Published in final edited form as:

Educ Psychol Meas. 2011 March 22; 71(2): 325–345. doi:10.1177/0013164410381272.

Constrained Maximum Likelihood Estimation for Two-level Mean and Covariance Structure Models²

Peter M. Bentler,

University of California, Los Angeles

Jiajuan Liang,

University of New Haven

Man-Lai Tang, and

Hong Kong Baptist University

Ke-Hai Yuan

University of Notre Dame

Abstract

Maximum likelihood is commonly used for estimation of model parameters in analysis of two-level structural equation models. Constraints on model parameters could be encountered in some situations such as equal factor loadings for different factors. Linear constraints are the most common ones and they are relatively easy to handle in maximum likelihood analysis. Nonlinear constraints could be encountered in complicated applications. In this paper we develop an EM-type algorithm for estimating model parameters with both linear and nonlinear constraints. The empirical performance of the algorithm is demonstrated by a Monte Carlo study. Application of the algorithm for linear constraints is illustrated by setting up a two-level mean and covariance structure model for a real two-level data set and running an EQS program.

Keywords

EM algorithm; maximum likelihood estimation; mean and covariance structure; linear and nonlinear constraints; two-level structural equation model

Two-level mean and covariance structure models (or structural equation models, SEM for simplicity) have been applied to data analysis in various fields such as education, medicine, psychology and sociology. In many practical situations, the hierarchical structure that is implicit in the collected data should not be ignored as various potentially important sources of variance can be identified. For example, SAT (Scholastic Attitude Test) scores from students nested in different high schools not only can be expected to reveal individual student differences, but also to reflect differences among the implicit variable “schools”. Schools can be expected to vary in the extent of student preparation, their socioeconomic environments, their teaching facilities, and their teacher training. Hence, the SAT scores will reflect school differences as well as individual differences, and as a result of the dependence introduced by school differences, the SAT scores are best not considered as independent observations. In this case, variation is at two levels, with the students called the level-1 units (or individuals) and the schools called the level-2 units (groups or clusters). Furthermore, when considering latent sources of variation giving rise to the SAT scores, it is probable that

²Correspondence should be addressed to Peter M. Bentler, University of California at Los Angeles, Box 951563, Los Angeles, CA 90095-1563. bentler@ucla.edu

the influences at the two levels are different. For instance, factors such as “general math ability” and “general writing ability” may explain performance at level one, while factors such as “school quality” or “teacher background” may explain school differences at level two. It is now widely recognized that to understand how the level-1 factors influence one another and their effects on the data, a level-1 model must be used to analyze the within (level-1) covariance structure; such a model may include the factor loadings, variances and covariances of factors and prediction errors. Similarly, to understand how the level-2 factors influence one another and their effects on the data, a level-2 model has to be set up to analyze the “between” (level-2) covariance structure. By observing the two different types of effects on two-level data, many researchers have proposed various formulations for two-level covariance structure models, see, for example, McDonald and Goldstein (1989), Muthén (1994), Lee and Poon (1998), Liang and Bentler (2004).

When a linear two-level SEM has been set up for analysis of the within and between covariance structures, the first problem to solve is estimation of model parameters. There are two basic methods for parameter estimation for two-level SEM in the literature: the maximum likelihood (ML) method and the generalized least squares (GLS) method. The ML method is closely related to the normal distributional assumption on the model and it is the most popular one in analysis of two-level SEM because of its easy implementation and some nice properties. Different ML methods have been proposed for analysis of two-level SEM and they are available in many standard statistical packages such as EQS (Bentler, 2006), LISREL (cf. du Toit & du Toit, 2008), and *Mplus* (Muthén & Muthén, 2004). Unfortunately, many existing algorithms for ML estimation do not allow general parameter constraints although they may allow simple linear constraints on parameters. Parameter constraints in ML analysis of statistical models are usually imposed as a result of substantive prior knowledge on the relationships among model parameters. For example, in a factor analysis model, if a factor is measured by indicator variables and prior knowledge implies that they are equally important, an equality constraint on the factor loadings may be imposed when implementing the ML estimation. Linear constraints are the simplest ones in ML analysis of statistical models with parameter constraints, and simple analytical solutions are usually available, see, for example, Kim and Taylor (1995), and Jamshidian (2004). Nonlinear functional constraints on parameters may appear in some complicated problems and this type of constraints has been less thoroughly studied. A detailed discussion on the necessity of constrained ML estimation can be dated back to Aitchison and Silvey (1958). Lee (1979) proposed an algorithm for both ML and weighted least squares estimation with general (linear and nonlinear) constraints for conventional (one-level) SEM. Lee and Tsang (1999) developed EM type algorithms for constrained ML estimation of two-level SEM with only covariance structures.

The importance of consideration of a nonsaturated mean structure in conventional SEM has been discussed by Yung and Bentler (1999). Liang and Bentler (2004) proposed a new formulation for two-level SEM and developed an EM algorithm that allows both mean and covariance structures in ML estimation without constraints. The study in this paper is a generalization of that in Liang and Bentler (2004) and an extension of Lee and Tsang (1999). The paper is organized as follows. Section 2 gives a simple review on the model in Liang and Bentler (2004). Section 3 presents the details of the algorithm associated with the Lagrange multiplier. Section 4 provides a limited Monte Carlo study on the performance of the proposed algorithm by using an artificial two-level SEM. Section 5 gives an illustration of the proposed algorithm with linear constraints by implementing a two-level SEM on EQS for a practical data set. Some concluding remarks are given in the last section.

A REVIEW ON THE EXISTING MODEL

Let $\{z_g : g = 1, \dots, G\}$ be a set of level-2 observations from G level-2 units such as financial sources for schools, and $\{y_{gi} : i = 1, \dots, N_g; g = 1, \dots, G\}$ a set of level-1 observations. For example, y_{gi} may stand for the observation from the i -th level-1 unit (e.g., student) nested in the g -th level-2 unit (e.g., schools). Liang and Bentler (2004) proposed the following data formulation of two-level SEM

$$\begin{pmatrix} z_g \\ y_{gi} \end{pmatrix} = \begin{pmatrix} z_g \\ v_g \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ v_{gi} \end{pmatrix} \quad (1)$$

associated with the assumptions:

- A1)** v_{gi} contains latent variables capturing level-1 effects. $\{v_{gi} : i = 1, \dots, N_g\}$ (for each fixed g) are i.i.d. (independently identically distributed) and $v_{gi} \sim N_p(\mathbf{0}, \Sigma_W)$, the p -dimensional normal distribution with $\Sigma_W > \mathbf{0}$ (positive definite) for $g = 1, \dots, G$;
- A2)** v_g contains latent variables capturing level-2 effects. $\{v_g : g = 1, \dots, G\}$ are i.i.d. and $v_g \sim N_p(\mathbf{0}, \Sigma_B)$ with $\Sigma_B > \mathbf{0}$;
- A3)** $\{z_g : g = 1, \dots, G\}$ are i.i.d. level-2 observations and $z_g \sim N_q(\mu_z, \Sigma_{zz})$ with $\Sigma_{zz} > \mathbf{0}$;
- A4)** the random vector $(z'_g, v'_g)' ((p+q) \times 1)$ has a joint nonsingular multivariate normal distribution $N_{p+q}(\mu, \Sigma_B)$ with $\Sigma_B > \mathbf{0}$ and

$$\mu(\theta) = \begin{pmatrix} \mu_z \\ \mu_y \end{pmatrix}, \quad \Sigma_B(\theta) = \text{cov} \begin{pmatrix} z_g \\ v_g \end{pmatrix} = \begin{pmatrix} \Sigma_{zz} & \Sigma_{zy} \\ \Sigma_{yz} & \Sigma_B \end{pmatrix}, \quad (2)$$

where $\Sigma_{zy} = \Sigma'_{yz} = \text{cov}(z_g, v_g)$;

- A5)** $\{z_g, v_g\}$ is uncorrelated with $\{v_{gi} : i = 1, \dots, N_g\}$.

