent normal distributions with means
$(\beta_0 + \beta_1 u_i, u_i)$ and common variance

$$\Sigma = \begin{bmatrix} \beta_1 \sigma_x^2 + \sigma_e^2 & \beta \sigma_x^2 \\ \beta \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{bmatrix}. \quad (1.10)$$

The parameter space is restricted by the inequalities
$\sigma_x^2 > 0$, $\sigma_e^2 > 0$, $\sigma_u^2 > 0$, and $|\Sigma| > 0$. Cox found the maximum likelihood estimates of the parameters of model (1.9)
and developed a method for testing hypotheses about $\beta_1$.

Browne (1974) considered the general problem of estimating the $r \times 1$ parameter vector, $\gamma_0$, of a $p \times p$
covariance matrix $\Sigma_0 = \Sigma(\gamma_0)$ when one has available independent observations of the $1 \times p$ random vectors $x_t^T$, $t = 1, 2, \ldots, n$. The $x_t$ are assumed to be identically distributed with mean $\mu_0$ and covariance matrix $\Sigma_0$. Letting $\Sigma_n$ represent the usual unbiased estimator of $\Sigma_0$ obtained from the $x_t$, Browne defined a generalized least squares (G.L.S.) estimator of $\gamma_0$ to be an estimator obtained by minimizing a weighted distance between the elements of $\Sigma_n$ and $\Sigma_0$. Anderson (1973), considering linear covariance structures, showed that a G.L.S. estimator converges in probability to the maximum likelihood estimator. We shall describe more fully the G.L.S. procedure in subsequent
chapters where we use this approach to estimate the parameters of the multivariate linear structural errors in variables model.

Others have written about the structural model, including Mandel (1959) who considered a model similar to Barnett's, and Draper and Guttman (1975) who found the Bayes estimators of the parameters of Grubbs' two instrument model.

2. Functional relationships

Considerably less has been written about the functional multivariate model than the structural multivariate model. Anderson (1976), using results from earlier research [Anderson (1951)], obtained maximum likelihood estimates of $\beta$ for the situation where repeated observations are available.

Let $(\tilde{y}_{tj}, \tilde{x}_{tj}), \ t = 1,2,...,n, \ j = 1,2,...,r$ denote $r$ repeated observations, where the true values, $(y_t, x_t)$, satisfy

$$y_t = x_t^\beta.$$  \hspace{1cm} (1.11)

By writing the model in terms of the observable random variables we have

$$\tilde{y}_{tj} = \tilde{x}_{tj}^\beta + \tilde{w}_{tj},$$
where \( w_{tj} = e_{tj} - u_{tj} \beta \). Assume that \( \xi_{tj} = (e_{tj}, u_{tj}) \), \( t = 1,2,\ldots,n, \ j = 1,2,\ldots,r, \) are independent drawings from a \((p+k)\)-variate normal distribution with mean vector zero and covariance matrix \( \Sigma_{\xi\xi} \). Then, \( w_{tj}, \ t = 1,2,\ldots,n, \ j = 1,2,\ldots,r, \) are independent drawings from a \( p \)-variate normal distribution with mean vector zero and covariance matrix

\[
\Sigma_{ww} = (I_p \times p', -\beta')', \quad \Sigma_{\xi\xi} (I_p \times p', -\beta').
\]

Define \( \Gamma' = (\Sigma_{ww}' - \gamma_{ww} \beta) \) where \( \gamma_{ww}^{-1} = \gamma_{ww}^{-1} \). Then the following relationships hold.

\[
\Gamma' \Sigma_{\xi\xi} \Gamma = I_p \times p' \tag{1.12}
\]

and

\[
\Gamma' \pi = 0_p \times n' \tag{1.13}
\]

where \( \pi \) is the \((p+k) \times n\) matrix with \( t^{th} \) column \((y_t, \bar{x}_t)\). Anderson derived the maximum likelihood estimator of \( \Gamma \), and hence, of \( \beta \).

A problem arising in the earth sciences was formulated as a multivariate functional relationship by Gleser and Watson (1973). Gleser and Watson derived the maximum likelihood estimators of the multivariate linear functional model, where \( p = k, \ n \geq 2p, \) and the \( 2p \times 2p \) block diagonal matrix \( \Sigma_{\xi\xi} = \sigma^2 \text{diag}(\bar{\Sigma}, \bar{\Sigma}) \) is known up to the multiple \( \sigma^2 \). The maximum likelihood estimator of \( \beta \) is consistent, but
the estimators for \( x_t \), \( t = 1,2,\ldots,n \), and \( \sigma^2 \) are inconsistent. The classical maximum likelihood large sample theory does not apply because, as \( n \to \infty \), the ratio of the number of parameters to the number of observations tends not to zero but to \( \frac{1}{2} \). Gleser and Watson were unable to derive either the asymptotic distribution or the variance of the estimator for \( \beta \).

Bhargava (1977) generalized the model of Gleser and Watson by letting \( \Sigma = \text{diag} ( \Sigma_1, \Sigma_2 ) \), be unknown. While able to prove the existence of a solution to the maximum likelihood equations, Bhargava could express the solution in closed form only if \( \Sigma \) and \( \beta \) have the same known eigenvectors.

A synthesis of the functional and structural relations was examined by Dolby (1976). Let \( Y_t \) and \( X_t \), \( t = 1,2,\ldots,n \) be \( 1 \times p \) random vectors satisfying the linear structural relationship. In addition, assume that the true independent values, \( x_t \), \( t = 1,2,\ldots,n \), are independently distributed as normal random vectors with distinct expectations \( \mu_t = \mu_{t,1} \), where \( \mu_{t,1} \) is a \( p \)-dimensional column vector of ones and \( \mu_t \) is an unknown scalar constant, and common covariance matrix \( \Sigma_{xx} = \sigma_x^2 \mathbf{I}_{p \times p} \). If \( \sigma_x^2 = 0 \), the above model specializes to the univariate linear functional model with replicated observations. Dolby termed this hybrid model the ultrastructural relation and derived
its maximum likelihood solution.

Höschel (1978) considered estimation of a general class of functional relationships. Let $F$ be a family of surfaces, or manifolds, defined on some subset of $\mathbb{R}^P$, which describes our knowledge about the type of functional relationship among the unobservable variables $u_t$, $t = 1, 2, \ldots, n$. The points $u_t$, $t = 1, 2, \ldots, n$ lie on a fixed but unknown "true" surface, $F \in F$. The surface $F$ is said to be identified if knowledge of $u_t$, $t = 1, 2, \ldots, n$, uniquely determines $F$. We observe $z_t$, $t = 1, 2, \ldots, n$, where

$$z_t = u_t + e_t, \quad t = 1, 2, \ldots, n, \quad (1.14)$$

and the $e_t$, $t = 1, 2, \ldots, n$ are independently and identically distributed as $p$-dimensional normal random vectors with mean zero and known covariance matrix $\Sigma_{ee}$. Hence, G.L.S. and M.L. estimation of $u_t$, $t = 1, 2, \ldots, n$, are equivalent. This model generalizes model (1.6) by allowing the true surface $F$ to be arbitrary. Höschel gave sufficient conditions to ensure that a G.L.S. solution for $u_t$, $t = 1, 2, \ldots, n$ exists and that the solution uniquely determines an estimate of the true functional relationship, $F$, with probability one.

For the multivariate linear functional model, where $\Sigma_{ee}$ is known, Nussbaum (1978) defined a class, $\phi$, of estimators of $\hat{\beta}$ which is contained within the class of consistent asymptotically normal estimators of $\hat{\beta}$. Let $M_k \times p$
be the set of all \(k \times p\) matrices. Define \(\mathcal{U}\) to be the set of all row spaces generated by \(k \times (p+k)\) matrices of the form \(\Theta = (\Theta_1, \Theta_2)\), where the \(k \times k\) submatrix \(\Theta_2\) is of full rank. The function \(f\), mapping \(M_{k \times p}\) into \(\mathcal{U}\), is defined by \(f(\beta) = R(\beta, I_k \times k)\), where \(R(A) = \) the row space of a matrix \(A\). Nussbaum claimed \(f\) is continuous and bi-unique. Define \(D_n = n^{-1} \sum_{t=1}^{n} Z_t'Z_t - Z_{\xi} Z_{\xi}'\). An estimator \(\hat{\beta}\) of \(\beta\) belongs to the class \(\phi\) if there is a sequence \(\{C_n\}\) of \((p + k) \times k\) random matrices depending on \(Z_t\), \(t = 1, 2, \ldots, n\) and a nonrandom \((p+k) \times k\) matrix \(C\) depending on \(\beta\) and \(x_t\), \(t = 1, 2, \ldots, n\) such that for some \(n_0\)

1. \(C_n\) converges to \(C\) in probability,
2. \(\text{rank } [(\beta, I_k \times k)C] = k\),
3. \(R(D_n C_n) \in \mathcal{U}\) a.e. for \(n = n_0, n_0 + 1, \ldots\),
4. \(\hat{\beta} = f^{-1}(R(D_n C_n))\) a.e. \(n = n_0, n_0 + 1, \ldots\) (1.15)

If a M.L. estimate of \(\beta\) exists, it is contained in \(\phi\).

By appropriate choice of \(C\), Nussbaum also demonstrated that an asymptotically optimal estimator, \(\beta^*\), exists within the class \(\phi\). The estimator \(\beta^*\) is optimal in the sense that the matrix which is the difference between the asymptotic covariance of any \(\hat{\beta} \in \phi\) and \(\beta^*\) is positive semidefinite.
3. Factor analysis

The multivariate linear model (1.1-1.5) can be viewed as the classical factor analysis model

\[ Z_t = X_t \Lambda + \xi_t, \quad t = 1, 2, \ldots, n, \]  

(1.16)

where \( Z_t = (Y_t, X_t), \Lambda = (\beta, I_{k \times k}), \) and the \( \xi_t = (\varepsilon_t, u_t) \) are independently and identically distributed as \((p + k)\)-variate normal vectors with mean zero and diagonal covariance matrix \( Z_{\xi \xi} \). The elements of the \( k \times (p + k) \)

matrix \( \Lambda \) are called the factor loadings and the elements of \( x_t \) are called common factors, or factors. As with the multivariate linear model, the common factors may be random or fixed. If the \( x_t \) are random, we assume they are independent normal random vectors with mean vector \( 0 \) and covariance matrix \( Z_{xx} \). Thus, if \( x_t \) is random it follows that the covariance matrix of \( Z_t \), say \( Z \), is

\[ Z = \Lambda' Z_{xx} \Lambda + Z_{\xi \xi}. \]  

(1.17)

For the random factor case let \( S_n \) denote the usual unbiased estimator of the sample covariance matrix of the \( Z_t, t = 1, 2, \ldots, n \). Lawley (1940) examined maximum likelihood estimation based on the distribution of \( S_n \) of model (1.16) with \( \Lambda \) unrestricted, but assuming \( \Lambda' Z_{xx}^{-1} \Lambda \) is diagonal and \( Z_{xx} = I_{k \times k} \). The likelihood equations do not
yield explicit solutions. Lawley suggested an iterative procedure which, while usually converging to a solution, does so slowly. In addition, Lawley gave a method, suitable for large samples, of testing hypotheses concerning the number of factors required. The estimation procedures were illustrated by Lawley (1943) using data from the field of education.

Maximum likelihood estimation when the common factors are fixed was investigated by Lawley (1942). However, the stationary points of the likelihood function obtained by Lawley are not absolute minima - a fact subsequently proven by Anderson and Rubin (1956). For the case where the common factors are fixed and \( \Lambda \) is diagonal with at least two distinct elements, Anderson and Rubin demonstrated that the likelihood function of the \( \tilde{Z}_t \) has no maximum.

Whittle (1952) considered model (1.16) with \( \tilde{Z} = \sigma^2 I (p+k) \times (p+k) \) and fixed factors. Least squares solutions were obtained by minimizing \( \sum_{t=1}^{n} (\tilde{Z}_t - x_t \Lambda) (\tilde{Z}_t - x_t \Lambda)' \) with respect to the factors and factor loadings. These solutions are maximum likelihood solutions if \( \varepsilon_t \) is normally distributed. The model considered by Gleser and Watson (1973) is similar to Whittle's model and, in fact, an indirect derivation of the results of Gleser and Watson can be obtained through Whittle's solution.
Statistical inference of model (1.16) was discussed by Anderson and Rubin (1956). They considered the following problems.

1. **Existence of the model**: If the factors are normally distributed, the variance of the observed variables is given by (1.17). Given a positive definite matrix $\Sigma^*$, conditions were given for the existence of a $\Lambda$, $\Sigma_{xx}$, and diagonal $\Sigma_{ee}$ such that $\Sigma^* = \Lambda' \Sigma_{xx} \Lambda + \Sigma_{ee}$.

2. **Identification**: Suppose there is some $\Lambda$, $\Sigma_{xx}$, and $\Sigma_{ee}$ such that $\Sigma^* = \Lambda' \Sigma_{xx} \Lambda + \Sigma_{ee}$. Anderson and Rubin enumerated both necessary and sufficient rank conditions for this equation to have a unique solution.

3. **Determination of the structure**: Given $\Sigma$, and if there exists a unique solution to (1.17), then one can determine $\Lambda$, $\Sigma_{xx}$, and $\Sigma_{ee}$. The method of determination depends on the identification conditions.

4. **Estimation of parameters**: A sample of $n$ observations, $Z_1, Z_2, \ldots, Z_n$, is drawn. Anderson and Rubin discussed properties of maximum likelihood estimation under various assumptions needed for identification.
5. **Test of hypothesis that the model fits:** One can test the hypothesis that \( \xi \) is of the form
\[
\Lambda' \xi \Lambda + \xi \varepsilon \varepsilon
\]
using a likelihood ratio test.

6. **Determination of the number of factors:** In situations where the number of factors cannot be specified in advance, ad hoc procedures for determining the "right" number of factors were outlined.

7. **Other tests of hypotheses:** Various hypotheses about the parameters, particularly \( \Lambda \), are of interest. Unfortunately, such hypotheses are difficult to test even in large samples, because the asymptotic variances of the parameters are complicated functions of the parameters. However, the hypothesis that the \( j \)th column of \( \Lambda \) is zero can be tested by using the multiple correlation between \( Z_j = (Z_{1j}, Z_{2j}, \ldots, Z_{nj}) \) and the remaining \( Z_{\ell} \), \( \ell = 1, 2, \ldots, p+k, \ell \neq j \).

8. **Estimation of factor scores:** If one postulates a model where the \( x_t, t = 1, 2, \ldots, n \) are nonrandom, Anderson and Rubin showed that a maximum likelihood solution of \( \Lambda, \xi \varepsilon \varepsilon \) and \( x_t \) does not exist in general. It is this result which disproves Lawley's (1942) claim that his stationary points of the likelihood function yield maximum likelihood estimates. Anderson and Rubin suggested first estimating \( \Lambda \)
and }\epsilon \text{ by maximum likelihood methods, treating } x_t \text{ as random. Treating the estimates of } \Lambda \text{ and } \Gamma_{\epsilon \epsilon} \text{ as known, maximum likelihood estimation could then be used again to estimate the } x_t. \text{ Other approaches suggested by Anderson and Rubin include the least squares approach of Whittle (1952), and an approach suggested by Thomson (1934).}

For estimating } \Lambda \text{ and } \Lambda' \Gamma_{\epsilon \epsilon} \Lambda \text{ when the factors are non-random and } \Lambda' \Gamma_{\epsilon \epsilon} \Lambda \text{ is diagonal, Anderson and Rubin applied the method of maximum likelihood to the distribution of } V_n = n^{-1} \sum_{t=1}^{n} (z_t - \bar{z})' (z_t - \bar{z}), \text{ which is noncentral Wishart. They proved that the estimates based on maximizing the non-central Wishart likelihood function are asymptotically equivalent to the maximum likelihood estimates for random factors in the sense that } n^2 \text{ times the difference of the two estimates converges to zero in probability. Thus, for large samples, one can treat nonrandom factors as if they are normally distributed.}

Jöreskog (1970) presented an iterative procedure for producing maximum likelihood estimates when the covariance matrix of observations taken from a normal population is of a general parametric form. The procedure subsumes maximum likelihood estimation of factor analysis models, systems of simultaneous equations, variance components, and path analysis models.
The generalized least squares principal was applied by Jöreskog and Goldberger (1972) to model (1.16) when 
\[ \xi_{xx} = I_{k \times k} \] and \( \Lambda \) is unrestricted. The covariance matrix of \( \xi_t \) is then \( \xi_t = \Lambda' \Lambda + \xi_{\varepsilon \varepsilon} \), \( \varepsilon \) where \( \xi_{\varepsilon \varepsilon} \) is diagonal. Assume that \( S_n = (n-1)^{-1} \sum_{t=1}^{n} (\xi_t - \bar{\xi}) (\xi_t - \bar{\xi})' \) converges stochastically to \( \xi \) and that the elements of \( \frac{1}{n^2}(S_n - \xi) \) have an asymptotic multivariate normal distribution with covariances given by 
\[ nE[(s_{ij} - \sigma_{ij})(s_{kl} - \sigma_{kl})] = \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{jk}, \] where \( s_{ij} \) and \( \sigma_{ij} \) are the \( (i,j) \)th elements of \( S_n \) and \( \xi \), respectively. Then Jöreskog and Goldberger showed that the G.L.S. and M.L. estimates of \( \xi_{\varepsilon \varepsilon} \) have the same asymptotic properties. Browne (1974) made the same distributional assumptions about \( S_n \) to derive asymptotic properties of G.L.S. estimates for covariance matrices, \( \xi \), which are general functions of unknown parameter vectors, \( \gamma \).

Numerous additional articles have been written on factor analysis models which relate to the multivariate errors in variables problem. These include Thurstone (1947), Anderson (1963), Jöreskog (1967, 1971, 1973), and Lawley and Maxwell (1971).

The remainder of this dissertation considers estimation of the multivariate linear errors in variables model. Chapter II presents definitions and preliminary algebraic results. Chapters III and IV deal with G.L.S. estimation of
structural and functional relationships, respectively, when the random variables are assumed to be normally distributed. In Chapter V we extend our results to nonnormal random variables. Finally, in Chapter VI, we present two examples which illustrate our results.
II. DEFINITIONS AND PRELIMINARY RESULTS

Proofs of many of the results in this dissertation involve performing various operations on the elements of matrices. We devote this chapter to developing a convenient notation for this purpose, and to presenting some preliminary algebraic results.

Let $A$ be an arbitrary $p \times q$ matrix with $(i,j)$th element $a_{ij}$. The vec of $A$, denoted vec $A$, is defined to be the $pq \times 1$ column vector obtained from stacking the columns of the matrix $A$ one beneath the other in a single column vector.

Definition 2.1: Let $A = (a_{ij})$ be a $p \times q$ matrix. Then

$$\text{vec } A = (a_{11}, a_{21}, \ldots, a_{p1}, a_{12}, a_{22}, \ldots, a_{p2}, a_{13}, a_{23}, a_{33}, \ldots, a_{pq})'. \quad (2.1)$$

The vector vech $A$ is defined similarly to vec $A$, except that the vech operator is confined to $A$ being square, and only that portion of each column of $A$ that is on and below the diagonal is put into vech $A$.

Definition 2.2: Let $A = (a_{ij})$ be a $p \times p$ matrix. Then

$$\text{vech } A = (a_{11}, a_{21}, a_{31}, \ldots, a_{p1}, a_{22}, a_{32}, \ldots, a_{p2}, a_{33}, a_{43}, \ldots, a_{pp})'. \quad (2.2)$$
If \( \tilde{A} \) is a \( p \times p \) symmetric matrix, then \( \text{vech}\tilde{A} \) contains the distinct elements of \( \tilde{A} \), and there exists a unique \( p^2 \times 2^{-1}p(p+1) \) matrix which transforms \( \text{vech}\tilde{A} \) into \( \text{vec}\tilde{A} \). We shall denote by \( G \) this unique \( p^2 \times 2^{-1}p(p+1) \) matrix.

**Definition 2.3:** Let \( \tilde{A} = (a_{ij}) \) be a \( p \times p \) symmetric matrix. We define the \( p^2 \times 2^{-1}p(p+1) \) matrix \( G \) to be that matrix which satisfies

\[
\text{vec}\tilde{A} = G \text{vech}\tilde{A}. \tag{2.3}
\]

We note that \( G \) is of full column rank.

While \( G \) is well-defined by (2.3), there are many linear transformations of \( \text{vec}\tilde{A} \) into \( \text{vech}\tilde{A} \). We define a particular matrix which transforms \( \text{vec}\tilde{A} \) into \( \text{vech}\tilde{A} \).

**Definition 2.4:** Let \( \tilde{A} = (a_{ij}) \) be a \( p \times p \) symmetric matrix. Let \( G \) be defined by (2.3), and let \( \tilde{H} = (G'G)^{-1}G' \). Then

\[
\tilde{H} \text{vec}\tilde{A} = (G'G)^{-1}G'(G \text{vech}\tilde{A}) = \text{vech}\tilde{A}. \tag{2.4}
\]

To illustrate these definitions, we consider a \( 3 \times 3 \) symmetric matrix \( \tilde{A} \),
\[ \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} , \quad a_{ij} = a_{ji} \text{ for } i \neq j. \]

Then

\[ \text{vec } \mathbf{A} = (a_{11}, a_{21}, a_{31}, a_{12}, a_{22}, a_{32}, a_{13}, a_{23}, a_{33})' \]

and

\[ \text{vech } \mathbf{A} = (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})' . \]

The 9x6 matrix \( \mathbf{G} \) is defined by the relationship

\[ \mathbf{vec} \mathbf{A} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{12} \\ a_{22} \\ a_{32} \\ a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{bmatrix} = \mathbf{G} \text{ vec } \mathbf{A}. \]

The 6x9 matrix \( \mathbf{H} \) is then defined to be
and we see that

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

\[H = H \sim \] 

\[H \sim \text{vec } A = \begin{bmatrix} a_{11} \\ \frac{1}{2}(a_{21} + a_{12}) \\ \frac{1}{2}(a_{31} + a_{13}) \\ a_{22} \\ \frac{1}{2}(a_{32} + a_{23}) \\ a_{33} \end{bmatrix} = \text{vech } \sim A, \]

since \( \sim A \) is symmetric.

Double subscripts, \( ij \), are used to denote elements of \( \sim \text{vec } A \) and \( \sim \text{vech } A \), the first subscript always being nested within the second. Double subscripts will also be used to represent rows or columns of certain matrices. Use of Kronecker's delta and the double subscript notation provide alternative definitions of the matrices \( \sim G \) and \( \sim H \) of Definitions 2.3 and 2.4. We first define Kronecker's delta.
Definition 2.5: Kronecker's delta, denoted $\delta_{ij}$, is defined to be

$$\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{otherwise}.
\end{cases}$$

We have the alternative, but equivalent, definitions of $G$ and $H = (G'G)^{-1}G'$.

Definition 2.6: The $(ij,kl)$th element of the matrix $H$ of Definition 2.4 can be expressed as

$$[H]_{ij,kl} = 2^{-1}(\delta_{kj}\delta_{li} + \delta_{ki}\delta_{lj}), \quad j \leq i \leq p, \quad k \leq l \leq p. \quad (2.6)$$

Definition 2.7: The $(ij,kl)$th element of the matrix $G$ of Definition 2.3 can be expressed as

$$[G]_{ij,kl} = (2^{-1}\delta_{ij})[H]_{kl,ij'}, \quad i \leq j \leq p, \quad k \leq l \leq p \quad (2.7)$$

Similarly, we have two equivalent ways of defining the direct, or Kronecker product of a $p \times q$ matrix $A$ and an $m \times n$ matrix $B$.

Definition 2.8: The Kronecker product of a $p \times q$ matrix $A = (a_{ij})$ and an $m \times n$ matrix $B = (b_{ij})$, denoted $A \otimes B$, is the $pm \timesqn$ matrix
\[ A \otimes B = \begin{bmatrix}
  a_{11}^B & a_{12}^B & \cdots & a_{1q}^B \\
  a_{21}^B & a_{22}^B & \cdots & a_{2q}^B \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{p1}^B & a_{p2}^B & \cdots & a_{pq}^B 
\end{bmatrix} \]

\[
\begin{bmatrix}
  a_{11}b_{11} & a_{11}b_{1n} & a_{12}b_{11} & a_{12}b_{1n} & \cdots & a_{1q}b_{11} & a_{1q}b_{1n} \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{11}b_{ml} & a_{11}b_{mn} & a_{12}b_{ml} & a_{12}b_{mn} & \cdots & a_{1q}b_{ml} & a_{1q}b_{mn} \\
  a_{21}b_{11} & a_{21}b_{1n} & a_{22}b_{11} & a_{22}b_{1n} & \cdots & a_{2q}b_{11} & a_{2q}b_{1n} \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{p1}b_{11} & a_{p1}b_{1n} & a_{p2}b_{11} & a_{p2}b_{1n} & \cdots & a_{pq}b_{11} & a_{pq}b_{1n} \\
  \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{p1}b_{ml} & a_{p1}b_{mn} & a_{p2}b_{ml} & a_{p2}b_{mn} & \cdots & a_{pq}b_{ml} & a_{pq}b_{mn} 
\end{bmatrix}
\]

(2.8)

**Definition 2.9:** The \((ij,kl)\)th element of the matrix which is the Kronecker product of a \(p \times q\) matrix \(A = (a_{ij})\) and an \(m \times n\) matrix \(B = (b_{ij})\) can be expressed as
\[
[A \otimes B]_{ij,kl} = a_{jk} b_{ik}.
\] (2.9)

We state the following properties of Kronecker matrix products.

**Theorem 2.1**: Assume the matrices \( A, B, C, \) and \( D \) are suitably conformable. Then

\[
(i) \quad (A \otimes B)(C \otimes D) = (A \otimes C) \otimes (B \otimes D) \quad (2.10)
\]

\[
(ii) \quad (A \otimes B)^{-1} = A^{-1} \otimes B^{-1} \quad (2.11)
\]

\[
(iii) \quad (A \otimes B)' = A' \otimes B' \quad (2.12)
\]

\[
(iv) \quad (A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D) \quad (2.13)
\]

**Proof**: The proofs of (i-iv) follow easily from the definitions of the Kronecker product.

We next define and state properties of the matrix which is the product of the matrices \( G \) and \( H \).

**Definition 2.10**: Let \( \sim_p \) be the \( p^2 \times p^2 \) symmetric idempotent matrix defined by

\[
\sim_p = \sim \sim = \sim \sim (G'G)^{-1} G',
\] (2.14)

where \( G \) and \( H \) are defined by (2.3) [or (2.7)] and (2.4) [or (2.6)]. We may express a typical element of \( \sim_p \) as
\[ [K_p]_{ij,kl} = 2^{-1}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \]

\[ i \leq p, \ j \leq p, \ k \leq p, \ l \leq p. \quad (2.15) \]

Theorem 2.2: Let \( A = (a_{ij}) \) be a \( p \times q \) matrix, and let \( K_p \) be defined by (2.14) [or (2.15)]. Then

\[ K_p (A \circ A) = (A \circ A) K_p. \quad (2.16) \]

Proof:

Using (2.15), we can write the \((ij,kl)\)th element of \( K_p(A \circ A) \) as

\[ [K_p(A \circ A)]_{ij,kl} = \sum_{s=1}^{p} \sum_{t=1}^{p} [K_p]_{ij,st} (A \circ A)_{st,kl} \]

\[ = \sum_{s=1}^{p} \sum_{t=1}^{p} 2^{-1}(\delta_{is}\delta_{jt} + \delta_{it}\delta_{js}) (A \circ A)_{st,kl} \]

\[ = 2^{-1} [(A \circ A)_{ij,kl} + (A \circ A)_{ji,kl}], \]

which, from (2.9),

\[ = 2^{-1} [(A \circ A)_{ij,kl} + a_{il}a_{jk}] \]

\[ = 2^{-1} [(A \circ A)_{ij,kl} + (A \circ A)_{ij,kl}] \]

\[ = \sum_{u=1}^{q} \sum_{v=1}^{q} 2^{-1}(\delta_{uk}\delta_{vl} + \delta_{ul}\delta_{vk}) (A \circ A)_{ij,uv} \]

\[ = \sum_{u=1}^{q} \sum_{v=1}^{q} (A \circ A)_{ij,uv} [K_q]_{uv,kl} \]

\[ = [(A \circ A) K_q]_{ij,kl}. \]
Theorem 2.3: Let $\sim A$ be a $p \times p$ nonsingular matrix, and let $\sim G$ and $\sim H$ be defined by (2.3) [or (2.7)] and (2.4) [or (2.6)]. Then

$$[\sim H(\sim A \otimes \sim A)\sim H']^{-1} = \sim G'(\sim A^{-1} \otimes \sim A^{-1})\sim G.$$ \hspace{1cm} (2.17)

Proof: The result may be verified by multiplication using (2.14), (2.16), and (2.11),

$$\sim H(\sim A \otimes \sim A)\sim H' \sim G'(\sim A^{-1} \otimes \sim A^{-1})\sim G = \sim H(\sim A \otimes \sim A)\sim (\sim A^{-1} \otimes \sim A^{-1})\sim G$$

$$= \sim H \sim G \sim H \sim G$$

$$= I_{p \times p}$$

Similarly it may be shown that $\sim G'(\sim A^{-1} \otimes \sim A^{-1})\sim G$ is a left inverse of $\sim H(\sim A \otimes \sim A)\sim H'$. The result follows.

A useful result involving the vec operator and the Kronecker product operator is given in Theorem 2.4.

Theorem 2.4: Let $\sim A = (a_{ij})$, $\sim B = (b_{ij})$, and $\sim C = (c_{ij})$ be $p \times q$, $q \times m$, and $m \times n$ matrices, respectively. Then

$$\text{vec}(\sim A \otimes \sim B \otimes \sim C) = (\sim C' \otimes \sim A) \text{ vec } \sim B.$$ \hspace{1cm} (2.18)

Proof: Let $\sim C. j$ be the vector formed from the $j^{th}$ column of $\sim C$, and let $\sim B. i$ be similarly defined. The $j^{th}$ sub-vector of $\text{ vec}(\sim A \otimes \sim B \otimes \sim C)$ equals...
\[ A \otimes B \otimes C \cdot j = \sum_i c_{ij} A \otimes B \cdot i \]

\[ = (C' \otimes A \otimes j) \cdot \text{vec } B. \]

The result follows.

It frequently will be useful to express a quadratic or bilinear form involving a Kronecker product as a trace using Theorem 2.5.

