韓國心理學會誌
Korean Journal of Psychology
1995. Vol. 14. No. 1, 72-84

Empirical Bayes Estimation for Unbalanced Multilevel Structural Equation Models via the EM algorithm

See-Heyon Jo

Mchigan State University

The question of how to analyze unbalanced hierarchical data generated from structural equation models has been a common problem for researchers and analysts. Among difficulties plaguing statistical modeling are removing estimation bias due to measurement error and incorporating variability associated with the social milieu in which individuals are situated. This paper presents empirical Bayes estimation by means of the EM algorithm in the context of unbalanced sampling designs. The EM algorithm is particularly useful when the analytic expressions exist for the conditional expectations of the missing data given complete data and for the maximum likelihood estimators (MLE) of the model parameters. The accuracy of the algorithm was tested using a set of artificial data. The numerical results suggest that this new methodology is a useful mean for studying hypothesized relations among latent variables varying at two levels of hierarchy.

1. Introduction

A distinguishing characteristic of the data encountered in many areas of psychological and educational research is that the sampling structure is hierarchical. Generally, students are taught in groups by a teacher, several classrooms are grouped together into a school, and schools into districts. Then students who attend the same school or classroom are expected to share certain educational experiences. As a result, the educational outcomes for these students will be, to varying degrees, correlated. These effects of clusters are sensibly described within the context of multilevel models (Goldstein, 1987, Bryk & Raudenbush, 1992; Longford, 1993). Under the standard assumption of I.I.D, covariance structure modeling (Joreskog, 1977) of such data misguides statistical inference by not taking into account the intracluster correlations arising from hierarchical data. An important implication of such structure is that the classical assumption of independence among nested observations is violated.

In the context of the linear model, statisticians (Lindley & Smith, 1972;

Raudenbush, 1984, 1988; Goldstein, 1986; Longford, 1987) developed hierarchical linear models (HLM) which are appropriate and powerful means of modeling hierarchical data. It was not until hierarchical linear modeling techniques 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 & Raudenbush, 1987; Goldstein, 1989), cross-national population studies (Mason, Wong & Entwistle, 1984), meta-analysis (Raudenbush & Bryk, 1985), and organization evaluation (Aitkin & Longford, 1986). However, none of the univariate hierarchical models offered methods for estimating measurement error. Also there have been no attempts of applying the EM methodology to the structural equation models for hierarchical data.

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. Then Muthen (1990) proposed an "ad hoc" estimator.

Lee and Poon(1992) also used the similar 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 & Sorbom, 1993) and EQS (Bentler, 1989).

More recently, Raudenbush (in press) proposed an alternate approach for the unbalanced case, He conceptualized the problem in the framework of a balanced design in which all groups have the same number of sampled cases but cases are missing at random. In particular, in the M-step (maximization) the method uses the standard program such as EQS (Bentler, 1989).

Current methods of the likelihood fitting function uses the Davidon-Fletcher /Powell algorithm. This algorithm often may yield one or more variance estimates with nonpositive values, and too-large off-diagonals yielding correlations larger than one. This has been referred to as an improper solution (Gerbing and Anderson, 1987). In that case, there is no way to assess the goodness-of-fit and to interpret parameter estimates. This has been a potential source of controversies in the application of the latent variables model (Joreskog, 1977). Another disappointingly common experience in using LISREL (Joreskog and Sorbom, 1993), EQS (Bentler,