The mean and covariance structures in (2) may be characterized by a common model parameter vector θ ($r \times 1$) with r functionally independent model parameters. Nonsaturated mean structure in (2) implies that the means μ_z and μ_y may be also characterized by the common model parameter vector θ . Existing algorithms for ML analysis of two-level SEM usually treat the means from manifest variables as individual parameters separated from the covariance parameters, see, for example, McDonald and Goldstein (1989), Muthén (1994), Raudenbush (1995), and du Toit and du Toit (2008). When the number of functionally independent model parameters r in θ in (2) is less than the total number of means from all manifest (observable) variables in z_g and y_{gi} in formulation (1) plus all variances and nonduplicated co-variances, a model with formulation (1) is called a nonsaturated model. The purpose of ML analysis of a two-level SEM under formulation (1) is to estimate the common parameter vector θ and validate the nonsaturated mean and covariance structures (2) (the null hypothesis) versus the saturated model (the alternative hypothesis) that considers all means, variances and nonduplicated covariances as independent model parameters. Under the null hypothesis, Liang and Bentler (2004) developed an EM algorithm for ML estimation of θ and validated the nonsaturated model.

Under assumptions A1)–A5) on model formulation (1) and considering $\{y_{gi}, z_g : g = 1, \dots, G\}$ as complete observations with missing values $\{v_g : g = 1, \dots, G\}$, Liang and Bentler (2004) obtained the E-step function

$$M(\theta^*|\theta) = N \left\{ \log |\Sigma_w(\theta^*)| + \text{tr}[\Sigma_w^{-1}(\theta^*) S_w(\theta)] \right\} + G \left\{ \log |\tilde{\Sigma}(\theta^*)| + \text{tr}[\tilde{\Sigma}^{-1}(\theta^*) \tilde{S}(\theta)] \right\}, \quad (3)$$

where both θ^* and θ are two arbitrarily specified values of the same parameter θ , and

$$\tilde{\Sigma}(\theta^*) = \begin{pmatrix} \tilde{\Sigma}_B^* + \mu^* \mu^{*'} & \mu^* \\ \mu^{*'} & 1 \end{pmatrix}, \tilde{S}(\theta) = \begin{pmatrix} \tilde{S}_B + d d' & d \\ d' & 1 \end{pmatrix},$$

$$N = \sum_{g=1}^G N_g, \tilde{\Sigma}_B^* = \tilde{\Sigma}_B(\theta^*), \mu^* = \mu(\theta^*), \quad (4)$$

where $\tilde{\Sigma}_B(\theta^*)$ and $\mu(\theta^*)$ are defined in (2) by taking $\theta = \theta^*$. The simplified formulas for computing the terms S_w , d and $\tilde{S}_B$ in (4) can be found in Liang and Bentler (2004).

THE ALGORITHM ASSOCIATED WITH THE LAGRANGE MULTIPLIER

By resorting to the simple E-step function (3), we can give an algorithm for ML estimation with general constraints for model formulation (1) with mean and covariance structures (2). Assume that there are s general parameter constraints defined by

$$h_i(\theta) = 0, i = 1, \dots, s, \text{ or } h(\theta) = \mathbf{0}, \mathbf{0}: s \times 1, \quad (5)$$

where $h(\theta) = (h_1(\theta), \dots, h_s(\theta))'$, $s < r$ and r is the dimension of the parameter vector θ ($r \times 1$) in the mean and covariance structures (2). $h_i(\theta)$ is a scalar function that permits up to the second partial derivatives with regard to θ . The standard steps for applying the Lagrange multiplier method can be summarized as follows (cf. Bertsekas, 1976; Lee & Tsang, 1999).

Step 1. Construct the augmented Lagrangian function

$$M(\theta^*|\theta) + \xi' h(\theta^*) + c \sum_{i=1}^s \phi[h_i(\theta^*)], \quad (6)$$

where $\xi = (\xi^{(1)}, \dots, \xi^{(s)})'$ contains s multipliers, c is a positive scalar constant and $\phi(\cdot)$ is a penalty (positive) function, $\phi(x) = 0$ if and only if $x = 0$, e.g., $\phi(t) = t^2/2$;

Step 2. For the current values of $\theta = \theta_j$, $c = c_j > 0$ and $\xi = \xi_j$, search a minimum point of $\theta^* = \theta_{j+1}$, say, such that the function

$$M_j(\theta^*|\theta_j) = M(\theta^*|\theta_j) + \xi_j' h(\theta^*) + c_j \sum_{i=1}^s \phi[h_i(\theta^*)], \quad (7)$$

is minimized at $\theta^* = \theta_{j+1}$;

Step 3. Increase c_j to another value $c_{j+1} > c_j > 0$ (e.g., $c_{j+1} = 1.5c_j$) and update

$\xi_j = (\xi_j^{(1)}, \dots, \xi_j^{(s)})'$ to $\xi_{j+1} = (\xi_{j+1}^{(1)}, \dots, \xi_{j+1}^{(s)})'$ by $\xi_{j+1}^{(i)} = \xi_j^{(i)} + c_{j+1} \phi[h_i(\theta_{j+1})]$ ($i = 1, \dots, s$),

where ϕ denotes the derivative of $\phi(\cdot)$. Update j to $j + 1$ and go to step 2. The process is terminated at the $(j + 1)$ -th iteration if the maximum absolute difference between θ_j and θ_{j+1} is less than a pre-assigned small value $\varepsilon > 0$.

The choice for the increasing constant series $\{c_j > 0\}$ in the above Step 3 is somewhat uncertain. It controls the magnitude of penalty from the penalty function $\phi(\cdot)$ that is related to convergent speed of the algorithm. There is no optional or definite rule for the choice. A rapidly increasing series $\{c_j > 0\}$ will generally result in faster convergence but may also lead to breakdown of the algorithm by generating non-positive definite estimates for the covariance matrices used in the algorithm. A convenient choice for updating the increasing constant series $\{c_j > 0\}$ is to choose $c_{j+1} = \gamma c_j$ with constant $\gamma > 1$. The role of $\{c_j > 0\}$ is like that of the step-halving constant in iterative algorithms, which can be chosen as a decreasing positive constant series if some step length is too big and thus makes the algorithm break down, while the purpose of $\{c_j > 0\}$ is to increase the penalty step by step in the iteration to make it converge faster. More discussions on the multiplier method are referred to Bertsekas (1976).

The above iteration process was proved to be convergent for sufficiently large c_j under mild conditions (Bertsekas, 1976). The key idea for minimizing $M_j(\theta^*|\theta_j)$ with respect to θ^* at the current value θ_j is to find an updated value $\theta^* = \theta_{j+1}$ by

$$\theta_{j+1} = \theta_j + \rho \Delta \theta_j, \quad (8)$$

where ρ is the step-halving parameter (Kim & Taylor, 1995) that takes values like 1, 1/2, 1/4, so that

$$M_j(\theta_{j+1}|\theta_j) \leq M_j(\theta_j|\theta_j), j=1, 2, \dots \quad (9)$$

This can be realized by using the EM gradient algorithm (Lange, 1995a, b) through choosing $\Delta \theta_j$ in (8) along the gradient direction at $\theta^* = \theta_j$. This is related to the calculation of the first and the second derivatives of the function $M_j(\theta^*|\theta_j)$ in (7). The details on the calculation of the increment $\Delta \theta_j$ in (8) are given as follows. Let

$$\begin{aligned} \dot{M}_j(\theta_j|\theta_j) &= \left. \frac{\partial M_j(\theta^*|\theta_j)}{\partial \theta^*} \right|_{\theta^*=\theta_j}, & \dot{h}(\theta_j) &= \left. \frac{\partial h'(\theta^*)}{\partial \theta^*} \right|_{\theta^*=\theta_j}, \\ \ddot{M}(\theta_j|\theta_j) &= \left. \frac{\partial^2 M(\theta^*|\theta_j)}{\partial \theta^{*2}} \right|_{\theta^*=\theta_j}, & \ddot{\phi}[h_i(\theta_j)] &= \left. \frac{\partial^2 \phi[h_i(\theta^*)]}{\partial \theta^{*2}} \right|_{\theta^*=\theta_j}. \end{aligned} \quad (10)$$

