A MULTIVARIATE LINEAR MODEL WITH LATENT FACTOR STRUCTURE

by

German Rodriguez

Institute of Statistics Mimeo Series #1014
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina

June 1975

RODRIGUEZ, GERMAN. A Multivariate Linear Model with Latent Factor Structure. (Under the direction of NORMAN LLOYD JOHNSON and HENRY BRADLEY WELLS.)

In the social and behavioral sciences, the variables of interest are frequently unobservable constructs, factors, or latent variables. The statistical analysis, on the other hand, must be based on observable indicators, responses, or manifest variables. In this dissertation we propose a general statistical model for the analysis of this type of data, designed to permit estimation of parameters and tests of hypotheses pertaining to unobservable constructs on the basis of observable indicators. The model results from combining a multivariate linear model for the latent variables with a factor analysis model relating these to the manifest variables, and is termed the *latent linear model*.

The model is stated in general form as a linear model where the regression and dispersion parameters are structured, and is shown to be analogous in its method of construction to growth curve and covariance structure models. The problem of identifying the structural parameters is discussed, and solutions are given along the same lines as in factor analysis. Some special cases of the model are considered, and used to illustrate its range of application and some relationships with other models, such as factor analysis in several populations, path analysis, and variance components.

The likelihood equations for estimating the parameters are derived, using some recent results on matrix differentiation. An iterative procedure based on the Fletcher-Powell method of function minimiza-

tion is proposed for their solution. The numerical procedure has been found to possess excellent convergence properties. A detailed treatment of large sample theory is given, including proofs of the consistency and asymptotic normality of the estimators, for a family of structural linear models which includes the latent linear model. In this process the concept of the limiting Fisher information matrix is introduced, and expressions for its elements are derived. These are used to obtain formulae for the asymptotic variances and covariances of the estimators of the structural parameters in the latent linear model. The likelihood ratio and Wald techniques are used to construct large sample tests of the goodness of fit of the model, and of a variety of hypotheses about the structural parameters; and the asymptotic distributions of the test statistics are derived.

A numerical example using simulated data is given, illustrating the proposed estimation and testing procedures. The results of a simulation study conducted to evaluate several large sample approximations are reported. These results indicate that the asymptotic results developed provide reasonable approximations for moderate size samples. Finally two extensions of the model are suggested as topics for further research.

ACKNOWLEDGEMENTS

It is a pleasure to express my appreciation to my co-advisors Dr. N.L. Johnson, for his patience, advice and encouragement during the course of this research; and Dr. H.B. Wells, for his continued guidance and support throughout my graduate studies.

I am also indebted to the other members of my advisory committee, Drs. J.E. Grizzle, D. Quade and P. Uhlenberg - as well as Dr. P.K. Sen, who participated in the committee before taking a sabbatical leave - for their assistance and interest in my work. Special thanks go to Dr. D. Quade for his many detailed comments and suggestions.

I would also like to thank Drs. W. Hoeffding and D.J. de Waal, for useful discussions on maximum likelihood estimation and on matrix derivatives.

During my stay at Chapel Hill I received continued financial support from the Population Council and the Ford Foundation, and I would like to express my debt of gratitude to these institutions. I am also grateful to the Carolina Population Center for providing a tuition grant in 1972.

I must warmly thank my family, and in particular my wife Pat, for the enduring love and support that made possible my graduate studies.

Finally, I wish to thank Mrs. J. Maxwell, for typing the manuscript with great patience, speed and accuracy.

CONTENTS

I. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

In the social and behavioral sciences the variables of interest are frequently unobservable constructs, factors (Thurstone, 1947) or *latent variables* (Lazarsfeld, 1950), such as attitudes, intelligence or socio-economic status. The statistical analysis, on the other hand, must be based on observable indicators, responses or *manifest variables*, such as verbal expressions of attitudes, performance on an intelligence test, or education and income. Furthermore, since measurement in the social sciences is usually inexact, a multiple indicator approach must frequently be employed in data collection, using several measurements of a few underlying factors of interest.

In this work we propose and study a general statistical model for the analysis of this type of data, designed to permit estimation of parameters and tests of hypotheses pertaining to unobservable constructs or latent variables, on the basis of observable indicators. The model results from combining a multivariate linear model for the latent variables with a factor analysis model relating these to the manifest variables. In this context the relationships of interest are represented in the underlying linear model and factor analysis is used to model the measurement process.

In Chapter 2 we state the general model and discuss its relationship with other models proposed in the literature. In Chapter 3 we derive

maximum likelihood equations for estimating the parameters and describe an iterative procedure for their solution. In Chapter 4 we study large sample properties of the estimators and obtain formulae for their asymptotic standard errors. In Chapter 5 we discuss hypothesis testing using likelihood ratio and Wald techniques. In Chapter 6 we provide numerical examples to illustrate the procedures. In Chapter 7 we describe a simulation study conducted to evaluate some large sample approximations and to assess the power of the proposed test procedures. Finally, in Chapter 8 we discuss some extensions and suggestions for further work.

In the derivation of our results we make extensive use of matrix derivatives. Since some of the techniques used have been developed very recently, we provide a brief review of matrix differentiation in the appendix.

The proposed model is an extension of the factor analysis model. In order to develop a proper background for its study, a review of factor analysis is provided in the rest of this introduction. The main results and developments are discussed, but proofs are omitted whenever they can be found in the literature or when more general results are proved later in this work.

1.2 The Factor Analysis Model

1.2.1 Historical Remarks

Factor analysis originated with Spearman (1904), who proposed that correlations among several tests of intelligence could be accounted for by a common factor underlying all tests (intelligence) plus factors specific to each test (errors). The theory was extended to multiple

factors by Thurstone (1931, 1947), who gave considerable impetus to the development of the field. Factor analysis has been developed mostly by psychologists and has always been a controversial subject, see for example the discussion in Kendall and Babington Smith (1950). Harman (1967) provides an authoritative account of the different schools of thought and methods of analysis that have evolved.

The statistical approach to factor analysis was initiated by Lawley (1940, 1942, 1943), who derived maximum likelihood equations for estimating the parameters of factor models. For several years the development of this approach was hindered by a lack of satisfactory numerical methods for obtaining the estimates. Howe (1955) and Bargmann (1957) proposed a Gauss-Seidel iterative method, and Rao (1955) proposed an estimation procedure based on canonical correlation analysis, which can be shown to be equivalent to maximum likelihood estimation; but experience with these methods was not satisfactory, see for example Maxwell (1961) and Browne (1965). More recently, however, Jöreskog (1967, 1969) developed a procedure based on the Fletcher and Powell (1963) method of function minimization which proved superior to previous developments and made efficient and accurate numerical solution of the likelihood equations possible, see also Lawley (1967) and Jöreskog and Lawley (1968).

Other statisticians who have contributed to the development of factor analysis are Bartlett (1937, 1938, 1950, 1953) and Anderson and Rubin (1956). An excellent discussion of the maximum likelihood approach appears in Lawley and Maxwell (1971).

1.2.2 The Factor Model

The general factor analysis model is

$$\underset{\sim}{x} = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\underset{\sim}{\chi} + \underset{\sim}{z} \ , \tag{1.2.1}$$

where $\underset{\sim}{x}$: p×1 is a stochastic vector of manifest variables or responses, $\underset{\sim}{\mu}$: p×1 is a vector of means, $\underset{\sim}{\Lambda}$:p×q is a matrix of parameters called factor *loadings* of full column rank q<p, $\underset{\sim}{\chi}$: q×1 is a stochastic vector of latent variables or factors with $E(\underset{\sim}{\chi}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{\chi}) = \underset{\sim}{\Phi}$ p.d. (positive definite) and $\underset{\sim}{z}$: p×1 is a vector of random errors with $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{z}) = \underset{\sim}{\Psi} = diag(\psi_1,\ldots,\psi_p)$ p.d. and $Cov(\underset{\sim}{\chi},\underset{\sim}{z}) = \underset{\sim}{0}$.

The model implies that $E(\underset{\sim}{x}) = \underset{\sim}{\mu}$ and $Var(\underset{\sim}{x}) = \underset{\sim}{V}$, where the variance-covariance matrix has the following structure:

$$\underset{\sim}{V} = \underset{\sim\sim}{\Lambda\Phi\Lambda}' + \underset{\sim}{\Psi} \ . \tag{1.2.2}$$

The diagonal elements of $\underset{\sim\sim}{\Lambda\Phi\Lambda}'$ are called *communalities*, and the diagonal elements of $\underset{\sim}{\Psi}$ are called *specificities*. These names refer to those parts of the variance of each response that can be attributed to the common factors and to the specific errors, respectively.

The essence of the model is to write the manifest variables as linear combinations of a smaller number of latent variables plus uncorrelated errors. Since from (1.2.1)

$$Var(\underset{\sim}{x}|\underset{\sim}{\chi}) = \underset{\sim}{\Psi} = diag(\psi_1,\ldots,\psi_p) \ , \tag{1.2.3}$$

the factors may be said to explain the covariance structure of the responses.

1.2.3 The Identification Problem

In (1.2.1) we assumed that $E(\underset{\sim}{\chi}) = \underset{\sim}{0}$. If $E(\underset{\sim}{\chi}) = \underset{\sim}{\xi}$ then $E(\underset{\sim}{x}) = \underset{\sim}{\mu} + \underset{\sim\sim}{\Lambda\xi}$ and clearly $\underset{\sim}{\mu}$ and $\underset{\sim}{\xi}$ cannot be separately identified.

The assumption that $E(\underset{\sim}{\chi}) = \underset{\sim}{0}$ implies no loss of generality, in the sense that if $E(\underset{\sim}{\chi}) = \underset{\sim}{\xi}$ we can redefine $\underset{\sim}{\chi}^* = \underset{\sim}{\chi} - \underset{\sim}{\xi}$ and $\underset{\sim}{\mu}^* = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\underset{\sim}{\xi}$, thus reducing the general case to (1.2.1). Since interest centers on the covariance structure, most authors write the model with both $E(\underset{\sim}{\chi}) = \underset{\sim}{0}$ and $E(\underset{\sim}{x}) = \underset{\sim}{0}$, effectively working with $\underset{\sim}{x}^* = \underset{\sim}{x} - \underset{\sim}{\mu}$, see for example Jöreskog (1967) or Lawley and Maxwell (1971, Ch. 2). We shall adopt this practice hereafter. The preceding remarks, however, will be important in the context of the general model proposed in §2.2.

Turning our attention to the covariance structure, note that the model is not identified unless restrictions are imposed on $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Phi}$, for if $\underset{\sim}{V}$ satisfies (1.2.2) and $\underset{\sim}{L}$ is any $q \times q$ non-singular matrix, then $\underset{\sim}{V}$ would still satisfy (1.2.2) if $\underset{\sim}{\Lambda}$ is replaced by $\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\underset{\sim}{L}^{-1}$ and $\underset{\sim}{\Phi}$ is replaced by $\underset{\sim}{\Phi}^* = \underset{\sim}{L}\underset{\sim}{\Phi}\underset{\sim}{L}'$. This indeterminacy of the model corresponds to a non-singular linear transformation of the factors $\underset{\sim}{\chi}^* = \underset{\sim}{L}\underset{\sim}{\chi}$. Since $\underset{\sim}{L}$ has q^2 elements, at least q^2 constraints need be imposed upon the parameters.

There are two approaches or types of solutions to this problem, known as exploratory or unrestricted factor analysis and confirmatory or restricted factor analysis.

In *exploratory factor analysis* $\frac{1}{2}q(q+1)$ constraints are imposed by defining the factors to be uncorrelated and have unit variances. In this case $\underset{\sim}{\Phi} = \underset{\sim}{I}_q$ and (1.2.2) becomes

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} . \qquad (1.2.2)'$$

The assumption $\mathrm{Var}(\underset{\sim}{\chi}) = \underset{\sim}{I}_q$ implies no loss of generality, for if $\mathrm{Var}(\underset{\sim}{\chi}) = \underset{\sim}{\Phi}$ we can redefine $\underset{\sim}{\chi}^* = \underset{\sim}{\Phi}^{-\frac{1}{2}}\underset{\sim}{\chi}$ and $\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}^{\frac{1}{2}}$ where $\underset{\sim}{\Phi}^{\frac{1}{2}}$ is any square root of $\underset{\sim}{\Phi}$ (symmetric or lower triangular). Then $\underset{\sim}{x} = \underset{\sim}{\Lambda}\underset{\sim}{\chi} + \underset{\sim}{z}$ can be written as $\underset{\sim}{x} = \underset{\sim}{\Lambda}^*\underset{\sim}{\chi}^* + \underset{\sim}{z}$ and $\underset{\sim}{V} = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}$ can be

written as $\underset{\sim}{V} = \underset{\sim}{\Lambda}^*\underset{\sim}{\Lambda}^{*\prime} + \underset{\sim}{\Psi}$, thus reducing the general case (1.2.2) to (1.2.2)'. Hence the name unrestricted factor analysis.

If $q > 1$ the model is still not identified, however, for if $\underset{\sim}{V}$ satisfies (1.2.2)' and $\underset{\sim}{M}$ is any orthogonal matrix of order q, then $\underset{\sim}{V}$ also satisfies (1.2.2)' with $\underset{\sim}{\Lambda}$ replaced by $\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\underset{\sim}{M}'$. This indeterminacy corresponds to a (rigid) rotation of the factors $\underset{\sim}{y}^* = \underset{\sim}{M}\underset{\sim}{y}$. Since $\underset{\sim}{M}$ has $\frac{1}{2}q(q-1)$ free elements, an additional $\frac{1}{2}q(q-1)$ constraints are needed. This is the problem of selecting a *basis* of the common factor space.

A set of restrictions that turns out to be quite convenient and interesting is to require $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ to be diagonal with its elements arranged in decreasing order of magnitude. This restriction will be related to canonical correlation analysis in §1.2.4 and to principal components in §1.2.5. The resulting basis will be called the *canonical basis* of the factor space, following Rao (1955).

In many behavioral science applications the canonical basis is used in estimation but the resulting estimates are then *rotated* to obtain a structure that is easier to interpret. Several criteria for rotation have been proposed, but these will not be reviewed here. The interested reader is referred to Kaiser (1958, 1959), Hendrickson and White (1964), Jennrich and Sampson (1966), Crawford and Ferguson (1970), Browne (1974a) and Lawley and Maxwell (1971, Ch. 6).

In *confirmatory factor analysis* q constraints are usually imposed by defining the factors as having unit variances, and at least an additional $q(q-1)$ constraints are imposed by requiring $q(q-1)$ or more elements of $\underset{\sim}{\Lambda}$ to be fixed, usually zero. The pattern of fixed values must be such that it would be destroyed by any variance-preserving non-singular

linear transformation of the factors other than the identity matrix. A simple way to achieve identifiability is to set to 0 at least (q-1) elements in each column of $\underset{\sim}{\Lambda}$. Other conditions are given by Anderson and Rubin (1956). This approach usually entails a restriction on the common factor space and rests on a hypothesis regarding the structure of $\underset{\sim}{\Lambda}$. Hence the names restricted or confirmatory factor analysis.

1.2.4. A Canonical Reduction of the Factor Model

Consider the unrestricted factor model with $\underset{\sim}{\Phi} = \underset{\sim}{I}_q$ and note that

$$\text{Var}\begin{bmatrix} \underset{\sim}{x} \\ \underset{\sim}{y} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{V} & \underset{\sim}{\Lambda} \\ \underset{\sim}{\Lambda}' & \underset{\sim}{I}_q \end{bmatrix} . \tag{1.2.4}$$

From canonical correlation theory, see Hotelling (1935, 1936), we know that there exist linear transformations $\underset{\sim}{x}^* = \underset{\sim}{L}'\underset{\sim}{x}$ and $\underset{\sim}{y}^* = \underset{\sim}{M}'\underset{\sim}{y}$ such that

$$\text{Var}\begin{bmatrix} \underset{\sim}{x}^* \\ \underset{\sim}{y}^* \end{bmatrix} = \begin{bmatrix} \underset{\sim}{I}_p & \underset{\sim}{\Gamma} \\ \underset{\sim}{\Gamma}' & \underset{\sim}{I}_q \end{bmatrix} , \tag{1.2.5}$$

where $\underset{\sim}{\Gamma}$: $p \times q = \begin{bmatrix} \underset{\sim}{P} \\ \underset{\sim}{0} \end{bmatrix}$ and $\underset{\sim}{P} = \text{diag}(\rho_1, \ldots, \rho_q)$. The ρ_i are the canonical correlations between factors and responses, and $\underset{\sim}{x}^*$, $\underset{\sim}{y}^*$ are the corresponding canonical variates. Furthermore $\rho_i^2 = \text{ch}_i(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})$ is the i-th largest characteristic root of $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$, $\underset{\sim}{L}$ is a matrix of eigenvectors of $\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Lambda}'$ standardized so that $\underset{\sim}{L}'\underset{\sim}{V}'\underset{\sim}{L} = \underset{\sim}{I}_p$, and $\underset{\sim}{M}$ is a matrix of orthonormal eigenvectors of $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$.

The canonical variates have the interesting property that

$$\underset{\sim}{x}^* = \underset{\sim}{\Gamma}\underset{\sim}{y}^* + \underset{\sim}{z}^* , \tag{1.2.6}$$

where $\underset{\sim}{z}^* = \underset{\sim}{L}'\underset{\sim}{z}$ with $\mathrm{Var}(\underset{\sim}{z}^*) = \underset{\sim}{L}'\underset{\sim}{\Psi}\underset{\sim}{L} = \mathrm{diag}(1-\rho_1^2,\ldots,1-\rho_q^2,1,\ldots,1) = \underset{\sim}{\Psi}^*$, say, and $\mathrm{Cov}(\underset{\sim}{\chi}^*,\underset{\sim}{z}^*) = \underset{\sim}{0}$. Thus (1.2.6) is a factor model. In view of the structure of $\underset{\sim}{\Gamma}$ this model has the property that x_i^* is loaded only on factor y_i^* for $i=1,\ldots,q$ and is independent of the factors y_i^* for $i = q+1,\ldots,p$. Also the loading of x_i^* on y_i^* is the i-th largest canonical correlation between factors and responses. Thus we have reduced the general model (1.2.1) to a particularly simple structure (1.2.6). This is called the *canonical reduction* of the factor model, see Rodriguez (1975).

Rao (1955) has proposed *defining* the factors $\underset{\sim}{\chi}$ as the canonical variates of the factor space with respect to the response space. In terms of our analysis this implies that $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ must be diagonal and have its elements arranged in decreasing order of magnitude, for then $\underset{\sim}{P}^2$ is $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ itself and $\underset{\sim}{M} = \underset{\sim}{I}_q$. It can be shown that $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda} = \underset{\sim}{\Delta}(\underset{\sim}{I}+\underset{\sim}{\Delta})^{-1}$ where $\underset{\sim}{\Delta} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ (see §3.4), and hence a sufficient condition for $\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}$ to be diagonal and have its elements ordered is that $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal and have its elements ordered. This shows that the canonical restrictions do indeed lead to Rao's canonical basis of the factor space.

Let $\underset{\sim}{x}_1^* = (x_1^*,\ldots,x_q^*)'$ be the first q canonical variates of the response space. If $\underset{\sim}{\Delta}$ is diagonal it can be shown that

$$\underset{\sim}{x}_1^* = \underset{\sim}{\Delta}^{-\frac{1}{2}}(\underset{\sim}{I}+\underset{\sim}{\Delta})^{-\frac{1}{2}}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{x} \ . \tag{1.2.7}$$

This result is related to methods for estimating factor scores proposed by Bartlett (1937, 1938) and Thompson (1951). For further details and a derivation of these results see Rodriguez (1975).

1.2.5 Factor Analysis and Principal Components

The unrestricted factor model is similar to, but should be distinguished from, principal components, introduced by Hotelling (1933). We now compare these models.

Let $E(\underset{\sim}{x}) = \underset{\sim}{0}$, $Var(\underset{\sim}{x}) = \underset{\sim}{V}$, $\underset{\sim}{D} = diag(d_1,\ldots,d_p)$ where $d_i = ch_i(\underset{\sim}{V})$ and let $\underset{\sim}{A}$: p×p be the matrix of orthonormal eigenvectors of $\underset{\sim}{V}$. Then $\underset{\sim}{A}'\underset{\sim}{x}$ is the vector of principal components of $\underset{\sim}{x}$. Let $\underset{\sim}{\Gamma} = \underset{\sim}{A}\underset{\sim}{D}^{\frac{1}{2}}$. Then since $\underset{\sim}{V} = \underset{\sim}{A}\underset{\sim}{D}\underset{\sim}{A}' = \underset{\sim}{\Gamma}\underset{\sim}{\Gamma}'$ we can write

$$\underset{\sim}{x} = \underset{\sim}{\Gamma}\underset{\sim}{y} \ , \tag{1.2.8}$$

where $E(\underset{\sim}{y}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{y}) = \underset{\sim}{I}_p$. Thus principal components may be considered a factor model with p factors and no errors. From (1.2.8) $\underset{\sim}{A}'\underset{\sim}{x} = \underset{\sim}{D}^{\frac{1}{2}}\underset{\sim}{y}$; thus the p factors are the principal components of $\underset{\sim}{x}$ standardized to unit variance.

So far we assumed that $\underset{\sim}{V}$ is of full rank. If rank $\underset{\sim}{V} = q < p$ we take $\underset{\sim}{D} = diag(d_1,\ldots,d_q)$, let $\underset{\sim}{A}$: p×q be the first q eigenvectors of $\underset{\sim}{V}$ and let $\underset{\sim}{\Gamma} = \underset{\sim}{A}\underset{\sim}{D}^{\frac{1}{2}}$: p×q. Then (1.2.8) holds with $\underset{\sim}{x}$ being a linear combination of q orthogonal factors. The q factors are the first q standardized principal components of $\underset{\sim}{x}$.

If $d_{q+1},\ldots,d_p$ are not zero but small, the the first q principal components approximate $\underset{\sim}{x}$ rather well. Okamoto (1969) has shown that the choice $\underset{\sim}{\Gamma} = \underset{\sim}{A}\underset{\sim}{D}^{\frac{1}{2}}$ minimizes the eigenvalues of $\underset{\sim}{\Sigma} = E(\underset{\sim}{x}-\underset{\sim}{\Gamma}\underset{\sim}{y})(\underset{\sim}{x}-\underset{\sim}{\Gamma}\underset{\sim}{y})'$ and thus its trace and norm, which are reasonable measures of information loss. The matrix $\underset{\sim}{\Sigma}$ may be interpreted as the error variance in the model

$$\underset{\sim}{x} = \underset{\sim}{\Gamma}\underset{\sim}{y} + \underset{\sim}{z} \ , \tag{1.2.9}$$

where $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{z}) = \underset{\sim}{\Sigma}$, $Cov(\underset{\sim}{y},\underset{\sim}{z}) = \underset{\sim}{0}$ and $\underset{\sim}{\Gamma}$ is chosen to minimize tr $\underset{\sim}{\Sigma}$. The covariance structure of $\underset{\sim}{x}$ is then $\underset{\sim}{V} = \underset{\sim}{\Gamma\Gamma}' + \underset{\sim}{\Sigma}$ and $\underset{\sim}{y}$ is given by the first q standardized principal components of $\underset{\sim}{x}-\underset{\sim}{z}$. The fundamental difference between this model and factor analysis is that $\underset{\sim}{\Sigma}$ need not be diagonal. We have thus shown that principal components may be interpreted in terms of factor models with p factors and no errors, or with q factors and correlated errors.

Let us now see in what sense factor analysis may be interpreted in terms of principal components. Let $\underset{\sim}{x}$ satisfy the unrestricted factor model (1.2.1) - (1.2.2)' and consider the random vector $\underset{\sim}{x}-\underset{\sim}{z} = \underset{\sim}{\Lambda y}$. Since $Var(\underset{\sim}{y}) = \underset{\sim}{I}$,

$$Var(\underset{\sim}{x}-\underset{\sim}{z}) = \underset{\sim}{\Lambda\Lambda}' = \underset{\sim}{V}-\underset{\sim}{\Psi} , \tag{1.2.10}$$

a matrix of rank q. Thus a basic feature of the factor model is that the rank of $\underset{\sim}{V}$ can be reduced by subtracting a matrix $\underset{\sim}{\Psi}$ of specific variances.

We now show that under the canonical restrictions that $\underset{\sim}{\Delta} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ be diagonal and have its elements ordered the factors are simply the first q standardized principal components of $\underset{\sim}{\Psi}^{-\frac{1}{2}}(\underset{\sim}{x}-\underset{\sim}{z})$. The scaling by $\underset{\sim}{\Psi}^{-\frac{1}{2}}$ has the effect of making the specific variances unity. Now

$$Var[\underset{\sim}{\Psi}^{-\frac{1}{2}}(\underset{\sim}{x}-\underset{\sim}{z})] = \underset{\sim}{\Psi}^{-\frac{1}{2}}\underset{\sim}{\Lambda\Lambda}'\underset{\sim}{\Psi}^{-\frac{1}{2}} , \tag{1.2.11}$$

by (1.2.10). If $\underset{\sim}{\Delta}$ is diagonal and ordered then $ch_i(\underset{\sim}{\Psi}^{-\frac{1}{2}}\underset{\sim}{\Lambda\Lambda}'\underset{\sim}{\Psi}^{-\frac{1}{2}}) = ch_i(\underset{\sim}{\Delta}) = \delta_i$ $(i=1,\ldots,q)$, and it may be verified that the first q eigenvectors of (1.2.11) are given by $\underset{\sim}{A} = \underset{\sim}{\Psi}^{-\frac{1}{2}}\underset{\sim}{\Lambda\Delta}^{-\frac{1}{2}}$. The first q standardized principal components of $\underset{\sim}{\Psi}^{-\frac{1}{2}}(\underset{\sim}{x}-\underset{\sim}{z})$ are then

$$\underset{\sim}{\Delta}^{-\frac{1}{2}}\underset{\sim}{A}'\underset{\sim}{\Psi}^{-\frac{1}{2}}(\underset{\sim}{x}-\underset{\sim}{z}) = \underset{\sim}{\Delta}^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda y} = \underset{\sim}{y} , \tag{1.2.12}$$

the *canonical* factors. For another discussion of this relationship see Lawley and Maxwell (1971, pp. 7-9).

The scaling by $\underset{\sim}{\Psi}^{-\frac{1}{2}}$ may be omitted and a principal component analysis may be conducted on the matrix (1.2.10) instead of (1.2.11). Since principal components are not scale-invariant, however, this leads to a different basis of the factor space. The first q standardized principal components of $\underset{\sim}{x}-\underset{\sim}{z}$ are called *principal* factors, see Rao (1955) and Harman (1967, Ch. 8).

This completes our review of the factor analysis model. We now consider estimation of the parameters and hypothesis testing in the exploratory and confirmatory cases. For this purpose we introduce the additional assumptions that $\underset{\sim}{\chi} \sim N_q(\underset{\sim}{0},\underset{\sim}{\Phi})$ and $\underset{\sim}{z} \sim N_p(\underset{\sim}{0},\underset{\sim}{\Psi})$ independently of $\underset{\sim}{\chi}$, so that $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu},\underset{\sim}{V})$.

1.3 Exploratory Factor Analysis

1.3.1 Maximum Likelihood Estimation

Let $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_{n+1}$ be a random sample from $N_p(\underset{\sim}{\mu},\underset{\sim}{V})$ where $\underset{\sim}{V}$ satisfies (1.2.2)'. We now consider estimation of $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ under the canonical restrictions. Define the unbiased estimator of $\underset{\sim}{V}$

$$\underset{\sim}{S} = \frac{1}{n}\sum_{\alpha=1}^{n+1}(\underset{\sim}{x}_\alpha-\bar{\underset{\sim}{x}})(\underset{\sim}{x}_\alpha-\bar{\underset{\sim}{x}})' , \tag{1.3.1}$$

and note that $n\underset{\sim}{S} \sim W_p(\underset{\sim}{V},n)$. The log-likelihood function is then

$$\log L = c - \frac{1}{2}n\log|\underset{\sim}{V}| - \frac{1}{2}n\,\mathrm{tr}\,\underset{\sim}{V}^{-1}\underset{\sim}{S} , \tag{1.3.2}$$

where c is a constant including terms on $\underset{\sim}{S}$ but not on $\underset{\sim}{V}$.

Maximizing log L is equivalent to minimizing

$$F(\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) = \log|\underset{\sim}{V}| + \text{tr } \underset{\sim}{V}^{-1}\underset{\sim}{S} - \log|\underset{\sim}{S}| - p \ . \tag{1.3.3}$$

The derivatives of F with respect to $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ are

$$\frac{\partial F}{\partial \underset{\sim}{\Lambda}} = 2\underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{S})\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda} \ , \quad \text{and} \tag{1.3.4}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Psi}_d} = \text{diag } \underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{S})\underset{\sim}{V}^{-1} \ . \tag{1.3.5}$$

It can be shown that given $\underset{\sim}{\Psi}$ the conditional m.l.e. (maximum likelihood estimator) of $\underset{\sim}{\Lambda}$ is

$$\tilde{\underset{\sim}{\Lambda}} = \underset{\sim}{\Psi}^{\frac{1}{2}} \underset{\sim}{\Omega}(\underset{\sim}{D}_\rho - \underset{\sim}{I})^{\frac{1}{2}} \ , \tag{1.3.6}$$

where $\underset{\sim}{D}_\rho = \text{diag}(\rho_1,\ldots,\rho_q)$, $\rho_i = \text{ch}_i(\underset{\sim}{\Psi}^{-\frac{1}{2}}\underset{\sim}{S}\underset{\sim}{\Psi}^{-\frac{1}{2}})$, $\underset{\sim}{\Omega} = (\underset{\sim}{\omega}_1,\ldots,\underset{\sim}{\omega}_q)$ and $\underset{\sim}{\omega}_i$ is the orthonormal eigenvector of $\underset{\sim}{\Psi}^{-\frac{1}{2}}\underset{\sim}{S}\underset{\sim}{\Psi}^{-\frac{1}{2}}$ corresponding to ρ_i $(i=1,\ldots,q)$.

The minimized value of F may be written as

$$\tilde{F}(\underset{\sim}{\Psi}) = \min_{\underset{\sim}{\Lambda}} F(\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) = \sum_{j=q+1}^{p} (\rho_j - \log \rho_j - 1) \ , \tag{1.3.7}$$

and the derivatives of this function with respect to $\underset{\sim}{\Psi}$, obtained evaluating (1.3.5) at $\underset{\sim}{\Lambda} = \tilde{\underset{\sim}{\Lambda}}$, can be written as

$$\frac{\partial F}{\partial \psi_i} = \frac{1}{\psi_i}\left[\sum_{j=1}^{q}(\rho_j - 1)\omega_{ij}^2 + 1 - \frac{s_{ii}}{\psi_i}\right] \ . \tag{1.3.8}$$

The m.l.e. $\hat{\underset{\sim}{\Psi}}$ of $\underset{\sim}{\Psi}$ is computed by numerical minimization of $\tilde{F}(\underset{\sim}{\Psi})$, and the m.l.e. $\hat{\underset{\sim}{\Lambda}}$ of $\underset{\sim}{\Lambda}$ is obtained by evaluating (1.3.6) at $\hat{\underset{\sim}{\Psi}}$. Iterative procedures for the numerical minimization are described in

§1.3.2 and §1.3.3 below. For a derivation of these results see Jöreskog (1967) or Lawley and Maxwell (1971, Ch. 4).

A nice property of the m.l.e.'s $\hat{\underset{\sim}{\Lambda}}$ and $\hat{\underset{\sim}{\Psi}}$ is that they are scale-invariant, see Morrison (1967, p. 268). Also we remark that in the special case $\underset{\sim}{\Psi} = \psi \underset{\sim}{I}_p$, closed-form expressions for the estimates exist, see Lawley and Maxwell (1971, p. 47).

Jöreskog and Goldberger (1972) have considered generalized least squares estimation in factor analysis. Computation of the estimates requires again iterative procedures.

1.3.2 The Fletcher-Powell Method

To obtain the estimates in maximum likelihood factor analysis Jöreskog (1967) has proposed using a numerical method of function minimization due to Davidon (1959) and further developed by Fletcher and Powell (1963).

Given a function $f(\underset{\sim}{\theta})$ depending on a parameter vector $\underset{\sim}{\theta}$ with first derivatives $\underset{\sim}{g}(\underset{\sim}{\theta})$, the method uses a symmetric p.d. matrix $\underset{\sim}{E}$ which is improved on each iteration and eventually converges to the inverse of the matrix of second derivatives $\underset{\sim}{H}(\underset{\sim}{\theta})$ evaluated at the minimum. Let $\underset{\sim}{\theta}^{(s)}$, $\underset{\sim}{g}^{(s)}$ and $\underset{\sim}{E}^{(s)}$ refer to the values of $\underset{\sim}{\theta}$, $\underset{\sim}{g}(\underset{\sim}{\theta})$ and $\underset{\sim}{E}$ at the start of the s-th iteration. The method uses a simple linear search along the direction $-\underset{\sim}{E}^{(s)}\underset{\sim}{g}^{(s)}$ to determine a point with positive gradient, and a cubic interpolation procedure to find $\underset{\sim}{\theta}^{(s+1)}$. Then the matrix $\underset{\sim}{E}^{(s)}$ is improved using

$$\underset{\sim}{E}^{(s+1)} = \underset{\sim}{E}^{(s)} + \frac{1}{\beta^{(s)}} \underset{\sim}{u}^{(s)}\underset{\sim}{u}^{(s)\prime} + \frac{1}{\gamma^{(s)}} \underset{\sim}{v}^{(s)}\underset{\sim}{v}^{(s)\prime}, \qquad (1.3.9)$$

where $\underset{\sim}{u}^{(s)} = \underset{\sim}{\theta}^{(s+1)} - \underset{\sim}{\theta}^{(s)}$, $\underset{\sim}{v}^{(s)} = \underset{\sim}{E}^{(s)}\underset{\sim}{w}^{(s)}$ with $\underset{\sim}{w}^{(s)} = \underset{\sim}{g}^{(s+1)} - \underset{\sim}{g}^{(s)}$, $\beta^{(s)} = \underset{\sim}{u}^{(s)\prime}\underset{\sim}{w}^{(s)}$ and $\gamma^{(s)} = \underset{\sim}{v}^{(s)\prime}\underset{\sim}{w}^{(s)}$. For further details and proofs of the convergence properties of the method see Fletcher and Powell (1963).

In our case $\underset{\sim}{\theta} = (\psi_1,\ldots,\psi_p)'$ and the initial estimate may be taken as $\psi_i^{(1)} = s_{ii}$. A better estimate recommended by Jöreskog (1963) is $\psi_i^{(1)} = (1-q/2p)/s^{ii}$ where s^{ii} is the (i,i)-th element of $\underset{\sim}{S}^{-1}$. The initial matrix $\underset{\sim}{E}$ may be taken simply as the identity matrix, in which case the first iteration is in the direction of steepest descent.

Lawley (1967) has given the following approximate second derivatives of $\tilde{F}(\underset{\sim}{\Psi})$ based on large sample considerations

$$\underset{\sim}{H} = \left(\frac{\partial^2\tilde{F}}{\partial\psi_i\partial\psi_j}\right) \doteq \underset{\sim}{G} = (\zeta_{ij}^2) \qquad (1.3.10)$$

where ζ_{ij} is the (i,j)-th element of $\underset{\sim}{\Psi}^{-\frac{1}{2}}(\underset{\sim}{I} - \underset{\sim}{\Omega}\underset{\sim}{\Omega}')\underset{\sim}{\Psi}^{-\frac{1}{2}}$. The matrix $\underset{\sim}{G}$ is p.d. for all $(p-q)^2 > p+q$ and $\underset{\sim}{G}^{-1}$ provides a good initial value for $\underset{\sim}{E}$. In practice, to speed up convergence it is recommended to start with two steepest descent iterations and then compute $\underset{\sim}{G}$ (which depends on $\underset{\sim}{\Psi}$) and switch to the Fletcher-Powell method with $\underset{\sim}{E}^{(1)} = \underset{\sim}{G}^{-1}$.

A difficulty that arises in applying this method is that elements of $\underset{\sim}{\Psi}$ may become negative during iteration. To avoid this the minimization is done under the restriction that $\psi_i > \varepsilon$ $(i=1,\ldots,p)$ where ε is a small positive constant, usually .005 or .001. If a value of ψ_i becomes $< \varepsilon$, it is changed to ε before evaluating the function. If its derivative at ε turns out to be positive, ψ_i is fixed at ε and minimization is continued with respect to the remaining elements of $\underset{\sim}{\Psi}$. Sets of estimates where $\hat{\psi}_i = \varepsilon$ for some i are termed improper

or Heywood cases. In these situations the variates with $\hat{\psi}_i = \varepsilon$ are partialled-out and a new model is fitted to the partial covariance matrix for the remaining variables. For further details see Jöreskog (1967).

1.3.3 The Newton-Raphson Method

Clarke (1970) has recently derived exact expressions for the second derivatives of $\tilde{F}$, which enabled him to use the Newton-Raphson method of function minimization. In the notation of §1.3.1, his result is

$$h_{ij} = \frac{\partial^2 \tilde{F}}{\partial \psi_i \partial \psi_j} = \zeta_{ij}^2 - \frac{2\delta_{ij}}{\psi_i} \frac{\partial \tilde{F}}{\partial \psi_i} + \sum_{k=1}^{q} \omega_{ik}\omega_{jk} \sum_{\ell=q+1}^{p} \frac{2\rho_k(\rho_k-1)}{\rho_\ell-\rho_k} \omega_{i\ell}\omega_{j\ell} , \tag{1.3.11}$$

where δ_{ij} is the Kronecker delta.

If $\underset{\sim}{\theta}^{(s)}$ is the s-th approximation to the minimum point, then the Newton-Raphson method takes

$$\underset{\sim}{\theta}^{(s+1)} = \underset{\sim}{\theta}^{(s)} - \underset{\sim}{H}^{-1}[\underset{\sim}{\theta}^{(s)}]\underset{\sim}{g}[\underset{\sim}{\theta}^{(s)}]. \tag{1.3.12}$$

The exact matrix of second derivatives $\underset{\sim}{H}$ is used in most iterations. Lawley's (1967) approximation $\underset{\sim}{G}$ is used (a) for the first iteration, (b) if $\max_i |\psi_i^{(s+1)} - \psi_i^{(s)}| > .1$, or (c) if $\underset{\sim}{H}[\underset{\sim}{\psi}^{(s)}]$ is not p.d. The latter condition is required because $\underset{\sim}{H}$ is p.d. only in an indeterminate neighborhood of the minimum. The first two conditions serve to improve convergence. Improper values of $\underset{\sim}{\Psi}$ are handled in the same manner as described above.

The Newton-Raphson method usually requires fewer iterations than the Fletcher-Powell method, but necessitates a somewhat greater amount of computation on each iteration. Current experience indicates that it

is more efficient unless p is fairly large.

1.3.4 Large Sample Standard Errors

Anderson and Rubin (1956) have proved that $\sqrt{n}(\hat{\underset{\sim}{\Lambda}}-\underset{\sim}{\Lambda})$ and $\sqrt{n}(\hat{\underset{\sim}{\Psi}}-\underset{\sim}{\Psi})$ are asymptotically normally distributed when $\underset{\sim}{S} \xrightarrow{p} \underset{\sim}{V}$ and $\sqrt{n}(\underset{\sim}{S}-\underset{\sim}{V})$ is asymptotically normal, a condition that is satisfied when $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu},\underset{\sim}{V})$. Derivation of asymptotic variances and covariances, however, is very involved.

Lawley (1953) has shown that if $\underset{\sim}{\Psi}$ is known, for sufficiently large n

$$\text{Var}(\hat{\underset{\sim}{\Lambda}}|\underset{\sim}{\Psi}) \doteq \frac{1}{n}\underset{\sim}{A} , \tag{1.3.13}$$

where $\underset{\sim}{A}$ is a matrix with element $a_{ir,js}$ in row $[p(r-1)+i]$ and column $[p(s-1)+j]$ and

$$a_{ir,js} = \begin{cases} \mu_r[v_{ij}^{-1}\mu_r\lambda_{ir}\lambda_{jr} + \sum_{\substack{m=1 \\ \neq r}}^{q}(\mu_m\gamma_{im}\lambda_{im}\lambda_{jm})], & r=s \\ -[\rho_r\rho_s/(\rho_r-\rho_s)]^2\lambda_{is}\lambda_{jr}, & r\neq s , \end{cases} \tag{1.3.14}$$

with $\mu_r = \rho_r/(\rho_r-1)$, $\gamma_{rm} = [(\rho_r-1)/(\rho_r-\rho_m)]^2 - 1$, and ρ_r as defined in (1.3.6).

Several years later, Lawley (1967) showed that if $\underset{\sim}{\Psi}$ is estimated then for large n

$$\text{Var}(\hat{\underset{\sim}{\Psi}}) \doteq \frac{2}{n}\underset{\sim}{G}^{-1} \tag{1.3.15}$$

$$\text{Var}(\hat{\underset{\sim}{\Lambda}}) \doteq \frac{1}{n}(\underset{\sim}{A} + 2\underset{\sim}{B}\underset{\sim}{G}^{-1}\underset{\sim}{B}) , \text{ and} \tag{1.3.16}$$

$$\text{Cov}(\hat{\underset{\sim}{\Psi}},\hat{\underset{\sim}{\Lambda}}) \doteq \frac{2}{n}\underset{\sim}{G}^{-1}\underset{\sim}{B} , \tag{1.3.17}$$

where $\underset{\sim}{G}$ is as defined in (1.3.10), $\underset{\sim}{B}$ is a matrix with $\underset{\sim}{b}_{ir}$ in

column $[p(r-1)+i]$, $\underset{\sim}{b}_{ir} = (b_{1,ir},\ldots,b_{p,ir})'$, and

$$b_{j,ir} = -\lambda_{jr}(\rho_r-1)^{-1}\psi_j^{-2}[\delta_{ij}\psi_j^{-\frac{1}{2}}\lambda_{ir}\lambda_{jr}/(\rho_r-1) + \rho_r \sum_{\substack{m=1 \\ \neq r}}^{q} \lambda_{im}\lambda_{jm}/(\rho_r-\rho_m)]. \tag{1.3.18}$$

For a derivation of these results see Lawley and Maxwell (1971, Ch. 5), except for expression (1.3.18) which is in error in the original and has been corrected by Jennrich and Thayer (1973).

If the loadings are rotated further complications arise. Lawley and Maxwell suggest treating the transformation matrix $\underset{\sim}{M}$ as known and thus the rotated loadings as linear functions of the original loadings. In practice, however, $\underset{\sim}{M}$ is usually derived from the data. Archer and Jennrich (1973) and Jennrich (1973) have obtained results for rotated loadings.