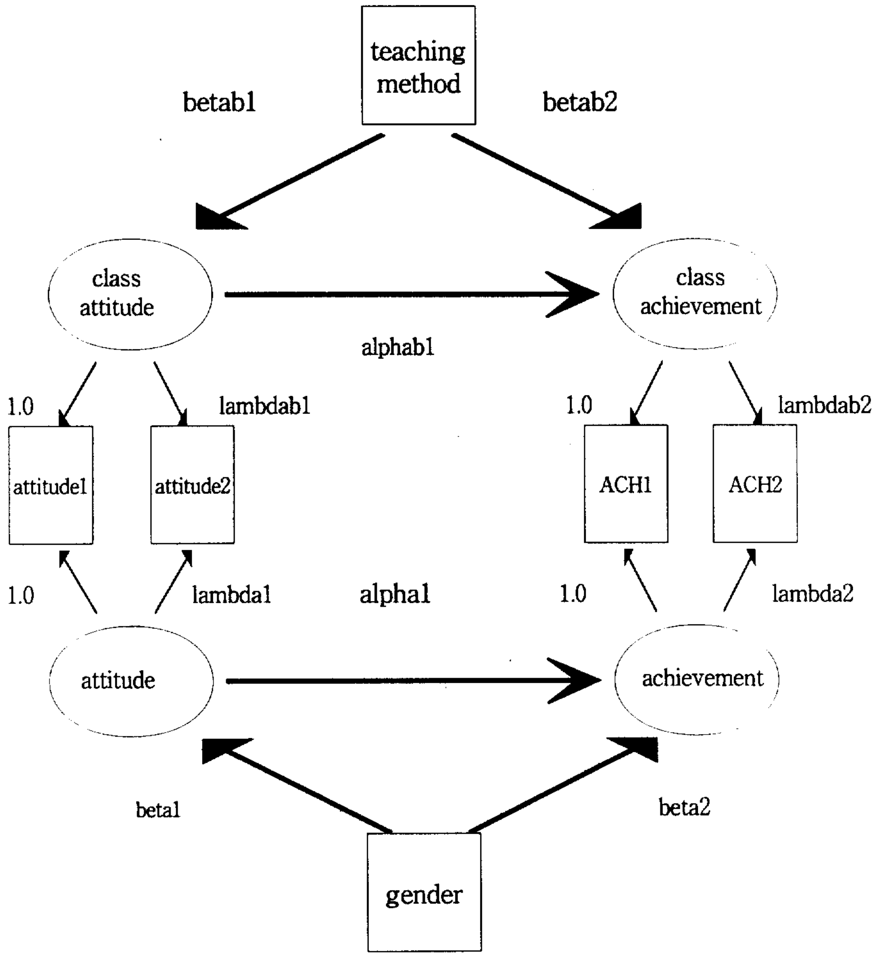

Figure1 A Path Diagram for Multilevel Structural Equation Model

1989) for assessing the goodness-of-fit is to conclude that the fit is not satisfactory, suggesting unknown specification errors. As a result, it is necessary to do a specification search, "a process of sequentially modifying a model so as to improve its fit and parsimony". However there is no reliable guide for respecification procedures (Herting and Costner, 1985). The purpose of this paper is to develop empirical Bayes estimation procedures for computing maximum likelihood (ML) estimates of the parameters in multilevel structural equation models in the context of unbalanced sampling design. This procedure does not require classifying level-2 units into subsets of of level-2 units having equal sample size. The special version of the EM algorithm is implemented for computing ML estimates. This new methodology provides completely new solution and insight into the above long-standing problems. In the following section 2 the general multilevel structural equation model is presented with a very simple example. In section 3 ML estimators of the parameters are developed in the framework of empirical Bayes approach. One numerical analysis of artificial data is made to check the accuracy of the algorithm. Finally the implications of this new methodology and some further research questions are discussed.

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 mathematics achievement is believed to be influenced by a student's gender and attitude toward mathematics, a classroom characteristic teaching style. Teaching styles are believed to influence classroom mean attitude and achievement. Gender also is believed to be related to students' attitude and achievement on the individual level. The path-diagram for this hypothetical model is shown in Figure 1.

A simple item level equation for each individual is

$$y_{ij} = \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \tag{1}$$

where j=1,2,...., J for classrooms, i=1,2,....., n_j for students nested in classroom j. The subscript "w" means the within-level, while "b" means the between-level. where $\varepsilon_{ij} \sim N(0, \Sigma)$; a typical form for Σ is Diagonal($\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2$). Assuming structural linear relationships among constructs, the theoretical relationships on the within-level depicted at the bottom part of Figure 1 can be expressed through the following structural equation :

$$\eta_{ij} = A\eta_{ij} + Bz_{ij} + u_{ij} \tag{2}$$

$$u_{ij} = [u_{1ij} u_{2ij}]^T, \quad u_{ij} \sim N(0, \Delta) \quad .$$

Equation (2) stipulates that on the individual-level the latent variables are a linear function of themselves and the exogenous predictor variables. In our example, gender is an exogenous predictor variable.