The following first order derivative was given by Liang and Bentler (2004):

$$\dot{M}(\theta|\theta) = N \Delta_w (\Sigma_w^{-1} \otimes \Sigma_w^{-1}) \text{vec}(\Sigma_w - S_w) + G \tilde{\Lambda} (\tilde{\Sigma}^{-1} \otimes \tilde{\Sigma}^{-1}) \text{vec}(\tilde{\Sigma} - \tilde{S}). \quad (11)$$

where,

$$\Delta_w = \frac{\partial (\text{vec} \Sigma_w)'}{\partial \theta}, \tilde{\Lambda} = \frac{\partial (\text{vec} \tilde{\Sigma})'}{\partial \theta}, \tilde{\Sigma} = \tilde{\Sigma}(\theta^*) \Big|_{\theta^*=\theta}, \quad (12)$$

with the terms S_w , $\tilde{\Sigma}$ and $\tilde{S}$ are the same as in (3). Hence the first derivative of the function in (7) at the current value $\theta^* = \theta_j$ is calculated by

$$\dot{M}_j(\theta_j|\theta_j) = \dot{M}(\theta_j|\theta_j) + \dot{h}(\theta_j)\xi_j + c_j \sum_{i=1}^s \dot{\phi}[h_i(\theta_j)], \quad (13)$$

with the terms given in (10) and (11).

Next, we need to compute the expected value of the Hessian matrix (the second derivative) of the function in (7) at the current value $\theta^* = \theta_j$. Let

$$\begin{aligned} \ddot{M}_j(\theta_j|\theta_j) &= \left. \frac{\partial^2 M_j(\theta^*|\theta)}{\partial \theta^* \partial \theta^{*'}} \right|_{\theta^* = \theta = \theta_j}, & \ddot{h}_i(\theta_j) &= \left. \frac{\partial^2 h_i(\theta)}{\partial \theta \partial \theta'} \right|_{\theta = \theta_j}, \\ \ddot{M}(\theta_j|\theta_j) &= \left. \frac{\partial^2 M(\theta^*|\theta)}{\partial \theta^* \partial \theta^{*'}} \right|_{\theta^* = \theta = \theta_j}, & \ddot{\phi}[h_i(\theta_j)] &= \left. \frac{\partial^2 \phi[h_i(\theta)]}{\partial \theta \partial \theta'} \right|_{\theta = \theta_j}. \end{aligned} \quad (14)$$

The following approximated expected Hessian matrix was given by Liang and Bentler (2004):

$$I(\theta) = E \left\{ \frac{\partial^2}{\partial \theta^* \partial \theta^{*'}} M(\theta^*|\theta) \right\} \Big|_{\theta^* = \theta} \approx N \Delta_w \left\{ 2(\Sigma_w^{-1} S_{ew} \Sigma_w^{-1}) \otimes \Sigma_w^{-1} - \Sigma_w^{-1} \otimes \Sigma_w^{-1} \right\} \Delta'_w + G \tilde{\Lambda} \left\{ 2(\tilde{\Sigma}^{-1} \tilde{S}_{eb} \tilde{\Sigma}^{-1}) \otimes \tilde{\Sigma}^{-1} - \tilde{\Sigma}^{-1} \otimes \tilde{\Sigma}^{-1} \right\} \tilde{\Lambda}', \quad (15)$$

where Δ_w and $\tilde{\Lambda}$ are given by (12),

$$\begin{aligned} S_{ew} &= 2\Sigma_B + \Sigma_w - \sum_{g=1}^G \frac{N_g}{N} (A_{gw} - A'_{gw}), \\ A_{gw} &= \Sigma_{2g} \Omega_g^{-1} \Sigma'_1, \Sigma_{2g} = (\Sigma_{yz}, \frac{1}{N_g} \Sigma_g), \end{aligned} \quad (16)$$

with

$$\Omega_g = \begin{pmatrix} \Sigma_{zz} & \Sigma_{zy} \\ \Sigma_{yz} & \frac{1}{N_g} \Sigma_g \end{pmatrix}, \Sigma_g = \Sigma_w + N_g \Sigma_B, \Sigma_1 = (\Sigma_{yz}, \Sigma_B), \quad (17)$$

for $g = 1, \dots, G$, and

$$S_{eb} = \begin{pmatrix} \Sigma_{zz} & A_b \\ A'_b & \Sigma_B \end{pmatrix}, A_b = (\Sigma_{zz}, \Sigma_{zy}) \bar{\Omega}^{-1} \Sigma'_1, \bar{\Omega}^{-1} = \frac{1}{G} \sum_{g=1}^G \Omega_g^{-1}. \quad (18)$$

Referring to (14) and (15), we obtain the approximated expected Hessian matrix of the function in (7):

$$H_j \stackrel{\text{def}}{=} E[\ddot{M}_j(\theta_j|\theta_j)] = I(\theta_j) + \sum_{i=1}^s \xi_j^{(i)} \ddot{h}_i(\theta_j) + c_j \sum_{i=1}^s \ddot{\phi}[h_i(\theta_j)], \quad (19)$$

where $(\xi_j^{(1)}, \dots, \xi_j^{(s)})' = \xi_j$ contains the current values of the components in the multiplier ξ_j in (7).

When taking $\phi(t) = t^2/2$ as a quadratic penalty function in (7), the terms $\dot{\phi}[h_i(\theta_j)]$ in (10) and $\ddot{\phi}[h_i(\theta_j)]$ in (14) reduce to

$$\dot{\phi}[h_i(\theta_j)] = h_i(\theta_j) \frac{\partial h_i(\theta)}{\partial \theta} \Big|_{\theta=\theta_j}, \quad (20)$$

and

$$\ddot{\phi}[h_i(\theta_j)] = \left\{ h_i(\theta) \frac{\partial^2 h_i(\theta)}{\partial \theta \partial \theta'} + \frac{\partial h_i(\theta)}{\partial \theta} \frac{\partial h_i(\theta)}{\partial \theta'} \right\} \Big|_{\theta=\theta_j}, \quad (21)$$

respectively.

According to the EM gradient algorithm in Lange (1995a, b) and Lee and Tsang (1999), we can summarize the EM algorithm for computing the ML estimation of θ in the mean and covariance structures (2) subject to the general constraints (5) as follows.

Step 1. Given a current value of $\theta = \theta_j$, a current value of the multiplier

$\xi_j = (\xi_j^{(1)}, \dots, \xi_j^{(s)})'$, and a constant $c_j > 0$, construct the Lagrangian function (6);

Step 2. Update θ_j to θ_{j+1} by (8) with

$$\Delta \theta_j = -H_j^{-1} \dot{M}(\theta_j | \theta_j), \quad (22)$$

where $\dot{M}(\theta_j | \theta_j)$ is computed by (13) and H_j by (19);

Step 3. Continue to update θ_{j+1} in Step 2 by (8) and (22) and the iteration stops until $\text{RMSE}(\theta_{j+1}, \theta_j) < \varepsilon$ (e.g., $\varepsilon = 10^{-6}$), where RMSE stands for the Root Mean Square Error defined by

$$\text{RMSE}(\theta_{j+1}, \theta_j) = \left\{ \frac{1}{r} \|\theta_{j+1} - \theta_j\|^2 \right\}^{1/2}, \quad (23)$$

and the sign " $\|\cdot\|$ " denotes the usual Euclidean distance, r is the dimension of θ ($r \times 1$).