**Theorem 2.5**: Let \( A = (a_{ij}) \), \( B = (b_{ij}) \), \( C = (c_{ij}) \), and \( D = (d_{ij}) \) be \( p \times q \), \( q \times m \), \( p \times n \), and \( n \times m \) matrices, respectively. Then, if the trace of a square matrix \( E \) is denoted by \( \text{tr } E \),

\[ (\text{vec } A)'(B \otimes C)(\text{vec } D) = \text{tr}(A \otimes B \otimes D'C'). \] (2.19)

**Proof**: From (2.18) we have
\[ \text{vec}(C \otimes D \cdot B') = (B \otimes C)(\text{vec } D). \]

But
\[ (\text{vec } A)' \cdot \text{vec}(C \otimes D \cdot B') = \text{tr}(A'C \otimes D \cdot B') = \text{tr}([C \otimes D)(B'A')] = \text{tr}([C \otimes D)(A B')]' = \text{tr}(A \otimes B \otimes D'C'). \]

Next we define various matrix derivatives and state some theorems regarding their use.

**Definition 2.11**: Let \( A = A(\gamma) \) be a \( p \times q \) matrix with typical element \( a_{ij} = a_{ij}(\gamma) \), where \( \gamma = (\gamma_1, \gamma_2, ..., \gamma_r)' \).

We define \( \frac{\partial A}{\partial \gamma_i} \) to be
\[
\begin{align*}
\frac{\partial a}{\partial \gamma_i} &= \begin{bmatrix}
\frac{\partial a_{11}}{\partial \gamma_i} & \frac{\partial a_{12}}{\partial \gamma_i} & \cdots & \frac{\partial a_{1q}}{\partial \gamma_i} \\
\frac{\partial a_{21}}{\partial \gamma_i} & \frac{\partial a_{22}}{\partial \gamma_i} & \cdots & \frac{\partial a_{2q}}{\partial \gamma_i} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial a_{pq}}{\partial \gamma_i} & \frac{\partial a_{p2}}{\partial \gamma_i} & \cdots & \frac{\partial a_{pq}}{\partial \gamma_i}
\end{bmatrix} 
\end{align*}
\tag{2.20}
\]

**Definition 2.12:** Let \( \tilde{a} = \tilde{a}(\gamma) \) be a \( p \times 1 \) vector with typical element \( a_{ij} = a_{ij}(\gamma) \), and where \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)' \). We define \( \frac{\partial \tilde{a}}{\partial \gamma} \) to be

\[
\frac{\partial \tilde{a}}{\partial \gamma} = \begin{bmatrix}
\frac{\partial a_1}{\partial \gamma_1} & \frac{\partial a_1}{\partial \gamma_2} & \cdots & \frac{\partial a_1}{\partial \gamma_r} \\
\frac{\partial a_2}{\partial \gamma_1} & \frac{\partial a_2}{\partial \gamma_2} & \cdots & \frac{\partial a_2}{\partial \gamma_r} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial a_p}{\partial \gamma_1} & \frac{\partial a_p}{\partial \gamma_2} & \cdots & \frac{\partial a_p}{\partial \gamma_r}
\end{bmatrix}
\tag{2.21}
\]

and \( \frac{\partial a'}{\partial \gamma} \) to be...
Theorem 2.6: Let $A = A(\gamma)$ be a $p \times p$ symmetric positive definite matrix with typical element $a_{ij}(\gamma)$, where $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_r)^T$. Then

$$\frac{\partial \log |A|}{\partial \gamma_i} = \text{tr}[A^{-1} \frac{\partial A}{\partial \gamma_i}].$$

**Proof:** The expansion of a determinant according to cofactors is

$$|A| = a_{i1}c_{i1} + \ldots + a_{ij}c_{ij} + \ldots + a_{ip}c_{ip},$$

where $c_{ij}$ is the cofactor of $a_{ij}$ in $A$. The only term in (2.24) which depends on $a_{ij}$ is $a_{ij}c_{ij}$, and $c_{ij}$ does not depend on $a_{ij}$. Hence

$$\frac{\partial |A|}{\partial a_{ij}} = c_{ij} = a_{ji} |A|,$$

is the $(i,j)$th element of $A^{-1}$. Thus,
Now,

\[ \frac{\partial \log |\sim A|}{\partial \gamma_i} = \sum_j \sum_k \frac{\partial \log |\sim A|}{\partial a_{jk}} \frac{\partial a_{jk}}{\partial \gamma_i} \]

\[ = \text{tr}[\sim A^{-1} \frac{\partial A}{\partial \gamma_i}] . \]

**Theorem 2.7:** Let \( \sim A = A(\gamma) \) be a \( p \times p \) nonsingular matrix with typical element \( a_{ij}(\gamma) \), where \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)' \). Then

\[ \frac{\partial \sim A^{-1}}{\partial \gamma_i} = -\sim A^{-1} \frac{\partial A}{\partial \gamma_i} \sim A^{-1} . \]

(2.27)
Proof: Let \( B = (b_{ij}) \) be a \( p \times p \) matrix. It is easily shown that

\[
\frac{\partial AB}{\partial \gamma_i} \sim = A \frac{\partial B}{\partial \gamma_i} \sim + B \frac{\partial A}{\partial \gamma_i} \sim .
\]

(2.28)

Suppose that \( B \) in (2.28) is \( A^{-1} \) so that \( A B \sim = I \). Then

\[
0 = A \frac{\partial A^{-1}}{\partial \gamma_i} \sim + (\frac{\partial A}{\partial \gamma_i} \sim ) A^{-1} .
\]

and the result follows.

The topics of vec and vech operators, Kronecker products, and matrix derivatives are discussed in Roth (1934), Neudecker (1969), Tracy and Dwyer (1969), Theil (1971), Browne (1974), and Searle and Quaas (1978). Our brief discussion reviews only a few of the results contained in these references.
III. ESTIMATION OF THE MULTIVARIATE LINEAR STRUCTURAL MODEL

A. Preliminary Notation

In this chapter we consider estimation of the parameters of the structural multivariate linear errors in variables model. We use a generalized least squares (G.L.S.) approach to analyze the sample covariance matrix of the observed random vectors \((Y_t, X_t), \ t = 1,2,\ldots,n\). For the structural relationship our results follow directly from results given in Browne (1974).

It will be convenient to use the form of the structural model given by Equation (1.6),

\[
Z_t = \mu_t + \varepsilon_t, \quad t = 1,2,\ldots,n, \tag{1.6}
\]

where \(Z_t = (Y_t, X_t)\), \(\mu_t = (x_t^\beta, x_t)\), and \(\varepsilon_t = (e_t, u_t)\) and where the definitions and assumptions of model (1.6) are given in Section B, Chapter I. Recall that the \(i^{th}\) elements of \(Z_t, \mu_t,\) and \(\varepsilon_t\) are denoted by \(Z_{it}, \mu_{it}\), and \(\varepsilon_{it}\) respectively.

We consider estimation of the \(r \times 1\) parameter vector \(\gamma_0\) which is composed of the unknown elements of \(\beta, Z_{xx},\) and \(Z_{\varepsilon \varepsilon}\). At times we shall treat the unknown parameter vector \(\gamma_0\) as a vector of mathematical variables. We distinguish this usage of \(\gamma_0\) by writing the unknown
parameter vector as $\gamma$.

Define the sample covariance matrix $\tilde{S}_n$ by

$$\tilde{S}_n = (n-1)^{-1} \sum_{t=1}^{n} (\tilde{Z}_t - \tilde{\bar{Z}})(\tilde{Z}_t - \tilde{\bar{Z}})' \quad (3.1)$$

where $\tilde{\bar{Z}} = n^{-1} \sum_{t=1}^{n} \tilde{Z}_t$. Because $x_t$, $t = 1, 2, \ldots, n$ are independent identically distributed as $k$-variate normal random vectors with mean vector $\mu_x$ and covariance matrix $\Sigma_{xx}$, $x_t$, $t = 1, 2, \ldots, n$ are independent identically distributed as $p$-variate normal random vectors with mean vector $\mu_x$ and covariance matrix $\Sigma_{xx}$. Note that if $p > k$ then the $x_t$, $t = 1, 2, \ldots, n$ have a singular normal distribution and $\Sigma_{xx}$ is a positive semidefinite matrix of rank $k$. Also, $\varepsilon_t$, $t = 1, 2, \ldots, n$ are $(p+k)$-variate normal random vectors with zero mean and covariance matrix $\Sigma_{\varepsilon\varepsilon}$ distributed independently of $x_t$, $t = 1, 2, \ldots, n$. Therefore $\tilde{Z}_t$, $t = 1, 2, \ldots, n$, are independent identically distributed as $(p+k)$-variate normal random vectors with mean $\mu_x$ and covariance matrix $(\beta, \varepsilon_t \otimes k) \Sigma_{xx} (\beta, \varepsilon_t \otimes k) + \Sigma_{\varepsilon\varepsilon}$, where $\varepsilon_t \otimes k$ is a $k \times k$ identity matrix. We also write $\tilde{Z}_{\mu\mu} = (\beta, \varepsilon_t \otimes k) \Sigma_{xx} (\beta, \varepsilon_t \otimes k)$.

The matrix $\tilde{S}_n$ is distributed as $(n-1)^{-1} A$, where $A$ is a Wishart matrix with $n-1$ degrees of freedom. We define $\tilde{Z}_0 = \tilde{Z}(\gamma_0)$ to be the expectation of $\tilde{S}_n$, so that
\begin{align*}
\mathcal{Z}(\gamma_0) &= E(S_n) \\
&= (\beta', I_k \times k)' \mathcal{Z}_{xx}(\beta', I_k \times k) + \mathcal{Z}_{\varepsilon \varepsilon} \\
&= \mathcal{Z}_{\mu \mu} + \mathcal{Z}_{\varepsilon \varepsilon}.
\end{align*}

(3.2)

A typical element of $\mathcal{Z}_0$ is denoted by $\sigma_{0ij}$, while a typical element of $\mathcal{Z} = \mathcal{Z}(\gamma)$ is denoted by $\sigma_{ij}$. Let the $1 \times q$ vector $s_n$, where $q = 2^{-1}(p+k)(p+k+1)$, be defined by

$$s_n = \text{vech } S_n,$$

(3.3)

and let $\mathcal{Z}(\gamma_0)$ represent the expectation of $s_n$. That is

$$\mathcal{Z}(\gamma_0) = E(s_n) = \text{vech } E(s_n) = \text{vech } \mathcal{Z}_0.$$

(3.4)

Estimation of $\gamma_0$ is based upon the minimization with respect to $\gamma$ of the residual quadratic form

$$g(\gamma|\gamma_0) = [s_n - \mathcal{Z}(\gamma)]' [2^{-1}(n-1)G'(V_0 \alpha V_0)'G] [s_n - \mathcal{Z}(\gamma)],$$

(3.5)

where $V_0$, of appropriate dimension, is a random matrix which converges in probability to $\mathcal{Z}_0^{-1}$, and $G$ is defined by (2.3). To motivate expression (3.5), we use the representation of the variance of $s_n$ given in Lemma 3.1. The following result is stated in Browne (1974).
Lemma 3.1: Assume the model defined by (1.6) with the \( z_t \), \( t = 1, 2, \ldots, n \), independently and identically distributed as \((p+k)\)-dimensional normal vectors with mean zero and covariance matrix \( \Sigma \). Let \( s_n \) be defined by (3.1) and (3.3) and let \( \tilde{H} \) be defined by (2.4). Then \( \Omega = \tilde{\Omega}(Y_0) = (n-1)\text{Var}(s_n) \) can be expressed as

\[
\tilde{\Omega} = 2H(\tilde{z}_0 \otimes \tilde{z}_0)H'.
\] (3.6)

The matrix \( \tilde{\Omega} \) is positive definite.

Proof: Consider elements \( s_{ij} \) and \( s_{kl} \) of the matrix \( s_n \). Since \( s_n \) is a multiple of a Wishart matrix, the covariance of \( s_{ij} \) and \( s_{kl} \) is given by [see Chapter 7, Anderson (1958)]

\[
\text{Cov}(s_{ij}, s_{kl}) = (n-1)^{-1}(\sigma_{0ik}\sigma_{0jl} + \sigma_{0il}\sigma_{0jk}).
\]

From (2.9) we see this can be written

\[
\text{Cov}(s_{ij}, s_{kl}) = (n-1)^{-1}[2^{-1}\left( (\tilde{z}_0 \otimes \tilde{z}_0)_{ij}, kl \\ + (\tilde{z}_0 \otimes \tilde{z}_0)_{ji}, kl \\ + 2^{-1}[ (\tilde{z}_0 \otimes \tilde{z}_0)_{ji}, kl + (\tilde{z}_0 \otimes \tilde{z}_0)_{ij}, kl \right)].
\]
or, using Kronecker deltas,

\[
\begin{aligned}
&= 2^{-(n-1)} \sum_{v=1}^{(p+k)^2} \sum_{u=1}^{(p+k)^2} \sum_{t=1}^{(p+k)^2} \sum_{s=1}^{(p+k)^2} \left( \delta_{si} \delta_{tj} \delta_{uk} \delta_{vl} \right) \\
&+ \delta_{sj} \delta_{ti} \delta_{ul} \delta_{vk} + \delta_{sj} \delta_{ti} \delta_{uk} \delta_{vl} \\
&+ \delta_{si} \delta_{tj} \delta_{ul} \delta_{vk} \left( \xi_0 \otimes \xi_0 \right)_{st,uv} \\
&= 2^{-(n-1)} \sum_{v=1}^{(p+k)^2} \sum_{u=1}^{(p+k)^2} \sum_{t=1}^{(p+k)^2} \sum_{s=1}^{(p+k)^2} \left( \delta_{uj} \delta_{vk} + \delta_{uk} \delta_{vl} \right) \left( \xi_0 \otimes \xi_0 \right)_{st,uv},
\end{aligned}
\]

which, from (2.6),

\[
\begin{aligned}
&= 2^{(n-1)} \sum_{v=1}^{(p+k)^2} \sum_{u=1}^{(p+k)^2} \sum_{t=1}^{(p+k)^2} \sum_{s=1}^{(p+k)^2} \xi_{ij, st} \\
&\times \left( \xi_0 \otimes \xi_0 \right)_{kl, uv} \\
&= 2^{(n-1)} \left[ H \left( \xi_0 \otimes \xi_0 \right) \right]_{ij, kl}. \\
\end{aligned}
\]

Therefore,

\[
(n-1) \text{Var}(\xi_n) = 2H \left( \xi_0 \otimes \xi_0 \right) H' = \Omega.
\]

Since $H$ is of full row rank, $\Omega$ is positive definite.

Thus, the matrix $\Omega$ defined by (2.6) provides a
convenient notation for expressing the covariance matrix of a vector composed of the distinct elements of \( S_n \).

Using (2.17) and (3.6), we can express \( \Omega^{-1} \) as

\[
\Omega^{-1} = 2^{-1} G'(X_n' - X_n) G. \tag{3.7}
\]

From (3.7) we see that the matrix \( 2^{-1} G'(V_n \circ V_n) G \) of the residual quadratic form \( g(\gamma | V_0) \) is a consistent estimator of \( \Omega^{-1} \), since \( V_0 \) is assumed to converge in probability to \( \Omega_0^{-1} \).

B. Generalized Least Squares Estimation of the Multivariate Linear Structural Model

Browne (1974) considers the class of estimators obtained by minimizing with respect to \( \gamma \) the quadratic form

\[
g(\gamma | V) = (S_n - s(\gamma))' 2^{-1}(n-1) G'(V \circ V) G (S_n - s(\gamma))
= 2^{-1}(n-1) [\text{vec}(S_n - \gamma(\gamma))]' (V \circ V) [\text{vec}(S_n - \gamma(\gamma))]. \tag{3.8}
\]

where \( V_n \), of appropriate dimension, is either a positive definite constant matrix \( (V = V_0) \), or a random matrix which converges in probability to a positive definite matrix \( \Omega_0 \), and \( G \) is defined by (2.3). Using (2.19) we can write (3.8) as

\[
g(\gamma | V) = 2^{-1}(n-1) \text{tr}\{[S_n - \gamma(\gamma)] V\}^2. \tag{3.9}
\]

Note that expression (3.8) differs from (3.5) in that the matrix \( V \) in \( g(\gamma | V) \) need not converge to \( V_0^{-1} \), while \( V_0 \)
was required to converge to $\hat{\gamma}_0^{-1}$. However, the estimators obtained from minimization of $g(\gamma|\gamma_0)$ are contained in the class of estimators which minimize $g(\gamma|\gamma)$.

Browne calls estimators which minimize (3.8) generalized least squares (G.L.S.) estimators. We denote G.L.S. estimators by $\hat{\gamma}$. The name generalized least squares estimator arises from the analogy to generalized least squares estimation of the parameters of the general linear model. Expression (3.5) parallels the Aitken equations of linear model theory with the elements of $s_n$ playing the role of the observed dependent variables, $\sigma(\gamma_0)$ the expected value of the vector of observed dependent variables, and 

\[ \{2^{-1}(n-1)\Sigma'(\gamma_0 \mathcal{A} \gamma_0)\Sigma\} \]

an estimate of the inverse of the covariance matrix of the observed dependent variables. In fact, were $g(\gamma_0)$ a known linear function of the elements of $\gamma_0'$, and if $\gamma_0 = \hat{\gamma}_0^{-1}$, then by the Gauss-Markov Theorem our estimation procedure would yield the best linear unbiased estimates of estimable functions of $\gamma_0$.

As with linear model theory, the concept of identifiability of the parameters we wish to estimate is important. Generally speaking, we say an unknown parameter vector is identified if knowledge of the distribution function of the observable random variables uniquely determines the value of the parameter vector. Since $s_n$ is a multiple of a
Wishart matrix, specification of \( \mathbf{Z}_0 \) is equivalent to specification of the distribution function of the vector \( \mathbf{s}_n \). Hence, we say that \( \mathbf{Y}_0 \) is identified if \( \mathbf{Z}(\mathbf{Y}_1) = \mathbf{Z}(\mathbf{Y}_0) \) implies \( \mathbf{Y}_1 = \mathbf{Y}_0 \).

The next theorem is due to Browne (1974).

**Theorem 3.1:** Consider the model of Lemma 3.1. Assume that \( \mathbf{Y}_0 \) is identified. Then the G.L.S. estimators are consistent.

**Proof:** We use representation (3.9) of \( g(\mathbf{y}|\mathbf{V}) \). Since \( \mathbf{Y}_0 \) is identified and \( \mathbf{V} \) is positive definite, \( \text{tr}([\mathbf{Z}_0 - \mathbf{Z}(\mathbf{Y})]\mathbf{V})^2 \) has its absolute minimum of zero at \( \mathbf{Y} = \mathbf{Y}_0 \). Now \( \mathbf{Z}_n \) and \( \mathbf{V} \) converge stochastically (in probability) to \( \mathbf{Z}_0 \) and \( \mathbf{V} \), and \( \mathbf{Z}(\mathbf{Y}) \) is bounded in a neighborhood of \( \mathbf{Y} = \mathbf{Y}_0 \). Consequently \( \text{tr}([\mathbf{Z}_n - \mathbf{Z}(\mathbf{Y})]\mathbf{V})^2 \) converges in probability to \( \text{tr}([\mathbf{Z}_0 - \mathbf{Z}(\mathbf{Y})]\mathbf{V})^2 \) uniformly in a neighborhood of \( \mathbf{Y} = \mathbf{Y}_0 \). Since \( \text{tr}([\mathbf{Z}_n - \mathbf{Z}(\mathbf{Y})]\mathbf{V})^2 \) is continuous in \( \mathbf{Y} \), the point \( \hat{\mathbf{Y}} \) where it has its absolute minimum converges stochastically to \( \mathbf{Y}_0 \).

The above proof by Browne is an adaptation of a proof by Anderson and Rubin (1956). We state the obvious corollary.

**Corollary 3.1.1:** Let the assumptions of Theorem 3.1 hold. Then the estimators \( \hat{\mathbf{Y}} \) are consistent.

The following theorem gives the limiting distribution of \( \frac{1}{n^2} [\mathbf{Z}_n - \mathbf{Z}(\mathbf{Y}_0)] \).
Theorem 3.2: Assume the model of Lemma 3.1. Then
\[ \frac{1}{n} [s_n - \gamma_0] \] converges in distribution to a normal vector random variable with mean zero and covariance matrix \( \Omega \).

Proof: It suffices to show that for each \( q \)-dimensional nonnull vector of constants \( a = (a_1, a_2, \ldots, a_q)' \) that
\[ \frac{1}{n^2} a'[s_n - \gamma_0] \to N(0, a'\Omega a). \] (3.10)

Define the \((p+k)\times(p+k)\) symmetric matrix \( W \) to be the matrix such that \( \text{vech}(W) = (G'G)^{-1}a \), where \( G \) is defined by \( \text{vec} S_n = Gs_n \). Then,
\[ a'[s_n - \gamma_0] = a'(G'G)^{-1}(G'G)[s_n - \gamma_0] \]
\[ = \text{vech}(W)'[s_n - \gamma_0] \]
\[ = \text{vec}(W)'\{\text{vec}[s_n - E(S_n)]\} \]
Thus,
\[ \frac{1}{n^2} a'[s_n - \gamma_0] = \frac{1}{n^2} \text{tr}(W[s_n - E(S_n)]). \] (3.11)

The matrix \( A = \sum_{t=1}^{n} (Z_t - \bar{Z})'(Z_t - \bar{Z}) \) is distributed as
\[ A = \sum_{t=1}^{n-1} V_t'V_t, \] where the \( V_t, t = 1, 2, \ldots, n-1 \) are independent, each with the distribution \( N(0, \bar{Z}_0) \) [Anderson (1958, pp. 51-54)]. Hence, we can rewrite (3.11) as
\[ \frac{1}{n^2} a'[s_n - \gamma_0] = \frac{1}{n^2} (n-1)^{-1} \text{tr}(W[ \sum_{t=1}^{n-1} V_t'V_t - \bar{Z}_0 ]). \]
Let $\sum_{i=1}^{p+k} \lambda_i \xi_i \xi_i' = \frac{1}{2} \mathbf{Z}_0 \mathbf{W} \mathbf{Z}_0'$ be the spectral decomposition of $\frac{1}{2} \mathbf{Z}_0 \mathbf{W} \mathbf{Z}_0'$, where $\mathbf{Z}_0 \mathbf{Z}_0' = \mathbf{Z}_0'$, and $\xi_i$ is a p+k column vector. Since $\mathbf{W} = \frac{1}{2} (\sum_{i=1}^{p+k} \lambda_i \xi_i \xi_i') \mathbf{Z}_0'$,

$$n \text{ tr} \{\mathbf{W}[\mathbf{S}_n - \mathbf{E}(\mathbf{S}_n)]\}$$

$$= \frac{1}{2} n^2 (n-1)^{-1} \text{ tr} \{ \sum_{i=1}^{n-1} \lambda_i \frac{1}{2} \mathbf{Z}_0 (\sum_{i=1}^{p+k} \xi_i \xi_i') \mathbf{Z}_0' \xi_i' \xi_i \mathbf{Z}_0 \mathbf{Z}_0' \}$$

$$= \frac{1}{2} n^2 (n-1)^{-1} \mathbf{W} \{ \sum_{i=1}^{p+k} \lambda_i (n-1) \frac{1}{2} \text{ tr} \{ \mathbf{Z}_0' \mathbf{V}_t \mathbf{V}_t' \mathbf{Z}_0' \xi_i' \xi_i \} \}$$

$$= \frac{1}{2} n^2 (n-1)^{-1} \mathbf{W} \{ \sum_{i=1}^{p+k} \lambda_i (n-1) \mathbf{Z}_0' \mathbf{V}_t \mathbf{V}_t' \mathbf{Z}_0' \xi_i' \xi_i \}$$

Because $\xi_i' \xi_j = 0$ for all $i \neq j$, the p+k outer summands are independent. Now $\sum_{t=1}^{n-1} (\xi_i' \mathbf{V}_t - \frac{1}{2} \mathbf{V}_t) (\xi_i' \mathbf{V}_t - \frac{1}{2} \mathbf{V}_t)^2$ is distributed as $\xi_i' \xi_i x_{n-1}^2$, where $x_{n-1}^2$ is a chi-square random variable with n-1 degrees of freedom. Thus, $(n-1) \frac{1}{2} \sum_{t=1}^{n-1} (\xi_i' \mathbf{V}_t - \frac{1}{2} \mathbf{V}_t) (\xi_i' \mathbf{V}_t - \frac{1}{2} \mathbf{V}_t)^2 - \xi_i' \xi_i$ is asymptotically normal, and $n \mathbf{a}'(\mathbf{S}_n - \mathbf{g})$ is the sum of p+k independent random variables each of which is asymptotically normal. Therefore, $n \frac{1}{2} \mathbf{a}'(\mathbf{S}_n - \mathbf{g})$ is asymptotically normal for
each nonnull vector \( \mathbf{a} \). By Lemma 3.1, \( \text{Var}\left[\frac{1}{\sqrt{n}}\mathbf{a}'(\mathbf{s}_n - \tilde{\mathbf{g}}(\tilde{\gamma}_0))\right] \) converges to \( \mathbf{a}'\tilde{\mathbf{\Omega}}\mathbf{a} \), so that \( \frac{1}{\sqrt{n}}[\mathbf{s}_n - \tilde{\mathbf{g}}(\tilde{\gamma}_0)] \) converges in distribution to a \( N(0, \tilde{\mathbf{\Omega}}) \) random vector.

Theorem 3.3 below is given in Browne (1974). In proving this theorem, only the asymptotic distribution as \( n \) approaches infinity of \( \mathbf{s}_n \) is needed. The limiting distribution is multivariate normal with mean \( \tilde{\mathbf{g}}(\tilde{\gamma}_0) \) and covariance matrix \( (n-1)^{-1}\tilde{\mathbf{\Omega}} \). In Chapter IV it is demonstrated that the results of Theorem 3.3 hold for the functional relationship when \( \tilde{\mathbf{\Omega}} \) is not of the form (3.6). And in Chapter V, the fact that only asymptotic normality of \( \mathbf{s}_n \) is required allows us to extend our results to nonnormal multivariate errors in variables models.

**Theorem 3.3**: Consider the model of Lemma 3.1. Assume that \( \tilde{\gamma}_0 \) is identified. Define the \((p+k)^2\times r\) matrix \( \tilde{\Delta} \) to be

\[
\tilde{\Delta} = \left[ \frac{\partial \text{vec} \tilde{\mathbf{Z}}(\mathbf{\gamma})}{\partial \tilde{\gamma}} \right] |_{\mathbf{\gamma} = \tilde{\gamma}_0},
\]

and suppose that \( \tilde{\Delta} \) has full column rank. Then the limiting distribution of \( \frac{1}{\sqrt{n}}(\tilde{\mathbf{\gamma}} - \tilde{\gamma}_0) \) is multivariate normal with zero mean and covariance matrix
\[ 2[\Lambda(\bar{\gamma})]^{-1} \Lambda(\bar{V}Z_0 \bar{V}) [\Lambda(\bar{\gamma})]^{-1}, \]

where

\[ \Lambda(\bar{\gamma}) = \Lambda'(\bar{\gamma} \circ \bar{\gamma}) \Lambda. \]

**Proof:** Let

\[ h(y|V) = -\frac{\partial g(y|V)}{\partial y} \]

\[ = 2^{-1} \frac{\partial \text{vec}^T \bar{Z}(y) \circ \bar{V}}{\partial y} \text{vec}(S_n - \bar{Z}(y)). \]

Using (2.19), a typical element of this vector may be expressed as

\[ h_i(y|V) = 2^{-1} \text{tr}\{[V(S_n - \bar{Z}(y))V]\frac{\partial \bar{Z}(y)}{\partial y_i}\}. \]

By Taylor's theorem

\[ h(\bar{y}|V) = h(y_0|V) - \Theta(\bar{y} - y_0), \]

where

\[ [\Theta]_{ij} = \left. \frac{\partial h_i}{\partial y_j} \right|_{y=y_0} - \left( \frac{1}{2} \right) \sum_{l=1}^{r} (\bar{y}_l - y_0, l) \frac{\partial^2 h_i}{\partial y_j \partial \gamma_l} \bigg|_{\gamma = \gamma^*} \]

and \( \gamma^* \) lies between \( \bar{y} \) and \( \gamma_0 \).

Now,
\[
\frac{\partial h_i}{\partial \gamma_j} = 2^{-1} \text{tr}\left\{ \tilde{V}[S_n - \tilde{Z}(\gamma)] \tilde{V} \left[ \frac{\partial^2 \tilde{Z}(\gamma)}{\partial \gamma_i \partial \gamma_j} \right] \right\} \\
- 2^{-1} \text{tr}\left\{ \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_i} \right] \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_j} \right] \right\}, \tag{3.16}
\]

and

\[
\frac{\partial^2 h_i}{\partial \gamma_j \partial \gamma_k} = 2^{-1} \text{tr}\left\{ \tilde{V}[S_n - \tilde{Z}(\gamma)] \tilde{V} \left[ \frac{\partial^3 \tilde{Z}(\gamma)}{\partial \gamma_i \partial \gamma_j \partial \gamma_k} \right] - \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_j} \right] \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_i} \right] \right\} \\
- \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_j} \right] \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_k} \right] - \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_i} \right] \tilde{V} \left[ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_j} \right]. \tag{3.17}
\]

Since the elements of \([S_n - \tilde{Z}(\gamma_0)]\) and \((\tilde{Y} - \gamma_0)\) converge to zero in probability, since the trace functions in (3.16) and (3.17) are continuous, and since the partial derivatives are asymptotically bounded in a neighborhood of \(\gamma_0\) it follows that \([\partial]_{ij}\) converges stochastically to

\[
\text{tr}\left\{ \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_i} \tilde{V} \left( \frac{\partial \tilde{Z}(\gamma)}{\partial \gamma_j} \right) \right\}_{\gamma = \gamma_0},
\]

or

\[
\text{plim} \ \theta = 2^{-1} \hat{\Delta}' (\tilde{\Sigma} \otimes \tilde{\Sigma}) \hat{\Delta} \\
= 2^{-1} \hat{\Delta}(\tilde{\Sigma}).
\]

This matrix is nonsingular since we have assumed \(\hat{\Delta}\) has full column rank and \(\tilde{\Sigma}\) is positive definite. It follows from (3.13) and the fact \(\hat{h}(\tilde{\Sigma} | \tilde{\Sigma}) = 0\) that
\[ \tilde{y} = \gamma_0 + 2e^{-1}h(\gamma_0 | \tilde{v}) \]

and \( \tilde{y} \) is asymptotically equivalent to

\[ \tilde{y} = \gamma_0 + 2[\tilde{\Delta}'(\tilde{v} \circ \tilde{v})\tilde{\Delta}]^{-1}\tilde{h}(\gamma_0 | \tilde{v}) \]

\[ = \gamma_0 + [\tilde{\Delta}'(\tilde{v} \circ \tilde{v})\tilde{\Delta}]^{-1}\tilde{\Delta}'(\tilde{v} \circ \tilde{v})[\text{vec}(S_n - \tilde{\zeta}_0)], \quad (3.18) \]

because

\[ \frac{1}{n^2}(\tilde{y} - \gamma) = \left\{ e^{-1}[\tilde{\Delta}'(\tilde{v} \circ \tilde{v})\tilde{\Delta}]^{-1} \right\} \tilde{\Delta}'(\tilde{v} \circ \tilde{v}) \]

\[ + e^{-1} \tilde{\Delta}'\left[ (\tilde{v} \circ \tilde{v}) - (\tilde{v} \circ \tilde{v}) \right] \frac{1}{n^2} \text{vec}(S_n - \tilde{\zeta}_0) \} \quad (3.19) \]

converges in probability to the null vector as \( n \to \infty \). Since \( \tilde{y} \) is a linear function of \( \text{vec} S_n \), the limiting distribution of \( n^2(\tilde{y} - \gamma_0) \) and of \( n^2(\tilde{y} - \gamma_0) \) is multivariate normal with mean vector \( 0 \) and dispersion matrix

\[ [\tilde{\Delta}(\tilde{v})]^{-1}\tilde{\Delta}'(\tilde{v} \circ \tilde{v})G \text{ Var}(S_n)G'(\tilde{v} \circ \tilde{v})\tilde{\Delta}[\tilde{\Delta}(\tilde{v})]^{-1} \]

\[ = 2[\tilde{\Delta}(\tilde{v})]^{-1}\tilde{\Delta}'(\tilde{v} \circ \tilde{v})G H(\tilde{\zeta}_0 \circ \tilde{\zeta}_0)H' G'(\tilde{v} \circ \tilde{v})\tilde{\Delta}[\tilde{\Delta}(\tilde{v})]^{-1}. \]

This dispersion matrix may be expressed in the form (3.12) after use of (3.13), (2.10), (2.16), and the fact that each column of \( \tilde{\Delta} \) is formed from a symmetric matrix,

\[ \left. \frac{\partial \tilde{\zeta}(\gamma)}{\partial \gamma_j} \right| _{\gamma = \gamma_0} \]
This proof is an adaptation of a proof by Malinvaud (1970).

