This is to certify that the

dissertation entitled

EMPIRICAL BAYES ESTIMATION FOR UNBALANCED MULTILEVEL STRUCTURAL EQUATION MODEL VIA THE EM ALGORITHM

presented by

See-Heyon Jo

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Counseling,
Educational Psychology & Special
Education (Statistics & Research Design)

Major professor

Date November 15, 1994

EMPIRICAL BAYES ESTIMATION FOR UNBALANCED MULTILEVEL STRUCTURAL EQUATION MODELS VIA THE EM ALGORITHM

By

See-Heyon Jo

A DISSERTATION

Submitted to
Michigan State University
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Counceling, Educational
Psychology and Special Education

1994

ABSTRACT

EMPIRICAL BAYES ESTIMATION FOR UNBALANCED MULTILEVEL STRUCTURAL EQUATION MODELS VIA THE EM ALGORITHM

By

See-Heyon Jo

The question of how to analyze unbalanced hierarchical data generated from structural equation models has been a common problem for educational researchers and analysts. Among difficulties plaguing statistical modeling are estimation bias due to measurement error and the estimation of the effects of hierarchical, social milieu where education takes place. Over the last two decades, substantial progress in multilevel structural modeling and estimation techniques has been made for the balanced sampling design.

This dissertation presents empirical Bayes estimation procedures for the multilevel structural equation models in the context of unbalanced sampling designs. The computational procedure is implemented via the EM algorithm. It is particularly useful for the problem of estimating a large number of parameters in multilevel structural equation models.

A multilevel structural equation modeling process with an example illustrates the general principles of the empirical Bayes estimation with the EM algorithm. The accuracy of the algorithm was tested using a set of artificial data. The numerical results suggest that this new methodology is a potentially useful means for studying hypothesized causal relations among latent variables varying at two levels of hierarchy.

To my parents, brothers, wife and friends

ACKNOWLEDGMENT

I wish to express my extreme gratitude to my major professor, Dr. Stephen W. Raudenbush, for his patient guidance over the last five years. His insights contributed greatly to the content of this dissertation. And his encouragement, intellectual and financial support helped me complete this dissertation.

On a more general note, I would like to thank the entire faculty of the Department of Counseling, Educational Psychology and Special Education at Michigan State University for providing me the opportunity to pursue an advanced degree in statistics and research design, and for making my stay here a very rewarding experience.

My special gratitude goes to Dr. William Schmidt for his support. In the winter of 1990, Dr. Raudenbush and Dr.Schmidt gave me a precious opportunity to study multilevel structural equation models under their supervision. That winter enriched the idea of this dissertation.

I would like to express my gratitude to Dr. Richard Houang for his insightful suggestions and comments. I wish to express my appreciation to the rest of my committee, Dr. James Stapleton and Dr. Frank for reviewing my work and providing suggestions for improvement.

To my father, mother, brothers, and relatives, I owe a special "thank you" for their support and encouragement. To my wife, Kyung-Nam Lee, I dedicate this dissertation. Without her unfailing love and support this study could not exist.

Finally, I believe William Faulkner (1897-1962) also deserves some thanks for saying :

> " I would like to think that there was someone there at that time too, to reassure them that man is tough, that nothing, nothing -war, grief, hopelessness, despair can last as long as man himself can last; that man himself will prevail over all his anguishes, provided he will make the effort to; make the effort to believe in man and in hope -to seek not for a mere crutch to lean on, but to stand erect on his own feet by believing in hope and in his own toughness and endurance."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

As a consequence of various theoretical developments and of improvements in computing, maximum likelihood (ML) estimation has become a viable procedure for estimating parameters in multilevel structural equation models under the balanced sampling design. Many of these developments were reviewed in detail by Jo (1993).

The initial interest in ML estimation of a multilevel covariance structure model was noted and developed by Schmidt (1969). Schmidt and Wisenbaker (1986) extended this work to the structural equation models (Joreskog, 1973) for balanced hierarchical data.

McDonald and Goldstein (1989) derived the likelihood equations and derivatives for a bilevel structural equation model which allows for variables measured strictly at a higher level, though no computational approach was made available. They also indicated that the procedure for computation of ML estimates is currently less well developed for the unbalanced sampling design.

Recently, based on the balanced-data theory provided by Schmidt (1969) and McDonald and Goldstein (1989), Muthen (1990) showed that the maximum likelihood fitting function could be rewritten such that the between and within structural models could be estimated by means of a multi-population analysis in LISCOMP (Muthen,1987) or other comparable structural equations software. In the case of balanced data this could be accomplished by treating the within-group deviations as sampled from one population and the between-group deviations as sampled from a second population. For the case of

unbalanced data, each cluster of groups which have the same number of observations is treated as one population.

Lee and Poon (1992) also used the strategy of classifying level-2 units into subsets of level-2 units having equal sample size. They proposed an estimator for such data which, though not maximum likelihood (ML), has the same asymptotic distribution as the ML estimator as the number of level-2 units per subset increases without bound. Computationally this estimator is available using standard software program, such as LISREL (Joreskog and Sorbom, 1993) or EQS (Bentler, 1989).

More recently, Raudenbush (in press) proposed an alternate approach for the unbalanced case. He conceptualized the problem in the framework of groups which could all have the same number of sampled cases but are missing data for some individuals. In particular, in the M-step (maximization) the method uses the standard program such as EQS (Bentler, 1989). Vredevooogd (1993) applied this general approach to the global models (where two indicators for a group-level latent variable are included) in her dissertation proposal. Jo (1993) also applied the general procedure to a set of linear structural equation models.

The purpose of this dissertation is to develop empirical Bayes estimation procedures for computing maximum likelihood (ML) estimates of the parameters in the multilevel structural equation models in the context of the unbalanced sampling design. The procedures do not require classifying level-2 units into subsets of level-2 units having equal sample size. We present a multilevel structural equation modeling process with an

artificial example, which illustrates the general principles of empirical Bayes estimation with the EM algorithm.

1.1 The General Problems

A distinguishing characteristic of the data encountered in many areas of educational, medical, social science (sociology, econometrics, management, marketing) and genetics research is that the sampling structure is hierarchical. For instance, students are nested within schools, workers within firms, patients within some treatment-specific medical programs, family members in a family tree, or residents within census tracks. Individuals also can take the role of independently observed groups.

Generally, students are taught in groups by a teacher, several classrooms and teachers are grouped together into a school, schools into districts and districts are clustered in states. Then students who attend the same school or classroom are expected to share certain educational policies and practices. As a result, the educational outcomes for these students will be, to varying degrees, intercorrelated. These effects of clusters are most validly viewed within the context of multilevel linear models.

Much of social science data comes from two-, or three- stage sampling designs. Large-scale educational assessment, for example, is typically conducted by drawing a sample of schools and, from those schools, sampling the students who will take the assessment test. This hierarchical fashion of sampling is frequently selected in large-scale surveys, such as the National Longitudinal Study with data gathered regarding the educational aspirations and attainment of high school seniors of 1972 and the Second International Mathematics Studies

(SIMS; Crosswhite et al., 1985), and the Third International Mathematics and Science Studies (TIMSS; Schmidt, 1993). Under the standard assumption of I.I.D, covariance structure modeling (Joreskog, 1973) of such data misguides statistical inference by not taking into account the intracluster correlations which are present in hierarchical data. Hence an important implication of such structure is that the classical assumption of independence among nested observations is violated. Ignoring the existence of hierarchy in model building gives rise to several methodological and substantive problems and that have been well-documented in the literature (Burstein,1980).

In the context of the linear model, statisticians (Lindley and Smith, 1972; Smith, 1973; Raudenbush, 1984, 1988; Aitkin and Longford, 1986; Goldstein, 1986) developed hierarchical linear models (HLM) which are appropriate and powerful means of modeling hierarchical data. Many of these developments and examples are found in the recent book written by Bryk and Raudenbush (1992).

It was not until hierarchical modeling techniques (Aitkin and Longford, 1986; Goldstein, 1986; Mason, Wong and Entwistle, 1984; Raudenbush, 1984; Raudenbush & Bryk, 1986) were developed that complex relationships among variables across all levels could be inferred. Such techniques have been widely used for various types of research topics such as cognitive growth and change (Bryk and Raudenbush, 1987; Goldstein, 1989), population studies (Mason et al., 1984), meta-analysis (Raudenbush and Bryk, 1985), and evaluation of educational effectiveness (Aitkin and Longford, 1986; Raudenbush and Bryk, 1986).

However there have been only rare attempts of applying the methodology to

the structural equation models for hierarchical data. In discussing the empirical Bayes approach, Muthen (1990) also indicated the plausibility of application to the multilevel structural equation model by estimating each group's factor value under the assumption of "exchangeability" (deFinetti, 1937; Lindley and Smith, 1972) of the groups.

The main tasks in this dissertation are (a) to incorporate the effects of two levels of social organization into statistical models for outcomes measured at the individual level or/and cluster level; (b) to develop latent variable models that simultaneously incorporate effects of structural relations and measurement error.

1.2 Objectives

The primary objectives of this dissertation are:

(1) To review previous relevant advances in statistical modeling and estimation procedures for the multilevel structural equation model,

(2) To describe a multilevel structural equation model and develop empirical Bayes estimation procedures for ML estimates via the EM algorithm,

(3) To write a necessary computer program to implement this new estimation procedure,

(4) To demonstrate by the use of simulated data that this estimation procedure produces accurate parameter estimates.

1.3 Brief History of Single Level Structural Equation Models

The measurement of latent constructs with multiple manifest variables began with the work of Spearman (1904) early in this century. In 1904, Spearman proposed the method of factor analysis to investigate Galton's theory (Galton, 1883) of "intelligence" that a single common factor and a specific factor constituted cognitive ability as measured educational tests. This model evolved to represent intelligence with a hierarchical structure (Vernon, 1950). Thurstone (1947) extended Spearman's theory to a multiple factor analysis model. Apart from Spearman's factor analysis there is the work of Wright (1934), who derived the path analysis for the research of genetics. Before Lawley's (1943) development of the maximum likelihood function for factor analysis, the classical method was not based on the statistical theory of random sampling. While computational methods were not available at that time, Lawley derived the partial derivatives of the logarithm of the likelihood function with respect to each element of the covariance matrix. Based on Lawley's ML estimation theory, Joreskog (1967, 1973, 1977) developed the structural equation model.

> "Linear structural equation modeling (1977) represents an important combining of the traditions of econometric and psychometric methods producing a set of procedures that enable researchers to separate the structural part of the model from the measurement properties of the variables. The structural part of the model represents hypothesized networks among latent construct variables imperfectly projected in the observed indicators. This scheme of formulation allows the separation of issues of measurement error from the assessment of the structural relationships that embody the actual purposes of the research. This tradition has seen

many applications in education, psychology, and sociology over the last 20 years." (Raudenbush and Schmidt, 1991)

Joreskog (1977) adapted two optimization algorithms of steepest descent and the Davidon-Flectcher-Powell. Recently he extended LISREL to nonlinear structural model (Joreskog and Sorbom, 1993). The recent version of LISREL8 with PRELIS2 (Jorekog and Sorbom, 1993) provides user-easy language SIMPLIS for the PC-window.

As noted by Austin and Wolfle (1991), structural equation modeling is not a recent development (Bentler, 1983; McDonald, 1978), nor is it the work of any one individual or disciplinary area. However, there have been other lines of inquiry.

The second line of inquiry is represented by the work of Bock (1960), Bock and Bargmann (1966), Wiley (1967), and Wiley, Schmidt and Bramble (1973). They addressed a set of models formally parameterized as the factor analysis model but with different notions as to the roles of the parameters themselves.

The model for the observed score vector of p tests is :

$$y = \mu + A\xi + e \tag{1.3.1}$$

The model implies that the vector y has a multivariate normal distribution with mean vector μ and covariance matrix Σ:

$$\Sigma = A\Phi A' + \Psi \qquad (1.3.2)$$

the general assumption for the model is (1) Φ is considered as a diagonal matrix, (2) Ψ is considered to be $I\sigma^2$ or heterogeneous, (3) A is completely specified or scaled by unknown but estimable matrix of scaling factor, Γ. Wiley (1967) developed a set of 16 models that can be hypothesized by applying different combinations of restrictions to the three main components of the general model. These models cover a number of Joreskog's confirmatory factor analysis models. Wiley, Schmidt and Bramble (1973) developed the maximum likelihood estimators for eight of these models using the restrictions : (1) A is completely specified and unscaled, (2) A is completely specified and scaled by an unknown but estimable matrix of scaling weights Γ.

The third line of inquiry has a different tradition from the other two procedures of modeling and analysis. Recently the application of the regression component decomposition (Schonemann and Steiger, 1976) approach is proposed to avoid the indeterminacy problems in the conventional LISREL model. In the RCD (regression component decomposition) model, the common factors are defined as linear combinations of the observed variables and the factor loadings as the regression coefficients, regressing the manifest variables on the so defined common factors. In the context of LISREL model, Haaggen and Vittadini (1991) presented a method to decompose the observed data into components which have analogous properties to those of the latent variables in the LISREL model.

1.4 The Prior Work of Multilevel Structural Equation Models

Several articles deal with the analysis of the structural equation model for hierarchical data. In this section we review the proposed methods for their models and computational methods.

In Schmidt and Wisenbaker (1986), the measurement and structural models are:

$$y = \mu_y + \Lambda\eta + \Lambda_b\eta_b + \varepsilon + \varepsilon_b \qquad (1.4.1)$$

$$x = \mu_x + \Gamma\xi + \Gamma_b\xi_b + \varpi + \varpi_b \qquad (1.4.2)$$

$$\eta = A\eta + B\xi + \theta \qquad (1.4.3)$$

$$\eta_b = A_b\eta_b + B_b\xi_b + \theta_b \qquad (1.4.4)$$

The vectors μ_y, μ_x, in the equations (1.4.1) and (1.4.2) are simply the expected values of y and x respectively. The matrix Λ contains coefficients relating the latent endogenous within-groups variables η to the manifest variables, y. Similarly, Λ_b relates the true endogenous between-groups variables, η_b, to the observed variables, y. The vectors $\varepsilon, \varepsilon_b$, are the errors of measurement associated with the within-groups and between-groups levels respectively. The coefficients matrices, Γ, Γ_b, and the vectors $\xi, \xi_b, \varpi, \varpi_b$ bear similar relationships to the observed vector, x.

Equation (1.4.3) stipulates that the latent within-groups endogenous variables are expressible as linear functions of themselves and the latent within-groups exogenous variables. The matrices A and A_b must be lower triangular such that $(I - A)$ and $(I - A_b)$ are invertible. The vector, θ, contains the errors in structural equation.

Equation (1.4.4) is composed of parallel constructs dealing with the expression of the between-groups latent endogenous variables.

By formulating the total variance-covariance as a simple additive function of the within and between group variance-covariance matrices, the within group covariance matrix, Σ_w, and between group variance covariance matrix, Σ_b, were expressed as follows:

$$\text{var}\begin{pmatrix} x \\ y \end{pmatrix} = \Sigma_w + \Sigma_b \tag{1.4.5}$$

where

$$\Sigma_w = \begin{bmatrix} \Gamma\Sigma_\xi\Gamma' + \Psi_\varpi & \Gamma\Sigma_\xi B'(I-A)^{-1}\Lambda' \\ \textit{symmetric} & \Lambda[(I-A)^{-1}B\Sigma_\xi B'(I-A)^{-1}]\Lambda' \\ & +\Lambda[(I-A)^{-1}\Sigma_\theta(I-A)^{-1}]\Lambda' + \Psi_\varepsilon \end{bmatrix} \tag{1.4.6}$$

$$\Sigma_b = \begin{bmatrix} \Gamma_b\Sigma_{\xi_b}\Gamma_b' + \Psi_{\varpi_b} & \Gamma_b\Sigma_{\xi_b}B_b'(I-A_b)^{-1}\Lambda_b' \\ \textit{symmetric} & \Lambda_b[(I-A_b)^{-1}B_b\Sigma_{\xi_b}B_b'(I-A_b)^{-1}]\Lambda_b' \\ & +\Lambda_b[(I-A_b)^{-1}\Sigma_{\theta_b}(I-A_b)^{-1}]\Lambda_b' + \Psi_{\varepsilon_b} \end{bmatrix} \tag{1.4.7}$$

with the following definitions:

Σ_ξ: the variance-covariance matrix of the latent within-group exogenous variables,

Λ_B : the matrix of factor loadings associating manifest variables and latent variables at the group level.

Σ_θ: the variance-covariance matrix of the within-groups errors in equations, θ.

Ψ_{ϖ}: the variance-covariance matrix of the within-groups measurement error associated with the observed exogenous variables, x.

Ψ_{ε}: the variance-covariance matrix of the within-groups measurement error associated with the observed endogenous variables, y.

Λ : the matrix of factor loadings connecting manifest variables and latent variables at the individual level.

Σ_{ξ_b} : the variance-covariance matrix among the exogenous latent variables at the group level.

Σ_{θ_b} : the variance-covariance matrix among the endogenous latent variables at the individual level.