When the model parameters in the mean and covariance structures (2) are subject to linear constraints, the above steps can be further simplified. General linear constraints can be expressed as

$$h(\theta) = C\theta - a = 0, \quad (24)$$

where C is a known $m \times r$ matrix with $\text{rank}(C) = m < r$ ($\theta : r \times 1$) and a is an $m \times 1$ constant vector. Because the Lagrange multiplier method for ML estimation with linear constraints in the two-level SEM formulation (1) with mean and covariance structures (2) has been coded into EQS Version 6.1 (Bentler, 2006), we do not present the details here and guide readers to the practical application of the method, see the Appendix for a sample EQS program for ML

estimation with linear constraints. Note that when the general constraint (5) becomes the linear constraint (24) with $\phi(t) = t^2/2$ in (7) as in EQS, the first order derivative (13) reduces to

$$\dot{M}_j(\theta_j|\theta_j) = \dot{M}(\theta_j|\theta_j) + C\xi_j + c_j C'(C\theta_j - a), \quad (25)$$

where $\dot{M}(\theta_j|\theta_j)$ is given by (11) with $\theta = \theta_j$, and the expected Hessian matrix (19) reduces to

$$H_j = I(\theta_j) + c_j C' C, \quad (26)$$

where $I(\theta_j)$ is given by (15) with $\theta = \theta_j$.

A MONTE CARLO STUDY

In this section, we construct an artificial two-level SEM according to formulation (1) and generate the empirical two-level data from the model with the assumptions A1)–A5) on formulation (1). Then we use the algorithm in Section 3 to estimate the model parameters. The model is illustrated by the path diagram in Figure 1.

In the two-level SEM presented by Figure 1, the level-1 model is a factor analysis model with only one within factor F_W that generates four indicator variables: Y_1, Y_2, Y_3 and Y_4 with independent residual errors. The level-2 model is not a factor analysis model, where the factor is affected by two correlated observable level-2 variables Z_1 and Z_2 with means μ_1 and μ_2 , respectively. It has only one between factor F_B that generates the same four indicator variables Y_i ($i = 1, 2, 3, 4$) with independent residual errors. The path diagram uses the standard conventions in presenting a two-level SEM: a one-way arrow pointing to a variable indicates that the variable is influenced by another variable. A double arrowed curve means a covariance (or correlation, if standardized) between two variables. An arrow pointing to a variable but without starting from another variable indicates an influence by a random residual or error. The four rectangles corresponding to Y_i ($i = 1, 2, 3, 4$) and Z_j ($j = 1, 2$) are the data variables to be analyzed. The notation V999 in Figure 1 is the constant unit vector, and its effects on Z_1 and Z_2 represent the intercepts of these two variables. The constant factor loadings “1” at the path from F_W to Y_1 in the within model and at the path from F_B to Y_1 in the between model in Figure 1 are fixed as a reference for the purpose of model identification.

Now there are six observable variables Y_i ($i = 1, 2, 3, 4$) and Z_j ($j = 1, 2$) but only two free mean parameters $\mu_1 = E(Z_1)$ and $\mu_2 = E(Z_2)$. So we will have a nonsaturated mean structure in the model given by Figure 1. The exact mean and covariance structures corresponding to those given by (2) can be expressed as

$$\begin{aligned} \mu &= \begin{pmatrix} \mu_z \\ \mu_y \end{pmatrix}, \mu_z = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mu_y = (\mu_b, \phi_1\mu_b, \phi_2\mu_b, \phi_3\mu_b)', \\ \mu_b &= E(F_B) = \phi_{11}\mu_1 + \phi_{12}\mu_2, \\ \Sigma_w &= \Lambda_w \Phi_w \Lambda_w' + \Psi_w, \Lambda_w = (1, \theta_1, \theta_2, \theta_3)', \\ \Psi_w &= \text{diag}(\theta_5, \theta_6, \theta_7, \theta_8), \Phi_w = \text{var}(F_w) = (\theta_4), \\ \tilde{\Sigma}_B &= \text{cov} \begin{pmatrix} z_g \\ v_g \end{pmatrix} = \begin{pmatrix} \Sigma_{zz} & \Sigma_{zy} \\ \Sigma_{yz} & \Sigma_B \end{pmatrix}, \Sigma_B = \Lambda_B \Phi_B \Lambda_B' + \Psi_B, \\ \Lambda_B &= (1, \phi_1, \phi_2, \phi_3)', \\ \Phi_B(\phi) &= \text{var}(F_B), \Psi_B(\phi) = \text{diag}(\phi_4, \phi_5, \phi_6, \phi_7), \\ \Sigma_{zz} &= \begin{pmatrix} \phi_8 & \phi_9 \\ \phi_9 & \phi_{10} \end{pmatrix}, \Sigma_{zy} = \Sigma'_{yz} = (\text{cov}(Z_j, Y_i)), i=1, 2, 3, 4; j=1, 2, \end{aligned} \quad (27)$$

where

$$\begin{aligned}\text{var}(F_B) &= v_b = \phi_8 \phi_{11}^2 + \phi_{10} \phi_{12}^2 + 2\phi_9 \phi_{11} \phi_{12} + \phi_{13}, \\ \text{cov}(Z_1, Y_1) &= v_1 = \phi_8 \phi_{11} + \phi_9 \phi_{12}, \quad \text{cov}(Z_1, Y_2) = \phi_1 v_1, \\ \text{cov}(Z_1, Y_3) &= \phi_2 v_1, \quad \text{cov}(Z_1, Y_4) = \phi_3 v_1, \\ \text{cov}(Z_2, Y_1) &= v_2 = \phi_9 \phi_{11} + \phi_{10} \phi_{12}, \quad \text{cov}(Z_2, Y_2) = \phi_1 v_2, \\ \text{cov}(Z_2, Y_3) &= \phi_2 v_2, \quad \text{cov}(Z_2, Y_4) = \phi_3 v_2.\end{aligned}$$

The first Monte Carlo experiment is on the performance of the algorithm for nonlinear constraints on the parameters in the model given by Figure 1. These constraints are defined by:

$$\begin{aligned}2\theta_1 + \theta_2 - \theta_3^2 &= 0, & \phi_{11} - 2\phi_{12} &= 0, \\ \phi_1^2 + 2\phi_2 - \phi_3 &= 0, & \theta_1^2 - \phi_1 - \phi_2 &= 0.\end{aligned}\tag{28}$$

We use the following sampling designs to generate the two-level data.

1. Design D_1 : level 2 sample size $G = 60$, level-1 sample sizes $N_g = 4$ for $g = 1, \dots, 20$; $N_g = 6$ for $g = 21, \dots, 40$; $N_g = 8$ for $g = 41, \dots, 60$;
2. Design D_2 : level-2 sample size $G = 60$, level-1 sample sizes $N_g = 8$ for $g = 1, \dots, 20$; $N_g = 12$ for $g = 21, \dots, 40$; $N_g = 16$ for $g = 41, \dots, 60$;
3. Design D_3 : level-2 sample size $G = 120$, level-1 sample sizes $N_g = 4$ for $g = 1, \dots, 40$; $N_g = 6$ for $g = 41, \dots, 80$; $N_g = 8$ for $g = 81, \dots, 120$;
4. Design D_4 : level-2 sample size $G = 120$, level-1 sample sizes $N_g = 8$ for $g = 1, \dots, 40$; $N_g = 12$ for $g = 41, \dots, 80$; $N_g = 16$ for $g = 81, \dots, 120$;
5. Design D_5 : level-2 sample size $G = 240$, level-1 sample sizes $N_g = 4$ for $g = 1, \dots, 80$; $N_g = 6$ for $g = 81, \dots, 160$; $N_g = 8$ for $g = 161, \dots, 240$;
6. Design D_6 : level-2 sample size $G = 240$, level-1 sample sizes $N_g = 8$ for $g = 1, \dots, 80$; $N_g = 12$ for $g = 81, \dots, 160$; $N_g = 16$ for $g = 161, \dots, 240$.