From our previous discussion of the analogy between G.L.S. estimation of covariance structures and G.L.S. estimation of linear models, we expect \( \hat{\gamma} \) to be "best" in some sense. Because the matrix \( \Lambda = \sum_{t=1}^{n} (\bar{Z}_t - \bar{Z} )' (\bar{Z}_t - \bar{Z} ) \) is a Wishart matrix, we can show in the following corollary that the "best" G.L.S. (B.G.L.S.) estimators, \( \hat{\gamma} \), are asymptotically efficient among the class of all consistent estimators which are asymptotically normal.

**Corollary 3.3.1:** Let the assumptions of Theorem 3.3 hold. The asymptotic dispersion matrix of \( \frac{1}{n} \sum_{i} (\hat{\gamma}_i - \gamma_0) \) is \( 2[\Lambda(\Sigma_0^{-1})]^{-1} \). Among the class of consistent asymptotically normal estimators of \( \gamma_0 \), the B.G.L.S. estimators are asymptotically efficient.

**Proof:** Setting \( \bar{V} = \bar{Z}_0^{-1} \) in formula (3.12) shows the asymptotic dispersion matrix of \( \frac{1}{n} \sum_{i} (\hat{\gamma}_i - \gamma_0) \) to be \( 2[\Lambda(\Sigma_0^{-1})]^{-1} \). To prove asymptotic efficiency of the B.G.L.S. estimators we must show that the difference between \( 2[\Lambda(\Sigma_0^{-1})]^{-1} \) and the inverse information matrix of the normalized B.G.L.S. estimators based on the exact distribution of \( S_n \) is \( o(1) \). The log of the likelihood function of \( S_n = (n-1)^{-1} \Lambda \) is
\[ F(\gamma) = \log L(\tilde{z}(\gamma); S_n) = K - 2^{-1}(n-1) \log |\tilde{z}(\gamma)| \]
\[-2^{-1}(n-1) \text{tr}[S_n \tilde{z}(\gamma)], \quad (3.20) \]

where \( K \) does not depend on \( \gamma \). Thus,

\[
\frac{\partial F(\gamma)}{\partial \gamma_i} = 2^{-1}(n-1) \text{tr}[\tilde{z}^{-1}(\gamma) [S_n - \tilde{z}(\gamma)] \tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_i}] ,
\]
\[i = 1, \ldots, r, \quad (3.21)\]

and

\[
\frac{\partial F(\gamma)}{\partial \gamma_i \partial \gamma_j} = -2^{-1}(n-1) \text{tr}[\tilde{z}^{-1}(\gamma) [2 S_n - \tilde{z}(\gamma)] \tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_i}]
\times \tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_j} \\
+ \tilde{z}^{-1}(\gamma) [S_n - \tilde{z}(\gamma)] \tilde{z}^{-1}(\gamma) \frac{\partial^2 \tilde{z}(\gamma)}{\partial \gamma_i \partial \gamma_j} ,
\]
\[i = 1, \ldots, r; \quad j = 1, \ldots, r. \quad (3.22)\]

A typical element of the information matrix is

\[
-\mathbb{E}\left\{ \frac{\partial F(\gamma)}{\partial \gamma_i \partial \gamma_j} \right\} = 2^{-1}(n-1) \text{tr}\left[\tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_i} \tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_j} \bigg| \gamma = \gamma_0 \right\}
\]
\[= 2^{-1}(n-1) \text{tr}\left[\frac{\partial \tilde{z}(\gamma)}{\partial \gamma_i} \tilde{z}^{-1}(\gamma) \frac{\partial \tilde{z}(\gamma)}{\partial \gamma_j} \tilde{z}^{-1}(\gamma) \bigg| \gamma = \gamma_0 \right\}.
\]
Letting $I_n(y)$ be the information matrix of $y$ and using (2.19), we may write

$$I_n(y) = 2^{-(n-1)}[A' \otimes \Sigma^{-1} \Sigma^{-1}]$$

Therefore the difference between $2[A(XQ)]^{-1}$ and the inverse information matrix of the normalized B.G.L.S. estimators is

$$2[1-n(n-1)^{-1}] [A(XQ)]^{-1} = -2(n-1)^{-1}[A(XQ)]^{-1}.$$ 

The result follows.

In practice, we can obtain a B.G.L.S. estimator by setting $V_0 = S_n^{-1}$. Because $\mathbb{E}(S_n) = \Sigma_0$ and $\text{Var}(S_n) = O(n^{-1})$, $S_n^{-1}$ is a consistent estimator of $\Sigma_0^{-1}$.

Theorems 3.4 and 3.5 are due to Browne. Theorem 3.4 shows that, in addition to yielding a B.G.L.S. estimator of $\gamma_0$, use of $V_0$ enables one to test the null hypothesis that $\Sigma = \Sigma_0$ against the alternative that $\Sigma \neq \Sigma_0$.

Theorem 3.4: Consider the model of Lemma 3.1. Assume $\Sigma = \Sigma(\gamma_0)$ and that $\gamma_0$ is identified. Then the limiting distribution of $g(\hat{\gamma}|V_0)$ is chi-square with $q-r$ degrees of freedom, where $q$ and $r$ are the dimensions of $S_n$ and $\gamma_0$, respectively.
Proof: By (3.19), \( \frac{1}{n^2} \hat{\gamma} - \bar{\gamma} \) converges in probability to a null vector. Also, \( \frac{1}{n^2} \{ \text{vec}[\bar{\gamma} - \bar{\gamma}(\gamma_0) - \Delta(\hat{\gamma} - \gamma_0)] \} \) converges in probability to a null vector since, by Taylor's theorem,

\[
\frac{1}{n^2} \{ \sigma_{ij}(\hat{\gamma}) - \sigma_{0ij} - \left[ \frac{\partial \sigma_{ij}(\gamma)}{\partial \gamma} \right]_{\gamma=\gamma_0} (\hat{\gamma} - \gamma_0) \} = (2^{-1}n) \frac{1}{2} (\hat{\gamma} - \gamma_0) \left[ \frac{\partial^2 \sigma_{ij}(\gamma)}{\partial \gamma \partial \gamma} \right]_{\gamma=\gamma*} (\hat{\gamma} - \gamma_0),
\]

where \( \gamma^* \) lies between \( \hat{\gamma} \) and \( \gamma_0 \).

Consequently \( \frac{1}{n^2} \{ \text{vec}[S_n - \bar{Z}(\hat{\gamma})] \} \) converges stochastically to

\[
\frac{1}{n^2} \{ \text{vec}[S_n - \bar{Z}(\gamma_0)] - \Delta(\hat{\gamma} - \gamma_0) \} = \frac{1}{n^2} \{ I - \Delta [\bar{Z}_0^{-1} \otimes \bar{Z}_0^{-1}]^{-1} \Delta' (\bar{Z}_0^{-1} \otimes \bar{Z}_0^{-1}) \} \text{vec}(S_n - \bar{Z}(\gamma_0))
\]

and

\[
g(\gamma|\gamma_0) = 2^{-1} (n-1) \{ \text{vec}[S_n - \bar{Z}(\gamma)] \}' (\gamma_0 \otimes \gamma_0) \text{vec}[S_n - \bar{Z}(\gamma)]
\]

converges stochastically to

\[
g_0 = 2^{-1} (n-1) (S_n - \sigma_0)' \Sigma_0 (S_n - \sigma_0),
\]

where
\[ G_0 = G' \left( (X_0^{-1} \otimes X_0^{-1}) - (X_0^{-1} \otimes X_0^{-1}) \Delta (X_0^{-1} \otimes X_0^{-1}) \Delta^{-1} \right) \times \Delta' (X_0^{-1} \otimes X_0^{-1}) G. \]

Since \( G_0 [H(X_0 \otimes X_0) H'] \) is idempotent of rank \( q-r \), the limiting distribution of \( g_0 \) and of \( g(\hat{y} \mid Y_0) \) is the central chi-square distribution with \( q-r \) degrees of freedom [Graybill (1976, pp. 135-136)].

The above result is analogous to the result of linear model theory that the mean square residual is distributed as a central chi-square random variable with degrees of freedom equal to the number of observations minus the number of parameters estimated.

Theorem 3.5 points out the asymptotic relationship of maximum likelihood (M.L.) estimation of \( \gamma_0 \) to G.L.S. estimation of \( \gamma_0 \).

Theorem 3.5: Consider the model of Lemma 3.1, and assume that \( \gamma_0 \) is identified. Suppose that \( \hat{\gamma}_1 \) is a M.L. estimate of \( \gamma_0 \) and that \( \hat{\gamma}_2 \) is a G.L.S. estimate where \( \gamma = [X(\hat{\gamma}_1)]^{-1} \). Then \( \hat{\gamma}_2 \) is a B.G.L.S. estimate and \( \Pr(\hat{\gamma}_1 \neq \hat{\gamma}_2) \to 0 \) as \( n \to \infty \).

Proof: From (3.20) we see that maximizing the Wishart likelihood function is equivalent to minimizing
\[ G(\gamma) = \log|Z(\gamma)| + \text{tr}[S_n Z^{-1}(\gamma)]. \] (3.23)

Consequently, the equations

\[
\frac{\partial G(\gamma)}{\partial \gamma_i} = -\text{tr}(Z^{-1}(\gamma) [S_n - \hat{Z}(\gamma)] Z^{-1}(\gamma) \left[ \frac{\partial Z(\gamma)}{\partial \gamma_i} \right]) = 0,
\]

\[ i = 1, \ldots, r \] (3.24)

and the condition that the matrix with typical element

\[
\frac{\partial^2 G(\gamma)}{\partial \gamma_i \partial \gamma_j} = \text{tr}(Z^{-1}(\gamma) [2S_n - \hat{Z}(\gamma)] Z^{-1}(\gamma) \left[ \frac{\partial Z(\gamma)}{\partial \gamma_i} \right] Z^{-1}(\gamma) \left[ \frac{\partial Z(\gamma)}{\partial \gamma_j} \right])
\]

\[ i = 1, \ldots, r, \quad j = 1, \ldots, r \] (3.25)

be positive definite will be satisfied at the point

\[ \gamma = \gamma_1. \] The equations

\[
\frac{\partial g(\gamma|\alpha)}{\partial \gamma_i} = -\text{tr}(\alpha[Y [S_n - \hat{Z}(\gamma)] Y] \left[ \frac{\partial Z(\gamma)}{\partial \gamma_i} \right]) = 0, \quad i = 1, \ldots, r
\] (3.26)

and the condition that the matrix with typical element

\[
\frac{\partial^2 g(\gamma|\alpha)}{\partial \gamma_i \partial \gamma_j} = \text{tr}(Y \left[ \frac{\partial Z(\gamma)}{\partial \gamma_i} \right] Y \left[ \frac{\partial Z(\gamma)}{\partial \gamma_j} \right]) - \text{tr}(Y [S_n - \hat{Z}(\gamma)] Y) \]

\[ x \left[ \frac{\partial^2 Z(\gamma)}{\partial \gamma_i \partial \gamma_j} \right], \quad i = 1, \ldots, r, \quad j = 1, \ldots, r \] (3.27)
be positive definite will be satisfied at the point $\gamma = \hat{\gamma}_2$ when $V = X^{\top}(\hat{\gamma}_1)$.

Using reasoning similar to that used in the proof of Theorem 3.1 it can be shown that the M.L. estimator, $\hat{\gamma}_1$, is a consistent estimator of $\gamma_0$. Consequently $X^{-1}(\hat{\gamma}_1)$ is a consistent estimator of $X^{-1}_0$ and $\hat{\gamma}_2$ is a B.G.L.S. estimator.

Equations (3.24) and (3.26) are equivalent when $V = X^{-1}(\gamma)$. Consequently $\gamma = \hat{\gamma}_1$ is always a stationary point of $g(\gamma | X^{-1}(\hat{\gamma}_1))$ and will not be a minimum only if the matrix with typical element given by (3.27) is not positive definite. Since the matrix with typical element (3.25) is positive definite at $\gamma = \hat{\gamma}_1$ and since the difference

$$
\frac{\partial^2 g(\gamma | X^{-1}(\hat{\gamma}_1))}{\partial \gamma_i \partial \gamma_j} - \frac{\partial^2 g(\gamma)}{\partial \gamma_i \partial \gamma_j} \bigg|_{\gamma = \hat{\gamma}_1} = -2\text{tr}\{X^{-1}(\gamma) [S_n - X(\gamma)] X^{-1}(\gamma) \frac{\partial^2 z(\gamma)}{\partial \gamma_i^2} X^{-1}(\gamma) \frac{\partial z(\gamma)}{\partial \gamma_j} \}
$$

converges stochastically to zero, the probability that the matrix with typical element (3.27) is not positive definite at the point $\gamma = \hat{\gamma}_1$ tends to zero as $n \to \infty$. This implies that
the probability that the point \( \hat{y}_1 \) at which \( G(y) \) has an absolute minimum does not give at least a relative minimum of \( g(y|\mathcal{Q}_0^{-1}(\hat{y}_1)) \) tends to zero as \( n \to \infty \). Since \( g(y|\mathcal{Q}_0^{-1}) \) is convex in a neighborhood of \( \gamma_0 \) and since \( \hat{y}_1 \) and \( \hat{y}_2 \) both converge stochastically to \( \gamma_0 \), the probability that there is a minimum at \( \hat{y}_1 \) which does not coincide with the absolute minimum at \( \hat{y}_2 \) tends to zero as \( n \to \infty \).

This result implies that M.L. estimators will have the same asymptotic properties as B.G.L.S. estimators. For small samples, however, we are unable to determine general conditions under which \( \hat{y}_1 = \hat{y}_2 \). If \( \hat{y}_1 = \hat{y}_2 \) we see from Corollary 3.3.1 that the estimated variances of \( \hat{y}_1 \) and \( \hat{y}_2 \) are identical, given by

\[
\text{Var}(\hat{y}_1) = \text{Var}(\hat{y}_2) = 2(n-1)^{-1} \left[ \frac{\partial \text{vec } \mathcal{Z}(\gamma)}{\partial \gamma} \right] \mathcal{Z}^{-1}(\gamma) \mathcal{Z}^{-1}(\gamma) \]

\[
+ \left[ \frac{\partial \text{vec } \mathcal{Z}(\gamma)}{\partial \gamma} \right] \mathcal{Z}^{-1}(\gamma) \mathcal{Z}^{-1}(\gamma) \mid \gamma = \hat{y}_1 = \hat{y}_2. \tag{3.29}
\]

For the special case where \( \mathcal{Z}(\gamma) \) is linear in \( \gamma \), Browne (1974) obtained a closed form solution for a G.L.S. estimator \( \hat{y} \). Applying (2.19) to (3.26) we have

\[
0 = \text{tr}\{\mathcal{V}[\mathcal{S}_n - \mathcal{Z}(\gamma)]\mathcal{V} [\frac{\partial \mathcal{Z}(\gamma)}{\partial \gamma}]\} = \text{tr}\{[\mathcal{S}_n - \mathcal{Z}(\gamma)]\mathcal{V} [\frac{\partial \mathcal{Z}(\gamma)}{\partial \gamma}]\mathcal{V}\}
\]

\[
= [\text{vec } \mathcal{S}_n - \text{vec } \mathcal{Z}(\gamma)]'(\mathcal{V} \otimes \mathcal{V})[\frac{\partial \text{vec } \mathcal{Z}(\gamma)}{\partial \gamma}]. \tag{3.30}
\]
For $\mathcal{Z}(\gamma)$ linear in $\gamma$ we can write $\text{vec} \mathcal{Z}(\gamma) = \Delta \gamma$, so that $\frac{\partial \text{vec} \mathcal{Z}(\gamma)}{\partial \gamma} = \Delta$. This allows us to express (3.30) as

$$\gamma' \Delta' (\mathcal{V} \otimes \mathcal{V}) \Delta = (\text{vec} \mathcal{S}_n)'(\mathcal{V} \otimes \mathcal{V}) \Delta$$

$$= [\text{vec}(\mathcal{V} \mathcal{S}_n \mathcal{V})]' \Delta,$$

using (2.18) and (2.12).

Thus,

$$\hat{\gamma} = [\Delta' (\mathcal{V} \otimes \mathcal{V}) \Delta]^{-1} \Delta' \text{vec}(\mathcal{V} \mathcal{S}_n \mathcal{V})$$

$$= [\Lambda(\mathcal{V})]^{-1} \Delta' \text{vec}(\mathcal{V} \mathcal{S}_n \mathcal{V}). \quad (3.31)$$

The following iterative procedure is suggested by these results.

1. Use (3.31) with $\mathcal{V} = S_n^{-1}$ to obtain $\hat{\gamma}_1$.
2. Use (3.31) with $\mathcal{V} = \gamma^{-1}(\mathcal{V}_1)$ to obtain $\hat{\gamma}_2$.
3. Repeat this procedure until the difference between $\hat{\gamma}_i$ and $\hat{\gamma}_{i+1}$ is suitably small.

This procedure, as Browne (1974) observed, is equivalent to the Fisher scoring method for obtaining M.L. estimates. However, Kendall and Stuart (1967) pointed out that convergence to the M.L. estimate is not guaranteed when using the Fisher scoring method, so that the successive G.L.S. estimators $\hat{\gamma}_1, \hat{\gamma}_2, \ldots$, need not converge to $\hat{\gamma}_{ML}$. 
We conclude this chapter by observing that the results derived thus far have depended upon the fact that the covariance matrix of \( \tilde{s}_n \) for the structural model has the special form \( \Omega = H'(\tilde{Z}_0 \alpha \tilde{Z}_0)'H \). We consider the functional model in Chapter IV. The covariance matrix of \( \tilde{s}_n \) in the functional case includes a component due to the fixed \( \tilde{x}_t \), \( t = 1, 2, \ldots, n \). We shall extend the principal results of this chapter to this covariance structure.
IV. ESTIMATION OF THE MULTIVARIATE LINEAR FUNCTIONAL MODEL

A. Preliminary Notation

Many of the results of Chapter III can be extended to the multivariate linear functional model. The notation used in this chapter will, when appropriate, be the same as that used for the structural model. We shall use the form of the functional model,

\[ z_t = \mu_t + \varepsilon_t, \quad t = 1,2,\ldots,n, \]

where now \( \mu_t = (x_t^\beta, x_t) \) is a \((p+k)\)-dimensional vector of unknown constants. The remaining definitions and assumptions of model (1.6) are given in Section B, Chapter I.

The unknown \( r \times 1 \) parameter vector \( \gamma_0 \) contains the unknown elements of \( \beta \) and \( z_{t' \varepsilon} \). If we regard \( \gamma_0 \) as a vector of mathematical variables we delete the subscript.

The vectors \( z_{t'}, t = 1,2,\ldots, \) are independent \((p+k)\)-dimensional normal random vectors with means \( \mu_t, t = 1,2,\ldots, \) and common covariance matrix \( z_{t' \varepsilon} \). We define the sample moment matrix \( S_n \) to be

\[ S_n = n^{-1} \sum_{t=1}^{n} z_t' z_t. \]

We define \( S_n \) by (4.1) rather than (3.1) because of notational convenience. The results of this chapter can be
derived using definition (3.1) with the only possible
difference being \( n \) replaced by \( n-1 \). Asymptotic results
are unaffected by the choice of definition. Because the \( x_t', \)
t = 1, 2, ..., \( n \) are vectors of fixed unknowns, the matrix
\( S_n \) is no longer a multiple of a Wishart matrix. Rather,
the distribution of \( nS_n \) is that of a noncentral Wishart
matrix.

We define \( Y_n(y_0) \) to be the expectation of \( S_n \). Thus,

\[
Y_n(y_0) = E(S_n)
\]

\[
= n^{-1} E \left( \sum_{t=1}^{n} x_t' x_t \right)
\]

\[
= n^{-1} E \left[ \sum_{t=1}^{n} (u_t + \varepsilon_t)' (u_t + \varepsilon_t) \right]
\]

\[
= n^{-1} \sum_{t=1}^{n} u_t' u_t + y \varepsilon \varepsilon . \quad (4.2)
\]

Now, \( u_t = x_t (\beta, \frac{I}{k} x k) \), so that

\[
n^{-1} \sum_{t=1}^{n} u_t' u_t = n^{-1} \sum_{t=1}^{n} (\beta, \frac{I}{k} x k)' x_t x_t (\beta, \frac{I}{k} x k)
\]

\[
= (\beta, \frac{I}{k} x k)' \sum_{t=1}^{n} x_t x_t (\beta, \frac{I}{k} x k)
\]

\[
= (\beta, \frac{I}{k} x k)' \sum_{n} (\beta, \frac{I}{k} x k), \quad (4.3)
\]
where $V_n = n^{-1} \sum_{t=1}^{n} x_t' x_t$. If we let

$$M_n = (\beta, I_{k \times k})' \mathcal{V}_n (\beta, I_{k \times k}),$$

we may then express $\mathcal{Z}_n (\gamma_0)$ as

$$\mathcal{Z}_n (\gamma_0) = M_n + \mathcal{Z}_n \varepsilon \varepsilon.$$

We shall denote typical elements of $\mathcal{Z}_n (\gamma_0)$ and $\mathcal{Z}_n (\gamma)$ by $\sigma_{ij,n}(\gamma_0)$ and $\sigma_{ij,n}(\gamma)$, respectively.

The $q \times 1$ vector, $s_n$, where $q = 2^{-1} (p+k)(p+k+1)$, is defined to be

$$s_n = \text{vech } S_n.$$

Let $s_n = s_n (\gamma_0)$ be defined by

$$s_n (\gamma_0) = E(s_n)$$

$$= \text{vech } E(s_n)$$

$$= \text{vech } \mathcal{Z}_n (\gamma_0).$$

B. Generalized Least Squares Estimation of the Multivariate Linear Functional Model

Estimation of $\gamma_0$ is based upon the minimization with respect to $\gamma$ of the residual quadratic form
The matrix \(D(\Omega)\) is a \((q \times q)\)-dimensional random matrix which converges in probability to \(\Omega^{-1}\), where
\[
\Omega = \lim_{n \to \infty} \frac{1}{n} \text{Var}[n^{s-a}(s_n^s - g_n^s)] \quad (4.9)
\]
We shall examine asymptotic properties of the estimator \(\hat{\gamma}\) which minimizes expression (4.8). To ensure that \(\hat{\gamma}\) and, hence, \(f(\gamma|D(\Omega))\) are well-defined, we impose restrictions upon the limiting behavior of \(s_n^s\), or, equivalently, on the limiting behavior of \(V_n\). Let \(v_{ij,n}\) be the \((i,j)\)th element of \(V_n\). We say that \(V_n\) converges (elementwise) to a matrix \(V = (v_{ij})\), or that \(V_n\) is the limit of \(V_n\), if \(v_{ij,n}\) converges to \(v_{ij}\) for all \(i = 1,2,...,p+k\) and \(j = 1,2,...,p+k\). If \(V_n\) converges to \(V\), then \(M_n = n^{-1} \sum_{t=1}^{n} \mu_t \mu_t^\top = n^{-1}(\beta, I_k \times k) V_n(\beta, I_k \times k)^\top\) converges to \(M = (\beta, I_k \times k)^\top V(\beta, I_k \times k)\). The following lemma considers the limiting behavior of \(g_n^s\) and \(\Omega_n = n \text{Var}(s_n^s)\) under the assumption that \(V_n\) converges to a positive definite matrix \(V\).

**Lemma 4.1**: Assume the model defined by (1.6) and let \(s_n^s\) and \(g_n^s\) be defined by (4.6) and (4.7), respectively. Let \(H\) be defined by (2.6) and assume that \(V_n\) converges to a positive definite matrix \(V\). Then \(g_n^s\) converges to...
\( g = \text{vech}(\varepsilon_{n} \varepsilon_{n} + M) \), and \( \Omega_{n} = n \text{Var}(s_{n}) \) converges to a positive definite matrix \( \Omega \), where

\[
\Omega = 2H(\varepsilon_{n} \varepsilon_{n} \preceq \varepsilon_{n} \varepsilon_{n} + 2M \preceq \varepsilon_{n} \varepsilon_{n})H'.
\] (4.10)

**Proof:** Consider an element of \( s_{n} \), \( s_{ij,n} = n^{-1} \sum_{t=1}^{n} z_{it} z_{jt} \).

We have that \( E(s_{ij,n}) = n^{-1} \sum_{t=1}^{n} (\sigma_{ij} + \mu_{it} \mu_{jt}) = \sigma_{ij} + m_{ij,n} \), where \( m_{ij,n} \) is the \((i,j)\)th element of \( M_{n} \).

But \( m_{ij,n} \) converges to \( m_{ij} \), the \((i,j)\)th element of \( M \), so that \( \Omega_{n} \) converges to \( \Omega = \text{vech}(\varepsilon_{n} \varepsilon_{n} + M) \).

To find the limit of \( \Omega_{n} \), observe that \( s_{n} = n^{-1} \sum_{t=1}^{n} \text{vec} \varepsilon_{t}^{2} \varepsilon_{t} \). By independence of the \( \varepsilon_{t} \), \( t = 1, 2, \ldots, n \), and using (2.18),

\[
\Omega_{n} = n^{-1} \sum_{t=1}^{n} \text{Var}(H \text{vec} \varepsilon_{t}^{2} \varepsilon_{t})
\]

\[
= n^{-1} \sum_{t=1}^{n} \text{Var}[H(\varepsilon_{t} \varepsilon_{t} + \varepsilon_{t} \varepsilon_{t}) \text{vec} I_{1} \times 1]
\]

\[
= n^{-1} \sum_{t=1}^{n} \text{Var}[H(\mu_{t} + \varepsilon_{t})' \preceq (\mu_{t} + \varepsilon_{t})']
\]

\[
= n^{-1} \sum_{t=1}^{n} \text{Var}[H(\mu_{t} + \varepsilon_{t})' \preceq (\mu_{t} + \varepsilon_{t}')(\mu_{t} + \varepsilon_{t})(\mu_{t} + \varepsilon_{t})]H'.
\]

Because the third moment about the mean of the normal distribution is zero,
\[ \Omega_n = n^{-1} \sum_{t=1}^{n} H[\text{Var}(\varepsilon'_t \alpha \varepsilon'_t) + \text{Var}(\varepsilon'_t \alpha \mu'_t) \nonumber \\
\quad + \text{Var}(\varepsilon'_t \alpha \varepsilon'_t) + \text{Cov}(\mu'_t \alpha \varepsilon'_t, \varepsilon'_t \alpha \mu'_t) \nonumber \\
\quad + \text{Cov}(\varepsilon'_t \alpha \mu'_t, \mu'_t \alpha \varepsilon'_t)]H' \nonumber \]

From Definition 2.6, it follows that \( H(\mu'_t \alpha \varepsilon'_t) = H(\varepsilon'_t \alpha \mu'_t) \), so that

\[ H \text{Var}(\varepsilon'_t \alpha \mu'_t)H' = H \text{Var}(\mu'_t \alpha \varepsilon'_t)H' \],

and

\[ H \text{Cov}(\mu'_t \alpha \varepsilon'_t, \varepsilon'_t \alpha \mu'_t)H' = E\{[H(\mu'_t \alpha \varepsilon'_t)] [H(\varepsilon'_t \alpha \mu'_t)]'\} \nonumber \\
\quad = E\{[H(\mu'_t \alpha \varepsilon'_t)] [H(\mu'_t \alpha \varepsilon'_t)]'\} \nonumber \\
\quad = H \text{Var}(\mu'_t \alpha \varepsilon'_t)H' \].

Similarly,

\[ H \text{Cov}(\varepsilon'_t \alpha \mu'_t, \mu'_t \alpha \varepsilon'_t)H' = H \text{Var}(\mu'_t \alpha \varepsilon'_t)H' \].

Thus,

\[ \Omega_n = n^{-1} \sum_{t=1}^{n} H[\text{Var}(\varepsilon'_t \alpha \varepsilon'_t) + 4 \text{Var}(\mu'_t \alpha \varepsilon'_t)]H' \].