Recently, Jennrich (1974) has obtained simplified results for the asymptotic variances and covariances by approaching the problem as one in constrained maximum likelihood estimation. Let $\hat{\underset{\sim}{\theta}}_n$: m×1 be the m.l.e. of a parameter $\underset{\sim}{\theta}$ which is assumed to satisfy constraints $\underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0}$, and let $\underset{\sim}{I}(\underset{\sim}{\theta})$ be the Fisher information matrix (which will generally be singular). Define the *augmented information matrix*

$$\underset{\sim}{I}^*(\underset{\sim}{\theta}) = \begin{bmatrix} \underset{\sim}{I}(\underset{\sim}{\theta}) & \partial \underset{\sim}{g}/\partial \underset{\sim}{\theta} \\ (\partial \underset{\sim}{g}/\partial \underset{\sim}{\theta})' & \underset{\sim}{0} \end{bmatrix}, \tag{1.3.19}$$

and let $\underset{\sim}{I}^-(\underset{\sim}{\theta})$ be the matrix in the upper left m×m block of $\underset{\sim}{I}^{*-1}(\underset{\sim}{\theta})$. Then under certain regularity conditions

$$\sqrt{n}\,(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} N_m[\underset{\sim}{\theta}, \underset{\sim}{I}^-(\underset{\sim}{\theta})]\ . \tag{1.3.20}$$

For a proof of this result see Silvey (1971, p. 81). We require that

$\underset{\sim}{I}(\underset{\sim}{\theta})$ and $\partial g/\partial\underset{\sim}{\theta}$ exist and be continuous in a neighborhood of $\underset{\sim}{\theta}$, that $\underset{\sim}{I}^{*-1}(\underset{\sim}{\theta})$ exist and that $\hat{\underset{\sim}{\theta}}_n$ be consistent.

In exploratory factor analysis the elements of the information matrix are given by

$$I(\lambda_{ij},\lambda_{k\ell}) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj}, \quad (1.3.21)$$

$$I(\lambda_{ij},\psi_k) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj} \quad \text{and} \quad (1.3.22)$$

$$I(\psi_i,\psi_k) = \frac{1}{2}(\underset{\sim}{V}^{-1})^2_{ik} \; . \quad (1.3.23)$$

The constraint functions associated with the canonical restrictions are

$$g_{uv}(\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) = (\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda})_{uv} \quad (1.3.24)$$

for $1 \le u < v \le q$, and have derivatives

$$\frac{\partial g_{uv}}{\partial\lambda_{ij}} = (\delta_{ju}\lambda_{iv} + \delta_{jv}\lambda_{iu})/\psi_i \quad , \text{ and} \quad (1.3.25)$$

$$\frac{\partial g_{uv}}{\partial\psi_k} = -\lambda_{ku}\lambda_{kv}/\psi_k^2 \; . \quad (1.3.26)$$

These derivatives can be arranged into a vector by ordering the subscripts in lexicographical fashion. The augmented information matrix is then

$$\underset{\sim}{I}^{*}(\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) = \begin{bmatrix} \underset{\sim}{I}(\underset{\sim}{\Lambda},\underset{\sim}{\Lambda}) & \underset{\sim}{I}(\underset{\sim}{\Lambda},\underset{\sim}{\Psi}) & \partial\underset{\sim}{g}/\partial\underset{\sim}{\Lambda} \\ & \underset{\sim}{I}(\underset{\sim}{\Psi},\underset{\sim}{\Psi}) & \partial\underset{\sim}{g}/\partial\underset{\sim}{\Psi} \\ \text{sym.} & & \underset{\sim}{0} \end{bmatrix} , \quad (1.3.27)$$

and $\underset{\sim}{I}^{-}(\underset{\sim}{\Lambda},\underset{\sim}{\Psi})$ is obtained inverting (1.3.27). Consistent estimators of the asymptotic variances and covariances are obtained substituting the m.l.e.'s $\hat{\underset{\sim}{\Lambda}}_n,\hat{\underset{\sim}{\Psi}}_n$ for $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$ in (1.3.27). The results have been

found to agree with results obtained using Lawley's formulae (after correction). Although the method is computationally less efficient, the formulae are pleasingly simple. Furthermore, the method can easily be applied to analytically rotated loadings by modifying the constraint functions (1.3.24). For further details see Jennrich (1974).

The results discussed so far apply to estimates derived from a variance-covariance matrix $\underset{\sim}{S}$. The modifications required when m.l.e.'s are obtained from a correlation matrix may be found in Lawley and Maxwell (1971, Ch. 5) and Jennrich (1974).

1.3.3 Hypothesis Testing

Let Ω denote the set of all symmetric p.d. matrices of order p and let ω denote the subset for which (1.2.2)' holds. We consider testing H_0: $\underset{\sim}{V}\epsilon\omega$ vs. H_1: $\underset{\sim}{V}\epsilon\Omega-\omega$, the goodness of fit of the model.

It is well known that the unrestricted m.l.e. of $\underset{\sim}{V}$ is $\underset{\sim}{S}$ and

$$\max_{\Omega} \log L = -\frac{1}{2}n(\log|\underset{\sim}{S}| + p) . \tag{1.3.28}$$

The m.l.e. of $\underset{\sim}{V}$ under (1.2.2)' is $\hat{\underset{\sim}{V}}_q = \hat{\underset{\sim}{\Lambda}}\hat{\underset{\sim}{\Lambda}}' + \hat{\underset{\sim}{\Psi}}$, where the subscript q emphasizes the number of factors fitted. It can be shown that

$$\max_{\omega} \log L = -\frac{1}{2}n(\log|\hat{\underset{\sim}{V}}_q| + p) , \tag{1.3.29}$$

and hence the goodness of fit likelihood ratio test statistic is

$$-2 \log \lambda = n(\log|\hat{\underset{\sim}{V}}_q| - \log|\underset{\sim}{S}|) , \tag{1.3.30}$$

which can be shown to be simply $n\tilde{F}(\hat{\underset{\sim}{\Psi}})$, hence the choice of F in (1.3.3).

The asymptotic distribution of $-2 \log \lambda$ is χ^2_ν with degrees of

freedom

$$\nu = \frac{1}{2}p(p-1) - [pq+p - \frac{1}{2}q(q-1)] = \frac{1}{2}[(p-q)^2 - (p+q)] , \qquad (1.3.31)$$

the number of parameters in $\underset{\sim}{V}$ minus the number of free parameters in $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$, see Lawley and Maxwell (1971, Ch. 4).

If $q = 0$ then $\hat{\underset{\sim}{V}}_0 = \text{diag } \underset{\sim}{S}$ and (1.3.30) reduces to the well-known likelihood ratio test of independence, see Anderson (1958, Ch. 9). In this case Box (1949) has shown that the χ^2 approximation is improved if n is replaced by $n - (2p+5)/6$ in (1.3.30). Similar corrections for $q > 0$ have not been established, but Bartlett (1951) has suggested using Box's correction with n and p replaced by $n-q$ and $p-q$.

From (1.3.29) it can be seen that the likelihood ratio statistic for testing q versus $q+1$ factors is

$$-2 \log \lambda = n \, (\log|\hat{\underset{\sim}{V}}_q| - \log|\hat{\underset{\sim}{V}}_{q+1}|) , \qquad (1.3.32)$$

and $-2 \log\lambda \xrightarrow{d} \chi^2_\nu$ with degrees of freedom

$$\nu = [p(q+1) + p - \frac{1}{2}(q+1)] - [pq + p - \frac{1}{2}q(q-1)] = p-q , \qquad (1.3.33)$$

the difference between the number of free parameters in $(q+1)$- and q-factor models.

1.4 Confirmatory Factor Analysis

1.4.1 Maximum Likelihood Estimation

We now consider estimation of $\underset{\sim}{\Lambda}, \underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ under a structural hypothesis. Parameters will be of three types: (1) free or unconstrained parameters, (2) parameters that are constrained to be equal to other

parameters in the model, and (3) fixed parameters that have been assigned given values. For example, if $\underset{\sim}{\Lambda} = \lambda \underset{\sim}{1}_p$ in a one-factor model, we may treat λ_1 as fixed and $\lambda_2,\ldots,\lambda_p$ as constrained to be equal to λ_1. This set-up is more general than the set-up of Jöreskog (1969) or Lawley and Maxwell (1971, Ch. 7), who consider only free and fixed parameters, and is in the spirit of more general models considered by Jöreskog (1970a).

As in §1.3.1 we will minimize the function

$$F(\underset{\sim}{\Lambda},\underset{\sim}{\Phi},\underset{\sim}{\Psi}) = \log|\underset{\sim}{V}| + \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{S} - \log|\underset{\sim}{S}| - p \ , \tag{1.4.1}$$

where $\underset{\sim}{V}$ satisfies (1.2.2) and a structural hypothesis.

The derivatives of F with respect to $\underset{\sim}{\Lambda},\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are

$$\frac{\partial F}{\partial \underset{\sim}{\Lambda}} = 2\underset{\sim}{\Omega}\underset{\sim}{\Lambda}\underset{\sim}{\Phi} \ , \tag{1.4.2}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Phi}_s} = 2\underset{\sim}{\Lambda}'\underset{\sim}{\Omega}\underset{\sim}{\Lambda} - \operatorname{diag} \underset{\sim}{\Lambda}'\underset{\sim}{\Omega}\underset{\sim}{\Lambda} \ , \quad \text{and} \tag{1.4.3}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Psi}} = \operatorname{diag} \ \underset{\sim}{\Omega} \ , \tag{1.4.4}$$

where

$$\underset{\sim}{\Omega} = \underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{S})\underset{\sim}{V}^{-1} \ . \tag{1.4.5}$$

Let the elements of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ be arranged into a vector

$$\gamma = (\lambda_{11},\ldots,\lambda_{p1};\ldots;\lambda_{1q},\ldots,\lambda_{pq};\phi_{11};\phi_{12},\phi_{22};$$
$$\ldots;\phi_{1q},\ldots,\phi_{qq};\psi_1,\ldots,\psi_p)' \ , \tag{1.4.6}$$

and let $\underset{\sim}{\theta}$: m×1 contain the free parameters in γ. Then

$$\frac{\partial F}{\partial \theta_i} = \sum_j \alpha_{ij} \ \frac{\partial F}{\partial \gamma_j} \ , \tag{1.4.7}$$

where $\alpha_{ij} = 1$ if $\theta_i = \gamma_j$ and $\alpha_{ij} = 0$ otherwise, and $\partial F/\partial \gamma_j$

may be obtained from (1.4.2) - (1.4.4).

The function F is now considered a function of $\underset{\sim}{\theta}$ and may be minimized using the Fletcher-Powell method described in §1.3.2. The initial matrix $\underset{\sim}{E}$ may be taken simply as the identity matrix. A better choice is provided by half the inverse of the information matrix, as explained in §1.4.2. Since initial estimates may be far from the minimum point and the information matrix depends on $\underset{\sim}{\theta}$, it is convenient to start with several steepest descent iterations and then switch to the Fletcher-Powell method. For further details and a derivation of these results see Jöreskog (1969) or Lawley and Maxwell (1971, Ch. 7).

1.4.2 Large Sample Standard Errors

By standard asymptotic theory, see Kendall and Stuart (1961, pp. 54-55), it can be shown that $\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}) \xrightarrow{d} N_m[\underset{\sim}{0}, \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})]$, where $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the Fisher information matrix. The elements of the information matrix with respect to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are given by

$$I(\lambda_{ij},\lambda_{k\ell}) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{j\ell} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{i\ell}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{kj} , \quad (1.4.8)$$

$$I(\lambda_{ij},\phi_{k\ell}) = \tfrac{1}{2}(2-\delta_{k\ell})[(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{\ell j} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{kj}], \quad (1.4.9)$$

$$I(\lambda_{ij},\psi_k) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Phi})_{kj} , \quad (1.4.10)$$

$$I(\phi_{ij},\phi_{k\ell}) = \tfrac{1}{4}(2-\delta_{ij})(2-\delta_{k\ell})[(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{jk}] , \quad (1.4.11)$$

$$I(\phi_{ij},\psi_k) = \tfrac{1}{2}(2-\delta_{ij})(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj} , \quad (1.4.12)$$

$$I(\psi_i,\psi_k) = \tfrac{1}{2}(\underset{\sim}{V}^{-1})^2_{ik} . \quad (1.4.13)$$

These results were derived independently by Lawley (1967) and Lockhart (1967). These authors, however, assume that diag $(\underset{\sim}{\Phi}) = \underset{\sim}{I}_q$ and hence do not give the Kronecker factors in (1.4.9), (1.4.11) and (1.4.13). See also Jöreskog (1969) and Lawley and Maxwell (1971, Ch. 7).

The information matrix with respect to the free parameters is simply

$$I(\theta_i,\theta_j) = \sum_k \sum_\ell \alpha_{ik}\alpha_{j\ell} I(\gamma_k,\gamma_\ell) \ , \tag{1.4.14}$$

where α_{ik} is 1 if $\theta_i = \gamma_j$ and 0 otherwise, and the $I(\gamma_k,\gamma_\ell)$ are obtained from (1.4.8) - (1.4.13).

Since log L is, except for a constant, $-\frac{1}{2}$ nF, we have

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = -E\left[\frac{1}{n}\frac{\partial^2 \log L}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'}\right] = \frac{1}{2}E\left[\frac{\partial^2 F}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'}\right] \ . \tag{1.4.15}$$

Expected values of the second derivatives of F are thus given by $2\underset{\sim}{I}(\underset{\sim}{\theta})$. If $\underset{\sim}{\tilde{\theta}}$ is a preliminary estimate of $\underset{\sim}{\theta}$ such that $\underset{\sim}{I}(\underset{\sim}{\tilde{\theta}})$ is non-singular, then the matrix $\frac{1}{2}\underset{\sim}{I}^{-1}(\underset{\sim}{\tilde{\theta}})$ can be used as a good initial value for the matrix $\underset{\sim}{E}$ in the Fletcher-Powell method.

Mulaik (1971) has derived exact expressions for the second derivatives of F with respect to the elements of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, which could be used in a Newton-Raphson algorithm for minimizing F. The expressions, however, are too complicated to give here. Furthermore, the large number of parameters usually involved in confirmatory factor analysis makes the Fletcher-Powell method generally more efficient than Newton-Raphson.

1.4.3 Hypothesis Testing

The hypothesis that $\underset{\sim}{V}$ has the structure (1.2.2) with the specified restrictions regarding free, fixed and constrained parameters may be tested in large samples using the likelihood ratio goodness of fit statistic

$$-2 \log \lambda = n[\log|\hat{\underset{\sim}{V}}| + \mathrm{tr}(\hat{\underset{\sim}{V}}^{-1}\underset{\sim}{S}) - \log|\underset{\sim}{S}| - p] , \qquad (1.4.16)$$

where $\hat{\underset{\sim}{V}} = \hat{\underset{\sim}{\Lambda}}\hat{\underset{\sim}{\Phi}}\hat{\underset{\sim}{\Lambda}}' + \hat{\underset{\sim}{\Psi}}$, which is asymptotically distributed as χ^2_ν with

$$\nu = \frac{1}{2} p(p-1) - m , \qquad (1.4.17)$$

where m is the number of free parameters. As in §1.3.5, the test statistic is simply n times $F(\hat{\underset{\sim}{\Lambda}},\hat{\underset{\sim}{\Phi}},\hat{\underset{\sim}{\Psi}})$, the minimum value of F. Multipliers other than n in (1.4.16) which may improve convergence in law to χ^2_ν have not been established. Lawley and Maxwell (1971, p. 93) take $n-(2p+5)/6$, which is appropriate if $q = 0$.

1.5 Related Models

We now briefly consider two extensions of factor analysis which are relevant to our work.

1.5.1 Factor Analysis in Several Populations

Several authors have proposed extensions of factor analysis to several populations, see for example Rash (1953), Lawley and Maxwell (1971, Ch. 9), Jöreskog (1971a) and more recently Please (1973). In the general case one can write for the j-th population (j=1,...,k)

$$\underset{\sim}{x}^{(j)} = \underset{\sim}{\mu}^{(j)} + \underset{\sim}{\Lambda}^{(j)}\underset{\sim}{\chi}^{(j)} + \underset{\sim}{z}^{(j)} , \qquad (1.5.1)$$

where $\underset{\sim}{x}^{(j)}$ is a vector of responses, $\underset{\sim}{\mu}^{(j)}$ is a vector of means, $\underset{\sim}{\Lambda}^{(j)}$

is a matrix of loadings, $E[\underset{\sim}{\chi}^{(j)}] = \underset{\sim}{\xi}^{(j)}$, $Var[\underset{\sim}{\chi}^{(j)}] = \underset{\sim}{\Phi}^{(j)}$ p.d., $E[\underset{\sim}{z}^{(j)}] = \underset{\sim}{0}$, $Var[\underset{\sim}{z}^{(j)}] = \underset{\sim}{\Psi}^{(j)}$ with $\underset{\sim}{\Psi}^{(j)}$ a diagonal p.d. matrix and $Cov[\underset{\sim}{\chi}^{(j)}, \underset{\sim}{z}^{(j)}] = \underset{\sim}{0}$. Clearly this general model is not identified.

Lawley and Maxwell (1971, Ch. 9) consider the case $k = 2$ with $\underset{\sim}{\xi}^{(j)} = \underset{\sim}{0}$, $\underset{\sim}{\mu}^{(j)} = \underset{\sim}{0}$, $\underset{\sim}{\Lambda}^{(j)} = \underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}^{(j)} = \underset{\sim}{\Psi}$ (j=1,2), set the diagonal elements of $\underset{\sim}{\Phi}^{(1)}$ to unity, and derive maximum likelihood equations for estimating $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Psi}$, $\underset{\sim}{\Phi}^{(1)}$ and $\underset{\sim}{\Phi}^{(2)}$ under normality assumptions.

Jöreskog (1971a) considers the case of general k with $\underset{\sim}{\xi}^{(j)} = \underset{\sim}{0}$ and requires that q^2 independent constraints be imposed on $\underset{\sim}{\Lambda}^{(j)}$ and $\underset{\sim}{\Phi}^{(j)}$ for j=1,...,k. Actually his model is more general in that the factors and responses need not be the same in all populations. Jöreskog (1971a) also studies in greater detail the case where $\underset{\sim}{\Lambda}^{(j)} = \underset{\sim}{\Lambda}$ (j=1,...,k).

Please (1973) considers the case of general k with $\underset{\sim}{\mu}^{(j)} = \underset{\sim}{0}$ and $\underset{\sim}{\Lambda}^{(j)} = \underset{\sim}{\Lambda}$ (j=1,...,k), and derives the likelihood equations under the set of restrictions $\underset{\sim}{\Lambda}'\underset{\sim}{\Lambda} = \underset{\sim}{I}_q$. While Jöreskog emphasizes the covariance structure ignoring the means, Please considers the factor means also. Note that the models are heteroscedastic in that $Var[\underset{\sim}{x}^{(j)}]$ need not be the same in all populations.

In all these models the Fletcher-Powell method is used to compute the estimates.

1.5.2 Analysis of Covariance Structures

Jöreskog (1970a) has proposed a general model for the analysis of covariance structures which can be written as

$$\underset{\sim}{X} = \underset{\sim}{P\Xi A} + \underset{\sim}{U} \qquad (1.5.2)$$

where $\underset{\sim}{X}$: p×n is a matrix of n observations on p variates, $\underset{\sim}{P}$: p×g and $\underset{\sim}{A}$: h×n are design matrices of known constants of rank g ≤ p and h < n respectively, $\underset{\sim}{\Xi}$: g×h is a matrix of unknown parameters and $\underset{\sim}{U}$ is a stochastic matrix of errors with $E(\underset{\sim}{U}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{U}) = \underset{\sim}{V} \otimes \underset{\sim}{I}_n$, where the variance-covariance matrix $\underset{\sim}{V}$ has the structure

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}(\underset{\sim}{\Gamma}\underset{\sim}{\Sigma}\underset{\sim}{\Gamma}' + \underset{\sim}{T})\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} \tag{1.5.3}$$

where $\underset{\sim}{\Lambda}$: p×q and $\underset{\sim}{\Gamma}$: q×r are matrices of factor loadings, $\underset{\sim}{\Sigma}$: r×r is a p.d. matrix of factor covariances and $\underset{\sim}{T}$: q×q and $\underset{\sim}{\Psi}$: p×p are diagonal p.d. matrices of specific variances. The model (1.5.2) without the parametric structure of $\underset{\sim}{V}$ is the growth curve model of Potthoff and Roy (1964).

There is a great deal of indeterminacy in the model, for if $\underset{\sim}{T}_1$ is a non-singular matrix of order p such that $\underset{\sim}{T}_1\underset{\sim}{T}\underset{\sim}{T}_1'$ is diagonal and $\underset{\sim}{T}_2$ is any non-singular matrix of order q, then $\underset{\sim}{\Lambda}$ may be replaced by $\underset{\sim}{\Lambda}\underset{\sim}{T}_1^{-1}$, $\underset{\sim}{\Gamma}$ by $\underset{\sim}{T}_1\underset{\sim}{\Gamma}\underset{\sim}{T}_2^{-1}$, $\underset{\sim}{\Sigma}$ by $\underset{\sim}{T}_2\underset{\sim}{\Sigma}\underset{\sim}{T}_2'$ and $\underset{\sim}{T}$ by $\underset{\sim}{T}_1\underset{\sim}{T}\underset{\sim}{T}_1'$ in (1.5.3) leaving $\underset{\sim}{V}$ unaffected. In order to achieve identifiability some restrictions must be imposed on the parameters.

Jöreskog (1970a) allows the parameters in $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Gamma}$, $\underset{\sim}{\Sigma}$, $\underset{\sim}{T}$, and $\underset{\sim}{\Psi}$ to be (1) fixed at given values, (2) constrained to be equal to other parameters, or (3) free, unconstrained. This generates a variety of interesting special cases, including models for the analysis of congeneric measurements, factor analysis, Wiener and Markov process models for repeated measurements, and some special cases of path analysis.

The likelihood equations for estimating the parameters when

$\underset{\sim}{U} \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ are given in Jöreskog (1970a). The estimates are again obtained using the Fletcher-Powell method.

Several applications of the model are considered in Jöreskog (1970b, 1971b). The review paper by Mukherjee (1973) compares the factor analysis and covariance structure models. Recently, Browne (1974b) has considered generalized least squares estimation in the analysis of linear and non-linear covariance structures, and has proved several optimality properties of the estimators.

II. THE MULTIVARIATE LATENT LINEAR MODEL

2.1 Introduction

In this chapter we propose a generalization of the factor analysis model which obtains by letting the factors satisfy a multivariate linear model. The resulting model is termed the multivariate *latent* linear model, and provides a general framework for the statistical analysis of unobservable constructs of latent variables on the basis of observable indicators or manifest variables. In §2.2 we provide a formal statement of the model and note that it is analogous in its method of construction to growth curve and covariance structure models. In §2.3 we consider the identification problem and the types of restrictions that must be imposed on the parameters to achieve identification. In §2.4 we consider some special cases and applications of the model, and discuss its relationship with factor analysis in several populations, path analysis, and variance components models.

2.2 Statement of the Model

Consider the multivariate general *linear model*

$$\underset{\sim}{Y} = \Xi\underset{\sim}{A} + \underset{\sim}{\varepsilon} , \tag{2.2.1}$$

where $\underset{\sim}{Y}$: $q \times n$ is a stochastic matrix of n observations on q variates, Ξ: $q \times r$ is a matrix of unknown regression parameters, $\underset{\sim}{A}$: $r \times n$ is a design

matrix of full row rank $r < n$ and $\underset{\sim}{\varepsilon}$ is a stochastic matrix of errors with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ and $\text{Var}(\underset{\sim}{\varepsilon}) = \underset{\sim}{\Phi} \otimes \underset{\sim}{I}_n$, where $\underset{\sim}{\Phi}$ is a p.d. matrix of unknown variances and covariances. For estimation and testing purposes we will assume that $\underset{\sim}{\varepsilon} \sim N_{q\times n}(\underset{\sim}{0}, \underset{\sim}{\Phi} \otimes \underset{\sim}{I}_n)$.

We are interested in estimating the parameters $\underset{\sim}{\Xi}$ and $\underset{\sim}{\Phi}$, and in testing the multivariate general linear hypothesis

$$H_0: \underset{\sim\sim}{C\Xi B} = \underset{\sim}{0} , \tag{2.2.2}$$

where $\underset{\sim}{C}$: $t\times q$ and $\underset{\sim}{B}$: $r\times s$ are matrices of fixed, known constants.

We suppose, however, that the matrix $\underset{\sim}{Y}$ is not observable. Instead we observe a data matrix $\underset{\sim}{X}$: $p\times n$ with $p > q$ related to $\underset{\sim}{Y}$ by the general *factor analysis model*

$$\underset{\sim}{X} = \underset{\sim\sim}{\Lambda Y} + \underset{\sim}{Z} , \tag{2.2.3}$$

where $\underset{\sim}{\Lambda}$: $p\times q$ is a matrix of unknown parameters called factor *loadings*, of full column rank $q < p$ and $\underset{\sim}{Z}$ is a stochastic matrix of errors with $E(\underset{\sim}{Z}) = \underset{\sim}{0}$, $\text{Var}(\underset{\sim}{Z}) = \underset{\sim}{\Psi} \otimes \underset{\sim}{I}_n$, where $\underset{\sim}{\Psi} = \text{diag}(\psi_1, \ldots, \psi_p)$ is a diagonal p.d. matrix of unknown parameters called *specificities*, and $\text{Cov}(\underset{\sim}{Y}, \underset{\sim}{Z}) = \underset{\sim}{0}$. For estimation and testing purposes we will assume that $\underset{\sim}{Z} \sim N_{p\times n}(\underset{\sim}{0}, \underset{\sim}{\Psi} \otimes \underset{\sim}{I}_n)$ independently of $\underset{\sim}{Y}$.

In this formulation $\underset{\sim}{Y}$ represents unobservable constructs, factors, or latent variables, and $\underset{\sim}{X}$ represents observable indicators, responses or manifest variables. The problem is to estimate the linear model parameters $\underset{\sim}{\Xi}$ and $\underset{\sim}{\Phi}$ and to test linear hypotheses about $\underset{\sim}{\Xi}$ using the data $\underset{\sim}{X}$. We will also be interested in estimating and testing hypotheses about the factor model parameters $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$.

Combining (2.2.1) and (2.2.3) we can write the model in general form

as

$$\underset{\sim}{X} = \underset{\sim}{\beta A} + \underset{\sim}{U} \ , \tag{2.2.4}$$

where $\underset{\sim}{U}$ is a stochastic matrix with $E(\underset{\sim}{U}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{U}) = \underset{\sim}{V} \otimes \underset{\sim}{I}_n$,

$$\underset{\sim}{\beta} = \underset{\sim}{\Lambda\Xi}, \quad \text{and} \tag{2.2.5}$$

$$\underset{\sim}{V} = \underset{\sim}{\Lambda\Phi\Lambda}' + \underset{\sim}{\Psi}. \tag{2.2.6}$$

Under the normality assumptions on the distributions of $\underset{\sim}{\varepsilon}$ and $\underset{\sim}{Z}$,

$$\underset{\sim}{U} \sim N_{p\times n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n) \ . \tag{2.2.7}$$

The proposed model is thus seen to be a multivariate general linear model where both the regression and dispersion parameters are structured. We will call $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ the *reduced form* parameters and $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ the *structural* parameters. The structure of $\underset{\sim}{\beta}$ is analogous to the types of structures found in econometric models, see for example Theil (1971), and the structure of $\underset{\sim}{V}$ is of the factor analysis type discussed in §1.2, with the important feature that the parameter $\underset{\sim}{\Lambda}$ is common to both structures.

Since the basis of the proposed model is a multivariate general linear model on unobservable or latent variables, it will be termed the *multivariate latent linear model.*

The model is analogous in its method of construction to growth curve and covariance structure models. In the growth curve model we assume that

$$E(\underset{\sim}{X}) = \underset{\sim}{P\Theta} \ , \tag{2.2.8}$$

with within-subjects design matrix $\underset{\sim}{P}$: p×q and subject parameters $\underset{\sim}{\Theta}$: q×n, and we assume that the subject parameters in turn satisfy the

linear model

$$\underset{\sim}{\Theta} = \underset{\sim\sim}{\Xi}\underset{\sim}{A} \ , \tag{2.2.9}$$

with structural parameters $\underset{\sim\sim}{\Xi}$:q×r and across-subjects design matrix $\underset{\sim}{A}$: r×n. Thus the growth curve is a second-order or *compounded linear model*, see Grizzle and Allen (1969).

Suppose now that $\underset{\sim}{X}$ is given by a factor model

$$\underset{\sim}{X} = \underset{\sim\sim}{\Lambda}\underset{\sim}{Y} + \underset{\sim}{Z} \ , \tag{2.2.10}$$

where $\underset{\sim}{\Lambda}$: p×q is a matrix of factor loadings, $\underset{\sim}{Z}$ is a stochastic matrix of errors with $E(\underset{\sim}{Z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{Z}) = \underset{\sim}{\Psi} \otimes \underset{\sim}{I}_n$ where $\underset{\sim}{\Psi}$ is a diagonal p.d. matrix of specifities, and $Cov(\underset{\sim}{Y},\underset{\sim}{Z}) = \underset{\sim}{0}$, and suppose that $\underset{\sim}{Y}$ in turn is given by a factor model

$$\underset{\sim}{Y} = \underset{\sim\sim}{\Gamma}\underset{\sim}{F} + \underset{\sim}{\varepsilon} \ , \tag{2.2.11}$$

where $\underset{\sim}{\Gamma}$: q×r is a matrix of loadings, $\underset{\sim}{F}$ is a stochastic matrix of factors with $E(\underset{\sim}{F}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{F}) = \underset{\sim}{\Sigma} \otimes \underset{\sim}{I}_n$ where $\underset{\sim}{\Sigma}$ is p.d., and $\underset{\sim}{\varepsilon}$ is a stochastic matrix of errors with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $Var(\underset{\sim}{\varepsilon}) = \underset{\sim}{T} \otimes \underset{\sim}{I}_n$ where $\underset{\sim}{T}$ is a diagonal p.d. matrix of specificities and $Cov(\underset{\sim}{F},\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$. Then $E(\underset{\sim}{X}) = \underset{\sim}{0}$ and $Var(\underset{\sim}{X}) = \underset{\sim}{V} \otimes \underset{\sim}{I}_n$ where

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}(\underset{\sim\sim\sim}{\Gamma\Sigma\Gamma}' + \underset{\sim}{T})\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} \ , \tag{2.2.12}$$

which is the type of covariance structure assumed in Jöreskog's (1970a) general model described in §1.5.2. Thus the covariance structure model is a second-order or *compounded factor model*.

The latent linear model, on the other hand, may be described as a *compounded linear-factor model*.

It may be noted, incidentally, that the covariance structure (2.2.6) of the latent linear model may be obtained as a special case of Jöreskog's

by setting $\underset{\sim}{\Gamma} = \underset{\sim}{I}_p$ and $\underset{\sim}{T} = \underset{\sim}{0}$ in (2.2.12). Jöreskog (1970a) assumes, however, that $E(\underset{\sim}{X}) = \underset{\sim}{P\Xi A}$ where $\underset{\sim}{P}$ is a known matrix, while from (2.2.4) and (2.2.5) we assume $E(\underset{\sim}{X}) = \underset{\sim}{\Lambda\Xi A}$, where $\underset{\sim}{\Lambda}$ is unknown and appears also in the structure of $\underset{\sim}{V}$.

2.3 The Identification Problem.

We now discuss the identification problem for the model described in §2.2 and consider restrictions on the structural parameters and on the design matrix.

2.3.1 Restrictions on the Structural Parameters.

Let $\underset{\sim}{X}$ satisfy the linear model (2.2.4) with reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ given by (2.2.5) and (2.2.6), respectively. To see that the model is not identified let $\underset{\sim}{L}$ be any non-singular matrix of order q and let

$$\underset{\sim}{\Xi}^* = \underset{\sim}{L\Xi} , \tag{2.3.1}$$

$$\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda L}^{-1} , \text{ and} \tag{2.3.2}$$

$$\underset{\sim}{\Phi}^* = \underset{\sim}{L\Phi L} . \tag{2.3.3}$$

Then $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Phi}$ may be replaced by $\underset{\sim}{\Xi}^*$, $\underset{\sim}{\Lambda}^*$ and $\underset{\sim}{\Phi}^*$ in (2.2.5) and (2.2.6) without modifying the structure of the model nor the reduced form parameters. This indeterminacy of the model corresponds to a non-singular linear transformation of the factors $\underset{\sim}{Y}^* = \underset{\sim}{LY}$. Since L has q^2 elements, at least q^2 constraints need be imposed upon $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Phi}$.

Except for the additional parameter matrix $\underset{\sim}{\Xi}$, the problem is the same as encountered in factor analysis, and can be solved along the same lines as in §1.2.2. We therefore consider unrestricted and restricted versions of the model.

In the *unrestricted case* $\frac{1}{2}q(q+1)$ constraints will be imposed by defining $\underset{\sim}{\Phi} = \underset{\sim}{I}_q$. This implies no loss of generality because if $\underset{\sim}{\Phi} \neq \underset{\sim}{I}_q$ we can introduce the linear transformation $\underset{\sim}{Y}^* = \underset{\sim}{\Phi}^{-\frac{1}{2}}\underset{\sim}{Y}$, where $\underset{\sim}{\Phi}^{\frac{1}{2}}$ is any square root of $\underset{\sim}{\Phi}$, and redefine $\underset{\sim}{\Xi}^* = \underset{\sim}{\Phi}^{-\frac{1}{2}}\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}^{\frac{1}{2}}$ and $\underset{\sim}{\Phi}^* = \underset{\sim}{I}_q$.

If $q > 1$, however, we could still replace $\underset{\sim}{\Xi}$ and $\underset{\sim}{\Lambda}$ by

$$\underset{\sim}{\Xi}^* = \underset{\sim}{M}'\underset{\sim}{\Xi} \quad \text{and} \tag{2.3.4}$$

$$\underset{\sim}{\Lambda}^* = \underset{\sim}{\Lambda}\underset{\sim}{M}\,, \tag{2.3.5}$$

where $\underset{\sim}{M}$ is any orthonormal matrix of order q. In this event, $\frac{1}{2}q(q-1)$ additional constraints are needed to achieve identification.

Proceeding as in factor analysis, we may require $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ to be diagonal with its elements arranged in decreasing order of magnitude, thus obtaining the *canonical basis* of the factor space, see §1.2.3. If $\underset{\sim}{\Lambda}$ is any matrix satisfying (2.2.5) and (2.2.6), then the choice of $\underset{\sim}{M}$ that transforms $\underset{\sim}{\Lambda}$ to satisfy the canonical restrictions is given by the matrix of orthonormal eigenvectors of $\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$, for then

$$\underset{\sim}{\Lambda}^{*\prime}\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}^* = \underset{\sim}{M}'\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{M} = \underset{\sim}{\Delta}\,, \tag{2.3.6}$$

say, where $\underset{\sim}{\Delta} = \operatorname{diag}(\delta_1,\ldots,\delta_q)$ and $\delta_i = \operatorname{ch}_i(\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda})$. Unfortunately, in our case these restrictions do not turn out to be very convenient in maximum likelihood estimation.

An alternative set of constraints is obtained by setting to 0 the first $(j-1)$ elements in the j-th column of $\underset{\sim}{\Lambda}$ for $j=2,\ldots,q$. In this case $\underset{\sim}{\Lambda}$ has a triangular pattern of zeros

$$\underset{\sim}{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 & \ldots & 0 \\ \lambda_{21} & \lambda_{22} & \ldots & 0 \\ & \cdot & \cdot & \cdot \\ \lambda_{q1} & \lambda_{q2} & \ldots & \lambda_{qq} \\ & \cdot & \cdot & \cdot \\ \lambda_{p1} & \lambda_{p2} & \ldots & \lambda_{pq} \end{bmatrix} . \qquad (2.3.7)$$

Let $\underset{\sim}{\Lambda}$ be any matrix satisfying (2.2.5) and (2.2.6), and let $\underset{\sim}{\Lambda}_q$ denote the first q rows of $\underset{\sim}{\Lambda}$. Then the choice of $\underset{\sim}{M}$ that transforms $\underset{\sim}{\Lambda}$ to satisfy (2.3.7) is given by the Gram-Schmidt factorization $\underset{\sim}{\Lambda}_q = \underset{\sim}{T}\underset{\sim}{M}'$, where $\underset{\sim}{T}$ is lower triangular and $\underset{\sim}{M}$ is orthonormal. The matrix $\underset{\sim}{M}$ is a product of Householder transformations and may be computed using the modified Gram-Schmidt algorithm as described in Golub (1969). The resulting basis of the factor space will be called the *Gram-Schmidt basis*, and has the property that the i-th response (i=1,...,q-1) is *not* loaded on factors i+1,...,q, a feature that may be helpful in factor interpretation.

Statistically, one basis is as good as another so long as it allows identification of the model, and may be rotated to satisfy any criteria that may simplify interpretation of the resulting factors. In particular, the Gram-Schmidt basis may be rotated to canonical form.

It may be noted that we have achieved identification by imposing constraints on $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Lambda}$ but not directly on $\underset{\sim}{\Xi}$, although the latter is of course affected by the choice of $\underset{\sim}{\Lambda}$.

In the *restricted case* we will consider a *structural hypothesis* that will usually impose q constraints by requiring $\text{diag}(\underset{\sim}{\Phi}) = \underset{\sim}{I}_q$, and at least an additional $q(q-1)$ constraints by setting to zero $(q-1)$ elements in each column of $\underset{\sim}{\Lambda}$. In the simplest case where only q^2

constraints are imposed in this manner, the common factor space is not truly restricted and may be transformed to an equivalent canonical or Gram-Schmidt form. In a more general case, all parameters in $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ will be allowed to be either: (1) fixed, known parameters, (2) parameters constrained to be equal to other parameters in the model, or (3) free, unconstrained parameters. In this case special care must be exercised in setting up the restrictions to ensure that they are sufficient to identify the model. It may be noted that again we have not imposed constraints directly on $\underset{\sim}{\Xi}$. In this regard we feel that the constraints should be based directly on the nature of the measurement process and hence should refer to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$.

2.3.2 Restrictions on the Design Matrix.

In (2.2.3) we have written the factor analysis part of the model as $\underset{\sim}{X} = \underset{\sim}{\Lambda}\underset{\sim}{Y} + \underset{\sim}{Z}$, without a separate location parameter $\underset{\sim}{\mu}$ for the columns of $\underset{\sim}{X}$, as we had in (1.2.1). In many applications it may be desirable to introduce such a location parameter. In order to do this, however, it is necessary to introduce some restrictions on the design matrix $\underset{\sim}{A}$.

To fix ideas, consider a one-sample problem with $\underset{\sim}{A} = \underset{\sim}{1}'_n$. Let $\underset{\sim}{x}$ and $\underset{\sim}{y}$ denote generically the columns of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$, and write $\underset{\sim}{\Xi} = \underset{\sim}{\xi} = E(\underset{\sim}{y})$. Then if a location parameter $\underset{\sim}{\mu}$ is introduced we have $E(\underset{\sim}{x}) = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\underset{\sim}{\xi}$ and, as we noted in §1.2.3, $\underset{\sim}{\mu}$ and $\underset{\sim}{\xi}$ are not separately identified. In this case we assume $\underset{\sim}{\xi} = \underset{\sim}{0}$ or, equivalently, set $r = 0$ in the linear model.

Consider now a two-sample problem with n observations in each

sample, let $\underset{\sim}{A} = \begin{bmatrix} \underset{\sim}{1}_n' & \underset{\sim}{1}_n' \\ \underset{\sim}{1}_n' & -\underset{\sim}{1}_n' \end{bmatrix}$ and write $\underset{\sim}{\Xi} = (\underset{\sim}{\xi}, \underset{\sim}{\delta})$. Then if a location parameter is introduced we have $E(\underset{\sim}{x}_\alpha) = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\underset{\sim}{\xi} + \underset{\sim}{\Lambda}\underset{\sim}{\delta}$ $(\alpha=1,\ldots,n)$ and $E(\underset{\sim}{x}_\alpha) = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}\underset{\sim}{\xi} - \underset{\sim}{\Lambda}\underset{\sim}{\delta}$ $(\alpha=n+1,\ldots,2n)$, and clearly $\underset{\sim}{\xi}$ and $\underset{\sim}{\mu}$ are not separately identified, for we could replace $\underset{\sim}{\xi}$ and $\underset{\sim}{\mu}$ by $\underset{\sim}{\xi}^*$ and $\underset{\sim}{\mu}^* = \underset{\sim}{\mu} + \underset{\sim}{\Lambda}(\underset{\sim}{\xi}^* - \underset{\sim}{\xi})$ without affecting $E(\underset{\sim}{x}_\alpha)$, $(\alpha=1,\ldots,2n)$. In this case we assume $\underset{\sim}{\xi} = \underset{\sim}{0}$, or equivalently set the linear model with r=1 and $\underset{\sim}{A} = [\underset{\sim}{1}_n', -\underset{\sim}{1}_n']$, and estimate only $\underset{\sim}{\mu}$ and the location difference $\underset{\sim}{\delta}$. From these remarks it is clear that $\underset{\sim}{A}$ should not contain a row of ones; i.e. the model should not specify an overall mean of $\underset{\sim}{\chi}$.

Note now that an alternative parameterization for the two sample problem with r=2 is obtained by setting $\underset{\sim}{A} = \begin{bmatrix} \underset{\sim}{1}_n' & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{1}_n' \end{bmatrix}$ and $\underset{\sim}{\Xi} = (\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2)$. Since this is equivalent to the previous model, it is not possible to identify $\underset{\sim}{\mu}$, $\underset{\sim}{\xi}_1$ and $\underset{\sim}{\xi}_2$, but we can estimate $\underset{\sim}{\mu}$ and $\underset{\sim}{\xi}_1 - \underset{\sim}{\xi}_2$. Although no overall mean of $\underset{\sim}{\chi}$ is specified, such a mean is implicit in $\underset{\sim}{\xi}_1$ and $\underset{\sim}{\xi}_2$.

In the general case, adding a location vector $\underset{\sim}{\mu}$ amounts to considering the model

$$E(\underset{\sim}{X}) = [\underset{\sim}{\mu}, \underset{\sim}{\beta}] \begin{bmatrix} \underset{\sim}{1}_n' \\ \underset{\sim}{A} \end{bmatrix} . \qquad (2.3.8)$$

The reduced form parameters $\underset{\sim}{\mu}$ and $\underset{\sim}{\beta}$ in this model are identified if and only if the augmented design matrix $\begin{bmatrix} \underset{\sim}{1}_n' \\ \underset{\sim}{A} \end{bmatrix}$ is of full row rank. Since $\underset{\sim}{A}$ itself is of full row rank, the only additional requirement is that all rows of $\underset{\sim}{A}$ be linearly independent of $\underset{\sim}{1}_n$.