Now we reduce equation(5) into

$$\eta_{ij} = Z_{ij}\pi_0 + v_{ij} \tag{3}$$

$$\begin{aligned} v_{ij} &= \begin{bmatrix} v_{1ij} \\ v_{2ij} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{1ij} \\ u_{2ij} \end{bmatrix} \quad , v_{ij} \sim N(0, T_\eta), \end{aligned} \tag{4}$$

$$T_\eta = (I-A)^{-1}\Delta(I-A)^{-1^T},$$

$$\begin{bmatrix} \pi_{10} \\ \pi_{20} \\ \pi_{30} \\ \pi_{40} \end{bmatrix} = vec^*\left(\begin{bmatrix} 1 & 0 \\ -\alpha_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_{10} & \beta_{20} \\ \beta_{30} & \beta_{40} \end{bmatrix} \right) \tag{5}$$

where vec* stacks the transpose of each row of a matrix into a vector. Now on the between-cluster level, the structural relationships depicted at the upper part of Figure 1 can be expressed as follows :

$$\eta_{bj} = A_b\eta_b + B_b w_j + u_{wj} \quad (6)$$

where

$$u_{bj} = [u_{1bj} u_{2bj}]^T, \quad u_{bj} \sim N(0, \Delta_b) \quad .$$

Equation (6) is the expression for the structural relationships among latent variables and the exogenous predictor variable on the cluster level. 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, current membership in a political party or sports club.

Then we have the reduced form for between-level structural equations (6).

$$\eta_{bj} = W_j\pi_{b0} + v_{bj} \quad (7)$$

$$v_{bj} \sim N(0, \ \mathrm{T}_{\eta b}), \quad (8)$$

$$\mathrm{T}_{\eta b} = (I - A_b)^{-1}\Delta_b(I - A_b)^{-1^T} \quad (9)$$

By substituting the structural equations (3) and (7) into (1), we have the following combined equation (10). This representation permits us to develop a special version of the EM algorithm for general multilevel structural equation models :

$$y_{ij} = [\Lambda_w Z_{ij} \mid \Lambda_b W_j]\begin{bmatrix}\pi_0 \\ \pi_{b0}\end{bmatrix} + [\Lambda_w \mid \Lambda_b]\begin{bmatrix}v_{ij} \\ v_{bj}\end{bmatrix} + \varepsilon_{ij} \quad (10)$$

In a more compact form we have :

$$y_{ij} = \Lambda_0 \widetilde{Z_{ij}}\pi + \Lambda_0 \widetilde{w_{ij}} + \varepsilon_{ij} \quad (11)$$

where

$$\Lambda_0 = [\Lambda_w \mid \Lambda_b], \widetilde{Z_j} = \begin{bmatrix}Z_{ij} & 0 \\ 0 & W_j\end{bmatrix},$$

$$\pi = \begin{bmatrix}\pi_0 \\ \pi_{b0}\end{bmatrix}, \widetilde{w_{ij}} = \begin{bmatrix}v_{ij} \\ v_{bj}\end{bmatrix}$$

The model equation (11) subsumes as a special case the hierarchical linear model (Raudenbush, 1988) :

$$Y = A_1\theta_1 + A_2\theta_2 + E, \ \theta_2 \sim N(0, \mathrm{T}), E \sim N(0, \Psi) \quad (12)$$

Based on this general model (12) we develop the empirical Bayes estimation procedure in section 3. In our structural equation model, we consider a population of n level-one units, indexed j (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 endogenous 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.

3. The EM Algorithm for Maximum Likelihood Estimates

Dempster, Laird and Rubin (1977) presented the EM algorithm as a general iterative method for computing maximum likelihood estimates in the presence of incomplete data or missing data. Wu (1983) presented its properties of convergence, 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 of the model parameters given the complete data.

Through casting the measurement model and the reduced form of the structural equation for latent variables into the general model, we can conceptualize our problem as having complete data and incomplete data.

Given

$$
\begin{aligned}
y_{ij} &= \Lambda_w \eta_{ij} + \Lambda_b \eta_{bj} + \varepsilon_{ij} \\
\eta_{ij} &= Z_{ij}\pi_0 + v_{ij} \\
\eta_{bj} &= W_j \pi_{b0} + v_{bj}
\end{aligned}
\tag{13}
$$

We have $\Theta = \{\Lambda_w, \Lambda_b, \Sigma, T_{\eta_b}, T_\eta\}$ is the set of parameters.