Ψ_{ε_b} : the variance-covariance matrix among the random measurement errors associated with the observed endogenous variables at the group level

Ψ_{ϖ_b} : the variance-covariance matrix among the random measurement errors associated with the observed exogenous variables at the group level

Given the assumption of a multivariate normal distribution, Schmidt and Wisenbaker derived the log likelihood function;

$$L(\theta) = \frac{(J - Jn)}{2} \ln|\Sigma_w| - \frac{J}{2} \ln|\Sigma_w + n\Sigma_b| - \frac{Jn}{2} tr[\Sigma_w^{-1} S_w] - \frac{n}{2} tr[(\Sigma_w + n\Sigma_b)^{-1} S_b] \qquad (1.4.8)$$

where J is the total number of groups while n is the number of level-1 units in each

group. And $S_w = \frac{1}{Jn}\sum_{j=1}^{J}\sum_{i=1}^{n}(y_{ij}-\bar{y}_j)(y_{ij}-\bar{y}_j)^T$, $S_b = \frac{n}{J}\sum_{j=1}^{J}(\bar{y}_j-\bar{y})(\bar{y}_j-\bar{y})^T$, and y_{ij} is the r by 1 observation vector for the i-th individual in the j-th group.

The maximum likelihood estimates for the parameters can be obtained by setting the first partial derivatives of the log likelihood function with respect to each parameter equal to zero and solving for the unknowns. Schmidt and Wisenbaker (1986) adapted the Fletcher-Powell method of optimization. This balanced model and theory are subsumed in the Muthen's muti-sample analysis (1990) using LISCOMP. Note that in Schmidt and Wisenbaker's formulation, all between group variables are aggregated versions of within-group variables. In Muthen's extension, variables observed strictly at the group level are included.

Muthen (1990) provided the ML fitting function for the unbalanced case through which computational strategies are explicitly identical to the multi-sample analysis in the standard LISREL, LISCOMP and EQS. He also described the necessary steps of analyzing the multilevel structural equation model.

Muthen (1990) postulated that "in the unbalanced case the number of groups with distinct group sizes may be rather small in any given application."
The measurement model for within group level is:

$$y_{ij} = \mu_j + \Lambda\eta_{ij} + \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim N(0,\Sigma_w) \tag{1.4.9}$$

$$\mu_j = \Lambda_b\eta_{b_j} + \varepsilon_{b_j}, \qquad \varepsilon_{b_j} \sim N(0,\Sigma_b) \tag{1.4.10}$$

where η_{ij} and η_{b_j} are p x 1 vectors of within latent variables and between latent variables respectively; Λ and Λ_b are r x p matrices of within-factor loadings and between-factor loadings for the y variables; and, ε_{ij} and ε_{b_j} are the vectors of within- and between-group measurement errors.

Based on the balanced-data theory, Muthen (1990) derived the ML fitting function :

$$\sum_{d=1}^{D} J_d \{\ln\left|\Sigma_d^{-1}\right| + tr[\Sigma_d^{-1}(S_{bd} + n_d(\bar{y}_d - \mu)(\bar{y}_d - \mu)^T)]\} + (N-J)(\ln\left|\Sigma_w\right| + tr[\Sigma_w^{-1} S_w]) \quad (1.4.11)$$

where

$$\Sigma_d = [\Sigma_w + n_d \Sigma_b] \quad (1.4.12)$$

$$S_{bd} = \frac{n_d}{J_d} \sum_{j=1}^{J_d} (\bar{y}_{.jd} - \bar{y}_{..d})(\bar{y}_{.jd} - \bar{y}_{..d})^T \quad (1.4.13)$$

$$S_w = \frac{1}{N-J} \sum_{d=1}^{D} \sum_{j=1}^{J_d} \sum_{i=1}^{n_{jd}} (y_{ijd} - \bar{y}_{.jd})(y_{ijd} - \bar{y}_{.jd})^T \quad (1.4.14)$$

with the following definitions:

$\bar{y}_{..d}$ = the sample mean vector for the d-th subset.

y_{ijd} = the outcome vector for the i-th individual in the j-th group classified into the d-th subset.

J_d = the number of groups of the d-th subset.

μ = the population grand mean vector.

n_d = the total number of individuals in d-th subset, d=1,...,D.

n_{jd} = the number of individuals in j-th group classified into the d-th subset.

j = index for the groups classified into a subset of distinct size

N = the total number of individuals in a study.

J = the total number of groups.

$\bar{y}_{jd}$ = the sample group mean vector for the j-th group.

Note that $i = 1,\ldots,n_{jd}$. $j = 1,\ldots,J_d$. $d = 1,\ldots,D$

From a structural equation modeling point of view, the multilevel data ML fitting function can be viewed as corresponding to a simultaneous analysis of independent observations from D + 1 heterogeneous "populations". with the D populations for S_{bd}'s plus the within-group "population". All of the parameters are constrained to be equal across the D between-group populations except for the scaling factor, n_d. In equation (1.4.41) "it should be noted that the between sample covariance matrices may be singular due to being created by summation over fewer units than variables. This may prevent the use of certain conventional structural modeling software where positive definite matrices are assumed." (Muthen, 1990). To find the ML estimates one has to set up a command file (see, Muthen, 1990) to run EQS or LISCOMP program according to the model equations.

By use of the statistical concept of "missing data" (Dempster, Laird, and Rubin, 1977) Raudenbush (in press) developed a new estimation procedure. In theory the balanced data is equivalent to the complete data, while unbalanced data is incomplete data. In the unbalanced case, the E-step computes the conditional

expectations of the complete data sufficient statistics given observed data and the current parameter estimates. In the M-step the standard EQS software can be used to find the ML estimates of the parameters iteratively. Raudenbush (in press) postulates that "by supposing one has sampled n units within each of J clusters, one can apply the Muthen's balanced-data method. However, within each group k, only n_{1j} observations are available with $n-n_{1j}$ observations missing. Then one might regard the balanced data as the complete data, the observed unbalanced data as incomplete data. Consequently the maximization of the complete data likelihood is the same as the Muthen's balanced data approach for hierarchical data using the LISCOMP (Muthen, 1986) or EQS (Bentler, 1989) and so on". The complete data sufficient statistics can not be observed but can be estimated by means of their conditional expectations given the observed data and a current estimates at the parameter values. This process constitutes the E-step.

At level 1 (within cluster) we have p observations on each of n units, collected in the p by 1 vector y_{ij}. These vary randomly around the cluster mean μ_j according to the model:

$$y_{ij} = \mu_j + e_{ij}, e_{ij} \sim N(0, \Sigma) \tag{1.4.15}$$

The model at level 1 can be written as :

$$\begin{bmatrix} y_{1j} \\ y_{2j} \end{bmatrix} = \begin{bmatrix} A_{1j} \\ A_{2j} \end{bmatrix} \mu_j + \begin{bmatrix} e_{1j} \\ e_{2j} \end{bmatrix} \tag{1.4.16}$$

where y_{1j} is pn_{1j} by 1 observed vector, y_{2j} is pn_{2j} by 1 missing vector.

$$\begin{aligned} A_{1j} &= 1_{n_{1j}} \otimes I_p \\ A_{2j} &= 1_{n_{2j}} \otimes I_p \\ \Psi_{1j} &= I_{n_{1j}} \otimes \Sigma \\ \Psi_{2j} &= I_{n_{2j}} \otimes \Sigma \\ n &= n_{1j} + n_{2j} \end{aligned} \tag{1.4.17}$$

At level 2 the model can be written as:

$$\begin{aligned} &\mu_j = \mu_y + u_j, u_j \sim N(0, T_{uu}) \\ &\begin{bmatrix} \mu_j \\ x_j \end{bmatrix} \sim N\left[\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} T_{uu} & T_{ux} \\ T_{xu} & T_{xx} \end{bmatrix}\right] \end{aligned} \tag{1.4.18}$$

x_j is a vector of group level observed variables.

The expected value of the complete data sufficient statistics for within group variance covariance matrix given the parameter estimates from the M step of the previous iteration is:

$$S_w = S_{1w} + \overline{w}(J-1)\Sigma_w + \sum_{j=1}^{J} w_j n_j [L_j + (\overline{y}_{1j} - \overline{y}_{2j}^*)(\overline{y}_{1j} - \overline{y}_{2j}^*)^T] \tag{1.4.19}$$

S_{1w} is the usual pooled within-group variance-covariance estimate based on the observed data;

$$\text{Let} \quad T = \begin{bmatrix} T_{uu} & T_{ux} \\ T_{xu} & T_{xx} \end{bmatrix} \tag{1.4.20}$$

$$L_j = (n_{1j}\Sigma_w^{-1} + T^{-1})^{-1} n_{1j}\Sigma_w^{-1} \tag{1.4.21}$$

$$\begin{aligned} w_j &= \frac{n_{2j}}{n} \\ \overline{w} &= N_2 / nJ \end{aligned} \tag{1.4.22}$$

N_2 is the total number of missing level-1 units. $\bar{y}_{1j}$ is the observed group mean vector and $\bar{y}_{2j}^*$ is the posterior mean vector for the missing observed data in each group.

$$\bar{y}_{2j}^* = L_j^{-1}\bar{y}_{1j} + (I - L_j^{-1})\Gamma x_j \tag{1.4.23}$$

$$\Gamma = T_{wx}T_{xx}^{-1} \tag{1.4.24}$$

The expected value of the complete data sufficient statistics for S_{yz} given the parameter estimates from M-step of the previous iteration is:

$$S_{yz} = S_{1yx} + \sum (x_j - \bar{x})[w_j(\bar{y}_{2j}^* - \bar{y}_{1j}) - \overline{w}(\bar{y}_2^* - \bar{y}_1)]^T \tag{1.4.25}$$

$$S_{1yx} = \sum x_j \bar{y}_{1j}^T - J\bar{x}\bar{y}_1^T \tag{1.4.26}$$

The sufficient statistic for Σ_{xx} is:

$$S_{xx} = \sum_{j=1}^{J} x_j x_j^T - J\bar{x}\bar{x}^T \tag{1.4.27}$$

The expected value of the complete data sufficient statistics for variance covariance matrix for y given the parameter estimates from the M step of the previous iteration is:

$$S_{y_b} = n\sum \bar{y}_j^{\bullet}\bar{y}_j^{\bullet\prime} + n\sum w_j^2 L_j^{-1} + J\bar{w}\Sigma_w - nJ\bar{y}^{\bullet}\bar{y}^{\bullet\prime} - (n/J)\sum w_j^2 L_j^{-1} + \bar{w}\Sigma \qquad (1.4.28)$$

where

$$\bar{y}_j^{\bullet} = (1-w_j)\bar{y}_{1j} + w_j\bar{y}_{2j}^{\bullet} \qquad (1.4.29)$$

$$\bar{y}^{\bullet} = (1-\bar{w})\bar{y}_1 + \bar{w}\bar{y}_2^{\bullet} \qquad (1.4.30)$$

$$\bar{y}_1 = \frac{1}{N_1}\sum_{j=1}^{J} n_{1j}\bar{y}_{1j} \qquad (1.4.31)$$

$$N_1 = \sum_{j=1}^{J} n_{1j} \qquad (1.4.32)$$

$$\bar{y}_2^{\bullet} = \frac{1}{N_2}\sum_{j=1}^{J} (n-n_{1j})\bar{y}_{2j}^{\bullet} \qquad (1.4.33)$$

Given the starting values produced by Muthen's ad hoc estimator, expected values for the complete data sufficient statistics are calculated by a Fortran program (Jo, 1993). These estimates will then be used to obtain maximum likelihood estimates of the parameters using the EQS program. An executive computer program provides the mechanism to switch on the Fortran program for the E-step and then the packaged program for the M-step.

The previous work in the field of multilevel structural equation models made substantial progress. In this dissertation we propose a new approach which does not require classifying level-2 units into a subset of equal sample size.

In conclusion of this chapter we provide a brief preview of subsequent chapters. In chapter 2, we present the general structural equation model with an

example. We also transform the model in terms of mixed model form. And we will briefly describe some typical research questions that may be addressed by means of multilevel structural equation models. In chapter 3, empirical Bayes estimation procedures are developed. The maximum likelihood estimators for the parameters are given, and we present the observed log-likelihood function. In chapter 4, artificial data are generated for checking the accuracy of the parameter estimates. The analysis is carried out by use of a program in Gauss. The index of goodness-of-fit of the model is presented, and the likelihood ratio for two alternate models is also given. In chapter 5, the summaries of each chapter are given, and the implications of the models are discussed. And future research questions are also presented.

CHAPTER 2

MULTILEVEL STRUCTURAL EQUATION MODELS

To illustrate how measurement and substantive theory can be integrated between and within levels in one overall framework, a hypothetical achievement model will be examined as an example. Consider a model where achievement scores of a mathematics test are believed to be influenced by a student's attitude toward mathematics, individual characteristics, e.g., gender, and class characteristics, e.g., teaching styles. The teaching styles such as discovery-oriented instruction or expository teaching are believed to influence attitude and achievement on the classroom level. Gender also is believed to be related to students' attitude and achievement on the individual level. The path-diagram for this hypothetical achievement model is shown in Figure 2.1.1. In our example attitude1 measures an individual's view on the usefulness of mathematics in our life and is based on the sum of scores on the four attitude items, each of them scaled as a Likert (1932) response with categories: strongly disagree (1), disagree (2), undecided (3), agree (4), and strongly agree (5). These items are: 1. I can get along well in everyday day life without using mathematics.

2. A knowledge of mathematics is not necessary in most occupations.

3. Mathematics is not needed in everyday day living.

4. Most people do not use mathematics in their jobs.

Attitude2 measuring "Attracted" to mathematics, is based on the sum of the scores on the five attitude items, each scaled as five-category Likert; strongly disagree (1), disagree (2), undecided (3), agree (4), and strongly agree (5). These items are:

1. I would like to work at a job that lets me use mathematics.

2. I think mathematics is fun.

3. Working with numbers makes me happy.

4. I am looking forward to taking more mathematics.

5. I refuse to spend a lot of my own time doing mathematics.

The ACH1 is the first part composed of basic facts and principles, while the ACH2 is composed of problem solving questions.

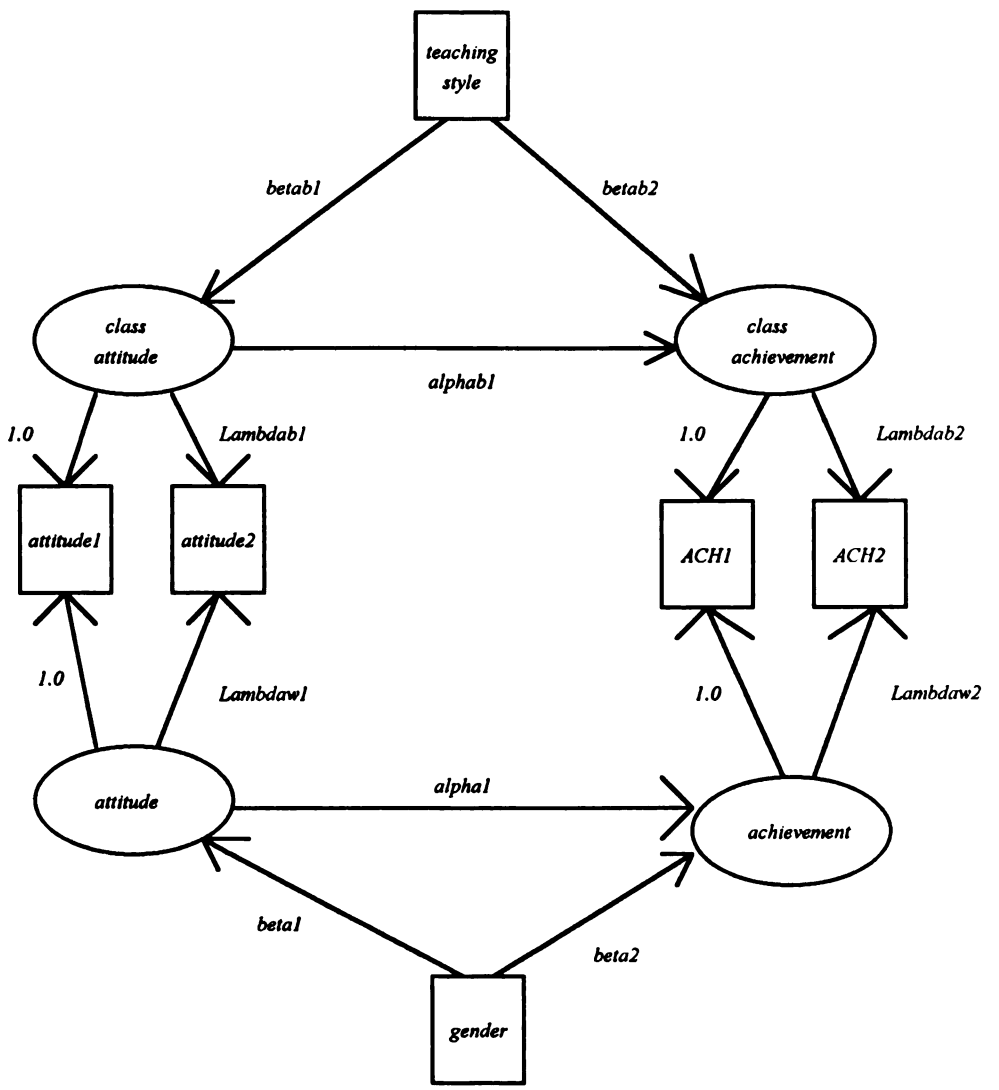

Figure 2.1.1 A Path Diagram for Multilevel Structural Equation Model

2.1 The Model and the Basic Notation

A simple item level equation for each individual:

$$y_{ij} = \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \tag{2.1.1}$$

$$\begin{bmatrix} y_{1ij} \\ y_{2ij} \\ y_{3ij} \\ y_{4ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{w1} & 0 \\ 0 & 1 \\ 0 & \lambda_{w2} \end{bmatrix} \begin{bmatrix} \eta_{1ij} \\ \eta_{2ij} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{b1} & 0 \\ 0 & 1 \\ 0 & \lambda_{b2} \end{bmatrix} \begin{bmatrix} \eta_{b1j} \\ \eta_{b2j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \\ \varepsilon_{3ij} \\ \varepsilon_{4ij} \end{bmatrix} \tag{2.1.2}$$

where j=1,2,..,J for classroom, i=1,2,..., n_j for students nested in classroom j. The subscript "w" means the within-level, while "b" means the between-level.