Estimation of $\underset{\sim}{\mu}$, discussed in §3.2, is considerably simplified by introducing the stronger requirement that $\underset{\approx}{A}$ be orthogonal to $\underset{\sim}{1}_n$, i.e.

$$\underset{\approx}{A}\underset{\sim}{1}_n = \underset{\sim}{0} . \qquad (2.3.9)$$

This entails no loss of generality, for if $\underset{\sim}{\bar{a}} = \frac{1}{n}\underset{\approx}{A}\underset{\sim}{1}_n \neq \underset{\sim}{0}$ we can always replace $\underset{\approx}{A}$ by $\underset{\approx}{A}^* = \underset{\approx}{A} - \underset{\sim}{\bar{a}}\underset{\sim}{1}_n'$ and $\underset{\sim}{\mu}$ by $\underset{\sim}{\mu}^* = \underset{\sim}{\mu} + \underset{\sim}{\beta}\bar{a}$. In practical terms this means that $\underset{\approx}{A}$ should contain only deviations of dummy variates or regressors from their means. The above requirement has been considered in a different context by Puri and Sen (1969).

2.4 Applications of the Model.

We now consider some special cases of the model to illustrate its range of application and some relationships with other models.

2.4.1 Factor Analysis in Several Populations.

In the k-sample problem the latent linear model may be written as

$$\underset{\sim}{x}^{(j)} = \underset{\sim}{\mu} + \underset{\approx}{\Lambda}\underset{\sim}{y}^{(j)} + \underset{\sim}{z}^{(j)} , \qquad (2.4.1)$$

where $\underset{\sim}{x}^{(j)}$ is an observation in the j-th population, $\underset{\sim}{y}^{(j)}$ is a vector of factors with $E[\underset{\sim}{y}^{(j)}] = \underset{\sim}{\xi}^{(j)}$ and $Var[\underset{\sim}{y}^{(j)}] = \underset{\approx}{\Phi}$, and $\underset{\sim}{z}^{(j)}$ is a vector of errors with $E[\underset{\sim}{z}^{(j)}] = \underset{\sim}{0}$, $Var[\underset{\sim}{z}^{(j)}] = \underset{\approx}{\Psi}$ diagonal and $Cov[\underset{\sim}{y}^{(j)}, \underset{\sim}{z}^{(j)}] = \underset{\approx}{0}$, (j=1,...,k). We thus obtain a model of the kind described in §1.5.1.

The models proposed by Jöreskog (1971a) and Please (1973) for this situation emphasize differences in the covariance structure among populations, and thus are more in the spirit of factor analysis, while the present model emphasizes differences in location, and thus is more in the spirit of linear models.

In a sense the model is more restrictive than Please's, for $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are assumed invariant over populations. It is more general, however, in that the parameters of the present model may be estimated under quite general identifying restrictions or structural hypotheses. Also Please (1973) assumes $\underset{\sim}{\mu} = \underset{\sim}{0}$ to obtain m.l.e.'s while, as indicated in §2.3.2, we can estimate $\underset{\sim}{\mu}$ and location differences among populations.

An example where the present model could be applied is if a scale containing 40 questions believed to measure 5 types of attitudes is applied to 4 groups and it is desired to test whether the groups differ on their attitudes. We conjecture that the multivariate latent linear model analysis proposed here is more powerful than a multivariate analysis of variance based on all 40 questions, or an analysis based on scores obtained for each factor by averaging the questions that best measure it.

2.4.2 A Multiple-Cause Multiple-Indicator Causal Model.

Path analysis, introduced by Wright (1918), is a technique frequently used in the social and behavioral sciences as well as in biometry to study causal relationships among variables in a non-experimental situation. See for example Blalock (1961), Boudon (1965), Duncan (1966) and Land (1969) on the social science side, and Turner and Stevens (1959) on the biometry side. The technique is closely related to structural equation models studied by econometricians, see for example Wold (1964) or Theil (1971).

The model proposed here may be regarded as a path analysis model

involving multiple causes of unobservable variates which are assessed using a multiple indicator approach. The causal structure for the case of two unobservable variates is illustrated in Figure 2.4.1.

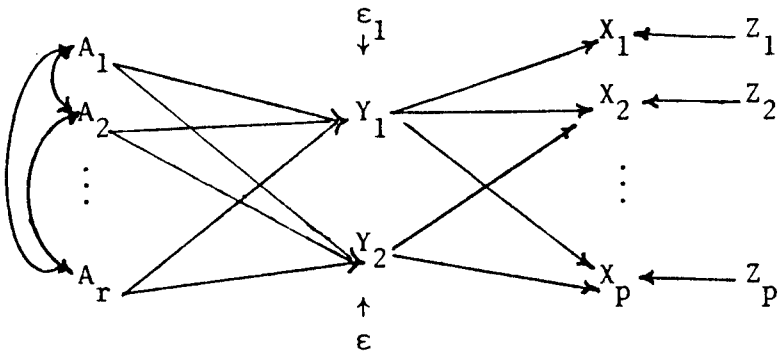

Fig. 2.4.1 A Multiple-Cause Multiple-Indicator Causal Model

The A_i represent causes of the unobservable Y_i and the X_i represent observable indicators. The ε_i and Z_i are uncorrelated error terms. Curved arrows represent correlations among independent variables and straight arrows represent causal relationships. By the nature of the linear model all independent variates (A) are assumed to affect the unobservable dependent variates (Y). By the nature of the factor model, however, some of the paths from unobservable variates (Y) to indicators (X) may be assumed non-existent. The model allows estimation of all regression coefficients and error variances under suitable identification restrictions.

Hauser and Goldberger (1971) consider causal models involving unobservable variates and give a one-factor version of the above model. They state incorrectly, however, that the estimates may be obtained using Jöreskog's (1970a) method for the analysis of covariance structures;

see §1.5.2 and the discussion at the end of §2.2. The present interpretation of the proposed model, however, was inspired by their work.

An example where the model could be applied is when one wishes to study the determinants of a latent variable such as risk of ischemic heart disease using indicators of risk such as blood pressure, serum cholesterol and blood glucose. Which variables are used as determinants and which as indicators would depend, of course, on the purpose of the study.

2.4.3 Variance Components and Mixed Models.

Since most applications of the proposed model are likely to involve multiple measurements of a few underlying characteristics of interest, it is convenient to consider in some detail the analysis of multiple indicator data. Since our interest centers on the measurement process we will ignore for simplicity the underlying linear model.

A model frequently employed when a set of subjects is measured independently by p observers is the variance components model

$$x_{ij} = \alpha + \beta_j + s_i + z_{ij} , \qquad (2.4.2)$$

where x_{ij} is the score assigned to the i-th subject by the j-th observer, α is an overall mean, β_j is a fixed effect due to the j-th observer with $\sum_j \beta_j = 0$, s_i is a random effect due to the i-th subject, with mean 0 and variance σ_s^2, and z_{ij} is an error term with mean 0 and variance σ_e^2. It is assumed that all random effects and errors are mutually uncorrelated. The correlation between any two measurements on the same subject is the intraclass correlation $\rho_c = \sigma_s^2/(\sigma_s^2+\sigma_e^2)$. For a discussion of this and related models see Landis and Koch (1974).

The model (2.4.2) may be rewritten within our framework and notation as

$$x_{ij} = \mu_j + \lambda y_i + z_{ij} \qquad (2.4.3)$$

where $\mu_j = \alpha + \beta_j$, $\lambda = \sigma_s$, $y_i = s_i/\sigma_s$ and $\text{Var}(z_{ij}) = \psi = \sigma_e^2$. The variate y may be interpreted as the characteristic being measured, standardized to have mean zero and variance one. The parameters μ_j and λ may be interpreted as the *intercept* and *slope* of the regression of the j-th measurement on y, and ψ as the corresponding *error variance*. The squared correlation between the j-th measurement and y is $\rho^2 = \lambda^2/(\lambda^2+\psi)$, and is the same as the intraclass correlation between any two measurements.

Note that (2.4.3) may be written in vector form as $\underset{\sim}{x} = \underset{\sim}{\mu} + \underset{\sim}{\lambda 1}_p y + \underset{\sim}{z}$, with $E(y) = 0$, $\text{Var}(y) = 1$, $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $\text{Var}(\underset{\sim}{z}) = \psi \underset{\sim}{I}_p$, $\text{Cov}(y,\underset{\sim}{z}) = \underset{\sim}{0}$ and thus $E(\underset{\sim}{x}) = \underset{\sim}{\mu}$ and $\text{Var}(\underset{\sim}{x}) = \underset{\sim}{V} = \lambda \underset{\sim}{1}_p \underset{\sim}{1}_p' + \psi \underset{\sim}{I}_p$. Thus (2.4.2) or (2.4.3) is seen to be a special case of a one-factor model with $\underset{\sim}{\Lambda} = \lambda \underset{\sim}{1}_p$ and $\underset{\sim}{\Psi} = \psi \underset{\sim}{I}_p$, where the regressions of all measurements on the characteristic of interest have the same slope and error variance, and differ only on their intercepts (biases). More general models may be obtained by relaxing the equal slope assumption, the homoscedasticity assumption, or both. The resulting structures are listed on Table 2.4.1.

In psychological test theory, tests that measure the same characteristic are called *parallel* if they satisfy Model 1 and *tau-equivalent* if they satisfy Model 2, see Lord and Novick (1968) and Jöreskog (1971b).

We have thus seen that the one-factor model provides a quite general model for the multiple indicator case, by allowing not only the intercepts μ_j but also the slopes λ_j and error variances ψ_j $(j=1,\ldots,p)$ in the regression of each measurement on the characteristic being measured to be

Table 2.4.1 Four Measurement Models

Model	$\underset{\sim}{\Lambda}$	$\underset{\sim}{\Psi}$
1. Variance Components	$\lambda\underset{\sim}{1}$	$\psi\underset{\sim}{I}$
2. Heteroscedastic Variance Components	$\lambda\underset{\sim}{1}$	$\underset{\sim}{\Psi}$
3. Homoscedastic Factor Analysis	$\underset{\sim}{\Lambda}$	$\psi\underset{\sim}{I}$
4. Factor Analysis	$\underset{\sim}{\Lambda}$	$\underset{\sim}{\Psi}$

different. In the general case the squared correlation between the i-th and j-th measurements is $\rho_{ij}^2 = \lambda_i^2\lambda_j^2/[(\lambda_i^2+\psi_i)(\lambda_j^2+\psi_j)]$ while that between the i-th measurement and y is $\rho_i^2 = \lambda_i^2/(\lambda_i^2+\psi_i)$, and these need not be the same for all i,j. Furthermore, the basic assumption that all observers measure the same characteristic and the more restrictive assumptions leading to Models 1-3 may be formally tested. These remarks may clearly be extended to multiple factor models.

The model proposed in §2.2 provides a further generalization by allowing the characteristics being measured to satisfy a linear model, so that other variables affecting them may be considered and studied.

III. MAXIMUM LIKELIHOOD ESTIMATION

3.1 Introduction.

We now consider maximum likelihood estimation of the parameters of the latent linear model under normality assumptions. The discussion is organized in five parts. In §3.2 we derive the likelihood equations using the matrix differentiation rules given in the appendix. In §3.3 we discuss the necessary modifications when a separate location parameter $\underset{\sim}{\mu}$ is introduced. In §3.4 we show that when $\underset{\approx}{\Xi}$ is unrestricted its conditional m.l.e. for fixed $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ may be obtained analytically. In §3.5 we consider estimation of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ under certain identifying restrictions. Finally, in §3.6 we describe an iterative procedure for computing the estimates.

3.2 The Likelihood Equations.

Let us write the latent linear model as

$$\underset{\sim}{X} = \underset{\approx}{\beta}\underset{\sim}{A} + \underset{\sim}{U} \tag{3.2.1}$$

where $\underset{\sim}{A}$ is of full row rank $r < n$, $\underset{\sim}{U} \sim N_{p\times n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n)$,

$$\underset{\approx}{\beta} = \underset{\sim}{\Lambda}\underset{\approx}{\Xi}\ , \quad \text{and} \tag{3.2.2}$$

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda} + \underset{\sim}{\Psi} \tag{3.2.3}$$

where $\underset{\sim}{\Lambda}$ is of full column rank $q < p$, $\underset{\sim}{\Phi}$ is symmetric p.d. and $\underset{\sim}{\Psi}$ is diagonal p.d.

The logarithm of the likelihood function may be written as

$$\text{LogL} = -\tfrac{1}{2}np\log(2\pi) - \tfrac{1}{2}n\log|\underset{\sim}{V}| - \tfrac{1}{2}\text{tr}[\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})'] \ . \qquad (3.2.4)$$

Let us use the notation

$$\underset{\sim}{T} = \frac{1}{n}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})' \ . \qquad (3.2.5)$$

Maximizing log L is then equivalent to minimizing the function

$$F = \log|\underset{\sim}{V}| + \text{tr}\ \underset{\sim}{V}^{-1}\underset{\sim}{T} \ , \qquad (3.2.6)$$

which will be used to obtain results directly comparable with factor analysis, see (1.3.3) and (1.4.1).

It may be noted that the likelihood function depends on the structural parameters $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ only through the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and thus remains invariant under transformations of the type discussed in §2.3.1. This shows that the problems of identifiability and estimability are equivalent. Henceforth we will assume that sufficient restrictions have been imposed on the model to identify the parameters and thus achieve estimability.

The first result we need is

Lemma 3.2.1. The derivatives of the function F defined in (3.2.6) with respect to the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ are given by

$$\frac{\partial F}{\partial \underset{\sim}{\beta}} = -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}' \ , \quad \text{and} \qquad (3.2.7)$$

$$\frac{\partial F}{\partial \underset{\sim}{V}} = \underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T})\underset{\sim}{V}^{-1} \ . \qquad (3.2.8)$$

For a proof of this result see §A.4, in particular (A.4.10) and (A.4.16).

Next we prove the following:

Lemma 3.2.2. The derivatives of the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ defined in (3.2.2) and (3.2.3) with respect to the structural parameters $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are given by

$$\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\Xi}} = (\underset{\sim}{\Lambda} \otimes \underset{\sim}{I}_q)\underset{\sim}{E}_{(q,r)} , \tag{3.2.9}$$

$$\frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\Lambda}} = \underset{\sim}{E}_{(p,q)}(\underset{\sim}{\Xi} \otimes \underset{\sim}{I}_q) , \tag{3.2.10}$$

$$\frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Lambda}} = \underset{\sim}{E}_{(p,q)}(\underset{\sim}{\Phi\Lambda}' \otimes \underset{\sim}{I}_q) + (\underset{\sim}{\Lambda\Phi} \otimes \underset{\sim}{I}_p)\underset{\sim}{I}_{(p,q)} , \tag{3.2.11}$$

$$\frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Phi}} = (\underset{\sim}{\Lambda} \otimes \underset{\sim}{I}_q)\underset{\sim}{E}_{(q,q)}(\underset{\sim}{\Lambda}' \otimes \underset{\sim}{I}_p) , \tag{3.2.12}$$

and

$$\frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Psi}_d} = \text{diag blocks } \underset{\sim}{E}_{(p,p)} = \begin{bmatrix} \underset{\sim}{E}_{11} & & \underset{\sim}{0} \\ & \ddots & \\ \underset{\sim}{0} & & \underset{\sim}{E}_{pp} \end{bmatrix} , \tag{3.2.13}$$

The concept of matrix derivative and the notations $\underset{\sim}{E}_{ij}$, $\underset{\sim}{E}_{(m,n)}$ and $\underset{\sim}{I}_{(m,n)}$ are defined in the appendix, see Definitions A.2.2 and A.2.3.

Proof: The proof is a straightforward application of the matrix differentiation rules given in §A.3. In particular (3.2.9) and (3.2.10) follow from (A.3.3); (3.2.11) follows from the sum rule (A.3.1) and (A.3.6). (3.2.12) may be obtained from the sum rule and (A.3.3), and (3.2.13) follows from the sum rule and the definition of matrix derivative (A.2.5). Note that in (3.2.12) we have ignored the symmetry of $\underset{\sim}{\Phi}$, but in (3.2.13) we have taken into account the fact that $\underset{\sim}{\Psi}$ is diagonal, as indicated by the use of the subscript d. □

We are now ready to prove the following result:

Theorem 3.2.3. The derivatives of the function F defined in (3.2.6) with respect to the structural parameters $\underset{\sim}{\Xi}$ $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are given by

$$\frac{\partial F}{\partial \underset{\sim}{\Xi}} = -\frac{2}{n}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}' \ , \tag{3.2.14}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Lambda}} = -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}'\underset{\sim}{\Xi}' + 2\underset{\sim}{\Omega}\underset{\sim}{\Lambda}\underset{\sim}{\Phi} \ , \tag{3.2.15}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Phi}_s} = 2\underset{\sim}{\Lambda}'\underset{\sim}{\Omega}\underset{\sim}{\Lambda} - \text{diag}\ \underset{\sim}{\Lambda}'\underset{\sim}{\Omega}\underset{\sim}{\Lambda} \ , \quad \text{and} \tag{3.2.16}$$

$$\frac{\partial F}{\partial \underset{\sim}{\Psi}_d} = \text{diag}\ \underset{\sim}{\Omega} \ , \tag{3.2.17}$$

where

$$\underset{\sim}{\Omega} = \underset{\sim}{V}^{-1}[\underset{\sim}{V}-\underset{\sim}{T}]\underset{\sim}{V}^{-1} \ . \tag{3.2.18}$$

Proof: The proof is based on a series of applications of the chain rule (A.3.15), together with Lemmas 3.2.1 and 3.2.2. Differentiating with respect to $\underset{\sim}{\Xi}$ we have from (A.3.15)

$$\frac{\partial F}{\partial \underset{\sim}{\Xi}} = \frac{\partial F}{\partial \underset{\sim}{\beta}} * \frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\Xi}} \tag{3.2.19}$$

$$= -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}' * (\underset{\sim}{\Lambda} \otimes \underset{\sim}{I}_q)\underset{\sim}{E}_{(q,r)} \tag{3.2.20}$$

$$= -\frac{2}{n}[\underset{\sim}{A}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}]' \ , \tag{3.2.21}$$

where (3.2.20) follows from (3.2.7) and (3.2.9), and (3.2.21) is obtained using identity (A.3.14) for star products. Then (3.2.14) follows by computing the transpose in (3.2.21).

Differentiating with respect to $\underset{\sim}{\Lambda}$ we have, from the chain rule,

$$\frac{\partial F}{\partial \underset{\sim}{\Lambda}} = \frac{\partial F}{\partial \underset{\sim}{\beta}} * \frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\Lambda}} + \frac{\partial F}{\partial \underset{\sim}{V}} * \frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Lambda}} \ . \tag{3.2.22}$$

The first term in (3.2.22) is

$$\frac{\partial F}{\partial \underset{\sim}{\beta}} * \frac{\partial \underset{\sim}{\beta}}{\partial \underset{\sim}{\Lambda}} = -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}' * \underset{\sim}{E}_{(p,q)}(\underset{\sim}{\Xi} \otimes \underset{\sim}{I}_q) \tag{3.2.23}$$

$$= -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}'\underset{\sim}{\Xi}' \ , \tag{3.2.24}$$

where (3.2.23) follows from (3.2.7) and (3.2.10), and (3.2.24) follows from identity (A.3.14) and $(\underset{\sim}{\underset{\sim}{AB}})' = \underset{\sim}{\underset{\sim}{B}}'\underset{\sim}{\underset{\sim}{A}}'$.

The second term in (3.2.22) is

$$\frac{\partial F}{\partial \underset{\sim}{V}} * \frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Lambda}} = \underset{\sim}{\underset{\sim}{\Omega}} * [\underset{\sim}{\underset{\sim}{E}}_{(p,q)}(\underset{\sim}{\underset{\sim}{\Phi\Lambda}}' \otimes \underset{\sim}{\underset{\sim}{I}}_q) + (\underset{\sim}{\underset{\sim}{\Lambda\Phi}} \otimes \underset{\sim}{\underset{\sim}{I}}_p)\underset{\sim}{\underset{\sim}{I}}_{(p,q)}] \tag{3.2.25}$$

$$= \underset{\sim}{\underset{\sim}{\Omega}} * \underset{\sim}{\underset{\sim}{E}}_{(p,q)}(\underset{\sim}{\underset{\sim}{\Phi\Lambda}}' \otimes \underset{\sim}{\underset{\sim}{I}}_q) + \underset{\sim}{\underset{\sim}{\Omega}} * (\underset{\sim}{\underset{\sim}{\Lambda\Phi}} \otimes \underset{\sim}{\underset{\sim}{I}}_p)\underset{\sim}{\underset{\sim}{I}}_{(p,q)} \tag{3.2.26}$$

$$= (\underset{\sim}{\underset{\sim}{\Phi\Lambda}}'\underset{\sim}{\underset{\sim}{\Omega}}')' + \underset{\sim}{\underset{\sim}{\Omega}}'\underset{\sim}{\underset{\sim}{\Lambda\Phi}} \tag{3.2.27}$$

$$= 2\underset{\sim}{\underset{\sim}{\Omega\Lambda\Phi}} \quad , \tag{3.2.28}$$

where (3.2.25) follows from (3.2.8), notation (3.2.18) and (3.2.11); (3.2.26) follows from the distributivity of star products, (3.2.27) may be obtained using identities (A.3.13) and (A.3.14), and (3.2.28) follows from the symmetry of $\underset{\sim}{\underset{\sim}{\Phi}}$ and $\underset{\sim}{\underset{\sim}{\Omega}}$. This proves (3.2.15).

Differentiating with respect to $\underset{\sim}{\underset{\sim}{\Phi}}$ we have

$$\frac{\partial F}{\partial \underset{\sim}{\underset{\sim}{\Phi}}} = \frac{\partial F}{\partial \underset{\sim}{V}} * \frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\underset{\sim}{\Phi}}} \tag{3.2.29}$$

$$= \underset{\sim}{\underset{\sim}{\Omega}} * (\underset{\sim}{\underset{\sim}{\Lambda}} \otimes \underset{\sim}{\underset{\sim}{I}}_q)\underset{\sim}{\underset{\sim}{E}}_{(q,q)}(\underset{\sim}{\underset{\sim}{\Lambda}}' \otimes \underset{\sim}{\underset{\sim}{I}}_q) \tag{3.2.30}$$

$$= \underset{\sim}{\underset{\sim}{\Lambda}}'\underset{\sim}{\underset{\sim}{\Omega\Lambda}} \quad , \tag{3.2.31}$$

where (3.2.29) follows from the chain rule, (3.2.30) follows from (3.2.8), (3.2.18) and (3.2.12), and (3.2.31) follows from identity (A.3.14) and the symmetry of $\underset{\sim}{\underset{\sim}{\Omega}}$. From (3.2.31) and the remarks in §A.3.3, the partial derivatives of F with respect to the distinct elements of $\underset{\sim}{\underset{\sim}{\Phi}}$ may be written as in (3.2.16).

Finally, differentiating with respect to $\underset{\sim}{\Psi}$,

$$\frac{\partial F}{\partial \underset{\sim}{\Psi}_d} = \frac{\partial F}{\partial \underset{\sim}{V}} * \frac{\partial \underset{\sim}{V}}{\partial \underset{\sim}{\Psi}_d} \tag{3.2.32}$$

$$= \underset{\sim}{\Omega} * \text{diag blocks } \underset{\sim}{E}_{(p,p)} \tag{3.2.33}$$

$$= \text{diag } \underset{\sim}{\Omega} , \tag{3.2.34}$$

where (3.2.32) follows from the chain rule, and (3.2.33) follows from (3.2.8), (3.2.18) and (3.2.13). To obtain (3.2.34) recall the definition of star product (A.3.11), note that ω_{ij} is multiplied by $\underset{\sim}{0}: p\times q$ when $i \neq j$ and that $\omega_{ii}E_{ii} = \text{diag}(0,\ldots,0,\omega_{ii},0,\ldots,0)$ where ω_{ii} is in the i-th diagonal position. Thus $\sum_i \omega_{ii}\underset{\sim}{E}_{ii} = \text{diag } \underset{\sim}{\Omega}$.

This completes the proof of the theorem. □

In the special case where $r = 0$ and the linear model part of the structure is ignored, the results in Theorem 3.2.3 reduce to the well-known results of factor analysis; compare (3.2.15) - (3.2.17) with (1.4.2) - (1.4.4).

3.3 Estimation of μ.

The results obtained so far apply to the case where no location parameter μ is specified. We now show that if μ is introduced its m.l.e. is $\bar{\underset{\sim}{X}} = \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}_n$ and that the results in §3.2 hold with $\underset{\sim}{X}$ replaced by $\underset{\sim}{X} - \bar{\underset{\sim}{X}}\underset{\sim}{1}_n'$, provided $\underset{\sim}{A}\underset{\sim}{1}_n = \underset{\sim}{0}$.

In the general case the function F has the same form (3.2.6) but with $\underset{\sim}{T}$ defined by

$$\underset{\sim}{T} = \frac{1}{n}(\underset{\sim}{X}-\mu\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A})(\underset{\sim}{X}-\mu\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A})' . \tag{3.3.1}$$

Proceeding in the same fashion as in (A.4.5) - (A.4.10) we have

$$\frac{\partial F}{\partial \underset{\sim}{\mu}} = \left(\frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{T}} * \frac{\partial \underset{\sim}{T}}{\partial (\underset{\sim}{X}-\underset{\sim}{\mu}\underset{\sim}{1}_n'-\underset{\sim}{\beta}\underset{\sim}{A})} \right) * \frac{\partial (\underset{\sim}{X}-\underset{\sim}{\mu}\underset{\sim}{1}_n'-\underset{\sim}{\beta}\underset{\sim}{A})}{\partial \underset{\sim}{\mu}}$$

$$= -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu}\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A}) * \underset{\sim}{E}_{(p,1)}(\underset{\sim}{1}_n' \otimes \underset{\sim}{I}_1)$$

$$= -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu}\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{1}_n \ . \qquad (3.3.2)$$

Setting (3.3.2) to zero and noting that $\underset{\sim}{1}_n'\underset{\sim}{1}_n = n$, a scalar, we obtain the equation

$$n\underset{\sim}{\mu} = \underset{\sim}{X}\underset{\sim}{1}_n - \underset{\sim}{\beta}\underset{\sim}{A}\underset{\sim}{1}_n \ , \qquad (3.3.3)$$

which under the restriction $\underset{\sim}{A}\underset{\sim}{1}_n = \underset{\sim}{0}$ leads to the estimator

$$\hat{\underset{\sim}{\mu}} = \frac{1}{n}\underset{\sim}{X}\underset{\sim}{1}_n = \bar{\underset{\sim}{X}} \ . \qquad (3.3.4)$$

Substituting this result into (3.3.1) we see that the function F niminized over $\underset{\sim}{\mu}$ has the usual form (3.2.6) with

$$\underset{\sim}{T} = \frac{1}{n}(\underset{\sim}{X}-\bar{\underset{\sim}{X}}\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A})(\underset{\sim}{X}-\bar{\underset{\sim}{X}}\underset{\sim}{1}_n' - \underset{\sim}{\beta}\underset{\sim}{A})' \ . \qquad (3.3.5)$$

Thus to minimize this function with respect to $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ we can use Theorem 3.2.3 with $\underset{\sim}{X}$ replaced by $\underset{\sim}{X} - \bar{\underset{\sim}{X}}\underset{\sim}{1}_n'$.

This result is analogous to factor analysis where one uses $\underset{\sim}{S} = \frac{1}{n}\underset{\sim}{X}\underset{\sim}{X}'$ if $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $\underset{\sim}{S} = \frac{1}{n-1}(\underset{\sim}{X}-\bar{\underset{\sim}{X}}\underset{\sim}{1}_n')(\underset{\sim}{X}-\bar{\underset{\sim}{X}}\underset{\sim}{1}_n')'$ in the general case, the difference in the denominator being due to the fact that the Wishart rather than the Normal likelihood is used.

It may also be noted that the restriction $\underset{\sim}{A}\underset{\sim}{1}_n = \underset{\sim}{0}$ discussed in §2.3.2 permits considerable simplification of the results, for we can see from (3.2.3) that if we only required $\underset{\sim}{A}$ to be linearly independent of $\underset{\sim}{1}_n$, then $\underset{\sim}{\mu}$ would have to be estimated iteratively.

3.4 Estimation of $\underset{\sim}{\Xi}$.

In the most general case, where all the parameters $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ may be restricted, the likelihood equations given in §3.2 must be solved numerically. If $\underset{\sim}{\Xi}$ is unrestricted, however, the value $\tilde{\underset{\sim}{\Xi}}$ that minimizes F conditional on fixed values $\underset{\sim}{\Lambda}$ $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ of the other structural parameters may be obtained analytically.

The following result will be useful.

Lemma 3.4.1. Let $\underset{\sim}{V}$ be any matrix satisfying (3.2.3). Then

$$\underset{\sim}{V}^{-1} = \underset{\sim}{\Psi}^{-1} - \underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Phi\Delta})^{-1}\underset{\sim}{\Phi\Lambda}'\underset{\sim}{\Psi}^{-1} , \tag{3.4.1}$$

$$\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda} = \underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{I} + \underset{\sim}{\Phi\Delta})^{-1} , \text{ and} \tag{3.4.2}$$

$$\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda} = \underset{\sim}{\Delta}(\underset{\sim}{I} + \underset{\sim}{\Phi\Delta})^{-1} , \tag{3.4.3}$$

where

$$\underset{\sim}{\Delta} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda} . \tag{3.4.4}$$

This result is well-known, see for example Lawley and Maxwell (1971, p. 89). Result (3.4.1) may be verified post-multiplying by $\underset{\sim}{V}$ in the form $\underset{\sim}{\Lambda\Phi\Lambda}' + \underset{\sim}{\Psi}$. Results (3.4.2) and (3.4.3) follow immediately.

We can now prove the following.

Theorem 3.4.2. Let F be defined by (3.2.6). If $\underset{\sim}{\Xi}$ is unrestricted, the value $\tilde{\underset{\sim}{\Xi}}$ that minimizes F conditional on fixed values of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ is given by

$$\tilde{\underset{\sim}{\Xi}} = (\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{XA}'(\underset{\sim}{AA}')^{-1} . \tag{3.4.5}$$

Proof: On setting to zero the partial derivatives of F with respect to $\underset{\sim}{\Xi}$ given in (3.2.14) and since $\underset{\sim}{\beta} = \underset{\sim}{\Lambda\Xi}$, we obtain

$$\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{XA}' = \underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\tilde{\underset{\sim}{\Xi}}\underset{\sim}{AA}' . \tag{3.4.6}$$

which under the full rank assumptions of the model gives

$$\tilde{\Xi} = (\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}X A'(AA')^{-1} . \qquad (3.4.7)$$

This result is sufficient to compute $\tilde{\Xi}$, but can be simplified as follows. Inverting (3.4.3) we obtain

$$\begin{aligned}(\Lambda'\Psi^{-1}\Lambda)^{-1} &= (I + \Phi\Delta)\Delta^{-1} = \Delta^{-1} + \Phi \\ &= \Delta^{-1}(I + \Delta\Phi) = \Delta^{-1}(I + \Phi\Delta)' . \qquad (3.4.8)\end{aligned}$$

On the other hand, transposing (3.4.2)

$$\Lambda'\Psi^{-1} = (I + \Phi\Delta)^{-T}\Lambda'\Psi^{-1} , \qquad (3.4.9)$$

where $A^{-T} = (A')^{-1}$. Now using (3.4.8) and (3.4.9)

$$(\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1} = \Delta^{-1}(I + \Phi\Delta)'(I + \Phi\Delta)^{-T}\Lambda'\Psi^{-1} = \Delta^{-1}\Lambda'\Psi^{-1}. \qquad (3.4.10)$$

Hence (3.4.7) may be written as in (3.4.5). This completes the proof. □

Note that $\tilde{\Xi}$ does not depend on Φ. A computational advantage of (3.4.5) over (3.4.7) is that inversion of Ψ is replaced by inversion of Ψ, a diagonal matrix.

3.5 Estimation of Λ, Φ and Ψ.

We consider now estimation of the remaining structural parameters, which will be subject to a set of identifying restrictions as discussed in §2.3.1. Unfortunately, there is no analytic solution of the likelihood equations (3.2.15) - (3.2.17), and the estimates must be obtained numerically.

The function to be optimized is F minimized over $\underset{\sim}{\Xi}$ for fixed $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, and will be written as

$$\tilde{F} = \tilde{F}(\underset{\sim}{\Lambda},\underset{\sim}{\Phi},\underset{\sim}{\Psi}) = \min_{\underset{\sim}{\Xi}} F(\underset{\sim}{\Xi},\underset{\sim}{\Lambda},\underset{\sim}{\Phi},\underset{\sim}{\Psi}) = F(\tilde{\underset{\sim}{\Xi}},\underset{\sim}{\Lambda},\underset{\sim}{\Phi},\underset{\sim}{\Psi}) \ . \tag{3.5.1}$$

We now give some results which are useful in evaluating $\tilde{F}$ and its derivatives. First we note that

$$\tilde{F} = \log|\underset{\sim}{V}| + \operatorname{tr} \underset{\sim}{V}^{-1}\tilde{\underset{\sim}{T}} \ , \tag{3.5.2}$$

where $\tilde{\underset{\sim}{T}}$ denotes $\underset{\sim}{T}$ evaluated at $\underset{\sim}{\Xi} = \tilde{\underset{\sim}{\Xi}}$.

Computation of $|\underset{\sim}{V}|$ is simplified by the following result given by Lawley and Maxwell (1971, p. 89).

Lemma 3.5.1 Let $\underset{\sim}{V}$ be given by (3.2.3). Then

$$|\underset{\sim}{V}| = |\underset{\sim}{\Psi}| \ |\underset{\sim}{I}_q + \underset{\sim}{\Phi\Delta}| \ , \tag{3.5.3}$$

where $\underset{\sim}{\Delta} = \underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$.

Proof: The proof is based on the following result from linear algebra. If $\underset{\sim}{A}$ is $p\times q$ and $\underset{\sim}{B}$ is $q\times p$ where $q < p$ then

$$|\underset{\sim}{I}_p + \underset{\sim}{AB}| = |\underset{\sim}{I}_q + \underset{\sim}{BA}| \ . \tag{3.5.4}$$

Premultiplying (3.2.3) by $\underset{\sim}{\Psi}^{-1}$, taking determinants on both sides and multiplying by $|\underset{\sim}{\Psi}|$ we obtain

$$|\underset{\sim}{V}| = |\underset{\sim}{\Psi}| \ |\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda\Phi\Lambda}' + \underset{\sim}{I}_p| \ . \tag{3.5.5}$$

Then (3.5.3) follows using (3.5.4) with $\underset{\sim}{A} = \underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda}$ and $\underset{\sim}{B} = \underset{\sim}{\Phi\Lambda}'$.

□

Lemmas 3.4.1 and 3.5.1 permit computation of the determinant and inverse of $\underset{\sim}{V}$ in terms of the determinants and inverses of $(\underset{\sim}{I} + \underset{\sim}{\Phi\Delta})$

and $\underset{\sim}{\Psi}$. Note that $\underset{\sim}{V}$ is p×p, while $(\underset{\sim}{I} + \underset{\sim}{\Phi\Delta})$ is q×q, with q usually considerably smaller than p, and $\underset{\sim}{\Psi}$ is diagonal.

To evaluate $\tilde{\underset{\sim}{T}}$ it is convenient to introduce the following notations. Let

$$\underset{\sim}{B} = \underset{\sim}{X\Lambda}'(\underset{\sim}{\Lambda\Lambda}')^{-1}\underset{\sim}{\Lambda X}' \quad \text{and} \tag{3.5.6}$$

$$\underset{\sim}{C} = (\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{\Lambda}'\underset{\sim}{\Psi}^{-1} \ . \tag{3.5.7}$$

Note that $\underset{\sim}{B}$ is symmetric and does not depend on the parameters.

Then $\tilde{\underset{\sim}{T}}$ may be written as

$$\tilde{\underset{\sim}{T}} = \frac{1}{n}[\underset{\sim}{XX}' - \underset{\sim}{\Lambda CB} - (\underset{\sim}{\Lambda CB})' + \underset{\sim}{\Lambda CBC}'\underset{\sim}{\Lambda}'] \ . \tag{3.5.8}$$

This result follows from expanding (3.2.5) and using (3.4.5) and the above notation, and has a minor computational advantage over (3.2.5) in that the product $(\underset{\sim}{\Lambda\Lambda}')^{-1}\underset{\sim}{\Lambda\Lambda}' = \underset{\sim}{I}_r$ drops out.

We now have the following result.

Lemma 3.5.2 The derivatives of $\tilde{F}$ defined in (3.5.1) with respect to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are given by

$$\frac{\partial \tilde{F}}{\partial \underset{\sim}{\Lambda}} = -\frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{BC}' - \underset{\sim}{\Lambda CBC}') + 2\tilde{\underset{\sim}{\Omega}}\underset{\sim}{\Lambda\Phi} \ , \tag{3.5.9}$$

$$\frac{\partial \tilde{F}}{\partial \underset{\sim s}{\Phi}} = 2\underset{\sim}{\Lambda}'\tilde{\underset{\sim}{\Omega}}\underset{\sim}{\Lambda} - \text{diag}\ \underset{\sim}{\Lambda}'\tilde{\underset{\sim}{\Omega}}\underset{\sim}{\Lambda} \ , \quad \text{and} \tag{3.5.10}$$

$$\frac{\partial \tilde{F}}{\partial \underset{\sim d}{\Psi}} = \text{diag}\ \tilde{\underset{\sim}{\Omega}} \ , \tag{3.5.11}$$

where $\tilde{\underset{\sim}{\Omega}}$ denotes $\underset{\sim}{\Omega}$ evaluated at $\underset{\sim}{T} = \tilde{\underset{\sim}{T}}$ and $\underset{\sim}{B},\underset{\sim}{C}$ are as defined in (3.5.6) - (3.5.7).

Proof: Derivatives of $\tilde{F}$ with respect to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are obtained evaluating (3.2.15) - (3.2.17) of Theorem 3.2.3 at $\underset{\sim}{\Xi} = \tilde{\underset{\sim}{\Xi}}$, for by

(3.5.1) and the chain rule

$$\frac{\partial \tilde{F}}{\partial \underset{\sim}{\Lambda}} = \frac{\partial F(\tilde{\underset{\sim}{\Xi}}, \underset{\sim}{\Lambda}, \underset{\sim}{\Phi}, \underset{\sim}{\Psi})}{\partial \underset{\sim}{\Xi}} * \frac{\partial \tilde{\underset{\sim}{\Xi}}}{\partial \underset{\sim}{\Lambda}} + \frac{\partial F(\tilde{\underset{\sim}{\Xi}}, \underset{\sim}{\Lambda}, \underset{\sim}{\Phi}, \underset{\sim}{\Psi})}{\partial \underset{\sim}{\Lambda}}$$

$$= \frac{\partial F}{\partial \underset{\sim}{\Lambda}}\Bigg|_{\underset{\sim}{\Xi}=\tilde{\underset{\sim}{\Xi}}} ,$$

since $\tilde{\underset{\sim}{\Xi}}$ is a root of $\partial F/\partial \underset{\sim}{\Xi} = \underset{\sim}{0}$. A similiar result holds for $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$. This gives immediately (3.5.10) and (3.5.11), while (3.5.9) is obtained evaluating (3.2.15) and using the notation in (3.5.6) and (3.5.7). □

Unfortunately, no convenient simplification of $\tilde{\underset{\sim}{\Omega}}$ has been found.

3.6 The Iterative Procedure

We are now ready to consider estimation of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ by numerical minimization of $\tilde{F}$ under the Gram-Schmidt restrictions or under a structural hypothesis. In these cases parameters may be (1) fixed *a priori*, (2) constrained to be equal to other parameters in the model, or (3) free, unconstrained parameters.

Let the elements of $\underset{\sim}{\Lambda}$, the upper triangular elements of $\underset{\sim}{\Phi}$ and the diagonal elements of $\underset{\sim}{\Psi}$ be arranged into a vector

$$\chi = (\lambda_{11},\ldots,\lambda_{p1};\ldots;\lambda_{1q},\ldots,\lambda_{pq};\phi_{11};\phi_{12},\phi_{22};\ldots;\phi_{1q},\ldots,\phi_{qq};\psi_1,\ldots,\psi_p)' \tag{3.6.2}$$

and let the free parameters in $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ form a vector

$$\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_m)' . \tag{3.6.2}$$

We now regard $\tilde{F}$ as a function $f(\underset{\sim}{\theta})$ of the free parameters $\underset{\sim}{\theta}$ with partial derivatives

$$g_i = \frac{\partial f}{\partial \theta_i} = \sum_j \alpha_{ij} \frac{\partial \tilde{F}}{\partial \gamma_j} \qquad (3.6.3)$$

where $\alpha_{ij} = 1$ if $\theta_i = \gamma_j$ and $\alpha_{ij} = 0$ otherwise, and the partial derivaties of $\partial\tilde{F}/\partial\gamma$ are obtained from Lemma 3.5.2. (The function $f(\underset{\sim}{\theta})$ and its derivatives $\underset{\sim}{g}(\underset{\sim}{\theta})$ also depend, of course, on the fixed parameters and the data, which are treated as constants).

The function $f(\underset{\sim}{\theta})$ may now be minimized using the method of Davidon (1959) and Fletcher and Powell (1963) described in §1.3.2.

The initial matrix $\underset{\sim}{E}$ may be taken as the identity matrix, in which case the first iteration is in the direction of steepest descent. A better choice of $\underset{\sim}{E}^{(1)}$ however, is given by the inverse of

$$\underset{\sim}{G} = E\left(\frac{\partial^2 f}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}'}\right) , \qquad (3.6.4)$$

where the second-order derivatives of $f(\underset{\sim}{\theta})$ are given by

$$h_{ij} = \frac{\partial^2 f}{\partial \theta_i \partial \theta_j} = \sum_k \sum_\ell \alpha_{ik}\alpha_{j\ell} \frac{\partial^2 \tilde{F}}{\partial \gamma_k \partial \gamma_\ell} , \qquad (3.6.5)$$

and the expected values of the second-order derivatives of $\tilde{F}$ are derived in §4.7.

