Computationally Efficient Estimation of Multilevel High-Dimensional Latent Variable Models

Tihomir Asparouhov[1], Bengt Muthen[2]
Muthen & Muthen[1]
UCLA[2]

Abstract

Multilevel analysis often leads to modeling with multiple latent variables on several levels. While this is less of a problem with Gaussian observed variables, maximum-likelihood (ML) estimation with categorical outcomes presents computational problems due to multidimensional numerical integration. We describe a new method that compared to ML is both computationally efficient and has similar MSE. The method is an extension of the Muthen (1984) weighted least squares (WLS) estimation method to multilevel multivariate latent variable models for any combination of categorical, censored, and normal observed variables. Using a new version of the Mplus program, we compare MSE and the computational time for the ML and WLS estimators in a simulation study.

KEY WORDS: Multilevel Models, Structural Equation Models, Weighted Least Square Estimation

1. Overview

Univariate and multivariate multilevel models with normally distributed dependent variables can be estimated with the maximum-likelihood estimator via the EM algorithm, see Raudenbush and Bryk (2002). This estimation is feasible even when the number of random effects is large and is implemented in the software package Mplus (Muthen & Muthen 1998-2007). For example, a multivariate random intercept model with p observed dependent variables has p random intercepts which are estimated simultaneously, so that the joint distribution of the p random effects is estimated.

When the dependent variables are not normally distributed, for example, ordered polytomous variables or censored variables, the multilevel models can be estimated by the EM algorithm however all random effects have to be numerically integrated, see Muthen and Asparouhov (2006). The numerical integration method is computationally very demanding when the number of random effects is large. In practical applications, one and two random effect models are straight forward to estimate, three and four random effect models are still feasible but typically will take a long time to complete. Models with five random effects and more can sometimes be estimated with numerical integration by using a smaller number of integration points for each dimension or by using monte-carlo integration, however in general such model estimation will typically have convergence problems and lack of precision in the estimates as well as substantial computational time which makes this approach not very practical. Some special models such as block-diagonal random effect variance/covariance models, which utilize the Cholesky parameterization for the random effects, implemented in Mplus, see Hedeker and Gibbons (1996), can be estimated via the EM algorithm with non-adaptive integration methods. In such settings the EM algorithm will have stable convergence regardless of the number of random effects and integration points per dimension, however the precision of the estimation still depends on the number of integration points. Bayesian estimation methods such as MCMC can also be used however they are usually very computationally demanding and similar in performance to the maximum-likelihood method with monte-carlo integration.

In this article we present a limited-information weighted least squares estimation method that can be used to estimate two-level latent variable models with binary, ordered polytomous, continuous and censored variables as well as combinations of such variables. The method can be used to estimate models with any number of random effects without increase in the computational time, i.e., the computational time is virtually independent of the number of random effects in the model. In addition the precision of the estimation is not compromised by the number of random effects. This method essentially replaces a complex model estimation with high dimensional integration by multiple simple models with one and two dimensional integration. The method is a direct generalization of Muthen (1984) weighted least squares estimation for single level models and is implemented in the upcoming release of Mplus 5.0. A number of models that would be of great importance in practice are now feasible to estimate. For example when the dependent variables are binary and ordered polytomous new feasible models are two-level structural equation models, three level growth models, exploratory and confirmatory factor analysis with multiple factors on both levels, unrestricted two level means and variance/covariance models that can also be used for comparison for structural models.

2. The Model

Let y_{pij} be the $p-$th observed dependent variable, $p = 1, ..., P$, for individual i, $i = 1, ..., N_j$, in cluster $j = 1, ..., C$, where N_j is the number of individuals in cluster j. Let x_{wqij} be the $q-$th observed individual predictor variable, $q = 1, ..., Q_1$, for individual i in cluster j and x_{bqj} be the $q-$th observed cluster level predictor variable, $q = 1, ..., Q_2$, observed for cluster j . We consider three types of dependent variables, categorical, which includes ordered polytomous and binary variables, normally distributed continuous variables and censored variables. To construct the latent variable model we proceed as in Muthen (1984) by defining an underlying normally distributed latent variable y^*_{pij}. If the p-th variable is normally distributed then $y^*_{pij} = y_{pij}$. If the p-th variable is categorical, for a set of parameters τ_{pk}