$y_{obs} = \{y_{ij}, Z_{ij}, W_j\}$ is the set of observed data.

$y_{mis} = \{\pi_0, \pi_{b0}, v_{ij}, v_{bj}, \varepsilon_{ij}\}$ is the set of missing data.

$c = \{y_{obs}, y_{mis}\}$ is the set of complete data.

Note that in the multilevel structural equation model the factor loadings are parameters rather than observed predictors.

The joint probability density function of the complete data is given by :

$$
\begin{aligned}
f(c, | \Theta, y_{obs}) &\propto (2\pi)^{-Nr/2} |\Sigma|^{-N/2} \\
&\times \exp\left[(-0.5)\sum\sum(y_{ij} - \Lambda_0 \widetilde{Z}_{ij}\pi - \Lambda_0 w_{ij})^T\right] \\
&\Sigma^{-1}(y_{ij} - \Lambda_0 \widetilde{Z}_{ij}\pi - \Lambda_0 w_{ij})\Big] \\
&\times (2\pi)^{-Jp/2} \exp\left[(-0.5)\sum\sum(v_{ij}^T T_\eta^{-1} v_{ij})\right] \\
&\times (2\pi)^{-Jp/2} |T_{\eta b}|^{-J/2} \exp\left[(-0.5)\sum(v_{bj}^T T_{\eta b}^{-1} v_{bj})\right] \\
&\times h(\pi)
\end{aligned}
\tag{14}
$$

where r=the number of indicators, p=the dimension of v_{ij}, and s=the dimension of v_{bj}. The Prior density $h(\pi)$is considered a very small constant and it can be ignored while the empirical Bayes estimators are calculated (Dempster, Rubin & Tsutakawa, 1981).

From Dempster, Laird and Rubin (1977), each iteration of the EM algorithm solves $\frac{\partial}{\partial \Theta}[Q(\Theta, \Theta^{(m)})|_{\Theta=\Theta}^{(m)}] = 0$ for Θ.

where

$$Q(\Theta, \Theta^{(m)}) = \int L(\Theta|c) f(y_{mis}|y_{obs}, \Theta = \Theta^{(m)}) dy_{mis},$$

and

$$\Theta^{(m)} = \{\Lambda_0^{(m)}, \Sigma^{(m)}, T_\eta^{(m)}, T_{\eta b}^{(m)}\}$$

The necessary posterior location vectors $(\pi^*, v_{ij}^* v_{bj}^*)$, and dispersion matrix, $(D_\pi^*, D_v^{*ij}, D_{v_{bj}})$ of the random vectors, π, v_{ij}, v_{bj}, follow from Raudenbush (1988, Appendix A) :

$$
\begin{aligned}
D_\pi^* = [\ &\sum A_{1j}^T \Psi_j^{-1} A_{1j} - \\
&\sum A_{1j}^T \Psi_j^{-1} A_{2j} V_{22j} A_{2j}^T \Psi_j^{-1} A_{1j} - \\
&\sum A_{1j}^T \Psi_j^{-1} A_{3j} V_{32j} A_{2j}^T \Psi_j^{-1} A_{1j} - \\
&\sum A_{1j}^T \Psi_j^{-1} A_{2j} V_{23j} A_{3j}^T \Psi_j^{-1} A_{1j} - \\
&\sum A_{1j}^T \Psi_j^{-1} A_{3j} V_{32j} A_{3j}^T \Psi_j^{-1} A_{1j}]^{-1}
\end{aligned}
$$

$$D_{v_{ij}}* = (A_{2ij}^T \Sigma^{-1} A_{2ij} + T_\eta^{-1})^{-1} + (A_{2ij}^T \Sigma^{-1} A_{2ij} + T_\eta^{-1})^{-1} A_{2ij}^{T\Sigma^{-1}} A_{3ij} V_{33j} A_{3ij}^T \Sigma^{-1} A_{2ij} \times (A_{2ij}^T \Sigma_{-1} A_{2ij} + T_\eta^{-1})^{-1}$$

$$D_v^* = [\ \Sigma(A_{3ij}^T \Sigma^{-1} A_{3ij} + T_{\eta_b^{-1}}) - \Sigma(A_{3ij}^{-1} \Sigma^{-1} A_{2ij})^T (A_{3ij}^T \Sigma^{-1} A_{3ij} + T_\eta^{-1})^{-1} \times (A_{2ij}^T \Sigma^{-1} A_{3ij})]^{-1}$$