In terms of our educational example, equation (2.1.2) can be expressed as follows :

$$\begin{bmatrix} attitude1_{ij} \\ attitude2_{ij} \\ ACH1_{ij} \\ ACH2_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{w1} & 0 \\ 0 & 1 \\ 0 & \lambda_{w2} \end{bmatrix} \begin{bmatrix} attitude_{ij} \\ achievement_{ij} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{b1} & 0 \\ 0 & 1 \\ 0 & \lambda_{b2} \end{bmatrix} \begin{bmatrix} attitude_{b1j} \\ achievement_{b2j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \\ \varepsilon_{3ij} \\ \varepsilon_{4ij} \end{bmatrix} \tag{2.1.3}$$

where $\varepsilon_{ij} \sim N(0, \Sigma)$, a typical form for Σ is : $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$

Assuming structural linear relationships among constructs, the theoretical

relationships on the within-level depicted at the bottom part of Figure 2.1.1 can be expressed through the following structural equation:

$$\begin{bmatrix} \eta_{1ij} \\ \eta_{2ij} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \alpha_1 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1ij} \\ \eta_{2ij} \end{bmatrix} + \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \begin{bmatrix} z_{1ij} \\ z_{2ij} \end{bmatrix} + \begin{bmatrix} u_{1ij} \\ u_{2ij} \end{bmatrix} \tag{2.1.4}$$

where $u_{ij} = [u_{1ij} u_{2ij}]^T$, $u_{ij} \sim N(0, \Delta)$, $\Delta = Diagonal(\delta_p, p = 1, \ldots, P)$.

In terms of our educational example, equation (2.1.4) can be expressed as follows :

$$\begin{bmatrix} attitude_{ij} \\ achievement_{ij} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \alpha_1 & 0 \end{bmatrix} \begin{bmatrix} attitude_{ij} \\ achievement_{ij} \end{bmatrix} + \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \begin{bmatrix} 1 \\ gender_{ij} \end{bmatrix} + \begin{bmatrix} u_{1ij} \\ u_{2ij} \end{bmatrix} \tag{2.1.5}$$

Equation (2.1.4) stipulates that on the individual-level the latent variables are captured as a structural linear function of themselves and the predictor variable. In our example, gender is used as a predictor variable. In equation (2.1.4) z_{2ij} is gender, while z_{1ij} is unit value so that the model has intercept terms.

Now we reduce equation (2.1.4) into the equation (2.1.6).

$$\begin{bmatrix} \eta_{1ij} \\ \eta_{2ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \begin{bmatrix} z_{1ij} \\ z_{2ij} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{1ij} \\ u_{2ij} \end{bmatrix} \tag{2.1.6}$$

Then the reduced form for the within-level structural equations (2.1.4) is :

$$\begin{bmatrix} \eta_{1ij} \\ \eta_{2ij} \end{bmatrix} = \begin{bmatrix} z_{1ij} & z_{2ij} & 0 & 0 \\ 0 & 0 & z_{1ij} & z_{2ij} \end{bmatrix} \begin{bmatrix} \pi_{10} \\ \pi_{20} \\ \pi_{30} \\ \pi_{40} \end{bmatrix} + \begin{bmatrix} v_{1ij} \\ v_{2ij} \end{bmatrix} \tag{2.1.7}$$

where

$$\begin{bmatrix} v_{1ij} \\ v_{2ij} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{1ij} \\ u_{2ij} \end{bmatrix}, \quad v_{ij} \sim N(0, \mathrm{T}_\eta)$$

$$\begin{bmatrix} \pi_{10} \\ \pi_{20} \\ \pi_{30} \\ \pi_{40} \end{bmatrix} = vec^{*}\left(\begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \right)$$

where vec^{*} stacks the transpose of each row of a matrix into a vector.

In terms of our educational example equation (2.1.7) is can be written:

$$\begin{bmatrix} attitude_{ij} \\ achievement_{ij} \end{bmatrix} = \begin{bmatrix} 1 & gender_{ij} & 0 & 0 \\ 0 & 0 & 1 & gender_{ij} \end{bmatrix} \begin{bmatrix} \pi_{10} \\ \pi_{20} \\ \pi_{30} \\ \pi_{40} \end{bmatrix} + \begin{bmatrix} v_{ij} \\ v_{2ij} \end{bmatrix} \tag{2.1.8}$$

Now on the between-cluster level, the structural relationships depicted at the upper part of Figure 2.2.1 can be expressed as follows :

$$\begin{bmatrix} \eta_{b1j} \\ \eta_{b2j} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \alpha_{b1} & 0 \end{bmatrix} \begin{bmatrix} \eta_{b1j} \\ \eta_{b2j} \end{bmatrix} + \begin{bmatrix} \beta_{b1} \\ \beta_{b2} \end{bmatrix} \begin{bmatrix} w_j \end{bmatrix} + \begin{bmatrix} u_{b1j} \\ u_{b2j} \end{bmatrix} \tag{2.1.9}$$

where $u_{bj} = [u_{b1j} u_{b2j}]^T$, $u_{bj} \sim N(0, \Delta_b)$, $\Delta_b = Diagonal(\delta_{bp}, p = 1, \ldots, P)$.

In terms of our educational example, equation (2.1.11) can be expressed as follows :

$$\begin{bmatrix} attitude_{bj} \\ achievement_{bj} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \alpha_{b1} & 0 \end{bmatrix} \begin{bmatrix} attitude_{bj} \\ achievement_{bj} \end{bmatrix} + \begin{bmatrix} \beta_{b1} \\ \beta_{b2} \end{bmatrix} \begin{bmatrix} teachingstyle_j \end{bmatrix} + \begin{bmatrix} u_{b1j} \\ u_{b2j} \end{bmatrix} \tag{2.1.10}$$

Equation (2.1.9) is the expression for the structural relationships among latent variables and the predictor variable on the group level. In equation (2.1.9) w_j is teaching style. All of exogenous predictor variables are observed directly without error, e.g., school location (rural, urban), school sector (public, nonpublic), religion, gender, ethnicity, family size (numbers of a household), individual's age in months and years, current membership in a political party or sports club.

Then we have the reduced form for between-level structural equations (2.1.9)

$$\begin{bmatrix} \eta_{b1j} \\ \eta_{b2j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_{b1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{b1} \\ \beta_{b2} \end{bmatrix} \begin{bmatrix} w_j \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -\alpha_{b1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{b1j} \\ u_{b2j} \end{bmatrix} \tag{2.1.11}$$

We can represent (2.1.11) in the following form:

$$\eta_{bj} = W_j \pi_{bj} + v_{bj} \tag{2.1.12}$$

or

$$\begin{bmatrix} \eta_{b1j} \\ \eta_{b2j} \end{bmatrix} = \begin{bmatrix} w_j & 0 \\ 0 & w_j \end{bmatrix} \begin{bmatrix} \pi_{b10} \\ \pi_{b20} \end{bmatrix} + \begin{bmatrix} v_{b1j} \\ v_{b2j} \end{bmatrix} \tag{2.1.13}$$

where

$$\begin{bmatrix} v_{b1j} \\ v_{b2j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_{b1} & 1 \end{bmatrix} \begin{bmatrix} v_{b1j} \\ v_{b2j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_{b1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{b1j} \\ u_{b2j} \end{bmatrix}, \quad v_{bj} \sim N(0, \mathrm{T}_{\eta_b}),$$

$$\begin{bmatrix} \pi_{b10} \\ \pi_{b20} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\alpha_{b1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{b1} \\ \beta_{b2} \end{bmatrix}$$

In terms of our educational example, equation (2.1.10) can be written as follows :

$$\begin{bmatrix} attitude_{bj} \\ achievement_{bj} \end{bmatrix} = \begin{bmatrix} teachingstyle_j & 0 \\ 0 & teachingstyle_j \end{bmatrix} \begin{bmatrix} \pi_{b10} \\ \pi_{b20} \end{bmatrix} + \begin{bmatrix} v_{b1j} \\ v_{b2j} \end{bmatrix} \tag{2.1.14}$$

Representation of the Equations in Matrix Form

We can represent the equation (2.1.1) in matrix form without subscripts:

$$y = \Lambda_W \eta + \Lambda_B \eta_b + \varepsilon \tag{2.2.1}$$

where

$y = [y_{111} y_{211} y_{311} y_{411}, \ldots, y_{1n_J J} y_{2n_J J} y_{3n_J J} y_{4n_J J}]^T$, in our example, 4N by 1 vector.

$\varepsilon = [\varepsilon_{111} \varepsilon_{211} \varepsilon_{311} \varepsilon_{411}, \ldots, \varepsilon_{1n_J J} \varepsilon_{2n_J J} \varepsilon_{3n_J J} \varepsilon_{4n_J J}]^T$,in our example, 4N by 1 vector.

$$N = \sum n_j$$

$$\Lambda_W = \begin{bmatrix} \Lambda_{w1} & 0 & 0 & \ldots & 0 \\ 0 & \Lambda_{w2} & 0 & \ldots & 0 \\ 0 & 0 & \Lambda_{w3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \Lambda_{wJ} \end{bmatrix},$$

where, $\Lambda_{wj} = \begin{bmatrix} \Lambda_{wj1} & 0 & 0 & \ldots & 0 \\ 0 & \Lambda_{wj2} & 0 & \ldots & 0 \\ 0 & 0 & \Lambda_{wj3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \Lambda_{wjn_j} \end{bmatrix}$, $\Lambda_{wji} = \begin{bmatrix} 1 & 0 \\ \lambda_{w1} & 0 \\ 0 & 1 \\ 0 & \lambda_{w2} \end{bmatrix}$

$$\Lambda_B = \begin{bmatrix} \Lambda_{b1} & 0 & 0 & \ldots & 0 \\ 0 & \Lambda_{b2} & 0 & \ldots & 0 \\ 0 & 0 & \Lambda_{b3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \Lambda_{bJ} \end{bmatrix}, \quad \Lambda_{bj} = \begin{bmatrix} 1 & 0 \\ \lambda_{b1} & 0 \\ 0 & 1 \\ 0 & \lambda_{b2} \end{bmatrix}$$

$\eta_{ij} = [\eta_{111} \eta_{211}, \ldots, \eta_{1nJ_{KJ}} \eta_{2n_J J}]^T$ is in our example, a 2N by 1 vector

$\eta_b = [\eta_{b11}\eta_{b21},\ldots,\eta_{b1J}\eta_{b2J}]^T$ is in our example, a 2J by 1 vector

The matrix form for the equation (2.1.4) without subscripts is:

$$\eta = \mathrm{A}\eta + Bz + u \tag{2.2.2}$$

where

$$A = \begin{bmatrix} A_1 & 0 & 0 & \ldots & 0 \\ 0 & A_2 & 0 & \ldots & 0 \\ 0 & 0 & A_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & A_J \end{bmatrix}, \; A_j = \begin{bmatrix} A_{j1} & 0 & 0 & \ldots & 0 \\ 0 & A_{j2} & 0 & \ldots & 0 \\ 0 & 0 & A_{j3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & A_{jn_j} \end{bmatrix}, \; A_{ji} = \begin{bmatrix} 0 & 0 \\ \alpha_1 & 0 \end{bmatrix}$$

A_{ji} is a lower triangular matrix with diagonal elements of zero.

$$B = \begin{bmatrix} B_1 & 0 & 0 & \ldots & 0 \\ 0 & B_2 & 0 & \ldots & 0 \\ 0 & 0 & B_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & B_J \end{bmatrix}, B_j = \begin{bmatrix} B_{j1} & 0 & 0 & \ldots & 0 \\ 0 & B_{j2} & 0 & \ldots & 0 \\ 0 & 0 & B_{j3} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & B_{jn_j} \end{bmatrix},$$

$$\mathrm{B}_{ji} = \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \quad \text{for all (j i).}$$

$u = [u_{111}u_{211},\ldots,u_{1n_J J}u_{2n_J J}]^T$ is in our example, a 2N by 1 vector.

The matrix form for the equation (2.1.9) without subscripts is:

$$\eta_b = \mathrm{A}_b \eta_b + B_b w + u_b \tag{2.2.3}$$

where

$$A_b = \begin{bmatrix} A_{b1} & 0 & 0 & \dots & 0 \\ 0 & A_{b2} & 0 & \dots & 0 \\ 0 & 0 & A_{b3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & A_{bJ} \end{bmatrix}, A_{bj} = \begin{bmatrix} A_{bj1} & 0 & 0 & \dots & 0 \\ 0 & A_{bj2} & 0 & \dots & 0 \\ 0 & 0 & A_{bj3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & A_{bjn_j} \end{bmatrix}, A_{bji} = \begin{bmatrix} 0 & 0 \\ \alpha_{b1} & 0 \end{bmatrix}$$

A_{bji} is a lower triangular matrix with diagonal elements are zeros.

$$B_b = \begin{bmatrix} B_{b1} & 0 & 0 & \dots & 0 \\ 0 & B_{b2} & 0 & \dots & 0 \\ 0 & 0 & B_{b3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & B_{bJ} \end{bmatrix}, \quad B_{bj} = \begin{bmatrix} \beta_{b10} \\ \beta_{b20} \end{bmatrix} \text{ across all j.}$$

$u_b = [u_{b11} u_{b21}, \dots, u_{b1J} u_{b2J}]^T$ is in our example a 2J by 1 vector

Now we can express the reduced form equation (2.1.7) in matrix form:

$$\eta = Z\pi + v \tag{2.2.4}$$

where, $\eta_{ij} = [\eta_{111}\eta_{211},\ldots,\eta_{1n_jJ}\eta_{2n_jJ}]^T$ is in our example a 2N by 1 vector

$$Z = \begin{bmatrix} Z_{11} \\ Z_{21} \\ \cdot \\ \cdot \\ Z_{n_jJ} \end{bmatrix}, \quad Z_{n_jj} = \begin{bmatrix} z_{1ij} & z_{2ij} & 0 & 0 \\ 0 & 0 & z_{1ij} & z_{2ij} \end{bmatrix},$$

$$\pi = \begin{bmatrix} \pi_{10} \\ \pi_{20} \\ \pi_{30} \\ \pi_{40} \end{bmatrix}$$

$\nu_{ij} = [\nu_{111}\nu_{211},\ldots,\nu_{1n_jJ}\nu_{2n_jJ}]^T$ is in our example a 2N by 1 vector.