Now,
\[
\text{Var}(\mu' \alpha \varepsilon'_{t}) = E[(\mu' \alpha \varepsilon'_{t})(\mu' \alpha \varepsilon'_{t})'] \\
= E[(\mu' \alpha \varepsilon'_{t})(\mu' \alpha \varepsilon'_{t})] \\
= E(\mu' \mu_{t} \alpha \varepsilon'_{t} \varepsilon'_{t}) \\
= \mu' \mu_{t} \alpha \varepsilon_{t} \varepsilon'_{t}.
\]

Also, from Lemma 3.1,

\[
H \text{Var}(\varepsilon'_{t} \alpha \varepsilon'_{t})H' = 2H(\varepsilon_{t} \alpha \varepsilon'_{t})H'.
\]

Therefore,

\[
\Omega_{n} = 2n^{-1} \sum_{t=1}^{n} H(\varepsilon_{t} \alpha \varepsilon'_{t} + 2(\mu' \mu_{t} \alpha \varepsilon_{t} \varepsilon'_{t}))H' \\
= 2H(\varepsilon_{t} \alpha \varepsilon'_{t} + 2M_{n} \alpha \varepsilon_{t} \varepsilon'_{t})H'.
\]

Since \( \Omega_{n} \) converges to a positive definite matrix \( \Omega \), \( \Omega = \Omega' \Omega' \) for some nonsingular matrix \( \Omega' \), implying that \( M = \Gamma' \Gamma' \), where \( \Gamma = \Omega(\beta, I_{k} x k) \). Also, \( \varepsilon_{t} \varepsilon'_{t} = \Gamma' \Gamma' \) for some nonsingular matrix \( \Gamma' \). Thus,

\[
M \alpha \varepsilon_{t} \varepsilon'_{t} = \Gamma' \Gamma' \alpha \Gamma' \Gamma' \\
= (\Gamma' \alpha \Gamma') (\Gamma' \alpha \Gamma') \\
= (\Gamma' \alpha \Gamma')' (\Gamma' \alpha \Gamma')
\]

and

\[
\varepsilon_{t} \varepsilon'_{t} \alpha \varepsilon_{t} \varepsilon'_{t} = \Gamma' \Gamma' \alpha \Gamma' \Gamma' \\
= (\Gamma' \alpha \Gamma')' (\Gamma' \alpha \Gamma'),
\]
where $T \otimes T$ is nonsingular since $T$ is nonsingular. Hence $Z_{\infty} \otimes Z_{\infty}$ is positive definite and $M \otimes Z_{\infty}$ is positive semidefinite, so that $\Omega = \lim_{n \to \infty} \Omega_n = 2H(Z_{\infty} \otimes Z_{\infty} + 2M \otimes Z_{\infty})H'$ is positive definite.

The following result is given in Nussbaumi (1978).

**Theorem 4.1:** Let the assumptions of Lemma 4.1 hold. Assume that $Y_n$ converges to a positive definite matrix $Y$. Then $\frac{1}{n^2}(s_n - \sigma_n)$ converges in distribution to a normal random vector with mean zero and covariance matrix $\Omega$.

**Proof:** It suffices to show that for each $q$-dimensional nonnull vector of constants $a = (a_1, a_2, \ldots, a_q)'$ that

$$\frac{1}{n} a'(s_n - \sigma_n) \xrightarrow{d} N(0, a' \Omega a). \quad (4.11)$$

Define the $(p+k) \times (p+k)$ symmetric matrix $W$ to be the matrix such that $\text{vech}(W) = (G'G)^{-1}a$, where $G$ is defined by $\text{vec} S_n = G S_n$. Then,

$$a'(s_n - \sigma_n) = a'(G'G)(s_n - \sigma_n)$$

$$= [\text{vech}(W)]'G'G(s_n - \sigma_n)$$

$$= [\text{vec}(W)]'\{\text{vec}[S_n - E(S_n)]\}$$

$$= \text{tr} W[S_n - E(S_n)].$$

Thus,
\[
\frac{1}{n^2} a'(s_n - \sigma_n) = \frac{1}{n^2} \text{tr}\{W[S_n - E(S_n)]\}
\]
\[
= \frac{-1}{n} \text{tr}\{W[\sum_{t=1}^{\infty} (Z'_t Z_t - E(Z'_t Z_t))]\}.
\]

Let \( \sum_{i=1}^{p+k} \lambda_i \xi_i \xi_i' = \frac{1}{2} \frac{1}{2} Z_{\xi \xi} Z_{\xi \xi} \) be the spectral decomposition of
\[
\frac{1}{2} \frac{1}{2} \text{tr}\{W[S_n - E(S_n)]\} = \frac{-1}{n} \text{tr}\{\sum_{t=1}^{\infty} \frac{-1}{2} \frac{-1}{2} \lambda_i \xi_i \xi_i' Z_{\xi \xi} Z_{\xi \xi} Z_t Z_t\}
\]

\[
\frac{1}{2} \frac{1}{2} \text{tr}\{W(S_n - E(S_n))\} = n \sum_{t=1}^{\infty} \frac{-1}{2} \frac{-1}{2} \sum_{i=1}^{p+k} \lambda_i \xi_i \xi_i' Z_{\xi \xi} Z_{\xi \xi} Z_t Z_t\}
\]

\[
\frac{1}{2} \frac{1}{2} \text{tr}\{W(S_n - E(S_n))\} = n \sum_{t=1}^{\infty} \frac{-1}{2} \frac{-1}{2} \sum_{i=1}^{p+k} \lambda_i \xi_i \xi_i' Z_{\xi \xi} Z_{\xi \xi} Z_t Z_t\}
\]

Because \( \xi_i \xi_j = 0 \) for all \( i \neq j \), the \( p+k \) outer summands of the last expression are independent random variables.
Thus, for proving asymptotic normality of (4.11) it suffices to prove that each of the $p+k$ outer summands is asymptotically normal. Now, $\sum_{t=1}^{n} \left( \xi_t^t Z_{t} + \frac{1}{2} Z_t \right)^2$ is distributed as a non-central chi-square random variable with noncentrality parameter $\omega_n$, where

$$\omega_n = \frac{1}{n} \sum_{t=1}^{n} \left( \xi_t^t Z_{t} + \frac{1}{2} Z_t \right)^2 = n \sum_{t=1}^{n} \xi_t^t Z_{t} + n \frac{1}{2} \xi_t Z_t.$$

By a suitable transformation [Anderson (1958, Exercise 3.7)], it can be shown that

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \left( (\xi_t^t Z_{t} + \frac{1}{2} Z_t)^2 - E(\xi_t^t Z_{t} + \frac{1}{2} Z_t)^2 \right)$$

$$= n^{-\frac{1}{2}} \sum_{t=1}^{n-1} \left( \gamma_t^2 - E(\gamma_t^2) \right) + n^{-\frac{1}{2}} (\gamma_0 + \omega_n)^2 - n^{-\frac{1}{2}} E(\gamma_0 + \omega_n)^2, \quad (4.13)$$

where $\gamma_i$, $i = 1, 2, \ldots, n$ are independent standard normal variables. Now $n^{-\frac{1}{2}} \sum_{t=1}^{n} \left( \gamma_t^2 - E(\gamma_t^2) \right)$ is asymptotically normal. Also,
\[ n^{-\frac{1}{2}} (\gamma_0 + \frac{1}{n} \omega_n^2) - n^{-\frac{1}{2}} \mathbb{E}(\gamma_0 + \frac{1}{n} \omega_n^2) \]

\[ = n^{-\frac{1}{2}} [\gamma_0^2 - \mathbb{E}(\gamma_0^2)] + 2 (\omega_n n^{-1})^{\frac{1}{2}} \gamma_0 \]

and

\[ \lim_{n \to \infty} n^{-1} \omega_n = \xi_i, \quad \xi_i \sim \mathcal{N}(0, \Sigma) \]

(say)

\[ = \omega_0. \]

Therefore, the last two summands in (4.13) converge in law to \( N(0, 4\omega_0) \), so that the entire expression (4.13) is asymptotically normal. Thus, expression (4.12), or equivalently

\[ n^{-\frac{1}{2}} a'(s_n - \sigma_n) \]

is asymptotically normal for all non-null \( a \). From Lemma 4.1, \( \text{Var}[n^{-\frac{1}{2}} (s_n - \sigma_n)] = a' \Omega_n a \) converges to \( a' \Omega a \). Therefore, \( n^{-\frac{1}{2}} (s_n - \sigma_n) \) converges in distribution to a normal vector random variable with mean zero and covariance matrix \( \Omega \).

The next three theorems extend to the functional model the results of Theorems 3.1, 3.3, and 3.4. The first theorem proves the consistency of a class of estimators defined to be those estimators which minimize

\[ f(\gamma \mid \bar{D}) = [s_n - \sigma_n(\gamma)]' \bar{D} [s_n - \sigma_n(\gamma)], \quad (4.14) \]
where $D$, of dimension $q \times q$, is either a random matrix which converges in probability to a positive definite matrix $D$ as $n$ approaches infinity, or a positive definite constant matrix ($D = \bar{D}$). Estimators which minimize (4.14) are called generalized least squares (G.L.S.) estimators, and will be denoted by $\hat{y}$. By Lemma 4.1 the estimators $\hat{y}$ are contained in the class of G.L.S. estimators.

Before stating Theorem 4.2, we need to extend the notion of identifiability to the functional model. For the structural model, $S_n$ is a multiple of a Wishart matrix, and $Z_0 = (\beta, I_{k \times k})'X_{xx}(\beta, I_{k \times k}) + Z_{ee}$ is a function of a finite number of parameters. Recall that $Y_0$ is said to be identified if knowledge of the distribution function of $S_n$ uniquely determines $Y_0$. For the structural model, the distribution function is specified once $Z_0$ is specified. Hence, we defined $Y_0$ to be identified if $Z_n(Y_1) = Z_n(Y_0)$ implies $Y_1 = Y_0$.

Specification of $Z_n(Y_0)$ does not determine the distribution of $S_n$ in the functional case, however. This is because $Z_n(Y_0) = (\beta, I_{k \times k})'V_n(\beta, I_{k \times k}) + Z_{ee}$ is a function of an indefinitely increasing number of parameters, $x_t$, $t = 1, 2, \ldots$. Because we are not interested in estimating the incidental parameters, $x_t$, we include only the elements of $V_n$ in the unknown parameter
vector, \( \gamma_0 \). Thus, \( \gamma_0 \) is composed of elements of \( \hat{y}, x, y, \) and \( v \). We define \( \bar{\gamma}_0 = \lim_{n \to \infty} \gamma_0 \) so that \( \bar{\gamma}_0 \) is composed of elements of \( \hat{y}, x, y, \) and \( v \). We say \( \bar{\gamma}_0 \) is identified if \( \bar{\gamma}(\gamma_0) = \bar{\gamma}(\bar{\gamma}_0) \) implies \( \bar{\gamma}_1 = \bar{\gamma}_0 \).

**Theorem 4.2:** Let the assumptions of Theorem 4.1 hold. Assume that \( \bar{\gamma}_0 \) is identified. Then the G.L.S. estimators are consistent.

**Proof:** Since \( \gamma \) is identified and \( \bar{D} \) is positive definite, \( [\bar{\sigma}(\bar{\gamma}_0) - \bar{\sigma}(\gamma)]' \bar{D} [\bar{\sigma}(\bar{\gamma}_0) - \bar{\sigma}(\gamma)] \) has its absolute minimum of zero at \( \gamma = \bar{\gamma}_0 \). Now from Lemma 4.1 we have that \( s_n \) converges in probability to \( \bar{\sigma}(\bar{\gamma}_0) \), since \( E(s_n) \) and \( \text{Var}(s_n) \) converge to \( \bar{\sigma}(\bar{\gamma}_0) \) and zero, respectively. Also, \( \bar{D} \) converges in probability to \( \bar{D} \) and \( \bar{\sigma}(\gamma) \) is bounded in a neighborhood of \( \gamma = \bar{\gamma}_0 \). Consequently, \( f(\gamma|D) \) converges in probability to \( [\bar{\sigma}(\bar{\gamma}_0) - \bar{\sigma}(\gamma)]' \bar{D} [\bar{\sigma}(\bar{\gamma}_0) - \bar{\sigma}(\gamma)] \) uniformly in a neighborhood of \( \gamma = \bar{\gamma}_0 \). Since \( f(\gamma|D) \) is continuous in \( \gamma \), the point \( \bar{\gamma} \) where it has its absolute minimum converges stochastically to \( \bar{\gamma}_0 \).

We state the obvious corollary.

**Corollary 4.2.1:** Let the assumptions of Theorem 4.2 hold. Then the estimators \( \hat{\gamma} \) are consistent.
Theorem 4.3: Assume the model defined by (1.6) and let $s_n$ and $\sigma_n(\gamma_0)$ be defined by (4.6) and (4.7), respectively. Let $V_n$ be defined by (4.3). Assume $V_n$ converges to a positive definite matrix $V$ and that $\gamma_0$ is identified. Define the $q \times r$ matrix $\Delta$ by

$$\Delta = \left[ \frac{\partial \sigma(\gamma)}{\partial \gamma} \right]_{\gamma = \gamma_0},$$

and suppose that $\Delta$ has full column rank. Then the limiting distribution of $n \left( \bar{\gamma} - \gamma_0 \right)$ is multivariate normal with zero mean and covariance matrix

$$(\Delta' \Delta)^{-1} \Delta' \Omega \Delta (\Delta' \Delta)^{-1}. \quad (4.15)$$

Proof:

Let

$$h(\gamma | \tilde{D}) = -\frac{\partial f(\gamma | \tilde{D})}{\partial \gamma} = 2\frac{\partial \sigma'(\gamma)}{\partial \gamma} \tilde{D}[s_n - \sigma_n(\gamma)].$$

We may express a typical element of the column vector
\[ h_i(\gamma|\tilde{\gamma}) \]

where \( \gamma_i \) is the \( i \)th row of \( \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_n \end{bmatrix}_i \).

By Taylor's theorem,

\[ h_i(\tilde{\gamma}|D) = h_i(\gamma_0|D) - o(\tilde{\gamma} - \gamma_0), \]

where the \((i,j)\)th element of \( \theta_n \) is given by

\[ \theta_{ij} = -\frac{\partial h_i}{\partial \gamma_j} \bigg|_{\gamma = \gamma_0} - \frac{1}{2} \sum_{\ell=1}^{r} (\tilde{\gamma}_\ell - \gamma_0, \ell) \frac{\partial^2 h_i}{\partial \gamma_j \partial \gamma_\ell} \bigg|_{\gamma = \gamma^*}, \quad (4.16) \]

and \( \gamma^* \) lies between \( \gamma_0 \) and \( \tilde{\gamma} \).

Now,

\[ \frac{\partial h_i}{\partial \gamma_j} = 2\left\{ \frac{\partial^2 \sigma_i^\prime(\gamma)}{\partial \gamma_j \partial \gamma_j} \right\} \left[ s_n - \sigma_n(\gamma) \right] - \frac{\partial \sigma_i^\prime(\gamma)}{\partial \gamma_j} \left[ \frac{\partial s_n}{\partial \gamma_i} \right] - \frac{\partial \sigma_i(\gamma)}{\partial \gamma_j} \left[ \frac{\partial \sigma_n}{\partial \gamma_i} \right] \]

and

\[ \frac{\partial^2 h_i}{\partial \gamma_j \partial \gamma_\ell} = 2\left\{ \frac{\partial^3 \sigma_i^\prime(\gamma)}{\partial \gamma_j \partial \gamma_\ell \partial \gamma_\ell} \right\} \left[ s_n - \sigma_n(\gamma) \right] - \frac{\partial^2 \sigma_i^\prime(\gamma)}{\partial \gamma_j \partial \gamma_\ell} \left[ \frac{\partial s_n}{\partial \gamma_i} \right] - \frac{\partial \sigma_i(\gamma)}{\partial \gamma_j} \left[ \frac{\partial^2 \sigma_n}{\partial \gamma_i \partial \gamma_\ell} \right] - \frac{\partial \sigma_i(\gamma)}{\partial \gamma_\ell} \left[ \frac{\partial^2 \sigma_n}{\partial \gamma_i \partial \gamma_j} \right]. \]

Since the elements of \( s_n - \sigma_n(\gamma_0) \) and \( \tilde{\gamma} - \gamma_0 \) converge to zero in probability, \( \tilde{\gamma} \) converges to \( \gamma_0 \) in probability,
and the partial derivatives are bounded in a neighborhood of \( \bar{Y}_0 \), it follows that \( \theta_{ij} \) converges stochastically to

\[
2 \left[ \frac{\partial \sigma'(\bar{Y}_0)}{\partial Y_i} \overline{D} - \frac{\partial \sigma(\bar{Y}_0)}{\partial Y_j} \right].
\]

That is,

\[
\lim_{n \to \infty} \theta_n = 2 \bar{\Delta}' \overline{D} \Delta. \tag{4.17}
\]

By assumption \( \Delta \) is of full column rank, implying that \( 2 \bar{\Delta}' \overline{D} \Delta \) is nonsingular.

Because \( h(\bar{Y} \mid \bar{D}) = 0 \), it follows that

\[
\bar{Y} = \gamma_0 + \Theta_n^{-1} h(c_\gamma \mid \bar{D}), \tag{4.18}
\]

and that \( \lim_{n \to \infty} n^2 (\bar{Y} - \gamma) = 0 \), where

\[
\bar{Y} = \gamma_0 + (2 \bar{\Delta}' \overline{D} \Delta)^{-1} h(c_\gamma \mid \bar{D})
\]

\[
= \gamma_0 + (\Delta' \overline{D} \Delta)^{-1} \Delta' \overline{D} \left[ s_n - c_n(c_\gamma) \right], \tag{4.19}
\]

because

\[
\frac{1}{n^2} (\bar{Y} - \gamma) = \left[ \Theta^{-1} \bar{\Delta}' \overline{D} - (\Delta' \overline{D} \Delta)^{-1} \Delta' \overline{D} \right] \frac{1}{n^2} \left[ s_n - c_n(c_\gamma) \right]
\]

\[
= \left\{ \Theta^{-1} (\Delta' \overline{D} \Delta)^{-1} \Delta' \overline{D} \right\} \frac{1}{n^2} \left[ s_n - c_n(c_\gamma) \right]. \tag{4.20}
\]

The limiting distribution of \( n^2 (\bar{Y} - Y_0) \), and hence of
\( \frac{1}{n^2} (\tilde{\gamma} - \gamma_0) \), is multivariate normal with zero mean vector and dispersion matrix given by

\[
(\Delta' \tilde{D} \Delta)^{-1} \Delta' \tilde{D} \Omega \tilde{D} \Delta (\Delta' \tilde{D} \Delta)^{-1}.
\]

We say for any two square matrices \( \tilde{A} \) and \( \tilde{B} \) that \( \tilde{A} \) is bounded below by \( \tilde{B} \) in the Loewner sense of inequality, denoted by \( \tilde{A} \geq \tilde{B} \), if \( \tilde{A} - \tilde{B} \) is positive semidefinite. Among G.L.S. estimators, the estimators \( \tilde{\gamma} \) are shown to be "best," in the Loewner sense of having minimum asymptotic variances, in the following corollary due to Browne (1974).

**Corollary 4.3.1:** The asymptotic dispersion matrix of the normalized G.L.S. estimator \( \frac{1}{n^2} (\tilde{\gamma} - \gamma_0) \) is bounded below by \( (\Delta' \tilde{\Omega}^{-1} \Delta)^{-1} \) in the Loewner sense of inequality. This bound is attained by the normalized G.L.S. estimators \( \frac{1}{n^2} (\tilde{\gamma} - \gamma_0) \).

**Proof:** Since \( \left[ n \text{Var}(\tilde{s}_n) \right]^{-1} = \tilde{\Omega}_n^{-1} \) converges to \( \tilde{\Omega}_0^{-1} \), \( \frac{1}{n^2} (\tilde{\gamma} - \gamma_0) \) has asymptotic dispersion matrix \( (\Delta' \tilde{\Omega}^{-1} \Delta)^{-1} \). Furthermore,

\[
(\Delta' \tilde{D} \Delta)^{-1} \Delta' \tilde{D} \Omega \tilde{D} \Delta (\Delta' \tilde{D} \Delta)^{-1} - (\Delta' \tilde{\Omega}^{-1} \Delta)^{-1} \\
= (\tilde{D} \Delta (\Delta' \tilde{D} \Delta)^{-1} - \tilde{\Omega}^{-1} \Delta (\Delta' \tilde{\Omega}^{-1} \Delta)^{-1})' \tilde{\Omega} [\tilde{D} \Delta (\Delta' \tilde{D} \Delta)^{-1} \\
- \tilde{\Omega}^{-1} \Delta (\Delta' \tilde{\Omega}^{-1} \Delta)^{-1}] \\
\geq 0,
\]
since \( \Omega \) is positive definite.

In addition to providing a "best" G.L.S. estimator, expression (4.14) can be used to test the null hypothesis that \( \varphi = \varphi(\gamma_0) \) against the alternative that \( \varphi \neq \varphi(\gamma_0) \).

**Theorem 4.4:**

Let the assumptions of Theorem 4.1 hold. Assume that \( \gamma_0 \) is identified and that \( \Delta \) has full column rank. Then the limiting distribution of \( nf(\gamma|D(\gamma)) \) is chi-square with \( q-r \) degrees of freedom.

**Proof:** It was seen from (4.20) in Theorem 4.3 that

\[
\frac{1}{n^2}(\hat{\gamma} - \gamma) \text{ converges in probability to a null vector. Also,}
\]

\[
\frac{1}{n^2}[\sigma_n(\hat{\gamma}) - \sigma_n(\gamma_0) - \Delta(\hat{\gamma} - \gamma_0)] \text{ converges in probability to a null vector since, by Taylor's theorem,}
\]

\[
\frac{1}{n^2}\{\varphi_i(\gamma) - \left[\varphi_i(\gamma_0) + \frac{\partial \varphi_i(\gamma)}{\partial \gamma} \frac{1}{2} (\hat{\gamma} - \gamma_0)\right] \}
\]

\[= \left(\frac{1}{2^2}\right)(\hat{\gamma} - \gamma_0)^t \left[\frac{\partial^2 \varphi_i(\gamma)}{\partial \gamma^t \partial \gamma} \right]_{\gamma = \gamma^*} (\hat{\gamma} - \gamma_0),\]

where

\[
\frac{\partial^2 \varphi_i(\gamma)}{\partial \gamma^t \partial \gamma} = \frac{\partial}{\partial \gamma} \left[\frac{\partial \varphi_i(\gamma)}{\partial \gamma}\right] \text{ and } \gamma^* \text{ lies between } \hat{\gamma} \text{ and } \gamma_0.\]
Consequently, $\frac{1}{n^2}[s_n - \sigma_n(\hat{\gamma})]$ converges stochastically to
\[
\frac{1}{n^2}[s_n - \sigma_n(\hat{\gamma})] + \frac{1}{n^2}[\Delta(\hat{\gamma} - \gamma)] + \frac{1}{n^2}[\sigma_n(\hat{\gamma}) - \sigma_n(\gamma_0) - \Delta(\hat{\gamma} - \gamma_0)]
\]
\[
= \frac{1}{n^2}[s_n - \sigma(\gamma_0) - \Delta(\hat{\gamma} - \gamma_0)]
\]
\[
= \frac{1}{n^2}[I - \Delta(\Omega^{-1}\Delta)^{-1} - \Omega^{-1}] [s_n - \sigma_n(\gamma_0)],
\]
the last equality holding because of (4.19) in Theorem 4.3. Then $nf(\hat{\gamma}|D(\Omega))$ converges stochastically to, say, $f_0'$, where
\[
f_0' = n[s_n - \sigma(\gamma_0)]'M_0 [s_n - \sigma_n(\gamma_0)],
\]
and,
\[
M_0 = [I - \Delta(\Omega^{-1}\Delta)^{-1} - \Omega^{-1}] [I - \Delta(\Omega^{-1}\Delta)^{-1} - \Omega^{-1}]
\]
\[
= \Omega^{-1} - \Omega^{-1}(\Delta'\Omega^{-1}\Delta)^{-1} - \Omega^{-1}.
\]
Since $M_0\Omega$ is idempotent of rank $q-r$, the limiting distribution of $nf_0'$ and of $nf(\hat{\gamma}|D(\Omega))$ is the central chi-square distribution with $q-r$ degrees of freedom [Graybill (1976, pp. 135-136)].

We have seen that within the class of G.L.S. estimators, the estimators $\hat{\gamma}$ have desirable properties. However, calculation of $\hat{\gamma}$ requires a consistent estimator of $\Omega^{-1}$. If $\nu_n$ converges elementwise to $\nu$, we
have the following.

\[ E(S_n) = \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} + M_n \]

\[ \Omega_n = 2H(\tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \otimes \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} + 2M_n \otimes \tilde{\mathbb{Z}}_{\varepsilon \varepsilon})^H \]

\[ \lim_{n \to \infty} \tilde{\Omega}_n = \Omega, \]

where \( M_n \) was defined in (4.4). This suggests the following procedure.

1. Obtain a consistent estimator of \( \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \), say \( \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \), where \( \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \) is positive definite.

2. Use \( S_n \) and \( \tilde{S}_{\varepsilon \varepsilon} \) to obtain a positive semidefinite matrix \( \tilde{M}_n \) which is a consistent estimator of \( M \).

3. Set \( \tilde{\Omega} = 2H(\tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \otimes \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} + 2\tilde{M}_n \otimes \tilde{\mathbb{Z}}_{\varepsilon \varepsilon})^H \) and solve for \( D = \tilde{\Omega}^{-1} \).

We consider each step in more detail.

**Step 1:** (Obtain \( \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \))

Suppose repeated observations on \((Y_t, X_t)\) are available, say \((\tilde{Y}_{ts}, \tilde{X}_{ts})\), where

\[ \tilde{Y}_{ts} = Y_t + e_{ts}, \]

\[ \tilde{X}_{ts} = X_t + u_{ts}, \]

and \((e_{ts}, u_{ts})\) are independent identically distributed as \((p+k)\)-variate normal vectors with covariance matrix \( \tilde{\mathbb{Z}}_{\varepsilon \varepsilon} \).
s = 1, 2, ..., r_0, \ t = 1, 2, ..., n. The true values \((y_t, x_t)\)
are assumed to satisfy

\[ y_t = x_t \beta. \]  \hspace{1cm} (4.22)

Then,

\[ \tilde{\Sigma}_{\varepsilon \varepsilon} = [n(r_0-1)]^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} (\bar{y}_{ts} - \bar{y}_t, \bar{x}_{ts} - \bar{x}_t) \]

\[ \times (\bar{y}_{ts} - \bar{y}_t, \bar{x}_{ts} - \bar{x}_t)'. \]  \hspace{1cm} (4.23)

where

\[ \bar{y}_t = r_0^{-1} \sum_{s=1}^{r_0} y_{ts} \quad \text{and} \quad \bar{x}_t = r_0^{-1} \sum_{s=1}^{r_0} x_{ts}, \]

is an unbiased estimator of \(\Sigma_{\varepsilon \varepsilon}\). In fact, \(n(r_0-1)\tilde{\Sigma}_{\varepsilon \varepsilon}\)
has a Wishart distribution with \(n(r_0-1)\) degrees of freedom, so that \(\tilde{\Sigma}_{\varepsilon \varepsilon} = \Sigma_{\varepsilon \varepsilon} + O_p(n^{-1})\). The matrix \(\tilde{\Sigma}_{\varepsilon \varepsilon}\) will be positive definite with probability one.

If repeated observations on \((\bar{y}_t, \bar{x}_t)\) are not available, but the unknown parameters of \(\Sigma_{\varepsilon \varepsilon}\) are identified, the G.L.S. procedure with \(D = I\) will yield a consistent estimator \(\tilde{\Sigma}_{\varepsilon \varepsilon}\). The unknown parameters of \(\Sigma_{\varepsilon \varepsilon}\) could be identified if we have prior knowledge of the structure of \(\Sigma_{\varepsilon \varepsilon}\). For example, it may be reasonable to assume that \(\Sigma_{\varepsilon \varepsilon}\)
is diagonal. Let \(\gamma\) include the unknown parameters of \(\Sigma_{\varepsilon \varepsilon}\) and assume the conditions of Theorem 4.4 are satisfied. From (4.18) in Theorem 4.3 we have that
\[ \tilde{y} = \tilde{y}_0 + \tilde{\theta}^{-1} h(\tilde{y}_0 | D). \]

For \( D = I \) and since \( \text{Var}(s_{\xi}) = O(n^{-1}) \),
\[ h(\tilde{y}_0 | D) = 2\tilde{\Delta}' [s_n - \tilde{\sigma}_n(\tilde{y}_0)] \]
\[ = O_p(n^{-2}). \]

Also, from (4.17) in Theorem 4.3, \( \lim_{n \to \infty} \tilde{\theta}^{-1} = 2^{-1}(\tilde{\Delta}'\tilde{\Delta})^{-1} \), so
\[ \tilde{\theta}^{-1} = O_p(1) \]. Thus, \( \tilde{y} = y_0 + O_p(n^{-2}) \), or,
\[ \tilde{y}_{\xi\xi} = y_{\xi\xi} + O_p(n^2). \]

Our choice of \( D = I \) is for illustrative purposes only. In practice one might set \( D = 2^{-1} \tilde{\Delta}'(\tilde{s}_n^{-1} \otimes \tilde{s}_n^{-1}) \tilde{\Delta} \). This choice of \( D \) provides a consistent estimator of \( \text{Var}[n^2(s_n - \sigma_n)] \)
for the structural model, and converges to a positive definite constant matrix for the functional model.

The G.L.S. estimator \( \tilde{y}_{\xi\xi} \) so obtained may not be positive definite, but it is possible to modify the estimator to produce a positive definite matrix. We shall henceforth assume that the estimator of \( \tilde{y}_{\xi\xi} \) is positive definite.

**Step 2:** (Obtain \( \tilde{M}_n \))

Define \( \tilde{M}_n \) as
\[ \tilde{M}_n = s_n - \tilde{\lambda}_{\xi\xi}, \]
(4.24)
where $\tilde{\lambda}$ is the smallest root of $|S_n - \tilde{\lambda}\tilde{Z}_{\varepsilon\varepsilon}| = 0$, and $\tilde{Z}_{\varepsilon\varepsilon}$ is the estimate of $Z_{\varepsilon\varepsilon}$ obtained from step 1. Then, as the next theorem shows, $\tilde{M}_n$ is a positive semi-definite matrix which is a consistent estimator of $\hat{M}$.

**Theorem 4.5:** Let the assumptions of Theorem 4.1 hold. Let $\tilde{M}_n$ be defined by (4.24). Then $\tilde{M}_n$ is a positive semi-definite matrix which converges in probability to $\hat{M}$.

**Proof:** We consider minimization with respect to $\tilde{\alpha}$ of

$$\frac{\tilde{\alpha}'S_n\tilde{\alpha}}{\tilde{\alpha}'\tilde{Z}_{\varepsilon\varepsilon}\tilde{\alpha}},$$

(4.25)

where $\tilde{\alpha}(\neq 0)$ is any $(p+k)$-dimensional column vector of unknown variables. Minimization of (4.25) with respect to $\tilde{\alpha}$ is equivalent to the following problem.