For each of the sampling designs, we use formulation (1) with assumptions A1)–A5) to generate the two-level data. The level-1 data for $\mathbf{v}_{gi}$ in formulation (1) are generated according to the level-1 model in Figure 1, and the level-2 data for $\mathbf{v}_g$ and $\mathbf{z}_g$ in formulation (1) are generated according to the level-2 model in Figure 1. Then the two-level data $\{\mathbf{y}_{gi} : i = 1, \dots, N_g; g = 1, \dots, G\}$ are obtained by $\mathbf{y}_{gi} = \mathbf{v}_g + \mathbf{v}_{gi}$ as defined by formulation (1). The true values of the model parameters are chosen as in Table 1 and they satisfy the nonlinear constraints (28). The normal samples are generated by the internal normal random number generator in *MATLAB Code*. In the simulation, the constant sequences ξ_j and c_j [see (7)] in the algorithm for nonlinear constraints are chosen as: the initial values are

$\xi_j = (\xi_j^{(1)}, \xi_j^{(2)}, \xi_j^{(3)}, \xi_j^{(4)})' = (1, 1, 1, 1)'$ (there are four constraints) and $c_1 = 1$ ($j = 1$), then $c_{j+1} = 1.5c_j$ for $j = 1, 2, 3, \dots$, and $\xi_j = (\xi_j^{(1)}, \xi_j^{(2)}, \xi_j^{(3)}, \xi_j^{(4)})'$ is updated by

$$\xi_{j+1}^{(i)} = \xi_j^{(i)} + c_{j+1} \phi[h_i(\theta^{(j)})] \quad (i = 1, \dots, 4)$$

for iteration $j = 1, 2, 3, \dots$ with $\phi(t) = t^2/2$. For each of the six sampling designs, the simulation was carried out with 200 replications. We take the average ML estimation for

each parameter obtained from 200 replications and compute the standard deviation (S.D.) to show the accuracy of the estimation. The results are summarized in Table 1.

The second simulation experiment is to show the performance of the algorithm for linear constraints on the parameters in the model given by Figure 1. The linear constraints are defined by:

$$\begin{aligned} \theta_1 + 2\theta_2 - \theta_3 &= 2, & \theta_3 - \phi_3 &= 0.5, \\ \theta_1 - 2\phi_1 &= 0, & \phi_1 + \phi_2 + \phi_3 - \phi_{11} &= 0, \\ \theta_2 + \phi_2 &= 2, & \phi_{11} - 2\phi_{12} &= 0.5. \end{aligned} \quad (29)$$

The sampling designs D_i ($i = 1, \dots, 6$) are the same as in Table 1. The results from the EM algorithm for the linear constraint (29) are summarized in Table 2.

The simulation results in Tables 1–2 show that the performance of the algorithm for both linear and nonlinear parameter constraints is generally acceptable. By comparing the standard deviation (S.D.) row by row among the six sampling designs D_i ($i = 1, \dots, 6$), we can summarize the following two empirical conclusions from Tables 1–2:

1. The algorithm performs better for a larger level-1 sample size than for a smaller level-1 sample size in the sense that the S.D. becomes smaller for a larger N_g with the same G . For example, for all cases of the within parameter estimates, $S.D.(D_2) < S.D.(D_1)$, $S.D.(D_4) < S.D.(D_3)$, and $S.D.(D_6) < S.D.(D_5)$. The between parameter estimates are not generally improved (i.e., with smaller S.D.) when simply increasing the level-1 sample sizes N_g . This verifies the theoretical results in Yuan and Bentler (2006): the level-1 sample size mainly affects the standard errors of within parameter estimates;
2. The algorithm performs better for a larger level-2 sample size than for a smaller level-2 sample size in the sense that the S.D. becomes smaller for a larger G with the same N_g . For example, for all cases of the within and between parameter estimates, $S.D.(D_5) < S.D.(D_3) < S.D.(D_1)$, and $S.D.(D_6) < S.D.(D_4) < S.D.(D_2)$. Both within and between parameter estimates are generally improved (i.e., with smaller S.D.) when increasing the level-2 sample sizes G . This also verifies the theoretical results in Yuan and Bentler (2006): the level-2 sample size affects the standard errors of both within and between parameter estimates.

PRACTICAL ILLUSTRATION ON LINEAR CONSTRAINTS

In this section we illustrate the algorithm in Section 3 by setting up a two-level SEM with formulation (1) and the mean and covariance structures (2) and running the model with the proposed linear constraints on EQS 6.1 (Bentler, 2006). The real data set is from the National Education Longitudinal Study (NELS: 88) and the full data set is available from the authors upon request. The data set contains measurements of some variables for schools and for 5,198 students nested in 235 schools. There are 21 columns in the data set. We only take the data in columns 7–10 that record the students' scores from four math tests (level-1 observations, denoted by indicator variables M_i , $i = 1, 2, 3, 4$); the data in columns 11–14 that record the students' scores from four science tests (level-1 observations, denoted by indicator variables S_i , $i = 1, 2, 3, 4$); and the data in columns 20–21 that record the school-level (level-2) observations from two level-2 observable variables: Z_1 =minority and Z_2 =school type. It is empirically known that the students' scores are affected by three within-school (level-1) factors: F_{w1} =general math ability, F_{w2} =general science ability, and F_{w3} =general writing ability; and two between-school (level-2) factors: F_{b1} =general background and F_{b2} =math background. The two-level SEM is set up as:

1. The within (level-1) model is a factor model: both F_{w1} and F_{w3} generate (subject to independent measurement errors without further mention as follows) the four math tests. In addition, F_{w1} also influences the second science test S_2 . The factor loadings of M_1 on F_{w1} and F_{w3} are fixed as the constant “1” as reference for model identification. Both F_{w2} and F_{w3} influence all four science tests S_i ($i = 1, 2, 3, 4$). All three level-1 factors have unknown variances as free parameters, the covariances $\text{cov}(F_{w1}, F_{w3})$ and $\text{cov}(F_{w2}, F_{w3})$ are free parameters, but $\text{cov}(F_{w1}, F_{w2}) = 0$;
2. The between (level-2) model is not a factor model: F_{b1} influences all eight tests M_i and S_i ($i = 1, 2, 3, 4$). F_{b2} influences the four math tests M_i ($i = 1, 2, 3, 4$) and the science test S_2 . The factor loadings of M_1 on F_{b1} and F_{b2} are fixed as the constant “1” as reference for model identification. In addition, the two factors F_{b1} and F_{b2} are predicted by the two school-level variables Z_1 and Z_2 with unknown intercepts and correlated disturbance errors. The additional unknown intercepts are also imposed on M_3 and S_1 , respectively, according to some prior knowledge. The observations from both Z_1 and Z_2 are obtained subject to correlated random residuals. The following covariances are free parameters: the covariance between the two disturbance errors for F_{b1} and F_{b2} , the covariance between the two residuals for M_3 and for Z_2 , the covariance between the two residuals for Z_1 and for Z_2 .

The above measurement relationships can be visualized in Figure 2, using the same conventions as in Figure 1. Some nonzero covariances in the within and between models as mentioned above are released as free parameters based on some preliminary EQS runs by using the *Lagrange Multiplier Test* (command LMTEST in EQS) for adding parameters. Note: this test should not be confused with the use of augmented Lagrangians in optimization as discussed above. See for example, Buse (1982) for a discussion of this test, which is asymptotically equivalent to the chi-square difference test and the Wald test.

In the model given by Figure 2, there are a total of ten indicator variables M_i and S_i ($i = 1, 2, 3, 4$), and Z_1 and Z_2 . But there are only six free mean parameters: the two means from Z_1 and Z_2 , the two nonzero intercepts from the prediction equations of F_{b1} and F_{b2} , and the two nonzero intercepts from the prediction equations of M_3 and S_1 . Therefore we have a nonsaturated mean structure for this model. The mean and covariance structures for the model given by Figure 2 can be clearly expressed as some functions of the model parameters (indicated by the star sign “*” in Figure 2) like those in (27). Since some of the latent factors in the model have almost the same sets of indicator variables, equality of factor loadings on a given factor is used to help differentiate or identify the various factors. For example, the four variables loading on F_{w1} are constrained to have equal loadings. These considerations, along with some preliminary EQS runs, yield 10 equality constraints overall. Thus we set up the model of Figure 2 with EQS incorporating these 10 restrictions, which are shown in the Appendix in the sections labeled /CONSTRAINTS. The EQS output provides information on the statistical adequacy of these constraints as follows.