We can express the between-level reduced form equation (2.1.12) in matrix form:

$$\eta_b = W\pi_b + V_b \qquad (2.2.5)$$

2J x 1

where

$\eta_b = [\eta_{b11}\eta_{b21},\ldots,\eta_{b1J}\eta_{b2J}]^T$ is in our example a 2J by 1 vector

$$W = \begin{bmatrix} W_1 \\ W_2 \\ . \\ . \\ W_J \end{bmatrix}, \quad W_j = \begin{bmatrix} W_{1J} & 0 \\ 0 & W_{1J} \end{bmatrix}, \quad \pi_b = \begin{bmatrix} \pi_{b10} \\ \pi_{b20} \end{bmatrix}$$

$v_b = [v_{b11} v_{b21}, \ldots, v_{b1J} v_{b2J}]^T$ is in our example a $2J$ by 1 vector

Table 2.1 Structural parts of the multilevel structural equation model

	Original Form	Reduced Form	where
Within group	$\eta = A\eta + Bz + v$	$\eta = Z\pi + v$	$\pi = vec[(\mathrm{I}_2 - \mathrm{A}_{ji})^{-1}\mathrm{B}_{ji}]$
Between group	$\eta_b = A_b\eta_b + B_b w + v_b$	$\eta_b = W\pi_b + v_b$	$\pi_b = (\mathrm{I}_2 - \mathrm{A}_{bji})^{-1} B_{bj}$

Transforming the Model into the Mixed Model Form

By substituting the structural equations (2.2.4) and (2.2.5) shown in Table 2.1 into (2.2.1) without subscripts, we have the following combined equation (2.3.1). This representation permits us to develop a special version of the EM algorithm for multilevel structural equation models.

$$y = [\Lambda_w Z \mid \Lambda_b W] \begin{bmatrix} \pi \\ \pi_b \end{bmatrix} + [\Lambda_w \mid \Lambda_b] \begin{bmatrix} v \\ v_b \end{bmatrix} + \varepsilon \qquad (2.3.1)$$

In a more compact form we have:

$$y = \Lambda_0 \tilde{Z}\pi + \Lambda_0 \varpi + \varepsilon \tag{2.3.2}$$

where

$$\Lambda_0 = [\Lambda_w | \Lambda_b],$$

$$\tilde{Z} = \begin{bmatrix} Z & 0 \\ 0 & W \end{bmatrix}$$

$$\pi = \begin{bmatrix} \pi_0 \\ \pi_{b0} \end{bmatrix}$$

$$\varpi = \begin{bmatrix} v \\ v_b \end{bmatrix}$$

The model equation (2.3.2) is a special case of the general mixed model (Raudenbush, 1988):

$$Y = A_1\theta_1 + A_2\theta_2 + E \tag{2.3.3}$$

$$A_1 = \Lambda_0 \tilde{Z},$$

$$A_2 = \Lambda_0,$$

$$\theta_2 = \varpi, \tag{2.3.4}$$

$$\theta_1 = \pi,$$

$$E = \varepsilon,$$

In equation (2.3.3),

$\theta_1 \sim N(0,\Gamma)$, since our prior knowledge about θ_1 is assumed null, the prior precision associated with θ_1 becomes null.

$$\theta_2 \sim N(0,\Omega_2), \quad \Omega_2 = subdiag(T_0),$$

$$T_0 = \begin{bmatrix} I_{n_k} \otimes T_\eta & 0 \\ 0 & I_K \otimes T_{\eta_b} \end{bmatrix}$$

$$E \sim N(0,\Psi),$$

$$\Psi = I_N \otimes \Sigma_r$$

Based on this general model(2.3.2) I develop the empirical Bayes estimation procedure in Chapter 3.

In our structural equation model, we consider a population of N level-one units, indexed k (group) and i (individual). Associated with each level-one unit are three vector-valued variables y, z and w. The values of the design variables, z and w, are completely known for all level-one units before observations are carried out, but the values of the outcome variables, y (the four indicators in our illustrative example), are not known at all. Design variables are considered fixed and known in our multilevel structural equation models.

Then the marginal distribution of y is:

$$y \sim N(\mu,\Phi) \tag{2.3.5}$$

where

$$\mu = \Lambda_0 \tilde{Z}\pi \tag{2.3.6}$$

$$\Phi = \Lambda_0 \tilde{Z} \Gamma (\Lambda_0 \tilde{Z})^T + \Lambda_0 \Omega \Lambda_0^T + \Sigma \qquad (2.3.7)$$

where

$$\Omega = \begin{bmatrix} T_\eta & 0 \\ 0 & T_{\eta_b} \end{bmatrix}$$

And the conditional distribution of y given η is:

$$y|\eta \sim N(\Lambda_0 \eta, \Sigma) \qquad (2.3.8)$$

Note that in the model there are not measurement errors at 2 levels that are distinct from the model residuals. This is a limitation on the illustrative example. To have an identifiable model we restrict the factor loadings for the first indicators of the latent variables to unit, the variance-covariance matrix Σ to a diagonal matrix, and "A" matrices to be lower triangular. In our model the total number of variance-covariance parameters to be estimated are 16, while the number of unique elements of the variance structure are 10 for each level.

One can also include the group-level observed variables (global variables) for latent constructs in the general multilevel structural equation model (2.1.2). For example, in the study of the United States SIMS data each of the constructs of teaching practices and training and experience are measured by two indicators (Vredevoogd, 1993). In that case we have the following form of the item-level equations for each individual:

$$\begin{bmatrix} y_{1ki} \\ y_{2ki} \\ y_{3ki} \\ y_{4ki} \\ x_{1ki} \\ x_{2ki} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1ki} \\ \eta_{2ki} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{b1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{b2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{b3} \end{bmatrix} \begin{bmatrix} \eta_{b1k} \\ \eta_{b2k} \\ \eta_{b3k} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1ki} \\ \varepsilon_{2ki} \\ \varepsilon_{3ki} \\ \varepsilon_{4ki} \\ \varepsilon_{5ki} \\ \varepsilon_{6ki} \end{bmatrix} \tag{2.3.9}$$

where x_{1ki}, x_{2ki} are the indicators for the group-level construct, teacher's teaching practices, which is supposed to influence the class posttest (Vredevoogd, 1993). Then one can see that this model is a special form subsumed into the general model (2.3.1).

In the conclusion of this section we note the measurement model specifications. There are three types of specifications. The first measurement model is implemented by requiring equal factor loadings for all manifest variables and equal unique variances (Joreskog, 1971). The second measurement model retains the assumption of identical error variances across measures, but allows factor loadings to differ. The second model provides a more realistic description of actual data where observed measures are similar in content but differ in difficulty. The third measurement model is that the observed measures have identical factor loadings but have unequal error variances. Of particular importance are the measurement models in which those measurements have different factor loadings and unequal error variances, but the manifest variables are highly correlated (i.e., they measure the same thing to somewhat high degree).

CHAPTER 3

EM ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATES

After a model has been formulated, the statistical problems are to estimate the parameters in the model and to test the fit of the model to the data. General descriptions of the EM algorithm for the multilevel structural equation models are given in this chapter. Dempster, Laird and Rubin (1977) presented the EM algorithm as a general iterative method for computing maximum likelihood estimates from "incomplete data". Wu (1983) presented it in a more general context, viewing it as a special optimization algorithm. The EM algorithm is particularly useful when analytic expressions exist for the conditional expectation of the missing data and for the maximum likelihood estimates (MLE) of the model parameters given the observed data and missing data. Although in the literature it has been known as a method for estimating parameters of a model when observed data can be regarded as incomplete data, there were early uses of EM notions by Hartley (1958), Healy and Westmoratt (1956), Baum et al (1970), Brown (1974) and Sundberg (1974). In Rubin (1991) the essential idea of EM algorithm is briefly depicted:

> "The basic idea behind the EM algorithm is very old and very intuitive and can be colloquially described follows:
>
> 1. Given a problem that is difficult to solve, formulate it so that if missing data were observed, then the solution would be at hand; in particular, formulate the problem so that a good estimate (e.g., the maximum likelihood estimate, MLE) of the

parameter $\theta, \hat{\theta}$, would be easy to find if the missing values, Y_{mis}, were observed in addition to the observed values, Y_{obs}. Notice that "missing data" is viewed quite broadly to include, for example, latent variables in psychometric models.

2. Consequently, fill in a set of values for Y_{mis} and solve the problem(i.e., find $\hat{\theta}$).
3. Using this $\hat{\theta}$, find better values of Y_{mis} to fill in, and then repeat Point 2 to find a new value of $\hat{\theta}$.
4. Iterate until the values of $\hat{\theta}$ converge."

Based on this basic notion, one can conceptualize the implementation of the EM algorithm for the multilevel structural equation model. In section 3.1, we discuss the concepts of incomplete and complete data as applied to the multilevel structural equation model. We also develop the posterior distributions of the random vectors in equation (2.1.3). In section 3.2, we present the iterates for the implementation of EM algorithm. We also present the maximum likelihood estimates. In section 3.3, we present the observed-data log likelihood function.

3.1 General Description and Application to the Multilevel Structural Equation Model

Through casting the measurement model and the reduced form of the structural equation for latent variables into the general mixed model (Raudenbush, 1988), we can conceptualize our problem as having complete data and incomplete data. Note that in the multilevel structural equation model the factor loadings are parameters rather than observed predictors.

Then to compute maximum likelihood estimates of the dispersion matrices for the random vectors in the model (2.3.1), we apply the EM algorithm (Dempster, Laird and Rubin, 1977). We now discuss the concepts of complete data and incomplete data, as applied to model (3.1.1)

From Chapter 2 we have the following combined equation:

$$y_{ij} = \left[\Lambda_w Z_{ij} | \Lambda_b W_{ij}\right] \begin{bmatrix} \pi_0 \\ \pi_{b0} \end{bmatrix} + \Lambda_w \theta_{ij} + \Lambda_b \theta_{bj} + \varepsilon_{ij} \qquad (3.1.1)$$

In more compact form we have :

$$y_{ij} = \Lambda_0 \tilde{Z}_{ij} \pi + \Lambda_0 \varpi_{ij} + \varepsilon_{ij} \qquad (3.1.2)$$

where

$$\Lambda_0 = \left[\Lambda_w | \Lambda_b\right], \quad \tilde{Z}_{ij} = \begin{bmatrix} Z_{ij} & 0 \\ 0 & W_{ij} \end{bmatrix} \qquad (3.1.3)$$

$$\pi = \begin{bmatrix} \pi_0 \\ \pi_{b0} \end{bmatrix}, \quad \pi \sim N(0, \Gamma), \quad \Gamma = \begin{bmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{bmatrix}.$$

Since our prior knowledge about π is assumed null, Γ^{-1}, the prior precision (Dempster, Rubin and Tsutakawa, 1981) associated with π, becomes null, that is, $\Gamma^{-1} \to 0$. And,

$$\varpi_{ij} = \begin{bmatrix} \theta_{ij} \\ \theta_{bj} \end{bmatrix}, \quad \varpi_{ij} \sim N(0, T),$$

$$\mathrm{T} = \begin{bmatrix} \mathrm{T}_{\eta} & 0 \\ 0 & \mathrm{T}_{\eta_b} \end{bmatrix}$$

$$\varepsilon_{ij} \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

In the model equations,

$$\begin{aligned} y_{ij} &= \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \\ \eta_{ij} &= Z_{ij} \pi_0 + \theta_{ij} \\ \eta_{bj} &= W_j \pi_{b0} + \theta_{bj} \end{aligned} \tag{3.1.4}$$

$\Theta = \{\Lambda_w, \Lambda_b, \Sigma, \mathrm{T}_{\eta_b}, T_{\eta}\}$ is the set of parameters.

$y_{obs} = \{\mathrm{Y}, \mathrm{Z}, \mathrm{W}\}$ is the set of observed data.

$\mathrm{c} = \{\pi_0, \pi_{b0}, \theta_{bj}, \theta_{ij}, \varepsilon_{ij}\}$ is the set of missing data.

fixed random

The conditional probability density function is proportional to the joint probability density function :

$$f(c|\Theta, y_{obs}) \propto (2\pi)^{-Nr/2}|\Sigma|^{-N/2}\exp[(-0.5)\sum\sum(y_{ij}-\Lambda_0\tilde{Z}_{ij}\pi-\Lambda_0\varpi_{ij})^T$$

$$\Sigma^{-1}(y_{ij}-\Lambda_0\tilde{Z}_{ij}\pi-\Lambda_0\varpi_{ij})]\times(2\pi)^{-Jp/2}|T_\eta|^{-N/2}\exp[(-0.5)\sum\sum(\theta_{ij}^T T_\eta^{-1}\theta_{ij})]$$

$$\times(2\pi)^{-Js/2}|T_{\eta_b}|^{-J/2}\exp[(-0.5)\sum(\theta_{bj}^T T_{\eta_b}^{-1}\theta_{bj})]\times h(\pi) \qquad (3.1.5)$$

where $c=\{\pi_0, \pi_{b0}, \theta_{bj}, \theta_{ij}, \varepsilon_{ij}\}$, $\Theta = \{\Lambda_w, \Lambda_b, \Sigma, T_{\eta_b}, T_\eta\}$, r=the number of indicators, p=the dimension of θ_{ij}. And s=the dimension of θ_{bj}. The prior distribution $h(\pi)$ is considered a very small constant and it can be ignored while the empirical Bayes estimators are calculated (Fotiu, 1989; Dempster, Rubin and Tsutakawa, 1981).

If "c" were observable, some function of "c", t(c) would be a vector of complete data sufficient statistics for the dispersion matrices. In reality, the vector "c" is unobservable; however, the vector y, whose elements are linear functions of the elements of "c", is observable. In the realm of the EM algorithm, we regard the elements of "c" as the "missing data" and those of y as the "incomplete" data.

Then we can develop the iterative E-step (expectation step) and M-step (maximizing step) for computing new parameter estimates. The E-step consists of estimating t(c), which would be a complete data sufficient statistics if the vector "c" of complete data were available, by its conditional mean given the observed data and current estimates of Θ. The equations that are solved for parameters in the M-step can be regarded as an approximation to what would be the likelihood equations

if the vector "c" of complete data were observable.

From Dempster et al. (1977),

Each iteration of the EM algorithm solves : $\frac{\partial}{\partial\Theta}[Q(\Theta,\Theta^{(i)})|_{\Theta=\Theta^{(i)}}]=0$

where $Q(\Theta,\Theta^{(i)})=E(\ln[f(c;\Theta)]|Y=y;\Theta^{(i)})$

The necessary posterior location vector and dispersion matrix of the random vectors in the model (2.3.3) is :

$$D^{*}=\begin{bmatrix} A_1'\Psi^{-1}A_1+\Gamma^{-1} & A_1'\Psi^{-1}A_2 \\ A_2'\Psi^{-1}A_1 & A_2'\Psi^{-1}A_2+\Gamma^{-1} \end{bmatrix}^{-1}=\begin{bmatrix} D_{\pi}^{*} & C_{\pi\theta}^{*} \\ C_{\theta\pi}^{*} & D_{\theta}^{*} \end{bmatrix}$$

$$D_{\pi}^{*}=[A_1'\Psi^{-1}A_1-A_1'\Psi^{-1}A_2C^{-1}(A_1'\Psi^{-1}A_2)^{T}]^{-1}$$

$$D_{\theta}^{*}=C^{-1}+C^{-1}(A_2'\Psi^{-1}A_1)D_{\pi}^{*}(A_1'\Psi^{-1}A_2)C^{-1}$$

$$C_{\pi\theta}^{*}=-D_{\pi}^{*}A_1'\Psi^{-1}A_2C^{-1}$$

$$C_{\theta\pi}^{*}=(C_{\pi\theta}^{*})'$$

where

$$C=A_2^{T}\Psi^{-1}A_2+T^{-1}$$

Location (Posterior mean)

$$\begin{cases} \pi^{*}=D_{\pi}^{*}A_1'(I-\Psi^{-1}A_2C^{-1}A_2')\Psi^{-1}Y \\ \theta^{*}=C^{-1}A_2'\Psi^{-1}(Y-A_1\theta) \end{cases}$$

These posterior distributions given parameters provide point estimates and intervals

Inside of

8th term

$$(y_{ij} - \Lambda_0 \hat{Z}_{ij}\pi^* - \Lambda_0 \hat{\omega}_{ij}^*)^T \Sigma^{-1} (y_{ij} - \Lambda_0 \hat{Z}_{ij}\pi^* - \Lambda_0 \hat{\omega}_{ij}^*)$$

$$= (y_{ij}^T - \pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T - \hat{\omega}_{ij}^{*T}\Lambda_0^T)(\Sigma^{-1}y_{ij} - \Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^* - \Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^*)$$

$$= y_{ij}^T\Sigma^{-1}y_{ij} - \pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T\Sigma^{-1}y_{ij} - \hat{\omega}_{ij}^{*T}\Lambda_0^T\Sigma^{-1}y_{ij} - y_{ij}^T\Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^* + \pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^* + \hat{\omega}_{ij}^{*T}\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^*$$
$$- y_{ij}^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^* + \pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^* + \hat{\omega}_{ij}^{*T}\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^*$$

$$= y_{ij}^T\Sigma^{-1}y_{ij} - \underset{(1)}{2y_{ij}^T\Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^*} - \underset{(2)}{2y_{ij}^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^*} + \underset{(4)}{2\pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^*}$$
$$+ \underset{(3)}{\pi^{*T}\hat{Z}_{ij}^T\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{Z}_{ij}\pi^*} + \underset{(5)}{\hat{\omega}_{ij}^{*T}\Lambda_0^T\Sigma^{-1}\Lambda_0\hat{\omega}_{ij}^*}$$

needed for inference about the random vectors. These derivations of the posterior distributions draws heavily on prior research (Raudenbush, 1988).

3.2 Computation of the Iterates

Each iteration of the EM algorithm is composed of two steps, the E-step (Expectation step) and the M-step (Maximization step). The E-step carries out estimating complete data sufficient statistics by their conditional expectations given incomplete data, and the current estimates of parameters. The M-step consists of solving the complete data log-likelihood function.