**minimize:** $\tilde{\alpha}'S_n\tilde{\alpha}$

**subject to:** $\tilde{\alpha}'\tilde{Z}_{\varepsilon\varepsilon}\tilde{\alpha} = 1$.

The Lagrangian for this problem is

$$f(\tilde{\alpha}, \lambda) = \tilde{\alpha}'S_n\tilde{\alpha} - \lambda(\tilde{\alpha}'\tilde{Z}_{\varepsilon\varepsilon}\tilde{\alpha} - 1).$$

Therefore, the $\tilde{\alpha}$ minimizing (4.25) satisfies

$$(S_n - \tilde{\lambda}\tilde{Z}_{\varepsilon\varepsilon})\tilde{\alpha} = 0,$$

where $\tilde{\lambda}$ is the smallest root of $|S_n - \tilde{\lambda}\tilde{Z}_{\varepsilon\varepsilon}| = 0$. The minimum value for the ratio (4.25) is
Thus, for any $\alpha \neq 0$,

$$\frac{\alpha' S_n \alpha}{\alpha' \tilde{\alpha}} \geq \lambda.$$

$S_n - \frac{\tilde{\lambda} \tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}}$ is positive semidefinite since

$$\alpha' (S_n - \frac{\tilde{\lambda} \tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}}) \alpha \geq 0.$$

Under the stated assumptions

$$\text{plim } S_n = \mathcal{Z} \tilde{\epsilon} \tilde{\epsilon} + M$$

and

$$\text{plim } \frac{\tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}} = \frac{\mathcal{Z}}{\mathcal{Z} \epsilon \epsilon}.$$

Because $\tilde{\lambda}$ is a locally continuous function of the elements of $S_n$ and $\frac{\tilde{\lambda} \tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}}$, and because $M_n$ is of rank $k$ for all $n$,

$$\text{plim } \tilde{\lambda} = 1.$$

Therefore,

$$\text{plim } S_n - \frac{\tilde{\lambda} \tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}} = \frac{\mathcal{Z}}{\mathcal{Z} \epsilon \epsilon} + M - \frac{\mathcal{Z}}{\mathcal{Z} \epsilon \epsilon} = M.$$

**Step 3:** (Calculate $\tilde{\mathcal{O}}^{-1}$)

Because $\frac{\tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}}$ is positive definite and $M_n$ is positive semidefinite, the matrix $\tilde{\mathcal{O}}$ is positive definite, where

$$\tilde{\mathcal{O}} = 2H(\frac{\tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}} \text{ } 2M_n \text{ } \frac{\tilde{\mathcal{Z}}}{\mathcal{Z} \tilde{\epsilon} \tilde{\epsilon}}) H'.$$
Further, since $\bar{Z}_{\Xi \Xi}$ and $\bar{M}_n$ are consistent estimators of $Z_{\Xi \Xi}$ and $M$, respectively,

$$\text{plim } \bar{\Omega} = \Omega.$$ 

Since the elements of $\bar{\Omega}^{-1}$ are continuous functions of the elements of $\bar{\Omega},$

$$\text{plim } \bar{\Omega}^{-1} = \Omega^{-1}.$$ 

Inversion of $\bar{\Omega}$ may present computational difficulties, since $\bar{\Omega}$ is of dimension $q \times q$, where $q = 2^{-1}(p+k)(p+k+1)$. We note that (4.8) can be written as

$$f(\gamma|D(\bar{\Omega})) = [s_n - \sigma_n(\gamma)]'T_n^-T_n^- [s_n - \sigma_n(\gamma)],$$

where $T_n^-T_n^- = [\text{Var}(s_n)]^{-1}$. The following theorem shows how to construct such a $T_n$, assuming $E(S_n)$ and $Z_{\Xi \Xi}$ are known.

**Theorem 4.6:** Let the assumptions of Theorem 4.1 hold. A matrix, $T_n^-$, such that $T_n^-T_n^- = [\text{Var}(s_n)]^{-1}$, is

$$T_n = L_n H(R_n \otimes R_n) G,$$  

(4.26)

where $G$ and $H$ are defined by $\text{vec } S_n = G \text{ vech } S_n$ and $H = (G'G)^{-1}G'$, the matrix $R_n$ satisfies

(i) $R_n (Z_{\Xi \Xi} + M_n) R_n' = I_{(p+k)x(p+k)}$
(ii) \( R_n Z_{\varepsilon \varepsilon} R_n' = \text{diag}(d_{11}, d_{22}, \ldots, d_{(p+k),(p+k)}) = D_n \) (say)

(iii) \( R_n M_{\varepsilon \varepsilon} R_n' = \text{diag}(c_{11}, c_{22}, \ldots, c_{(p+k),(p+k)}) = C_n \) (say)

and \( L_n = \text{diag}(b_{11}, b_{22}, \ldots, b_{qq}) \), where \( B_n = \text{diag}(b_{11}, b_{22}, \ldots, b_{qq}) \) is given by

\[
B_n = 2n^{-1} H(D_n \otimes D_n + C_n \otimes C_n)^{H'}.
\]

Proof: From (4.5) we have that \( E(S_n) = Z_{\varepsilon \varepsilon} + M_n \),

where \( M_n = n^{-1} \sum_{t=1}^{n} u_t u_t' \). Since \( Z_{\varepsilon \varepsilon} \) is positive definite, there exists a nonsingular matrix, \( Q \), such that \( Q Z_{\varepsilon \varepsilon} Q' = I_{(p+k) \times (p+k)} \). Also, since \( QM_n Q' \) is a real, positive semi-definite matrix, there exists an orthogonal matrix, \( P_n \), such that \( P_n Q M_n Q' P_n' = \text{diag}(\delta_{11}, \delta_{22}, \ldots, \delta_{(p+k),(p+k)}) \),

where all \( \delta_{ii} > 0 \). Letting \( D_n = \text{diag}(d_{11}, d_{22}, \ldots, d_{(p+k),(p+k)}) \), where

\[
\frac{1}{2} d_{ii} = \left( \frac{1}{1 + \delta_{ii}} \right)^{\frac{1}{2}}, \quad i = 1, 2, \ldots, p+k,
\]

we then have that \( R_n = D_n^{\frac{1}{2}} \) satisfies

(i) \( R_n (Z_{\varepsilon \varepsilon} + M_n) R_n' = I_{(p+k) \times (p+k)} \) (4.27)

(ii) \( R_n Z_{\varepsilon \varepsilon} R_n' = \text{diag}(d_{11}, d_{22}, \ldots, d_{(p+k),(p+k)}) \) (say)

\[
= D_n
\]
(iii) $\mathbf{R}_n \mathbf{M} \mathbf{R}_n' = \text{diag}(c_{11}, c_{22}, \ldots, c_{(p+k),(p+k)})$ \hspace{1cm} (4.29) \\

(say) \\

$= C_n'$ \\

where $c_{ii} = \delta_{ii} \mathbf{d}_{ii}$, $i = 1,2,\ldots, p+k$. Note that the $\mathbf{d}_{ii}$ and $c_{ii}$, $i = 1,2,\ldots, p+k$, are functions of the eigenvalues of $\mathbf{Q}_n \mathbf{Q}_n'$. We have not subscripted $\mathbf{d}_{ii}$ and $c_{ii}$ with $n$ for notational convenience.

Consider now the $(i,j)$th element of $S_n' S_{ij,n}'$ where

$$s_{ij,n} = n^{-1} \sum_{t=1}^{n} \mathbf{z}_i \mathbf{z}_j'$$

and $\mathbf{z}_i$ is the $i$th element of $\mathbf{z}_t$. Then,

$$\text{Cov}(s_{ij,n}', s_{gh,n}') = (\frac{1}{n^2}) \sum_{t=1}^{n} (\sigma_{ig} \sigma_{jh} + \sigma_{ih} \sigma_{jg} + \mu_{it} \mu_{jt} \sigma_{ig} + \mu_{it} \mu_{jt} \sigma_{ig})$$

where $\mu_{it}$ is the $i$th element of $\mu_t$. Let $v_{ij,n}$ denote the $(i,j)$th element of $R_n S_n R_n'$. From (4.1) we have that

$$R_n S_n R_n' = n^{-1} \sum_{t=1}^{n} R_n \mathbf{z}_t \mathbf{z}_t' R_n'$$

$$= n^{-1} \sum_{t=1}^{n} \xi_t' \check{\xi}_t'$$

where $\xi_t' = R_n \mathbf{z}_t'$ is distributed as a $(p+k)$-variate normal random vector with mean vector $R_n \mu_t' = \check{\mu}_t$ and covariance matrix $R_n \mathbf{z}_t R_n' = D_n$. Thus,
\[ \text{Cov}(v_{ij}, v_{gh}, n) = \frac{1}{n^2} \sum_{t=1}^{n} (d_{ij}^t d_{jh}^t + d_{ih}^t d_{jg}^t + v_{it}^t v_{gt}^t d_{jh}^t) \]

\[ + v_{it}^t v_{ht}^t d_{jg}^t + v_{jt}^t v_{gt}^t d_{ih}^t + v_{jt}^t v_{ht}^t d_{ig}^t) \]

\[ = \frac{1}{n} (d_{ig}^t d_{jh}^t + d_{ih}^t d_{jg}^t + c_{ig}^t d_{jh}^t + c_{ih}^t d_{jg}^t \]

\[ + c_{jg}^t d_{ih}^t + c_{jh}^t d_{ig}^t), \]

where \( v_{it} \) is the \( i \)th element of \( v_t \) and \( c_{ij} \) is the \((i,j)\)th element of \( C_n \). Hence, from (4.28) and (4.29),

\[
\text{Cov}(v_{ij}, v_{gh}, n) = \begin{cases} 
   n^{-1} (2d_{ii}^2 + 4c_{ii}^t d_{ii}) & \text{if } i = j = g = h \\
   n^{-1} (d_{ii}^t d_{jj}^t + c_{ii}^t d_{jj}^t + c_{jj}^t d_{ii}) & \text{if } i = g, j = h \\
   0 & \text{otherwise}
\end{cases}
\]

We have defined \( H \) so that \( \text{vech} \ R \frown S \frown R' = H \text{ vec} \ R \frown S \frown R' \).

From (2.18), and Definition 2.7,

\[ \text{vech} \ R \frown S \frown R' = H (R \frown R) \text{ vec} S \frown \]

\[ = H (R \frown R) G S \frown. \]
Now,
\[
\text{Cov}(\text{vech } R_{\sim \! n} S_{\sim \! n} R_{\sim \! n}^t) = \text{diag}(b_{11}, b_{22}, \ldots, b_{qq})
\]
(say)
\[
= B_n,
\]
where \( B_n = 2n^{-1} H(D_n \otimes D_n + C_n \otimes D_n^2) \).

(4.30)

Let
\[
L_n = \text{diag}(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}).
\]

Then
\[
\text{Cov}(L_{\sim \! n} \text{ vech } R_{\sim \! n} S_{\sim \! n} R_{\sim \! n}^t) = I_{q \times q}.
\]

Thus, a matrix \( T_n \), such that \( T_n^t T_n = [\text{Var}(\tilde{s}_n)]^{-1} \) is
\[
T_n = L_n H(R_n \otimes R_n) G.
\]

We can use the results of Theorem 4.6 to avoid the direct computation of \( \Omega^{-1} \).

Corollary 4.6.1: Let the assumptions of Theorem 4.6 hold.

Let \( \tilde{Z}_{\varepsilon \varepsilon} \) and \( \tilde{M}_n \) be positive definite and positive semi-definite matrices, respectively, satisfying \( \text{plim}_{n \to \infty} \tilde{Z}_{\varepsilon \varepsilon} = Z_{\varepsilon \varepsilon} \) and \( \text{plim}_{n \to \infty} \tilde{M}_n = M_n \). Then a matrix, \( \tilde{T}_n \), such that \( \tilde{T}_n^t \tilde{T}_n \) is a consistent estimator of \( [\text{Var}(\tilde{s}_n)]^{-1} \) is
\[
\tilde{T}_n = L_n H(\tilde{R}_n \otimes \tilde{R}_n) G,
\]
where \( G \) and \( H \) are defined by \( \text{vec } \tilde{s}_n = G \text{ vech } \tilde{s}_n \) and
$H = (G'G)^{-1}G'$, and the definitions of $\tilde{\Sigma}_n$ and $\tilde{\Omega}_n$ follow directly from the statement of Theorem 4.6 with $\tilde{\Sigma}_n$ and $\tilde{\Omega}_n$ replacing $\tilde{\Sigma}_n$ and $\tilde{\Omega}_n$, respectively.

**Proof:** By construction,

$$\tilde{T}_n = \left(2n^{-1}H(\tilde{\Sigma}_n \otimes \tilde{\Sigma}_n + \tilde{\Omega}_n \otimes \tilde{\Sigma}_n \tilde{\Sigma}_n)H'\right)^{-1}$$

so that

$$\tilde{T}_n = \left[2n^{-1}H(\tilde{\Sigma}_n \otimes \tilde{\Sigma}_n + \tilde{\Omega}_n \otimes \tilde{\Sigma}_n \tilde{\Sigma}_n)H'\right]^{-1}.$$

Thus,

$$\operatorname{plim} \tilde{T}_n = \left[\operatorname{Var}(\tilde{\Omega})\right]^{-1}.$$

The advantage of computing $\tilde{T}_n$ instead of $\tilde{\Omega}_n$ is that the largest matrix we invert in calculating $\tilde{T}_n$ is of dimension $(p+k) \times (p+k)$, whereas $\tilde{\Omega}_n$ is of dimension $q \times q$. For large $p+k$, say $p+k = 10$, use of Corollary 4.6.1 entails inversion of a $10 \times 10$ matrix. To compute $\tilde{\Omega}_n$ directly, we must invert a $55 \times 55$ matrix.

The models in Chapters III and IV assume normally distributed errors $\varepsilon_t$, and, in this chapter, normally distributed $x_t$. While some of our results (such as Theorem 4.6) depend on these normality assumptions, many theorems depend only on the asymptotic normality of $\tilde{\Sigma}_n$. Chapter V generalizes certain results of Chapters III and IV to multivariate linear errors in variables models where $x_t$ and $\varepsilon_t$ are not normally distributed.
V. ESTIMATION OF NONNORMAL MULTIVARIATE LINEAR ERRORS IN VARIABLES MODELS

A. Introduction

So far we have assumed that the error vectors $\xi_t$, $t = 1, 2, \ldots$, are independent identically distributed as (p+k)-variate normal vectors with mean zero and covariance matrix $\Sigma_{\varepsilon\varepsilon}$. For the structural model of Chapter III, we also assumed the $x_t$, $t = 1, 2, \ldots$, to be independent identically distributed as k-variate normal vectors with mean zero and covariance matrix $\Sigma_{xx}$. However, many of the results of Chapters III and IV depend only on the asymptotic normality of $\frac{1}{n^2} (s_n - \gamma_n)$. The assumption of normality of $\xi_t$ (and $x_t$) is unnecessarily restrictive and will be relaxed in this chapter. In Section B we consider estimation of the structural model when neither $x_t$ nor $\xi_t$ is normally distributed and in Section C we consider estimation of the functional model when $\xi_t$ is not normally distributed. We let $\gamma$ and $\Omega$ denote, as before, the asymptotic mean of $s_n$ and the asymptotic covariance matrix of $\frac{1}{n^2} s_n$, respectively. The matrix $\Omega$ is always taken to be positive definite. We will again find it convenient to use the form of the multivariate linear errors in variables model given by (1.6).
B. Estimation of the Structural Model when $\varepsilon_t$ and $x_t$, $t = 1, 2, \ldots$ are Nonnormal $z_t$

It is necessary for the asymptotic distribution of $\frac{1}{n} (s^2 - \sigma^2)$ to be multivariate normal if we are to extend the results of Chapter III to nonnormally distributed $\varepsilon_t$ and $x_t$. The following lemma states sufficient conditions for the asymptotic normality of $\frac{1}{n} (s_n - \sigma_n)$.

**Lemma 5.1:** Assume the model defined by (1.1-1.5) with the $\varepsilon_t$ and $x_t$, $t = 1, 2, \ldots$, independent identically distributed $(p+k)$-variate and $k$-variate random vectors, respectively, with means zero and $(4+\eta)$th moments. Then $\frac{1}{n^2} (s_n - \sigma_n)$ converges in distribution to a normal vector $\tilde{\Omega}$.

**Proof:** To show the asymptotic normality of $\frac{1}{n} (s_n - \sigma_n)$, we show that for any $q$-dimensional nonnull vector of constants $a = (a_1, a_2, \ldots, a_q)'$ that

$$\frac{1}{n} (W_n - \tilde{w}_n) \xrightarrow{D} N(0, 1), \quad (5.1)$$

where

$$W_n = \sum_{t=1}^{n} \left( \sum_{i=1}^{q} a_i s_{it} \right) = na's_n'$$

and

$$\tilde{w}_n = \sum_{t=1}^{n} \left( \sum_{i=1}^{q} a_i \sigma_i \right) = na'\sigma_n'$$
\[ d_n = \sum_{t=1}^{n} \text{Var} \left( \sum_{i=1}^{q} a_i s_{i,t} \right) = \sum_{t=1}^{n} \text{Var}(a's_t), \]

where \( s_{i,t} \) is the \( i^{th} \) element of \( s_t = \text{vech } Z'Z_t \) and \( \sigma_i = \text{E}(s_{i,t}) \) is the \( i^{th} \) element of \( \sigma = \text{E}(s_t) \). For (5.1) to hold it is sufficient to show that the Liapounov condition is satisfied. That is, for all \( \varepsilon > 0 \), and for some \( \delta > 0 \),

\[
\lim_{n \to \infty} d_n \sum_{t=1}^{n} \left[ \frac{1}{2} \varepsilon + \left(1 + \frac{\varepsilon}{2}\right) \right] \frac{1}{2} \sum_{t=1}^{n} \left| z - a' \sigma_t \right|^2 dF_t(z) = 0,
\]

where \( F_t \) is the distribution function of \( a's_t \). Since the \( s_{i,t} \), \( t = 1, 2, \ldots, n \) are independent,

\[
d_n = \sum_{t=1}^{n} \text{Var}(a's_t) = a' \left[ \text{Var} \left( \sum_{i=1}^{n} s_{i,t} \right) \right]a
\]

\[
= n^2 a' \text{Var}(s_n) a.
\]

By assumption we have that \( na' \text{Var}(s_n) a \) converges to some positive constant \( d = a'Qa \). Thus, for, \( \delta = 2^{-1}n \), we have

\[
\lim_{n \to \infty} d_n \sum_{t=1}^{n} \left[ \frac{1}{2} \varepsilon + \left(1 + \frac{\varepsilon}{2}\right) \right] \frac{1}{2} \sum_{t=1}^{n} \left| z - a' \sigma_t \right|^2 dF_t(z) = d \lim_{n \to \infty} \left(1 + \frac{n}{4}\right) \sum_{t=1}^{n} \left| z - a' \sigma_t \right|^2 dF_t(z),
\]
provided the limit on the right exists. Because the
(4+n)th moments of \( \varepsilon_t \) and \( x_t \) exist,
\( n^{-1} \sum_{t=1}^{n} |z-a'\sigma|^{2} + \frac{n}{2} dF_{t}(z) \) is bounded, so that
\[
\lim_{n \to \infty} n \sum_{t=1}^{n} \left[ |z-a'\sigma|^{2} + \frac{n}{2} dF_{t}(z) \right] = 0.
\]
Thus, the Liapounov condition is satisfied, implying
\[
d_n^{-\frac{1}{2}} (s_n - \sigma) \text{ converges in distribution to a } N(0,1) \text{ random variable.}
\]
Because \( \Omega_n \) converges to \( \Omega \),
\[
\lim n \frac{1}{2} n^{-\frac{1}{2}} (a's_n - a'\sigma) - \frac{1}{2} n^{-\frac{1}{2}} (a's_n - a'\sigma) = 0.
\]
Hence, for all nonnull \( a \), \( d_n^{-\frac{1}{2}} (a's_n - a'\sigma) \) has limiting
distribution \( N(0,1) \), which implies \( n^{-\frac{1}{2}} (s_n - \sigma) \) converges
in distribution to a normal vector random variable with
zero mean vector and covariance matrix \( \Omega \).
If the conditions of Lemma 5.1 are met, we can obtain
G.L.S. estimators, \( \tilde{\gamma} \), of the parameters of the linear
structural relationship. The properties of G.L.S. esti-
mators outlined in Chapter III apply to these G.L.S.
estimators. However, because \( \Omega^{-1}_n \) need not be of the
special form \( \Omega^{-1}_n = G'(V \otimes V)G \) of Chapter III, the results
of that chapter must be restated in notation appropriate for
arbitrary positive definite $\Omega$. A G.L.S. estimator, $\tilde{\gamma}$, is obtained by minimizing $g(\gamma|\Omega)$ with respect to $\gamma$, where

$$g(\gamma|\Omega) = [s - a(y)]'D[s - a(y)], \quad (5.2)$$

and where $D$, of dimension $q \times q$, is either a stochastic matrix which converges in probability to a positive definite matrix $\bar{D}$ as $n$ approaches infinity, or a positive definite constant matrix ($D = \bar{D}$).

Theorem 5.1: Let the assumptions of Lemma 5.1 hold. Let $\varepsilon_t$ be distributed independently of $x_t'$ for all $t = 1, 2, \ldots$, and $t' = 1, 2, \ldots$. Assume that $\gamma_0$ is identified. Define the $q \times r$ matrix $A$ by

$$A = \left. \left[ \frac{\partial g(\gamma)}{\partial \gamma} \right] \right|_{\gamma = \gamma_0}.$$

Suppose that $A$ has full column rank. A G.L.S. estimator, $\tilde{\gamma}$, is a consistent estimator of $\gamma_0$, and the limiting distribution of $\frac{1}{n} \tilde{\gamma} - \gamma_0$ is multivariate normal with mean zero and covariance matrix

$$(\Delta'D\Delta)^{-1}\Delta'D\Omega \Delta^\prime(\Delta'D\Delta)^{-1}.$$  

Proof: The proof is identical to the proofs of Theorems 4.2 and 4.3, except that $g(\gamma|\Omega)$ replaces $f(\gamma|\Omega)$.

To obtain B.G.L.S. estimates we need a consistent estimator of $\Omega^{-1}$. Lemma 5.2 shows how the observations $Z_t$, ...
t = 1, 2, ..., n can be used to estimate \( \Omega \), and hence \( \Omega^{-1} \), consistently. The assumption that the eighth moments of \( \xi_t \) and \( x_t \) exist is overly restrictive, but is used to simplify the presentation.

**Lemma 5.2:** Assume the model defined by (1.1-1.5) with the \( \xi_t \) and \( x_t \), \( t = 1, 2, ..., \) independent identically distributed \((p+k)\)-variate and \( k \)-variate random vectors, respectively, with means zero and bounded eighth moments.

Let \( \xi_t \) be distributed independently of \( x_t \), for all \( t = 1, 2, ... \), and \( t' = 1, 2, ... \). Suppose \( Z_{it} \), \( Z_{jt} \), \( Z_{kt} \), and \( Z_{mt} \) are (possibly nondistinct) elements of \( Z_t \), where \( Z_t \) is defined by (1.6). Let \( s_{ij} \) denote the following estimator of \( \text{Cov}(Z_{it}, Z_{jt}) \),

\[
s_{ij} = \frac{1}{n} \sum_{t=1}^{n} (Z_{it} - \bar{Z}_i)(Z_{jt} - \bar{Z}_j),
\]

where, for instance, \( \bar{Z}_i = \frac{1}{n} \sum_{t=1}^{n} Z_{it} \). Define \( \tilde{\omega}_{ij,lm} \) as

\[
\tilde{\omega}_{ij,lm} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \left( Z_{it} - \bar{Z}_i \right) \left( Z_{jt} - \bar{Z}_j \right) - s_{ij} \right\} x \left[ \left( Z_{kt} - \bar{Z}_k \right) \left( Z_{mt} - \bar{Z}_m \right) - s_{km} \right].
\]

Then,

\[
\tilde{\omega}_{ij,lm} = \text{Cov}(\frac{1}{n^2}s_{ij}, \frac{1}{n^2}s_{lm}) + O_p(\frac{1}{n^{3/2}}).
\]
Proof: We note that under the assumptions, the $\mathbf{Z}_t$, $t = 1, 2, \ldots$, are independent random vectors with bounded eighth moments. Also, observe that

$$(\mathbf{Z}_i - \mathbf{Z}_i^*) (\mathbf{Z}_j - \mathbf{Z}_j^*) = s_{ij} = \mathbf{Z}_i \mathbf{Z}_j - \mathbf{Z}_i \mathbf{Z}_j^*$$

$$- \mathbf{Z}_i^n (\mathbf{Z}_j - \mathbf{Z}_j^*) - \mathbf{Z}_j^n (\mathbf{Z}_i - \mathbf{Z}_i^*),$$

where $\mathbf{Z}_i \mathbf{Z}_j^* = \frac{1}{n} \sum_{t=1}^{n} \mathbf{Z}_i \mathbf{Z}_j^*$. We can express $\tilde{\omega}_{ij}$ in the form,

$$\tilde{\omega}_{ij}, \mathbf{Z}_m = -n^{-1} \sum_{t=1}^{n} \left[ \mathbf{Z}_i \mathbf{Z}_j - \mathbf{Z}_i \mathbf{Z}_j^* - \mathbf{Z}_i (\mathbf{Z}_j - \mathbf{Z}_j^*) - \mathbf{Z}_j (\mathbf{Z}_i - \mathbf{Z}_i^*) \right]$$

$$\times \left[ \mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^* - \mathbf{Z}_j \mathbf{Z}_m^* \right]$$

$$= n^{-1} \left\{ \sum_{t=1}^{n} (\mathbf{Z}_i \mathbf{Z}_j - \mathbf{Z}_i \mathbf{Z}_j^*) (\mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^*) \right. - \frac{1}{n} \sum_{t=1}^{n} (\mathbf{Z}_i (\mathbf{Z}_j - \mathbf{Z}_j^*) + \mathbf{Z}_j (\mathbf{Z}_i - \mathbf{Z}_i^*)) (\mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^*)$$

$$- \sum_{t=1}^{n} (\mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^*) (\mathbf{Z}_i \mathbf{Z}_j - \mathbf{Z}_i \mathbf{Z}_j^*)$$

$$+ \sum_{t=1}^{n} (\mathbf{Z}_j (\mathbf{Z}_i - \mathbf{Z}_i^*) + \mathbf{Z}_i (\mathbf{Z}_i - \mathbf{Z}_i^*))$$

$$\times \left. \left( \mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^* \right) \left( \mathbf{Z}_i \mathbf{Z}_m - \mathbf{Z}_i \mathbf{Z}_m^* \right) \right\}.$$ 

(5.5)

Since $E(\mathbf{Z}_i) = 0$ and $\text{Var}(\mathbf{Z}_i) = O(n^{-1})$, by Chebyshev's...
inequality, \( E_i = O_p(n^{-1/2}) \). Similarly \( \bar{Z_j}, \bar{Z_k}, \) and \( \bar{Z_m} \) are \( O_p(n^{1/2}) \). Also, for instance

\[
E[n^{-1} \sum_{t=1}^{n} (Z_{jt} - \bar{Z_j})(Z_{mt} - \bar{Z_m})] = n^{-1}(n-1)\text{Cov}(Z_{jt}, Z_{mt})
\]

(say)

\[
= n^{-1}(n-1)\text{Cov}(Z_j, Z_m),
\]

and

\[
\text{Var}[n^{-1} \sum_{t=1}^{n} (Z_{jt} - \bar{Z_j})(Z_{mt} - \bar{Z_m})] = O(n^{-1}).
\]

This implies \( n^{-1/2} \sum_{t=1}^{n} (Z_{jt} - \bar{Z_j})(Z_{mt} - \bar{Z_m}) \) is \( O_p(1) \). Thus, the last summation in (5.5) is \( O_p(n^{-1}) \).