As expected, the p -values in Table 3 verify that all chi-square tests for the equality constraints are insignificant at the usual statistical levels (1%, 5% and 10%). So all null hypotheses of equality constraints should not be rejected, i.e., they are suitable for the data. The ML estimates with the above ten constraints are provided by the following EQS output for the measurement equations, where the number with the star sign “*” is an estimate for a factor loading parameter, the number with “*” before the variable V999 means that it is an estimate of an intercept or a mean parameter, $F1=F_{w1}$, $F2=F_{w2}$, and $F3=F_{w3}$ in the within model; $F1=F_{b1}$, and $F2=F_{b2}$ in the between model.

Measurement equations for the within model:

$$\begin{aligned}
 M1 &= V7 = 1.000F1 + 1.000F3 + 1.000E1 \\
 M2 &= V8 = 1.039F1 + 1.047F3 + 1.000E2 \\
 M3 &= V9 = 1.039F1 + .682F3 + 1.000E3 \\
 M4 &= V10 = 1.039F1 + 1.047F3 + 1.000E4 \\
 S1 &= V11 = 1.000F2 + .036F3 + 1.000E5 \\
 S2 &= V12 = 1.039F1 + .699F2 + .180F3 + 1.000E6 \\
 S3 &= V13 = .699F2 + .180F3 + 1.000E7 \\
 S4 &= V14 = .699F2 + .180F3 + 1.000E8
 \end{aligned} \tag{30}$$

Measurement equations for the between model:

$$\begin{aligned}
 M1 &= V7 = 1.000F1 + 1.000F2 + 1.000E1 \\
 M2 &= V8 = 1.070F1 + 1.276F2 + 1.000E2 \\
 M3 &= V9 = .596F1 + .454F2 + .655V999 + 1.000E3 \\
 M4 &= V10 = 1.123F1 + .498F2 + 1.000E4 \\
 S1 &= V11 = .961F1 + .324V999 + 1.000E5 \\
 S2 &= V12 = .855F1 + .332F2 + 1.000E6 \\
 S3 &= V13 = .855F1 + 1.000E7 \\
 S4 &= V14 = .855F1 + 1.000E8 \\
 Z1 &= V20 = 4.630V999 + 1.000E9 \\
 Z2 &= V21 = 1.170V999 + 1.000E10 \\
 F1 &= - .159V20 + 0.19V21 + 3.373V999 + 1.000D1 \\
 F2 &= - .061V20 - .048V21 - .107V999 + 1.000D2
 \end{aligned} \tag{31}$$

As can be seen above, the equality constraints imposed on the model were implemented during estimation, that is, the relevant parameter estimates are in fact equal. For example, the value 1.039 describes four of the within factor loadings (on F1, or F_{w1} in the diagram). The ML estimates for the variance-covariance parameters in the within and between models are summarized in Table 4.

In addition to the ML parameter estimates subject to the linear constraints in Table 3, EQS provides additional information on the model in Figure 2. An important statistic is the model chi-square=53.8 with 48 degrees of freedom and a p -value=.261. This implies that the model is suitable for the selected data. The model contains the nonzero covariances presented in Table 4, all of which are statistically significant at level 5%. The LMTEST furthermore showed that all remaining covariances between any two residuals are statistically insignificant at level 5%. This implies that it is not necessary to release more covariance parameters. Finally, the overall necessity for using a two-level model rather than a one-level model can be evaluated by the model-based intraclass correlations. In this example, variables M1–M4 and S1–S4 respectively have .200, .224, .067, .092, .167, .150, .150, and .064 proportions of the total (within plus between) variance. The between-school effects on individual performance clearly should not be ignored. The above interpretation on the output may not be easily understandable for the readers who are not familiar with structural equation models. A good reference is Bentler (2006).

CONCLUDING REMARKS

The algorithm in this paper is a mixture of the EM approach (Dempster, Laird, & Rubin, 1977) and the Lagrange multiplier method. There is no guarantee for the final solution to be the global maximum point for the likelihood function under the parameter constraints. A

theoretical discussion on the global convergence of this kind of mixed algorithm is beyond the scope of this paper. We are also unable to provide a comparison between our algorithm and some possible alternative ones, as we have not found parallel algorithms for handling the same model in the literature. However, the simulation experiments and the practical illustration provide some evidence that the current approach is feasible. Since complicated functional constraints with mean and covariance structures are not usually proposed in two-level statistical models, our evaluation was limited to simulation results. Linear constraints, especially equality constraints, are much more frequent in application, and we were able to illustrate them with the real school data set using the publicly available EQS program (Bentler, 2006). As a result, anyone with a working knowledge of EQS can easily run the program given in the Appendix, or, by referring to the EQS manual (Bentler, 2006), they can write their own setups for analysis of two-level SEM with both mean and covariance structures subject to general linear constraints on parameters.

This paper is focused on developing an algorithm for estimating parameters subject to general constraints in more complicated two-level SEM. Asymptotic properties such as the asymptotic normality and standard errors of the parameter estimates, and goodness-of fit test for justifying the imposed constraints can also be developed. Due to limited space of this paper, we will guide interested readers to our further research.

Acknowledgments

The research was supported by National Institute on Drug Abuse grant DA01070, a grant from the Hong Kong Baptist University (project number FRG/07-08/II-35), and University of New Haven 2009 and 2010 Summer Research Grants.

REFERENCES

- Aitchison J, Silvey SD. Maximum likelihood estimation of parameters subject to restraints. *The Annals of Mathematical Statistics*. 1958; 29:813–828.
- Bentler, PM. EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc.; 2006. ISBN: 1-885889-03-7 (www.mvsoft.com).
- Bertsekas DP. Multiplier method: A survey. *Automatica*. 1976; 12:133–145.
- Buse A. The likelihood ratio, Wald and Lagrange multiplier tests: An expository note. *American Statistician*. 1982; 36:153–157.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via EM algorithm (with discussion). *Journal of the Royal Statistical Society (Series B)*. 1977; 39:1–38.
- du Toit, S.; du Toit, M. Multilevel structural equation modeling. In: de Leeuw, J.; Meijer, E., editors. *Handbook of Multilevel Analysis* (Chapter 12). New York: Springer Verlag; 2008.
- Jamshidian M. On algorithms for restricted maximum likelihood estimation. *Computational Statistics & Data Analysis*. 2004; 45:137–157.
- Kim DK, Taylor JMG. The restricted EM algorithm for maximum likelihood estimation under linear restrictions on parameters. *Journal of the American Statistical Association*. 1995; 90:708–716.
- Lange K. A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society (Series B)*. 1995a; 57:425–437.
- Lange K. A Quasi-Newton acceleration of the EM algorithm. *Statistica Sinica*. 1995b; 5:1–8.
- Lee SY. Constrained estimation in covariance structure analysis. *Biometrika*. 1979; 66:539–545.
- Lee SY, Poon WY. Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica*. 1998; 8:749–766.
- Lee SY, Tsang SY. Constrained maximum likelihood estimation of two-level covariance structure model via EM type algorithms. *Psychometrika*. 1999; 64:435–450.
- Liang J, Bentler PM. A new EM algorithm for fitting two-level structural equation models. *Psychometrika*. 2004; 69:101–122.

- McDonald RP, Goldstein H. Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical & Statistical Psychology*. 1989; 42:215–232.
- McLachlan, GJ.; Krishnan, T. *The EM Algorithm and Extensions*. New York: Wiley; 1997.
- Muthén BO. Multilevel covariance structure analysis. *Sociological Methods & Research*. 1994; 22:376–398.
- Muthén, LK.; Muthén, BO. *Mplus Version 3 User's Guide*. Los Angeles: Muthén & Muthén; 2004.
- Raudenbush SW. Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical & Statistical Psychology*. 1995; 48:359–370.
- Yuan K-H, Bentler PM. Asymptotic robustness of standard errors in multilevel structural equation models. *Journal of Multivariate Analysis*. 2006; 97:1121–1141.
- Yung YF, Bentler PM. On added information for ML factor analysis with mean and covariance structures. *Journal of Educational & Behavioral Statistics*. 1999; 24:1–20.