To find the ML estimates, we have to maximize the function $Q(\Theta^{(i)},\Theta^{(i-1)})$. By definition $Q(\Theta^{(i)},\Theta^{(i-1)}) = E(\log[f(c;\Theta^{(i)})]|Y = y;\Theta^{(i-1)})$. As shown in Appendix 3, we have:

$$
\begin{aligned}
Q(\Theta^{(i)},\Theta^{(i-1)}) \;=\; & (-Nr/2)\ln 2\pi + (-N/2)\ln|\Sigma| \\
& + (-Np/2)\ln(2\pi) + (-N/2)\ln|\mathrm{T}_{\eta}| + (-Js/2)\ln(2\pi) + (-J/2)\ln|\mathrm{T}_{\eta_b}| \\
& -\frac{1}{2}\sum\sum tr[\Sigma^{-1}\{(\Lambda_0\tilde{Z}_{ij})D^{*}_{\pi}(\Lambda_0\tilde{Z}_{ij})^{T} + (\Lambda_0 D^{*}_{\varpi_{ij}}\Lambda_0^{T}) \\
& + (\Lambda_0\tilde{Z}_{ij}C^{*}_{\pi\varpi_{ij}}\Lambda_0^{T}) + (\Lambda_0 C^{*T}_{\pi\varpi_{ij}}\tilde{Z}_{ij}^{T}\Lambda_0^{T})\}] \\
& -\frac{1}{2}\sum\sum(y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi^{*} - \Lambda_0\varpi^{*}_{ij})^{T}\Sigma^{-1}(y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi^{*} - \Lambda_0\varpi^{*}_{ij}) \\
& -\frac{1}{2}\sum\sum tr[\mathrm{T}_{\eta}^{-1}E(\theta_{ij}\theta_{ij}^{T}|Y = y,\Theta^{(i-1)})] - \frac{1}{2}\sum tr[\mathrm{T}_{\eta_b}^{-1}E(\theta_{bj}\theta_{bj}^{T}|Y = y,\Theta^{(i-1)})] \qquad (3.2.1)
\end{aligned}
$$

In order to maximize, we take the first derivatives of $Q(\Theta^{(i)},\Theta^{(i-1)})$ with respect to $\mathrm{T}_{\eta}, \mathrm{T}_{\eta_b}, \Sigma, \Lambda_0$, respectively.

(1) $$\frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \mathrm{T}_{\eta}} = -\frac{N}{2}[2\mathrm{T}_{\eta}^{-1} - D(\mathrm{T}_{\eta}^{-1})]$$

$$+\frac{1}{2}\sum\sum[2\mathrm{T}_{\eta}^{-1}E(\theta_{ij}\theta_{ij}^{T}|Y=y,\Theta^{(i-1)})\mathrm{T}_{\eta}^{-1} - D\{\mathrm{T}_{\eta}^{-1}E(\theta_{ij}\theta_{ij}^{T}|Y=y,\Theta^{(i-1)})\mathrm{T}_{\eta}^{-1}\}] \quad (3.2.2)$$

where D means a diagonal matrix (Graybill, 1983). Thus D(T) is a diagonal matrix with i-th diagonal element equal to the i-th diagonal element of a matrix T. Setting the derivative equal to null matrix and solving gives the ML estimate (see, Press, 1982; Magnus and Neudecker, 1986).

Then the ML estimate is :

$$\hat{\mathrm{T}}_{\eta_b} = \frac{1}{N}\left[\sum\sum(\theta_{ij}^{*}\theta_{ij}^{*T} + D_{\theta_{ij}}^{*})\right] \quad (3.2.3)$$

(2) $$\frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \mathrm{T}_{\eta_b}} = -\frac{N}{2}[2\mathrm{T}_{\eta_b}^{-1} - D(\mathrm{T}_{\eta_b}^{-1})]$$

$$+\frac{1}{2}\sum[2\mathrm{T}_{\eta_b}^{-1}E(\theta_{bj}\theta_{bj}^{T}|Y=y,\Theta^{(i-1)})\mathrm{T}_{\eta_b}^{-1} - D\{\mathrm{T}_{\eta_b}^{-1}E(\theta_{bj}\theta_{bj}^{T}|Y=y,\Theta^{(i-1)})\mathrm{T}_{\eta_b}^{-1}\}] \quad (3.2.4)$$

Setting the derivative equal to null matrix and solving gives the ML estimate.

Then the ML estimate is :

$$\hat{\mathrm{T}}_{\eta_b} = \frac{1}{J}\left[\sum(\theta_{bj}^{*}\theta_{bj}^{*T} + D_{\theta_{bj}}^{*})\right] \quad (3.2.5)$$

$$\varepsilon_{ij} = Y_{ij} - \Lambda_0 \tilde{Z}_j \pi \Lambda \tilde{w}_j$$

$$E\left(\varepsilon_i \varepsilon_j^T \mid \psi, \theta^{(t)}\right) = E\left((Y_i - \Lambda_0 \tilde{z}_i \pi \Lambda \tilde{w}_i)(Y_j - \Lambda_0 \tilde{z}_j \pi \Lambda \tilde{w}_j)^T \mid \psi, \theta^{(t)}\right)$$

$$= E\left[(Y_i - \Lambda_0 \tilde{z}_i \pi \Lambda \tilde{w}_i)(Y_j - \pi^T \tilde{z}_j^T \Lambda_0^T - \tilde{w}_j^T \Lambda^T \Lambda_0^T) \mid \psi, \theta^{(t)}\right]$$

$$= \frac{1}{n} \left[Y_i Y_j^T - \Lambda_0 \tilde{z}_i \pi Y_j^T - \Lambda_0 \tilde{w}_i^T Y_j^T - Y_i \pi^T \tilde{z}_j^T \Lambda_0^T + \Lambda_0 \tilde{z}_i \pi \pi^T \tilde{z}_j^T \Lambda_0^T + \Lambda_0 \tilde{w}_i \pi^T \tilde{z}_j^T \Lambda_0^T \right.$$
$$\left. - Y_i \tilde{w}_j^T \Lambda^T \Lambda_0^T + \Lambda_0 \tilde{z}_i \pi \tilde{w}_j^T \Lambda^T \Lambda_0^T + \Lambda_0 \tilde{w}_i \tilde{w}_j^T \Lambda^T \Lambda_0^T \mid \psi, \theta^{(t)} \right]$$

$$= \Sigma_{ij}$$

$$(3)\quad \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \Sigma_f}\frac{\partial \Sigma_f}{\partial \sigma_r^2}=\Big[-\frac{N}{2}[2\Sigma^{-1}-D(\Sigma^{-1})]$$

$$+\frac{1}{2}\sum\sum[2\Sigma^{-1}E(\varepsilon_{ij}\varepsilon_{ij}^T|Y=y,\Theta^{(i-1)})\Sigma^{-1}-D\{\Sigma^{-1}E(\varepsilon_{ij}\varepsilon_{ij}^T|Y=y,\Theta^{(i-1)})\Sigma^{-1}\}]\Big]\times I_r^* \qquad (3.2.6)$$

where I_r^* is the column indicator vector which has a 1 in the r-th position and zeros in other positions. And Σ_f is a full matrix. Setting the derivative equal to zero and solving gives the ML estimate,. $\hat{\sigma}_r^2$. In appendix 4 we present the ML estimator for each element of Σ.

$$\hat{\Sigma}=Diagonal(\hat{\sigma}_1^2,\ldots,\hat{\sigma}_r^2) \qquad (3.2.7)$$

$$(4)\quad \frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \Lambda'}\frac{\partial \Lambda'}{\partial \lambda_{gk}}=\Big[-[\sum\sum\Sigma^{-1}\Lambda\tilde{Z}_{ij}C^*_{\pi\varpi_{ij}}]-[\Sigma^{-1}\Lambda C^{*T}_{\pi\varpi_{ij}}\tilde{Z}_{ij}^T]$$

$$-[\sum\sum\Sigma^{-1}\Lambda\tilde{Z}_{ij}D^*_{\pi}\tilde{Z}_{ij}^T]-[\sum\sum\Sigma^{-1}\Lambda D^*_{\varpi_{ij}}]$$

$$+[\sum\sum\Sigma^{-1}y_{ij}\pi^{*T}\tilde{Z}_{ij}^T]+[\sum\sum\Sigma^{-1}y_{ij}\varpi_{ij}^{*T}]-[\sum\sum\Sigma^{-1}\Lambda\tilde{Z}_{ij}\pi^*\pi^{*T}\tilde{Z}_{ij}^T]$$

$$-[\sum\sum\Sigma^{-1}(\Lambda\varpi_{ij}^*\pi^{*T}\tilde{Z}_{ij}^T+\Lambda\tilde{Z}_{ij}\pi^*\varpi_{ij}^{*T})]-[\sum\sum\Sigma^{-1}\Lambda\varpi_{ij}^*\varpi_{ij}^{*T}\Lambda']]\times I^*_{kgk} \qquad (3.2.8)$$

where Λ is a full matrix. The details of derivation of $\frac{\partial Q(\Theta^{(i)},\Theta^{(i-1)})}{\partial \Lambda}$ are given in appendix 5. Setting the derivative equal to zero and solving gives the ML estimate, $\hat{\lambda}_{gk}$. In appendix 5 we present the ML estimate for $\hat{\Lambda}_0$.

$$\hat{\Lambda}_0 = \left[\hat{\lambda}_{gk}\right]_{\Lambda} \tag{3.2.9}$$

where $\left[a_{gk}\right]_A$ means placing element a_{gk} in the g-th row and k-th column of matrix A, and zeros elsewhere. In our example, g= 2, 4, k= 1,2,3,4.

In sum the E-steps and M-steps are:

(1) E-step : Find $E\log[L(c,\Theta)|y, \mathrm{T}_{\eta}^{(i-1)}]$,

M-step : Substitute the equation (3.2.3) with these quantities, and then we obtain new T_{η}, set $\mathrm{T}_{\eta}^{(i)}$ equal to this new T_{η},

(2) E-step : Find $E\log[L(c,\Theta)|y, \mathrm{T}_{\eta_b}^{(i-1)}]$,

M-step : Substitute the equation (3.2.5) with these quantities, and then we obtain new T_{η_b}, set $\mathrm{T}_{\eta_b}^{(i)}$ equal to this new T_{η_b}.

(3) E-step : Find $E\log[L(c,\Theta)|y, \Sigma^{(i-1)}]$,

M-step : Substitute the equation (3.2.7) with these quantities, and then we obtain new Σ, set $\Sigma^{(i)}$ equal to this new Σ.

(4) E-step: Find π^{*T}, ϖ_{ki}^{*T}, $D_{\varpi_{ki}}^{*}$, D_{π}^{*}, $C_{\pi\varpi_{ki}}^{*}$. Note that these are all functions of $\Lambda_0^{(i-1)}$.

M-step : Substitute the equation (3.2.8) with these quantities, and then we obtain new Λ_0, set $\Lambda_0^{(i)}$ equal to this new Λ_0.

Then here the first iteration of the E and M step is completed. This algorithm proceeds until some user-specified termination criteria are met. For example, the algorithm might terminated when successive iterates differ from each other by no more than some number (i.e., $= 0^{-6}$).

3.3 Likelihood Function

We conclude this chapter with expressions for the observed log-likelihood function which is numerically simple to evaluate. Although the EM algorithm does not require an evaluation of the likelihood function, successive values of the function can be useful in monitoring the progress of the algorithm toward convergence at each iteration. And it's used in testing fit of alternate models.

Note the relationship among probability density functions :

$$P(y) = \frac{P(y|\theta)P(\theta)}{P(\theta|y)} \tag{3.3.1}$$

In the framework of the general mixed linear model, the equation (3.3.1) is rewritten as:

$$P(y|\Omega,\Psi,\Lambda_0)=\frac{P(y|\theta,\Omega,\Psi,\Lambda_0)P(\theta|\Omega,\Psi,\Lambda_0)}{P(\theta|y,\Omega,\Psi,\Lambda_0)} \tag{3.3.2}$$

The specific expressions for each of the density functions stated in equation (3.3.2):

$$P(y|\theta,\Omega,\Psi,\Lambda_0)=[(2\pi)^N|\Psi|]^{(-\frac{1}{2})}\exp[(-0.5)(y-A\theta)^T\Psi^{-1}(y-A\theta)] \tag{3.3.3}$$

$$P(\theta|\Omega,\Psi,\Lambda_0)=[(2\pi)^G|\Omega|]^{(-\frac{1}{2})}\exp[(-0.5)(\theta^T\Omega^{-1}\theta)] \tag{3.3.4}$$

where

y : the observed outcome vector for an individual

A : the design matrix for the multilevel structural equation model

$\theta=\begin{bmatrix}\pi^T & \varpi^T\end{bmatrix}^T$

Finally, the denominator part in equation (3.3.2) can be specified as:

$$P(\theta|y,\Omega,\Psi,\Lambda_0)=[(2\pi)^G|D_\theta^*|]^{(-\frac{1}{2})}\exp[(-0.5)(\theta-\theta^*)^TD_\theta^*(\theta-\theta^*)] \tag{3.3.5}$$

In particular, when $\theta=\theta^*$, we have:

$$P(y|\Omega,\Psi,\Lambda_0)=(2\pi)^{-N/2}|D_\theta^*|^{1/2}|\Psi|^{-1/2}|\Omega|^{-1/2}\exp[(-0.5)S(\theta^*)] \tag{3.3.6}$$

where:

$$S(\theta^*)=y^T\Psi^{-1}(y-A_1\theta_1^*-A_2\theta_2^*) \tag{3.3.7}$$

Now the log-likelihood function for the structural equation model may be:

$$LLR(\Omega, \Psi, \Lambda_0 | y) \propto (-0.5)\log|\Psi| + (0.5)|D_\theta^*| - (0.5)\log|\Omega| - (0.5)S(\theta^*) \tag{3.3.8}$$

First we evaluate:

$$\begin{aligned} &\det(\Psi) = \det(\Sigma \otimes I_N) = [\det(\Sigma)]^N \\ &\det(\Sigma) = (\sigma_1^2)\ (\sigma_2^2)\ \ldots(\sigma_r^2) \end{aligned} \tag{3.3.9}$$

$$\log(\det(\Psi)) = N[\log\sigma_1^2 + \log\sigma_2^2 + \ldots + \log\sigma_r^2] \tag{3.3.10}$$

And also:

$$\begin{aligned} &\det(\Omega) = \det(\Omega_\eta)\det(\Omega_{\eta_b}) \\ &\log[\det(\Omega)] = N\log[\det(\mathrm{T}_\eta)] + J\log[\det(\mathrm{T}_\beta)] \end{aligned} \tag{3.3.11}$$

Γ is considered large but fixed, from Dempster, Rubin and Tsutakawa (1981).

Finally we have:

$$\det(D_\theta^*) = \det\begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{bmatrix} = [\det(V_{11})][\det(V_{22} - V_{21}V_{11}^{-1}V_{12})][\det(d_{33} - d_{32}d_{22}^{-1}d_{23})] \tag{3.3.12}$$

The second term in equation (3.3.12) is given in appendix 2 as $\det(Q_3^{-1})$. Let the third term be $\det(U^{-1})$

where

$$d_{22} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, d_{23} = \begin{bmatrix} V_{13} \\ V_{23} \end{bmatrix}, d_{32} = [d_{23}]', d_{33} = V_{33} \tag{3.3.13}$$

$$S(\theta^*) = \sum\sum [y_{ki}^T \Sigma^{-1} (y_{ki} - A_{1ki}\pi^* - A_{2ki}\varpi_{ki}^*)] \tag{3.3.14}$$

Then the log-likelihood function for the structural equation model is :

$$\begin{aligned} LLR(\Psi, \Omega, \Lambda_0 | y) &= (N/2)\log(\det \Sigma) - (N/2)\log(\det T_\eta) - (K/2)\log(\det T_\beta) \\ &+ (1/2)\log(\det V_{11}) + (1/2)\sum \log(\det Q_{3k}^{-1}) + (1/2)\sum \log(\det U_k^{-1}) \\ &- \sum\sum [y_{ki}^T \Sigma^{-1} (y_{ki} - A_{1ki}\pi^* - A_{2ki}\varpi_{ki}^*)] \end{aligned} \tag{3.3.15}$$

At each iteration the algorithm evaluates the log-likelihood function to monitor the progress of the estimation.

CHAPTER 4

NUMERICAL RESULTS

In this chapter, I use a computer program written in Gauss (Version 2.2) to compute ML estimates from a set of artificial data. To verify that the produced estimates of the parameters are accurate the data are randomly generated with known (predetermined) population parameters.

The analysis was done for the balanced case and the unbalanced case. The Gauss program is designed to use cross-product matrices and initial starting values as input data and to perform computing over numerous iterations of the EM algorithm.

The path diagram for the model is given as a figure 2.1.1. In the example the two indicators for the ATT (attitude) latent variable are ATT1, ATT2. They are the student-reported responses to the questions in the attitude scale. The ACH1 variable measures achievement score in the "principle" parts, while the second indicator ACH2 measures achievement score in "problem solving" part.

4.1 Generating the Data

Before creating the necessary data we have to consider several issues. For the balanced data 10 subjects are selected per group. The distribution of the number of groups per group size is given in Table 4.1. Due to the heavy computational load

of estimating these model via the EM algorithm, only a single sample data will be generated.