Since initial estimates may be far from the minimum point and $\underset{\sim}{G}$ depends on $\underset{\sim}{\theta}$, it is convenient to start with several steepest descent iterations, which work better in the first stages of the minimization process. The results from this stage may then be used to compute $\underset{\sim}{E}^{(1)}$ using (3.6.4), and the process may be continued using Fletcher-Powell iterations until a convergence criterion such as $\max_{1\le i\le m} |g_i| < \delta$ for $\delta = .001$, say, is satisfied.

Until experience in using this procedure accumulates, an optimum changeover point can not be determined. On the basis of Jöreskog's (1970a) experience on the analysis of covariance structures, however, it is recommended to change over when f decreases by less than 5% between two consecutive steepest-descent iterations, see also §6.2.3.

Since during iteration the elements of $\underset{\sim}{\Psi}$ may become negative, we proceed as in factor analysis introducing the restriction $\psi_i > \varepsilon$ $(i=1,\ldots,p)$ for small $\varepsilon > 0$ and handle the problem in exactly the same manner as described in §1.3.2.

It may be noted that in factor analysis the canonical restrictions are quite convenient in maximum likelihood estimation, for they lead to an analytical solution for the m.l.e. of $\underset{\sim}{\Lambda}$ for fixed $\underset{\sim}{\Psi}$, and thus reduce the numerical problem to minimization of a function of $\underset{\sim}{\Psi}$. Unfortunately, in our case the structure of $\partial\tilde{F}/\partial\underset{\sim}{\Lambda}$ in (3.4.9) makes such analytic solution impossible. To use these restrictions in estimation would considerably complicate the numerical procedure. We have thus preferred to use the Gram-Schmidt restrictions in estimation, leaving the option of rotating to canonical estimates afterwards.

In the case where $\underset{\sim}{\Xi}$ is restricted, the estimates may be computed minimizing F using the same procedure described here, except that the partial derivatives are as given in Theorem 3.2.3, and the number of parameters to be estimated iteratively may increase considerably.

IV. LARGE SAMPLE PROPERTIES OF THE ESTIMATORS

4.1 Introduction

The main purpose of this chapter is to establish the consistency and asymptotic normality of the maximum likelihood estimators of the free parameters in the latent linear model. It is well known that if $\hat{\underset{\sim}{\theta}}_n$ is the m.l.e. of a parameter $\underset{\sim}{\theta}$ based on n i.i.d. observations then, under certain regularity conditions, as $n \to \infty$ $\hat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}$ and

$$\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] \ , \tag{4.1.1}$$

where $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the Fisher information matrix, see for example Kendall and Stuart (1961, pp. 54-55) or Zacks (1971, pp. 246-257). Unfortunately in linear models the columns of $\underset{\sim}{X}$, though independent, are not identically distributed random vectors. Thus the standard results mentioned above cannot be applied to our problem; but additional definitions and assumptions must be introduced and existing proofs must be adapted.

The approach adopted is the following. In §4.2 we state a basic assumption regarding the nature of the limiting process, and establish conditions for the consistency and asymptotic normality of least squares estimators in linear models. In §4.3 we consider a family of structural linear models where the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ are regular functions of a structural parameter $\underset{\sim}{\theta}$, and use the results of the previous section to establish the consistency of an estimator

$\hat{\underset{\sim}{\theta}}_n$ of $\underset{\sim}{\theta}$ obtained by minimizing the function F of Chapter 3. In §4.4 we introduce the concept of the limiting Fisher information matrix $\underset{\sim}{I}(\underset{\sim}{\theta})$, which plays a central role in our treatment of asymptotic distribution theory; obtain the second derivatives of the function F; use these to derive expressions for the elements of $\underset{\sim}{I}(\underset{\sim}{\theta})$ for the family of structural linear models introduced in the previous section, and show that $\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n)$ is a consistent estimator of $\underset{\sim}{I}(\underset{\sim}{\theta})$. In §4.5 we establish some additional convergence results and show that the asymptotic distribution of $\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta})$ in the family of structural linear models under consideration is as given in (4.1.1), but with $\underset{\sim}{I}(\underset{\sim}{\theta})$ denoting the limiting (rather than the ordinary) Fisher information matrix. In §4.6 we proceed to specialize the results obtained to the case of the latent linear model. The main task here is to obtain expressions for the elements of $\underset{\sim}{I}(\underset{\sim}{\theta})$ for the latent linear model. Finally, in §4.7 we prove an asymptotic result which provides large sample approximations to the second derivatives of the function $\tilde{F}$ of §3.5. These approximations are used in the Fletcher-Powell procedure for minimizing $\tilde{F}$ in §3.6.

4.2 Asymptotic Results for Linear Models

Consider the multivariate general linear model

$$\underset{\sim}{X}_n = \underset{\sim}{\beta}\underset{\sim}{A}_n + \underset{\sim}{U}_n \,, \tag{4.2.1}$$

where $\underset{\sim}{A}_n$ is of full row rank $r < n$ and the columns of $\underset{\sim}{U}_n$ are i.i.d. random vectors with mean vector $\underset{\sim}{0}$ and p.d. variance-covariance matrix $\underset{\sim}{V}$.

Define the least squares estimators of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$

$$\overline{\underset{\sim}{\beta}}_n = \underset{\sim}{X}_n\underset{\sim}{A}_n'(\underset{\sim}{A}_n\underset{\sim}{A}_n')^{-1} \quad , \quad \text{and} \tag{4.2.2}$$

$$\overline{\underset{\sim}{V}}_n = \frac{1}{n}\underset{\sim}{X}_n[\underset{\sim}{I}-\underset{\sim}{A}_n'(\underset{\sim}{A}_n\underset{\sim}{A}_n')^{-1}\underset{\sim}{A}_n]\underset{\sim}{X}_n' \quad , \tag{4.2.3}$$

which are also m.l.e.'s if $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$, see §A.4.

We now consider the large sample behavior of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$.

4.2.1 Conditions on the Limiting Process

The linear model (4.2.1) is conditional on a design matrix $\underset{\sim}{A}_n$ of fixed, known constants. As n increases, however, this matrix is modified by the addition of new columns. Clearly some stability conditions must be imposed on this process. In this regard we introduce

Assumption 4.2.1. Let $\{\underset{\sim}{A}_n\}$ be a sequence of $r\times n$ matrices of fixed, known constants, and define $\underset{\sim}{Q}_n = \frac{1}{n}\underset{\sim}{A}_n\underset{\sim}{A}_n'$. We assume that

$$\lim_{n\to\infty} \underset{\sim}{Q}_n = \underset{\sim}{Q} \tag{4.2.4}$$

exists and is positive definite.

This assumption (or an equivalent one) has been used by Eicker (1963) and Sen and Puri (1970) in studies of asymptotic properties of least squares estimators in linear models. Also, it is frequently considered in the econometric literature, see for example Theil (1971, Ch. 8). We now consider some implications of this assumption in applied situations.

Consider a multivariate one-way analysis of variance problem for which the design matrix $\underset{\sim}{A}_n$ consists of dummy variates such that $a_{ij} = 1$ if observation j is on treatment i and $a_{ij} = 0$ otherwise.

Let n_i denote the number of observations on the i-th treatment $(i=1,\ldots,r)$ and $n = \Sigma n_i$. Then

$$\underset{\sim}{Q}_n = \frac{1}{n}\underset{\sim}{A}_n\underset{\sim}{A}_n' = \text{diag}\,\frac{1}{n}(n_1,\ldots,n_r) \ , \tag{4.2.5}$$

and Assumption 4.2.1 is equivalent to requiring that as $n \to \infty$ $\frac{n_i}{n} \to \eta_i$: $0 < \eta_i < 1$ $(i=1,\ldots,r)$, so that the sample size increases for all treatments.

Or consider a multivariate multiple regression problem where the design matrix $\underset{\sim}{A}_n$ consists of regressors or explanatory variables. Although the analysis is conditional on fixed values of the regressors the columns of $\underset{\sim}{A}_n$, say $\underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_n$, may be regarded as values taken by a stochastic vector. Suppose that $E(\underset{\sim}{a}_\alpha) = \underset{\sim}{\mu}$ and $\text{Var}(\underset{\sim}{a}_\alpha) = \underset{\sim}{\Sigma}$, so that $E(\underset{\sim}{a}_\alpha\underset{\sim}{a}_\alpha') = \underset{\sim}{\Sigma} + \underset{\sim}{\mu}\underset{\sim}{\mu}'$ $(\alpha=1,\ldots,n)$. Then by Kolmogorov's strong law of large numbers, see e.g. Rao (1965, p. 94), as $n \to \infty$

$$\underset{\sim}{Q}_n = \frac{1}{n}\sum_{\alpha=1}^{n}\underset{\sim}{a}_\alpha\underset{\sim}{a}_\alpha' \xrightarrow{a.s.} \underset{\sim}{\Sigma} + \underset{\sim}{\mu}\underset{\sim}{\mu}'. \tag{4.2.6}$$

Hence if $\underset{\sim}{\Sigma}$ is p.d., Assumption 4.2.1 will be satisfied for almost all sample sequences $\{\underset{\sim}{A}_n\}$.

An alternative way to conceptualize the limiting process is to consider a sequence of replications of a basic experiment with design matrix $\underset{\sim}{A}$: $r\times k$ such that the n-th experiment in the sequence has design matrix

$$\underset{\sim}{A}_n = [\underset{\sim}{A},\underset{\sim}{A},\ldots,\underset{\sim}{A}] \ , \tag{4.2.7}$$

with n matrices on the right hand side. If $\underset{\sim}{A}$ is of full row rank $r \le k$, then this sequence satisfies Assumption 4.2.1 trivially, for $\underset{\sim}{Q}_n = \frac{1}{k}\underset{\sim}{A}\underset{\sim}{A}'$ for all n. While this more restrictive assumption leads to

somewhat simpler proofs of asymptotic results, it becomes difficult to justify use of those results as approximations in applied situations, such as analysis of covariance problems, where the columns of $\underset{\sim}{A}_n$ may be all different, and hence $\underset{\sim}{A}_n$ may not be considered a member of a sequence of form (4.2.7). In this regard it is important to bear in mind that asymptotic results are useful to the extent that they can be used as approximations in large samples, and that this may well depend on the conceptual framework used in the derivations.

4.2.2 Consistency of $\bar{\underset{\sim}{\beta}}_n$ and $\bar{\underset{\sim}{V}}_n$.

We are now ready to prove the following result.

Lemma 4.2.1. Let $\underset{\sim}{X}_n$ satisfy the linear model (4.2.1) with design matrix $\underset{\sim}{A}_n$ $(n=r+1, r+2,\ldots)$. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 then as $n \to \infty$

$$\bar{\underset{\sim}{\beta}}_n \xrightarrow{p} \underset{\sim}{\beta} \quad \text{and} \quad \bar{\underset{\sim}{V}}_n \xrightarrow{p} \underset{\sim}{V}, \tag{4.2.8}$$

i.e. $\bar{\underset{\sim}{\beta}}_n$ and $\bar{\underset{\sim}{V}}_n$ are consistent estimators of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$.

Proof. Consider first $\bar{\underset{\sim}{\beta}}_n$. From (4.2.2), noting that $E(\underset{\sim}{X}_n) = \underset{\sim}{\beta}\underset{\sim}{A}_n$,

$$E(\bar{\underset{\sim}{\beta}}_n) = \underset{\sim}{\beta}. \tag{4.2.9}$$

On the other hand, from Anderson (1958, p. 182),

$$\mathrm{Var}(\bar{\underset{\sim}{\beta}}_n) = \underset{\sim}{V} \otimes (\underset{\sim}{A}_n\underset{\sim}{A}_n)^{-1} = \frac{1}{n}\underset{\sim}{V} \otimes \underset{\sim}{Q}_n^{-1}. \tag{4.2.10}$$

From Assumption 4.2.1 we have, as $n \to \infty$

$$\mathrm{Var}(\bar{\underset{\sim}{\beta}}_n) \to \underset{\sim}{0}, \tag{4.2.11}$$

which in view of (4.2.9) implies that $\bar{\underset{\sim}{\beta}}_n$ converges to $\underset{\sim}{\beta}$ in quadratic mean, and hence also in probability, see Rao (1965, p. 90).

Consider now $\overline{V}_n$. From (4.2.2) and (4.2.3),

$$n\overline{V}_n = X_n X_n' - \overline{\beta}_n A_n A_n' \overline{\beta}_n' \ . \tag{4.2.12}$$

Let us define

$$T_n = \frac{1}{n}(X_n - \beta A_n)(X_n - \beta A_n)' \tag{4.2.13}$$

Expanding (4.2.13) and using (4.2.2) we obtain

$$nT_n = X_n X_n' - \overline{\beta}_n A_n A_n' \beta' - \beta A_n A_n' \overline{\beta}_n' + \beta A_n A_n' \beta' \ , \tag{4.2.14}$$

and on substituting this into (4.2.12) we find, after simplification,

$$\overline{V}_n = T_n - (\overline{\beta}_n - \beta) Q_n (\overline{\beta}_n - \beta)' \ . \tag{4.2.15}$$

Now from (4.2.13) and (4.2.1), T_n may be written as

$$T_n = \frac{1}{n} U_n U_n' = \frac{1}{n} \sum_{\alpha=1}^{n} u_\alpha u_\alpha' \ , \tag{4.2.16}$$

where the $u_\alpha u_\alpha'$ are i.i.d. random matrices with $E(u_\alpha u_\alpha') = V$ for all α. Hence by Khinchin's law of large numbers, see Rao (1965, p. 92), as $n \to \infty$

$$T_n \xrightarrow{p} V \ . \tag{4.2.17}$$

On the other hand, since $\overline{\beta} \xrightarrow{p} \beta$ and for each n, the quadratic form in the second term of (4.2.15) is a continuous function of $\overline{\beta}_n$ we have, using Assumption 4.2.1 and result (xiii) in Rao (1965, p. 104), that as $n \to \infty$

$$(\overline{\beta}_n - \beta) Q_n (\overline{\beta}_n - \beta)' \xrightarrow{p} 0 \ . \tag{4.2.18}$$

In view of (4.2.15), the lemma follows from (4.2.17) and (4.2.18).

□

This lemma generalizes Theorem 1 in Eicker (1963), who proves consistence of $\overline{\underset{\sim}{\beta}}_n$ in the univariate case. For related results see Jennrich (1969) and Sen and Puri (1970).

4.2.3 Asymptotic Normality of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$.

It is well known that if $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ then $\overline{\underset{\sim}{\beta}}_n \sim N_{p\times r}[\underset{\sim}{\beta},\underset{\sim}{V} \otimes (\underset{\sim}{A}_n\underset{\sim}{A}_n)^{-1}]$ and $n\overline{\underset{\sim}{V}}_n \sim W_p(\underset{\sim}{V},n-r)$ independently of $\overline{\underset{\sim}{\beta}}_n$; see for example Anderson (1958, p. 183). Asymptotically we have

Lemma 4.2.2. Let $\underset{\sim}{X}_n$ satisfy the linear model (4.2.1) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \times \underset{\sim}{I}_n)$, $(n=r+1,r+2,\ldots)$. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, then as $n \to \infty$

$$\sqrt{n}(\overline{\underset{\sim}{\beta}}_n-\underset{\sim}{\beta}) \xrightarrow{d} N_{p\times r}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{Q}^{-1}) , \tag{4.2.19}$$

and, independently of $\overline{\underset{\sim}{\beta}}_n$,

$$\sqrt{n}(\overline{\underset{\sim}{V}}_n-\underset{\sim}{V}) \xrightarrow{d} N[\underset{\sim}{0},((v_{ik}v_{j\ell} + v_{i\ell}v_{jk}))] , \tag{4.2.20}$$

where the notation used in (4.2.20) indicates that the asymptotic covariance of the (i,j)-th and (k,ℓ)-th elements of $\sqrt{n}(\overline{\underset{\sim}{V}}_n-\underset{\sim}{V})$ is $v_{ik}v_{j\ell} + v_{i\ell}v_{jk}$.

Thus $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ have asymptotically a joint normal distribution.

Proof: The first part of the lemma follows from the fact that for every fixed n

$$\sqrt{n}(\overline{\underset{\sim}{\beta}}_n-\underset{\sim}{\beta}) \sim N_{p\times r}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{Q}_n^{-1}) , \tag{4.2.21}$$

and from Assumption 4.2.1.

To prove the second part note that from (4.2.15)

$$\sqrt{n}(\bar{\underset{\sim}{V}}_n - \underset{\sim}{V}) = \sqrt{n}(\underset{\sim}{T}_n - \underset{\sim}{V}) - \sqrt{n}(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})\underset{\sim}{Q}_n(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})' \ . \tag{4.2.22}$$

From (4.2.16), $\underset{\sim}{T}_n$ is the average of n i.i.d. random matrices $\underset{\sim}{u}_\alpha \underset{\sim}{u}_\alpha'$ with $E(\underset{\sim}{u}_\alpha \underset{\sim}{u}_\alpha') = \underset{\sim}{V}$ and covariances given by

$$E(u_{i\alpha}u_{j\alpha} - v_{ij})(u_{k\alpha}u_{\ell\alpha} - v_{k\ell}) = v_{ik}v_{j\ell} + v_{i\ell}v_{jk} \tag{4.2.23}$$

$(\alpha = 1, \ldots, n)$, see Anderson (1958, p. 39). Thus by the multivariate version of the Lindeberg-Levy central limit theorem, see Rao (1965, pp. 107-109), as $n \to \infty$

$$\sqrt{n}\ (\underset{\sim}{T}_n - \underset{\sim}{V}) \xrightarrow{d} N[\underset{\sim}{0}, ((v_{ik}v_{j\ell} + v_{i\ell}v_{jk}))] \ . \tag{4.2.24}$$

Consider now the quadratic form on the right hand side of (4.2.22). In Lemma 4.2.1 we proved that $\bar{\underset{\sim}{\beta}}_n \xrightarrow{p} \underset{\sim}{\beta}$. This result, however, can be strengthened to

$$n^{\frac{1}{4}}(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}) \xrightarrow{p} \underset{\sim}{0} \ , \tag{4.2.25}$$

because from (4.2.9) the expectation of the left hand side is $\underset{\sim}{0}$, and from (4.2.10) its variance is

$$n^{\frac{1}{2}}\mathrm{Var}(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta}) = \frac{1}{\sqrt{n}}\underset{\sim}{V} \otimes \underset{\sim}{Q}_n^{-1} \ , \tag{4.2.26}$$

which converges to $\underset{\sim}{0}$ as $n \to \infty$. Since the quadratic form of interest is a continuous function of $n^{\frac{1}{4}}(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})$ for each n, from Assumption 4.2.1 and (4.2.25) we obtain that as $n \to \infty$

$$\sqrt{n}(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})\underset{\sim}{Q}_n(\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})' \xrightarrow{p} \underset{\sim}{0} \ . \tag{4.2.27}$$

In view of (4.2.22), the rest of the proof follows from (4.2.24), (4.2.27) and Slutzky's theorem, see for example result (ix) in Rao (1965, p. 101). □

The second part of Lemma 4.2.2 is analogous to Theorem 4.2.4 in Anderson (1958, p. 75), establishing the asymptotic normality of the sample covariance matrix.

The results obtained in §4.5 below, regarding the asymptotic normality of m.l.e.'s in structural linear models, depend only on $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ having the asymptotic joint distribution specified in Lemma 4.2.2. It may be of interest to investigate whether the assumption of normality in the lemma may be weakened, so that estimates obtained minimizing the function F of (3.2.6) may be asymptotically normally distributed even if the distribution of $\underset{\sim}{U}_n$ is not normal. In this regard we note the following.

Sen and Puri (1970) have shown that if the columns of $\underset{\sim}{U}_n$ in (4.2.1) are i.i.d. with mean vector $\underset{\sim}{0}$ and variance-covariance matrix $\underset{\sim}{V}$, and if the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 and the *generalized Noether condition*

$$\max_{\substack{1\le i\le r\\ 1\le j\le n}} \{a_{ij}^2 / \sum_{\alpha=1}^{n} a_{i\alpha}^2\} \to 0 \quad \text{as } n \to \infty , \tag{4.2.28}$$

then $\sqrt{n}(\overline{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})$ has the asymptotic distribution (4.2.19).

From the proof of the second part of the lemma, on the other hand, it is seen that if Assumption 4.2.1 is satisfied and if the fourth moments of $\underset{\sim}{u}_\alpha$ exist then $\sqrt{n}(\overline{\underset{\sim}{V}}_n - \overline{\underset{\sim}{V}})$ is asymptotically normal with mean $\underset{\sim}{0}$ and covariances depending on the fourth moments. If the fourth cumulants of $\underset{\sim}{u}_\alpha$ are zero then the covariances are as given in (4.2.20); see Cramer (1946, pp. 365-366) or Kendall and Stuart (1969, p. 321).

Finally, proceeding as in the proof of Theorem 4.3.2 in Anderson

(1958, p. 83), it can be shown that $\mathrm{Cov}(\bar{\beta}_n, \bar{V}_n) = 0$ for all n. Thus $\bar{\beta}_n$ and $\bar{V}_n$ are uncorrelated, and their asymptotic *marginal* distributions are as given in Lemma 4.3.2 under assumptions weaker than normality. These weaker conditions, however, do not appear to be sufficient for the asymptotic *joint* distribution of $\bar{\beta}_n$ and $\bar{V}_n$ to be as given in the lemma.

4.3 Consistency of $\hat{\theta}_n$ in Structural Linear Models

We now consider a family of structural linear models where X_n satisfies (4.2.1) and the reduced form parameters β and V are continuous differentiable functions of a structural parameter vector θ.

Consider an estimator $\hat{\theta}_n$ of θ obtained by minimizing the function F first defined in (3.2.6),

$$F = \log|V| + \mathrm{tr}\, V^{-1} T_n \,. \tag{4.3.1}$$

Note that $\hat{\theta}_n$ is the m.l.e. of θ if $U_n \sim N_{p\times n}(0, V \otimes I_n)$.

Using (4.2.2), (4.2.3) and proceeding as in (4.2.15), we can write

$$F = \log|V| + \mathrm{tr} V^{-1}[\bar{V}_n + (\bar{\beta}_n - \beta) Q_n (\bar{\beta}_n - \beta)'] \,. \tag{4.3.2}$$

Thus F depends on the observations X_n only through $\bar{\beta}_n$ and $\bar{V}_n$. Under normality assumptions this means that $\bar{\beta}_n$ and $\bar{V}_n$ are sufficient, though not necessarily minimal sufficient, statistics for the structural parameter θ. We now take advantage of this fact in proving consistency of $\hat{\theta}_n$.

Theorem 4.3.1. Let X_n satisfy (4.2.1) with design matrix A_n ($n=r+1, r+2, \ldots$), let β and V be continuous differentiable functions

of a structural parameter $\underset{\sim}{\theta}$, and let $\hat{\underset{\sim}{\theta}}_n$ minimize the function (4.3.1). If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 and if $\underset{\sim}{\theta}$ is identified then as $n \to \infty$

$$\hat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta} , \tag{4.3.3}$$

i.e. $\hat{\underset{\sim}{\theta}}_n$ is a consistent estimator of $\underset{\sim}{\theta}$.

Proof: Let us use the notations $\underset{\sim}{\theta}^*$, $\underset{\sim}{\beta}^*$ and $\underset{\sim}{V}^*$ for arbitrary values assigned to the parameters $\underset{\sim}{\theta}$, $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$.

Since by (4.3.2) F is a continuous function of $\bar{\underset{\sim}{\beta}}_n$ and $\bar{\underset{\sim}{V}}_n$, and since by Lemma 4.2.1 $\bar{\underset{\sim}{\beta}}_n \xrightarrow{p} \underset{\sim}{\beta}$ and $\bar{\underset{\sim}{V}}_n \xrightarrow{p} \underset{\sim}{V}$, we have using Assumption 4.2.1 and result (xiii) in Rao (1965, p. 104), that as $n \to \infty$

$$F(\underset{\sim}{\theta}^*) \xrightarrow{p} \log|\underset{\sim}{V}^*| + \mathrm{tr}\underset{\sim}{V}^{*-1}[\underset{\sim}{V} - (\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)\underset{\sim}{Q}(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)']. \tag{4.3.4}$$

We now prove that if $\underset{\sim}{\theta}^* \neq \underset{\sim}{\theta}$ then

$$\underset{n\to\infty}{\text{p lim}}\, F(\underset{\sim}{\theta}^*) > \underset{n\to\infty}{\text{p lim}}\, F(\underset{\sim}{\theta}) , \tag{4.3.5}$$

which implies that in the probability limit F has a unique minimum at $\underset{\sim}{\theta}^* = \underset{\sim}{\theta}$. Now

$$\underset{n\to\infty}{\text{p lim}}\, F(\underset{\sim}{\theta}^*) - \underset{n\to\infty}{\text{p lim}}\, F(\underset{\sim}{\theta})$$

$$= \log|\underset{\sim}{V}^*| + \mathrm{tr}\underset{\sim}{V}^{*-1}[\underset{\sim}{V}+(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)\underset{\sim}{Q}(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)'] - \log|\underset{\sim}{V}|-p \tag{4.3.6}$$

$$\geq \log|\underset{\sim}{V}^*| - \log|\underset{\sim}{V}| + \mathrm{tr}\underset{\sim}{V}^{*-1}\underset{\sim}{V} - p \tag{4.3.7}$$

$$= \mathrm{tr}\underset{\sim}{V}^{*-1}\underset{\sim}{V} - \log|\underset{\sim}{V}^{*-1}\underset{\sim}{V}| - p$$

$$= \sum_{i=1}^{p} (\lambda_i - \log \lambda_i - 1) \tag{4.3.8}$$

$$\geq 0 , \tag{4.3.9}$$

where (4.3.6) follows from (4.3.4), (4.3.7) follows from the fact that

$(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)\underset{\sim}{Q}(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)'$ is a non-negative definite quadratic form, (4.3.8) follows by letting $\lambda_i = ch_i(\underset{\sim}{V}^{*-1}\underset{\sim}{V})$ and the relations of traces and determinants to latent roots, and (4.3.9) follows from the fact that if x is non-negative $x-\log x-1 \geq 0$. This part of the proof is based on an argument of Watson (1964) reported by Rao (1965, p. 449) in connection with maximum likelihood estimation of the parameters of a multinormal distribution.

If $\underset{\sim}{\theta}$ is identified, $\underset{\sim}{\theta}^* \neq \underset{\sim}{\theta}$ implies that $\underset{\sim}{\beta}^* \neq \underset{\sim}{\beta}$ or $\underset{\sim}{V}^* \neq \underset{\sim}{V}$ (or both). If $\underset{\sim}{\beta}^* \neq \underset{\sim}{\beta}$ the quadratic form $(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)\underset{\sim}{Q}(\underset{\sim}{\beta}-\underset{\sim}{\beta}^*)'$ is positive definite and thus the inequality in (4.3.7) is strict. If $\underset{\sim}{V}^* \neq \underset{\sim}{V}$ then the latent roots of $\underset{\sim}{V}^{*-1}\underset{\sim}{V}$ are not all unity and thus the inequality in (4.3.9) is strict. Hence the strict inequality in (4.3.5).

Since F is a continuous differentiable function of the structural parameter, and from (4.3.5) in the probability limit it has a unique minimum at $\underset{\sim}{\theta}^* = \underset{\sim}{\theta}$, its partial derivatives must vanish there. Thus there exists a root of $\partial F/\partial\underset{\sim}{\theta} = \underset{\sim}{0}$ which is consistent for $\underset{\sim}{\theta}$. If there are multiple roots care must be exercised to avoid local minima. To this end we define $\hat{\underset{\sim}{\theta}}_n$ as a root satisfying

$$F(\hat{\underset{\sim}{\theta}}_n) \leq F(\underset{\sim}{\theta}^*) \tag{4.3.10}$$

for all permissible $\underset{\sim}{\theta}^*$. By (4.3.5), a root so determined is consistent, and for sufficiently large n is unique. This part of the proof is adapted from Rao (1965, p. 300) and Kendall and Stuart (1961, pp. 40-41). Note that the proof does not depend on distributional assumptions. □

In our experience with the latent linear model, roots of $\partial F/\partial \underset{\sim}{\theta} = \underset{\sim}{0}$ found so far have been unique. On the basis of experience accumulated in factor analysis, see Jöreskog (1967), we conjecture that estimates such that $\hat{\psi}_i > \varepsilon$ $(i=1,\ldots,p)$ are unique, while in improper cases where $\hat{\psi}_i = \varepsilon$ for some i there may be multiple solutions giving equal minimum values of F.

4.4 The Information Matrix for Structural Linear Models

We now introduce the concept of the limiting Fisher information matrix $\underset{\sim}{I}(\underset{\sim}{\theta})$, and derive expressions for the elements of $\underset{\sim}{I}(\underset{\sim}{\theta})$ for the family of structural linear models considered in the previous section.

4.4.1 The Limiting Fisher Information Matrix

Let $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$ be independent random vectors and let $\underset{\sim}{x}_\alpha$ have density $f_\alpha(\underset{\sim}{x};\underset{\sim}{\theta})$ depending on a parameter $\underset{\sim}{\theta}$, $(\alpha=1,\ldots,n)$. It is not necessary that each density depend on all the elements of $\underset{\sim}{\theta}$, but each θ_i should appear in some of the densities.

The Fisher information matrix relative to $\underset{\sim}{\theta}$ associated with the α-th observation is

$$\underset{\sim}{I}_\alpha(\underset{\sim}{\theta}) = E\left[\frac{\partial \log f_\alpha}{\partial \underset{\sim}{\theta}} \frac{\partial \log f_\alpha}{\partial \underset{\sim}{\theta}'}\right], \tag{4.4.1}$$

which under certain regularity conditions, see Zacks (1971, pp. 182-183), may be written as

$$\underset{\sim}{I}_\alpha(\underset{\sim}{\theta}) = -E\left[\frac{\partial^2 \log f_\alpha}{\partial \underset{\sim}{\theta}\partial \underset{\sim}{\theta}'}\right]. \tag{4.4.2}$$

Since $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$ are independent, the Fisher information matrix associated with the first n observations is $\sum_{\alpha=1}^{n} \underset{\sim}{I}_\alpha(\underset{\sim}{\theta})$. We now introduce

Definition 4.4.1. The *limiting* Fisher information matrix relative to $\underset{\sim}{\theta}$ associated with the sequence of independent random vectors $\{\underset{\sim}{\chi}_\alpha\}$ is defined as

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \frac{1}{n} \sum_{\alpha=1}^{n} \underset{\sim}{I}_\alpha(\underset{\sim}{\theta}) \ , \tag{4.4.3}$$

provided the indicated limit exists.

If the $\underset{\sim}{\chi}_\alpha$ are i.i.d., then $\underset{\sim}{I}(\underset{\sim}{\theta})$ reduces to the ordinary Fisher information matrix. This definition was motivated by the work of Wald (1948).

By analogy with the standard result (4.1.1) for the case of i.i.d. observations, it becomes natural to ask under what conditions on the sequence of densities $\{f_\alpha\}$ as $n \to \infty$

$$\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] \ , \tag{4.4.4}$$

where $\hat{\underset{\sim}{\theta}}_n$ is the m.l.e. of $\underset{\sim}{\theta}$ based on n independent observations (obtained maximizing $\log L = \sum_\alpha \log f_\alpha$) , and $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix. This problem has been considered by Bradley and Gart (1962), but the general conditions given there are rather difficult to verify in specific cases. In §4.5 we therefore give a direct proof of (4.4.4) for the case of structural linear models. First, however, we obtain the second derivatives of F and derive expressions for the elements of $\underset{\sim}{I}(\underset{\sim}{\theta})$.

4.4.2 Second Derivatives of F.

Lemma 4.4.1. Let F be the function defined in (3.2.6) or (4.3.1), and let the reduced form parameters be twice differentiable functions of a structural parameter $\underset{\sim}{\theta}$. Then

$$\frac{\partial F}{\partial \theta_i} = -2\mathrm{tr}\left[\frac{1}{n}\underset{\sim}{A}_n(\underset{\sim}{X}_n-\underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T}_n)\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\right], \qquad (4.4.5)$$

and

$$\frac{\partial^2 F}{\partial \theta_i \partial \theta_j} = 2\mathrm{tr}\left[\frac{1}{n}\underset{\sim}{A}_n\underset{\sim}{A}_n'\frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_j}\right] + 2\mathrm{tr}\left[\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\frac{1}{n}\underset{\sim}{A}_n(\underset{\sim}{X}_n-\underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right]$$

$$-2\mathrm{tr}\left[\frac{1}{n}\underset{\sim}{A}_n(\underset{\sim}{X}_n-\underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial^2 \underset{\sim}{\beta}}{\partial \theta_i \partial \theta_j}\right]$$

$$-2\mathrm{tr}\left[\frac{1}{n}\underset{\sim}{A}_n(\underset{\sim}{X}_n-\underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right] - \mathrm{tr}\left[\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right]$$

$$+2\mathrm{tr}\left[\underset{\sim}{V}^{-1}\underset{\sim}{T}_n\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T}_n)\underset{\sim}{V}^{-1}\frac{\partial^2 \underset{\sim}{V}}{\partial \theta_i \partial \theta_j}\right]. \qquad (4.4.6)$$

Proof: In our proof we use repeatedly several results on matrix differentiation given in the appendix, particularly (A.3.16), (A.3.20) (A.3.21) and (A.3.22).

Differentiating F with respect to θ_i using (A.3.16) gives

$$\frac{\partial F}{\partial \theta_i} = \mathrm{tr}\left[\frac{\partial F}{\partial \underset{\sim}{\beta}'}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right] + \mathrm{tr}\left[\frac{\partial F}{\partial \underset{\sim}{V}'}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\right], \qquad (4.4.7)$$

and using Lemma 3.2.1 for the derivatives of F with respect to $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ we obtain

$$\frac{\partial F}{\partial \theta_i} = -\frac{2}{n}\mathrm{tr}\left[\underset{\sim}{A}_n(\underset{\sim}{X}_n-\underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T}_n)\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}\right], \qquad (4.4.8)$$

where $\underset{\sim}{T}_n$ is defined in (4.2.13). This proves (4.4.5).

Let us introduce the notations

$$\underset{\sim}{C}_i = \underset{\sim}{A}_n(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}, \tag{4.4.9}$$

and

$$\underset{\sim}{D}_i = \underset{\sim}{V}^{-1}(\underset{\sim}{V} - \underset{\sim}{T}_n)\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial \theta_i}; \tag{4.4.10}$$

so that on differentiating (4.4.8) with respect to θ_j we obtain

$$\frac{\partial^2 F}{\partial \theta_i \partial \theta_j} = -\frac{2}{n}\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \theta_j} + \frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial \theta_j}. \tag{4.4.11}$$

We now differentiate $\mathrm{tr}\underset{\sim}{C}_i$ with respect to θ_j. Using (A.3.16)

$$\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \theta_j} = \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \underset{\sim}{\beta}'}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_j}\right] + \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \underset{\sim}{V}'}\frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right]$$

$$+ \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial(\partial \underset{\sim}{\beta}/\partial \theta_i)'}\frac{\partial^2 \underset{\sim}{\beta}}{\partial \theta_i \partial \theta_j}\right]. \tag{4.4.12}$$

The derivatives needed in (4.4.12) are as follows

$$\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \underset{\sim}{\beta}'} = -\frac{\partial}{\partial \underset{\sim}{\beta}'}\mathrm{tr}\left[\underset{\sim}{A}_n\underset{\sim}{A}_n'\underset{\sim}{\beta}'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right]$$

$$= -\underset{\sim}{A}_n\underset{\sim}{A}_n'\frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i}\underset{\sim}{V}^{-1} \quad \text{by (A.3.21)}, \tag{4.4.13}$$

$$\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial \underset{\sim}{V}} = \frac{\partial}{\partial \underset{\sim}{V}}\mathrm{tr}\left[\underset{\sim}{A}_n(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right]$$

$$= -\underset{\sim}{V}^{-1}\left\{\frac{\partial}{\partial \underset{\sim}{V}^{-1}}\mathrm{tr}\left[\underset{\sim}{A}_n(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right]\right\}\underset{\sim}{V}^{-1} \quad \text{by (A.3.20)}$$

$$= -\underset{\sim}{V}^{-1}(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)\underset{\sim}{A}_n'\frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i}\underset{\sim}{V}^{-1} \quad \text{by (A.3.21)}, \tag{4.4.14}$$

and

$$\frac{\partial \mathrm{tr}\underset{\sim}{C}_i}{\partial(\partial\underset{\sim}{\beta}/\partial\theta_i)} = \frac{\partial}{\partial(\partial\underset{\sim}{\beta}/\partial\theta_i)} \mathrm{tr}\left[\underset{\sim}{A}_n(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)'\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{\beta}}{\partial\theta_i}\right]$$

$$= \underset{\sim}{V}^{-1}(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)\underset{\sim}{A}_n' \quad \text{by (A.3.21).} \qquad (4.4.15)$$

Substituting (4.4.13) - (4.4.15) into (4.4.12) and taking care to use the transposes in the last two cases, gives $\partial \mathrm{tr}\underset{\sim}{C}_i/\partial\theta_j$. Using this in conjunction with (4.4.11) gives the first three terms of (4.4.6).

We now differentiate $\mathrm{tr}\underset{\sim}{D}_i$ with respect to θ_j. Using (A.3.16)

$$\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial\theta_j} = \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial\underset{\sim}{\beta}'}\frac{\partial\underset{\sim}{\beta}}{\partial\theta_j}\right] + \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial\underset{\sim}{V}'}\frac{\partial\underset{\sim}{V}}{\partial\theta_j}\right]$$

$$+ \mathrm{tr}\left[\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial(\partial\underset{\sim}{V}/\partial\theta_i)'}\frac{\partial^2\underset{\sim}{V}}{\partial\theta_i\partial\theta_j}\right]. \qquad (4.4.16)$$

The derivatives needed in (4,4.16) are as follows.

$$\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial\underset{\sim}{\beta}} = \frac{\partial}{\partial\underset{\sim}{\beta}} \mathrm{tr}\left[\underset{\sim}{V}^{-1}\underset{\sim}{T}_n\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\right]$$

$$= -\frac{2}{n}\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\underset{\sim}{V}^{-1}(\underset{\sim}{X}_n - \underset{\sim}{\beta}\underset{\sim}{A}_n)\underset{\sim}{A}_n', \qquad (4.4.17)$$

by an argument similar to that used in (A.4.5) - (A.4.10);

$$\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial\underset{\sim}{V}} = \frac{\partial}{\partial\underset{\sim}{V}} \mathrm{tr}\left[\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i} - \underset{\sim}{V}^{-1}\underset{\sim}{T}_n\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\right]$$

$$= -\underset{\sim}{V}^{-1}\left\{\frac{\partial}{\partial\underset{\sim}{V}^{-1}}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\right]\right.$$

$$\left. - \frac{\partial}{\partial\underset{\sim}{V}^{-1}}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\underset{\sim}{T}_n\underset{\sim}{V}^{-1}\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\right]\right\}\underset{\sim}{V}^{-1}, \text{ by (A.3.20)}$$

$$= -\underset{\sim}{V}^{-1}\left[\frac{\partial \underset{\sim}{V}}{\partial\theta_i} - 2\frac{\partial \underset{\sim}{V}}{\partial\theta_i}\underset{\sim}{V}^{-1}\underset{\sim}{T}_n\right]\underset{\sim}{V}^{-1}, \quad \text{by (A.3.21) and (A.3.22)}$$

$$= -\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\theta_i}V^{-1} + 2\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\theta_i}\underset{\sim}{V}^{-1}\underset{\sim}{T}_n\underset{\sim}{V}^{-1}, \qquad (4.4.18)$$

and finally

$$\frac{\partial \mathrm{tr}\underset{\sim}{D}_i}{\partial(\partial \underset{\sim}{V}/\partial\theta_i)} = \frac{\partial}{\partial(\partial \underset{\sim}{V}/\partial\theta_i)}\mathrm{tr}\left[\underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T}_n)\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\theta_i}\right]$$

$$= \underset{\sim}{V}^{-1}(\underset{\sim}{V}-\underset{\sim}{T}_n)\underset{\sim}{V}^{-1} \quad \text{by (A.3.21).} \qquad (4.4.19)$$

Substituting (4.4.17) - (4.4.19) into (4.4.16) and taking the transposes as required gives $\partial \mathrm{tr}\underset{\sim}{D}_i/\partial\theta_j$. Using this in (4.4.11) gives the last four terms in (4.4.6). This completes the proof of the lemma. □

4.4.3 The Limiting Fisher Information Matrix of Structural Linear Models.

We can now prove the following result.

Theorem 4.4.2. Let $\underset{\sim}{X}_n$ satisfy (4.2.1) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n)$, (n=r+1,r+2,...), and let the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ be twice differentiable functions of the structural parameter $\underset{\sim}{\theta}$. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 then the elements of the limiting Fisher information matrix of Definition 4.4.1 are given by

$$I(\theta_i,\theta_j) = \mathrm{tr}\left[\underset{\sim}{Q}\frac{\partial \underset{\sim}{\beta}'}{\partial\theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{\beta}}{\partial\theta_j}\right] + \frac{1}{2}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\theta_i}\underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\theta_j}\right]. \qquad (4.4.20)$$

Further, if $\underset{\sim}{\theta}$ is identified then $\underset{\sim}{I}(\underset{\sim}{\theta})$ is positive definite.

Proof: Under the assumed regularity conditions, $\underset{\sim}{I}(\underset{\sim}{\theta})$ is given by

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = - \lim_{n\to\infty} \frac{1}{n} E\left[\frac{\partial^2 \log L}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'}\right] , \tag{4.4.21}$$

and if $\underset{\sim}{\theta}$ is identified then $\underset{\sim}{I}(\underset{\sim}{\theta})$ is non-singular, see Silvey (1959, pp. 81-82).

Since $\log L = c - \frac{1}{2}nF$ where c is a constant we have

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = \frac{1}{2}\lim_{n\to\infty} E\left[\frac{\partial^2 F}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'}\right] . \tag{4.4.22}$$

Consider the second derivatives of F given in Lemma 4.4.1. By Assumption 4.2.1, as $n \to \infty$ the first term in (4.4.6), which contains no random variables, converges to

$$2\mathrm{tr}\left[\underset{\sim}{Q} \frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{\beta}}{\partial \theta_j}\right] . \tag{4.4.23}$$

Under the assumptions of tne model, for each n

$$E(\underset{\sim}{X}_n) = \underset{\sim}{\beta}\underset{\sim}{A}_n . \tag{4.4.24}$$

Using this result in (4.4.6) we see that on taking expectations the second, third and fourth terms vanish.