APPENDIX

EQS Input Program for the model in Figure 2

```

/TITLE

Two-level analysis for the school data

(Two factors in the between model, three factors in the within model, 8 y-variables)

WITHIN MODEL FIRST

/SPECIFICATION

data='school.dat'; case =5198; variable=21; method=ml;

matrix=raw; GROUP=2; analysis=covariance; MULTILEVEL=ML; CLUSTER= V19;

/LABELS

V7=M1; V8=M2; V9=M3; V10=M4; V11=S1; V12=S2; V13=S3; V14=S4;

V19=SCHOOL; V20=Z1; V21=Z2;

! F1=Fw1=math ability factor; F2=Fw2=science ability factor;

! F3=Fw3=general writing ability factor

/EQUATIONS

M1=1F1+1F3+E1; M2=*F1+*F3+E2; M3=*F1+*F3+E3; M4=*F1+*F3+E4;

S1=1F2+*F3+E5; S2=*F1+*F2+*F3+E6; S3=*F2+*F3+E7; S4=*F2+*F3+E8;

/VARIANCES

E1-E8=*; F1=*; F2=*; F3=*;

/COVARIANCES

F1,F2=0; F2,F3=0*; F1,F3=0*;

```

/CONSTRAINTS

(M2,F1)=(M3,F1)=(M4,F1)=(S2,F1); (M2,F3)=(M4,F3);

(S2,F3)=(S3,F3)=(S4,F3); (S2,F2)=(S3,F2)=(S4,F2);

/END

/TITLE

BETWEEN MODEL

/LABELS

V7=M1; V8=M2; V9=M3; V10=M4; V11=S1; V12=S2; V13=S3; V14=S4;

V19=SCHOOL; V20=Z1; V21=Z2;

! F1=Fb1==general background factor; F2=Fb2==math background factor;

/EQUATIONS

M1=1F1+1F2+E1; M2=*F1+*F2+E2; M3=*V999+*F1+*F2+E3; M4=*F1+*F2+E4;

S1=*V999+*F1+E5; S2=*F1+*F2+E6; S3=*F1+E7; S4=*F1+E8;

F1=*V999+*Z1+*Z2+D1; F2=*V999+*Z1+*Z2+D2; Z1=*V999+E9; Z2=*V999+E10;

/VARIANCES

E1-E10=0*; D1-D2=*;

/COVARIANCES

D1,D2=0*; E3,E10=0*; E9,E10=0*;

/CONSTRAINTS

(S2,F1)=(S3,F1)=(S4,F1);

/tech

itr=200; con=.000001;

/LMTEST

set=pee;

/END

FIGURE 1.
Path diagram of the model in the simulation study

FIGURE 2.
Path diagram of the model for the selected school data

TABLE 1

Average Estimates with Nonlinear Constraints

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
True Value	1.0000	2.0000	2.0000	0.1600	0.2500	0.2500	0.2500	0.2500
A.E.(D_1)	0.9977	1.9819	1.9935	0.1640	0.2466	0.2518	0.2543	0.2454
A.E.(D_2)	1.0036	2.0067	2.0033	0.1604	0.2485	0.2479	0.2481	0.2491
A.E.(D_3)	0.9991	1.9890	1.9966	0.1601	0.2486	0.2480	0.2535	0.2493
A.E.(D_4)	1.0004	2.0123	2.0032	0.1585	0.2494	0.2492	0.2488	0.2494
A.E.(D_5)	1.0001	2.0015	2.0003	0.1593	0.2505	0.2507	0.2500	0.2496
A.E.(D_6)	1.0006	2.0029	2.0009	0.1594	0.2480	0.2499	0.2508	0.2500
S.D.(D_1)	0.0372	0.1687	0.0526	0.0201	0.0182	0.0228	0.0415	0.0348
S.D.(D_2)	0.0218	0.0911	0.0273	0.0129	0.0161	0.0155	0.0266	0.0245
S.D.(D_3)	0.0174	0.1037	0.0295	0.0139	0.0179	0.0157	0.0296	0.0266
S.D.(D_4)	0.0159	0.0662	0.0200	0.0092	0.0117	0.0109	0.0168	0.0141
S.D.(D_5)	0.0123	0.0686	0.0192	0.0082	0.0107	0.0106	0.0179	0.0166
S.D.(D_6)	0.0135	0.0479	0.0154	0.0066	0.0082	0.0081	0.0124	0.0113

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
True Value	0.5000	0.5000	1.2500	0.3600	0.3600	0.3600	0.3600	0.4900
A.E.(D_1)	0.4899	0.5068	1.2595	0.3621	0.3720	0.3760	0.3429	0.4851
A.E.(D_2)	0.5074	0.5002	1.2596	0.3626	0.3500	0.3611	0.3461	0.4788
A.E.(D_3)	0.5003	0.4981	1.2475	0.3578	0.3565	0.3753	0.3467	0.4887
A.E.(D_4)	0.4982	0.5030	1.2550	0.3488	0.3643	0.3573	0.3676	0.4910
A.E.(D_5)	0.5002	0.5001	1.2508	0.3614	0.3589	0.3662	0.3596	0.4885
A.E.(D_6)	0.5011	0.5003	1.2522	0.3477	0.3621	0.3520	0.3622	0.4876
S.D.(D_1)	0.0765	0.0608	0.1089	0.1119	0.1101	0.0985	0.1295	0.0862
S.D.(D_2)	0.0434	0.0314	0.0628	0.0759	0.0766	0.0853	0.0982	0.0867
S.D.(D_3)	0.0306	0.0246	0.0521	0.0584	0.0653	0.0551	0.0844	0.0572
S.D.(D_4)	0.0273	0.0245	0.0497	0.0537	0.0489	0.0644	0.0741	0.0589

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
True Value	0.5000	0.5000	1.2500	0.3600	0.3600	0.3600	0.3600	0.4900
S.D.(D_3)	0.0212	0.0167	0.0366	0.0478	0.0421	0.0440	0.0575	0.0460
S.D.(D_6)	0.0238	0.0161	0.0374	0.0439	0.0387	0.0397	0.0490	0.0429

	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{13}	μ_1	μ_2
True Value	0.3000	0.4900	1.0000	0.5000	0.1600	1.0000	1.0000
A.E.(D_1)	0.2925	0.4908	0.9965	0.4983	0.1500	1.0030	1.0131
A.E.(D_2)	0.2899	0.4751	0.9911	0.4956	0.1470	1.0066	0.9977
A.E.(D_3)	0.3028	0.4974	1.0052	0.5026	0.1599	0.9973	0.9910
A.E.(D_4)	0.2985	0.4854	0.9918	0.4959	0.1594	1.0115	1.0037
A.E.(D_5)	0.2994	0.4918	0.9994	0.4997	0.1563	0.9970	0.9960
A.E.(D_6)	0.3020	0.4916	0.9982	0.4991	0.1621	1.0037	1.0008
S.D.(D_1)	0.0746	0.0927	0.0698	0.0349	0.0660	0.0851	0.0924
S.D.(D_2)	0.0747	0.0894	0.0491	0.0246	0.0629	0.0986	0.0872
S.D.(D_3)	0.0509	0.0680	0.0376	0.0188	0.0516	0.0606	0.0562
S.D.(D_4)	0.0523	0.0637	0.0369	0.0184	0.0445	0.0575	0.0592
S.D.(D_5)	0.0362	0.0443	0.0307	0.0154	0.0345	0.0433	0.0435
S.D.(D_6)	0.0383	0.0487	0.0284	0.0142	0.0287	0.0451	0.0502

Note: In Table 1, A.E.(D_i)=Average Estimate from 200 simulation replications, $i = 1, \dots, 6$; S.D.(D_i)=Standard Deviation from 200 simulation replications, $i = 1, \dots, 6$. RMSE in (23) is $\leq 10^{-4}$.