Table 4.1 Number of Groups per Group Size

Group Size	Balanced Data	Unbalanced Data
6		10
7		10
8		10
9		20
10	500	450
Total	500	500

To create samples to be fit to the multilevel structural equation model specified in chapter 2, we modified the covariance structure by setting $\mathrm{var}(\pi) = 0$. Then the observed outcome vector y is calculated by using equation (4.1.1) :

$$\begin{bmatrix} y_{1ij} \\ y_{2ij} \\ y_{3ij} \\ y_{4ij} \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0 & 1.0 & 0.0 \\ 0.82 & 0.0 & 0.75 & 0.0 \\ 0.0 & 1.0 & 0.0 & 1.0 \\ 0.0 & 0.73 & 0.0 & 0.66 \end{bmatrix} \begin{bmatrix} 1.0 & z_{2ij} & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & z_{2ij} & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & w_j & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & w_j \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.1 \\ 0.31 \\ 0.33 \\ 0.25 \\ 0.35 \end{bmatrix}$$

$$+ \begin{bmatrix} 1.0 & 0.0 & 1.0 & 0.0 \\ 0.82 & 0.0 & 0.75 & 0.0 \\ 0.0 & 1.0 & 0.0 & 1.0 \\ 0.0 & 0.73 & 0.0 & 0.66 \end{bmatrix} \begin{bmatrix} v_{1ij} \\ v_{2ij} \\ v_{b1j} \\ v_{b2j} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \qquad (4.1.1)$$

The values corresponding to the vector π in (2.3.2) have determined by using the formula as shown in Appendix 1. The necessary values for $\alpha's$ and $\beta's$ are :

$\beta_1 = 0.20, \beta_2 = 0.10, \beta_3 = 0.25, \beta_4 = 0.30, \beta_{b1} = 0.25, \beta_{b2} = 0.30, \alpha_{w1} = 0.30, \alpha_{b1} = 0.20.$

The values for $z_{2ij}, w_j, v_{1ij}, v_{2ij}, v_{b1j}, v_{b2j}, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ are generated by the following:

(1) first we generate 5000 "z" variable from a standard normal distribution. Then if the value is bigger than 0.0 we assign 1 to it, while the value is less than 0.0, we assign 0 to it,

(2) do the same for "w" variable for 500 groups,

(3) generate 500 between level random vectors from the population VC (variance covariance) matrix, T_{η_b},

(4) generate 5000 within level random vectors from the population VC matrix, T_{η},

(5) generate 5000 measurement error vectors from the population VC matrix, Σ,

(6) then we use the equation (4.1.1) to obtain a balanced raw data.

The IMSL FORTRAN library contains the necessary several subroutines. The dimension of the observed variables is four (r=4). The dimension of the latent variables is two (p=2). And then we create an unbalanced data (4890 data points) by randomly deleting 4 for each of 10 groups, deleting 3 for each another 10 groups, and deleting 2 for each of another 10 groups, and finally delete 1 for each of 20 groups.

Table 4.2 Descriptive Statistics for Samples

(1) Balanced Data

TOTAL SAMPLE SIZE = 5000

	MEAN	ST.DEV	SKEWNESS	KURTOSIS	MIN	FREQ.	MAX	FREQ.
Y1	-0.291	7.735	-0.081	-0.010	-33.444	1	27.565	1
Y2	-0.187	6.477	-0.070	-0.005	-23.317	1	23.612	1
Y3	-0.145	7.913	-0.003	-0.125	-27.877	1	25.107	1
Y4	-0.097	6.055	-0.021	-0.085	-20.807	1	24.764	1

SAMPLE COVARIANCE MATRIX

Y1	59.824			
Y2	39.570	41.951		
Y3	16.487	12.887	62.620	
Y4	11.125	8.927	35.242	36.667

(2) Unbalanced Data

TOTAL SAMPLE SIZE = 4890

	MEAN	ST. DEV.	SKEWNESS	KURTOSIS	MIN	FREQ.	MAX	FREQ.
Y1	-0.285	7.805	-0.075	0.045	-29.918	1	28.897	1
Y2	-0.187	6.533	-0.092	-0.018	-22.221	1	24.147	1
Y3	-0.167	7.957	-0.010	-0.029	-27.866	1	26.820	1
Y4	-0.100	6.041	-0.017	-0.137	-21.051	1	19.738	1

ESTIMATED COVARIANCE MATRIX

Y1	60.922			
Y2	40.477	42.686		
Y3	16.370	13.360	63.309	
Y4	10.376	8.746	35.342	36.492

4.2 Results of the Analysis

The output of the Table 4.3 and 4.4 are the result of fitting balanced data and

unbalanced data to the same model. The focus of investigation is in discovering and testing the estimates of parameters are close to the predetermined population parameters within some what sampling error.

Table 4.3 Results of Analysis for Balanced Data

	Population Parameters	Starting Values	Estimated Parameters
λ_{w1}	0.82	0.582	0.857253
λ_{w2}	0.73	0.573	0.720532
λ_{b1}	0.75	0.575	0.738310
λ_{b2}	0.66	0.566	0.652453
$t_{\eta_{11}}$	30.00	20.000	29.006845
$t_{\eta_{12}}$	9.00	7.000	9.432951
$t_{\eta_{22}}$	32.70	20.000	34.328632
$t_{\eta_{b11}}$	20.00	10.000	18.407147
$t_{\eta_{b12}}$	4.00	3.000	3.528167
$t_{\eta_{b22}}$	20.80	10.000	19.754972
σ_1^2	10.00	8.000	10.851423
σ_2^2	12.00	8.000	11.2542916
σ_3^2	14.00	10.000	13.7741274
σ_4^2	16.00	10.000	16.1793495
α_{w1}	0.30	0.350	0.3251942
α_{b1}	0.20	0.300	0.1916734

Table 4.3.1 Conditional expectations of regression coefficients

β_1	0.20	0.100	0.3550013
β_2	0.10	0.005	0.1492152
β_3	0.25	0.100	0.6579540
β_4	0.30	0.200	0.1218303
β_{b1}	0.25	0.100	0.4485403
β_{b2}	0.30	0.100	0.2161860

Table 4.4 Results of Analysis for Unbalanced Data

	Population Parameters	Starting Values	Estimated Parameters
λ_{w1}	0.82	0.582	0.855920
λ_{w2}	0.73	0.573	0.719029
λ_{b1}	0.75	0.575	0.736739
λ_{b2}	0.66	0.566	0.651504
$t_{\eta_{11}}$	30.00	20.000	29.563270
$t_{\eta_{12}}$	9.00	7.000	9.569872
$t_{\eta_{22}}$	32.70	20.000	34.697681
$t_{\eta_{b11}}$	20.00	10.000	18.153446
$t_{\eta_{b12}}$	4.00	3.000	3.465942
$t_{\eta_{b22}}$	20.80	10.000	19.38584
σ_1^2	10.00	8.000	10.87648
σ_2^2	12.00	8.000	11.29153
σ_3^2	14.00	10.000	14.04039
σ_4^2	16.00	10.000	16.18727
α_{w1}	0.30	0.200	0.32372
α_{b1}	0.20	0.150	0.19092

Table 4.4.1 Conditional expectations of regression coefficients

β_1	0.20	0.150	0.29023
β_2	0.10	0.500	0.13452
β_3	0.25	0.150	0.74056
β_4	0.30	0.200	0.10906
β_{b1}	0.25	0.150	0.47312
β_{b2}	0.30	0.200	0.27258

Table 4.5 The Values of the Observed Log-likelihood at Convergence

Balanced Data	
Iteration 1353	-2902.3472413
Iteration 1354	-2902.3472410
Iteration 1356	-2902.3472408
Iteration 1357	-2902.3472405
Iteration 1358	-2902.3472402
Iteration 1359	-2902.3472400
Iteration 1360	-2902.3472397

Table 4.6 The Values of the Observed Log-likelihood at Convergence

Unbalanced Data	
Iteration 1321	-2768.5733562
Iteration 1322	-2768.5733550
Iteration 1323	-2768.5733537
Iteration 1324	-2768.5733522
Iteration 1325	-2768.5733511
Iteration 1326	-2768.5733508

In discussing the results reported in Table 4.3 and 4.4 , we say that the EM algorithm recovered the population parameters values well. The criterion used for convergence of the observed log-likelihood is that log-likelihood is smaller than 0.1^6($\delta \leq 10^{-6}$). In the 486 IBM PC with the speed of 66 mhrz for the convergence it took about 45 hours. For the unbalanced case the hours spent are about 48 hours. The Table 4.5 and Table 4.6 show the list of log-likelihood values at the neighborhood of convergence. As the Table 4.4 shows the number of iterations is very large and the spent hours is very long. The slowness of convergence of the EM algorithm is the repeatedly criticized property of the algorithm. This seems to be caused by the fact that missing information (Meng and Rubin, 1991) is relatively large in the multilevel structural equation model. To obtain estimates of $\alpha' s$ and $\beta' s$ we assume that the matrices, Δ and Δ_b are diagonal matrices. Then as shown in Appendix 1, we obtain the estimates of structural parameters. When we use extremely poor starting values, the estimates are not similar. But if we use moderately poor starting values, then the results are very similar (almost same) across various sets of starting values.

After estimating parameters, the likelihood ratio tests are available to test more and less complex models. It is known that the statistic, $-2\ln(\frac{L_1}{L_2})$, has an asymptotic chi-square distribution, where L_1 is the maximum likelihood value of a less complex model and L_2 is the value of a more complex model. The degrees of freedom of chi-square statistic is the difference of the number of parameters to be estimated in each model. In our educational example, the model for L_1 may be constrained as follows.

$$(1)\ \mathrm{T}_{\eta} = \mathrm{T}_{n_b}, \qquad (2)\ \Lambda_w = \Lambda_b$$

As the Table 4.5 shows the number of parameters of the simpler model is 10 while the

complex model has 16 parameters to be estimated. The degrees of freedom of chi-square is 6. The deviance between the two values of -2ln(Likelihood) is 662.24. This value is large enough to reject the adequacy of the simpler model. In this likelihood ratio test we do not know which of the restrictions is not adequate. Thus for each of the restrictions we may test the adequacy of the simpler model rather than complex model.

Table 4.5 Results of Analysis for Unbalanced Data for Restricted Model

	Population Parameters	Starting Values	Estimated Parameters
λ_{w1}	0.82	0.75	0.841037
λ_{w2}	0.73	0.70	0.692301
λ_{b1}	0.75	0.70	0.720219
λ_{b2}	0.66	0.50	0.837227
$t_{\eta_{11}}$	30.00	15.00	32.647209
$t_{\eta_{12}}$	9.00	5.00	10.538582
$t_{\eta_{22}}$	30.00	15.00	36.943765
$t_{\eta_{b11}}$	20.00	15.00	16.459542
$t_{\eta_{b12}}$	4.00	5.00	7.358164
$t_{\eta_{b22}}$	20.00	15.00	25.980356
σ_1^2	10.00	8.00	9.236718
σ_2^2	12.00	8.00	12.450677
σ_3^2	14.00	8.00	12.120453
σ_4^2	16.00	8.00	13.863560
α_{w1}	0.30	0.200	0.323054
α_{b1}	0.20	0.150	0.447046
The value of log-likelihood			-3104.68428

Table 4.5.1 Conditional expectations of regression coefficients

β_1	0.20	0.150	0.92709
β_2	0.10	0.500	0.17796
β_3	0.25	0.150	2.05197
β_4	0.30	0.200	1.17230
β_{b1}	0.25	0.150	2.64902
β_{b2}	0.30	0.200	1.48537

The multilevel structural equation model postulates that the causal relationship on the within level and between level are different because the explanatory variable is different, in our example, gender on the within level, teaching methods on the classroom level. We may calculate the root mean square residual (Joreskog and Sorbom, 1993) for an overall goodness of fit measure. In the literature of LISREL, we found the root mean squared residual can be used to compare the fit of two different models for the same data.

$$RMQR = \left[\sum\sum(s_{ij} - \hat{\sigma}_{ij})^2 / (r^2 + r)\right]^{\frac{1}{2}} \qquad (4.2.1)$$

where s_{ij} is the i th row and j-th column variance-covariance element of the sample total variance-covariance matrix. And $\hat{\sigma}_{ij}$ is the corresponding element of the model predicted total variance-covariance matrix, $\hat{\Sigma}$. RMQR is a measure of the average of the fitted residuals. The model predicted total variance covariance is given by $\hat{\Sigma} = \hat{\Sigma}_w + \hat{\Sigma}_b$. then we obtain, RMQR = 0.73850 for complex model, RMQR= 5.98066 for simple model. Judged by these index we may conclude that the fit of the simple model to the hierarchical data is not as adequate as in the complex model. This conclusion was expected because the data were generated in a hierarchical fashion,

and the restrictions makes the model a single level structural equation model. Now we make inferences about $\pi's$ in balanced data case. Specifically, from the posterior variance estimates of $\pi's$ we obtained the standard errors for $\pi's$ shown in Table 4.6. and 4.7. For the balanced data the t-statistics are: [1.247 0.8373 0.5566 0.8806 1.09108 0.6998]. Thus the expectation of the "attitude" latent variable for the girl students in classrooms of which instruction style is expository is significantly different from zero. And expectation of the "ACH" latent variable for the girl students in classrooms of which instruction style is expository is not significantly different from zero. The total effect of "gender" on the "ACH" latent variable is not significantly different from zero. The "gender gap" on the "attitude" latent variable is not significantly different from zero. The total effect of discovery teaching style on "ACH" latent variable is not significantly different zero at the between group level. And the effect of discovery teaching style on the "attitude" latent variable at the between group level is significantly different from zero.

Table 4.6 Estimates of $\pi's$ for Two Sets of Data

		Balanced Data		Unbalanced Data	
	Population Parameters	Estimated Parameters	Standard Error	Estimated Parameters	Standard Error
π_1	0.20	0.3550	0.28560	0.2902	0.2137
π_2	0.10	0.1492	0.17818	0.1345	0.1845
π_3	0.31	0.7734	0.30602	0.8345	0.2827
π_4	0.33	0.1703	0.19344	0.1526	0.2216
π_{b1}	0.25	0.4485	0.41105	0.5316	0.4363
π_{b2}	0.35	0.3021	0.43176	0.3629	0.5262

CHAPTER 5

CONCLUSION

Research in the field of education provides various challenges. For example, the random assignment of students to a set of conditions is not realistic in most cases. Even in the experimental setting the outcomes will have a positive intracluster correlations due to the fact that (1) students do not receive their instruction individually but in groups, (2) interactions exist between treatments and students (Lumsdaine, 1963). This situation often makes the application of the linear structural equation models (Joreskog, 1977) to the real data inappropriate.

As Cronbach (1976) pointed out, many studies in the field of education have produced inappropriate analysis, especially in the field of evaluation studies because they failed to recognize the nature of hierarchical data. The difficulty of analyzing data arising from two levels is in assessing the nature of intervariable relationships at both levels simultaneously. During the last two decades researchers have developed multilevel structural equation models for hierarchical data. Previous work on the multilevel structural equation models involved a minimization fitting function or /and balanced sampling design. And most ultilized standard software. These minimization fitting function approaches have made substantial methodological advances. They require classifying groups into subsets of groups having equal sample size.

This dissertation has shown how multilevel structural equation models can be formulated for hierarchical data and how they can be analyzed by using empirical Bayes with the EM algorithm. The model equations are linear at each level, the direct

direct connection between variables is typically specified by the value of the coefficients as in the LISREL (Joreskog, 1977) or EQS (Bentler, 1983) models.

The major outputs of this thesis are four:

1. I presented the general multilevel structural equation model in the mode of hierarchical empirical Bayes modeling.
2. I developed the empirical Bayes estimation procedure via the EM algorithm to find maximum likelihood estimates of the model.
3. A computer program for numerical analysis of hierarchical data for structural equation models was developed in Gauss.
4. The accuracy of the computing algorithm has been tested across sampling designs and models.

5.1 Summary, Implications and Conclusions

I now summarize the major points that emerged in each chapter and discuss their implications for fitting multilevel structural equation model to hierarchical data. In chapter one, the problems confronting single level and multilevel structural equation models were identified. These problems are old ones. Traditional single level structural equation models do not incorporate the random errors from the multilevel structure. Therefore reserachers face the dillema "What should be our unit of analysis?" That is, they have to choose individual level analysis or group level analysis. Ignoring the inherent hierarchical structure in our data sets results in the confounding of group level effects with individual level effects. Of course, the

individual level analysis violates the independence assumption and results in over estimation of precision. In program evaluation studies, for example, curriculuum evaluation, we are interested in the interaction effects between the treatments and the individuals across individual level background information and group level variables.

In chapter two, the multilevel structural equation model and a few basic model assumptions are presented and then translated into the general mixed model (hierarchical model of linear equations) on both the cluster level and individual level. In particular, the model explicitly utilizes the socio-demographic information on both group and individual levels. We also provided an example for model specification. In chapter three, empirical Bayes estimation procedure via the EM algorithm was generalized to the multilevel structural equation model. In the "hierarchical prior distribution" specification we adopted the MLR (Dempster, Rubin and Tsutakawa, 1981) approach. We also presented the E and M steps for ML estimates.

Much of Chapter four assessed the accuracy of the EM algorithm for an artificial model for different sampling designs. The results showed that the EM algorithm for the multilevel structural equation model is quite accurate for the data generated. The results were drawn from the simulation under the balanced and unbalanced sampling design.

Ignoring the second level effects can yield a misleading interpretation of the structure of the causal map among latent constructs under the study. The presented methodology allows the simultaneous estimation of the prarmeters in the unbalanced multilevel structural equation models.