$$C_{\pi v_{ij}} = -D_\pi^* A_{1ij}^T \Sigma^{-1} \widetilde{A_{2ij}} (\widetilde{A_{2ij}^T} \Sigma^{-1} \widetilde{A_{2ij}} + T_\beta^{-1})^{-1}$$

where, $\widetilde{A_{2ij}} = \Lambda_0,\ T_\beta = \begin{bmatrix} T_\eta & 0 \\ 0 & T_{\eta_b} \end{bmatrix}$

These posterior distributions given parameters provide point estimates and intervals needed for inference about the random vectors.

To find the ML estimates, we have to maximize the function $Q(\Theta, \Theta^{(m)})$ by taking its first derivatives with respect to $T_\eta, T_{\eta_b}, \Sigma, \Lambda_0$,

(1) $$\frac{dQ(\Theta, \Theta^{(m)})}{dT_\eta} = -\frac{N}{2}[2T_\eta^{-1} - D(T_\eta^{-1})]$$

$$\frac{1}{2}\Sigma\Sigma[\ 2T_\eta^{-1} E(v_{ij} v_{ij}^T | Y=y, \Theta^{(m)}) T_\eta^{-1} - D\{T_\eta^{-1} E(v_{ij} v_{ij}^T | Y=y, \Theta^{(m)}) T_\eta^{-1}\}] \quad (15)$$

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

$$\widehat{T}_\eta = \frac{1}{N}[\Sigma\Sigma(v_{ij}^{*(m)} v_{ij}^{*(m)T} + (D_{v_{ij}})^*] \quad (16)$$

(2) $$\frac{Q(\Theta, \Theta^{(m)})}{QT_{\eta_b}} = -\frac{J}{2}[2T_{\eta b}^{-1} - D(T_{\eta b}^{-1})] +$$

$$\frac{1}{2}\Sigma\Sigma[2T_{\eta b}^{-1} E(v_{bj} v_{bj}^T | Y=y, \Theta^{(m)}) T_{\eta b}^{-1} - D\{T_{\eta b}^{-1} E(v_{bj} v_{bj}^T | Y=y, \Theta^{(m)}) T_{\eta b}^{-1}\}] \quad (17)$$

Setting the derivative equal to null matrix and solving gives the complete-data ML estimate. Then the complete-data ML estimate is :

$$\widehat{T_{\eta b}} = \frac{1}{J}[\Sigma\Sigma(v_{bj}^{*(m)} v_{bj}^{*(m)T} + (D_{v_i})^{*(m)})] \quad (18)$$

(3)

$$\frac{dQ(\Theta, \Theta^{(m)})}{d\Sigma} \frac{d\Sigma}{\sigma_r^2} = \Big[-\frac{N}{2}[2\Sigma^{-1} - D(\Sigma^{-1})] +$$

$$\frac{1}{2}\Sigma\Sigma[2\Sigma^{-1} E(\varepsilon_{ij}\varepsilon_{ij}^T | Y=y, \Theta^{(m)})\Sigma^{-1} - D[\Sigma^{-1} E(\varepsilon_{ij}\varepsilon_{ij_T} | Y=y, \Theta^{(m)})\Sigma^{-1}]]\Big] \times e_r^*$$

(19)

where e_r^* is the column indicator vector which has a 1 in the r-th position and zeros in other positions. Setting the derivative equal to zero and solving gives the complete-data ML estimate.

$$\Sigma = Diagonal(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) \quad (20)$$

(4) By setting $\frac{dQ(\Theta, \Theta^{(m)})}{d(\Lambda_0^T)} \frac{d(\Lambda_0)}{d\lambda_{gk}}$ equal to zero and solving gives the complete-data ML estimate, $\widehat{\lambda_{gk}}$.

Then, $\widehat{\Lambda_0} = [\widehat{\lambda_{gk}}]$ (21)

The element $\widehat{\lambda_{gk}}$ is in the g-th row and k-th column of the matrix $\widehat{\Lambda_0}$, and zeros elsewhere.