Also under the assumptions of the model, for each n

$$E[\underset{\sim}{T}_n] = \frac{1}{n} \Sigma E(\underset{\sim}{u}_\alpha \underset{\sim}{u}_\alpha') = \underset{\sim}{V} . \tag{4.4.25}$$

Using this result in (4.4.6) we see that on taking expectations the seventh term vanishes, while the sixth becomes

$$2\ \mathrm{tr}\left[\underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right] . \tag{4.4.26}$$

On subtracting from this the fifth term in (4.4.6), which is non-

stochastic and does not depend on n, and recalling expression (4.4.23) for the first term, we obtain

$$\lim_{n\to\infty} E\,\frac{\partial^2 F}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'} = 2\mathrm{tr}\left[\underset{\sim}{Q}\,\frac{\partial\underset{\sim}{\beta}'}{\partial\theta_i}\,\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{\beta}}{\partial\theta_j}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\,\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\right]. \quad (4.4.27)$$

In view of (4.4.22), the theorem follows from (4.4.27) after dividing by 2. □

Jennrich (1970) asserted that if $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ are i.i.d. $N_p(\underset{\sim}{\mu},\underset{\sim}{V})$ where $\underset{\sim}{\mu}$ and $\underset{\sim}{V}$ depend on a structural parameter $\underset{\sim}{\theta}$, then the elements of the (ordinary) Fisher information matrix are given by

$$I(\theta_i,\theta_j) = \left[\frac{\partial\underset{\sim}{\mu}'}{\partial\theta_i}\,\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{\mu}}{\partial\theta_j}\right] + \tfrac{1}{2}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{V}}{\partial\theta_i}\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{V}}{\partial\theta_j}\right], \quad (4.4.28)$$

and indicated that the result can be proved by appealing to a result concerning the expected value of the product of four jointly normally distributed random variables. In a later paper, Jennrich (1974) used this result to derive simplified formulae for asymptotic standard errors in factor analysis, in which case $\underset{\sim}{\mu}$ does not depend on structural parameters and thus the first term on the right hand side of (4.4.28) drops out. See in this regard §1.3.4.

We remark that (4.4.28) may be obtained as a special case of Theorem 4.4.2 by letting $r = 1$ and $\underset{\sim}{A}_n = \underset{\sim}{1}_n'$, and noting that $\underset{\sim}{Q}_n = \frac{1}{n}\,\underset{\sim}{1}_n'\underset{\sim}{1}_n = 1$ for all n, and $\underset{\sim}{\beta} = \underset{\sim}{\mu}$ is $p\times 1$ so that the quantity inside the brackets in the first term of (4.4.20) is a scalar. Note that by taking second derivatives we do not need to appeal in the proof to results other than the first two moments of $\underset{\sim}{U}_n$.

We now consider estimation of $\underset{\sim}{I}(\underset{\sim}{\theta})$.

Lemma 4.4.3. Let $\hat{\underset{\sim}{\theta}}_n$ be the m.l.e. of $\underset{\sim}{\theta}$. Then under the assumptions of Theorem 4.4.2, as $n \to \infty$

$$\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\theta}) , \tag{4.4.29}$$

i.e. $\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n)$ is a consistent estimator of $\underset{\sim}{I}(\underset{\sim}{\theta})$.

Proof: By Theorem 4.3.1, $\hat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}$.

Since $\partial\underset{\sim}{\beta}/\partial\theta_i$, $\partial\underset{\sim}{V}/\partial\theta_i$ and $\underset{\sim}{V}^{-1}$ are continuous functions of $\underset{\sim}{\theta}$, their values at $\hat{\underset{\sim}{\theta}}_n$ will converge stochastically to their values at the true parameter value $\underset{\sim}{\theta}$.

Since the elements of $\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n)$ are in turn continuous functions of $\partial\underset{\sim}{\beta}/\partial\theta_i$, $\partial\underset{\sim}{V}/\partial\theta_i$ and $\underset{\sim}{V}^{-1}$ evaluated at $\hat{\underset{\sim}{\theta}}_n$, they will converge stochastically to $\underset{\sim}{I}(\underset{\sim}{\theta})$. This completes the proof. □

In actual practice, the matrix $\underset{\sim}{Q}$ appearing in $\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n)$ would be replaced by $\underset{\sim}{Q}_n$, which by Assumption 4.2.1 converges to $\underset{\sim}{Q}$ as $n \to \infty$.

4.5 Asymptotic Normality of $\hat{\underset{\sim}{\theta}}_n$ in Structural Linear Models.

We are now ready to prove asymptotic normality of the m.l.e. $\hat{\underset{\sim}{\theta}}_n$ in structural linear models. We first note that since by (4.3.2) the function F depends on the observations only through $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$, the estimator $\hat{\underset{\sim}{\theta}}_n$ is an implicit function of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ defined by $\partial F(\overline{\underset{\sim}{\beta}}_n, \overline{\underset{\sim}{V}}_n; \hat{\underset{\sim}{\theta}}_n)/\partial\underset{\sim}{\theta} = \underset{\sim}{0}$. If first and second derivatives of this implicit function exist in a neighborhood of $\underset{\sim}{\theta}$, and if the asymptotic joint distribution of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ is multivariate normal, then the limiting distribution of $\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta})$ is normal; see for example Theorem 4.2.5 in Anderson (1958, pp. 76-77). While this approach

shows rather clearly the nature of the problem, derivation of the asymptotic distribution is rather difficult, for it involves the derivatives of the implicit function defining $\hat{\underset{\sim}{\theta}}_n$ in terms of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$. We shall therefore adopt a more indirect approach. For clarity in the argument we prove first

Lemma 4.5.1. Let F be the function defined in (3.2.6) or (4.3.1), and let the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ be twice differentiable functions of structural parameter $\underset{\sim}{\theta}$ which is identified. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, then as $n \to \infty$

$$\frac{1}{2} \frac{\partial^2 F}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'} \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\theta}) , \tag{4.5.1}$$

where $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix given in Theorem 4.4.2 and $\underset{\sim}{\theta}$ is any permissible value of the structural parameter vector.

Proof: Consider the second derivatives of F given in (4.4.6) of Lemma 4.4.1. By Assumption 4.2.1, as $n \to \infty$ the first term in (4.4.6) converges to

$$2\mathrm{tr}\left[\underset{\sim}{Q} \frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{\beta}}{\partial \theta_j}\right] . \tag{4.5.2}$$

Using (4.2.2) we can write

$$\frac{1}{n} \underset{\sim}{A}_n (\underset{\sim}{X}_n - \underset{\sim}{\beta} \underset{\sim}{A}_n)' = \underset{\sim}{Q}_n (\overline{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})' . \tag{4.5.3}$$

Since by Lemma 4.2.1 $\overline{\underset{\sim}{\beta}}_n \xrightarrow{p} \underset{\sim}{\beta}$ and since $\underset{\sim}{Q}_n \to \underset{\sim}{Q}$, under the assumptions of the lemma, as $n \to \infty$

$$\frac{1}{n} \underset{\sim}{A}_n (\underset{\sim}{X}_n - \underset{\sim}{\beta} \underset{\sim}{A}_n)' \xrightarrow{p} \underset{\sim}{0} . \tag{4.5.4}$$

This result implies that as $n \to \infty$ the second, third and fourth terms in (4.4.6) converge stochastically to zero.

Recall now from (4.2.17) that as $n \to \infty$

$$\underset{\sim}{T}_n \xrightarrow{p} \underset{\sim}{V} . \tag{4.5.5}$$

This result implies that as $n \to \infty$ the seventh term in (4.4.6) converges stochastically to zero, while the sixth term converges stochastically to

$$2\mathrm{tr}\left[\underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_i}\right] . \tag{4.5.6}$$

On subtracting from this the fifth term, which does not depend on n, and recalling expression (4.5.2) for the first term, we find that as $n \to \infty$

$$\frac{\partial^2 F}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'} \xrightarrow{p} 2\mathrm{tr}\left[\underset{\sim}{Q} \frac{\partial \underset{\sim}{\beta}'}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{\beta}}{\partial \theta_j}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_i} \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_j}\right] \tag{4.5.7}$$

This lemma then follows by comparing (4.5.7) with (4.4.20) in Theorem 4.4.2, which gives $\underset{\sim}{I}(\underset{\sim}{\theta})$. □

Suppose now that we evaluate the second partial derivatives of F at a value $\tilde{\underset{\sim}{\theta}}_n$ which converges stochastically to $\underset{\sim}{\theta}$. Let $\tilde{\underset{\sim}{\beta}}_n$ and $\tilde{\underset{\sim}{V}}_n$ correspond to $\tilde{\underset{\sim}{\theta}}_n$. Since these are continuous functions of $\tilde{\underset{\sim}{\theta}}_n$ we have $\tilde{\underset{\sim}{\beta}}_n \xrightarrow{p} \underset{\sim}{\beta}$ and $\tilde{\underset{\sim}{V}}_n \xrightarrow{p} \underset{\sim}{V}$, and thus as $n \to \infty$

$$(\tilde{\underset{\sim}{\beta}}_n - \bar{\underset{\sim}{\beta}}_n) \xrightarrow{p} 0 \tag{4.5.8}$$

$$(\tilde{\underset{\sim}{V}}_n - \bar{V}_n) \xrightarrow{p} 0 . \tag{4.5.9}$$

Using these results and proceeding as in the proof of Lemma 4.5.1 we obtain

Corollary 4.5.2. Under the assumptions of Lemma 4.5.1, if $\tilde{\underset{\sim}{\theta}}_n$ is a vector converging stochastically to $\underset{\sim}{\theta}$ then as $n \to \infty$

$$\frac{1}{2}\frac{\partial^2 F(\tilde{\underset{\sim}{\theta}}_n)}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'} \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\theta}) , \qquad (4.5.10)$$

where $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix of Theorem 4.4.2.

We are now ready to establish the main result of this section.

Theorem 4.5.3. Let $\underset{\sim}{X}_n$ satisfy (4.2.1) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V}\otimes\underset{\sim}{I}_n)$, $(n=r+1,r+2,\ldots)$ and let the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ be twice differentiable functions of a structural parameter $\underset{\sim}{\theta}$ which is identified. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, then as $n \to \infty$

$$\sqrt{n}\,(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} N[\underset{\sim}{0},\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] , \qquad (4.5.11)$$

where $\hat{\underset{\sim}{\theta}}_n$ is the m.l.e. of $\underset{\sim}{\theta}$ and $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix given in Theorem 4.4.2.

Proof: Consider the Taylor Series expansion of $\partial F(\hat{\underset{\sim}{\theta}}_n)/\partial\underset{\sim}{\theta}$ about the true parameter value $\underset{\sim}{\theta}$.

$$\frac{\partial F(\hat{\underset{\sim}{\theta}}_n)}{\partial\underset{\sim}{\theta}} = \frac{\partial F(\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}} + \frac{\partial F(\tilde{\underset{\sim}{\theta}}_n)}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) = \underset{\sim}{0} , \qquad (4.5.12)$$

where $\tilde{\underset{\sim}{\theta}}_n$ is a point in the line segment connecting $\hat{\underset{\sim}{\theta}}$ and $\underset{\sim}{\theta}$. On inserting a factor of $\frac{\sqrt{n}}{2}$ we obtain

$$-\frac{\sqrt{n}}{2}\frac{\partial F(\underset{\sim}{\theta})}{\partial\underset{\sim}{\theta}} = \frac{1}{2}\frac{\partial F(\tilde{\underset{\sim}{\theta}}_n)}{\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}'}\sqrt{n}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) . \qquad (4.5.13)$$

Since by Theorem 4.3.1, $\hat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}$ as $n \to \infty$ and since $\tilde{\underset{\sim}{\theta}}_n = \nu\hat{\underset{\sim}{\theta}}_n + (1-\nu)\underset{\sim}{\theta}$ for some $0 \le \nu \le 1$, it follows that as $n \to \infty$ $\tilde{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta}$. Then by Corollary 4.5.2, as $n \to \infty$

$$\frac{\partial F(\tilde{\underset{\sim}{\theta}}_n)}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'} \xrightarrow{p} \frac{1}{2} \underset{\sim}{I}(\underset{\sim}{\theta}) \quad . \tag{4.5.14}$$

Using this result on (4.5.13) and noting that by identifiability $I(\theta)$ is non-singular, we obtain that

$$\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \overset{p}{=} - \underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \cdot \frac{\sqrt{n}}{2} \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \quad , \tag{4.5.15}$$

where the symbol $\overset{p}{=}$ in (4.5.15) indicates that the vectors on the left and right hand side are asymptotically stochastically equivalent, i.e. as $n \to \infty$ their difference converges in probability to zero.

Consider now the random vector $\frac{\sqrt{n}}{2} \partial F(\underset{\sim}{\theta})/\partial \underset{\sim}{\theta}$. From (4.4.5) and (4.5.3),

$$\frac{\partial F(\underset{\sim}{\theta})}{\partial \theta_i} = 2\mathrm{tr}\left[\underset{\sim}{Q}_n (\bar{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})' \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{\beta}}{\partial \theta_i}\right] + \mathrm{tr}\left[\underset{\sim}{V}^{-1} (\underset{\sim}{T}_n - \underset{\sim}{V}) \underset{\sim}{V}^{-1} \frac{\partial \underset{\sim}{V}}{\partial \theta_i}\right] \quad . \tag{4.5.16}$$

Since $E(\bar{\beta}_n) = \beta$ and $E(T_n) = V$, we have from (4.5.16)

$$E\left[\frac{\sqrt{n}}{2} \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right] = \underset{\sim}{0} \quad . \tag{4.5.17}$$

Consider now the variance-covariance matrix of $\frac{\sqrt{n}}{2} \partial F(\underset{\sim}{\theta})/\partial \underset{\sim}{\theta}$. This depends on the variances and covariances of $\underset{\sim}{\beta}_n$ and $\underset{\sim}{T}_n$, but it is easier to derive from the relationship between F and $\log L$. Since $\log L = c - \tfrac{1}{2} n F$, $\partial F(\underset{\sim}{\theta})/\partial \underset{\sim}{\theta} = \frac{2}{n} \partial \log L / \partial \underset{\sim}{\theta}$ and thus

$$\mathrm{Var}\left[\frac{\sqrt{n}}{2} \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right] = \frac{n}{4} E\left[\frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \cdot \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}\right] = \frac{1}{n} E\left[\frac{\partial \log L}{\partial \underset{\sim}{\theta}} \cdot \frac{\partial \log L}{\partial \underset{\sim}{\theta}'}\right] \quad . \tag{4.5.18}$$

But under the assumed regularity conditions

$$\frac{1}{n} E\left[\frac{\partial \log L}{\partial \underset{\sim}{\theta}} \cdot \frac{\partial \log L}{\partial \underset{\sim}{\theta}'}\right] = -\frac{1}{n} E\left[\frac{\partial^2 \log L}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'}\right] = \frac{1}{n} \sum_{\alpha=1}^{n} \underset{\sim}{I}_\alpha(\underset{\sim}{\theta}) \quad , \tag{4.5.19}$$

where $\underset{\sim}{I}_\alpha(\underset{\sim}{\theta})$ is the information matrix associated with the α-th observation, see (4.4.2).

Using (4.5.19) in (4.5.18) we find that the variance-covariance matrix is

$$\mathrm{Var}\left[\frac{\sqrt{n}}{2}\frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right] = \frac{1}{n}\sum_{\alpha=1}^{n}\underset{\sim}{I}_\alpha(\underset{\sim}{\theta}) \quad , \tag{4.5.20}$$

which converges to $\underset{\sim}{I}(\underset{\sim}{\theta})$, the limiting Fisher information, as $n \to \infty$.

Having obtained the means and covariances, note from (4.5.16) that the vector $\partial F(\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ is a function of $\overline{\underset{\sim}{\beta}}_n$ and $\overline{V}_n$ with first and second derivatives existing for all $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ in a neighborhood of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$. Furthermore, by Lemma 4.2.2 the asymptotic joint distribution of $\sqrt{n}(\overline{\underset{\sim}{\beta}}_n - \underset{\sim}{\beta})$ and $\sqrt{n}(\overline{\underset{\sim}{V}}_n - \underset{\sim}{V})$ is multivariate normal under the conditions of the present theorem. Therefore by Theorem 4.2.5 in Anderson (1958, p. 77), the limiting distribution of $\frac{\sqrt{n}}{2}\partial F(\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ is multivariate normal. In view of (4.5.17) and (4.5.20),

$$\frac{\sqrt{n}}{2}\frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \xrightarrow{d} N[\underset{\sim}{0},\ \underset{\sim}{I}(\underset{\sim}{\theta})] \ , \tag{4.5.21}$$

where $I(\underset{\sim}{\theta})$ is the limiting Fisher information matrix. From this result we obtain that

$$\underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \cdot \frac{\sqrt{n}}{2}\frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] \ . \tag{4.5.22}$$

Then the theorem follows by recalling (4.5.15) and by Slutzky's theorem, see for example result (ix) in Rao (1965, p. 101). The basic argument in the proof is adapted from Kendall and Stuart (1961, pp. 54-55). □

In this theorem we have assumed that $\underset{\sim}{U} \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ and have taken advantage of this condition in obtaining the asymptotic variance-covariance matrix. From (4.5.16), however, it is clear that $\sqrt{n}(\hat{\underset{\sim}{\theta}}_n-\underset{\sim}{\theta})$ will have the asymptotic distribution (4.5.11) whenever the *joint* asymptotic distribution of $\sqrt{n}(\bar{\underset{\sim}{\beta}}_n-\underset{\sim}{\beta})$ and $\sqrt{n}(\bar{\underset{\sim}{V}}_n-\underset{\sim}{V})$ is as given in Lemma 4.2.2. If interest centers on structural parameters that appear only in $\underset{\sim}{\beta}$ or only in $\underset{\sim}{V}$, as it is the case in factor analysis, then it is only required that the asymptotic *marginal* distributions of $\sqrt{n}(\bar{\underset{\sim}{\beta}}_n-\underset{\sim}{\beta})$ and $\sqrt{n}(\bar{\underset{\sim}{V}}_n-\underset{\sim}{V})$ be as given in Lemma 4.2.2. See in this regard the discussion in §4.2.3.

4.6 Large Sample Theory for the Latent Linear Model

Let us now specialize the results obtained in the previous section to the case of the latent linear model. We summarize our results in the following

Corollary 4.6.1. Let $\underset{\sim}{X}_n$ satisfy the latent linear model (2.2.4) - (2.2.6) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \times \underset{\sim}{I}_n)$ $(n=r+1,r+2,\ldots)$. Assume that the model has been identified in the manner described in §2.3.1, let $\underset{\sim}{\theta}$ denote a vector containing the free parameters in $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, and let $\hat{\underset{\sim}{\theta}}_n$ be the m.l.e. of $\underset{\sim}{\theta}$ discussed in Chapter 3. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, then as $n \to \infty$

$$\hat{\underset{\sim}{\theta}}_n \xrightarrow{p} \underset{\sim}{\theta} , \tag{4.6.1}$$

i.e., $\hat{\underset{\sim}{\theta}}_n$ is a consistent estimator of $\underset{\sim}{\theta}$;

$$\sqrt{n}(\hat{\underset{\sim}{\theta}}_n-\underset{\sim}{\theta}) \xrightarrow{d} N[\underset{\sim}{0},\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] , \tag{4.6.2}$$

where $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix of Definition

4.4.1; and furthermore

$$\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\theta}) \ , \tag{4.6.3}$$

i.e. $\underset{\sim}{I}(\underset{\sim}{\theta})$ is estimated consistently by $\underset{\sim}{I}(\hat{\underset{\sim}{\theta}}_n)$.

Proof: The proof follows from Theorem 4.3.1, Lemma 4.4.3 and Theorem 4.5.3, and the fact that in (2.2.5) and (2.2.6) $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ are twice differentiable functions of $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, and hence of $\underset{\sim}{\theta}$. □

It remains only to obtain expressions for the elements of the limiting Fisher information matrix. To this end we give two auxiliary lemmas.

Lemma 4.6.2. The derivatives of the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ with respect to the elements of the structural parameter matrices $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ of (2.2.5) and (2.2.6) are given by

$$\frac{\partial \underset{\sim}{\beta}}{\partial \xi_{ij}} = \underset{\sim}{\Lambda} \underset{\sim}{E}_{ij} \ , \tag{4.6.4}$$

$$\frac{\partial \underset{\sim}{\beta}}{\partial \lambda_{ij}} = \underset{\sim}{E}_{ij} \underset{\sim}{\Xi} \ , \tag{4.6.5}$$

$$\frac{\partial \underset{\sim}{V}}{\partial \lambda_{ij}} = \underset{\sim}{E}_{ij} \underset{\sim}{\Phi} \underset{\sim}{\Lambda}' + \underset{\sim}{\Lambda} \underset{\sim}{\Phi} \underset{\sim}{E}_{ji} \ , \tag{4.6.6}$$

$$\frac{\partial \underset{\sim}{V}}{\partial \phi_{ij}} = \frac{1}{2}(2-\delta_{ij}) \underset{\sim}{\Lambda} (\underset{\sim}{E}_{ij} + \underset{\sim}{E}_{ji}) \underset{\sim}{\Lambda}' = \begin{cases} \underset{\sim}{\Lambda} \underset{\sim}{E}_{ii} \underset{\sim}{\Lambda}' & \text{if } i = j \\ \underset{\sim}{\Lambda} \underset{\sim}{E}_{ij} \underset{\sim}{\Lambda}' + \underset{\sim}{\Lambda} \underset{\sim}{E}_{ji} \underset{\sim}{\Lambda}' & \text{if } i \neq j \ , \end{cases} \tag{4.6.7}$$

and

$$\frac{\partial \underset{\sim}{V}}{\partial \psi_i} = \underset{\sim}{E}_{ii} \ . \tag{4.6.8}$$

Proof: This result is a corollary to Lemma 3.2.2, giving the analogous matrix derivatives. Results (4.6.4), (4.6.5) and (4.6.7) follow from (A.3.4), (4.6.6) follows from (A.3.7), and (4.6.8) follows from first principles.

Lemma 4.6.3. Let $\underset{\sim}{E}_{ij}$: p×q be a matrix with 1 in the (i,j)-th position and 0 elsewhere, and let $\underset{\sim}{E}_{k\ell}$: r×s be similarly defined. Also let $\underset{\sim}{A} = (a_{ij})$ be q×r and $\underset{\sim}{B} = (b_{ij})$ be r×p. Then

$$\mathrm{tr}[\underset{\sim}{E}_{ij}\underset{\sim}{A}\underset{\sim}{E}_{k\ell}\underset{\sim}{B}] = a_{jk}b_{\ell i} \ . \tag{4.6.9}$$

Proof: By direct computation

$$\underset{\sim}{E}_{ij}\underset{\sim}{A} = i \begin{bmatrix} 0 & ,\ldots, & 0 \\ & \vdots & \\ a_{j1} & ,\ldots, & a_{jr} \\ & \vdots & \\ 0 & ,\ldots, & 0 \end{bmatrix} : p\times r \ , \tag{4.6.10}$$

with a similar result holding for $\underset{\sim}{E}_{k\ell}\underset{\sim}{B}$. Thus

$$\underset{\sim}{E}_{ij}\underset{\sim}{A}\underset{\sim}{E}_{k\ell}\underset{\sim}{B} = i \begin{bmatrix} 0 & ,\ldots, & 0 \\ & \vdots & \\ a_{jk}b_{\ell 1} & ,\ldots, & a_{jk}b_{\ell p} \\ & \vdots & \\ 0 & ,\ldots, & 0 \end{bmatrix} : p\times p \ , \tag{4.6.11}$$

and the only non-zero diagonal element of (4.6.11) is $a_{jk}b_{\ell i}$. Hence the lemma. □

We are now ready to prove

Theorem 4.6.4. Let $\underset{\sim}{X}_n$ satisfy the latent linear model (2.2.4) - (2.2.6) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ (n=r+1,r+2,...). If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, then the elements of the limiting Fisher information matrix with respect

to $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are given by

$$I(\xi_{ij},\xi_{k\ell}) = (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{Q})_{j\ell} \quad , \tag{4.6.12}$$

$$I(\xi_{ij},\lambda_{k\ell}) = (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki}(\underset{\sim}{Q\Xi}')_{j\ell} \quad , \tag{4.6.13}$$

$$I(\lambda_{ij},\lambda_{k\ell}) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{\Xi Q\Xi}')_{j\ell} + (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{j\ell} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{i\ell}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj} \quad , \tag{4.6.14}$$

$$I(\lambda_{ij},\phi_{k\ell}) = \tfrac{1}{2}(2-\delta_{k\ell})[(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{\ell j} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj}] \quad , \tag{4.6.15}$$

$$I(\lambda_{ij},\psi_{k}) = (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj} \quad , \tag{4.6.16}$$

$$I(\phi_{ij},\phi_{k\ell}) = \tfrac{1}{4}(2-\delta_{ij})(2-\delta_{k\ell})[(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{jk}] \quad , \tag{4.6.17}$$

$$I(\phi_{ij},\psi_{k}) = \tfrac{1}{2}(2-\delta_{ij})(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj} \quad , \text{ and} \tag{4.6.18}$$

$$I(\psi_{i},\psi_{k}) = \tfrac{1}{2}(\underset{\sim}{V}^{-1})^{2}_{ik} \quad , \tag{4.6.19}$$

the remaining elements, such as $I(\xi_{ij},\phi_{k\ell})$ being all zero.

Proof: The proof follows from Theorem 4.4.2 and repeated application of Lemmas 4.6.2 and 4.6.3, as follows:

$$I(\xi_{ij},\xi_{k\ell}) = \operatorname{tr}\left[\underset{\sim}{Q}\,\frac{\partial\underset{\sim}{\beta}'}{\partial\xi_{ij}}\,\underset{\sim}{V}^{-1}\,\frac{\partial\underset{\sim}{\beta}}{\partial\xi_{k\ell}}\right] \quad \text{by (4.4.20)} \quad ,$$

$$= \operatorname{tr}[\underset{\sim}{Q}\underset{\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{E}_{k\ell}] \quad \text{by (4.6.4)} \quad ,$$

$$= \operatorname{tr}[\underset{\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{E}_{k\ell}\underset{\sim}{Q}] \quad ,$$

$$= (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{Q})_{\ell j} \quad \text{by (4.6.9)} \quad .$$

$$I(\xi_{ij},\lambda_{k\ell}) = \operatorname{tr}\left[\underset{\sim}{Q}\,\frac{\partial \underset{\sim}{\beta}'}{\partial \xi_{ij}}\,\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{\beta}}{\partial \lambda_{k\ell}}\right] \quad \text{by (4.4.20) ,}$$

$$= \operatorname{tr}[\underset{\sim\sim}{Q}\underset{\sim\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Xi}] \quad \text{by (4.6.4) and (4.6.5) ,}$$

$$= (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{ik}(\underset{\sim\sim}{\Xi}\underset{\sim\sim}{Q})_{\ell j} \quad \text{by (4.6.9) .}$$

$$I(\lambda_{ij},\lambda_{k\ell}) = \operatorname{tr}\left[\underset{\sim}{Q}\,\frac{\partial \underset{\sim}{\beta}'}{\partial \lambda_{ij}}\,\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{\beta}}{\partial \lambda_{k\ell}}\right] + \frac{1}{2}\operatorname{tr}\left[\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial \lambda_{ij}}\,\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial \lambda_{k\ell}}\right]$$

by (4.4.20) ,

$$= \operatorname{tr}[\underset{\sim\sim}{Q}\underset{\sim\sim}{\Xi}'\underset{\sim\sim}{E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Xi}]$$
$$+ \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}(\underset{\sim\sim}{E}_{ij}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'+\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{ji})\underset{\sim}{V}^{-1}(\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'+\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{\ell k})]$$

by (4.6.5) and (4.6.6) ,

$$= \operatorname{tr}[\underset{\sim\sim}{Q}\underset{\sim\sim}{\Xi}'\underset{\sim\sim}{E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Xi}]$$
$$+ \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{ij}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'] + \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{ij}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{\ell k}]$$
$$+ \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{k\ell}\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'] + \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi}\underset{\sim\sim}{E}_{\ell k}] ,$$

$$= (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim\sim}{\Xi}\underset{\sim\sim}{Q}\underset{\sim\sim}{\Xi}')_{\ell j}$$
$$+ \tfrac{1}{2}(\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{jk}(\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{\ell i} + \tfrac{1}{2}(\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{j\ell}(\underset{\sim}{V}^{-1})_{ki}$$
$$+ \tfrac{1}{2}(\underset{\sim}{V}^{-1})_{ik}(\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{\ell j} + \tfrac{1}{2}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{i\ell}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{kj}$$

by (4.6.9) ,

$$= (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim\sim}{\Xi}\underset{\sim\sim}{Q}\underset{\sim\sim}{\Xi}')_{\ell j}$$
$$+ (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim\sim}{\Phi}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{\ell j} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{i\ell}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{\Phi})_{kj} .$$

$$I(\lambda_{ij},\phi_{k\ell}) = \tfrac{1}{2}\mathrm{tr}\ \underset{\sim}{V}^{-1}\left[\frac{\partial \underset{\sim}{V}}{\partial\lambda_{ij}}\ \underset{\sim}{V}^{-1}\frac{\partial \underset{\sim}{V}}{\partial\phi_{k\ell}}\right] \quad \text{by (4.4.20)},$$

$$= \tfrac{1}{4}(2-\delta_{k\ell})\ \mathrm{tr}[\underset{\sim}{V}^{-1}(\underset{\sim}{E}_{ij}\underset{\sim}{\Phi\Lambda}'+\underset{\sim}{\Lambda\Phi E}_{ji})\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{E}_{k\ell}+\underset{\sim}{E}_{\ell k})\underset{\sim}{\Lambda}']$$

by (4.6.6) and (4.6.7),

$$= \tfrac{1}{4}(2-\delta_{k\ell})\{\mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{E}_{ij}\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{k\ell}\underset{\sim}{\Lambda}'] + \mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{E}_{ij}\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{\ell k}\underset{\sim}{\Lambda}']$$

$$+ \mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{k\ell}\underset{\sim}{\Lambda}'] + \mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{\ell k}\underset{\sim}{\Lambda}']\},$$

$$= \tfrac{1}{4}(2-\delta_{k\ell})[(\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{jk}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{\ell i} + (\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{ki}$$

$$+ (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{\ell j} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj}]$$

by (4.6.9),

$$= \tfrac{1}{2}(2-\delta_{k\ell})[(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{\ell j} + (\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj}].$$

$$I(\lambda_{ij},\psi_k) = \tfrac{1}{2}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\ \frac{\partial \underset{\sim}{V}}{\partial\lambda_{ij}}\ \underset{\sim}{V}^{-1}\ \frac{\partial \underset{\sim}{V}}{\partial\psi_k}\right] \quad \text{by (4.4.20)}$$

$$= \tfrac{1}{2}\mathrm{tr}[\underset{\sim}{V}^{-1}(\underset{\sim}{E}_{ij}\underset{\sim}{\Phi\Lambda}'+\underset{\sim}{\Lambda\Phi E}_{ji})\underset{\sim}{V}^{-1}\underset{\sim}{E}_{kk}] \quad \text{by (4.6.6) and (4.6.8)}$$

$$= \tfrac{1}{2}\mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{E}_{ij}\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{E}_{kk}] + \tfrac{1}{2}\mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi E}_{ji}\underset{\sim}{V}^{-1}\underset{\sim}{E}_{kk}],$$

$$= \tfrac{1}{2}(\underset{\sim}{\Phi\Lambda}'\underset{\sim}{V}^{-1})_{jk}(\underset{\sim}{V}^{-1})_{ki} + \tfrac{1}{2}(\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj} \quad \text{by (4.6.9)}$$

$$= (\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda\Phi})_{kj}.$$

$$I(\phi_{ij},\phi_{k\ell}) = \tfrac{1}{2}\mathrm{tr}\left[\underset{\sim}{V}^{-1}\ \frac{\partial \underset{\sim}{V}}{\partial\phi_{ij}}\ \underset{\sim}{V}^{-1}\ \frac{\partial \underset{\sim}{V}}{\partial\phi_{k\ell}}\right] \quad \text{by (4.4.20)}$$

$$= \tfrac{1}{8}(2-\delta_{ij})(2-\delta_{k\ell})\mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{E}_{ij}+\underset{\sim}{E}_{ji})\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}(\underset{\sim}{E}_{k\ell}+\underset{\sim}{E}_{\ell k})\underset{\sim}{\Lambda}'] \quad \text{by (4.6.7)}$$

$$= \tfrac{1}{8}(2-\delta_{ij})(2-\delta_{k\ell})\{\mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{ij}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{k\ell}\underset{\sim}{\Lambda}'] + \mathrm{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{ij}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda E}_{\ell k}\underset{\sim}{\Lambda}']$$

$$+ \operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{k\ell}\underset{\sim}{\Lambda}'] + \operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{\ell k}\underset{\sim}{\Lambda}']\}$$

$$= \tfrac{1}{8}(2-\delta_{ij})(2-\delta_{k\ell})[(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{jk}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{\ell i} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki}$$

$$+ (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{\ell j} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj}] \text{ by (4.6.9)}$$

$$= \tfrac{1}{4}(2-\delta_{ij})(2-\delta_{k\ell})[(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{j\ell} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{i\ell}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{jk}] \,.$$

$$I(\phi_{ij},\psi_k) = \tfrac{1}{2}\operatorname{tr}\,\underset{\sim}{V}^{-1}\left[\frac{\partial \underset{\sim}{V}}{\partial\phi_{ij}}\,\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial\psi_k}\right] \quad \text{by (4.4.20)} \,,$$

$$= \tfrac{1}{4}(2-\delta_{ij})\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}(\underset{\sim\sim}{E}_{ij}+\underset{\sim\sim}{E}_{ji})\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{kk}] \quad \text{by (4.6.7) and (4.6.8)}$$

$$= \tfrac{1}{4}(2-\delta_{ij})\{\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{ij}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{kk}] + \operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim\sim}{E}_{ji}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{kk}]\}$$

$$= \tfrac{1}{4}(2-\delta_{ij})[(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{jk}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki} + (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj}] \quad \text{by (4.6.9)}$$

$$= \tfrac{1}{2}(2-\delta_{ij})(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ki}(\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{kj} \,.$$

$$I(\psi_i,\psi_k) = \tfrac{1}{2}\operatorname{tr}\left[\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial\psi_i}\,\underset{\sim}{V}^{-1}\,\frac{\partial \underset{\sim}{V}}{\partial\psi_k}\right] \quad \text{by (4.4.20)} \,,$$

$$= \tfrac{1}{2}\operatorname{tr}[\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{ii}\underset{\sim}{V}^{-1}\underset{\sim\sim}{E}_{kk}] \qquad \text{by (4.6.8)} \,,$$

$$= \tfrac{1}{2}(\underset{\sim}{V}^{-1})_{ik}(\underset{\sim}{V}^{-1})_{ki} \qquad \text{by (4.6.9)} \,,$$

$$= \tfrac{1}{2}(\underset{\sim}{V}^{-1})^2_{ik} \,.$$

This completes the proof. □

It may be noted that all terms not involving $\underset{\sim}{Q}$ in the above expressions agree with the corresponding results for confirmatory factor analysis given in §1.4.2. Compare (4.6.14) - (4.6.19) with (1.4.8)-(1.4.13) which have been obtained by a different method.

The theorem gives the elements of the limiting information matrix in terms of $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$. To obtain the elements with respect to the free parameters $\underset{\sim}{\theta}$, let $\underset{\sim}{\gamma}$ be a vector containing all the elements of $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$, and $\underset{\sim}{\Psi}$ ordered in lexicographical fashion. Then

$$I(\theta_i,\theta_j) = \sum_{k,\ell} \alpha_{ik}\alpha_{j\ell} I(\gamma_k,\gamma_\ell) \ , \tag{4.6.20}$$

where $\alpha_{ik} = 1$ if $\theta_i = \gamma_k$ and 0 otherwise, and the $I(\gamma_k,\gamma_\ell)$ are obtained from Theorem 4.6.4.

The same method of proof of Theorem 4.6.4 could be used to obtain the second derivatives of F with respect to $\underset{\sim}{\theta}$, using Lemma 4.4.1 in combination with the auxiliary Lemmas 4.6.2 and 4.6.3. These derivatives could then be used in a Newton-Raphson algorithm for computing the estimates. The results, however, are too complicated to be of much practical use, and the usually large number of parameters to be estimated iteratively would generally make the Fletcher-Powell method more efficient than Newton-Raphson.

An alternative approach that could be used to obtain second derivatives is to differentiate the first derivatives given in Theorem 3.2.3. In all but one case, however, this is an arduous task, even with the help of the matrix differentiation rules given in the Appendix, primarily because of the lack of a convenient chain rule for matrix functions of matrices. We now give one second derivative that serves to check the results obtained in this section. From (3.2.14)

$$\frac{\partial F}{\partial \underset{\sim}{\Xi}} = -\frac{2}{n} \underset{\sim}{\Lambda}' \underset{\sim}{V}^{-1} (\underset{\sim}{X} - \underset{\sim}{\Lambda}\underset{\sim}{\Xi}\underset{\sim}{A}_n)\underset{\sim}{A}_n' \ . \tag{4.6.21}$$

Differentiating this with respect to $\underset{\sim}{\Xi}$ we obtain

$$\frac{\partial}{\partial \underset{\sim}{\Xi}}\left[\frac{\partial F}{\partial \underset{\sim}{\Xi}}\right] = 2\,\frac{\partial}{\partial \underset{\sim}{\Xi}}[\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{\Xi}\,\frac{1}{n}\,\underset{\sim}{A}_n\underset{\sim}{A}_n'] \ , \tag{4.6.22}$$

$$= 2(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda} \otimes \underset{\sim}{I}_q)\underset{\sim}{E}_{(q,r)}(\underset{\sim}{Q}_n \otimes \underset{\sim}{I}_r) \ , \tag{4.6.23}$$

where (4.6.23) follows from (A.3.3). Collecting elements in different blocks of (4.6.23), or using (A.3.4), we obtain

$$\frac{\partial}{\partial \xi_{k\ell}}\left(\frac{\partial F}{\partial \underset{\sim}{\Xi}}\right) = 2\,\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda}\underset{\sim}{E}_{k\ell}\underset{\sim}{Q}_n \ . \tag{4.6.24}$$

But the (i,j)-th element of $\underset{\sim}{A}\underset{\sim}{E}_{k\ell}\underset{\sim}{B}$ is $a_{ik}b_{\ell j}$. Thus

$$\frac{\partial^2 F}{\partial \xi_{ij}\partial \xi_{k\ell}} = 2(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{Q}_n)_{j\ell} \ . \tag{4.6.25}$$

Since under Assumption 4.2.1 $\underset{\sim}{Q}_n \to \underset{\sim}{Q}$ as $n \to \infty$, we find, recalling (4.4.22) and noting that (4.6.25) is non-stochastic, that

$$I(\xi_{ij},\xi_{k\ell}) = (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})_{ik}(\underset{\sim}{Q})_{j\ell}, \tag{4.6.26}$$

which agrees with (4.6.12) of Theorem 4.6.4. We remark that this result could not be verified by comparison with analogous results for factor analysis, because it involves $\underset{\sim}{Q}$.

4.7 Approximate Second Derivatives of $\tilde{F}$

Recall from Lemma 4.5.1 that as $n \to \infty$

$$\frac{1}{2}\,\frac{\partial^2 F}{\partial \underset{\sim}{\theta}\partial \underset{\sim}{\theta}'} \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\theta}) \ . \tag{4.7.1}$$

Thus for sufficiently large n, approximate vaues of the second derivatives of F with respect to the free parameters are given by twice the limiting Fisher information matrix. Hence $\frac{1}{2}\,\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})$

provides a good initial value for the matrix $\underset{\sim}{E}$ in the Fletcher-Powell iterative procedure for minimizing F discussed in §1.3.2.

When $\underset{\approx}{\Xi}$ is unrestricted, however, estimates of the parameters are computed by minimizing the function $\tilde{F}$, which is F minimized over $\underset{\approx}{\Xi}$ for fixed $\underset{\sim}{\Lambda}$ $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, as indicated in §3.5 and §3.6. We now show how approximate vaues of the second derivatives of $\tilde{F}$ may be obtained from $\underset{\sim}{I}(\underset{\sim}{\theta})$. For this purpose we prove

Lemma 4.7.1. Let F be the function defined in (3.2.6) or (4.3.1), and let the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$ be twice differentiable functions of a structural parameter $\underset{\sim}{\theta}$ which is identified. Let $\underset{\sim}{\theta}$ be partitioned into components $\underset{\sim}{\theta}_1$: $m_1 \times 1$ and $\underset{\sim}{\theta}_2$: $m_2 \times 1$, and let $\tilde{F}$ be the function F minimized over $\underset{\sim}{\theta}_2$ for a fixed $\underset{\sim}{\theta}_1$. Suppose the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1, so that the limiting Fisher information matrix $\underset{\sim}{I}(\underset{\sim}{\theta})$ exists, and let $\underset{\sim}{I}(\underset{\sim}{\theta})$ be partitioned into blocks $\underset{\sim}{I}_{ij}(\underset{\sim}{\theta})$: $m_i \times m_j$ for i,j=1,2. Then as $n \to \infty$

$$\frac{1}{2} \frac{\partial^2 \tilde{F}}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_1'} \xrightarrow{p} \underset{\sim}{I}_{11}(\underset{\sim}{\theta}) - \underset{\sim}{I}_{12}(\underset{\sim}{\theta}) \underset{\sim}{I}_{22}^{-1}(\underset{\sim}{\theta}) \underset{\sim}{I}_{21}(\underset{\sim}{\theta}) . \tag{4.7.2}$$

Proof: The function $\tilde{F}$ is given by

$$\tilde{F}(\underset{\sim}{\theta}_1) = \min_{\underset{\sim}{\theta}_2} F(\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2) = F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)] , \tag{4.7.3}$$

where $\tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)$ is the value of $\underset{\sim}{\theta}_2$ that minimizes F for given $\underset{\sim}{\theta}_1$.