The primary advantage of the EM method for multilevel structural equation models lies in the three key facts: (1) it does not require the calculation of the 2nd-order partial derivatives of the maximum likelihood function. Even though the models studied by several researchers are different in terms of balanced or unbalanced sampling designs, the calculation of partial derivatives are essential for the methods proposed by Schmidt and Wisenbaker (1986), MacDonald and Goldstein (1989), Lee and Poon(1992), Raudenbush (press). (2) it does not require classifying level-2 units into subsets having equal sample size. However it has the shortcoming of slowness of convergence. Another shortcoming of the EM algorithm is that it does not provide the standard errors for the estimates of parameters.

When we apply the method to real-world problems we need further elaboration of the models that carefully takes into account the special features of a certain subject matter. Typical applications of structural equation models involve (1) the development of a prior model, representing hypothetical causal associations among a set of latent variables and manifest variables. (2) fitting the prior model to sample data, (3) the evaluation of the solution in terms of its parameters estimates and goodness of fit, (4) the modification of the model so as to improve its parsimony and its fit to the data. This last step has been known "a specification search" or "respecification". During such a search the researcher alters the model specification in search of substantively meaningful model that fits the data well. The structural misspecifications are involved the specification of the elements of matrix A and A_b. It is also closely related to the identifiable model.

The purpose of latent variables model is to improve the accuracy and validity of inferences from empirical data. In order to accomplish this goal, several assumptions about the structure of the data and the meaning of the associations between variables must be made. Ideally, each of these assumptions will be based either on special features of subject-matter application area or on the knowledge, derived from past empirical evidence.

5.2 Future Work

One of the potential fields to which the mutilevel structural equation model is extended is the model where the slope for the exogenous variables varies randomly across groups. Note that there are numerous settings in which multilevel structural equation models consisting of random slopes for exgenous variables are needed in order to represent adequately the variance-covariance structure of the data. Thus, for example, the gender gap in the SAT mathematics test scores can be explained by the group-level characteristics.

As is well known, the EM algorithm is simple to implement and numerically stable, but is slow. Recently Jamshidian and Jennrich (1993) developed a conjugate gradient scheme for accelerating the speed of the EM algorithm. In their AEM (accelerating EM) algorithm the evaluation of the gradient of the likelihood function is essential. When the number of parameters is moderate the implementation of the AEM algorithm seems not so burdensome. To obtain the standard error of the

maximum likelihood estimates one may apply the SEM (supplemented EM algorithm) algorithm developed by Meng and Rubin (1991).

The current approach to drawing inferences concerning variance components is based on large sample theory for ML estimators. When the number of groups is small the normal approximation will be invalid. For that case, one may apply the Data Augmentation (Tanner and Wong, 1987) approach.

Future models can be expanded to larger and more complicated models. Also robustness of the model remain to be studied.

APPENDICES

Appendix 1

For finding the values of $\beta's$ and $\alpha's$, we have the elementwise algebraic relations :

$$\pi_1 = \beta_1 \tag{A.1.1}$$

$$\pi_2 = \alpha_{21}\beta_1 + \beta_2 \tag{A.1.2}$$

Then we can derive estimates of the structural coefficients from this first within-group algebraic relations. Especially the identification of $\alpha's$ is carried out on the base of the following assumption (A.1.3) and the second within-group algebraic relationship (A.1.4).

$$\text{var}(u_{ij}) = \Delta = diag(\delta_p, p = 1,\ldots,P) \tag{A.1.3}$$

$$(I - A)^{-1}\Delta(I - A)^{-1^T} = \text{var}(v_{ij}) = \text{T}_\eta \tag{A.1.4}$$

The elementwise relations of (A.1.4) are:

$$\delta_{11} = \tau_{\eta_{11}}$$

$$\delta_{11}\alpha_{21} = \tau_{\eta_{12}} \tag{A.1.5}$$

After finding 's from the above algebraic relationship (A.1.5), we find the $\beta's$ from the first within-group algebraic relationship (A.1.2).

Similarly the first between-group algebraic relations are:

$$(I - A_b)^{-1} B_b = \Pi_b \,, \tag{A.1.6}$$

The corresponding elementwise relationships are:

$$\pi_{b011} = \beta_{b11}$$
$$\pi_{b021} = \alpha_{b21}\beta_{b11} + \beta_{b21} \tag{A.1.7}$$

In the same fashion, the identification of $\alpha_b's$ is carried out on the base of the following assumption (A.1.8) and the second within-group algebraic relationship (A.1.9).

$$\text{var}(u_{bj}) = \Delta_b = diag(\delta_{bp}, p = 1,\ldots,P) \tag{A.1.8}$$

$$(I - A_b)^{-1} \Delta_b (I - A_b)^{-1^T} = \text{var}(v_{bj}) = \text{T}_{\eta_b} \tag{A.1.9}$$

The elementwise relations of (A.1.9) are:

$$\delta_{b11} = \tau_{\eta_{b11}}$$
$$\delta_{b11}\alpha_{b21} = \tau_{\eta_{b12}} \tag{A.1.10}$$

After finding $\alpha_b's$ from the above algebraic relationship (A.1.10), we find the $\beta_b's$ from the first between-group algebraic relationship (A.1.7).

Appendix 2

For the computational convenience of the necessary posterior expectations and dispersion matrices for the random vector the equation (3.1.1) without subscript becomes :

$$y = A_1\pi + A_2\theta_2 + A_3\theta_3 + \varepsilon$$

$$\varepsilon \sim N(0, I_N \otimes \Sigma), \qquad N = \sum n_K$$

$$\pi \sim N(0, \Gamma)$$

$$\theta_2 \sim N(0, \Omega_\pi),$$

$$\theta_3 \sim N(0, \Omega_\eta),$$

let $y = A\theta + r$

where $A = [A_1 | A_2 | A_3]$

$$\theta = \begin{bmatrix} \pi \\ \theta_2 \\ \theta_3 \end{bmatrix}, \pi = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \cdot \\ \cdot \\ \cdot \\ \pi_t \end{bmatrix}, \theta_3 = \begin{bmatrix} \theta_{31} \\ \theta_{32} \\ \cdot \\ \cdot \\ \cdot \\ \theta_{3k} \end{bmatrix},$$

$$\theta_2 = [\theta_{211}\theta_{212},\ldots,\theta_{21n_1},\ldots,\theta_{2K1}\theta_{2K2},\ldots,\theta_{2Kn_K}]'$$

t : the dimension of π

$$\text{var}(\theta) = \begin{bmatrix} \Gamma & 0 & 0 \\ 0 & \Omega_\eta & 0 \\ 0 & 0 & \Omega_{\eta_b} \end{bmatrix} \qquad \begin{array}{l} \Omega_\eta = subdiag(\mathrm{T}_\eta) \\ \Omega_{\eta_b} = subdiag(\mathrm{T}_{\eta_b}) \end{array}$$

By using the results of the standard multivariate normal distribution (Searle, 1971), one obtain the following joint distribution of the multilevel structural equation model :

$$\begin{bmatrix} y \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_0 & A_1\Gamma & A_2\Omega_\eta & A_3\Omega_{\eta_b} \\ \Gamma A_1' & \Gamma & 0 & 0 \\ \Omega_\eta A_2' & 0 & \Omega_\eta & 0 \\ \Omega_{\eta_b} A_3' & 0 & 0 & \Omega_{\eta_b} \end{bmatrix}\right]$$

where $\Sigma_0 = A_1\Gamma A_1' + A_2\Omega_\eta A_2' + A_3\Omega_{\eta_b} A_3' + \mathrm{I}_N \otimes \Sigma$

The posterior dispersion matrix D_θ^* of θ is translated from $A'\Psi^{-1}A + \Omega^{-1}]^{-1}$ into the context of the multilevel structural equation model. The posterior dispersion matrix can be written as :

$$D_\theta^* = [\mathrm{A}'\Psi^{-1}\mathrm{A} + \Omega^{-1}] \Leftrightarrow \begin{bmatrix} \mathrm{A}_1'\Psi^{-1}\mathrm{A}_1 + \Gamma^{-1} & \mathrm{A}_1'\Psi^{-1}\mathrm{A}_2 & \mathrm{A}_1'\Psi^{-1}A_3 \\ \mathrm{A}_2'\Psi^{-1}\mathrm{A}_1 & \mathrm{A}_2'\Psi^{-1}\mathrm{A}_2 + \Omega_\eta^{-1} & \mathrm{A}_2'\Psi^{-1}A_3 \\ A_3'\Psi^{-1}\mathrm{A}_1 & A_3'\Psi^{-1}\mathrm{A}_2 & A_3'\Psi^{-1}A_3 + \Omega_{\eta_b}^{-1} \end{bmatrix}^{-1}$$

For the convenience of applying the inversion formula of the partitioned matrix (Graybill, 1983), one can rewrite the matrix D_θ^*:

$$\begin{bmatrix} \mathrm{B}_1 & \mathrm{B}_2 & L_1 \\ \mathrm{B}_2' & \mathrm{B}_3 & L_2 \\ L_1' & L_2' & U \end{bmatrix}^{-1} = \begin{bmatrix} Q^{-1} & -(GQ^{-1})' \\ -(GQ^{-1}) & U^{-1} + GQ^{-1}G' \end{bmatrix}$$

where $$Q = B_+ - LU^{-1}L' , \quad G = U^{-1}L'$$

where $$B_+ = \begin{bmatrix} B_1 & B_2 \\ B_2' & B_3 \end{bmatrix}, \qquad L' = [L_1' L_2']$$

$$Q^{-1} = \begin{bmatrix} Q_1 & Q_2 \\ Q_2' & Q_3 \end{bmatrix}^{-1} = \begin{bmatrix} R_1 & R_2 \\ R_2' & R_3 \end{bmatrix}$$

$$Q_1 = A_1'\Psi^{-1}A_1 - A_1'\Psi^{-1}A_3(A_3'\Psi^{-1}A_3 + \Omega_{\eta_b})^{-1}A_3'\Psi^{-1}A_1$$

$$Q_2 = A_1'\Psi^{-1}A_2 - A_1'\Psi^{-1}A_3(A_3'\Psi^{-1}A_3 + \Omega_{\eta_b})^{-1}A_3'\Psi^{-1}A_2$$

$$Q_3 = \mathrm{A}_2'\Psi^{-1}\mathrm{A}_2 + \Omega_\eta^{-1} - \mathrm{A}_2'\Psi^{-1}A_3(A_3'\Psi^{-1}A_3 + \Omega_{\eta_b}^{-1})^{-1}A_3'\Psi^{-1}\mathrm{A}_2$$

$$D_{\theta}^{*}=\begin{bmatrix} R_1 & R_2 & -R_1L_1U^{-1}-R_2L_2U^{-1} \\ R_2' & R_3 & -R_2'L_1U^{-1}-R_3L_2U^{-1} \\ \textit{symmetric} & & U^{-1}+U^{-1}(L_1'R_1L+L_2'R_2'L_1+L_1'R_2L_2+L_2'R_3L_2)U^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{12}' & V_{22} & V_{23} \\ V_{13}' & V_{23}' & V_{33} \end{bmatrix} = \begin{bmatrix} D_{\pi}^{*} & C_{\pi\theta} & C_{\pi\theta_b} \\ C_{\pi\theta}^{T} & D_{\theta}^{*} & C_{\theta\theta_b} \\ C_{\pi\theta_b}^{T} & C_{\theta\theta_b}^{T} & D_{\theta_b}^{*} \end{bmatrix}$$

Thus in the E-step the needed conditional density of θ_{3ki} , θ_{2k} , ε_{2k} given the data and the parameters, $\pi_0, \pi_{b0}, \mathrm{T}_{\pi}, \mathrm{T}_{\eta}, \mathrm{T}_{\eta_b}, \Sigma_r$ have the following locations and dispersion matrices:

$$\theta_1^{*} = V_{11}[(\mathrm{A}_2' - Q_2Q_3^{-1}\mathrm{A}_2')\Psi^{-1}(\mathrm{I} - A_3U^{-1}A_3'\Psi^{-1})](y)$$

$$\theta_2^{*} = (V_{22}\mathrm{A}_2' - Q_3^{-1}Q_2V_{11}\mathrm{A}_1')\Psi^{-1}[\mathrm{I} - A_3(A_3'\Psi^{-1}A_3 + \mathrm{T}_{\eta_b}^{-1})^{-1}A_3'\Psi^{-1}](y)$$

$$\theta_3^{*} = U^{-1}A_3'\Psi^{-1}[y - (\mathrm{A}_1\theta_1^{*} + \mathrm{A}_2\theta_2^{*})]$$

$$_{11} = [Q_1 - Q_2Q_3^{-1}Q_2']^{-1}$$

$$_{12} = -V_{11}Q_2Q_3^{-1}$$

$$_{13} = -[V_{11}A_1'\Psi^{-1}A_3 + V_{12}A_2'\Psi^{-1}A_3]U^{-1}$$

$$_{22} = Q_3^{-1} + Q_3^{-1}Q_2'V_{11}Q_2Q_3^{-1}$$

$$_{23} = -[V_{12}'A_1'\Psi^{-1}A_3 + V_{22}A_2'\Psi^{-1}A_3]U^{-1}$$

$$_{33} = U^{-1} - [V_{13}'A_1'\Psi^{-1}A_3 + V_{23}'A_2'\Psi^{-1}A_3]U^{-1}$$

Appendix 3

We show the explicit derivation of $Q(\Theta^{(i)},\Theta^{(i-1)})$ defined in chapter 3.

By definition from Dempster, Laird and Rubin (1977),

$$Q(\Theta^{(i)},\Theta^{(i-1)}) = E(\log[f(c;\Theta^{(i)})]|Y = y;\Theta^{(i-1)}).$$

By applying the theorem on the expectation of the quadratic function we have:

$$E[\theta_{ij}^T \mathrm{T}_\eta^{-1} \theta_{ij}|Y = y;\Theta^{(i-1)}] = tr[\mathrm{T}_\eta^{-1} E(\theta_{ij}\theta_{ij}^T|Y = y;\Theta^{(i-1)})]$$

$$E[\theta_{bj}^T \mathrm{T}_{\eta_b}^{-1} \theta_{bj}|Y = y;\Theta^{(i-1)}] = tr[\mathrm{T}_{\eta_b}^{-1} E(\theta_{bj}\theta_{bj}^T|Y = y;\Theta^{(i-1)})]$$

$$E[\varepsilon_{ij}^T \Sigma^{-1} \varepsilon_{ij}|Y = y;\Theta^{(i-1)}] = tr[\Sigma^{-1} E(\varepsilon_{ij}\varepsilon_{ij}^T|Y = y;\Theta^{(i-1)})]$$

After some algebraic calculations we have:

$$\begin{aligned}
Q(\Theta^{(i)},\Theta^{(i-1)}) \;=\; & (\text{-Nr/2})\ln 2\pi + (\text{-N/2})\ln|\Sigma| \\
& + (-Np/2)\ln(2\pi) + (-N/2)\ln|\mathrm{T}_\eta| + (-Js/2)\ln(2\pi) + (-J/2)\ln|\mathrm{T}_{\eta_b}| \\
& -\frac{1}{2}\sum\sum tr[\Sigma^{-1}((\Lambda_0\tilde{Z}_{ij})D_\pi^*(\Lambda_0\tilde{Z}_{ij})^T + (\Lambda_0 D_{\varpi_{ij}}^* \Lambda_0^T) \\
& + (\,(\Lambda_0\tilde{Z}_{ij}C_{\pi\varpi_{ij}}^* \Lambda_0^T) + (\Lambda_0 C_{\pi\varpi_{ij}}^{*T} \tilde{Z}_{ij}^T \Lambda_0^T))] \\
& -\frac{1}{2}\sum\sum (y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi^* - \Lambda_0\varpi_{ij}^*)^T \Sigma^{-1}(y_{ij} - \Lambda_0\tilde{Z}_{ij}\pi^* - \Lambda_0\varpi_{ij}^*) \\
& -\frac{1}{2}\sum\sum tr[\mathrm{T}_\eta^{-1} E(\theta_{ij}\theta_{ij}^T|Y = y,\Theta^{(i-1)})] - \frac{1}{2}\sum tr[\mathrm{T}_{\eta_b}^{-1} E(\theta_{bj}\theta_{bj}^T|Y = y,\Theta^{(i-1)})] \qquad (3.2.1)
\end{aligned}$$

Appendix 4

In this appendix we provide the formula for the element of $\hat{\Sigma}$ is :

$$\hat{\sigma}_m^2 = \frac{1}{N}\sum\sum[2\lambda_{wm}D^{*}_{\pi ff} + 2\lambda_{wm}z_{2ij}D^{*T}_{\pi fg} + 2\lambda_{bm}w_j D^{*T}_{\pi fg} + 2\lambda_{1m}z_{2ij}D^{*}_{\pi fg} + 2\lambda_{wm}z_{2ij}^2 D^{*}_{\pi ff} +$$

$$2\lambda_{bm}w_j D^{*T}_{\pi fg}z_{2ij} + 2\lambda_{wm}D^{*}_{\varpi cc} + 2\lambda_{bm}D^{*}_{\varpi cd} + 2\{2\lambda_{wm}C^{*}_{\pi\varpi fc} + 2\lambda_{w1}z_{2ij}C^{*}_{\pi\varpi fc} + \lambda_{bm}w_j C^{*}_{\pi\varpi fc} +$$

$$\lambda_{bm}C^{*}_{\pi\varpi fc} + \lambda_{bm}z_{2ij}C^{*}_{\pi\varpi fc}\} + \{y_{mij} - \lambda_{wm}(\pi^{*}_f + z_{2ij}\pi^{*}_f) - \lambda_{wm}v^{*}_{cij} - \lambda_{bm}w_j\pi^{*}_{bf} - \lambda_{bm}v^{*}_{bcj}\}^2$$

where m=1,..,4.