Using the independence of \( Z_t, t = 1,2,\ldots \), and the fact that the sixth moments of \( Z_t \) are bounded, we can show that an expression such as \( n^{-1/2} \sum_{t=1}^{n} (Z_{jt} - \bar{Z_j})(Z_{lt} Z_{mt} - \bar{Z_j} \bar{Z_m}) \) is \( O_p(1) \). Therefore the middle two summations in (5.5) are \( O_p(n^{-1/2}) \), implying

\[
\tilde{\omega}_{ij,km} = n^{-1} \sum_{t=1}^{n} (Z_{it} Z_{jt} - \bar{Z_i} \bar{Z_j})(Z_{lt} Z_{mt} - \bar{Z_j} \bar{Z_m})
\]

\[
+ O_p(n^{-1/2}) \quad (5.6)
\]

Define \( u_t = Z_{it} Z_{jt} - E(Z_i Z_j) \) and \( v_t = Z_{lt} Z_{mt} - E(Z_j Z_m) \). Then \( u_t \) and \( v_t, t = 1,2,\ldots \), are independently and identically distributed with expectation zero and variances denoted by \( \text{Var}(Z_i Z_j) \) and \( \text{Var}(Z_j Z_m) \),
respectively. Let \( \bar{u} = n^{-1} \sum_{t=1}^{n} u_t = \bar{Z}_{i} \bar{Z}_{j} - \bar{E}(Z_{i} Z_{j}) \) and
\( \bar{v} = n^{-1} \sum_{t=1}^{n} v_t = \bar{Z}_{\kappa} \bar{Z}_{m} - \bar{E}(Z_{\kappa} Z_{m}) \), so that \( \bar{u} \) and \( \bar{v} \) have means zero and variances \( n^{-1} \text{Var}(Z_{i} Z_{j}) \) and \( n^{-1} \text{Var}(Z_{\kappa} Z_{m}) \). Therefore,

\[
E[n^{-1} \sum_{t=1}^{n} (Z_{it} Z_{jt} - \bar{Z}_{i} \bar{Z}_{j})(Z_{\kappa t} Z_{mt} - \bar{Z}_{\kappa} \bar{Z}_{m})] \\
= n^{-1}(n-1) \text{Cov}(Z_{i} Z_{j}, Z_{\kappa} Z_{m})
\]

and

\[
\text{Var}[n^{-1} \sum_{t=1}^{n} (Z_{it} Z_{jt} - \bar{Z}_{i} \bar{Z}_{j})(Z_{\kappa t} Z_{mt} - \bar{Z}_{\kappa} \bar{Z}_{m})] \\
= \text{Var}[n^{-1} \sum_{t=1}^{n} (u_t - \bar{u})(v_t - \bar{v})] \\
= \text{Var}(n^{-1} \sum_{t=1}^{n} u_t v_t - \bar{u} \bar{v}) \\
= n^{-1} \text{Var}(uv) + \text{Var}(\bar{u} \bar{v}) - 2 \text{Cov}[n^{-1} \sum_{t=1}^{n} u_t v_t, \bar{u} \bar{v}] \\
= 0(n^{-1}).
\]

Because the eighth moments of \( Z_{i} \) are bounded, the first term of (5.7) is \( O(n^{-1}) \). The second term can be shown to be \( O(n^{-2}) \) using Theorem 5.4.1 of Fuller (1976). The Cauchy-Schwarz inequality implies that the last term of expression (5.7) is at most \( O(n^{-1}) \). Hence,

\[
\text{Var}[n^{-1} \sum_{t=1}^{n} (Z_{it} Z_{jt} - \bar{Z}_{i} \bar{Z}_{j})(Z_{\kappa t} Z_{mt} - \bar{Z}_{\kappa} \bar{Z}_{m})] = O(n^{-1}),
\]
and,
\[ n^{-1} \sum_{t=1}^{n} (Z_{it}^2 - \overline{Z}_{i}^2)(Z_{\lambda t}^2 - \overline{Z}_{\lambda}^2) = \text{Cov}(Z_{i}^2, Z_{\lambda}^2) + O_p(n^{-1/2}). \]

Thus, from (5.6),
\[ \tilde{\omega}_{ij,\lambda m} = \text{Cov}(Z_{i}^2, Z_{\lambda}^2) + O_p(n^{-1/2}). \]  (5.8)

Now we look at the covariance of \( s_{ij} \) and \( s_{\lambda m} \).

\[ \text{Cov}(s_{ij}, s_{\lambda m}) = \text{Cov}[n^{-1} \sum_{t=1}^{n} Z_{it}Z_{jt} \]

\[ - n^{-1} \{ \sum_{t=1}^{n} Z_{\lambda t}Z_{\lambda t} - \overline{Z}_{\lambda}^2 \} \]

\[ = n^{-2} \text{Cov}(\sum_{t=1}^{n} Z_{it}Z_{jt}, \sum_{t=1}^{n} Z_{\lambda t}Z_{\lambda t}) \]

\[ - n^{-1} \text{Cov}(\overline{Z}_{i}^2, \sum_{t=1}^{n} Z_{\lambda t}Z_{\lambda t}) \]

\[ - n^{-1} \text{Cov}(\sum_{t=1}^{n} Z_{it}Z_{jt}, \overline{Z}_{\lambda}^2) \]

\[ + \text{Cov}(\overline{Z}_{i}^2, \overline{Z}_{\lambda}^2) \]. \]  (5.9)

We look at each term of (5.9) in turn. The first term is
\[ n^{-2} \text{Cov}(\sum_{t=1}^{n} Z_{it}Z_{jt}, \sum_{t=1}^{n} Z_{lt}Z_{mt}) \]

\[ = n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{ms}) \]

\[ = n^{-2} \left[ \sum_{t=1}^{n} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{mt}) \right] \]

\[ + \sum_{t \neq s} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{ms}) \]

\[ = n^{-2} \sum_{t=1}^{n} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{mt}) \]

\[ = n^{-1} \text{Cov}(Z_iZ_j, Z_{lt}Z_{mt}) \] (5.10)

The next two terms are of the same form. We only need to look at the first of these two terms,

\[ n^{-1} \text{Cov}(\overline{Z_iZ_j}, \sum_{t=1}^{n} Z_{lt}Z_{mt}) \]

\[ = n^{-3} \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{u=1}^{n} \text{Cov}(Z_{iu}Z_{js}, Z_{lt}Z_{mt}) \]

\[ = n^{-3} \left\{ \sum_{t=1}^{n} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{mt}) \right\} \]

\[ + \sum_{t \neq u} \text{Cov}(Z_{iu}Z_{ju}, Z_{lt}Z_{mt}) + \sum_{t \neq u} \text{Cov}(Z_{iu}Z_{jt}, Z_{lt}Z_{mt}) \]

\[ + \sum_{t \neq u} \text{Cov}(Z_{it}Z_{ju}, Z_{lt}Z_{mt}) + \sum_{t \neq s \neq u} \text{Cov}(Z_{iu}Z_{js}, Z_{lt}Z_{mt}) \]
\[ n^{-3} \sum_{t=1}^{n} \text{Cov}(Z_{it}Z_{jt}, Z_{lt}Z_{mt}) = n^{-2} \text{Cov}(Z_iZ_j, Z_lZ_m). \]

\[ = o(n^{-2}). \]  (5.11)

Also,

\[ n^{-1} \text{Cov} \left( \sum_{t=1}^{n} Z_{it}Z_{jt}, \bar{Z}_l\bar{Z}_m \right) = n^{-2} \text{Cov}(Z_iZ_j, Z_lZ_m). \]

\[ = o(n^{-2}). \]  (5.12)

The last term of expression (5.9) is at most \( o(n^{-2}) \), again using Theorem 5.4.1 of Fuller (1976). Thus, from (5.9), (5.10), and (5.11) we have

\[ \frac{1}{n^2} \text{Cov}(n^2s_{ij}, n^2s_{lm}) = \text{Cov}(Z_iZ_j, Z_lZ_m) + o(n^{-1}). \]  (5.13)

It follows from (5.8) and (5.13) that for all \( i, j, l, m \leq p+k \),

\[ \tilde{\omega}_{ij, lm} = \text{Cov}(n^2s_{ij}, n^2s_{lm}) + o_p(n^{-\frac{1}{2}}). \]

If the assumptions of Lemma 5.2 are satisfied, we can define a consistent estimator of \( \Omega \), say \( \tilde{\Omega} \), using the double subscript notation discussed in Chapter II to define the \((ij,lm)\)th element of \( \tilde{\Omega} \) by
Note that \( \hat{\Omega} \) is positive semidefinite. From Lemma 5.2, 
\[
\hat{\Omega} = \Omega + O_p \left( n^{-\frac{1}{2}} \right).
\]
Replacing \( D \) of expression (3.6) with \( \hat{\Omega}^{-1} \) we obtain
\[
g(\gamma|\hat{\Omega}^{-1}) = [s_n - \sigma(\gamma)]' \hat{\Omega}^{-1} [s_n - \sigma(\gamma)].
\] (5.15)

Any \( \gamma \) which minimizes \( g(\gamma|\hat{\Omega}^{-1}) \) is a B.G.L.S. estimator of \( \gamma_0 \). We denote this B.G.L.S. estimator by \( \hat{\gamma} \). All the properties of B.G.L.S. estimators proven in Chapter III which depend only upon the asymptotic normality of
\[
\frac{1}{n^2} (s_n - \bar{\gamma})
\]
apply to a B.G.L.S. estimator obtained from minimization of \( g(\gamma|\hat{\Omega}^{-1}) \). We summarize this discussion in the following theorem.

**Theorem 5.2:** Assume the model defined by (1.1-1.5) with the \( \varepsilon_t \) and \( x_t \), \( t = 1, 2, \ldots \), independent identically distributed \((p+k)\)-variate and \( k\)-variate random vectors, respectively, with means zero and bounded eighth moments. Let \( \varepsilon_t \) be distributed independently of \( x_t \) for all \( t = 1, 2, \ldots \), and \( t' = 1, 2, \ldots \). Assume \( \gamma_0 \) is identified and let \( \hat{\Delta} \) be as defined in Theorem 5.1. Let \( \hat{\Omega} \) be defined by (5.14) and suppose \( \hat{\gamma} \) is a B.G.L.S. estimator obtained from minimization of 
\[
g(\gamma|\hat{\Omega}^{-1}),
\]
where \( g(\gamma|\hat{\Omega}^{-1}) \) is defined by (5.15). Then \( \hat{\gamma} \).
is a G.L.S. estimator and among all G.L.S. estimators, \( \hat{\gamma} \) is asymptotically efficient. That is, the asymptotic variance of any G.L.S. estimator is bounded below in the Loewner sense of inequality by the asymptotic variance of \( \hat{\gamma} \), which is
\[
(\Lambda'\hat{\Omega}^{-1}\Lambda)^{-1}.
\]
(5.16)
Furthermore, the limiting distribution of \( g(\hat{\gamma}|\hat{\Omega}^{-1}) \) is chi-square with \( q-r \) degrees of freedom.

Proof: The proof is identical to the proofs of Corollary 4.3.1 and Theorem 4.4, with \( g(\gamma|\hat{\Omega}) \) replacing \( f(\gamma|D(\Omega)) \).

C. Estimation of the Functional Model when the \( x_t \), \( t = 1,2,... \) are Nonnormal

We state the analogue to Lemma 5.1 for the functional model.

Lemma 5.3: Assume the model defined by (1.1-1.5) with the \( \xi_t', \ t = 1,2,... \) independent identically distributed \((p+k)\)-variate random vectors with mean zero and bounded \((4+n)\)th moments. Further assume that the \((4+n)\)th moments of the vector sequence \( \xi_t', \ t = 1,2,... \) are bounded. That is, for \( 1 \leq i_1 \leq i_2 \leq k \), \( |n^{-1}\sum_{t=1}^{n} x_{i_1 t} x_{i_2 t}|^2 + \frac{n}{2} < L \) for all \( n \) for some real number \( L \). Then \( n^{2}(s_n-\varnothing) \) converges in distribution to a normal vector random variable with mean zero and covariance matrix \( \hat{\Omega} \).
Proof: The proof of Lemma 5.3 is essentially the same as the proof of Lemma 5.1, and is therefore omitted.

The properties of G.L.S. estimators for \( \gamma_0 \) of the functional model outlined in Chapter IV apply to G.L.S. estimators obtained under the assumptions of Lemma 5.3. A G.L.S. estimator, \( \tilde{\gamma} \), is defined to be any \( \gamma \) which minimizes \( f(\gamma|D) \), where \( f(\gamma|D) \) is defined by (4.14).

**Theorem 5.3:** Let the assumptions of Lemma 5.3 hold. Assume that \( \gamma_0 \) is identified. Define the \( q \times r \) matrix \( \Delta \) by

\[
\Delta = \left[ \frac{\partial \sigma(\gamma)}{\partial \gamma} \right]_{\gamma=\gamma_0}^{-1}.
\]

Suppose that \( \Delta \) has full column rank. A G.L.S. estimator, \( \tilde{\gamma} \), is a consistent estimator of \( \gamma_0 \), and the limiting distribution of \( n^{1/2}(\tilde{\gamma}-\gamma_0) \) is multivariate normal with mean zero and covariance matrix

\[
(\Delta'\Omega\Delta)^{-1} \Delta' \Omega \tilde{\Delta} \Delta (\Delta'\Omega\Delta)^{-1}.
\]

Proof: The proof is identical to the proofs of Theorems 4.2 and 4.3.

To obtain a B.G.L.S. estimator we need a consistent estimator of \( \Omega^{-1} \). The procedure suggested by Lemma 5.2 does not produce a consistent estimator of \( \Omega^{-1} \) for the
functional model. However, if we have available repeated observations, $Z_{ts} = (Y_{ts}, X_{ts})$, $t = 1, 2, \ldots, n$, $s = 1, 2, \ldots, r_0$ of the form (4.21) and (4.22), we can obtain a B.G.L.S. estimator of $\gamma_0$.

We define $S_n$, for the case of repeated observations, to be

$$S_n = (nr_0)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} Z_{ts}' Z_{ts}.$$  \hspace{1cm} (5.17)

For the special case of repeated observations, a more convenient definition of $S_n$ might be $S_n = S_n^*$, with

$$S_n^* = (n-1)^{-1} \sum_{t=1}^{n} (\bar{Z}_{t}. - \bar{Z})' (\bar{Z}_{t}. - \bar{Z}),$$

where $\bar{Z}_{t} = r_0^{-1} \sum_{s=1}^{r_0} Z_{ts}$ and $\bar{Z} = n^{-1} \sum_{t=1}^{n} \bar{Z}_{t}$. The advantage of defining $S_n = S_n^*$ would be that terms involving averages of elements of the $\mu_t$, $t = 1, 2, \ldots, n$ would not appear in expression (5.23). We proceed with $S_n$ defined by (5.17) to be consistent with the notation used in Chapter IV.

A typical element of $S_n = \text{vech } S_n$ is $s_{ij}$, where

$$s_{ij} = (nr_0)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} Z_{its} Z_{jts}, \quad j \leq i \leq p+k, \hspace{1cm} (5.18)$$

and $Z_{its}$ is the $i^{th}$ element of $Z_{ts}$. We may write $Z_{ts}$ as

$$Z_{ts} = \mu_t + \varepsilon_{ts}, \quad t = 1, 2, \ldots, n, \quad s = 1, 2, \ldots, r_0.$$  \hspace{1cm} (5.19)
where $\mathbf{u}_t = (x_t^\beta, x_t^\gamma)$ and $\mathbf{e}_{ts} = (e_{ts}, u_{ts})$. The expected value of $S_n$ is of the form (4.5),

$$
S_n(y_0) = M_n + \mathbf{Z}_\mathbf{e}_\mathbf{e},
$$

(5.20)

where $\mathbf{Z}_\mathbf{e}_\mathbf{e}$ is the covariance matrix of $\mathbf{e}_{ts}$ and

$$
M_n = (\beta, I_k \times k)' \mathbf{V}_n(\beta, I_k \times k)
$$

$$
= (\beta, I_k \times k)' (n^{-1} \sum_{t=1}^n x_t' x_t)(\beta, I_k \times k).
$$

If $\mathbf{V}_n = n^{-1} \sum_{t=1}^n x_t' x_t$ converges elementwise to a constant matrix $\mathbf{V}$, then $S_n(y_0)$ converges elementwise to $\mathbf{s}(\gamma_0) = M + \mathbf{Z}_e e$, where $M = (\beta, I_k \times k)' \mathbf{V}(\beta, I_k \times k)$. We define $\sigma_n = \sigma_n(\gamma_0)$ by

$$
\sigma_n(\gamma_0) = E(s_n),
$$

(5.21)

and, if $\mathbf{V}_n$ converges to $\mathbf{V}$, we define $\sigma = \sigma(\gamma_0)$ by

$$
\sigma(\gamma_0) = \lim_{n \to \infty} \sigma_n(\gamma_0)
$$

$$
= \text{vech} \mathbf{s}(\gamma_0).
$$

(5.22)

The definitions and properties of G.L.S. estimators of $\gamma_0$ summarized in Theorem 5.3 hold for the case of repeated observations. We shall now demonstrate how one might use repeated observations to obtain a B.G.L.S. estimator. First, we find $\mathbf{W}$, the covariance matrix of $s_n$. The $(ij,lm)$th element of $\mathbf{W}$ is given in Lemma 5.4.
Lemma 5.4: Assume the model defined by (5.19) with the $\xi_{ts}$, $t = 1, 2, \ldots, n$, $s = 1, 2, \ldots, r_0$, independently and identically distributed as $(p+k)$-variate random vectors with mean zero and bounded fourth moments. Let $Z_{its}$, $Z_{jts}$, $Z_{lts}$, and $Z_{mts}$ be (possibly nondistinct) elements of $Z_{ts}$, and let $s_{ij}$ and $s_{lm}$ be defined by (5.18). Then

$$
\text{Cov}[\frac{1}{n^{r_0}}s_{ij}, \frac{1}{n^{r_0}}s_{lm}] = \mu_j \mu_l \text{Cov}(\xi_i, \xi_m) + \mu_j \mu_m \text{Cov}(\xi_i, \xi_l) + \mu_i \mu_l \text{Cov}(\xi_j, \xi_m) + \mu_i \mu_m \text{Cov}(\xi_j, \xi_l) + \mu_l \text{E}(\xi_i \xi_j \xi_m) + \mu_m \text{E}(\xi_i \xi_j \xi_l) + \mu_j \text{E}(\xi_i \xi_j \xi_l)
$$

$$+ \mu_i \text{E}(\xi_j \xi_l \xi_m) + \text{Cov}(\xi_i, \xi_j, \xi_l, \xi_m), \quad (5.23)
$$

where, for instance, $\mu_j \mu_l = n^{-1} \sum_{t=1}^{n} \mu_j \mu_l \xi_{ts}$, $\text{Cov}(\xi_i, \xi_m) = \text{Cov}(\xi_{its}, \xi_{mts})$, $\mu_l = n^{-1} \sum_{t=1}^{n} \mu_l \xi_{ts}$, $\text{E}(\xi_i \xi_j \xi_l) = \text{E}(\xi_{its} \xi_{jts} \xi_{lts})$, and $\text{Cov}(\xi_i \xi_j \xi_l \xi_m) = \text{Cov}(\xi_{its} \xi_{jts} \xi_{lts} \xi_{mts})$.

Proof: Proceeding in a straightforward manner, we have
\[
\begin{align*}
\text{Cov}[(nr_0)^{\frac{1}{2}}s_{ij}, (nr_0)^{\frac{1}{2}}s_{lm}] & \\
& = (nr_0)^{-1} \text{Cov}(\sum_{t=1}^{r_0} \sum_{s=1}^{r_0} Z_{its} Z_{jts}', \sum_{t=1}^{r_0} \sum_{s=1}^{r_0} Z_{lts} Z_{mts}') \\
& = (nr_0)^{-1} \sum_{t=1}^{r_0} \sum_{s=1}^{r_0} \text{Cov}(Z_{its} Z_{jts}', Z_{lts} Z_{mts}') \\
& = (nr_0)^{-1} \sum_{t=1}^{r_0} \sum_{s=1}^{r_0} \text{Cov}[(\mu_{it} + \epsilon_{its}) (\mu_{jt} + \epsilon_{jts}), \\
& \quad (\mu_{lt} + \epsilon_{lts})(\mu_{mt} + \epsilon_{mts})] \\
& = (nr_0)^{-1} \sum_{t=1}^{r_0} \sum_{s=1}^{r_0} [\mu_{jt} \mu_{mt} \text{Cov}(\epsilon_i, \epsilon_m) \\
& \quad + \mu_{jt} \mu_{mt} \text{Cov}(\epsilon_j, \epsilon_m) + \mu_{it} \mu_{mt} \text{Cov}(\epsilon_i, \epsilon_m)] \\
& = \mu_{ij} \mu_{kl} \text{Cov}(\epsilon_i, \epsilon_m) + \mu_{ij} \mu_{lm} \text{Cov}(\epsilon_i, \epsilon_m) \\
& \quad + \mu_{ij} \mu_{ml} \text{Cov}(\epsilon_j, \epsilon_m) \\
& \quad + \mu_{ij} \mu_{lm} \text{Cov}(\epsilon_j, \epsilon_m) + \mu_{il} \mu_{mj} \text{Cov}(\epsilon_i, \epsilon_m) + \mu_{il} \mu_{mj} \text{Cov}(\epsilon_j, \epsilon_m) \\
& \quad + \mu_{il} \mu_{mj} \text{Cov}(\epsilon_i, \epsilon_m) + \mu_{il} \mu_{mj} \text{Cov}(\epsilon_j, \epsilon_m) + \mu_{il} \mu_{mj} \text{Cov}(\epsilon_i, \epsilon_m).
\end{align*}
\]
Because we have repeated observations, we can estimate each term of (5.23) consistently.

Lemma 5.5: Assume the model defined by (5.19) with the \( \xi_{ts} \), \( t = 1, 2, \ldots, \ s = 1, 2, \ldots, r_0 \), independently and identically distributed as \( (p+k) \)-variate random vectors with mean zero and bounded eighth moments. Also assume that \( n^{-1} \sum_{t=1}^{n} x_t^t x_t \) converges to a constant matrix \( \Sigma \). Define

\[
(\text{i}) \quad \bar{\mu}_i = (nr_0)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} Z_{its} \\
(\text{ii}) \quad \sqrt{\mu_i \mu_k} = (nr_0)^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} Z_{its} Z_{kts} - \text{Cov}(\xi_i, \xi_k) \\
(\text{iii}) \quad \text{Cov}(\xi_i, \xi_k) = [n(r_0-1)]^{-1} \sum_{t=1}^{n} \sum_{s=1}^{r_0} (Z_{its} - \bar{Z}_{it})(Z_{kts} - \bar{Z}_{kt}) \\
(\text{iv}) \quad \tilde{E}(\xi_i \xi_k) = \frac{r_0}{n(r_0-1)(r_0-2)} \sum_{t=1}^{n} \sum_{s=1}^{r_0} (Z_{its} - \bar{Z}_{it})(Z_{kts} - \bar{Z}_{kt}) \\
(\text{v}) \quad \text{Cov}(\xi_i \xi_j, \xi_k \xi_m) = \frac{r_0}{n(r_0-1)(r_0^2-3r_0+3)} \sum_{t=1}^{n} \sum_{s=1}^{r_0} (Z_{its} - \bar{Z}_{it})(Z_{jts} - \bar{Z}_{jt})(Z_{kts} - \bar{Z}_{kt})(Z_{mts} - \bar{Z}_{mt})
Then,

\( (i) \quad \bar{\mu}_i = \bar{\mu}_i + O_p(n^{-\frac{1}{2}}), \)

\( (ii) \quad \frac{\bar{\mu}_i \bar{\mu}_j}{\sqrt{\mu_i \mu_j}} = \frac{\mu_i \mu_j}{\sqrt{\mu_i \mu_j}} + O_p(n^{-\frac{1}{2}}), \)

\( (iii) \quad \text{Cov}(\epsilon_i, \epsilon_j) = \text{Cov}(\epsilon_i, \epsilon_j) + O_p(n^{-\frac{1}{2}}), \)

\( (iv) \quad \tilde{E}(\epsilon_i \epsilon_j \epsilon_m) = E(\epsilon_i \epsilon_j \epsilon_m) + O_p(n^{-\frac{1}{2}}), \)

\( (v) \quad \text{Cov}(\epsilon_i \epsilon_j, \epsilon_j \epsilon_m) = \text{Cov}(\epsilon_i \epsilon_j, \epsilon_j \epsilon_m) + O_p(n^{-\frac{1}{2}}). \)

\( (5.25) \)

**Proof:** The proof is straightforward, but tedious. Therefore, we omit the proof.

**Lemma 5.6:** Let the assumptions of Lemma 5.5 hold and let \( s_{ij} \) be defined by (5.18). Then a consistent estimator of \( \frac{1}{2} \text{Cov}(n^{\frac{1}{2}}s_{ij}, n^{\frac{1}{2}}s_{lm}) \) is
\[
\begin{align*}
\text{Cov}(n^2 s_{ij}, n^2 s_{\lambda m}) &= \frac{1}{2} \text{Cov}(\varepsilon_i, \varepsilon_m) + \frac{1}{2} \text{Cov}(\varepsilon_i, \varepsilon_{\lambda}) \\
&+ \frac{1}{2} \text{Cov}(\varepsilon_j, \varepsilon_m) + \frac{1}{2} \text{Cov}(\varepsilon_j, \varepsilon_{\lambda}) \\
&+ \frac{1}{2} \text{Cov}(\varepsilon_i \varepsilon_j, \varepsilon_m) + \frac{1}{2} \text{Cov}(\varepsilon_i \varepsilon_j, \varepsilon_{\lambda}) \\
&+ \frac{1}{2} \text{Cov}(\varepsilon_j \varepsilon_{\lambda}, \varepsilon_m) + \frac{1}{2} \text{Cov}(\varepsilon_j \varepsilon_{\lambda}, \varepsilon_{\lambda m}).
\end{align*}
\]

**Proof:** The proof follows immediately from Lemma 5.5, since each term on the right-hand side of expression (5.26) is a consistent estimator of the corresponding term on the right-hand side of expression (5.23).

The \((ij, \lambda m)\)th element of \(\text{Cov}(s_n) = \Omega\) is

\[
\Omega_{ij, \lambda m} = \text{Cov}(n^2 s_{ij}, n^2 s_{\lambda m}), \quad j \leq i \leq p+k, \quad m \leq \lambda \leq p+k.
\]

(5.27)

A consistent estimator of \(\Omega\), say \(\tilde{\Omega}\), is defined by

\[
\tilde{\Omega}_{ij, \lambda m} = \text{Cov}(n^2 s_{ij}, n^2 s_{\lambda m}), \quad j \leq i \leq p+k, \quad m \leq \lambda \leq p+k.
\]

(5.28)

Replacing \(D(\Omega)\) in expression (4.8) with \(\tilde{\Omega}^{-1}\), we obtain

\[
f(\gamma | \tilde{\Omega}^{-1}) = \left[ s_n - \bar{\sigma}_n(\gamma) \right] \tilde{\Omega}^{-1} \left[ s_n - \bar{\sigma}_n(\gamma) \right],
\]

(5.29)

where now \(s_n\) is a vector containing the distinct elements of \(s_n = (n r_0)^{-1} \sum_{t=1}^{r_0} \sum_{s=1}^{r} z_t z_s\), and \(\bar{\sigma}_n(\gamma_0) = E(s_n)\). Any \(\gamma\) which minimizes \(f(\gamma | \tilde{\Omega}^{-1})\) is a B.G.L.S. estimator of \(\gamma_0\). We
denote this B.G.L.S. estimator by $\hat{y}$. Any property of B.G.L.S. estimators proven in Chapter IV which depends only upon the asymptotic normality of $\frac{1}{n^2}(s_n - \sigma_n)$ applies to a B.G.L.S. estimator obtained from minimization of $f(\gamma | \tilde{\Sigma}^{-1})$. Theorem 5.4 summarizes these properties.

**Theorem 5.4:** Let the assumptions of Lemma 5.5 hold, and let $s_n$, $\sigma_n(\gamma)$, and $\tilde{\Sigma}$ be defined by (5.15), (5.21), and (5.28), respectively. In addition, let the \((4+n)\)th moments of the vector sequence $x_t, t = 1, 2, \ldots$ be bounded. Assume $\tilde{\Sigma}$ is identified and let $\tilde{\Delta}$, as defined in Theorem 5.3, be of full column rank. The asymptotic variance of any G.L.S. estimator is bounded below in the Loewner sense of inequality by the asymptotic variance of $\hat{y}$, which is

$$(\tilde{\Delta}'\tilde{\Omega}^{-1}\tilde{\Delta})^{-1}. \quad (5.30)$$

Furthermore, the limiting distribution of $f(\hat{y} | \tilde{\Sigma}^{-1})$ is chi-square with $q-r$ degrees of freedom.

**Proof:** Because the \((4+n)\)th moments of $x_t$ and $\varepsilon_t$ exist, we can conclude from Lemma 5.3 that $\frac{1}{n^2}(s_n - \sigma)$ is asymptotically normal with mean zero and covariance matrix $\tilde{\Sigma}$. From Lemma 5.6 we know that $\tilde{\Sigma}^{-1}$ is a consistent estimator of $\tilde{\Sigma}^{-1}$. Hence, the conditions of Corollary 4.3.1 and Theorem 4.4 are satisfied. The results follow.
We conclude this dissertation by presenting two examples which illustrate the application of the G.L.S. approach to the estimation of errors in variables models.
VI. EXAMPLES

We illustrate the application of the G.L.S. principle by presenting two examples. The first example arose from an experiment conducted by agronomists at Iowa State University to determine the effect of soil moisture on corn yields. The second example uses data from an experiment conducted at the U.S.D.A. Beef Cattle Research Station in Fort Robinson, Nebraska. One purpose of this second experiment was to estimate the effects of genotype and environment on phenotype in beef cattle.

A. An Internal Estimate of Measurement Error in Available Soil Moisture

An experiment to measure the effect of soil moisture on corn yield was initiated in the fall of 1956 at the Moody Experimental Farm near Doon, Iowa. A detailed description of this experiment is found in Mowers, et al. (Forthcoming). We briefly describe the experiment.

The experiment was designed as a long term fixed rotation experiment using the rotation corn-oats-meadow-meadow. The experimental area was divided into four fields. In a given year each field carried one of the crops in the rotation. To provide replicates, each field was subdivided into three blocks which in turn were subdivided into four individual plots. Treatments were randomized within the
blocks at the start of the experiment, but because of the fixed rotation, were not rerandomized in subsequent years.

Treatments were applied to the individual plots during the second year meadow as follows.

Treatment 1 (control): Meadow cut two or three times for hay, land plowed in spring before planting corn.

Treatment 2 (short fallow): Meadow killed in early fall at a height of 6 to 8 inches following second cutting of hay, land plowed in spring before planting corn.

Treatment 3 (long fallow): Meadow killed in midsummer at a height of 6 to 8 inches following first cutting of hay, land plowed in spring before planting corn.

Soil moisture measurements were taken in the spring prior to planting of the corn plots. Two sample cores of soil were taken from each plot to a depth of five feet. Samples were taken from each core in one foot increments to determine available soil moisture. An indication of the relationship of available stored moisture to corn yield is given in Table 1. This table presents the yields obtained in this experiment for several ranges of available moisture in the soil profile prior to planting. Average corn yields increased with increased spring soil moisture.

The effect of stored soil moisture on corn yields varies with precipitation levels during the growing season.
Table 1. Observed corn yields for ranges of stored soil moisture, Moody soil experiment, 1958-1977

<table>
<thead>
<tr>
<th>Spring soil moisture (inches)</th>
<th>Number of observations</th>
<th>Average yield (bushels/acre)</th>
<th>Range of yield (bushels/acre)</th>
<th>Yields less than 40 bushels/acre (%)</th>
<th>Yields greater than 80 bushels/acre (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3.0</td>
<td>73</td>
<td>39</td>
<td>0-111</td>
<td>51</td>
<td>15</td>
</tr>
<tr>
<td>3.0-5.0</td>
<td>50</td>
<td>60</td>
<td>1-124</td>
<td>32</td>
<td>36</td>
</tr>
<tr>
<td>5.0-7.0</td>
<td>35</td>
<td>110</td>
<td>54-162&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3</td>
<td>66</td>
</tr>
<tr>
<td>Over 7.0</td>
<td>22</td>
<td>119</td>
<td>76-163</td>
<td>0</td>
<td>91</td>
</tr>
</tbody>
</table>

<sup>a</sup>One observation with no corn yield and 5.2 inches stored soil moisture was recorded for 1970. This observation was not included in the range because there was a crop failure in 1970.
A weather index was created using a preliminary regression equation which included the total rainfall in the four months of May, June, July, and August as variables. The weather index is a weighted average with weights in the proportions $\frac{1}{2}:1:2:1$ for May, June, July, and August, respectively.

The data for our analysis consists of 180 observations on corn yield and observed soil moisture, and 20 observations on the yearly weather index. Observations have been gathered over the period 1958-1977 with three replications for each of the three treatments.

A model for corn yield is

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_1 x_{ijk} + \gamma_2 w_j + \epsilon_{ijk},$$

$$i = 1,2,3; \quad j = 1,2,...,20; \quad k = 1,2,3,$$  \hspace{1cm} (6.1)

where

- $y_{ijk}$ = corn yield for the $i^{th}$ treatment, $j^{th}$ year, and $k^{th}$ replication,
- $\mu$ = intercept,
- $\alpha_i$ = $i^{th}$ treatment effect,
- $\beta_j$ = $j^{th}$ year effect,
- $(\alpha\beta)_{ij}$ = $(i,j)^{th}$ year by treatment interaction effect,
- $x_{ijk}$ = amount of available soil moisture prior to planting for the $(i,j,k)^{th}$ plot,
- $\gamma_1$ = linear regression coefficient for $x_{ijk}$,
\[ W_j = \text{weather index computed for the } j^{th} \text{ year}, \]
\[ \gamma_2 = \text{linear regression coefficient for } W_j, \]
\[ \epsilon_{ijk} = \text{error for the (i,j,k)th plot}. \]

The intercept and treatment effects are regarded as fixed, while the year effects, year by treatment interaction effects, available soil moistures, weather indices, and plot errors are considered to be random samples drawn from normal populations with means and variances given by \((0, \sigma^2_\beta), (0, \sigma^2_\alpha), (\mu_x, \sigma^2_x), (\mu_W, \sigma^2_W), \text{ and } (0, \sigma^2_\epsilon),\) respectively.