Interpret $\tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)$ as the root of

$$\frac{\partial F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2} = \underset{\sim}{0} . \tag{4.7.4}$$

Differentiating (4.7.4) with respect to $\underset{\sim}{\theta}_1'$ using the chain rule

for vector derivatives (A.3.23), we find

$$\frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_1'} + \frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_2'} \frac{\partial \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1'} = \underset{\sim}{0} \quad . \tag{4.7.5}$$

Differentiating (4.7.3) in the same fashion we obtain

$$\frac{\partial \tilde{F}(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1'} = \frac{\partial F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_1'} + \frac{\partial F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2'} \frac{\partial \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1'} , \tag{4.7.6}$$

$$= \frac{\partial F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_1'} \quad \text{by (4.7.4).} \tag{4.7.7}$$

Transposing (4.7.7) and differentiating with respect to $\underset{\sim}{\theta}_1'$ using (A.3.23) gives

$$\frac{\partial^2 \tilde{F}(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_1'} = \frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_1'} + \frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_2'} \frac{\partial \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1'} \quad . \tag{4.7.8}$$

But from (4.7.5),

$$\frac{\partial \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1'} = - \left[\frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_2'} \right]^{-1} \frac{\partial^2 F[\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)]}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_1'} , \tag{4.7.9}$$

provided the indicated inverse exists. Substituting this into (4.7.8) and writing $\tilde{\underset{\sim}{\theta}}_2$ for $\tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1)$ we have

$$\frac{\partial^2 \tilde{F}(\underset{\sim}{\theta}_1)}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_1'} = \frac{\partial^2 F(\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2)}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_1'} - \frac{\partial^2 F(\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2)}{\partial \underset{\sim}{\theta}_1 \partial \underset{\sim}{\theta}_2'} \left[\frac{\partial^2 F(\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2)}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_2'} \right]^{-1} \frac{\partial^2 F(\underset{\sim}{\theta}_1, \tilde{\underset{\sim}{\theta}}_2)}{\partial \underset{\sim}{\theta}_2 \partial \underset{\sim}{\theta}_1'} , \tag{4.7.10}$$

provided the indicated inverse exists.

From Theorem 4.3.1, $\tilde{\underset{\sim}{\theta}}_2(\underset{\sim}{\theta}_1) \xrightarrow{p} \underset{\sim}{\theta}_2$ as $n \to \infty$. Thus from Corollary 4.5.2 , as $n \to \infty$

$$\frac{1}{2}\frac{\partial^2 F(\underset{\sim}{\theta}_1,\tilde{\underset{\sim}{\theta}}_2)}{\partial\underset{\sim}{\theta}_i\partial\underset{\sim}{\theta}_j'} \xrightarrow{p} \underset{\sim}{I}_{ij}(\underset{\sim}{\theta}) \ , \quad (i,j=1,2) \ . \tag{4.7.11}$$

Since $\underset{\sim}{\theta}$ is identified, $\underset{\sim}{I}(\underset{\sim}{\theta})$ is non-singular. In view of (4.7.11) this implies that the inverse required in (4.7.9) - (4.7.10) will exist for sufficiently large n. On using (4.7.11) on (4.7.10) we find that as $n \to \infty$

$$\frac{1}{2}\frac{\partial^2 \tilde{F}(\underset{\sim}{\theta}_1)}{\partial\underset{\sim}{\theta}_1\partial\underset{\sim}{\theta}_1'} \xrightarrow{p} \underset{\sim}{I}_{11}(\underset{\sim}{\theta}) - \underset{\sim}{I}_{12}(\underset{\sim}{\theta})\underset{\sim}{I}_{22}^{-1}(\underset{\sim}{\theta})\underset{\sim}{I}_{21}(\underset{\sim}{\theta}) \ . \tag{4.7.12}$$

This completes the proof. (We are grateful to Professor Hoeffding for outlining the proof of a similar result for the case of θ_1 and θ_2 scalars. The above proof is a straightforward extension of his argument to the multiparameter case.) □

In view of the theorem, a good initial value for the matrix $\underset{\sim}{E}$ in the Fletcher-Powell iterative procedure of §3.5 for minimizing $\tilde{F}$ is given by

$$\frac{1}{2}[\underset{\sim}{I}_{11}(\underset{\sim}{\theta}) - \underset{\sim}{I}_{12}(\underset{\sim}{\theta})\underset{\sim}{I}_{22}^{-1}(\underset{\sim}{\theta})\underset{\sim}{I}_{21}(\underset{\sim}{\theta})]^{-1} \ . \tag{4.7.13}$$

By a well-known result concerning the inverses of partitioned matrices, see Morrison (1967, p. 66); the matrix (4.7.13) is half the matrix in the upper left $m_1 \times m_1$ block of $\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})$.

V. HYPOTHESIS TESTING

5.1 Introduction

In this chapter we consider large sample tests of hypotheses in the latent linear model. The discussion is organized in three parts. In §5.2 we develop a likelihood ratio goodness of fit test for a family of structural linear models, obtain the asymptotic distribution of the test statistic if the model fits, and discuss application of the procedure in the latent linear model. In §5.3 we consider briefly testing hypotheses about structural parameters after the structure has been established, with special reference to hypotheses involving the parameters $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ in the latent linear model. In §5.4 we discuss testing linear hypotheses about the parameter $\underset{\sim}{\Xi}$ in the latent linear model and give a procedure for computing the restricted m.l.e. of $\underset{\sim}{\Xi}$, which is needed to construct the likelihood ratio test statistic. We also consider an alternative test procedure using Wald's statistic, which does not require computation of the restricted m.l.e., and derive the large sample distributions of the test statistic under the null hypothesis and under a sequence of alternative hypotheses.

5.2 Testing Goodness of Fit

Let the matrix $\underset{\sim}{X}_n$ be given by

$$\underset{\sim}{X}_n = \underset{\sim}{\beta}\underset{\sim}{A}_n + \underset{\sim}{U}_n \ , \tag{5.2.1}$$

where $\underset{\sim}{A}_n$ is of full row rank $r < n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0},\underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ with

$\underset{\sim}{V}$ p.d.

Let Ω be the Cartesian product of the set of all $p \times r$ matrices $\underset{\sim}{\beta}$ and the set of all $p \times p$ symmetric p.d. matrices $\underset{\sim}{V}$, and let ω be the subset of Ω where $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ are specified functions of a vector $\underset{\sim}{\theta}$.

We consider testing

$$H_0: \{\underset{\sim}{\beta},\underset{\sim}{V}\} \in \omega \qquad \text{vs.} \qquad H_1: \{\underset{\sim}{\beta},\underset{\sim}{V}\} \in \Omega - \omega , \tag{5.2.2}$$

i.e., the goodness of fit of a structural linear model where the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ are functions of a structural parameter $\underset{\sim}{\theta}$.

Let $\ell = pr + \frac{1}{2}p(p+1)$ denote the number of parameters in $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and let m denote the number of parameters in $\underset{\sim}{\theta}$. Then H_0 imposes $\ell - m$ constraints upon $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and we require $m < \ell$ for H_0 to be non-trivial.

Let $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$ be the unrestricted m.l.e.'s of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ given in (4.2.2) and (4.2.3), and let $\overline{\underset{\sim}{\gamma}}_n: \ell \times 1$ be a vector containing all distinct elements in $\overline{\underset{\sim}{\beta}}_n$ and $\overline{\underset{\sim}{V}}_n$. Then the maximum value of log L in Ω is

$$\max_{\Omega} \log L(\underset{\sim}{\beta},\underset{\sim}{V}) = c - \tfrac{1}{2}nF(\overline{\underset{\sim}{\gamma}}_n) , \tag{5.2.3}$$

$$= c - \tfrac{1}{2}n[\log|\overline{\underset{\sim}{V}}_n| + p] , \tag{5.2.4}$$

where c is a constant and F is the function defined in (3.2.6).

Let $\hat{\underset{\sim}{\theta}}_n: m \times 1$ denote the m.l.e. of the structural parameter $\underset{\sim}{\theta}$, let $\hat{\underset{\sim}{\beta}}_n$ and $\hat{\underset{\sim}{V}}_n$ be the corresponding restricted m.l.e.'s of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and let $\hat{\underset{\sim}{T}}_n = \frac{1}{n}(\underset{\sim}{X}_n - \hat{\underset{\sim}{\beta}}_n\underset{\sim}{A}_n)(\underset{\sim}{X}_n - \hat{\underset{\sim}{\beta}}_n\underset{\sim}{A}_n)'$. Then the maximum value of log L

in ω is

$$\max_{\omega} \log L(\underset{\sim}{\beta},\underset{\sim}{V}) = c - \tfrac{1}{2}nF(\hat{\underset{\sim}{\theta}}_n) , \tag{5.2.5}$$

$$= c - \tfrac{1}{2}n[\log|\hat{\underset{\sim}{V}}_n| + \operatorname{tr} \hat{\underset{\sim}{V}}_n^{-1}\hat{\underset{\sim}{T}}_n], \tag{5.2.6}$$

where we have written F as a function of $\underset{\sim}{\theta}$.

The likelihood ratio statistic for testing (5.2.2) is then given by

$$-2 \log \lambda_n = -2[\max_{\omega} \log L - \max_{\Omega} \log L] \tag{5.2.7}$$

$$= n[F(\hat{\underset{\sim}{\theta}}_n) - F(\bar{\underset{\sim}{\gamma}}_n)] \tag{5.2.8}$$

$$= n[\log|\hat{\underset{\sim}{V}}_n| + \operatorname{tr} \hat{\underset{\sim}{V}}_n^{-1}\hat{\underset{\sim}{T}}_n - \log|\hat{\underset{\sim}{V}}_n| - p] . \tag{5.2.9}$$

We now obtain the asymptotic null distribution of $-2 \log \lambda_n$.

Theorem 5.2.1. Let $\underset{\sim}{X}_n$ satisfy (5.2.1) for $(n=r+1,r+2,\ldots)$, and let the reduced form parameters $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ be twice differentiable functions of a structural parameter $\underset{\sim}{\theta}$ which is identified. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 then as $n \to \infty$

$$-2 \log \lambda_n \xrightarrow{d} \chi^2_\nu , \tag{5.2.10}$$

where $-2 \log \lambda_n$ is as defined in (5.2.7) - (5.2.9) and $\nu = \ell - m$, the difference between the number of reduced form and structural parameters.

Proof: The proof is adapted from the argument used by Rao (1965, pp. 347-351) to obtain the asymptotic null distribution of likelihood ratio test statistics based on i.i.d. observations.

Let $\underset{\sim}{\gamma}$ be a vector containing the distinct elements of $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and let $\bar{\underset{\sim}{\gamma}}_n$ be its unrestricted m.l.e. Consider a Taylor series expansion of $F(\underset{\sim}{\gamma})$ about $\underset{\sim}{\gamma} = \bar{\underset{\sim}{\gamma}}_n$:

$$F(\underset{\sim}{\gamma}) = F(\bar{\underset{\sim}{\gamma}}_n) + (\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n)' \frac{\partial F(\bar{\underset{\sim}{\gamma}}_n)}{\partial \underset{\sim}{\gamma}} + \frac{1}{2}(\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n)' \frac{\partial^2 F(\tilde{\underset{\sim}{\gamma}}_n)}{\partial \underset{\sim}{\gamma}\partial \underset{\sim}{\gamma}'}(\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n) \qquad (5.2.11)$$

where $\tilde{\underset{\sim}{\gamma}}_n$ is a point on the line segment connecting $\bar{\underset{\sim}{\gamma}}_n$ and $\underset{\sim}{\gamma}$.

Since $\bar{\underset{\sim}{\gamma}}_n$ is a root of $\partial F/\partial \underset{\sim}{\gamma} = \underset{\sim}{0}$, the second term of the right hand side of (5.2.11) is zero.

Since $\tilde{\underset{\sim}{\gamma}}_n = \nu\bar{\underset{\sim}{\gamma}}_n + (1-\nu)\underset{\sim}{\gamma}$ for some $0 \le \nu \le 1$, and since by Lemma 4.2.1 $\bar{\underset{\sim}{\gamma}}_n \xrightarrow{p} \underset{\sim}{\gamma}$ as $n \to \infty$, we have that $\tilde{\underset{\sim}{\gamma}}_n \xrightarrow{p} \underset{\sim}{\gamma}$ as $n \to \infty$. Therefore by Corollary 4.5.2, as $n \to \infty$

$$\frac{1}{2} \frac{\partial^2 F(\tilde{\underset{\sim}{\gamma}}_n)}{\partial \underset{\sim}{\gamma}\partial \underset{\sim}{\gamma}'} \xrightarrow{p} \underset{\sim}{I}(\underset{\sim}{\gamma}) , \qquad (5.2.12)$$

where $\underset{\sim}{I}(\underset{\sim}{\gamma})$ denotes the limiting Fisher information matrix with respect to $\underset{\sim}{\gamma}$, which may be obtained from Theorem 4.2.2.

Using these results in (5.2.11), we obtain the following asymptotic stochastic equivalence

$$F(\underset{\sim}{\gamma}) - F(\bar{\underset{\sim}{\gamma}}_n) \overset{p}{=} (\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n)' \underset{\sim}{I}(\underset{\sim}{\gamma}) (\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n) . \qquad (5.2.13)$$

On the other hand, using expression (4.5.15) in the proof of Theorem 4.5.3, and noting that by Lemma 4.2.2 $\underset{\sim}{I}(\underset{\sim}{\gamma})$ is non-singular when $\underset{\sim}{V}$ is p.d., we have

$$(\underset{\sim}{\gamma}-\bar{\underset{\sim}{\gamma}}_n) \overset{p}{=} -\frac{1}{2}\underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} . \qquad (5.2.14)$$

On using this in (5.2.13) we obtain

$$F(\underset{\sim}{\gamma}) - F(\bar{\underset{\sim}{\gamma}}_n) \overset{p}{=} \frac{1}{4} \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}'} \underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} . \qquad (5.2.15)$$

Furthermore, from expression (4.5.21) in the proof of Theorem 4.5.3, as $n \to \infty$

$$\frac{\sqrt{n}}{2} \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{I}(\underset{\sim}{\gamma})] \ . \tag{5.2.16}$$

Under H_0, $\underset{\sim}{\gamma}$ is a function of a vector $\underset{\sim}{\theta}$ of structural parameters, admitting first derivatives

$$\frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'} = \left[\frac{\partial \underset{\sim}{\gamma}_i}{\partial \underset{\sim}{\theta}_j}\right] : \ell \times m \ . \tag{5.2.17}$$

Let $\hat{\underset{\sim}{\theta}}_n$ denote the m.l.e. of $\underset{\sim}{\theta}$ under H_0. Proceeding as in (5.2.11) - (5.2.15) we obtain the asymptotic stochastic equivalence

$$F(\underset{\sim}{\theta}) - F(\hat{\underset{\sim}{\theta}}_n) \overset{p}{=} \frac{1}{4} \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'} \underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} \ , \tag{5.2.18}$$

where $\underset{\sim}{I}(\underset{\sim}{\theta})$ denotes the limiting Fisher information matrix with respect to $\underset{\sim}{\theta}$, which is given in Theorem 4.2.2 and is non-singular if $\underset{\sim}{\theta}$ is identified.

Using the chain rule for vector derivatives (A.3.23),

$$\frac{\partial F(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'} = \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}'} \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'} \ , \tag{5.2.19}$$

and substituting this into (5.2.18) we obtain

$$F(\underset{\sim}{\theta}) - F(\hat{\underset{\sim}{\theta}}_n) \overset{p}{=} \frac{1}{4} \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}'} \left[\frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'} \underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}\right] \frac{\partial (\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} \tag{5.2.20}$$

In view of (5.2.15) and (5.2.20), the likelihood ratio statistic (5.2.8) may be written as

$$-2\log\lambda_n \overset{p}{=} \frac{n}{4} \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}'} \left[\underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'} \underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}\right] \frac{\partial F(\underset{\sim}{\gamma})}{\partial \underset{\sim}{\gamma}} \ , \tag{5.2.21}$$

since $F(\underset{\sim}{\gamma})$ and $F(\underset{\sim}{\theta})$ are the same.

Let $\underset{\sim}{D}$ denote the matrix in brackets in (5.2.21). In view of (5.2.16), the right hand side of (5.2.21) is a quadratic form in

asymptotically normally distributed random variables, and will therefore have an asymptotic χ^2_ν distribution with degrees of freedom $\nu = \text{tr } \underset{\sim}{D}\,\underset{\sim}{I}(\underset{\sim}{\gamma})$ if and only if $\underset{\sim}{D}\,\underset{\sim}{I}(\underset{\sim}{\gamma})\underset{\sim}{D} = \underset{\sim}{D}$.

To verify this condition note that by (4.4.2) and (4.4.4)

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \frac{1}{n} E\left[\frac{\partial \log L}{\partial \underset{\sim}{\theta}} \frac{\partial \log L}{\partial \underset{\sim}{\theta}'}\right],$$

$$= \lim_{n\to\infty} \frac{1}{n} E\left[\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} \frac{\partial \log L}{\partial \underset{\sim}{\gamma}} \frac{\partial \log L}{\partial \underset{\sim}{\gamma}'} \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\right] \quad \text{by (5.2.19)}$$

$$= \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} \underset{\sim}{I}(\underset{\sim}{\gamma}) \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}. \qquad (5.2.22)$$

Now

$$\underset{\sim}{D}\underset{\sim}{I}(\underset{\sim}{\gamma})\underset{\sim}{D} = \left[\underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}\right]\underset{\sim}{I}(\underset{\sim}{\gamma})\left[\underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}\right]$$

$$= \underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - 2\frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} + \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\left[\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} I(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\right]\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}$$

$$= \underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - 2\frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} + \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\underset{\sim}{I}(\underset{\sim}{\theta})\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}'} \quad \text{by (5.2.22)}$$

$$= \underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) \frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} = \underset{\sim}{D}. \qquad (5.2.23)$$

Hence as $n \to \infty$ the right hand side of (5.2.21) $\xrightarrow{d} \chi^2_\nu$ with

$$\nu = \text{tr } \underset{\sim}{D}\,\underset{\sim}{I}(\underset{\sim}{\gamma}) = \text{tr}\left[\underset{\sim}{I}^{-1}(\underset{\sim}{\gamma}) - \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}}\right]\underset{\sim}{I}(\underset{\sim}{\gamma})$$

$$= \text{tr } \underset{\sim}{I}_\ell - \text{tr } \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\left[\frac{\partial \underset{\sim}{\gamma}'}{\partial \underset{\sim}{\theta}} I(\underset{\sim}{\gamma}) \frac{\partial \underset{\sim}{\gamma}}{\partial \underset{\sim}{\theta}'}\right]$$

$$= \ell - \text{tr}[\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})\underset{\sim}{I}(\underset{\sim}{\theta})] \quad \text{by (5.2.22)}$$

$$= \ell - \text{tr } \underset{\sim}{I}_m = \ell - m. \qquad (5.2.24)$$

The theorem then follows by (5.2.21) and Slutzky's theorem.

□

Thus for sufficiently large n, an approximate size α test of (5.2.2) has critical region

$$-2 \log \lambda_n > \chi^2_{\nu,1-\alpha} . \tag{5.2.25}$$

The theorem provides a large sample goodness of fit test for the latent linear model, where $\underset{\sim}{\beta}$ and $\underset{\sim}{y}$ satisfy (2.2.5) and (2.2.6) for some structural parameter matrices $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, which are subject to a set of identifying restrictions as discussed in §2.3.1.

In exploratory studies the investigator will usually fit a sequence of unrestricted models with q latent variables, $\underset{\sim}{\Phi} = \underset{\sim}{I}_q$ and $\underset{\sim}{\Lambda}$ satisfying the Gram-Schmidt pattern (2.3.7), for different values of q. The goodness of fit statistic may then be used as an informal criterion to determine an appropriate value of q, much in the same spirit as in polynomial regression. The number of free structural parameters in a q-factor unrestricted latent linear model is $m = qr + pq + p - \frac{1}{2}q(q-1)$, and the number of reduced form parameters is $\ell = pr + \frac{1}{2}p(p+1)$. Hence the number of degrees of freedom is

$$\nu = (p-q)r + \frac{1}{2}[(p-q)^2 - (p+q)] . \tag{5.2.26}$$

This expression may be used to determine the maximum number of factors that may be fitted for given p and r. Note that on setting $r = 0$, (5.2.26) reduces to the corresponding result (1.3.31) for exploratory factor analysis.

In confirmatory studies the investigator will usually be able to specify in advance the number of latent variables, and a set of identi-

fying restrictions that is better suited to the nature of the problem at hand. As indicated in §2.3.1, these restrictions may take the form of a structural hypothesis specifying free, fixed and constrained elements in $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and possibly $\underset{\sim}{\Psi}$. The goodness of fit of the model may be tested using (5.2.25) with degrees of freedom

$$\nu = pr + \frac{1}{2}p(p+1) - m \qquad (5.2.27)$$

where m is the number of free parameters. On setting $r = 0$, (5.2.27) reduces to the corresponding result (1.4.17) for confirmatory factor analysis.

5.3 Testing Hypotheses about $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$.

So far we have discussed testing the goodness of fit of a structural linear model where $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$ are functions of a structural parameter $\underset{\sim}{\theta}$. Having found a model that fits, we may use the likelihood ratio technique to test a variety of hypotheses about the structural parameter $\underset{\sim}{\theta}$. Let g be an $s\times 1$ vector function of $\underset{\sim}{\theta}$. Then we consider testing

$$H_0: \underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0} \qquad \text{v/s} \qquad H_1: \underset{\sim}{g}(\underset{\sim}{\theta}) \neq \underset{\sim}{0}\ . \qquad (5.3.1)$$

Let $\hat{\underset{\sim}{\theta}}_n$ be the unrestricted m.l.e. of $\underset{\sim}{\theta}$, and let $\hat{\hat{\underset{\sim}{\theta}}}_n$ be the m.l.e. under the restrictions imposed by H_0. Then the likelihood ratio test statistic is

$$-2 \log \lambda_n = n[F(\hat{\hat{\underset{\sim}{\theta}}}_n) - F(\hat{\underset{\sim}{\theta}}_n)]\ . \qquad (5.3.2)$$

The hypothesis (5.3.1) is clearly equivalent to a hypothesis specifying that $\underset{\sim}{\theta}$ is a function of an $(m-s)\times 1$ vector $\underset{\sim}{\tau}$, see Rao (1965, p. 350). By the same argument used in the proof of Theorem 5.2.1,

it then follows that as $n \to \infty$

$$-2 \log \lambda_n \xrightarrow{d} \chi^2_s \ . \qquad (5.3.3)$$

From (5.2.8) it can be seen that the statistic (5.3.2) is actually the difference between the goodness of fit statistics for two structural linear models, specifying $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$ as functions of $\underset{\sim}{\tau}$ and $\underset{\sim}{\theta}$, respectively.

From a theoretical point of view the testing problem considered here is essentially the same as that considered in the previous section. From a practical point of view, however, it is convenient to distinguish between the problems of finding a structure that fits and studying characteristics of the parameters of that structure.

In the context of the latent linear model, this technique may readily be applied to test hypotheses about the structural parameters $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, of the type considered in §2.3.1, where the elements of $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ are specified to be (1) free, (2) fixed at given values, or (3) constrained to be equal to other parameters in the model. In this case the procedures discussed in Chapter 3 may be used to obtain both the unrestricted and restricted m.l.e.'s of the structural parameters, by fitting in effect two models under two sets of restrictions. Two examples of hypotheses that may be of interest are $H_0: \underset{\sim}{\Lambda} = \lambda \underset{\sim}{1}_p$ (in a one-factor model) and $H_0: \underset{\sim}{\Psi} = \psi \underset{\sim}{I}_p$, which have been considered in §2.4.3.

The same approach may be used to test hypotheses about the structural parameter $\underset{\sim}{\Xi}$, but we have not yet discussed in detail the problem of estimating $\underset{\sim}{\Xi}$ under a set of restrictions. We now consider this problem for the case of a linear hypothesis.

5.4 Testing Linear Hypotheses About $\underset{\sim}{\Xi}$.

Consider testing the general linear hypotheses

$$H_0: \underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B} = \underset{\sim}{0} \qquad \text{v/s} \qquad H_1: \underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B} \neq \underset{\sim}{0} \tag{5.4.1}$$

where $\underset{\sim}{C}$: t×q and B: r×s are matrices of known constants of full row rank $t \leq q$ and full column rank $s \leq r$ respectively.

5.4.1 The Likelihood Ratio Test

To construct a likelihood ratio test of H_0 we require the m.l.e. $\hat{\hat{\underset{\sim}{\Xi}}}_n$ of $\underset{\sim}{\Xi}$ under the restrictions $\underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B} = \underset{\sim}{0}$. In this regard we prove

Theorem 5.4.1. Let F be defined by (3.2.6). If $\underset{\sim}{\Xi}$ is subject to the linear restrictions $\underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B} = \underset{\sim}{0}$, then the value $\tilde{\tilde{\underset{\sim}{\Xi}}}$ that minimizes F conditional on fixed values of $\underset{\sim}{\Lambda}$ $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ is given by

$$\tilde{\tilde{\underset{\sim}{\Xi}}} = \tilde{\underset{\sim}{\Xi}} - (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{C}'[\underset{\sim}{C}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{C}']^{-1}\underset{\sim}{C}\tilde{\underset{\sim}{\Xi}}\underset{\sim}{B}[\underset{\sim}{B}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1}\underset{\sim}{B}]^{-1}\underset{\sim}{B}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1}, \tag{5.4.2}$$

where $\tilde{\underset{\sim}{\Xi}}$ is the value that minimizes F conditional on fixed $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ when $\underset{\sim}{\Xi}$ is unrestricted, given in (3.4.5).

Proof: Let $\underset{\sim}{L}$: s×t be a matrix of Lagrange multipliers. We minimize the function

$$f(\underset{\sim}{\Xi},\underset{\sim}{L}) = F(\underset{\sim}{\Xi}) + \text{tr}[\underset{\sim}{L}'\underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B}] \tag{5.4.3}$$

with respect to $\underset{\sim}{\Xi}$ and $\underset{\sim}{L}$.

Differentiating (5.4.3) with respect to $\underset{\sim}{\Xi}$ using (3.2.14) and (A.3.19) gives

$$\frac{\partial f}{\partial \underset{\sim}{\Xi}} = -\frac{2}{n}\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\Lambda}\underset{\sim}{\Xi}\underset{\sim}{A})\underset{\sim}{A}' + \underset{\sim}{C}'\underset{\sim}{L}\underset{\sim}{B}' , \tag{5.4.4}$$

and differentiating (5.4.3) with respect to $\underset{\sim}{L}'$ using (A.3.19) and transposing gives

$$\frac{\partial f}{\partial \underset{\sim}{L}} = \underset{\sim\sim}{C\Xi B} \ . \tag{5.4.5}$$

On setting these equations to $\underset{\sim}{0}$ we obtain, from (5.4.4)

$$\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{XA}' - \frac{n}{2}\underset{\sim}{C}'\underset{\sim\sim}{\tilde{L}B}' = \underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{\tilde{\Lambda}\Xi AA}' \ , \tag{5.4.6}$$

which under the full rank assumptions of the model gives

$$\begin{aligned}\underset{\approx\approx}{\tilde{\Xi}} &= (\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}[\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim\sim}{XA}' - \frac{n}{2}\underset{\sim}{C}'\underset{\sim\sim}{\tilde{L}B}'](\underset{\sim\sim}{AA}')^{-1} \\ &= \underset{\approx\approx}{\Xi} - \frac{n}{2}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{C}'\underset{\sim\sim}{\tilde{L}B}'(\underset{\sim\sim}{AA}')^{-1} \ , \end{aligned} \tag{5.4.7}$$

where (5.4.7) follows from expression (3.4.5) giving $\underset{\approx\approx}{\tilde{\Xi}}$.

On the other hand, from (5.4.5) we obtain

$$\underset{\sim\sim}{C\tilde{\Xi}B} = \underset{\sim}{0} \ . \tag{5.4.8}$$

On using (5.4.8) on (5.4.7) we obtain

$$\underset{\sim\sim}{C\tilde{\Xi}B} = \frac{n}{2}\underset{\sim}{C}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{C}'\underset{\sim\sim}{\tilde{L}B}'(\underset{\sim\sim}{AA}')^{-1}\underset{\sim}{B} \ , \tag{5.4.9}$$

and hence

$$\underset{\sim}{\tilde{L}} = \frac{2}{n}[\underset{\sim}{C}(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{C}']^{-1}\underset{\sim\sim}{C\tilde{\Xi}B}[\underset{\sim}{B}'(\underset{\sim\sim}{AA}')^{-1}\underset{\sim}{B}]^{-1} \ . \tag{5.4.10}$$

Substituting this into (5.4.7) gives (5.4.2). This completes the proof. □

Note that if H_0: $\underset{\approx}{\Xi} = \underset{\sim}{0}$ then $\underset{\approx\approx}{\tilde{\Xi}} = \underset{\sim}{0}$ by (5.4.2), as we would expect. Also computation of $(\underset{\sim}{\Lambda}'\underset{\sim}{V}^{-1}\underset{\sim}{\Lambda})^{-1}$ may be simplified using (3.4.3) in Lemma 3.4.1.

To obtain the unconditional restricted m.l.e. $\underset{\approx}{\overset{\approx}{\Xi}}$ of $\underset{\approx}{\Xi}$, minimize $F(\underset{\approx}{\overset{\approx}{\Xi}},\underset{\sim}{\Lambda},\underset{\sim}{\Phi},\underset{\sim}{\Psi})$ with respect to $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$ proceeding in the same manner as in §3.5 and §3.6, but using $\underset{\approx}{\overset{\approx}{\Xi}}$ given by (5.4.2) instead of $\underset{\approx}{\overset{\sim}{\Xi}}$ given by (3.4.5), and then evaluate (5.4.2) at the m.l.e.'s $\underset{\sim}{\overset{\approx}{\Lambda}}_n$, $\underset{\sim}{\overset{\approx}{\Phi}}_n$ and $\underset{\sim}{\overset{\approx}{\Psi}}_n$.

The likelihood ratio test statistic is then

$$-2 \log \lambda_n = n[F(\underset{\approx}{\overset{\approx}{\Xi}}_n,\underset{\sim}{\overset{\approx}{\Lambda}}_n,\underset{\sim}{\overset{\approx}{\Phi}}_n,\underset{\sim}{\overset{\approx}{\Psi}}_n) - F(\underset{\approx}{\hat{\Xi}}_n,\underset{\sim}{\hat{\Lambda}}_n,\underset{\sim}{\hat{\Phi}}_n,\underset{\sim}{\hat{\Psi}}_n)] \qquad (5.4.11)$$

and is asymptotically distributed as χ^2_ν with $\nu = ts$, the number of restrictions imposed by H_0.

5.4.2 The Wald Statistic

Note that computation of the likelihood ratio statistic requires fitting the model with $\underset{\approx}{\Xi}$ unrestricted, and with $\underset{\approx}{\Xi}$ restricted, and at this later stage (5.4.2) must be evaluated in each iteration. We now consider an alternative test procedure suggested by Wald (1943), which does not require calculation of restricted m.l.e.'s and thus is computationally more convenient.

Let us introduce the following definition. For any matrix $\underset{\sim}{A}$: m×n let vec $\underset{\sim}{A}$ denote an mn×1 vector containing the columns of $\underset{\sim}{A}$ packed one below another. If the matrices $\underset{\sim}{A}$, $\underset{\sim}{B}$ and $\underset{\sim}{C}$ are conformable, then it can be shown that

$$\text{vec}(\underset{\sim}{ABC}) = (\underset{\sim}{C}' \otimes \underset{\sim}{A})\text{vec } \underset{\sim}{B}, \qquad (5.4.12)$$

see for example result (2.10) in Neudecker (1969).

Let $\underset{\approx}{\xi} = \text{vec } \underset{\approx}{\Xi}$ and $\underset{\approx}{\hat{\xi}}_n = \text{vec } \underset{\approx}{\hat{\Xi}}_n$, where $\underset{\approx}{\hat{\Xi}}_n$ is the unrestricted m.l.e. of $\underset{\approx}{\Xi}$. Using (5.4.12) the linear hypothesis (5.4.1) may be

written in terms of $\underset{\sim}{\xi}$ as

$$H_0: (\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\xi} = \underset{\sim}{0} \qquad v/s \qquad H_1: (\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\xi} \neq \underset{\sim}{0}. \tag{5.4.13}$$

Let $\underset{\sim}{\theta}$ denote the vector of (free) structural parameters in the model, and suppose the elements of $\underset{\sim}{\xi}$ are the first qr elements of $\underset{\sim}{\theta}$. Let the limiting Fisher information matrix $\underset{\sim}{I}(\underset{\sim}{\theta})$ of Theorem 4.6.3 be partitioned into four blocks $\underset{\sim}{I}_{ij}(\underset{\sim}{\theta})$ $(i,j=1,2)$ such that $\underset{\sim}{I}_{11}(\underset{\sim}{\theta})$ is $qr \times qr$, and define

$$\underset{\sim}{\Sigma}(\underset{\sim}{\theta}) = [\underset{\sim}{I}_{11}(\underset{\sim}{\theta}) - \underset{\sim}{I}_{12}(\underset{\sim}{\theta})\underset{\sim}{I}_{22}^{-1}(\underset{\sim}{\theta})\underset{\sim}{I}_{21}(\underset{\sim}{\theta})]^{-1}, \tag{5.4.14}$$

the upper-left $qr \times qr$ block of $\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})$. Let $\underset{\sim}{\Sigma}(\hat{\underset{\sim}{\theta}}_n)$ denote (5.4.14) evaluated at the unrestricted m.l.e. $\hat{\underset{\sim}{\theta}}_n$ of $\underset{\sim}{\theta}$. Then the Wald (1943) statistic for testing (5.4.13) is

$$W_n = n\hat{\underset{\sim}{\xi}}_n'(\underset{\sim}{B} \otimes \underset{\sim}{C}')[(\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\hat{\underset{\sim}{\theta}}_n)(\underset{\sim}{B} \otimes \underset{\sim}{C}')]^{-1}(\underset{\sim}{B}' \otimes \underset{\sim}{C})\hat{\underset{\sim}{\xi}}_n. \tag{5.4.15}$$

We now obtain the asymptotic distribution of W_n under the null hypothesis and under a sequence of alternative hypotheses.

Theorem 5.4.2. Let $\underset{\sim}{X}_n$ satisfy the latent linear model (2.2.4) - (2.2.6) with design matrix $\underset{\sim}{A}_n$ and $\underset{\sim}{U}_n \sim N_{p\times n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n)$, $(n=r+1, r+2, \ldots)$; and with structural parameters $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, where $\underset{\sim}{\Xi}$ satisfies $\underset{\sim}{C}\underset{\sim}{\Xi}\underset{\sim}{B} = \underset{\sim}{0}$ for fixed matrices $\underset{\sim}{C}$ and $\underset{\sim}{B}$. Suppose the model is identified and let $\underset{\sim}{\theta}$ denote the vector of free parameters. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 then as $n \to \infty$

$$W_n \xrightarrow{d} \chi^2_\nu, \tag{5.4.16}$$

with degrees of freedom $\nu = ts$, the number of restrictions imposed by H_0.

Proof: By Theorem 4.5.3, as $n \to \infty$

$$\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}) \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{I}^{-1}(\underset{\sim}{\theta})] \tag{5.4.17}$$

where $\hat{\underset{\sim}{\theta}}_n$ is the m.l.e. of $\underset{\sim}{\theta}$ and $\underset{\sim}{I}(\underset{\sim}{\theta})$ is the limiting Fisher information matrix.

If $\underset{\sim}{\xi} = \text{vec}\ \underset{\approx}{\Xi}$ is given by the first qr elements of θ and $\hat{\underset{\sim}{\xi}}_n$ is its m.l.e., then as $n \to \infty$

$$\sqrt{n}(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}) \xrightarrow{d} N[\underset{\sim}{0}, \underset{\sim}{\Sigma}(\underset{\sim}{\theta})] \ ; \tag{5.4.18}$$

for $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$, defined at (5.4.14), is the variance-covariance matrix of the asymptotic marginal distribution of the first qr elements of $\sqrt{n}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta})$.

Since a linear function of asymptotically normally distributed random variables is asymptotically normal, as $n \to \infty$

$$\sqrt{n}(\underset{\sim}{B}' \otimes \underset{\sim}{C})(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}) \xrightarrow{d} N[\underset{\sim}{0}, (\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\underset{\sim}{\theta})(\underset{\sim}{B} \otimes \underset{\sim}{C}')] \ . \tag{5.4.19}$$

Now by Lemma 4.4.3, as $n \to \infty$

$$\underset{\sim}{\Sigma}(\hat{\underset{\sim}{\theta}}_n) \xrightarrow{p} \underset{\sim}{\Sigma}(\underset{\sim}{\theta}) \ . \tag{5.4.20}$$

Using this in (5.4.15), we obtain the following asymptotic stochastic equivalence

$$W_n \overset{p}{=} n\hat{\underset{\sim}{\xi}}_n'(\underset{\sim}{B} \otimes \underset{\sim}{C}')[(\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\underset{\sim}{\theta})(\underset{\sim}{B} \otimes \underset{\sim}{C}')]^{-1}(\underset{\sim}{B}' \otimes \underset{\sim}{C})\hat{\underset{\sim}{\xi}}_n \ . \tag{5.4.21}$$

In view of (5.4.19), the right-hand side of (5.4.21) is a quadratic form in asymptotically normally distributed random variables with mean $\underset{\sim}{0}$ and variance-covariance matrix equal to the inverse of the discriminant matrix in the quadratic form; and therefore converges in distribution to a central χ^2_ν distribution with $\nu = ts$.

The theorem then follows by Slutzky's Theorem. □

In view of this result, a large sample test of the hypothesis (5.4.1) of approximate size α has critical region

$$W_n > \chi^2_{\nu,1-\alpha} \cdot \tag{5.4.22}$$

Consider now a Pitman sequence of alternative hypotheses

$$H_1^{(n)}: (\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\xi} = \frac{1}{\sqrt{n}} \underset{\sim}{\eta} \quad , \tag{5.4.23}$$

where $\underset{\sim}{\eta}$ is a fixed vector. We now obtain the asymptotic distribution of W_n under this sequence of alternatives.

Theorem 5.4.3. Let $\underset{\sim}{X}_n$ satisfy the latent linear model (2.2.4) - (2.2.6) with design matrix $\underset{\sim}{A}_n$, $\underset{\sim}{U}_n \sim \underset{\sim}{N}_{p_x n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n)$, and structural parameters $\underset{\sim}{\Xi}_n, \underset{\sim}{\Lambda}, \underset{\sim}{\Phi}$ and $\underset{\sim}{\Psi}$, where $\underset{\sim}{\Xi}_n$ satisfies (5.4.23), for $(n=r+1, r+2, \ldots)$. If the sequence of design matrices $\{\underset{\sim}{A}_n\}$ satisfies Assumption 4.2.1 then as $n \to \infty$

$$W_n \xrightarrow{d} \chi^2_\nu(\delta) \quad , \tag{5.4.24}$$

with degrees of freedom $\nu = ts$ and non-centrality parameter

$$\delta = \underset{\sim}{\eta}'[(\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\underset{\sim}{\theta})(\underset{\sim}{B} \otimes \underset{\sim}{C}')]^{-1}\underset{\sim}{\eta} \quad . \tag{5.4.25}$$

Proof: Proceeding as in the proof of Theorem 5.4.2, but taking into account the fact that $\underset{\sim}{\xi}_n$ satisfies (5.4.23), we obtain that as $n \to \infty$

$$\sqrt{n}(\underset{\sim}{B}' \otimes \underset{\sim}{C})(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}_n) \xrightarrow{d} N[\underset{\sim}{\eta}, (\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\underset{\sim}{\theta})(\underset{\sim}{B} \otimes \underset{\sim}{C}')] \quad , \tag{5.4.26}$$

where $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$ is defined at (5.4.14). Note that since $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$ is a function of the limiting information matrix, and since the sequence of alternatives (5.4.23) converges to the null hypothesis, $\underset{\sim}{\Sigma}(\underset{\sim}{\theta})$ is evaluated under the null hypothesis.

Consider now the asymptotic stochastic equivalence (5.4.21). In view of (5.4.26), the right-hand side of (5.4.21) is now a quadratic form in asymptotically normally distributed random variables with mean $\underset{\sim}{\eta}$ and variance-covariance matrix equal to the inverse of the discriminant matrix in the quadratic form, and therefore converges in distribution to a non-central $\chi^2_\nu(\delta)$ distribution with $\nu = ts$ and δ as given in (5.4.25).

The theorem then follows by Slutzky's Theorem. □

In view of this result, for sufficiently large n, the Wald test of the null hypothesis (5.4.1) with critical region (5.4.22) has approximate power against the (fixed) alternative H_1: $(\underset{\sim}{B}' \times \underset{\sim}{C})\underset{\sim}{\xi} = \underset{\sim}{\eta}$ given by

$$\Pr\{\chi^2_\nu(\delta) > \chi^2_{\nu,1-\alpha}\} , \tag{5.4.27}$$

where the non-centrality parameter δ is given by

$$\delta = n\underset{\sim}{\eta}'[(\underset{\sim}{B}' \otimes \underset{\sim}{C})\underset{\sim}{\Sigma}(\underset{\sim}{\theta})(\underset{\sim}{B} \otimes \underset{\sim}{C}')]^{-1}\underset{\sim}{\eta} . \tag{5.4.28}$$

This completes our discussion of hypothesis testing in latent linear models.

VI. A NUMERICAL EXAMPLE

6.1 Introduction

In this chapter we use simulated data to illustrate estimation and testing in latent linear models. In §6.2 we describe the population model and the procedure used to simulate the data. In §6.3 we illustrate maximum likelihood estimation of the parameters in the model, and in §6.4 we consider testing hypotheses about the structural parameters.

6.2 Simulation of Data

Let us consider a two sample latent linear model with $p = 5$ manifest variables and $q = 2$ latent variables. Let the elements of the design matrix be given by $a_{ij} = 1$ if observation j is from the i-th population and $a_{ij} = 0$ otherwise, and let the structural parameters be $\underset{\sim}{\Phi} = \underset{\sim}{I}_2$,

$$\underset{\sim}{\Xi} = \begin{bmatrix} .2 & 1.2 \\ .8 & 2.8 \end{bmatrix}, \qquad \underset{\sim}{\Lambda} = \begin{bmatrix} .9 & 0 \\ .7 & .2 \\ .5 & .4 \\ .3 & .6 \\ .1 & .8 \end{bmatrix}, \text{ and } \underset{\sim}{\Psi} = \text{diag} \begin{bmatrix} .3 \\ .4 \\ .5 \\ .6 \\ .7 \end{bmatrix}, \qquad (6.2.1)$$

In order to achieve identification, the parameters λ_{12} and $\underset{\sim}{\Phi}$ are fixed at 0 and $\underset{\sim}{I}_2$, respectively, while the others are free parameters. As indicated in §2.3.1, this implies no loss of generality, for any set of parameter values may be transformed to conform to these restrictions.