Appendix 5

In this appendix we present the derivations of the elements of the loading matrices. First by expanding the 8-th term in the right of the equation (3.2.1) we have the following five terms:

$$(1)(-2.0)y_{ij}^T\Sigma^{-1}\Lambda_0\tilde{Z}_{ij}\pi^*$$

$$(2)(-2.0)y_{ij}^T\Sigma^{-1}\Lambda_0\varpi_{ij}^*$$

$$(3)(\Lambda_0\tilde{Z}_{ij}\pi^*)^T\Sigma^{-1}\Lambda_0\tilde{Z}_{ij}\pi^*$$

$$(4)(\Lambda_0\tilde{Z}_{ij}\pi^*)^T\Sigma^{-1}\Lambda_0\varpi_{ij}^*$$

$$(5)(\Lambda_0\varpi_{ij}^*)^T\Sigma^{-1}\Lambda_0\varpi_{ij}^*$$

For each term we take the derivatives with respect to Λ_0. Note that in taking the derivatives we regard Λ_0 as a full matrix. Then we have:

$$(1)(-2.0)\Sigma^{-1}y_{ij}\pi^{*T}\tilde{Z}_{ij}^T$$

$$(2)(-2.0)\Sigma^{-1}y_{ij}\varpi_{ij}^{*T}$$

$$(3)[(2.0)\Sigma^{-1}\Lambda_0\tilde{Z}_{ij}\pi^*(\pi^{*T}\tilde{Z}_{ij}^T)]$$

$$(4)(\Sigma^{-1}\Lambda_0\varpi_{ij}^*\pi^{*T}\tilde{Z}_{ij}^T+\Sigma^{-1}\Lambda_0\tilde{Z}_{ij}\pi^*\varpi_{ij}^{*T})$$

$$(5)(\Sigma^{-1}\Lambda_0\varpi_{ij}^*\varpi_{ij}^{*T})$$

And we take the derivatives of 7-th term in right side of the equation (3.2.1) with respect to Λ_0. Then we obtain:

$$\frac{\partial Q}{\partial \Lambda_0} = \frac{\partial \Theta}{\partial \Lambda_0} = 2\,tr\left[\Sigma^{-1}\{\Lambda_0 \tilde{Z}_{ij} D_{\pi}^{*} \hat{Z}_{ij}^{T} + \Lambda_0 D_{\tilde{w}_j}^{*} + \Lambda_0 \hat{Z}_{ij} G_{\tilde{w}_j}^{*} + \Lambda_0 G_{\tilde{w}_j}^{*} \tilde{Z}_{ij}\}\right]$$

(6) $tr[\Sigma^{-1}\{\Lambda_0 \tilde{Z}_{ij} D_{\pi}^{*} (\Lambda_0 \tilde{Z}_{ij})^T + \Lambda_0 D_{w}^{*} \Lambda_0^T + \Lambda_0 \tilde{Z}_{ij} C_{\pi w}^{*} \Lambda_0^T + \Lambda_0 C_{w\pi}^{*} (\Lambda_0^T \tilde{Z}_{ij})^T\}]$

By arranging these six equations we obtain $\frac{\partial Q(\Theta^{(i)}, \Theta^{(i-1)})}{\partial \Lambda}$ as we write in chapter 3.

Let A= $\frac{\partial Q(\Theta^{(i)}, \Theta^{(i-1)})}{\partial \Lambda}$. $\left[A\right]_{\Lambda_0}$ is the matrix A with zeros inserted in the places of fixed parameters in Λ_0. Finally by the equation (3.2.9) we obtain the following ML estimators for each elements of Λ_0:

$$\hat{\lambda}_{2,1} = [-2\lambda_{2,3} w_j (D^{*}_{\pi 15} z_{1ij} + D^{*}_{\pi 25} z_{2ij} + C^{*}_{\pi \omega_{ij} 51}) - 2\lambda_{2,3} (D^{*}_{\omega_{ij} 13} + z_{1ij} C^{*}_{\pi \omega_{ij} 13} + z_{2ij} C^{*}_{\pi \omega_{ij} 23})$$
$$+2(y_{2ij} - 2\lambda_{2,3} w_j \pi^{*}_{b1} - \lambda_{2,3} \theta^{*}_{b1j})(z_{1ij} \pi^{*}_{1} + z_{2ij} \pi^{*}_{2} + \theta^{*}_{1ij})]$$
$$\times\left[2[D^{*}_{\pi 11} + 2z_{2ij} D^{*}_{\pi 12} + z^{2}_{2ij} D^{*}_{\pi 22} + D^{*}_{\omega_{ij} 11} + 2C^{*}_{\pi \omega_{ij} 11} + 2z_{2ij} C^{*}_{\pi \omega_{ij} 21} + (z_{1ij} \pi^{*}_{1} + z_{2ij} \pi^{*}_{2} + \theta^{*}_{1ij})^2]\right]^{-1}$$

$$\hat{\lambda}_{4,2} = [-2\lambda_{4,4} w_j (D^{*}_{\pi 36} z_{1ij} + D^{*}_{\pi 46} z_{2ij} + C^{*}_{\pi \omega_{ij} 62}) - 2\lambda_{4,4} (D^{*}_{\omega_{ij} 24} + z_{1ij} C^{*}_{\pi \omega_{ij} 34} + z_{2ij} C^{*}_{\pi \omega_{ij} 44})$$
$$+2(y_{4ij} - 2\lambda_{4,4} w_j \pi^{*}_{b2} - \lambda_{4,4} \theta^{*}_{b2j})(z_{1ij} \pi^{*}_{3} + z_{2ij} \pi^{*}_{4} + \theta^{*}_{2ij})]$$
$$\times\left[2[D^{*}_{\pi 33} + 2z_{2ij} D^{*}_{\pi 34} + z^{2}_{2ij} D^{*}_{\pi 44} + D^{*}_{\omega_{ij} 22} + 2C^{*}_{\pi \omega_{ij} 32} + 2z_{2ij} C^{*}_{\pi \omega_{ij} 42} + (z_{1ij} \pi^{*}_{3} + z_{2ij} \pi^{*}_{4} + \theta^{*}_{2ij})^2]\right]^{-1}$$

$$\hat{\lambda}_{2,3} = [-\lambda_{2,1} w_j (2D^{*}_{\pi 15} z_{1ij} + 2D^{*}_{\pi 25} z_{2ij} + C^{*}_{\pi \omega_{ij} 51}) - 2\lambda_{2,1} (D^{*}_{\omega_{ij} 13} + z_{1ij} C^{*}_{\pi \omega_{ij} 13} + z_{2ij} C^{*}_{\pi \omega_{ij} 23})$$

$$+2(w_j\pi^*_{b1}+\theta^*_{b1j})(\lambda_{2,1}(z_{1ij}\pi^*_1+z_{2ij}\pi^*_2+\theta^*_{1ij})+y_{2ij}\Big]$$

$$\times\Big[2[w_j^2D^*_{\pi 55}+D^*_{\varpi_{ij}33}+2w_jC^*_{\pi\varpi_{ij}53}+2(w_j\pi^*_{b1}+\theta^*_{b1j})^2]\Big]^{-1}$$

$$\hat{\lambda}_{4,4}=[-\lambda_{4,2}w_j(2D^*_{\pi 36}z_{1ij}+2D^*_{\pi 46}z_{2ij}+2C^*_{\pi\varpi_{ij}62})\ -2\lambda_{4,2}(D^*_{\varpi_{ij}24}+z_{1ij}C^*_{\pi\varpi_{ij}34}+z_{2ij}C^*_{\pi\varpi_{ij}44})$$

$$+2(w_j\pi^*_{b2}+\theta^*_{b2j})(y_{4ij}-\lambda_{4,2}(z_{1ij}\pi^*_3+z_{2ij}\pi^*_4+\theta^*_{2ij}))]$$

$$\times\Big[2[w_j^2D^*_{\pi 66}+D^*_{\varpi_{ij}66}+2w_jC^*_{\pi\varpi_{ij}64}+2(w_j\pi^*_{b2}+\theta^*_{b2j})^2]\Big]^{-1}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

Aitkin, M., Anderson, D., & Hinde, J. (1981). Statistical modeling of data on teaching styles. Journal of the Royal Statistical Society, Series A, 144(4), 419-461.

Aitkin, M., & Longford, N (1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society, Series A, 149, 1, 1-43.

Bentler, P. M (1983). Simultaneous equations as moment structure models: With an introduction to latent variable models. Journal of Econometrics, 22, 13-42.

Bentler, P. M (1989). EQS: Structural Equation Program Manual, Los Angeles, CA: BMDP Statistical Software, Inc.

Bock, R.D. (1960). Componets of variance analysis as a structural and discriminant analysis for psychological tests. British Journal of Statistical Psychology, 13, 151-163.

Bock, R.D. (1989). Addendum-measurement of variation: A two-stage model. In R.D. Bock (Ed.), Multilevel analysis of educational data (pp. 319-342). New York: Academic.

Bock, R. D., and Bargmann, R. E., (1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

Bryk, A.S., and Raudenbush, S.W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 1, 147-158.

Bryk, A.S., Raudenbush, S.W. (1992). Hierarchical Linear Models in Social and Behavioral Research: Applications and data analysis Methods. Newbury Park, CA: Sage Publications.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D.C. Berliner(Ed.), Review of research in education, vol. 8. Washington, DC: American Educational Research Association.

Cronbach, L. J. (1976). Research on classrooms and schools: Formulation of questions, design and analysis. Occasional paper of the Stanford Evaluation Consortium, School of Education, Stanford University.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., MxKnight, C.C., & Cooney, T. J. (1985). Second international mathematics study: Summary report for the United States. Champaign, IL: Stipes.

deFinetti , B. (1937). Foresight: its logical laws, its subjective sources. Annals de

l'Institute Henri Poincare. Reprinted (in translation) in Kyburg and Smokler (1964).

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, series B, 34, 1-8.

Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in covariance components models. Journal of the American Statistical Assoc., 76, 341-353.

Foitu, R.,P (1989). A comparison of the EM and Data Augmentation algorithms on simulated small sample hierarchical data from research on education. Unpublished doctoral dissertation, Michigan State University.

Galton, F. (1883). Inquiries into human faculty and its development. London: Macmillan.

Gauss system version 2.2 (1984-1990), Aptech systems, Inc.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometika, 73, 43-56.

Goldstein, H. (1987). Multilevel models in educational and social research. London: Oxford University Press.

Goldstein, H. (1989). Models for multilevel response variables with an application to growth curves. In R.D. Bock (Ed.), Multilevel analysis of educational data (pp. 107-125). New York: Academic.

Graybill, F. A. (1983). Matrices with applications in statistics. Wadsworth Publishing Company, Inc. Belmont, California.

Haaggen, K. and Vittadiini, G. (1991). Regression component decomposition in structural analysis. Communications of Statistical Theory and Method, 20, 1153-1161.

Hartley, H. O., (1958). Maximum likelihood estimation from incomplete data. Biometrics, 14, 174-194.

Hartley, H. O., and Hocking, R.R., (1971). The analysis of incomplete data. Biometrics, 27, 783-808.

International Mathematical and Statistics Libraries (1987). Math/Library and Stat/Library, version 2.0. Houaston: IMSL Problem-Solving Software Systems.

Jamshidian, M. and Jennrich, R. I. (1993). Conjugate gradient acceleration of the EM algorithm. Journal of the American Statistical Association, 88,221-228.

Jo, S. H (1993). Multilevel analysis of structural equation model with E-step and M-step with EQS. The apprenticeship paper presented to the Ph.D guidance committee, Michigan State University.

Joreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. Psychometrika, 32, 443-482.

Joreskog, K.G. (1973). A general method for estimating a linear structural equation sysyem. In A.S. Goldberger and O. D. Duncan (Eds.), Structural equation models in the social sciences. New York: Seminar Press.

Joreskog, K.G (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), Applications of Statistics. Amsterdam: North-Holland.

Joreskog, K.G., & Sorbom, D. (1993). LISREL 8 with PRELIS 2: A guide to the program and applications. Chicago, IL: SPSS.

Lawley, D. N. (1943). On problems connected with item selection and test construction. Proceedings of the Royal Socoety of Edinburgh, Section A, 61, 273-287.

Lee, S. Y. and Poon, W. Y (1992). Two-level analysis of covariance structures for unbalanced designs with small level-one samples. British Journal of Mathematical and statistical Psychology. 45, 109-123.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of psychology, 40.

Lindley, D,V, and Smith, A.F.M. (1972) Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society, Series B, 34, 1-41.

Longford, N. T. (1985). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. unpublished manuscript, Institute for Applied Statistics. Lancaster University, Lancaster, England.

Lumsdaine, A. A. (1963). Instruments and media of instruction. In N.L. Gage (Ed.), Handbook of research on teaching. Rand Mcnally & Company, Chicago.

Magnus, J. R., and Neudecker, H. (1986). Matrix differential calculus and static optimization. John Wiley, Chichester.

Mason, W. M., Wong, G. Y., & Entwistle, B. (1984). Contextual analysis through the multilevel linear model. Sociological Methodology, Sanfrancisco, CA: Jossey-Bass.

McDonald, R.P. (1978). A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and sytatistical Psychology, 31, 59-72.

McDonald, R.P., & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. British Journal of mathematical and Statistical Psychology, 42. 215-232.

Meng, X. L., and Rubin, D. B., (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. Journal of the American Statistical Association, 86, 899-909.

Muthen, B.O. (1987). LISCOMP. Analysis of linear structural equations with a comprehensive measurement model. User's guide [computer program]. Chicago, IL.: Scientific Software.

Muthen, B.O. (1988). Some uses of structural equation modeling in validity studies: Extending IRT to external variables. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 213-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Muthen, B.O. (1990).Means and covariance structure analysis of hierarchical data. University of California at Los Angeles Statistics Series #62. Los Angeles: UCLA.

Press, S. J (1982). Applied Multivariate Analysis : Using Bayesian and Frequentist Methods of Inference. Robert E. Krieger Publishing Company. Malabar, Florida.

Raudenbush,S.W (1984). Applications of a hierarchical linear model in educational research. Unpublished doctoral dissertation, Havard University.

Raudenbush,S.W. and Bryk, A.S.(1985). Empirical Bayes meta-analysis. Journal of Educational Statistics. 10, 75-98.

Raudenbush, S.W., & Bryk, A.S. (1986). A hierarchical model for studying school effects. Sociology of Education, 1-17.

Raudenbush, S.W. (1988). Estimational applications of hierarchical linear models: A review. Journal of Educational Statistics, 13, 85-116.

Raudenbush, S. W (press). Maximum likelihood estimation for unbalanced multilevel structural equation models via the EM algorithm.
British Journal of Mathematical and Statistical Psychology

Rubin, D.B. (1991). EM and Beyond , Psychometrika 56, 241-254.

Schmidt, W.H. (1969). Covariance structure analysis of the multivariate random effects model. Unpublished doctoral dissertation, University of Chicago.

Schmidt, W. H., and Wisenbaker, J. (1986). Hierarchical data analysis: An approach based on structural equations. (CEPSE, No. 4., Research Series). Michigan State University.

Schmidt, W. H., (1993). Survey of mathematics and science opportunities. Research report series no. 57, TIMSS: curriculum analysis a content analytic approach.

Schonemann, P. H., and Steiger,J. H., (1976). Regression component analysis. British Journal of Mathematical and Statistical Psychology, 29, 175-189.

Shonemann, P. H., and Haagen, K. (1987). On the use of factor scores for prediction. Biometrical Journal, 29, 835-847.

Searle, S. R. (1971). Linear Models. John Wiley & Sons, Inc. , New York.

Smith, A. M. F. (1973). A general Bayesian linear model. Journal of the Royal Statistical Society, Series B, 35, 61-75.

Spearman, C., (1904). General intelligence, objectively dtermined and measured. American Journal of Psychology, 15, 201-293.

Tanner, M. A., and Wong, W. H. (1987). The calculation of posterior distribution by data augmentation (with discussion). Journal of the American Statistical Association, 82, 528-550.

Thurstone, L. L. (1947). Multiple factor analysis. Chicago: University of Chicago Press.

Vernon, P. E. (1950). The structure of human abilities. New York: Wiley.

Vredevoogd, Janet (1989) . Split-scale path analysis: The advantages of the full measurement model without some of the problems. Paper presented at the annual meeting of the American Educational Research Assoc., San Francisco, California.

Wiley, D. E., (1967). Analysis of covariance structures. N.S.F. Research Grant Application.

Wiley, D. E., Schmidt, W. H., and Bramble W. J., (1973). Studies of a class of covariance structure models. Journal of American Statistical Associations, 68, 317-323.

Wright, S,. (1934). The methods of path coefficients. Annals of mathematical statistics, 5, 161-215.

Wu, C.F.J (1983). On the convergence properties of the EM algorithm. Annals of Statistics, 11, 95-103.