If we could observe the true amount of stored soil moisture, \(x_{ijk},\) ordinary least squares would yield a best linear unbiased estimator for all estimable functions of the unknown parameters. However, we are not able to observe the true available moistures. We actually observe the random variable \(X_{ijk}\) composed of the true stored moisture, \(x_{ijk},\) masked by measurement error, \(u_{ijk}.\) A model for the observed available soil moisture is

\[ X_{ijk} = u^* + \alpha_i^* + \beta_j^* + (\alpha \beta)^*_{ij} + z_{ijk} + u_{ijk} \]
\[ = x_{ijk} + u_{ijk}, \quad i = 1, 2, 3; \quad j = 1, 2, \ldots, 20; \]
\[ k = 1, 2, 3, \quad (6.2) \]

where

\[ u^* = \text{intercept}, \]
\[ \alpha_i^* = i^{th} \text{ treatment effect}, \]
\( \beta_j \) = \( j \)th year effect,
\((\alpha \beta)_ij\) = \((i,j)\)th year by treatment interaction effect,
\(z_{ijk}\) = plot error,
\(x_{ijk} = u^* + \alpha_i^* + \beta_j^* + (\alpha \beta)_ij^* + z_{ijk}\)
= true available moisture,
\(u_{ijk}\) = measurement error.

The intercept and treatment effects are regarded as fixed, while the year effects, year by treatment interaction effects, plot errors, and measurement errors are considered to be random samples drawn from normal populations with means and variances \((0, \sigma^2_{\beta^*}), (0, \sigma^2_{\alpha \beta^*}), (0, \sigma^2_z),\) and \((0, \sigma^2_u)\), respectively. We allow \(\text{Cov}(x_{ijk}, W_j)\) to be nonzero, but assume that \(\beta_j^*, (\alpha \beta)_ij^*, \beta_j^*, (\alpha \beta)_ij^*, \epsilon_{ijk}, z_{ijk},\) and \(u_{ijk}\) are pairwise independent. In addition, we assume that \(W_j\) is independent of \(\beta_j^*, (\alpha \beta)_ij^*, \epsilon_{ijk},\) and \(u_{ijk}\). We shall call equations (6.1) and (6.2), along with the associated assumptions, Model I. Model I is a univariate linear errors in variables model.

A reparameterization of Model I will be helpful in evaluating expected mean squares. We define Model II to be

\( Y_{ijk} = m + a_i + b_j + (ab)_{ij} + e_{ijk'}\)
\(X_{ijk} = u^* + \alpha_i^* + \beta_j^* + (\alpha \beta)_ij^* + v_{ijk'}\)
\(i = 1,2,3; \ j = 1,2,...,20; \ k = 1,2,3, \)  \(6.4\)
where
\[ m = \mu + \gamma_1 \mu^*, \]
\[ a_i = \alpha_i + \gamma_1 \alpha_i^*, \]
\[ b_j = \beta_j + \gamma_1 \beta_j^* + \gamma_2 W_j \]
\[ (ab)_{ij} = (\alpha \beta)_{ij} + \gamma_1 (\alpha \beta)_{ij}^* \]
\[ \epsilon_{ijk} = \gamma_1 z_{ijk} + \epsilon_{ijk}' \]
\[ v_{ijk} = z_{ijk} + u_{ijk}. \]

We denote the variances of \( b_j, (ab)_{ij}, \epsilon_{ijk}, \) and \( v_{ijk} \) by \( \sigma_b^2, \sigma_{ab}^2, \sigma_e^2, \) and \( \sigma_v^2, \) respectively. The form of Model II is that of the standard 2-way classification mixed model, except that \( b_j \) contains \( W_j \) which is correlated with \( (\alpha \beta)_{ij}^* \) and \( z_{ijk}. \) Hence, \( b_j \) is correlated with \((ab)_{ij}\) and \( \epsilon_{ijk}. \) However, \((ab)_{ij}\) and \( \epsilon_{ijk} \) are distributed independently of each other, so that the expected mean squares of the error and year by treatment interaction effects for corn yield under Model II are those obtained for the standard 2-way classification mixed model. The analysis of variance for the standard mixed model of the form (6.4) is given in [Searle (1971), Chapter 9]. Tables 2 and 3 give the AOV for the corn yield and available soil moisture, respectively. Define
\[ s_{11} = (120)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (Y_{ijk} - \overline{Y}_{ij}).^2, \]
\[ s_{22} = (120)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (X_{ijk} - \overline{X}_{ij}).^2. \]
and

\[ s_{12} = (120)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (Y_{ijk} - \bar{Y}_{ij}) (X_{ijk} - \bar{X}_{ij}), \quad (6.5) \]

where \( \bar{Y}_{ij} = (3)^{-1} \sum_{k=1}^{3} Y_{ijk} \) and \( \bar{X}_{ij} = (3)^{-1} \sum_{k=1}^{3} X_{ijk} \).

From Tables 2 and 3 we see that \( s_{11} = 44.2919 \) and \( s_{22} = 0.6594 \). The expected values of \( s_{11} \) and \( s_{22} \) are the expected mean squares of the error source in the 2-way classification mixed model. Thus,

\[ E(s_{11}) = \sigma_e^2 \]
\[ = \gamma_1^2 \sigma_z^2 + \sigma_e^2, \quad (6.6) \]

and

\[ E(s_{22}) = \sigma_v^2 \]
\[ = \sigma_z^2 + \sigma_u^2. \quad (6.7) \]

Straightforward evaluation of \( s_{12} \) and its expectation gives us that \( s_{12} = 1.9350 \) and

\[ E(s_{12}) = \gamma_1 \sigma_z^2. \quad (6.8) \]

The parameters of major interest, \( \gamma_1 \) and \( \sigma_u^2 \), cannot be estimated from \( (s_{11}, s_{12}, s_{22}) \) using the G.L.S. procedure because the parameters \( \gamma_1, \sigma_z^2, \sigma_e^2, \) and \( \sigma_u^2 \) are not identified by Equations (6.6-6.8). Additional information is required to identify \( \gamma_1 \) and \( \sigma_u^2 \). Define
Table 2. Corn yield analysis of variance table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>19</td>
<td>( \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{Y}<em>{ijk} - \bar{Y}</em>{...})^2 ) = 391,093.75</td>
<td>20,583.88</td>
<td></td>
</tr>
<tr>
<td>Tmt</td>
<td>2</td>
<td>( \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{Y}<em>{ij} - \bar{Y}</em>{...})^2 ) = 3,866.62</td>
<td>1,933.31</td>
<td>43.65**</td>
</tr>
<tr>
<td>Year x Tmt</td>
<td>38</td>
<td>( \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{Y}<em>{ij} - \bar{Y}</em>{...})^2 ) = 7,684.18</td>
<td>202.21</td>
<td>4.57**</td>
</tr>
<tr>
<td>Error</td>
<td>120</td>
<td>( \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (Y_{ijk} - \bar{Y}_{ij})^2 ) = 5,315.03</td>
<td>44.29</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>179</td>
<td>( \sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (Y_{ijk} - \bar{Y}_{...})^2 ) = 407,959.57</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Significant at 0.01 level.
### Table 3. Available moisture at planting analysis of variance table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>19</td>
<td>$\sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{X}<em>{ikj} - \bar{X}</em>{..})^2 = 779.4982$</td>
<td>41.0262</td>
<td></td>
</tr>
<tr>
<td>Tmt</td>
<td>2</td>
<td>$\sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{X}<em>{i} - \bar{X}</em>{..})^2 = 48.7904$</td>
<td>24.3953</td>
<td>36.99**</td>
</tr>
<tr>
<td>Year x Tmt</td>
<td>38</td>
<td>$\sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (\bar{X}<em>{ij} - \bar{X}</em>{i..} - \bar{X}<em>{.j} + \bar{X}</em>{..})^2 = 51.6506$</td>
<td>1.3592</td>
<td>2.06**</td>
</tr>
<tr>
<td>Error</td>
<td>120</td>
<td>$\sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (X_{ijk} - \bar{X}_{ij})^2 = 79.1306$</td>
<td>0.6594</td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>179</td>
<td>$\sum_{i=1}^{3} \sum_{j=1}^{20} \sum_{k=1}^{3} (X_{ijk} - \bar{X}_{..})^2 = 959.0698$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** Significant at 0.01 level.
\[ s_{33} = (38)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} (\overline{Y}_{ij} - \overline{Y}_{i..} - \overline{Y}_{..j} + \overline{Y}_{..})^2, \]

\[ s_{44} = (38)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} (\overline{X}_{ij} - \overline{X}_{i..} - \overline{X}_{..j} + \overline{X}_{..})^2, \]

and

\[ s_{34} = (38)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} (\overline{Y}_{ij} - \overline{Y}_{i..} - \overline{Y}_{..j} + \overline{Y}_{..})(\overline{X}_{ij} - \overline{X}_{i..} - \overline{X}_{..j} + \overline{X}_{..}), \quad (6.9) \]

where

\[ \overline{Y}_{i..} = (20)^{-1} \sum_{j=1}^{20} \overline{Y}_{ij}, \quad \overline{Y}_{..j} = (3)^{-1} \sum_{i=1}^{3} \overline{Y}_{ij}. \]

\[ \overline{Y}_{..} = (3)^{-1} \sum_{i=1}^{3} \sum_{j=1}^{20} \overline{Y}_{i..}, \quad \overline{X}_{i..} = (20)^{-1} \sum_{j=1}^{20} \overline{X}_{ij}, \quad \overline{X}_{..j} = (3)^{-1} \sum_{i=1}^{3} \overline{X}_{ij}. \]

and

\[ \overline{X}_{..} = (3)^{-1} \sum_{i=1}^{3} \overline{X}_{i..}, \quad \overline{X}_{..} = (20)^{-1} \sum_{j=1}^{20} \overline{X}_{..j}. \]

From Tables 2 and 3 we see that \( s_{33} = (3)^{-1}(202.2150) = 67.4050 \) and \( s_{44} = (3)^{-1}(1.3592) = 0.4531 \). The expected values of \( s_{33} \) and \( s_{44} \) are one-third times the expected mean squares of the interaction source in the 2-way classification mixed model. Thus,
\[ E(s_{33}) = (3)^{-1} \left( 3\sigma_{ab}^2 + \sigma_e^2 \right) \]
\[ = \sigma_{ab}^2 + \gamma_1 \sigma_{ab}^2 + (3)^{-1} \left( \gamma_1 \sigma_z^2 + \sigma_e^2 \right) \]
\[ = \gamma_1 \sigma_{ab}^2 + \sigma_e^2 \quad (6.10) \]

and

\[ E(s_{44}) = (3)^{-1} \left( 3\sigma_{ab}^2 + \sigma_v^2 \right) \]
\[ = \sigma_{ab}^2 + (3)^{-1} \left( \sigma_z^2 + \sigma_u^2 \right) \]
\[ = \sigma_{ab}^2 + \sigma_v^2 \quad (6.11) \]

where \( \sigma_p^2 = \sigma_{ab}^2 + 3^{-1} \sigma_z^2 \) and \( \sigma_q^2 = \sigma_{ab}^2 + 3^{-1} \sigma_e^2 \). Straightforward evaluation of \( s_{34} \) and its expectation gives us that \( s_{34} = 3.1639 \) and

\[ E(s_{34}) = \gamma_1 \sigma_p^2. \quad (6.12) \]

The expectations of the six sample covariances \( s_{11} \), \( s_{12} \), \( s_{22} \), \( s_{33} \), \( s_{34} \), and \( s_{44} \) are functions of the six parameters \( \gamma_1, \sigma_z, \sigma_e, \sigma_u, \sigma_p, \) and \( \sigma_q \). Application of the G.L.S. procedure at this point would yield consistent estimates of the parameters. However, we have not yet exploited the information contained in the weather index and in the year mean squares. Define

\[ s_{55} = (19)^{-1} \sum_{j=1}^{20} (\bar{Y}_j - \bar{Y})^2, \]
\[ s_{66} = (19)^{-1} \sum_{j=1}^{20} (\bar{X}_{j} - \bar{X})^2, \]
\[ s_{77} = (19)^{-1} \sum_{j=1}^{20} (\bar{W}_{j} - \bar{W})^2, \]
\[ s_{56} = (19)^{-1} \sum_{j=1}^{20} (\bar{X}_{j} - \bar{X})(\bar{Y}_{j} - \bar{Y}), \]
\[ s_{57} = (19)^{-1} \sum_{j=1}^{20} (\bar{X}_{j} - \bar{X})(\bar{W}_{j} - \bar{W}), \]
and
\[ s_{67} = (19)^{-1} \sum_{j=1}^{20} (\bar{X}_{j} - \bar{X})(\bar{W}_{j} - \bar{W}), \] (6.13)
where \( \bar{W} = (20)^{-1} \sum_{j=1}^{20} W_{j} \). The values of the sample covariances of (6.13) are
\[ s_{55} = 2287.0980, \]
\[ s_{66} = 4.5585, \]
\[ s_{77} = 14.9380, \]
\[ s_{56} = 67.3160, \]
\[ s_{57} = 147.5393, \]
and,
\[ s_{67} = 1.8343. \]
Direct evaluation of the expectations of Equation (6.13) gives
\[ E(s_{55}) = \gamma_{1}^{2} \sigma_{\bar{X}}^{2} + \gamma_{2}^{2} \sigma_{\bar{W}}^{2} + 2\gamma_{1}\gamma_{2} \sigma_{\bar{X} \bar{W}} + \sigma_{S}^{2}, \] (6.14)
\[ E(s_{66}) = \sigma_{\bar{X}}^{2} + 9^{-1} \sigma_{\bar{U}}^{2}, \] (6.15)
\[ E(s_{77}) = \sigma_{\bar{W}}^{2}. \] (6.16)
\[ E(s_{56}) = \gamma_1 \sigma_r^2 + \gamma_2 \sigma_{xW}^2, \]  
(6.17)

\[ E(s_{57}) = \gamma_1 \sigma_{xW}^2 + \gamma_2 \sigma_{W}^2, \]  
(6.18)

and

\[ E(s_{67}) = \sigma_{xW}' \]  
(6.19)

where \( \sigma_r^2 = \sigma_{x}^2 + 3^{-1} \sigma_{2x} + 9^{-1} \sigma_{z}^2 \), \( \sigma_{xW} = \text{Cov}(x_j, W_j) \), and \( \sigma_s^2 = \sigma_{x}^2 + 3^{-1} \sigma_{2x} + 9^{-1} \sigma_{\varepsilon}^2 \). The set of unknown parameters we shall estimate using the G.L.S. procedure is

\[ \gamma_0 = (\gamma_1, \gamma_2, \sigma_r^2, \sigma_p^2, \sigma_z^2, \sigma_s^2, \sigma_q^2, \sigma_r^2, \sigma_{xW}^2, \sigma_w^2, \sigma_u^2)' . \]  
(6.20)

Define \( s_n \) to be

\[ s_n = (s_{11}', s_{12}', s_{22}', s_{33}', s_{34}', s_{44}', s_{55}', s_{56}', s_{57}', s_{66}', s_{67}', s_{77}') , \]  
(6.21)

and

\[ \sigma(\gamma_0) = E(s_n)' . \]  
(6.22)

Table 4 lists the elements of \( s_n \) and \( \varrho = \sigma(\gamma_0) \). To obtain a B.G.L.S. estimate, \( \hat{\gamma} \), of \( \gamma_0 \) we require a consistent estimator of \( \Omega = \text{Var}(s_n) \). Let \( s_1, s_2, \) and \( s_3 \) be defined by

\[ s_1 = \begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix} , \quad s_2 = \begin{bmatrix} s_{33} & s_{34} \\ s_{34} & s_{44} \end{bmatrix} , \quad \text{and} \quad s_3 = \begin{bmatrix} s_{55} & s_{56} & s_{57} \\ s_{56} & s_{66} & s_{67} \\ s_{57} & s_{67} & s_{77} \end{bmatrix} . \]  
(6.23)
Table 4. Sample covariance estimates and expected values for the Moody soil moisture experiment

<table>
<thead>
<tr>
<th>Estimate ($s_{ij}$)</th>
<th>Expected value ($\gamma_{ij}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{11} = 44.291$</td>
<td>$\gamma_1\sigma_z^2 + \sigma_c^2$</td>
</tr>
<tr>
<td>$s_{12} = 1.935$</td>
<td>$\gamma_1\sigma_z^2$</td>
</tr>
<tr>
<td>$s_{22} = 0.659$</td>
<td>$\sigma_z^2 + \sigma_u^2$</td>
</tr>
<tr>
<td>$s_{33} = 67.405$</td>
<td>$\gamma_1\sigma_p^2 + \sigma_q^2$</td>
</tr>
<tr>
<td>$s_{34} = 3.164$</td>
<td>$\gamma_1\sigma_p^2$</td>
</tr>
<tr>
<td>$s_{44} = 0.453$</td>
<td>$\sigma_p^2 + 3\gamma_0^2\sigma_u^2$</td>
</tr>
<tr>
<td>$s_{55} = 2287.098$</td>
<td>$\gamma_1\sigma_r^2 + \gamma_2\sigma_w^2 + 2\gamma_1\sigma_r\sigma_w + \sigma_s^2$</td>
</tr>
<tr>
<td>$s_{56} = 67.316$</td>
<td>$\gamma_1\sigma_r^2 + \gamma_2\sigma_w^2$</td>
</tr>
<tr>
<td>$s_{57} = 147.539$</td>
<td>$\gamma_1\sigma_w^2 + \gamma_2\sigma_w^2$</td>
</tr>
<tr>
<td>$s_{66} = 4.558$</td>
<td>$\sigma_r^2 + 9\gamma_0^2\sigma_u^2$</td>
</tr>
<tr>
<td>$s_{67} = 1.834$</td>
<td>$\sigma_w^2$</td>
</tr>
<tr>
<td>$s_{77} = 14.938$</td>
<td>$\sigma_w^2$</td>
</tr>
</tbody>
</table>

The matrices $\Sigma_1$, $\Sigma_2$, and $\Sigma_3$ are sample covariance matrices which are distributed as multiples of Wishart matrices. Furthermore, $\Sigma_1$, $\Sigma_2$, and $\Sigma_3$ are distributed independently of each other. Thus, with the data configuration of Table 4, $\Sigma$ is block diagonal,

$$
\Sigma = \begin{bmatrix}
\Sigma_1 & 0 & 0 \\
0 & \Sigma_2 & 0 \\
0 & 0 & \Sigma_3
\end{bmatrix},
$$

(6.24)
where

\[ \Omega_1 = \text{Var}[(s_{11}, s_{12}, s_{22})'], \]
\[ \Omega_2 = \text{Var}[(s_{33}, s_{34}, s_{44})'], \]

and

\[ \Omega_3 = \text{Var}[(s_{55}, s_{56}, s_{57}, s_{66}, s_{67}, s_{77})']. \]

Estimates of \( \Omega_1, \Omega_2, \) and \( \Omega_3 \) were calculated using formula (3.7) with \( Z_0 \) replaced by \( Z_1, Z_2, \) and \( Z_3, \) respectively. The estimated covariance matrix of \( s_n, \hat{n}, \) is given in Table 5. Table 6 lists the B.G.L.S. estimate, \( \hat{y}, \) of \( \gamma_0 \) and the estimated standard errors of the elements of \( \hat{y}. \) The value of \( g(\hat{y}|\hat{n}^{-1}) \) is 0.026. Recall from Theorem 3.4 that under the assumption that \( x = x(\gamma_0), \) \( g(\hat{y}|\hat{n}^{-1}) \) is approximately distributed as a chi-square random variable with one degree of freedom. Thus, Model I appears to be an acceptable model for these data. The estimates for \( \gamma_1 \) and \( \sigma_u^2 \) are 11.38 and 0.49 with estimated standard errors 1.683 and 0.080, respectively. If one replaces observed soil moisture with true soil moisture in model (5.1) one obtains an estimate of \( \gamma_1 \) equal to 2.34 with an estimated standard error of 0.702 using ordinary least squares. A second ordinary least squares estimate of \( \gamma_1 \) can be obtained by fitting the model
Table 5. Estimated covariance matrix of $s_n$ for the Moody soil moisture experiment

\[
\begin{bmatrix}
550612.3 & 16206.13 & 35519.67 & 476.99 & 1045.45 & 2291.35 \\
787.22 & 743.54 & 32.30 & 41.90 & 28.49 \\
2943.82 & 13.00 & 67.17 & 231.99 \\
2.19 & 0.88 & 0.35 \\
3.76 & 2.88 \\
23.49 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
239.1 & 11.22 & 0.5269 \\
1.07 & 0.0754 \\
0.0108 \\
\end{bmatrix}
\]

Symmetric

\[
\begin{bmatrix}
32.70 & 1.428 & 0.0624 \\
0.275 & 0.0213 \\
0.0072 \\
\end{bmatrix}
\]
Table 6. B.G.L.S. estimate of $\gamma_0$ and estimated standard errors of the elements of $\gamma_0$ for the Moody soil moisture experiment

<table>
<thead>
<tr>
<th>Parameter ($\gamma_0$)</th>
<th>Estimate ($\hat{\gamma}$)</th>
<th>Estimated standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_1$</td>
<td>11.38</td>
<td>1.683</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>8.48</td>
<td>0.999</td>
</tr>
<tr>
<td>$\sigma_r^2$</td>
<td>4.50</td>
<td>1.479</td>
</tr>
<tr>
<td>$\sigma_p^2$</td>
<td>0.28</td>
<td>0.095</td>
</tr>
<tr>
<td>$\sigma_z^2$</td>
<td>0.17</td>
<td>0.053</td>
</tr>
<tr>
<td>$\sigma_s^2$</td>
<td>264.23</td>
<td>88.027</td>
</tr>
<tr>
<td>$\sigma_q^2$</td>
<td>31.88</td>
<td>11.348</td>
</tr>
<tr>
<td>$\sigma_e^2$</td>
<td>22.26</td>
<td>6.832</td>
</tr>
<tr>
<td>$\sigma_{xW}^2$</td>
<td>1.83</td>
<td>1.939</td>
</tr>
<tr>
<td>$\sigma_{w}^2$</td>
<td>14.94</td>
<td>4.847</td>
</tr>
<tr>
<td>$\sigma_{u}^2$</td>
<td>0.49</td>
<td>0.080</td>
</tr>
</tbody>
</table>
\[ \bar{Y}_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_1 \bar{X}_{ij} + \gamma_2 \bar{W}_j + \bar{e}_{ij}, \quad (6.25) \]

\[ i = 1, 2, 3, \quad j = 1, 2, \ldots, 20. \]

The measurement error of \( \bar{X}_{ij} \) is \( 3^{-1/2} \sigma_u \). The estimated value of \( \gamma_1 \) for model (6.25) is 6.983 with an estimated standard error of 1.644. The presence of measurement error in the covariate \( X_{ijk} \) appears to induce a downward bias in the ordinary least squares estimates of \( \gamma_1 \).

B. Estimation of Heritability in Beef Cattle

The data for this example were gathered from an experiment initiated in 1957 at the U.S.D.A. Beef Cattle Research Station in Fort Robinson, Nebraska. Detailed descriptions of this experiment can be found in Gregory, K. E., et al. (1965), Gregory, K. E., et al. (1966a), and Gregory, K. E., et al. (1966b). We briefly describe the experiment.

Eighty Hereford, 80 Angus, and 80 Shorthorn yearling heifers were bred to Hereford, Angus, and Shorthorn bulls. Heifers were randomly assigned to bulls so that the nine mating combinations Hereford bull-Hereford heifer, Angus bull-Angus heifer, Shorthorn bull-Shorthorn heifer, Hereford bull-Angus heifer, Angus bull-Hereford heifer, Hereford bull-Shorthorn heifer, Shorthorn bull-Hereford heifer, Shorthorn bull-Angus heifer, and Shorthorn bull-Shorthorn heifer occurred in the proportions 1/6, 1/6, 1/6, 1/12, 1/12, 1/12, 1/12, 1/12,
and 1/12, respectively. If a heifer failed to produce offspring in a particular year, that animal was culled from the herd for the remainder of the experiment.

Data on calves born in the years 1961 through 1965 are analyzed in this example. The cows were four through seven years of age in these years. Because the age of a cow remained constant through a given year, age of dam and year effects on the calves are confounded. Thus removal of age of dam effects is accomplished with removal of year effects. All calves were weighed within 24 hours of birth and male calves were castrated and dehorned within the same period. Calving took place between February 10 and May 1 of each year, with most calves born between February 10 and March 20. The calves ran with the cows on pasture until they were weaned in the first week of October each year.

Steer calves were group fed and treated identically for a 2-1/2 week period after weaning. These calves were then randomly assigned to individual feeding stalls where they received identical feed, which was approximately 65 percent total digestible nutrients (T.D.N.). Individual feed consumption was recorded daily over a 252 day period. The steer calves were grouped into a big pen and sent to slaughter at the end of the 252 day period. All animals were sent to slaughter at the same time within a given year, although the
length of confinement in the big pen prior to slaughter ranged from one to two weeks in different years. Slaughtering was performed at the same location each year. The right-hand side of each carcass was sent to Kansas State University in Manhattan, Kansas for detailed cut-out data. Measurements were taken on 372 calves. These 372 calves were sired by 51 bulls. Data for the following variables were collected for each calf.

- **WW** = weaning weight (kgs.) of an animal at a standardized 200 day of weaning.
- **ADG** = 100 times the average daily weight gain over the 252 day feeding period.
- **FC** = one-tenth the weight (kgs.) of T.D.N. consumed by an animal during the 252 day feeding period.
- **RP** = retail product of an animal, which is defined to be twice the weight (kgs.) of closely trimmed, nearly boneless meat from the right-hand side of the carcass of an animal.

Closely trimmed meat has less than 0.75 centimeters of fat on any surface. Nearly boneless meat is less than one percent bone by weight. Definitions of these and other variables are contained in Gregory et al. (1966b).

We shall analyze jointly measurements on the four variables WW, ADG, FC, and RP. To simplify the presentation we first consider a single variable. Let \( Y_{jkts} \) be a measurement on the \( s^{th} \) calf of the \( t^{th} \) sire where the calf was born in the \( k^{th} \) year to a sire and dam belonging
to the $j$th mating combination. Let $r_t$ be the number of calves in our data which were sired by the $t$th bull. A model for $Y_{jkts}$ is

$$Y_{jkts} = \mu + \theta_j + \beta_k + \eta_t + \delta_{ts},$$

$$j = 1,2,\ldots,9; \ k = 1,2,\ldots,5;$$

$$t = 1,2,\ldots,51; \ s = 1,2,\ldots,r_t, \ (6.26)$$

where

- $\mu$ = intercept,
- $\theta_j$ = breed effect of sire and dam on $s$th offspring of $t$th bull, corresponding to the $j$th mating combination.
- $\beta_k$ = effect of year of birth and age of dam on $s$th calf of $t$th sire,
- $\eta_t$ = effect of $t$th sire on its $s$th offspring,
- $\delta_{ts}$ = error for the $s$th offspring of $t$th sire.

We assume $\theta_j$ and $\beta_k$ are fixed effects, while $\eta_t$ and $\delta_{ts}$ are independently and identically distributed random variables with zero means and variances $\sigma_{\eta}^2$ and $\sigma_{\delta}^2$, respectively. Furthermore, $\eta_t$ and $\delta_t$'s are assumed to be uncorrelated for $t = 1,2,\ldots,51; \ t' = 1,2,\ldots,51,$

s = 1,2,\ldots,r_t.

Breed, year, and age of dam effects were removed from $Y_{jkts}$ using the computer package SUPER CARP [Hidiroglou et al. (1978)] developed at Iowa State University.

Generalized least squares estimates of the $\theta_j$ and $\beta_k$
were obtained taking into account the nested error structure of Equation (6.26). Let \( y_{ts} \) denote the value of \( Y_{jkts} \) with the estimated year and age effects removed. We treat the estimated values of \( \theta_j \) and \( \beta_k \) as the true values of \( \theta_j \) and \( \beta_k \) so that the model for \( y_{ts} \) is assumed to be

\[
y_{ts} = \mu + \eta_t + \delta_{ts}, \quad t = 1,2,\ldots,51; \quad s = 1,2,\ldots,r_t,
\]

(6.27)

where \( \mu, \eta_t, \) and \( \delta_{ts} \) are defined in (6.26). The \( y_{ts} \) are identically distributed random variables with mean \( \mu \) and variance \( \sigma_p^2 = \sigma_\eta^2 + \sigma_\delta^2 \). Table 7 presents an analysis of variance for model (6.27).

We shall call the model described by (6.27) Model I. Model I is useful for analyzing our data in terms of sire effects. However, we wish to express \( \sigma_\delta^2 \) and \( \sigma_\eta^2 \) in terms of genetic components of variance. To aid in this endeavor we shall define a second model for \( y_{ts} \). First we summarize some basic concepts of quantitative genetics. Our presentation borrows heavily from Kempthorne (1957) and Falconer (1960).

The value observed when some character, or trait, is measured on an individual is the phenotypic value of that individual. The phenotypic value may be divided into components attributable to the influence of genotype and environment. The genotype is the particular assemblage of genes possessed by an individual, and the environment is
Table 7. Analysis of variance for the nested model (Model I) of the beef cattle experiment

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>EMS&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among sires</td>
<td>n-1</td>
<td>$\sum_{t=1}^{n} \sum_{s=1}^{r_t} \left( Y_t \bar{y}_t \right)^2$</td>
<td>(n-1)-1 $\sum_{t=1}^{n} \sum_{s=1}^{r_t} \left( Y_t \bar{y}_t \right)^2$</td>
<td>$\sigma_0^2 + \kappa \sigma_0^2$</td>
</tr>
<tr>
<td>Within sires</td>
<td>$\sum_{t=1}^{n} r_t$</td>
<td>$\sum_{t=1}^{n} \sum_{s=1}^{r_t} \left( Y_{ts} - \bar{Y}_t \right)^2$</td>
<td>$\sum_{t=1}^{n} \sum_{s=1}^{r_t} \left( Y_{ts} - \bar{Y}_t \right)^2$</td>
<td>$\sigma_0^2$</td>
</tr>
<tr>
<td>Corrected total</td>
<td>N-1</td>
<td>$\sum_{t=1}^{n} \sum_{s=1}^{r_t} \left( Y_{ts} - \bar{y}_t \right)^2$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>The value of $\kappa$ is calculated as $\kappa = (n-1)^{-1}[N-N^{-1}(n_t=1 r_t^2)]$ where $r_t$ = the number of offspring sired by the $t^{th}$ bull, $n$ = the number of bulls and $N = \sum_{t=1}^{n} r_t$ = the total number of calves.
all the nongenetic circumstances that influence the phenotypic value. Symbolically,

\[ P = G + E, \]

where \( P \) is the phenotypic value, \( G \) is the genotypic value, and \( E \) is the environmental deviation. Genotypic value may be subdivided further into additive, dominance deviation, and interaction components. We shall assume there is no interaction component. It is necessary that we define the average effect of a gene before discussing the components of genotypic value.