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

(1) E-step : Find

$E\log[L(c,\Theta)|y,T_{\eta}^{(m)}]$,

M-step : Substitute the equation (16) with these quantities, and then we obtain new

T_{η}, set $T_{\eta}^{(m+1)}$ equal to this new T_{η},

(2) E-step : Find

$E\log[L(c,\Theta)|y,T_{\eta b}^{(m)}]$,

M-step : Substitute the equation (18) with these quantities, and then we obtain new

T_{η_b}, set $T_{\eta b}^{(m+1)}$ equal to this new T_{η_b},

(3) E-step : Find

$E\log[L(c,\Theta)|y,\Sigma^{(m)}]$,

M-step : Substitute the equation (20) with these quantities, and then we obtain new

Σ, set $\Sigma^{(m+1)}$equal to this new Σ.

(4) E-step : Find

$\pi^{*(m)T}$, $\omega_{ij}^{*(m)T}$, $D_{(\omega_{ij}^{*(m)}}$, $D_{\pi}^{8(m)}$, $C_{\pi\omega ij}^{*(m)}$, acting as if $\Lambda_0=\Lambda_0^{(m)}$

M-step : Substitute the equation (21) with these quantities, and then we obtain new

Λ_0, set $\Lambda_0^{(m+1)}$ 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 the log likelihood of successive iterates differ from each other by no more than some number (i.e., = 10^{-6}). I conclude this section with the expression for the observed log-likelihood function. Monitoring convergence of the algorithm and making likelihood ratio test monitored are done by computing the observed data log likelihood on each iteration, which is given by

$$
\begin{aligned}
LLR(\Theta|y) &= \frac{N}{2}\log(det\Sigma)-\frac{N}{2}\log(det T_{\eta}) \\
&-\frac{J}{2}\log(det T_{\eta b})+\frac{1}{2}\log(det V_{11}) \\
&+\frac{1}{2}\log(det Q_{3j})+\frac{1}{2}\log(det U_j^{-1}) \\
&-\sum\sum[y_{ij}^{T}\Sigma^{-1}(y_{ij}-A_{1ij}\pi^{*}-A_{2ij}\omega_{ij}^{*})]
\end{aligned}
$$

4. Numerical Results

The output of the Table 2 is the result of fitting unbalanced data to the model. The distribution of the number of groups per group size is given in Table 1. By a subroutine (Jo, 1993) generating for hierarchical data, one sample of 4890 observational units is created for this study. 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.

Group size	Unbalanced data
6	10
7	10
8	10
9	20
10	450

Table 1 Number of Groups per Group Size

	Mean	st. dev.	Sample covariance matrix			
Y1	-0.285	7.805	60.922			
Y2	-0.187	6.533	40.477	42.686		
Y3	-0.167	7.957	16.370	13.360	63.309	
Y4	-0.160	6.041	10.376	8.746	35.342	36.492

Table 2 Descriptive Statistics for Sample

In discussing the results reported in Table 3 and Table 4, we say that the EM algorithm recovered the population parameters values fairly well. The criterion used for convergence of the observed log-likelihood is that the difference of (i-1)th log-likelihood and i-th log-likelihood

	Population Parameters	Starting Values	Estimated Parameters
λ_{w1}	0.82	0.582	0.85592
λ_{w2}	0.73	0.573	0.71902
λ_{b1}	0.75	0.575	0.73673
λ_{b2}	0.66	0.753	0.6337
$t_{\eta_{11}}$	30.00	20.000	32.71043
$t_{\eta_{12}}$	10.00	7.000	10.53190
$t_{\eta_{22}}$	30.00	20.000	32.54095
$t_{\eta_{b11}}$	20.00	24.010	22.70168
$t_{\eta_{b12}}$	4.00	3.000	3.69027
$t_{\eta_{b22}}$	20.00	25.000	22.83989
σ_1^2	10.00	8.00	7.81079
σ_2^2	10.00	8.000	11.09361
σ_3^2	12.00	9.000	10.24275
σ_4^2	12.00	9.000	13.10782
a_{w1}	0.33	0.200	0.31947
a_{b1}	0.30	0.150	0.25065
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 3 Parameter estimates for unbalanced data

is smaller than $(0.1)^6$, that is, ($\delta \leq (0.1)^6$). In the 486 IBM-PC with the speed of 66 mhz for convergence it took about 87.5 hours. The Table 4 shows the list of log-likelihood values. As the table shows the number of iterations is very large and the spent hours is quite long. This seems to be caused by the fact that missing information is relatively large in the multilevel structural equation model.