The reduced form parameters, computed from (6.2.1) using (2.2.5) and (2.2.6), are

$$\underset{\sim}{\beta}' = \begin{bmatrix} .18 & .30 & .42 & .54 & .66 \\ 1.08 & 1.40 & 1.72 & 2.04 & 2.36 \end{bmatrix},$$

and

$$\underset{\sim}{V} = \begin{bmatrix} 1.11 & & & & \\ .63 & .93 & & & \\ .45 & .43 & .91 & & \\ .27 & .33 & .39 & 1.05 & \\ .09 & .23 & .37 & .51 & 1.35 \end{bmatrix}. \qquad (6.2.2)$$

Note that there are 25 distinct elements in $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and 18 free parameters in $\underset{\sim}{\Xi}$, $\underset{\sim}{\Lambda}$ and $\underset{\sim}{\Psi}$. Hence the number of degrees of freedom associated with the model is 7.

To generate data satisfying this model, recall (2.2.1) and (2.2.3), and proceed as follows: (1) generate two standard normal random variates and form a vector $\underset{\sim}{\xi}_i$; (2) add to $\underset{\sim}{\xi}_i$ the first [second] column of $\underset{\sim}{\Xi}$ as given in (6.2.1), to obtain an observation $\underset{\sim}{y}_i$ from the first [second] population; (3) generate five standard normal variates and multiply each by the square root of one of the diagonal elements of $\underset{\sim}{\Psi}$ as given in (6.2.1), to obtain a vector $\underset{\sim}{z}_i$ of random errors; (4) premultiply $\underset{\sim}{y}_i$ by $\underset{\sim}{\Lambda}$ as given in (6.2.1) and add $\underset{\sim}{z}_i$, to obtain a vector $\underset{\sim}{x}_i$ of manifest variables corresponding to $\underset{\sim}{y}_i$; and (5) repeat the above steps for i=1,...,n.

This method was used to generate a sample of 100 observations from each population. The standard normal random variables required in steps (1) and (3) above were generated using subroutine VARGEN, available at the University of North Carolina Computation Center.

The data, given in terms of the sum of products matrices, are

as follows:

$$\underset{\sim}{AA}' = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

$$\underset{\sim}{AX}' = \begin{bmatrix} 31.2911 & 58.5773 & 63.6883 & 59.1566 & 76.1187 \\ 110.0359 & 137.2624 & 174.0489 & 213.5502 & 233.8271 \end{bmatrix},$$

and

$$\underset{\sim}{XX}' = \begin{bmatrix} 335.0615 & & & & \\ 280.9319 & 390.6948 & & & \\ 295.2279 & 362.0408 & 525.2303 & & \\ 304.7186 & 410.8762 & 506.5956 & 762.9457 & \\ 318.3725 & 430.5187 & 540.4315 & 683.0306 & 881.1164 \end{bmatrix}. \qquad (6.2.3)$$

The least-squares estimates of the reduced form parameters obtained from these data are

$$\underset{\sim}{\bar{\beta}}' = \begin{bmatrix} .3129 & .5858 & .6369 & .5916 & .7612 \\ 1.1004 & 1.3726 & 1.7405 & 2.1355 & 2.3383 \end{bmatrix},$$

and

$$\underset{\sim}{\bar{V}} = \begin{bmatrix} 1.0210 & & & & \\ .5578 & .8399 & & & \\ .4198 & .4292 & .9087 & & \\ .2561 & .4155 & .4862 & 1.3596 & \\ .1863 & .3249 & .4249 & .6933 & 1.3821 \end{bmatrix}, \qquad (6.2.4)$$

which may be compared with (6.2.2).

6.3 Maximum Likelihood Estimation

A computer program has been written to do all necessary calculations in fitting latent linear models. The program is written in Fortran IV-G for the IBM System/360, and has been tested extensively at the University of North Carolina Computation Center.

A choice of four descent methods for minimizing $\tilde{F}$ is provided in the program: (1) steepest descent, (2) Fletcher-Powell with initial matrix $\underset{\sim}{E}$ equal to the identity matrix, (3) Fletcher-Powell with initial matrix $\underset{\sim}{E}$ obtained from the information matrix in accordance with the results of §4.7, and (4) a combination of steepest descent for the initial iterations, followed by Fletcher-Powell using the information matrix for the remaining iterations, the change-over point being when $\tilde{F}$ fails to decrease by more than 5% between two consecutive steepest descent iterations. These methods will be referred to hereafter as SD, FP-I, FP-II and SD/FP-II, respectively. In most applications the SD/FP-II combination will be preferred. When good initial estimates of the parameters are available, however, the FP-II method may be used right from the start.

The program was used to compare the SD, FP-I, and SD/FP-II methods for the data given in (6.2.3). To start the iterative procedure each free parameter was set to 1, thus providing arbitrary initial estimates. Iteration was stopped when all partial derivatives of $\tilde{F}$ were less than .001 in absolute value.

Table 6.3.1 shows the behavior of the function and its derivatives during the iterative procedure. For each method the column $\tilde{F}$ shows the value of the function, and the column G-max shows the largest derivative in absolute value, at selected stages of the process. At the bottom of the table we give the CPU time in seconds required by each method on the IBM System/360-75.

Table 6.4.1 Steepest Descent and Fletcher Powell Function Minimization

	SD		FP-I		SD/FP-II	
Iteration	$\tilde{F}$	G-max	$\tilde{F}$	G-max	$\tilde{F}$	G-max
0	5.717378	.895047	5.717378	.895047	5.717378	.895047
1	5.089637	.446847	5.089637	.446847	5.089637	.446847
2	4.691579	.681971	4.486940	.628014	4.691579	.681971
3	4.440242	.352651	4.390288	.548307	4.440242	.352651
4	4.318538	.590315	4.258196	.481806	4.318538	.590315
5	4.246603	.298247	4.211948	.646492	4.068019	.212787
6	4.210801	.325221	4.131595	.400649	4.055455	.411462
8	4.164438	.206044	4.092626	.272744	4.044735	.027394
10	4.128719	.182090	4.068557	.084633	4.044586	.003847
12	4.099288	.154941	4.057798	.167238	4.044574	.000889
15	4.066047	.291028	4.045418	.052400		
18	4.049712	.037323	4.044612	.007396		
21	4.046359	.033712	4.044574	.000458		
182	4.044577	.000985				
Time (seconds)	32.44		4.89		1.63	

The steepest descent method works well in the initial stages of the procedure, but requires 182 iterations and 32.44 seconds to converge, indicating that the function $\tilde{F}$ is probably relatively flat in a neighborhood of its minimum. The Fletcher-Powell method is clearly superior, requiring only 21 iterations and 4.89 seconds to

converge. Although SD is usually faster than FP-I for the first iterations, it would appear that in this example the initial estimates are sufficiently close to the m.l.e.'s to offset this advantage.

The combined method SD/FP-II is the best one for this example, requiring only 12 iterations and 1.63 seconds to obtain the estimates of all 14 parameters (not counting the 4 elements of $\underset{\sim}{\Xi}$). The change-over from SD to FP-II occurs after the 4th iteration, and the large sample approximations to the second derivatives of $\tilde{F}$ appear to work well, for the process converges very rapidly after that. To obtain some feeling for the accuracy of the approximation we let the FP-I method continue for several additional iterations, thus obtaining a very close approximation to the inverse of the matrix of second derivatives evaluated at the minimum. This was compared with the large sample approximation of §4.7, and found very close.

The maximum likelihood estimates of the parameters are

$$\hat{\underset{\sim}{\Xi}} = \begin{bmatrix} .5078 & 1.3459 \\ .7566 & 2.4067 \end{bmatrix}, \quad \hat{\underset{\sim}{\Lambda}} = \begin{bmatrix} .7956 & 0 \\ .6900 & .1986 \\ .5325 & .4233 \\ .3265 & .7017 \\ .2315 & .8392 \end{bmatrix}, \text{ and } \hat{\underset{\sim}{\Psi}} = \text{diag} \begin{bmatrix} .3875 \\ .3237 \\ .4831 \\ .7088 \\ .6469 \end{bmatrix}, \quad (6.3.1)$$

and are reasonably close to the true parameter values (6.2.1). These estimates were verified using different starting points for the iterative procedure, in particular starting from the true parameter values, and were found to be correct to within .001.

The goodness of fit statistic is 9.51, and is approximately distributed as a chi-square variate with 7 degrees of freedom. This gives an approximate p-value of .2181, which would lead to accepting

the hypothesis that the model fits.

The maximum likelihood estimates of the reduced form parameters, obtained from (6.3.1) using (2.2.5) and (2.2.6), are

$$\hat{\underset{\sim}{\beta}}' = \begin{bmatrix} .4040 & .5006 & .5907 & .6967 & .7525 \\ 1.0708 & 1.4066 & 1.7354 & 2.1282 & 2.3313 \end{bmatrix},$$

and

$$\hat{\underset{\sim}{V}} = \begin{bmatrix} 1.0205 & & & & \\ .5490 & .8392 & & & \\ .4237 & .4515 & .9458 & & \\ .2598 & .3646 & .4709 & 1.3078 & \\ .1842 & .3664 & .4785 & .6645 & 1.4047 \end{bmatrix}. \quad (6.3.2)$$

These estimates take into account the latent factor structure of the model, while those of (6.2.2) do not. The residuals, or differences between these two sets of estimates, may provide valuable insight in fitting latent linear models. For this example the residuals are all small, as would be expected. The average and largest absolute differences between $\bar{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\beta}}$ are .0419 and .1051, respectively, and those between $\bar{\underset{\sim}{V}}$ and $\hat{\underset{\sim}{V}}$ are .0205 and .0518, respectively.

The estimated large sample standard errors of the estimates of the structural parameters, obtained from the information matrix evaluated at the m.l.e.'s (6.3.1), are

$$\hat{s.e.}(\hat{\underset{\sim}{\Xi}}) = \begin{bmatrix} .1285 & .1992 \\ .1493 & .2657 \end{bmatrix}, \quad \hat{s.e.}(\hat{\underset{\sim}{\Lambda}}) = \begin{bmatrix} .0923 & 0 \\ .0814 & .0717 \\ .0778 & .0597 \\ .0990 & .0708 \\ .1038 & .0808 \end{bmatrix}, \quad \hat{s.e.}(\hat{\underset{\sim}{\Psi}}) = \text{diag} \begin{bmatrix} .1205 \\ .0708 \\ .0590 \\ .0973 \\ .1198 \end{bmatrix}, \quad (6.3.3)$$

where $\hat{s.e.}(\hat{\underset{\sim}{\Xi}})$ is a matrix whose (i,j)-th element is the estimated standard error of the (i,j)-th element of $\hat{\underset{\sim}{\Xi}}$, and $\hat{s.e.}(\hat{\underset{\sim}{\Lambda}})$ and $\hat{s.e.}(\hat{\underset{\sim}{\Psi}})$ are similarly defined.

On the other hand, the true large sample standard errors, obtained from the information matrix evaluated at the parameter values (6.2.1), are

$$s.e.(\hat{\underset{\sim}{\Xi}}) = \begin{bmatrix} .1155 & .1700 \\ .1440 & .2849 \end{bmatrix}, \quad s.e.(\hat{\underset{\sim}{\Lambda}}) = \begin{bmatrix} .0937 & 0 \\ .0801 & .0539 \\ .0742 & .0495 \\ .0825 & .0583 \\ .0938 & .0759 \end{bmatrix}, \quad s.e.(\hat{\underset{\sim}{\Psi}}) = \begin{bmatrix} .1332 \\ .0733 \\ .0601 \\ .0791 \\ .1216 \end{bmatrix}. \tag{6.3.4}$$

Note that the estimated standard errors are very close to the true values.

Although in actual applications of the model the latent variates are not observable, the method of data simulation used in this example is such that the values of the latent variates are known. Using those values we obtain the following estimate of $\underset{\sim}{\Xi}$, from ordinary linear model analysis,

$$\underset{\sim}{\Xi} = \begin{bmatrix} .3344 & 1.1531 \\ .9395 & 2.9170 \end{bmatrix}, \text{ with } s.e.(\underset{\sim}{\Xi}) = \begin{bmatrix} .0990 & .0990 \\ .1002 & .1002 \end{bmatrix}. \tag{6.3.5}$$

These estimates are somewhat closer to the true parameter values (6.2.1) than the estimates $\hat{\underset{\sim}{\Xi}}$ of (6.3.1) based on the latent linear model analysis, and their estimated standard errors are smaller than the estimated standard errors of $\hat{\underset{\sim}{\Xi}}$ given in (6.3.3). These results illustrate the fact that some information is lost when the latent variates can not be observed, but they also indicate that parameters

pertaining to the latent variates can still be estimated reliably on the basis of observable indicators.

6.4 Hypothesis Testing

Let us now consider testing hypotheses about the structural parameters. We consider first testing

$$H_0: \underset{\sim}{\Psi} = \psi \underset{\sim}{I}_5 \qquad \text{v/s} \qquad H_1: \underset{\sim}{\Psi} \neq \psi \underset{\sim}{I}_5 \ , \tag{6.4.1}$$

the hypothesis that all specificities are equal.

To test (6.4.1) we fit a model where the first element of $\underset{\sim}{\Psi}$, say ψ, is a free parameter, and all other elements are constrained to be equal to ψ. The total number of free parameters in the model is now 14 instead of 18, and the number of degrees of freedom associated with H_0 is 4.

To compute the estimates we used the FP-II method, with initial estimates equal to the true parameter values for $\underset{\sim}{\Lambda}$ and to the value .5 for ψ. The procedure converged in 5 iterations to a function value of 4.157385 and a largest absolute derivative of .000506. The restricted m.l.e.'s of the structural parameters are

$$\hat{\hat{\underset{\sim}{\Xi}}} = \begin{bmatrix} .5489 & 1.4695 \\ .6740 & 2.2232 \end{bmatrix}, \quad \hat{\hat{\underset{\sim}{\Lambda}}} = \begin{bmatrix} .7384 & 0 \\ .6395 & .2079 \\ .5742 & .3967 \\ .3305 & .7455 \\ .2280 & .8981 \end{bmatrix}, \text{ and } \hat{\hat{\underset{\sim}{\Psi}}} = .5038 \underset{\sim}{I}_5 \ . \tag{6.4.2}$$

The goodness of fit statistic for the restricted model is 32.07, and would be approximately distributed as a chi-square variate with 11 degrees of freedom if the model fitted. The approximate p-value

is .0007, indicating that the more restricted model does not fit the data.

To test (6.4.1) we form the difference of the goodness of fit statistics, in accordance with the procedure of §5.3, obtaining a test statistic of 22.56. Under H_0 the test statistic is approximately distributed as a chi-square variate with 4 degrees of freedom. The approximate p-value is thus .0002 and H_0 would be rejected at the .0002 significance level. Since we know H_0 to be false, this result gives us some assurance with respect to the power of the test procedure.

Let us now consider testing a simple linear hypothesis about $\underset{\sim}{\Xi}$ using the Wald statistic of §5.4. Let $\underset{\sim}{\xi}_j$ denote the j-th column of $\underset{\sim}{\Xi}$ (j=1,2), and consider testing

$$H_0: \underset{\sim}{\xi}_2 - \underset{\sim}{\xi}_1 = 0 \qquad v/s \qquad H_1: \underset{\sim}{\xi}_2 - \underset{\sim}{\xi}_1 \neq 0 \ , \tag{6.4.3}$$

the hypothesis of no location difference between the populations. Note that H_0 imposes two constraints upon $\underset{\sim}{\Xi}$.

This hypothesis may be put in the framework of §5.4 by letting $\underset{\sim}{B} = \underset{\sim}{I}_2$ and $\underset{\sim}{C}' = [-1,1]$. Note that

$$\underset{\sim}{C}' \otimes \underset{\sim}{B} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \ . \tag{6.4.4}$$

The unrestricted m.l.e. of $\underset{\sim}{\Xi}$ given in (6.3.1) may be written as the vector

$$\hat{\underset{\sim}{\xi}} = [.5078,\ .7566,\ 1.3459,\ 2.4067]' \ . \tag{6.4.5}$$

The estimated large sample variance covariance matrix of $\hat{\underset{\sim}{\xi}}$, obtained from the information matrix evaluated at the m.l.e.'s (6.3.1) is

$$\hat{\underset{\sim}{\Sigma}} = \begin{bmatrix} .0165 & & & \\ -.0041 & .0223 & & \\ .0090 & -.0054 & .0397 & \\ -.0047 & .0186 & -.0163 & .0706 \end{bmatrix} . \qquad (6.4.6)$$

Premultiplying (6.4.5) by (6.4.4) we find the estimated differences in location

$$(\underset{\sim}{C}' \otimes \underset{\sim}{B})\hat{\underset{\sim}{\xi}}_n = [.9381, 1.6501]' , \qquad (6.4.7)$$

The estimated large sample variance-covariance matrix of (6.4.7), obtained on premultiplying (6.4.6) by (6.4.4) and post-multiplying by the transpose of (6.4.4), is

$$(\underset{\sim}{C}' \otimes \underset{\sim}{B})\hat{\underset{\sim}{\Sigma}}(\underset{\sim}{C} \otimes \underset{\sim}{B}') = \begin{bmatrix} .0383 & -.0103 \\ -.0103 & .0558 \end{bmatrix}, \qquad (6.4.8)$$

and we see immediately that the differences shown in (6.4.7) are too large relative to their variances for H_0 to be true.

Inverting (6.4.8), premultiplying by the transpose of (6.4.7) and post-multiplying by (6.4.7), we obtain a value of 84.67 for the Wald statistic. Since the approximate distribution of the statistic under H_0 is a central chi-square with 2 degrees of freedom, we reject H_0.

The reason that the statistic is so large is that the Mahalanobis distance between the populations is very large. The true differences, obtained from (6.2.1), are

$$(\underset{\sim}{C}' \otimes \underset{\sim}{B})\underset{\sim}{\xi} = [1.0 , 2.0]' , \qquad (6.4.9)$$

and the true large sample variance-covariance matrix of (6.4.7), obtained from the information matrix evaluated at the true parameter

values (6.2.1), is

$$(\underset{\sim}{C}' \otimes \underset{\sim}{B})\underset{\sim}{\Sigma}(\underset{\sim}{C} \otimes \underset{\sim}{B}') = \begin{bmatrix} .0352 & -.0080 \\ -.0080 & .0682 \end{bmatrix} . \qquad (6.4.10)$$

Inverting (6.4.10), pre-multiplying by the transpose of (6.4.9) and post-multiplying by (6.4.9), we find the non-centrality parameter to be $\delta = 103.16$. According to the results of §5.4, the true large sample distribution of the Wald statistic is a non-central chi-square with 2 degrees of freedom and non-centrality 103.16. The probability that a $\chi_2^2(103.16)$ variate exceeds 84.67 is .8430; thus the results obtained are not at all surprising, given the actual difference between the populations.

Since the method used to simulate the data gives the values of the latent variates, a direct test of H_0 is possible. From ordinary linear models analysis we obtain a value of 151.14 for the likelihood ratio statistic, and 228.44 for the Hotelling trace statistic, which is equivalent to the Wald statistic for this problem. Under H_0 these statistics are each distributed as chi-square with 2 degrees of freedom in large samples, and lead to rejection of H_0. The fact that they are larger than the statistic arising from the latent linear model analysis illustrates the fact that some power is lost when the latent variates cannot be observed. The fact that the latter statistic is nonetheless highly significant, however, gives some reassurance with respect to the power of tests of hypotheses pertaining to latent variables based on observable indicators.

VII. A SIMULATION STUDY

7.1 Introduction

We now report the results of a simulation study conducted to evaluate the large sample approximations of Chapters 4 and 5. In §7.2 we describe the population model and the characteristics of the study, and in §7.3 we consider the distributions of the goodness of fit statistic, of the estimators of the structural parameters, and of the Wald statistic.

7.2 Simulation of Data

Let us consider a two-sample latent linear model with 3 measurements and 1 factor. Let the design matrix have elements $a_{ij} = 1$ if observation j is from the i-th population and $a_{ij} = 0$ otherwise, and let the structural parameters be $\phi = 1$,

$$\underset{\sim}{\xi} = [1,2], \quad \underset{\sim}{\lambda} = \begin{bmatrix} .3 \\ .5 \\ .7 \end{bmatrix}, \quad \text{and} \quad \underset{\sim}{\Psi} = \text{diag} \begin{bmatrix} .91 \\ .75 \\ .51 \end{bmatrix}. \qquad (7.2.1)$$

In order to achieve identification, the parameter ϕ is fixed at 1, while all others are free parameters.

The reduced form parameters are

$$\underset{\sim}{\beta} = \begin{bmatrix} .3 & .6 \\ .5 & 1.0 \\ .7 & 1.4 \end{bmatrix}, \quad \text{and} \quad \underset{\sim}{V} = \begin{bmatrix} 1 & & \\ .15 & 1 & \\ .21 & .35 & 1 \end{bmatrix}. \qquad (7.2.2)$$

The structural parameters have been chosen so that $\text{Var}(x_i) = 1$ for all i. Thus the correlation between the i-th and j-th measurements

of the latent variable y is simply $\mathrm{Cov}(x_i,x_j) = \lambda_i\lambda_j$ as given in (7.2.2), and the correlation between the i-th measurement and y is simply $\mathrm{Cov}(x_i,y) = \lambda_i$, (see in this regard §2.4.3). Thus we use 3 measurements which have correlations of .3, .5 and .7 with the underlying factor.

Note that there are 12 elements in $\underset{\sim}{\beta}$ and $\underset{\sim}{V}$, and 8 free parameters in $\underset{\sim}{\xi}$, $\underset{\sim}{\lambda}$ and $\underset{\sim}{\Psi}$. Hence the number of degrees of freedom associated with the model is 4.

Let the elements of $\underset{\sim}{\xi}$, $\underset{\sim}{\lambda}$ and $\underset{\sim}{\Psi}$ be ordered in a vector $\underset{\sim}{\theta}$ as follows:

$$\underset{\sim}{\theta} = [\lambda_1,\lambda_2,\lambda_3,\psi_1,\psi_2,\psi_3,\xi_1,\xi_2]' \ . \qquad (7.2.3)$$

The information matrix evaluated at the true parameter values using (4.6.12) - (4.6.19) is

$$\underset{\sim}{I}(\underset{\sim}{\theta}) = \begin{bmatrix} 3.1007 & & & & & & & \\ -.2323 & 3.6127 & & & & & & \\ -.4782 & -1.0202 & 3.9411 & & & & & \\ .1310 & -.0114 & -.0285 & .5008 & & & & \\ -.0244 & .3203 & -.1070 & .0038 & .6580 & & & \\ -.1032 & -.2203 & .6736 & .0161 & .0734 & .6866 & & \\ .0654 & .1396 & .2874 & 0 & 0 & 0 & .2906 & \\ .1309 & .2792 & .5748 & 0 & 0 & 0 & 0 & .2906 \end{bmatrix} \qquad (7.2.4)$$

and its inverse is

$$\underset{\sim}{I}^{-1}(\underset{\sim}{\theta}) = \begin{bmatrix} .4459 & & & & & & & \\ .2187 & .6228 & & & & & & \\ .3407 & .6567 & 1.1235 & & & & & \\ -.0893 & -.0053 & .0045 & 2.0225 & & & & \\ -.0124 & -.1689 & .0166 & -.0043 & 1.6316 & & & \\ -.1936 & -.3051 & -.8710 & -.0666 & -.2467 & 2.2119 & & \\ -.5424 & -.9088 & -1.4600 & .0182 & .0675 & 1.0515 & 5.4436 & \\ -1.0848 & -1.8177 & -2.9200 & .0364 & .1350 & 2.1030 & 4.0051 & 11.4513 \end{bmatrix} \tag{7.2.5}$$

We take samples of size 50 from each population, so that n = 100. The data were generated using the same procedure as in §6.2 The sum of products were accumulated as the observations were generated, and punched out for analysis. A total of 1000 sets of data were generated in this manner.

To estimate the parameters efficiently a special computer program was written to evaluate the function $\tilde{F}$ and its derivatives using all simplifications that obtain in a one-factor model. For example from (3.4.1),

$$(\underset{\sim}{V}^{-1})_{ij} = \begin{cases} 1/\psi_i - \lambda_i^2/[\psi_i^2(1+\delta)], & i=j \\ -\lambda_i\lambda_j/[\psi_i\psi_j(1+\delta)], & i\neq j \end{cases} \tag{7.2.6}$$

where

$$\delta = \sum_{i=1}^{3} \lambda_i^2/\psi_i \quad , \tag{7.2.7}$$

and from (3.5.3),

$$|\underset{\sim}{V}| = \psi_1\psi_2\psi_3(1+\delta) \quad . \tag{7.2.8}$$

After experimenting with different numerical procedures to compute the estimates for ten sets of data, it was decided to use the FP-II method with initial estimates given by the true parameter values, and with initial $\underset{\sim}{E}$ matrix obtained by dividing by 2 the elements of the

inverse of the information matrix evaluated at the true parameter values. The method was found to require between 3 and 5 iterations to converge for the test samples, the convergence criterion being that all partial derivatives of F be less than .001 in absolute value. Note that the number of parameters being estimated iteratively is 6.

The efficiency of the method is further illustrated by the fact that computation of the estimates for 1000 sets of data took only 23.9 seconds of CPU time in the IBM System/360-75. The iterative procedure converged in all but 3 cases (out of 1000) for which the estimate of the smallest specificity, ψ_3, tended to become negative. As mentioned in §1.3.2 this is quite common in factor analysis, particularly in small samples when a specificity is not very large relative to the standard error of its estimate. For the sake of simplicity, these sets of data were replaced.

7.3 Empirical Distributions

7.3.1 The Goodness of Fit Statistic

According to the results of §5.2, we expect the goodness of fit statistic to be approximately distributed as a central chi-square variate with 4 degrees of freedom.

Table 7.3.1 gives the number of statistics falling below selected percentile values of the χ^2_4 distribution. Note that the fit is remarkably good, particularly in the upper tail, which is used in setting up critical regions. A test of the goodness of fit of the model at the 95% significance level based on the asymptotic theory would accept the null hypothesis in 95.1% of the samples.

Table 7.3.1 Number of Goodness of Fit Statistics Falling Below Selected Percentiles of χ^2_4

Percentile	Number of Statistics
20	185
40	375
60	589
80	790
90	903
95	951
99	989

The Kolmogorov-Smirnov goodness of fit statistic is 0.0256, and the associated p-value is .5305. The chi-square goodness of fit test using 5 categories with expected frequencies of 20% gives $\chi^2_4 = 3.11$, with associated p-value of .5396.

These results indicate that the χ^2_4 distribution provides a very good approximation to the distribution of the goodness of fit statistic for moderate size samples.

7.3.2 Estimates of Structural Parameters

According to the results of Chapter 4, we expect the maximum likelihood estimators of the structural parameters to be approximately normally distributed with means equal to the true parameter values (7.2.1) and variances and covariances given by the inverse of the information matrix (7.2.5) divided by the sample size, which is n=100.

Table 7.3.2 shows the mean, variance, skewness, kurtosis, and the Kolmogorov-Smirnov goodness of fit statistic together with its p-value for each estimate.

Table 7.3.2 Descriptive Statistics for Empirical Distribution of the Maximum Likelihood Estimators

Estimator	Mean	Variance	Skewness	Kurtosis	D	p-value
$\hat{\lambda}_1$	.2945	.0045	.212	.176	.0502	.0129
$\hat{\lambda}_2$	.4897	.0068	-.023	-.068	.0599	.0015
$\hat{\lambda}_3$	.6870	.0125	-.023	-.133	.0549	.0048
$\hat{\psi}_1$	.8963	.0180	.305	.174	.0704	.0001
$\hat{\psi}_2$	.7376	.0165	.455	.710	.0656	.0004
$\hat{\psi}_3$	.4979	.0252	.142	.081	.0625	.0008
$\hat{\xi}_1$	1.0572	.0666	.888	2.261	.0661	.0003
$\hat{\xi}_2$	2.0901	.1660	1.134	2.376	.0772	.0000

Note that the empirical means are very close to the true parameter values, and that the empirical variances are very close to the asymptotic variances given by the inverse of the information matrix. The empirical variance-covariance matrix of the estimates is

$$\hat{\mathrm{Var}}(\hat{\underset{\sim}{\theta}}) = \begin{bmatrix} .0045 & & & & & & & \\ .0024 & .0068 & & & & & & \\ .0039 & .0064 & .0125 & & & & & \\ -.0005 & .0003 & -.0115 & .0180 & & & & \\ .0001 & -.0014 & .0008 & .0000 & .0165 & & & \\ -.0024 & -.0031 & -.0100 & -.0001 & -.0042 & .0252 & & \\ -.0063 & -.0109 & -.0175 & .0002 & .0003 & .0130 & .0666 & \\ -.0145 & -.0240 & -.0368 & .0006 & .0022 & .0263 & .0593 & .1660 \end{bmatrix} \tag{7.3.1}$$

Comparison of (7.3.1) with the inverse of the information matrix (7.2.5) indicates that all empirical variances and covariances are very close to the asymptotic values. This result gives us assurance as to

the correctness of the formulae for the elements of the information matrix given in §4.6, and indicates a fast rate of convergence of the first two moments.

Convergence of higher order moments, however, appears to be slower, judging from the values of the skewness and kurtosis coefficients. The Kolmogorov-Smirnov statistics given in the sixth column of Table 7.3.2, on the other hand, range between .0502 and .0772, and seven of them indicate a lack of fit significant at the .001 level or better. Thus for a sample of size 100 the distributions of the estimate appears not to be normal.

The actual values of the statistics are not very large, though, and the question of most practical import is to what extent the asymptotic distribution can serve as a reasonable approximation for a sample of this size. To investigate this matter the information matrix evaluated at the m.l.e.'s was computed and inverted for each sample, and an approximate 95% confidence interval of the form

$$\hat{\theta}_i \pm 1.96\sqrt{\underset{\sim}{I}^{-1}_{ii}(\hat{\underset{\sim}{\theta}})/100}\ , \qquad (7.3.2)$$

where $\underset{\sim}{I}^{-1}_{ii}(\underset{\sim}{\theta})$ denotes the i-th diagonal element of $\underset{\sim}{I}^{-1}(\underset{\sim}{\theta})$, was constructed for each of the parameters. Table 7.3.3 gives the number of such intervals that actually contained the parameter.

Table 7.3.3 Number of 95% Confidence Intervals Found to Contain True Parameter Value

Parameter	Number
λ_1	937
λ_2	937
λ_3	942
ψ_1	927
ψ_2	931
ψ_3	940
ξ_1	963
ξ_2	964

Note that the empirical confidence coefficients range from 92.7% to 96.4%, and thus are close to the asymptotic value. Therefore, confidence intervals computed from moderate size samples on the basis of the asymptotic theory would be acceptable for most practical purposes.

7.3.3 The Wald Test for Linear Hypotheses

Let us now consider testing the hypothesis of no location difference between the populations using the Wald statistic of §5.4. The true difference between the populations, obtained from (7.2.1) is

$$\xi_2 - \xi_1 = 1 \ , \tag{7.3.3}$$

and the large sample variance of the estimated difference obtained from (7.2.5), is

$$\mathrm{Var}(\hat{\xi}_2 - \hat{\xi}_1) = .0888. \tag{7.3.4}$$

From the results of §5.4 we expect the Wald statistic to be

approximately distributed as a non-central chi-square variate with 1 degree of freedom and non-centrality parameter $\delta = 8.88$.

For each of the 1000 samples, the Wald statistic was computed using the inverse of the information matrix evaluated at the m.l.e.'s. Table 7.3.4 shows the number of statistics falling below selected percentile values of the χ^2_1 (8.88) distribution.

Table 7.3.4 Number of Wald Statistics Falling Below Selected Percentiles of χ^2_1 (8.88)

Percentile	Number
5	45
10	101
20	238
40	426
60	567
80	715

The chi-square goodness of fit statistic using 5 categories with expected frequencies of 20% is 74.99, and indicates a very poor overall fit. Inspection of the data shows that the empirical distribution of the Wald statistic has a very long upper tail. However, it is the lower tail that would be used in computing approximate powers, and the fit is considerably better there. We now assess the practical significance of this result.

An approximate test of H_0 at the 95% significance level based on the asymptotic central distribution of W rejects if $W \geq \chi^2_{1,.95} = 3.8415$. The approximate power of the test computed from the asymptotic

non-central distribution of W is therefore given by

$$\Pr\{\chi_1^2(8.88) \geq 3.8415\} = .8461 \ . \qquad (7.3.5)$$

On the other hand, the empirical power, given by the number of statistics exceeding 3.8415 in the simulation study, was found to be .825.

In conclusion, the results of this section indicate that the asymptotic distribution theory provides quite reasonable approximations for moderate sample sizes, particulary for the goodness of fit statistic and for the standard errors of the estimates. More extensive simulation studies would be needed, however, to compare results for alternative sets of parameter values and to establish empirically the rates of convergence of the distributions.

VIII. SUGGESTIONS FOR FURTHER RESEARCH

8.1 Introduction

In this chapter we provide some suggestions for future research by proposing two extensions of the latent linear model: the latent growth curve model, stated in §8.2; and the latent covariance structure model, stated in §8.3. These models are similar to the latent linear model in that observable indicators are used to estimate parameters and test hypotheses pertaining to unobservable constructs, and belong to the general family of structural linear models considered in Chapters 4 and 5.

8.2 The Latent Growth Curve Model

Consider the growth curve model

$$\underset{\sim}{Y} = \underset{\sim}{P}\underset{\sim}{\Xi}\underset{\sim}{A} + \underset{\sim}{\varepsilon}, \tag{8.2.1}$$

where $\underset{\sim}{Y}$: $q \times n$ is a matrix of n observations on q variates, $\underset{\sim}{P}$: $q \times q'$ is a within-subjects design matrix of full column rank $q' \leq q$, $\underset{\sim}{A}$: $r \times n$ is an across-subjects design matrix of full row rank $r < n$, $\underset{\sim}{\Xi}$: $q' \times r$ is a matrix of unknown regression parameters, and $\underset{\sim}{\varepsilon}$: $q \times n$ is a stochastic matrix of errors with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ and $\mathrm{Var}(\underset{\sim}{\varepsilon}) = \underset{\sim}{\Phi} \otimes \underset{\sim}{I}_n$.

Suppose now that $\underset{\sim}{Y}$ is not observable. Instead we observe a data matrix $\underset{\sim}{X}$: $p \times n$ related to $\underset{\sim}{Y}$ by the factor analysis model

$$\underset{\sim}{X} = \underset{\sim}{\Lambda}\underset{\sim}{Y} + \underset{\sim}{Z} \tag{8.2.2}$$

where $\underset{\sim}{\Lambda}$: p×q is a matrix of factor loadings and $\underset{\sim}{Z}$: p×n is a matrix of errors with $E(\underset{\sim}{Z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{Z}) = \underset{\sim}{\Psi} \otimes \underset{\sim}{I}_n$, where $\underset{\sim}{\Psi}$ is a diagonal matrix of specificities, and $Cov(\underset{\sim}{Y},\underset{\sim}{Z}) = \underset{\sim}{0}$.

Combining (8.2.1) and (8.2.2) we can write the model as

$$\underset{\sim}{X} = \underset{\sim}{\beta}\underset{\sim}{A} + \underset{\sim}{U}\,, \tag{8.2.3}$$

where $E(\underset{\sim}{U}) = \underset{\sim}{0}$, $Var(\underset{\sim}{U}) = \underset{\sim}{V} \otimes \underset{\sim}{I}_n$,

$$\underset{\sim}{\beta} = \underset{\sim}{\Lambda}\underset{\sim}{P}\underset{\sim}{\Xi}\,, \quad \text{and} \tag{8.2.4}$$

$$\underset{\sim}{V} = \underset{\sim}{\Lambda}\underset{\sim}{\Phi}\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}\,. \tag{8.2.5}$$

The model is thus seen to be a multivariate linear model where the regression and dispersion parameters are structure and related to each other. Since its basis is a growth curve model on unobservable variates, it will be termed the *latent growth curve model*.

The model differs from the latent linear model only in the addition of the new design matrix $\underset{\sim}{P}$, and hence the modifications required in the results of Chapter 3 to obtain the maximum likelihood estimators of the structural parameters are relatively simple. Also, since the model belongs to the family of structural linear models considered in Chapters 4 and 5, the general asymptotic results given there can be applied.

8.3 The Latent Covariance Structure Model

Consider now the covariance structure model

$$\underset{\sim}{Y} = \underset{\sim}{P}\underset{\sim}{\Xi}\underset{\sim}{A} + \underset{\sim}{\varepsilon}, \tag{8.3.1}$$

where $\underset{\sim}{Y}$, $\underset{\sim}{P}$, $\underset{\sim}{\Xi}$ and $\underset{\sim}{A}$ are as defined in §8.2. $\underset{\sim}{\varepsilon}$ is a matrix of errors with $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ and $\text{Var}(\underset{\sim}{\varepsilon}) = \underset{\sim}{\Phi} \otimes \underset{\sim}{I}_n$, where

$$\underset{\sim}{\Phi} = \underset{\sim}{\Gamma\Sigma\Gamma}' + \underset{\sim}{T} , \tag{8.3.2}$$

where $\underset{\sim}{\Gamma}$:q×s is a matrix of loadings, $\underset{\sim}{\Sigma}$:s×s is a symmetric p.d. matrix of factor variances and covariances, and $\underset{\sim}{T}$:q×q is a diagonal matrix of specificities. This part of the model is similar to, but less general than, Jöreskog's (1970a) covariance structure model, considered in §1.5.

Suppose, as before, that $\underset{\sim}{Y}$ is not observable but instead we observe

$$\underset{\sim}{X} = \underset{\sim}{\Lambda Y} + \underset{\sim}{Z} , \tag{8.3.3}$$

where $\underset{\sim}{\Lambda}$: p×q is a matrix of loadings and $\underset{\sim}{Z}$ is a stochastic matrix of errors with $E(\underset{\sim}{Z}) = \underset{\sim}{0}$, $\text{Var}(\underset{\sim}{Z}) = \underset{\sim}{\Psi} \otimes \underset{\sim}{I}_n$, where $\underset{\sim}{\Psi}$: p×p is a diagonal matrix of specificities, and $\text{Cov}(\underset{\sim}{Y},\underset{\sim}{Z}) = 0$.

Combining (8.3.1) and (8.3.3) we obtain

$$\underset{\sim}{X} = \underset{\sim}{\beta A} + \underset{\sim}{U} , \tag{8.3.4}$$

where $E(\underset{\sim}{U}) = \underset{\sim}{0}$, $\text{Var}(\underset{\sim}{U}) = \underset{\sim}{V} \times \underset{\sim}{I}_n$,

$$\underset{\sim}{\beta} = \underset{\sim}{\Lambda P \Xi A} \quad , \quad \text{and} \tag{8.3.5}$$

$$\underset{\sim}{V} = \underset{\sim}{\Lambda(\Gamma\Sigma\Gamma}' + \underset{\sim}{T})\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi} . \tag{8.3.6}$$

The resulting model is thus a structural linear model of the general type considered in Chapters 4 and 5. The structure of $\underset{\sim}{V}$ is identical with the covariance structure (1.5.3) in Jöreskog's model, but the structure of $\underset{\sim}{\beta}$ involves the parameter $\underset{\sim}{\Lambda}$, which does not appear in (1.5.2). Since the basis of the model is a

covariance structure model on unobservable variates it will be termed the *latent covariance structure model.*

The relationship between this model and Jöreskog's (1970a) model for the analysis of covariance structures can best be discussed in terms of an example given in Jöreskog (1970a). Consider the Wiener stochastic process

$$y_t = \xi_t + u_1 + \ldots + u_t \tag{8.3.7}$$

$(t=1,\ldots,q)$ where $\xi_t = E(y_t)$, and $u_1,\ldots,u_t$ are independent increments. This may be written in matrix form as

$$\underset{\sim}{\chi} = \underset{\sim}{\xi} + \underset{\sim}{M}\underset{\sim}{u} \ , \tag{8.3.8}$$

where $\underset{\sim}{M}$ is a lower triangular matrix whose nonzero elements are all unity. The variance-covariance matrix of $\underset{\sim}{\chi}$ is

$$\underset{\sim}{\Phi} = \underset{\sim}{M}\underset{\sim}{\Sigma}\underset{\sim}{M}' \ , \tag{8.3.9}$$

where $\underset{\sim}{\Sigma}$ is a diagonal matrix whose elements are the variances of the independent increments.

Jöreskog (1970a) assumes that instead of $\underset{\sim}{\chi}$ we observe a vector $\underset{\sim}{x}$ of q measurements, given by

$$\underset{\sim}{x} = \underset{\sim}{\chi} + \underset{\sim}{z} \ , \tag{8.3.10}$$

where $\underset{\sim}{z}$ is a q-vector of measurement errors with $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{z}) = \underset{\sim}{\Psi}$, a diagonal matrix, and $Cov(\underset{\sim}{\chi},\underset{\sim}{z}) = \underset{\sim}{0}$. Thus

$$E(\underset{\sim}{x}) = \underset{\sim}{\xi} \ , \quad \text{and} \tag{8.3.11}$$

$$V(\underset{\sim}{x}) = \underset{\sim}{M}\underset{\sim}{\Sigma}\underset{\sim}{M}' + \underset{\sim}{\Psi} \ , \tag{8.3.12}$$

which is a special case of (1.5.2) - (1.5.3) with $\underset{\sim}{P} = \underset{\sim}{I}$, $\underset{\sim}{A} = \underset{\sim}{1}_n'$,

$\underset{\sim}{\Gamma} = \underset{\sim}{M}$, $\underset{\sim}{\Sigma}$ restricted to be diagonal, $\underset{\sim}{T} = \underset{\sim}{0}$ and $\underset{\sim}{\Lambda} = \underset{\sim}{I}$.

Suppose now that instead of $\underset{\sim}{y}$ we observe

$$\underset{\sim}{x} = \underset{\sim}{\Lambda y} + \underset{\sim}{z} \ , \tag{8.3.13}$$

where $\underset{\sim}{x}$ is a vector of $p>q$ measurements, $\underset{\sim}{\Lambda}$: $p \times q$ is a matrix of factor loadings, and $\underset{\sim}{z}$ is a p-vector of measurement errors with $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $Var(\underset{\sim}{z}) = \underset{\sim}{\Psi}$, and $Cov(\underset{\sim}{y},\underset{\sim}{z}) = \underset{\sim}{0}$. Then

$$E(\underset{\sim}{x}) = \underset{\sim}{\Lambda\xi} \quad , \text{ and} \tag{8.3.14}$$

$$V(\underset{\sim}{x}) = \underset{\sim}{\Lambda M \Sigma M' \Lambda'} + \underset{\sim}{\Psi}, \tag{8.3.15}$$

which is a special case of (8.3.4) - (8.3.6) with $\underset{\sim}{P} = \underset{\sim}{I}$, $\underset{\sim}{A} = \underset{\sim}{1}_n'$, $\underset{\sim}{\Gamma} = \underset{\sim}{M}$, $\underset{\sim}{\Sigma}$ restricted to be diagona, and $\underset{\sim}{T} = \underset{\sim}{0}$. This model will be termed the *latent Wiener Process*.