TABLE 2

Average Estimates with Linear Constraints

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
True Value	1.0000	1.0000	1.0000	1.0000	0.2500	0.2500	0.2500	0.2500
A.E.(D_1)	0.9997	0.9994	0.9986	1.0073	0.2459	0.2496	0.2508	0.2500
A.E.(D_2)	1.0002	0.9993	0.9988	1.0008	0.2486	0.2495	0.2496	0.2495
A.E.(D_3)	0.9981	0.9996	0.9974	1.0004	0.2495	0.2510	0.2504	0.2479
A.E.(D_4)	0.9988	1.0000	0.9988	0.9980	0.2495	0.2490	0.2495	0.2498
A.E.(D_5)	0.9978	1.0016	1.0011	1.0017	0.2499	0.2517	0.2515	0.2499
A.E.(D_6)	0.9994	1.0003	0.9999	1.0023	0.2503	0.2503	0.2502	0.2491
S.D.(D_1)	0.0212	0.0116	0.0179	0.0896	0.0220	0.0271	0.0261	0.0252
S.D.(D_2)	0.0150	0.0092	0.0145	0.0598	0.0168	0.0191	0.0209	0.0172
S.D.(D_3)	0.0145	0.0071	0.0125	0.0627	0.0184	0.0195	0.0185	0.0181
S.D.(D_4)	0.0108	0.0059	0.0093	0.0450	0.0133	0.0121	0.0124	0.0113
S.D.(D_5)	0.0102	0.0052	0.0091	0.0406	0.0134	0.0146	0.0142	0.0133
S.D.(D_6)	0.0081	0.0042	0.0069	0.0316	0.0100	0.0089	0.0097	0.0095

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
True Value	0.5000	1.0000	0.5000	0.1600	0.1600	0.1600	0.1600	0.3600
A.E.(D_1)	0.4999	1.0006	0.4986	0.1655	0.1651	0.1663	0.1561	0.3567
A.E.(D_2)	0.5001	1.0007	0.4988	0.1614	0.1539	0.1621	0.1575	0.3549
A.E.(D_3)	0.4990	1.0004	0.4974	0.1621	0.1586	0.1593	0.1573	0.3535
A.E.(D_4)	0.4994	1.0000	0.4988	0.1623	0.1610	0.1553	0.1613	0.3575
A.E.(D_5)	0.4989	0.9984	0.5011	0.1554	0.1605	0.1625	0.1624	0.3637
A.E.(D_6)	0.4997	0.9997	0.4999	0.1603	0.1547	0.1619	0.1606	0.3482
S.D.(D_1)	0.0106	0.0116	0.0179	0.0591	0.0481	0.0621	0.0488	0.0634
S.D.(D_2)	0.0075	0.0092	0.0145	0.0501	0.0392	0.0495	0.0392	0.0655
S.D.(D_3)	0.0073	0.0071	0.0125	0.0369	0.0345	0.0387	0.0324	0.0457
S.D.(D_4)	0.0054	0.0059	0.0093	0.0378	0.0313	0.0394	0.0285	0.0417

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8
True Value	0.5000	1.0000	0.5000	0.1600	0.1600	0.1600	0.1600	0.3600
S.D.(D_3)	0.0051	0.0052	0.0091	0.0261	0.0231	0.0271	0.0222	0.0328
S.D.(D_6)	0.0041	0.0042	0.0069	0.0266	0.0204	0.0231	0.0193	0.0338

	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}	ϕ_{13}	μ_1	μ_2
True Value	0.2000	0.3600	2.0000	0.7500	0.2500	1.0000	1.0000
A.E.(D_1)	0.1951	0.3580	1.9990	0.7495	0.2444	1.0006	1.0114
A.E.(D_2)	0.2008	0.3569	1.9997	0.7498	0.2391	1.0057	1.0014
A.E.(D_3)	0.1983	0.3589	1.9968	0.7484	0.2466	1.0001	0.9963
A.E.(D_4)	0.2021	0.3646	1.9982	0.7491	0.2526	1.0067	1.0103
A.E.(D_5)	0.2007	0.3540	1.9983	0.7492	0.2545	1.0000	1.0027
A.E.(D_6)	0.1913	0.3539	1.9994	0.7497	0.2487	1.0055	1.0095
S.D.(D_1)	0.0553	0.0700	0.0254	0.0127	0.0978	0.0769	0.0801
S.D.(D_2)	0.0552	0.0695	0.0180	0.0090	0.0706	0.0790	0.0781
S.D.(D_3)	0.0384	0.0484	0.0183	0.0091	0.0611	0.0538	0.0565
S.D.(D_4)	0.0384	0.0469	0.0131	0.0065	0.0570	0.0492	0.0490
S.D.(D_5)	0.0207	0.0291	0.0129	0.0064	0.0493	0.0380	0.0426
S.D.(D_6)	0.0277	0.0372	0.0100	0.0050	0.0373	0.0345	0.0323

Note: In Table 2, A.E.(D_i)=Average Estimate from 200 simulation replications, $i = 1, \dots, 6$; S.D.(D_i)=Standard Deviation from 200 simulation replications, $i = 1, \dots, 6$. RMSE in (23) is $\leq 10^{-4}$.

TABLE 3

LMTEST for the Equality Constraints

CONSTRAINTS FROM GROUP 1 (within model)			
CONSTRAINT #	CONSTRAINT	$\chi^2(1)$ -statistic	p-Value
CONSTR: 1	(M2,F1)–(M3,F1)=0;	.001	.979
CONSTR: 2	(M2,F1)–(M4,F1)=0;	.000	.985
CONSTR: 3	(M2,F1)–(S2,F1)=0;	.288	.592
CONSTR: 4	(M2,F3)–(M4,F3)=0;	.024	.876
CONSTR: 5	(S2,F3)–(S3,F3)=0;	.033	.857
CONSTR: 6	(S2,F3)–(S4,F3)=0;	.033	.857
CONSTR: 7	(S2,F2)–(S3,F2)=0;	.001	.972
CONSTR: 8	(S2,F2)–(S4,F2)=0;	.003	.956
CONSTRAINTS FROM GROUP 2 (between model)			
CONSTR: 9	(S2,F1)–(S3,F1)=0;	.849	.357
CONSTR: 10	(S2,F1)–(S4,F1)=0;	.849	.357

Note: In Table 3, $F1=F_{w1}$, $F2=F_{w2}$, and $F3=F_{w3}$ in the within model; $F1=F_{b1}$, and $F2=F_{b2}$ in the between model. EQS notation such as (M2,F1) denotes the factor loading parameter of the second math test M_2 on the first within-level factor $F1=F_{w1}$ =general math ability.

TABLE 4

ML Estimates for the Variance-Covariance Parameters

Within Model Variances-Covariances Parameter Estimates												
	E1	E2	E3	E4	E5	E6	E7	E8	F1	F2	F3	
Variance	.344	.257	1.328	2.463	.605	.629	.607	2.403	.153	.495	.887	
Covariance	cov(F1,F3)=-.146 cov(F2,F3)=-.464											
Between Model Variances-Covariances Parameter Estimates												
	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	D1	D2
Variance	.016	.004	.033	.037	.017	.028	.009	.025	4.437	.142	.114	.023
Covariance	cov(D1,D2)=-.005 cov(E3,E10)=-.024 cov(E9,E10)=-.365											

Note: In Table 4, the residuals E1, E2, ..., E8, and the three level-1 factors F1, F2, and F3 in the within model are the same as those in the measurement equations given by (30); the residuals E1, E2, ..., E10, and the two disturbances D1 and D2 in the between model are the same as those in the measurement equations given by (31). All nonzero covariances are statistically significant at level 5% based on the individual t-test provided by EQS.