Assume there is some trait (A) at a single locus on the chromosome with two alleles, \( A_1 \) and \( A_2 \). The three possible genotypes resulting from a mating are \( A_1A_1 \), \( A_1A_2 \), and \( A_2A_2 \). These genotypes are arbitrarily assigned values of +a, d, and -a, respectively. Assume the frequency of the genes is \( p \) for \( A_1 \) and \( q = 1-p \) for \( A_2 \). The mean genotypic value of A in a random breeding population is

\[
m = p^2(a) + 2pq(d) + q^2(-a) = a(p-q) + 2dpq
\]

The average effect of a gene is defined to be the mean deviation from \( m \) of genotypic values for individuals which receive that gene from one parent, the gene received
from the other parent having come at random from the population. The average effect of the gene $A_i$ is denoted by $\alpha_i$, $i = 1, 2$. Suppose that an individual has received an $A_1$ gene from one parent in a random breeding population. The individual's genotype will be $A_1A_1$ with probability $p$ and $A_1A_2$ with probability $q$. The mean genotypic value of this individual is $pa + qd$. Thus, the average effect of the gene $A_1$ is

$$\alpha_1 = pa + qd - m$$

$$= pa + qd - [a(p-q) + 2dpq]$$

$$= q[a + d(q-p)].$$  \hspace{1cm} (6.28)

Similarly, the average effect of the gene $A_2$ is

$$\alpha_2 = p[-a + d(p-q)]$$

$$= -p[a + d(q-p)].$$  \hspace{1cm} (6.29)

Suppose $g_{ij}$ is the genotypic value of an individual with genotype $A_iA_j$, $i,j = 1,2$, where $g_{ij}$ is expressed as a deviation from the population mean of genotypic values. We have the identity

$$g_{ij} = \alpha_i + \alpha_j + (g_{ij} - \alpha_i - \alpha_j)$$

$$= a_{ij} + d_{ij},$$  \hspace{1cm} (6.30)

where $a_{ij} = \alpha_i + \alpha_j$ is defined to be the additive effect of the individual's genes and $d_{ij}$ is defined to be the
dominance deviation of the individual. The additive effect is sometimes called the breeding value of the individual. It is easily shown that \( \alpha_1 \) and \( \alpha_2 \) defined by (6.28) and (6.29) are the least squares fit of \((2\alpha_1, \alpha_1 + \alpha_2, 2\alpha_2)'\) to \(\left(g_{11}, g_{12}, g_{22}\right)'\) [Kempthorne (1957), Chapter 15]. Hence, the dominance deviation is the residual of the least squares fit, so that

\[
s_g^2 = s_a^2 + s_d^2,
\]

where the two components of genotypic variance, \(s_g^2\), are the additive variance, \(s_a^2\), and the dominance deviation variance, \(s_d^2\).

Let \(P_{ij}\) denote the phenotypic value of an individual with genotype \(A_i A_j\). We may express \(P_{ij}\) as

\[
P_{ij} = \mu + a_{ij} + d_{ij} + e, \quad (6.31)
\]

where \(\mu\) is the population mean, \(e\) is an environmental error uncorrelated with \(a_{ij}\) and \(d_{ij}\), and \(a_{ij}\) and \(d_{ij}\) are defined by Equation (6.30). Under model (6.30) the components of phenotypic variance, \(s_p^2\), are additive variance, \(s_a^2\), dominance deviation variance, \(s_d^2\), and environmental variance, \(s_e^2\). Thus,

\[
s_p^2 = s_a^2 + s_d^2 + s_e^2. \quad (6.32)
\]
The ratio $\frac{\sigma^2_a}{\sigma^2_p}$ is called the heritability and is of interest to animal breeders.

The above results may be generalized to the case where there are several genes $A_1, A_2, \ldots, A_s$, with corresponding frequencies $p_1, p_2, \ldots, p_s$. The gene effects $(\alpha_1, \alpha_2, \ldots, \alpha_s)$ are defined to be the least squares coefficients obtained by fitting $(2\alpha_1, \alpha_1 + \alpha_2, \ldots, \alpha_i + \alpha_j, \ldots, 2\alpha_s)'$ to $(g_{11}, g_{12}, \ldots, g_{ij}, \ldots, g_{ss})'$ and the dominance deviations are defined by $d_{ij} = g_{ij} - \alpha_i - \alpha_j$. The partitioning of variance in Equation (6.32) remains valid. We define Model II for the phenotypic value $y_{ts}$ to be

$$y_{ts} = \mu + a_{ts} + d_{ts} + e_{ts}, \quad t = 1, 2, \ldots, 51;$$
$$s = 1, 2, \ldots, r_t,$$

where

$\mu$ = intercept,

$a_{ts}$ = additive component of phenotypic value for the $s^{th}$ offspring of the $t^{th}$ sire,

$d_{ts}$ = dominance deviation component of phenotypic value for the $s^{th}$ offspring of the $t^{th}$ sire,

$e_{ts}$ = environmental error for the $s^{th}$ offspring of the $t^{th}$ sire.

We assume the additive, dominance deviation, and environmental components of the phenotypic values are uncorrelated.
Suppose we select at random two individuals X and Y which are a random pair of members of a population with a certain pattern of relationship, e.g., cousins, half-sibs, double first cousins, and so on. Let X have the genotype $A_{x_1}A_{x_2}$, and Y the genotype $A_{y_1}A_{y_2}$. Then the phenotypic values of X and Y are

$$X = \mu + \alpha_{x_1} + \alpha_{x_2} + d_{x_1x_2} + e$$

$$Y = \mu + \alpha_{y_1} + \alpha_{y_2} + d_{y_1y_2} + f,$$  \hspace{1cm} (6.34)

where e and f are environmental errors uncorrelated with each other and with the genotypic values. Let $P(x_1=y_1)$ equal the probability that genes $A_{x_1}$ and $A_{y_1}$ are identical by descent. In the particular case when we may identify $A_{x_1}, A_{y_1}$ as genes contributed to X and Y on one chromosome and $A_{x_2}, A_{y_2}$ as genes contributed to X and Y on the other chromosome, with no relationship between the two chromosomes received by X or by Y, the covariance between X and Y is

$$\text{Cov}(X, Y) = \sigma^2 + \phi' \sigma^2_a$$  \hspace{1cm} (6.35)

where $\phi = P(x_1=y_1)$ and $\phi' = P(x_2=y_2)$. For half-sibs $\phi = \frac{1}{2}$ and $\phi' = 0$, so that the covariance of half-sibs is $\frac{1}{4} \sigma^2_a$.

Models I and II represent the phenotypic value of an
individual chosen from a population consisting of 51 groups of half-sibs. That is, $y_{ts}$ and $y_{ts'}$, $s \neq s'$, represent phenotypic values of half-sibs sired by the $t^{th}$ bull.

From (6.27) and (6.35),

$$\text{Cov}(y_{ts}, y_{ts'}) = \text{Cov}(\eta_t + \delta_{ts}, \eta_t + \delta_{ts'})$$

$$= \sigma^2_\eta$$

$$= \left(\frac{1}{4}\right)\sigma^2_a$$

(6.36)

But, from (6.27) and (6.32),

$$\text{Var}(y_{ts}) = \text{Var}(\eta_t + \delta_{ts})$$

$$= \sigma^2_\eta + \sigma^2_\delta$$

$$= \sigma^2_a + \sigma^2_d + \sigma^2_e$$

(6.37)

Combining (6.36) and (6.37), we obtain

$$\sigma^2_\delta = \left(\frac{3}{4}\right)\sigma^2_a + \sigma^2_f$$

(6.38)

where $\sigma^2_f = \sigma^2_d + \sigma^2_e$. Thus, we can express the expected mean squares of Table 7 in terms of phenotypic components of variance using (6.36) and (6.38). The formula for $\gamma$ in Table 7 is obtained by straight-forward calculation of the expected mean squares.

The four phenotypic values, WW, ADG, FC, and RP, recorded for the $s^{th}$ offspring of the $t^{th}$ sire are denoted
by $W_{ts}$, $ADG_{ts}$, $FC_{ts}$, and $RP_{ts}$, respectively. The multivariate version of Model I for these values is

$$y_{ts} = \mu + \eta_t + \delta_{ts}, \quad t = 1, 2, \ldots, 51;$$

$$s = 1, 2, \ldots, r_t,$$

(6.39)

where

$$y_{ts} = \begin{bmatrix} Y_{1,ts} \\ Y_{2,ts} \\ Y_{3,ts} \\ Y_{4,ts} \end{bmatrix},$$

$$\begin{bmatrix} Y_{1,ts} \\ Y_{2,ts} \\ Y_{3,ts} \\ Y_{4,ts} \end{bmatrix} = \begin{bmatrix} W_{ts} \text{ adjusted for breed, age of dam, and year effects} \\ ADG_{ts} \text{ adjusted for breed, age of dam, and year effects} \\ FC_{ts} \text{ adjusted for breed, age of dam, and year effects} \\ RP_{ts} \text{ adjusted for breed, age of dam, and year effects} \end{bmatrix},$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix},$$

$$\eta_t = \begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \end{bmatrix},$$

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} = \begin{bmatrix} \text{intercept for } W_{ts} \\ \text{intercept for } ADG_{ts} \\ \text{intercept for } FC_{ts} \\ \text{intercept for } RP_{ts} \end{bmatrix},$$

$$\begin{bmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \end{bmatrix} = \begin{bmatrix} \text{random effect of } t^{th} \text{ sire on } W_{ts} \\ \text{random effect of } t^{th} \text{ sire on } ADG_{ts} \\ \text{random effect of } t^{th} \text{ sire on } FC_{ts} \\ \text{random effect of } t^{th} \text{ sire on } RP_{ts} \end{bmatrix}.$$
and

\[
\begin{bmatrix}
\delta_{1,ts} \\
\delta_{2,ts} \\
\delta_{3,ts} \\
\delta_{4,ts}
\end{bmatrix} = \begin{bmatrix}
\text{error for WW}_{ts} \\
\text{error for ADG}_{ts} \\
\text{error for FC}_{ts} \\
\text{error for RP}_{ts}
\end{bmatrix}.
\]

We assume the \( \eta_t \) and \( \delta_{ts} \) are independently and identically distributed with zero means and covariance matrices \( \Sigma_\eta \) and \( \Sigma_\delta \), respectively. Furthermore, we assume \( \eta_t \) and \( \delta_{ts} \) are uncorrelated for \( t = 1, 2, \ldots, 51; t' = 1, 2, \ldots, 51; s = 1, 2, \ldots, r_t \). Thus, the \( \gamma_{ts} \) are identically distributed random vectors with mean zero and covariance matrix \( \Sigma_p = \Sigma_\eta + \Sigma_\delta \). The matrix \( \Sigma_p \) is the phenotypic covariance matrix of \( \gamma_{ts} \).

Let

\[
S_W = (308)^{-1} \sum_{t=1}^{51} \sum_{s=1}^{r_t} (\gamma_{ts} - \bar{\gamma}_t)(\gamma_{ts} - \bar{\gamma}_t),
\]

(6.40)

and

\[
S_A = (50)^{-1} \sum_{t=1}^{51} \sum_{s=1}^{r_t} (\bar{\gamma}_t - \bar{\gamma})(\bar{\gamma}_t - \bar{\gamma}),
\]

(6.41)

where \( \bar{\gamma}_t = r_t^{-1} \sum_{s=1}^{r_t} \gamma_{ts} \) and \( \bar{\gamma} = (\sum_{t=1}^{51} r_t)^{-1} \sum_{t=1}^{51} r_t \gamma_{ts} \).

The 308 degrees of freedom (df) for \( S_W \) are calculated as follows:
308 df = 372 observations - 1 df for intercept - 8 df for breed effects - 5 df for year effects - 51 df for sire effects.

The matrices $S_W$ and $S_A$ are multivariate versions of "within sires" and "among sires" mean squares, respectively. Tables 8 and 9 display the calculated values of $S_W$ and $S_A$. The diagonal elements of $S_W$ were calculated using SUPER CARP and are unbiased estimates of the diagonal elements of $Z_\delta$. An approximation used in the calculation of the off-diagonal elements of $S_W$ may produce a slight bias in these estimators of the off-diagonal elements of $Z_\delta$. We neglect this slight bias and write

$$E(S_W) = Z_\delta'$$

and

$$E(S_A) = Z_\delta' + \kappa Z_\eta', \quad (6.42)$$

where $\kappa$ is defined in Table 7.

The multivariate version of Model II is

$$Y_{ts} = \mu + \alpha_{ts} + \delta_{ts} + \epsilon_{ts}, \quad (6.43)$$

where

$$a_{ts} = \begin{bmatrix} a_{1,ts} \\ a_{2,ts} \\ a_{3,ts} \\ a_{4,ts} \end{bmatrix} = \begin{bmatrix} \text{additive component of WW}_{ts} \\ \text{additive component of ADG}_{ts} \\ \text{additive component of FC}_{ts} \\ \text{additive component of RP}_{ts} \end{bmatrix}.$$
Table 8. The within sires sample covariance matrix, $S_W$, of the beef cattle experiment

\[
S_W = \begin{bmatrix}
606.00 & 5.90 & 116.50 & 157.20 \\
93.50 & 64.18 & 66.89 \\
135.59 & 64.79 \\
Symmetric & 190.26
\end{bmatrix}
\]

Table 9. The among sires sample covariance matrix, $S_A$, of the beef cattle experiment

\[
S_A = \begin{bmatrix}
863.25 & 113.80 & 193.04 & 287.59 \\
126.00 & 103.04 & 164.10 \\
147.33 & 148.85 & 399.50
\end{bmatrix}
\]

\[
d_{ts} = \begin{bmatrix}
d_{1,ts} \\
d_{2,ts} \\
d_{3,ts} \\
d_{4,ts}
\end{bmatrix} = \begin{bmatrix}
\text{dominance deviation component of } WW_{ts} \\
\text{dominance deviation component of } ADG_{ts} \\
\text{dominance deviation component of } FC_{ts} \\
\text{dominance deviation component of } RP_{ts}
\end{bmatrix},
\]

and

\[
e_{ts} = \begin{bmatrix}
e_{1,ts} \\
e_{2,ts} \\
e_{3,ts} \\
e_{4,ts}
\end{bmatrix} = \begin{bmatrix}
\text{environmental error for } WW_{ts} \\
\text{environmental error for } ADG_{ts} \\
\text{environmental error for } FC_{ts} \\
\text{environmental error for } RP_{ts}
\end{bmatrix}.
\]
We assume the additive, dominance deviation, and environmental components of the phenotypic values are uncorrelated.

As in the univariate case we have that

\[ \bar{z}_\delta = \left(\frac{3}{4}\right)z_a + z_f', \]

and

\[ \bar{z}_\eta = \left(\frac{1}{4}\right)z_a, \]

where \( z_a \) is the additive component of \( z_p \) and \( z_f = z_d + z_e' \), where \( z_d \) and \( z_e \) are the dominance deviation and environmental components of \( z_p \), respectively.

We have allowed for dominance deviation, or deviation of the additive value from the genotypic value, of an individual in our development of Model II. The development of Model II was more general, perhaps, than the data of this example warrant. Recall that \( y_{ts} \) represents the phenotypic value of the \( s \)th calf sired by the \( t \)th bull adjusted for breed, age of dam, and year effects. Removal of breed effects eliminates dominance deviation [Willham (1970)]. The component \( d_{ts} \) can be removed from expression (6.43) and \( \bar{z}_\delta \), expressed in genetic components of variance, becomes

\[ \bar{z}_\delta = \left(\frac{3}{4}\right)z_a + z_e'. \]  

(6.44)
We wish to test the hypothesis that the four phenotypic values WW, ADG, FC, and RP depend upon a single common additive component of genotypic value which we denote by $x_1$.

Under this hypothesis $a_{ts}$ has the form

$$a_{ts} = x_{1,ts}B,$$  \hspace{1cm} (6.45)

where $B = (\beta_1 \beta_2 \beta_3 1)$ and the $x_{1,ts}$ are independently and identically distributed with mean zero and variance $\sigma^2_{x_1}$.

Model II then may be expressed as

$$Y_{ts} = x_{1,ts} + u_{ts}, \quad t = 1,2,\ldots,51;$$ $s = 1,2,\ldots,r_t,$  \hspace{1cm} (6.46)

where $Y_{ts} = (y_{1,ts}, y_{2,ts}, y_{3,ts}, y_{4,ts})$, $\beta_\sim = (\beta_1, \beta_2, \beta_3)$, $X_{ts} = x_{1,ts}$, and $(v_{ts}, u_{ts}) = e_{ts}$ is distributed independently of $x_{1,ts}$, $t = 1,2,\ldots,51$; $t' = 1,2,\ldots,51$; $s = 1,2,\ldots,r_t$, and $s' = 1,2,\ldots,r_{t'}$. Model (6.46) is a multivariate structural errors in variables model.

A preliminary check to determine the adequacy of model (6.45) can be made using three simple errors in variables models, the $m^{th}$ of which is defined by

$$y_{m,ts} = x_{1,ts}z_m + e_{m,ts}$$
\[ y_{4,ts} = x_{1,ts} + e_{4,ts}, \quad t = 1,2,\ldots,51; \]
\[ s = 1,2,\ldots, r_t, \quad (6.47) \]

where \( m = 1, 2, \text{or 3} \). The adequacy of models such as (6.47) can be tested using a statistic \( \hat{\lambda} \) defined by Fuller (Forthcoming). We denote by \( \hat{\lambda}_m \) the value of \( \hat{\lambda} \) calculated for the \( m^{\text{th}} \) model defined by (6.47). The approximate distribution of each \( \hat{\lambda}_m \) is Snedecor's F with 49 and infinity degrees of freedom under the null hypotheses that the models (6.47) are the true models. The values of \( \hat{\lambda}_m \), \( m = 1,2,3 \), were calculated using SUPER CARP and are listed in Table 10. The values of \( \hat{\lambda}_2 \) and \( \hat{\lambda}_3 \) are nonsignificant at the 0.10 level, while the value of \( \hat{\lambda}_1 \), which is 1.39, is significant at the 0.05 level. This suggests the following model may be more appropriate than model (6.45).

\[ z_{ts} = x_{ts} A + e_{ts}, \quad t = 1,2,\ldots,51; \quad s = 1,2,\ldots, r_t, \quad (6.48) \]

where

\[ z_{ts} = (y_{ts}, x_{ts}), \quad x_{ts} = (x_{1,ts}, x_{2,ts}), \quad \text{and} \]

\[ A = \begin{bmatrix} \beta_1 & 0 \\ \beta_2 & 0 \\ \beta_3 & 0 \\ 1 & 0 \end{bmatrix}. \]
Table 10. Preliminary tests for the adequacy of Model II in the beef cattle experiment when additive variation is assumed to be due to a single underlying variable

<table>
<thead>
<tr>
<th>m</th>
<th>$\hat{\lambda}_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.39*</td>
</tr>
<tr>
<td>2</td>
<td>0.84</td>
</tr>
<tr>
<td>3</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*Approximate distribution of $\hat{\lambda}_m$ is Snedecor's F with 49 and infinity degrees of freedom.

*Significant at 0.05 level.

The $x_{ts}$ are assumed to be distributed independently of each other, and of any $e_{ts}$, with common distribution $N(0, \Sigma_{xx})$, where $\Sigma_{xx} = \text{diag}(\sigma^2_{x_1}, \sigma^2_{x_2})$. Model (6.48) is the classical factor analysis model defined by (1.16), except we allow $\text{Var}(e_{ts}) = \Sigma_{ee}$ to be nondiagonal. Although not a multivariate errors in variables model, the G.L.S. approach is applicable to model (6.48).

We apply the G.L.S. procedure to models (6.45) and (6.48) under the additional assumption that the $y_{ts}$ are normally distributed. Define

$$\gamma_0 = (\beta_1, \beta_2, \beta_3, \sigma^2_{x_1}, \text{vech} \Sigma_a),$$

(6.49)
and

$$\gamma_0 = (\beta_1, \beta_2, \beta_3, \sigma_{x_1}^2, \sigma_{x_2}^2, \text{vech } z_a), \quad (6.50)$$

where \( \text{vech } z_a = (\sigma_{11}, \sigma_{21}, \sigma_{31}, \sigma_{41}, \sigma_{22}, \sigma_{32}, \sigma_{42}, \sigma_{33}, \sigma_{43}, \sigma_{44})' \). Let \( s_n \) be defined by

$$s_n = \begin{bmatrix} \text{vech } S_W \\ \text{vech } S_A \end{bmatrix}. \quad (6.51)$$

The expected values of \( s_n \) under models (6.45) and (6.48) are denoted by \( \sigma(\gamma_0^1) \) and \( \sigma(\gamma_0^2) \), respectively.

We require a consistent estimator of \( [\text{Var}(s_n)]^{-1} \) to obtain B.G.L.S. estimates of \( \gamma_0^1 \) and \( \gamma_0^2 \). Under our normality assumption, \( \text{vech } S_W \) and \( \text{vech } S_A \) are distributed independently of each other. Since \( S_W \) is a multiple of a Wishart matrix, formula (3.7) was used to calculate the estimate \( \tilde{\Omega}_W \) of \( [\text{Var}(\text{vech } S_w)] \) shown in Table 11. Because the \( t^{th} \) sire produced \( r_t \) offspring, \( \text{Var}(r_t \gamma_{t}) = r_t \gamma_t + z_0'. \) Thus, \( S_A \) is not a multiple of a Wishart matrix and use of formula (3.7) need not produce a consistent estimator of \( [\text{Var}(\text{vech } S_A)]^{-1} \) under our assumptions. The estimate of \( [\text{Var}(\text{vech } S_A)] \), \( \tilde{\Omega}_A \), was computed using formula (5.3) and is shown in Table 12. Let

$$\tilde{\Omega} = \text{diag}(\tilde{\Omega}_W, \tilde{\Omega}_A). \quad (6.52)$$
Table 11. The estimated covariance matrix of vech $\tilde{S}_W$

\[
\begin{bmatrix}
2384.65 & 23.22 & 458.44 & 618.59 & 0.23 & 4.46 & 6.02 & 88.13 & 118.92 & 160.47 \\
184.08 & 128.51 & 134.62 & 3.58 & 36.60 & 49.00 & 48.55 & 58.06 & 68.28 \\
310.84 & 186.94 & 2.46 & 26.87 & 34.00 & 102.57 & 93.71 & 66.14 \\
454.65 & 2.56 & 26.54 & 37.79 & 49.01 & 105.05 & 194.25 \\
56.77 & 38.97 & 40.61 & 26.75 & 27.88 & 29.05 \\
54.53 & 33.61 & 56.51 & 42.95 & 28.14 \\
72.30 & 27.00 & 53.72 & 82.66 \\
119.38 & 57.04 & 27.26 \\
97.40 & 80.06 \\
235.16
\end{bmatrix}
\]

$\tilde{\Omega}_W$ is symmetric.

Symmetric

154
Table 12. The estimated covariance matrix of vech $S$, $\tilde{\Omega}_A$

$$
\tilde{\Omega}_A = \\
\begin{bmatrix}
26,829.10 & 974.05 & 4511.02 & 7796.65 & 167.07 & 264.86 & 218.06 & 464.52 & 1293.85 & 2072.10 \\
1976.86 & 1527.30 & 2275.48 & 201.78 & 310.54 & 617.72 & 308.49 & 647.22 & 254.60 \\
2158.67 & 2669.91 & 72.68 & 249.88 & 368.95 & 453.34 & 716.29 & 148.92 \\
7026.19 & 285.08 & 407.39 & 221.54 & 458.32 & 796.85 & -2121.92 \\
633.72 & 419.77 & 776.74 & 219.65 & 595.93 & 1262.81 \\
369.89 & 631.02 & 298.93 & 581.67 & 1151.10 \\
1706.36 & 414.59 & 1467.41 & 3664.37 \\
430.75 & 504.70 & 789.88 \\
Symmetric & & & & & & & & & \\
1483.42 & 3107.63 \\
11,264.00
\end{bmatrix}
$$
Denote by $\hat{y}_i^i$ the B.G.L.S. estimate of $\gamma_i^0$ obtained using $\tilde{n}^{-1}$ to estimate $[\text{Var}(s_i)]^{-1}$, $i = 1, 2$. Tables 13 and 14 list the elements of $\hat{\gamma}_1$ and $\hat{\gamma}_2$, their estimated standard errors, and the approximate t-statistic associated with each element. The values of $g(\hat{\gamma}_1|\tilde{n}^{-1})$ and $g(\hat{\gamma}_2|\tilde{n}^{-1})$ are 7.72 and 8.24, respectively, where $g(\hat{\gamma}|\tilde{n}^{-1})$ is defined by (5.15). Recall from Theorem 3.4 that the distribution of $g(\hat{\gamma}|\tilde{n}^{-1})$ is approximately chi-square with $q-r$ degrees of freedom under the hypothesis that $\gamma = \gamma_0$, where $q$ is the dimension of $s_n$ and $r$ is the dimension of $\gamma_0$. Thus, $g(\hat{\gamma}_1|\tilde{n}^{-1})$ and $g(\hat{\gamma}_2|\tilde{n}^{-1})$ are distributed as approximate chi-square random variables with six and five degrees of freedom, respectively.

Because both $g(\hat{\gamma}_1|\tilde{n}^{-1})$ and $g(\hat{\gamma}_2|\tilde{n}^{-1})$ are nonsignificant at the 0.10 level, models (6.45) and (6.48) both appear to be given an adequate fit to the data. However, the approximate t-tests for the estimates of $\beta_1$ and $\sigma_{x_2}$ in model (6.48) are nonsignificant at the 0.10 level, where, by analogy to ordinary least squares, levels of significance are based on a t-distribution with $q-r$ degrees of freedom. This suggests that (6.45) may be a more appropriate model for these data than (6.48). Thus, we cannot reject our hypothesis that phenotypic values of WW, ADG, FC, and RP are influenced by a single common additive component.
Table 13. B.G.L.S. estimates of the elements of \( Y_0^1 \), the estimated standard errors, and approximate t-statistics

<table>
<thead>
<tr>
<th>Parameter ((Y_0^1))</th>
<th>Estimate ((\hat{Y}_0^1))</th>
<th>Estimated standard error</th>
<th>Approximate t-statistic^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>1.017</td>
<td>0.609</td>
<td>1.67</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.564</td>
<td>0.145</td>
<td>3.89*</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>0.507</td>
<td>0.134</td>
<td>3.78*</td>
</tr>
<tr>
<td>( \sigma_{x_1} )</td>
<td>8.399</td>
<td>2.437</td>
<td>3.45*</td>
</tr>
<tr>
<td>( \sigma_{x_2} )</td>
<td>7.786</td>
<td>3.95</td>
<td>1.97</td>
</tr>
<tr>
<td>( \sigma_{11} )</td>
<td>495.960</td>
<td>77.432</td>
<td>6.405**</td>
</tr>
<tr>
<td>( \sigma_{21} )</td>
<td>-20.089</td>
<td>23.104</td>
<td>-0.87</td>
</tr>
<tr>
<td>( \sigma_{31} )</td>
<td>83.140</td>
<td>26.347</td>
<td>3.16*</td>
</tr>
<tr>
<td>( \sigma_{41} )</td>
<td>100.010</td>
<td>39.185</td>
<td>2.55</td>
</tr>
<tr>
<td>( \sigma_{22} )</td>
<td>74.904</td>
<td>13.283</td>
<td>5.64**</td>
</tr>
<tr>
<td>( \sigma_{32} )</td>
<td>50.429</td>
<td>12.632</td>
<td>3.99*</td>
</tr>
<tr>
<td>( \sigma_{42} )</td>
<td>38.174</td>
<td>18.873</td>
<td>2.02</td>
</tr>
<tr>
<td>( \sigma_{33} )</td>
<td>114.147</td>
<td>15.466</td>
<td>7.38**</td>
</tr>
<tr>
<td>( \sigma_{43} )</td>
<td>36.913</td>
<td>19.364</td>
<td>1.91</td>
</tr>
<tr>
<td>( \sigma_{44} )</td>
<td>136.762</td>
<td>36.825</td>
<td>3.71*</td>
</tr>
</tbody>
</table>

^aLevels of significance are based on a t-distribution with five degrees of freedom.

* Significant at 0.05 level.

** Significant at 0.01 level.
Table 14. B.G.L.S. estimates of the elements of $\chi_0^2$ the estimated standard errors, and approximate t-statistics

<table>
<thead>
<tr>
<th>Parameter $(\chi_0^2)$</th>
<th>Estimate $(\hat{\chi}_0^2)$</th>
<th>Estimated standard error</th>
<th>Approximate t-statistic$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.763</td>
<td>0.618</td>
<td>2.85*</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.607</td>
<td>0.166</td>
<td>3.66*</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.589</td>
<td>0.159</td>
<td>3.70*</td>
</tr>
<tr>
<td>$\sigma_{x_1}$</td>
<td>6.759</td>
<td>1.917</td>
<td>3.53*</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>488.594</td>
<td>86.997</td>
<td>5.62**</td>
</tr>
<tr>
<td>$\sigma_{21}$</td>
<td>-27.675</td>
<td>21.856</td>
<td>1.27</td>
</tr>
<tr>
<td>$\sigma_{31}$</td>
<td>71.926</td>
<td>28.035</td>
<td>2.57*</td>
</tr>
<tr>
<td>$\sigma_{41}$</td>
<td>90.636</td>
<td>39.290</td>
<td>2.31</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>79.776</td>
<td>12.954</td>
<td>6.16**</td>
</tr>
<tr>
<td>$\sigma_{32}$</td>
<td>53.933</td>
<td>11.784</td>
<td>4.58**</td>
</tr>
<tr>
<td>$\sigma_{42}$</td>
<td>48.494</td>
<td>15.932</td>
<td>3.04*</td>
</tr>
<tr>
<td>$\sigma_{33}$</td>
<td>116.333</td>
<td>14.827</td>
<td>7.85**</td>
</tr>
<tr>
<td>$\sigma_{43}$</td>
<td>43.349</td>
<td>16.205</td>
<td>2.68*</td>
</tr>
<tr>
<td>$\sigma_{44}$</td>
<td>155.255</td>
<td>27.900</td>
<td>5.56**</td>
</tr>
</tbody>
</table>

$^a$Levels of significance are based on a t-distribution with six degrees of freedom.

* Significant at 0.05 level.

** Significant at 0.01 level.
VII. BIBLIOGRAPHY


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


Wald, A. 1940. Fitting of straight lines if both variables are subject to error. Ann. Math. Statist. 11:284-300.


VIII. ACKNOWLEDGMENTS

I wish to thank Professor Wayne A. Fuller for his helpful guidance in the preparation of this dissertation. The past three years spent working with Professor Fuller have been an extremely stimulating and enjoyable experience. Professors Dean Isaacson, Glen Meeden, Paul Hinz, Malay Ghosh, and Ed Pollak were also of great help in my graduate studies.

Margaret Nichols and Linda Price have provided skilled typing of several drafts of this dissertation. Helen Nelson has aided me through the efficient performance of her secretarial duties. I am grateful for their help and friendship.

Thanks go to Pat Gunnells for her excellent typing of the final version of the dissertation.

Appreciation is also given to Diane Dahm and Brian Dahm for their encouragement and support, and to our unborn child who provided additional stimulus to finish my graduate work.