Iteration	1019	-23905.70543
Iteration	1020	-23905.70516
Iteration	1021	-23905.70488
Iteration	1022	-23905.70461
Iteration	1023	-23905.70434
Iteration	1024	-23905.70407

Table 4 The Values of the Observed Log-likelihood

5. Conclusion

Research in the field of social science provides various challenges. For example, the random assignment of individuals to a set of conditions is not realistic in most cases. Even an experimental setting (Lumsdaine, 1963) the outcomes may show 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. This situation makes the application of the conventional linear structural equation models to the real data inappropriate. As Cronbach(1976) pointed out, due to by not recognizing the nature of hierarchical data, many studies in the field of education have used inappropriate analyses. The difficulty of analyzing data arising from two levels is in assessing the nature of intervariable relationships at both levels simultaneously. This paper 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.

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 of exogenous 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. The approach presented here can be generalized to incorporate random slopes. As is well known, the EM algorithm is numerically stable, but is slow. Recently Jamshidian and Jennrich (1993) developed a conjugate gradient scheme for accelerating. In that case the evaluation of the gradient of the likelihood function is essential. In the case of small number of groups one can employ the full Bayesian approach can be implemented with Gibbs sampler algorithm.

References

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

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

Bryk, A. S., & 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.

Cronbach, L. J.(1976). *Research on classsrooms and schools: formulation of questions, design and analysis.* Occaional paper of the Stanford Evaluation Consortium, School of Education, Stanford University.

Dempster, A. P., Laird, N. M & rubin, D. B.(1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series* B, 39, 1-38.

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

Gerbing, D. W. & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: their interpretability and a comparison of alternate respecifications. Psychometrika, vol, 52, 99-111.

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

Goldstein, H. (1987). *Multilevel models in educational and social research.* C. Griffin & Co., London and Oxford University Press, New York.

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 Press, San Diego.

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

Herting, J. R. & Costner, H. L. (1985). Respecification in the multiple indicator models. in H. M. Blalock Jr. (Ed.), Causal models in the social science (2nd ed., 321-393). Hawthorne, NY: Aldine.

Jamshidian, M. & Jennrich, R. I. (1993). Conjugate gradient acceleration of the EM algorithm. *Journal of the American Statistical Association*, 88, 221-229.

Jo, S. H. (1993). Multilevel analysis of structural equation models with E-step and M-step with EQS. The apprenticeship paper, Michigan state University.

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

Joreskog, K. G., & Sorbom, D. (1993). *LISREL8 with PRELIS : A guide to* *the program and applications.* Chicago, IL : SPSS.

Lee, S. Y. & 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.

Lindley, D, B, & 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. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika,* 74, 817-827

Longford, N. T. (1993). *Random Coefficient Models.* Clarendon Press, Oxford

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

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., & 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.

Muthen, B. O. (1988). *LISCOMP : User's guide.* Mooresville, IN : Scientific Software.

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

Press, S. J. (1982). *Applied Multivariate Analysis : Using Byesian 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. & Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics.* 10, 75-98.

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

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

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

韓國心理學會誌
Korean Journal of Psychology
1995. Vol. 14. No. 1, 72-84

EM 알고리즘에 의한 다층 선형구조방정식 모형의 ML 추론

조 세 현

Mchigan State University

사회과학의 제반 연구분야에서 다루는 분석자료들은 대부분 위계 구조(hierarchical) 또는 다층구조(multilevel)의 특성을 갖는 것으로 알려져 있다. 다층구조의 자료를 분석하기 위하여 여하히 선형 구조방정식 모형을 세우고, ML 추론이론을 정립하며, 계산수행을 위해 어떻게 알고리즘을 설계하여 컴퓨터 프로그램을 만드느냐 하는 것은 사회과학 연구자들과 통계자료 분석자들의 공통된 주요 연구과제 가운데 하나이다. 이 논문의 목적은 불균형자료라는 보다 일반적인 조건에서 EM 알고리즘에 의한 empirical Bayes 추론방법으로 다층선형구조방정식모형에 대한 ML 해법을 보여주는 것이다.