While Jöreskog (1970a) considers the case of a single fallible measurement x_t of y_t, with regression coefficient 1 and error variance ψ_t, the more general model given here considers a set of indicators $\underset{\sim}{x}_t$ of $\underset{\sim}{y}_t$, with regression coefficients $\underset{\sim}{\Lambda}_t$ and error variance $\underset{\sim}{\Psi}_t$.

APPENDIX

ON MATRIX DERIVATIVES

A.1 Introduction

Matrix derivatives were introduced in the statistical literature by Dwyer and MacPhail (1948). A number of useful results for matrix differentiation have been given by Deemer and Olkin (1951), Olkin (1953), Dwyer (1967), Neudecker (1969), Tracy and Dwyer (1969), Vetter (1970) and more recently MacRae (1974). Some elementary results appear in the text by Graybill (1970, pp. 260-70).

The derivation of these results, for the most part, has not been based on a unified matrix differential calculus. As MacRae (1974) has noted, either a typical element of a matrix array is examined in hopes of inferring a matrix expression for the entire array, or matrices of total differentials are obtained and transformed into arrays of derivatives using special theorems. Compare Anderson (1958) and Neudecker (1969).

Furthermore, most authors consider derivatives of a scalar with respect to a matrix and of a matrix with respect to a scalar, but do not consider explicitly the derivative of a matrix with respect to a matrix. The latter is usually described in terms of the derivatives of a matrix with respect to each of the elements of another matrix. See for example Dwyer (1967).

Recently, MacRae (1974) has proposed a unified approach to matrix differentiation and has introduced new matrix operations and identities. Her approach simplifies the notation and facilitates application by providing a formal matrix differential calculus, i.e., a set of general rules for dealing with matrix differentiation. A similar approach may be found in the paper by Vetter (1970).

In this appendix we review some basic definitions and theorems and collect a number of results that are used throughout the dissertation.

A.2 Definition of Matrix Derivatives

Definition A.2.1 We consider first derivatives of scalar functions of matrices. Let f be a scalar function of a matrix $\underset{\sim}{X}$: m×n. Then we define the derivative of f with respect to $\underset{\sim}{X}$ as the m×n matrix

$$\frac{\partial f}{\partial \underset{\sim}{X}} = \left(\frac{\partial f}{\partial x_{ij}} \right) . \tag{A.2.1}$$

Example A.2.1. The following two results are well known. If $\underset{\sim}{X}$ is an m×m matrix, then

$$\frac{\partial \mathrm{tr} \underset{\sim}{X}}{\partial \underset{\sim}{X}} = \underset{\sim}{I}_m \quad , \text{ and} \tag{A.2.2}$$

$$\frac{\partial \log |\underset{\sim}{X}|}{\partial \underset{\sim}{X}} = \underset{\sim}{X}^{-1} \quad , \ \underset{\sim}{X} \text{ non-singular.} \tag{A.2.3}$$

Result (A.2.2) follows directly from the definition. Result (A.2.3) may be obtained by writing the determinant in terms of co-factors, see for example Graybill (1970, pp. 266-267).

We now consider the derivative of a matrix function of a matrix

Definition A.2.2. Let $\underset{\sim}{Y}$: p×q be a matrix whose elements are functions of a matrix $\underset{\sim}{X}$: m×n. Let $\frac{\partial}{\partial \underset{\sim}{X}} = \left(\frac{\partial}{x_{ij}}\right)$ be a matrix of partial derivative operators. Then we define the matrix derivative of $\underset{\sim}{Y}$ with respect to $\underset{\sim}{X}$ as

$$\frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} = \underset{\sim}{Y} \otimes \frac{\partial}{\partial \underset{\sim}{X}} \quad , \tag{A.2.4}$$

where multiplication by a partial derivative operator corresponds to the operation of partial differentiation, i.e., $y_{k\ell} \frac{\partial}{\partial \underset{\sim}{X}} = \frac{\partial y_{k\ell}}{\partial \underset{\sim}{X}}$ and thus

$$\frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial \underset{\sim}{X}}, \ldots, \frac{\partial y_{1q}}{\partial \underset{\sim}{X}} \\ \ldots \\ \frac{\partial y_{p1}}{\partial \underset{\sim}{X}}, \ldots, \frac{\partial y_{pq}}{\partial \underset{\sim}{X}} \end{bmatrix} : pm \times qn \tag{A.2.5}$$

This definition is due to MacRae (1974). Neudecker (1969) and Vetter (1970) arrange matrix derivatives in a pattern that corresponds to $\partial/\partial \underset{\sim}{X} \otimes \underset{\sim}{Y}$, but do not use the concept of a partial derivative operator.

Definition A.2.3. Let $\underset{\sim}{E}_{ij}$ be an m×n matrix with 1 in the (i,j)-th position and 0 elsewhere, and define

$$\underset{\sim}{E}_{(m,n)} = \begin{bmatrix} \underset{\sim}{E}_{11}, \ldots, \underset{\sim}{E}_{1n} \\ \ldots \\ \underset{\sim}{E}_{m1}, \ldots, \underset{\sim}{E}_{mn} \end{bmatrix} : m^2 \times n^2 \tag{A.2.6}$$

$$\underset{\sim}{I}_{(m,n)} = \begin{bmatrix} \underset{\sim}{E}'_{11}, \ldots, \underset{\sim}{E}'_{1n} \\ \ldots \\ \underset{\sim}{E}'_{m1}, \ldots, \underset{\sim}{E}'_{mn} \end{bmatrix} : mn \times mn \ . \tag{A.2.7}$$

The matrix $\underset{\sim}{I}_{(m,n)}$ is called the *permuted identity matrix*, see Tracy and Dwyer (1969) and MacRae (1974). The notation $\underset{\sim}{E}_{(m,n)}$ has been introduced by deWaal (1974).

Theorem A.2.1. Let $\underset{\sim}{X}$ be an m×n matrix. Then

$$\frac{\partial \underset{\sim}{X}}{\partial \underset{\sim}{X}} = \underset{\sim}{E}_{(m,n)} , \tag{A.2.8}$$

$$\frac{\partial \underset{\sim}{X}'}{\partial \underset{\sim}{X}} = \underset{\sim}{I}_{(m,n)} . \tag{A.2.9}$$

This theorem is given by MacRae (1974) and Vetter (1970) using a different notation. The proof follows from Definitions A.2.2 and A.2.3.

A.3 Rules for Matrix Differentiation

We now review some rules for matrix differentiation and illustrate their application.

A.3.1 The Sum, Product and Inverse Rules.

Theorem A.3.1. (The Sum Rule). Let $\underset{\sim}{Y}$: p×q and $\underset{\sim}{Z}$: p×q be matrices whose elements are functions of a matrix $\underset{\sim}{X}$: m×n. Then

$$\frac{\partial (\underset{\sim}{Y}+\underset{\sim}{Z})}{\partial \underset{\sim}{X}} = \frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} + \frac{\partial \underset{\sim}{Z}}{\partial \underset{\sim}{X}} . \tag{A.3.1}$$

The proof follows directly from Definition A.2.2.

Theorem A.3.2. (The Product Rule). Let $\underset{\sim}{Y}$ and $\underset{\sim}{Z}$ be conformable matrices whose elements are functions of $\underset{\sim}{X}$: m×n. Then

$$\frac{\partial \underset{\sim}{Y}\underset{\sim}{Z}}{\partial \underset{\sim}{X}} = \frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} (\underset{\sim}{Z} \otimes \underset{\sim}{I}_n) + (\underset{\sim}{Y} \otimes \underset{\sim}{I}_m) \frac{\partial \underset{\sim}{Z}}{\partial \underset{\sim}{X}} . \tag{A.3.2}$$

For a proof see MacRae (1974). An analogous result is given by Vetter (1970) using a different notation.

Example A.3.1. (Linear Forms). Let $\underset{\sim}{A}$ be p×m, $\underset{\sim}{X}$ be m×n and $\underset{\sim}{B}$ be n×q where $\underset{\sim}{A}$ and $\underset{\sim}{B}$ do not depend on $\underset{\sim}{X}$. Then

$$\frac{\partial \underset{\sim}{AXB}}{\partial \underset{\sim}{X}} = (\underset{\sim}{A} \otimes \underset{\sim}{I}_m)\underset{\sim}{E}_{(m,n)}(\underset{\sim}{B} \otimes \underset{\sim}{I}_n) \ . \tag{A.3.3}$$

Proof: Using the product rule, since $\partial \underset{\sim}{B}/\partial \underset{\sim}{X} = \underset{\sim}{0}$ we have

$$\frac{\partial \underset{\sim}{AXB}}{\partial \underset{\sim}{X}} = \frac{\partial \underset{\sim}{AX}}{\partial \underset{\sim}{X}} (\underset{\sim}{B} \otimes \underset{\sim}{I}_n) \ ,$$

and using the product rule again, since $\partial \underset{\sim}{A}/\partial \underset{\sim}{X} = 0$,

$$\frac{\partial \underset{\sim}{AXB}}{\partial \underset{\sim}{X}} = (\underset{\sim}{A} \otimes \underset{\sim}{I}_m) \frac{\partial \underset{\sim}{X}}{\partial \underset{\sim}{X}} (\underset{\sim}{B} \otimes \underset{\sim}{I}_n) \ ;$$

then (A.3.3) follows using (A.2.8). □

Collecting elements in different blocks of (A.3.3) we obtain

$$\frac{\partial \underset{\sim}{AXB}}{\partial \underset{\sim}{X}_{ij}} = \underset{\sim}{A}\underset{\sim}{E}_{ij}\underset{\sim}{B} \ . \tag{A.3.4}$$

This result is well known, see for example Tracy and Dwyer (1969).

Example A.3.2. (Quadratic Forms). Let $\underset{\sim}{X}$ be m×n, $\underset{\sim}{B}$ be n×n , and $\underset{\sim}{C}$ be m×p where $\underset{\sim}{B}$ and $\underset{\sim}{C}$ do not depend on X. Then

$$\frac{\partial \underset{\sim}{XBX'C}}{\partial \underset{\sim}{X}} = \underset{\sim}{E}_{(m,n)}(\underset{\sim}{BX'C} \otimes \underset{\sim}{I}_n) + (\underset{\sim}{XB} \otimes \underset{\sim}{I}_m)\underset{\sim}{I}_{(m,n)}(\underset{\sim}{C} \otimes \underset{\sim}{I}_n) \ . \tag{A.3.5}$$

Proof: Using the product rule

$$\frac{\partial \underset{\sim}{XBX'C}}{\partial \underset{\sim}{X}} = \frac{\partial \underset{\sim}{XB}}{\partial \underset{\sim}{X}} (\underset{\sim}{X'C} \otimes \underset{\sim}{I}_n) + (\underset{\sim}{XB} \otimes \underset{\sim}{I}_m)\frac{\partial \underset{\sim}{X'C}}{\partial \underset{\sim}{X}} \ ,$$

and using the product rule again, since $\partial \underset{\sim}{B}/\partial \underset{\sim}{X} = \underset{\sim}{0}$ and $\partial \underset{\sim}{C}/\partial \underset{\sim}{X} = \underset{\sim}{0}$,

$$\frac{\partial \underset{\sim\sim}{X}\underset{\sim\sim}{B}\underset{\sim}{X}'\underset{\sim}{C}}{\partial \underset{\sim}{X}} = \frac{\partial \underset{\sim}{X}}{\partial \underset{\sim}{X}} (\underset{\sim}{B} \otimes \underset{\sim}{I}_n)(\underset{\sim}{X}'\underset{\sim}{C} \otimes \underset{\sim}{I}_n) + (\underset{\sim\sim}{X}\underset{\sim}{B} \otimes \underset{\sim}{I}_m)\frac{\partial \underset{\sim}{X}'}{\partial \underset{\sim}{X}}(\underset{\sim}{C} \otimes \underset{\sim}{I}_n) .$$

Then (A.3.5) follows from the well-known identity for Kronecker products $(\underset{\sim}{A} \otimes \underset{\sim}{B})(\underset{\sim}{C} \otimes \underset{\sim}{D}) = \underset{\sim\sim}{AC} \otimes \underset{\sim\sim}{BD}$ and Theorem A.2.1. □

An important special case of (A.3.5) is obtained setting $\underset{\sim}{C} = \underset{\sim}{I}_m$ and noting that $\underset{\sim}{I}_m \otimes \underset{\sim}{I}_n = \underset{\sim}{I}_{mn}$. Then

$$\frac{\partial \underset{\sim\sim}{X}\underset{\sim\sim}{B}\underset{\sim}{X}'}{\partial \underset{\sim}{X}} = \underset{\sim}{E}_{(m,n)}(\underset{\sim\sim}{B}\underset{\sim}{X}' \otimes \underset{\sim}{I}_n) + (\underset{\sim\sim}{X}\underset{\sim}{B} \otimes \underset{\sim}{I}_m)\underset{\sim}{I}_{(m,n)} . \qquad (A.3.6)$$

Collecting terms in different blocks of (A.3.6) we obtain

$$\frac{\partial \underset{\sim\sim}{X}\underset{\sim\sim}{B}\underset{\sim}{X}'}{\partial x_{ij}} = \underset{\sim}{E}_{ij}\underset{\sim\sim}{B}\underset{\sim}{X}' + \underset{\sim\sim}{X}\underset{\sim\sim}{B}\underset{\sim}{E}_{ji} , \qquad (A.3.7)$$

a well-known result, see Tracy and Dwyer (1969).

Theorem A.3.3. (The Inverse Rule). Let $\underset{\sim}{Y}$ be a non-singular matrix function of $\underset{\sim}{X}$: m×n. Then

$$\frac{\partial \underset{\sim}{Y}^{-1}}{\partial \underset{\sim}{X}} = - (\underset{\sim}{Y}^{-1} \otimes \underset{\sim}{I}_m)\frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}}(\underset{\sim}{Y}^{-1} \otimes \underset{\sim}{I}_n) . \qquad (A.3.8)$$

The proof follows from the product rule writing $\underset{\sim}{Y}^{-1} = \underset{\sim}{Y}^{-1}\underset{\sim\sim}{Y}\underset{\sim}{Y}^{-1}$.

An important special case of (A.3.8) is obtained when $\underset{\sim}{Y} = \underset{\sim}{X}$ for $\underset{\sim}{X}$ non-singular. Then

$$\frac{\partial \underset{\sim}{X}^{-1}}{\partial \underset{\sim}{X}} = - (\underset{\sim}{X}^{-1} \otimes \underset{\sim}{I}_m)\underset{\sim}{E}_{(m,n)}(\underset{\sim}{X}^{-1} \otimes \underset{\sim}{I}_m) . \qquad (A.3.9)$$

Collecting terms in different blocks of (A.3.9) we obtain

$$\frac{\partial \underset{\sim}{X}^{-1}}{\partial x_{ij}} = - \underset{\sim}{X}^{-1}\underset{\sim}{E}_{ij}\underset{\sim}{X}^{-1} , \qquad (A.3.10)$$

a result given in Tracy and Dwyer (1969).

A.3.2 The Star Product and the Chain Rule

MacRae (1974) has introduced the following operation:

Definition A.3.1. (Star Product). Let $\underset{\sim}{A}$ be p×q and let $\underset{\sim}{B}$: pm×qn be partitioned into pq blocks $\underset{\sim}{B}_{ij}$: m×n. Then the star product of $\underset{\sim}{A}$ and $\underset{\sim}{B}$ is defined as the m×n matrix

$$\underset{\sim}{A} * \underset{\sim}{B} = \sum_{ij} a_{ij}\underset{\sim}{B}_{ij} . \tag{A.3.11}$$

If $\underset{\sim}{A}$ and $\underset{\sim}{B}$ have the same dimensions then

$$\underset{\sim}{A} * \underset{\sim}{B} = \text{tr } \underset{\sim}{A}'\underset{\sim}{B} . \tag{A.3.12}$$

Theorem A.3.4. Let $\underset{\sim}{A}$ be m×p, $\underset{\sim}{B}$ be p×q and $\underset{\sim}{C}$ be q×n. Then

$$\underset{\sim\sim}{ABC} = \underset{\sim}{B}' * (\underset{\sim}{C} \otimes \underset{\sim}{I}_m)\underset{\sim}{I}_{(m,n)}(\underset{\sim}{A} \otimes \underset{\sim}{I}_n) \tag{A.3.13}$$

$$(\underset{\sim\sim}{ABC})' = \underset{\sim}{B}' * (\underset{\sim}{C} \otimes \underset{\sim}{I}_n)\underset{\sim}{E}_{(n,m)}(\underset{\sim}{A} \otimes \underset{\sim}{I}_m) . \tag{A.3.14}$$

Identity (A.3.13) is given by MacRae (1974) and (A.3.14) in the present notation is given by deWaal (1974). The proofs can be established by direct computation. These results are very useful in applications. Other identities that we have not used in our work may be found in MacRae (1974) and deWaal (1974).

The importance of the star product is given by the following result due to MacRae (1974).

Theorem A.3.5. (The Chain Rule). Let f be a scalar function of a matrix $\underset{\sim}{Y}$: p×q whose elements are functions of a matrix $\underset{\sim}{X}$: m×n. Then

$$\frac{\partial f}{\partial \underset{\sim}{X}} = \frac{\partial f}{\partial \underset{\sim}{Y}} * \frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} \tag{A.3.15}$$

The proof follows directly from the ordinary chain rule and the definition of star product.

If $\underset{\sim}{X}$ is a scalar x then, by the simplification of star products of matrices of the same dimension noted in (A.3.12),

$$\frac{\partial f}{\partial x} = \mathrm{tr}\left[\frac{\partial f}{\partial \underset{\sim}{Y}'} \cdot \frac{\partial \underset{\sim}{Y}}{\partial x}\right] . \tag{A.3.16}$$

We also note that by repeated use the theorem applies to the case where f is a scalar function of $\underset{\sim}{Z}$ which is a matrix function of $\underset{\sim}{Y}$ which in turn depends on $\underset{\sim}{X}$. In this case

$$\frac{\partial f}{\partial \underset{\sim}{X}} = \left[\frac{\partial f}{\partial \underset{\sim}{Z}} * \frac{\partial \underset{\sim}{Z}}{\partial \underset{\sim}{Y}}\right] * \frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}} , \tag{A.3.17}$$

where the star products must be evaluated as indicated, for $\frac{\partial \underset{\sim}{Z}}{\partial \underset{\sim}{X}} \neq \frac{\partial \underset{\sim}{Z}}{\partial \underset{\sim}{Y}} * \frac{\partial \underset{\sim}{Y}}{\partial \underset{\sim}{X}}$, indeed the star product on the right hand side of this expression is not defined unless $\underset{\sim}{Z}$ is a scalar.

In the examples below let f be a scalar function of a matrix $\underset{\sim}{Y}$ with matrix derivative $\partial f/\partial \underset{\sim}{Y}$.

Example A.3.3. Let $\underset{\sim}{A}$ be $p\times m$, $\underset{\sim}{X}$ be $m\times n$ and $\underset{\sim}{B}$ be $n\times q$. Then

$$\frac{\partial f(\underset{\sim\sim}{AXB})}{\partial \underset{\sim}{X}} = A' \frac{\partial f}{\partial (\underset{\sim\sim}{AXB})} \underset{\sim}{B}' . \tag{A.3.18}$$

The proof follows from the chain rule, Example A.3.1 and identity (A.3.14).

Example A.3.4. Let $\underset{\sim}{X}$ be $m\times n$, $\underset{\sim}{B}$ be $n\times n$ and $\underset{\sim}{C}$ be $m\times p$. Then

$$\frac{\partial f(\underset{\sim\sim}{XBX'C})}{\partial \underset{\sim}{X}} = \frac{\partial f}{\partial (\underset{\sim\sim}{XBX'C})} \underset{\sim}{C}' \underset{\sim\sim}{XB}' + \underset{\sim}{C} \frac{\partial f}{\partial (\underset{\sim\sim}{XBX'C})} \underset{\sim\sim}{XB} . \tag{A.3.19}$$

The proof follows from the chain rule, Example A.3.2 and identities (A.3.13) and (A.3.14).

Example A.3.5. Let $\underset{\sim}{X}$ be non-singular. Then

$$\frac{\partial f(\underset{\sim}{X}^{-1})}{\partial \underset{\sim}{X}} = -\underset{\sim}{X}^{-T} \frac{\partial f}{\partial \underset{\sim}{X}^{-1}} \underset{\sim}{X}^{-T} \tag{A.3.20}$$

where $\underset{\sim}{X}^{-T} = (\underset{\sim}{X}^{-1})'$. The proof follows from the chain and inverse rules and identity (A.3.14).

A number of important special cases are obtained when $f(\underset{\sim}{Y}) = \text{tr } \underset{\sim}{Y}$ for square $\underset{\sim}{Y}$. In particular note that from Example A.3.3

$$\frac{\partial \text{ tr } \underset{\sim}{A}\underset{\sim}{X}\underset{\sim}{B}}{\partial \underset{\sim}{X}} = \underset{\sim}{A}'\underset{\sim}{B}' , \tag{A.3.21}$$

and from Example A.3.4

$$\frac{\partial \text{ tr } \underset{\sim}{X}\underset{\sim}{B}\underset{\sim}{X}'\underset{\sim}{C}}{\partial \underset{\sim}{X}} = \underset{\sim}{C}'\underset{\sim}{X}\underset{\sim}{B}' + \underset{\sim}{C}\underset{\sim}{X}\underset{\sim}{B} . \tag{A.3.22}$$

Vetter (1970) has given a chain rule for matrix functions of matrices. His result, however, involves packing the columns of the matrices one below another in vector form, and it appears unlikely to be very useful in applications until further identities are established to simplify the results.

In the case of vector functions of vectors, however, the following result is easily established.

Theorem A.3.6. Let $\underset{\sim}{z}$: t×1 be a vector function of a vector $\underset{\sim}{y}$: s×1 whose elements depend on a vector $\underset{\sim}{x}$: r×1. Then

$$\frac{\partial \underset{\sim}{z}}{\partial \underset{\sim}{x}'} = \frac{\partial \underset{\sim}{z}}{\partial \underset{\sim}{y}'} \frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{x}'} , \tag{A.3.23}$$

where $\partial \underset{\sim}{z}/\partial \underset{\sim}{x}' = (\partial z_i/\partial x_j)$: t×r , by Definition A.2.2.

A.3.3 A Note on Symmetric and Diagonal Matrices

So far in our discussion we have assumed that the elements of $\underset{\sim}{X}$ are functionally independent. An important case of functional dependence that arises in our work is that of $\underset{\sim}{X}$ *symmetric*. In this case the best strategy is to apply the formal matrix differentiation rules ignorming the symmetry and to bring this in the last step by appropriately modifying the final result. For our purposes it will suffice to discuss the case of derivatives of scalar functions of symmetric matrices.

According to Definition A.2.1, the (i,j)-th element of $\partial f/\partial \underset{\sim}{X}$ is the partial derivative of f with respect to x_{ij} treating all other elements of $\underset{\sim}{X}$ as constants. If $\underset{\sim}{X}$ is symmetric, however, it may be desirable to compute partial derivatives of f with respect to the *distinct* elements x_{ij} $(1 \leq i < j \leq m)$ of $\underset{\sim}{X}$ treating all other elements except x_{ji} as constants. These derivatives may be arranged in convenient matrix form by defining

$$\frac{\partial f}{\partial \underset{\sim}{X}_S} = 2 \frac{\partial f}{\partial \underset{\sim}{X}} - \text{diag} \frac{\partial f}{\partial \underset{\sim}{X}} . \qquad \text{(A.3.24)}$$

In numerical optimization problems involving $f(\underset{\sim}{X})$ the required partial derivatives may be obtained from (A.3.24). Note, however, that (A.3.24) is not a *matrix* derivative in the sense of Definition A.2.1, just a convenient notation. This distinction is important in some applications, see for example Travinsky and Bargmann (1964).

Another important case where the elements of $\underset{\sim}{X}$ are restricted is when $\underset{\sim}{X}$ is *diagonal*. In the case of a scalar function f of a diagonal matrix $\underset{\sim}{X}$ one may apply the formal matrix differentiation

rules ignoring the fact that $\underset{\sim}{X}$ is diagonal, and take account of this in the last step by defining

$$\frac{\partial f}{\partial \underset{\sim}{X}_d} = \text{diag} \frac{\partial f}{\partial \underset{\sim}{X}} . \tag{A.3.25}$$

Alternatively, in many applications involving diagonal matrices it is possible to proceed directly from first principles.

A.4 Maximum Likelihood Estimation in the Linear Model

We illustrate the application of matrix derivatives in maximum likelihood estimation in the multivariate linear model

$$\underset{\sim}{X} = \underset{\sim\sim}{\beta A} + \underset{\sim}{U} , \tag{A.4.1}$$

where $\underset{\sim}{X}$ is $p \times n$, $\underset{\sim}{\beta}$ is $p \times r$, $\underset{\sim}{A}$ is $r \times n$ of full row rank $r < n$ and $\underset{\sim}{U} \sim N_{p \times n}(\underset{\sim}{0}, \underset{\sim}{V} \otimes \underset{\sim}{I}_n)$ with $\underset{\sim}{V}$ positive definite.

The log-likelihood function may be written as

$$\log L = -\tfrac{1}{2}pn \log(2\pi) - \tfrac{1}{2}n \log|\underset{\sim}{V}| - \tfrac{1}{2} \operatorname{tr} \underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim\sim}{\beta A})(\underset{\sim}{X}-\underset{\sim\sim}{\beta A})' . \tag{A.4.2}$$

Let us use the notation

$$\underset{\sim}{T} = \frac{1}{n}(\underset{\sim}{X}-\underset{\sim\sim}{\beta A})(\underset{\sim}{X}-\underset{\sim\sim}{\beta A})' . \tag{A.4.3}$$

Maximizing $\log L$ is equivalent to minimizing

$$F = \log|\underset{\sim}{V}| + \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T} . \tag{A.4.4}$$

To differentiate this function with respect to $\underset{\sim}{\beta}$ use the chain rule (A.3.17), obtaining

$$\frac{\partial F}{\partial \underset{\sim}{\beta}} = \left(\frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{T}} * \frac{\partial \underset{\sim}{T}}{\partial (\underset{\sim}{X}-\underset{\sim\sim}{\beta A})} \right) * \frac{\partial (\underset{\sim}{X}-\underset{\sim\sim}{\beta A})}{\partial \underset{\sim}{\beta}} . \tag{A.4.5}$$

Using (A.3.21)

$$\frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{T}} = \underset{\sim}{V}^{-1} , \tag{A.4.6}$$

and using (A.3.6)

$$\frac{\partial \underset{\sim}{T}}{\partial(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})} = \frac{1}{n}\{\underset{\sim}{E}_{(p,n)}[(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})' \otimes \underset{\sim}{I}_p] + [(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A}) \otimes \underset{\sim}{I}_n]\underset{\sim}{I}_{(p,n)}\}. \tag{A.4.7}$$

Substituting these results into (A.4.5) gives

$$\frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{T}} * \frac{\underset{\sim}{T}}{(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})} = \frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A}) , \tag{A.4.8}$$

by identities (A.3.13) and (A.3.14). On the other hand

$$\frac{\partial(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})}{\partial \underset{\sim}{\beta}} = - \underset{\sim}{E}_{(p,r)}(\underset{\sim}{A} \otimes \underset{\sim}{I}_n) \tag{A.4.9}$$

by the sum rule and (A.3.3). Using (A.4.8) and (A.4.9) on (A.4.5)

$$\frac{\partial F}{\partial \underset{\sim}{\beta}} = - \frac{2}{n}\underset{\sim}{V}^{-1}(\underset{\sim}{X}-\underset{\sim}{\beta}\underset{\sim}{A})\underset{\sim}{A}' , \tag{A.4.10}$$

by identity (A.3.14).

On setting (A.4.10) to zero we obtain the well known result

$$\hat{\underset{\sim}{\beta}} = \underset{\sim}{X}\underset{\sim}{A}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1} . \tag{A.4.11}$$

To differentiate with respect to $\underset{\sim}{V}$, use the sum and product rules, obtaining

$$\frac{\partial F}{\partial \underset{\sim}{V}} = \frac{\partial \log|\underset{\sim}{V}|}{\partial \underset{\sim}{V}} + \frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{V}^{-1}} * \frac{\partial \underset{\sim}{V}^{-1}}{\partial \underset{\sim}{V}} . \tag{A.4.12}$$

Using (A.1.3)

$$\frac{\partial \log|\underset{\sim}{V}|}{\partial \underset{\sim}{V}} = \underset{\sim}{V}^{-1} . \tag{A.4.13}$$

On the other hand

$$\frac{\partial \operatorname{tr} \underset{\sim}{V}^{-1}\underset{\sim}{T}}{\partial \underset{\sim}{V}^{-1}} * \frac{\partial \underset{\sim}{V}^{-1}}{\partial \underset{\sim}{V}} = - \underset{\sim}{T} * (\underset{\sim}{V}^{-1} \otimes \underset{\sim}{I}_p)\underset{\sim}{E}_{(p,p)}(\underset{\sim}{V}^{-1} \otimes \underset{\sim}{I}_p) , \qquad (A.4.14)$$

$$= -\underset{\sim}{V}^{-1}\underset{\sim}{T}\underset{\sim}{V}^{-1} \qquad (A.4.15)$$

where (A.4.14) follows from (A.3.21) and (a.3.9); and (A.4.15) follows from identity (A.3.14). Combining (A.4.13) and (A.4.15) we obtain

$$\frac{\partial F}{\partial \underset{\sim}{V}} = \underset{\sim}{V}^{-1}[\underset{\sim}{V}-\underset{\sim}{T}]\underset{\sim}{V}^{-1} , \qquad (A.4.16)$$

So far we have ignored the symmetry of $\underset{\sim}{V}$. Setting to 0 each of the elements of $\partial f/\partial \underset{\sim}{V}$, however, is the same as setting to 0 each of the elements of $\partial f/\partial \underset{\sim}{V}_s = 2\partial f/\partial \underset{\sim}{V}$ - diag $\partial f/\partial \underset{\sim}{V}$ and gives $\underset{\sim}{V} = \underset{\sim}{T}$.

To obtain $\hat{\underset{\sim}{V}}$ note that $\underset{\sim}{T}$ is a function of $\underset{\sim}{\beta}$ and evaluate it at $\underset{\sim}{\beta} = \hat{\underset{\sim}{\beta}}$. Using the fact that $[\underset{\sim}{I} - \underset{\sim}{A}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1}\underset{\sim}{A}]$ is symmetric and idempotent gives the well known result

$$\hat{\underset{\sim}{V}} = \frac{1}{n}\underset{\sim}{X}[\underset{\sim}{I}-\underset{\sim}{A}'(\underset{\sim}{A}\underset{\sim}{A}')^{-1}\underset{\sim}{A}]\underset{\sim}{X}' . \qquad (A.4.17)$$

An alternative derivation of these results, without using matrix derivatives, may be found in Anderson (1958, pp. 44-49 and 179-181).

BIBLIOGRAPHY

Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons.

Anderson, T.W. and Rubin, H. (1956). Statistical inference in factor analysis. In Neyman, J. (Editor). *Proceedings Third Berkeley Symposium on Mathematical Statistics and Probability, 5*: 111-150. Berkeley: University of California Press.

Archer, C.O. and Jennrich, R.I. (1973). Standard errors for rotated factor loadings. *Research Bulletin*, 73-77. Princeton, New Jersey: Educational Testing Service.

Bargmann, R.E. (1957). A study of independence and dependence in multivariate normal analysis. Ph.D. Thesis, University of North Carolina. (Also *Institute of Statistics Mimeo Series No. 186.*)

Bartlett, M.S. (1937). The statistical conception of mental factors. *British Journal of Psychology, 28*: 97-104.

Bartlett, M.S. (1938). Methods of estimating mental factors. *Nature-London, 141*: 609-610.

Bartlett, M.S. (1950). Tests of significance in factor analysis. *British Journal of Psychology - Statistics Section, 3*: 77-85.

Bartlett, M.S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology - Statistics Section, 4*: 1-2.

Bartlett, M.S. (1953). Factor analysis as a statistician sees it. *Uppsala Symposium on Psychological Factor Analysis.* Stockholm: Almquist and Wiksell.

Blalock, H.M. (1961). *Causal Inferences in Nonexperimental Research.* Chapel Hill: University of North Carolina Press.

Boudon, R. (1965). A method of linear causal analysis: dependence analysis. *American Sociological Review, 35*: 101-111.

Box, G.E.P. (1949). A general distribution theory for a class of likelihood ratio criteria. *Biometrika, 36:* 317-346.

Bradley, R.A. and Gart, J.A. (1969). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika, 49*: 205-214.

Browne, M.W. (1965). A comparison of factor analytic techniques. Master's Thesis, University of Witwatersrand, South Africa.

Browne, M.W. (1974a). Gradient methods for analytic rotation. *British Journal of Mathematical and Statistical Psychology, 27*: 115-121.

Browne, M.W. (1974b). Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*: 1-24.

Clarke, M.R.B. (1970). A rapidly convergent method for maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*: 43-52.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Crawford, C.B. and Ferguson, G.A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika, 35*: 321-332.

Davidon, W.C. (1959). A variable metric method for minimization. AEC Research and Development report, ANL-5990.

Deemer, W.L. and Olkin, I. (1951). Jacobians of matrix transformations useful in multivariate analysis. *Biometrika, 38*: 345-367.

de Waal, D.J. (1974). Parametric multivariate analysis - Part 2. Supplement B: matrix derivatives. Mimeographed notes, Department of Statistics, University of North Carolina.

Duncan, O.D. (1966). Path analysis: sociological examples. *American Journal of Sociology, 72*: 3-16.

Dwyer, P.S. (1967). Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association, 62*: 607-625.

Dwyer, P.S. and MacPhail, M.S. (1948). Symbolic matrix derivatives. *Annals of Mathematical Statistics, 19*: 517-534.

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Annals of Mathematical Statistics, 34*: 447-456.

Fletcher, R. and Powell, M.J.D. (1963). A rapidly convergent descent method for minimization. *Computer Journal, 6*: 163-168.

Golub, G.A. (1969). Matrix decompositions and statistical calculations. In Milton, R.C. and Nelder, A. (Editors). *Statistical Computation*. New York: Academic Press.

Graybill, F.A. (1970). *Introduction to Matrices with Applications in Statistics*. Belmont, California: Wadsworth.

Grizzle, D.J. and Allen, D.M. (1969). Analysis of growth and Bose response curves. *Biometrics, 25:* 307-318.

Harman, H.H. (1967). *Modern Factor Analysis*. Chicago: University of Chicago Press.

Hauser, R.M. and Goldberger, A.S. (1971). The treatment of unobservable variables in path analysis. In Costner, H.L. (Editor). *Sociological Methodology: 1971*. San Francisco: Jossey Bass.

Hendrickson, A.E. and White, P.O. (1964). Promax: a quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17*: 65-70.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24:* 417-441, 498-520.

Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology, 26*: 139-142.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika, 28*: 321-377.

Howe, W.G. (1955). Some contributions to factor analysis. Ph.D. Thesis, University of North Carolina. (Also Report No. ORNL-1919, Oak Ridge, Tennessee: Oak Ridge National Laboratory.)

Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics, 40*: 633-643.

Jennrich, R.I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association, 65*: 904-912.

Jennrich, R.I. (1973). Standard errors for obliquely rotated factor loadings. *Research Bulletin*, 73-28. Princeton, New Jersey: Educational Testing Service.

Jennrich, R.I. (1974). Simplified formulae for standard errors in maximum-likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 27:* 122-131.

Jennrich, R.I. and Sampson, P.F. (1966). Rotation for simple loadings. *Psychometrika, 31*: 313-323.

Jennrich, R.I. and Thayer, D.I. (1973). A note on Lawley's formulae for standard errors in maximum likelihood factor analysis. *Research Bulletin* 73-31. Princeton, New Jersey: Educational Testing Service.

Jöreskog, K.G. (1963). *Statistical Estimation in Factor Analysis.* Stockholm: Almquist and Wiksell.

Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 32*: 443-482.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*: 183-202.

Jöreskog, K.G. (1970a). A general method for analysis of covariance structures. *Biometrika, 57*: 239-251.

Jöreskog, K.G. (1970b). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*: 121-145.

Jöreskog, K.G. (1971a). Simultaneous factor analysis in several populations. *Psychometrika, 36*: 409-426.

Jöreskog, K.G. (1971b). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*: 109-133.

Jöreskog, K.G. and Goldberger, A.S. (1972). Factor analysis by generalized least squares. *Psychometrika, 37*: 243-260.

Jöreskog, K.G. and Lawley, D.N. (1968). New methods in maximum likelihood factor analysis. *British Journal of Mathematical and Statistical Psychology, 21*: 85-96.

Kaiser, H.F. (1958). The variance criterion for analytic rotation in factor analysis. *Psychometrika, 23*: 187-200.

Kaiser, H.F. (1959). Computer program for varimax rotation in factor analysis. *Educational and Psychological Measurement, 19*:413-420.

Kendall, M.G. and Babington-Smith, B. (1950). Factor analysis. *Journal of the Royal Statistical Society - Series B, 12*: 60-94.

Kendall, M.G. and Stuart, A. (1961). *The Advanced Theory of Statistics* - Vol. II: Inference and Relationship. London: Charles Griffin and Co.

Kendall, M.G. and Stuart, A. (1969). *The Advanced Theory of Statistics* - Vol. I: Distribution Theory. (3rd Edition). London: Charles Griffin and Co.

Land, K.C. (1969). Principles of path analysis. In Borgatta, E.F. (Editor). *Sociological Methodology: 1969.* San Francisco: Jossey-Bass.

Landis, J.R. and Koch, G.G. (1974). A review of statistical methods in the analysis of data arising from observer reliability studies. University of North Carolina: *Institute of Statistics Mimeo Series No. 956.*

Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh - Series A, 60*: 64-82.

Lawley, D.N. (1942). Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh - Series A, 62*: 176-185.

Lawley, D.N. (1943). The application of the maximum likelihood method to factor analysis. *British Journal of Psychology, 33*: 172-175.

Lawley, D.N. (1953). A modified method of estimation in factor analysis and some large sample results. *Uppsala Symposium on Psychological Factor Analysis*. Stockholm: Almquist and Wiksell.

Lawley, D.N. (1967). Some new results in maximum likelihood factor analysis. *Proceedings of the Royal Society of Edinburgh - Series A, 67*: 256-264.

Lawley, D.N. and Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method*. (2nd Edition). New York: American Elsevier.

Lazarfeld, P.F. (1950). The logic and mathematical foundation of latent structure analysis. In Stouffer *et. al.* (Editors). *Measurement and Prediction*. Princeton, New Jersey: Princeton University Press.

Lockhart, R.S. (1967). Asymptotic sampling variances for factor analytic models identified by specified zero parameters. *Psychometrika, 32*: 265-277.

Lord, F.M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Massachusetts: Addison-Wesley.

MacRae, E.C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Annals of Statistics, 2*: 337-346.

Maxwell, D.N. (1961). Recent rends in factor analysis. *Journal of the Royal Statistical Society - Series A, 124*: 49-59.

Morrison, D.F. (1967). *Multivariate Statistical Methods*. New York: McGraw Hill.

Mukherjee, B.N. (1973). Analysis of covariance structures and exploratory factor analysis. *British Journal of Mathematical and Statistical Psychology, 26*: 125-154.

Mulaik, S.A. (1971). A note on some equations of confirmatory factor analysis. *Psychometrika, 36*: 63-70.

Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association, 64*: 953-963.

Okamoto, M. (1969). Optimality of principal components. In Krishnaiah, P.R. (Editor). *Multivariate Analysis II:* Proceedings of the Second International Symposium on Multivariate Analysis. New York: Academic Press.

Olkin, I. (1953). Note on "Jacobians of matrix transformations useful in multivariate analysis." *Biometrika, 40:* 43-46.

Please, N.W. (1973). Comparison of factor loadings in different populations. *British Journal of Mathematical and Statistical Psychology, 26:* 61-89.

Potthoff, R.R. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful specially for growth curves. *Biometrika, 51:* 313-326.

Puri, M.L. and Sen, P.K. (1969). A class of rank order tests for a general linear hypothesis. *Annals of Mathematical Statistics, 40:* 1325-1343.

Rao, C.R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika, 20:* 93-111.

Rao, C.R. (1965). *Linear Statistical Inference and its Applications.* New York: John Wiley and Sons.

Rash, G. (1953). On simultaneous factor analysis in several populations. *Uppsala Symposium on Psychological Factor Analysis.* Stockholm: Almquist and Wicksell.

Rodriguez, G. (1975). A canonical reduction of the factor analysis model. University of North Carolina: *Institute of Statistics Mimeo Series No. 992.*

Sen, P.K. and Puri, M.L. (1970). Asymptotic theory of likelihood ratio and rank order tests in some multivariate linear models. *Annals of Mathematical Statistics, 41:* 87-100.

Silvey, S.D. (1971). *Statistical Inference.* Baltimore: Penquin Books.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15:* 201-293.

Theil, H. (1971). *Principles of Econometrics.* New York: John Wiley and Sons.

Thompson, G.H. (1951). *The Factorial Analysis of Human Ability.* London: London University Press.

Thurstone, L.L. (1931). Multiple factor analysis. *Psychological Review, 38:* 406-427.

Thurstone, L.L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Tracy, D.S. and Dwyer, P.S. (1969). Multivariate maxima and minima with matrix derivatives. *Journal of the American Statistical Association, 64*: 1576-1594.

Travinsky, I. and Bargmann, R.E. (1964). Maximum likelihood estimation with incomplete data. *Annals of Mathematical Statistics, 35*: 647-657.

Turner, M.E. and Stevens, C.D. (1959). The regression analysis of causal paths. *Biometrics, 15*: 236-258.

Vetter, W.J. (1970). Derivatives operations on matrices. *IEEE Transactions on Automatic Control - AC, 15*: 241-244.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54*: 426-483.

Wald, A. (1948). Estimation of a parameter when the number of unknown parameters increases indefinitely with the number of observations. *Annals of Mathematical Statistics, 19*: 220-227.

Watson, G.S. (1964). A note on maximum likelihood. *Sankhyā - Series A, 26*: 303-304.

Wold, H. (Editor) (1964). *Econometric Model Building:* Essays on the causal chain approach. Amsterdam: North Holland.

Wright, S. (1918). On the nature of size factors. *Genetics, 3*: 367-374.

Zacks, S. (1971). *The Theory of Statistical Inference.* New York: John Wiley and Sons.