$$y_{pij} = k \Leftrightarrow \tau_{pk} < y^*_{pij} < \tau_{pk+1}. \quad (1)$$

If the p-th variable is censored at the censoring point c_p

$$y_{pij} = \begin{cases} y^*_{pij} & \text{if } y^*_{pij} > c_p \\ c_p & \text{if } y^*_{pij} \leq c_p \end{cases}. \quad (2)$$

A linear regression for y^*_{pij} is thus equivalent to a linear, probit and censored regression for y_{pij}, depending on the type of the $p-$th variable.

The two-level model is constructed as in Muthen (1994)

$$y^*_{pij} = y_{wpij} + y_{bpj} \quad (3)$$

where y_{wpij} and y_{bpj} are normally distributed independent latent variables. The interpretation of equation (3) is that the latent variable y^*_{pij} is simply composed of a cluster level effect y_{bpj}, i.e., random intercept, and an individual effect y_{wpij}.

Two separate latent variable models are defined for y_{wpij} and y_{bpj}. Suppose that η_{wmij} are normally distributed latent variables defined on the individual level, $m = 1, ...M_1$ and η_{bmj} are normally distributed latent variables defined on the cluster level, $m = 1, ...M_2$. We define the vector variables $y_{wij} = (y_{w1ij}, ..., y_{wPij})$ and similarly y_{bj}, η_{wij}, η_{bj}, x_{wij}, and x_{bj}. The structural model on the within (individual) level is given by

$$y_{wij} = \Lambda_w \eta_{wij} + \varepsilon_{wij} \quad (4)$$

$$\eta_{wij} = B_w \eta_{wij} + \Gamma_w x_{wij} + \xi_{wij}. \quad (5)$$

The structural model on the between (cluster) level is given by

$$y_{bj} = \nu_b + \Lambda_b \eta_{bj} + \varepsilon_{bj} \quad (6)$$

$$\eta_{bj} = \alpha_b + B_b \eta_{bj} + \Gamma_b x_{bj} + \xi_{bj}. \quad (7)$$

The vector and matrix parameters Λ_w, Γ_w, B_w, Λ_b, Γ_b, B_b, ν_b, α_b as well as the thresholds parameters τ_{pk} are to be estimated. Not all of these parameters are identified, different parameters are fixed to 0 to obtain various structural models. The residual variables ε_{wij}, ξ_{wij}, ε_{bj} and ξ_{bj} are zero mean normally distributed independent vector variables with full variance/covariance matrices Θ_w, Ψ_w, Θ_b, Ψ_b respectively. For identification purposes the variance of ε_{wpij} is fixed to 1 if the $p-$th variable is categorical. Note that the maximum-likelihood method described in Muthen and Asparouhov (2006) for this model requires that Θ_w is diagonal. This restriction is not needed for the new estimation method. In single level models there are two separate parameterizations for this model, which are refereed to as the delta and the theta parameterizations, see Muthen & Asparouhov (2002). In this two level model we use only the theta parameterization, which means that the parameters for the categorical variables are defined on the scale where the residual variance is 1.

3. The Unrestricted Two-Level Model

The weighted least squares estimation described in this article is based on the estimation of the following saturated model. For categorical variables we define the threshold parameters t_{pk} by

$$y_{pij} = k \Leftrightarrow t_{pk} < y^*_{pij} < t_{pk+1}. \quad (8)$$

The multilevel decomposition is given again by

$$y^*_{pij} = y_{wpij} + y_{bpj}. \quad (9)$$

The structural part of the model is defined by

$$y_{wij} = \Pi_w x_{wij} + \epsilon_{wij} \quad (10)$$

$$y_{bj} = \mu_b + \Pi_b x_{bj} + \epsilon_{bj}. \quad (11)$$

The residual variables ϵ_{wij} and ϵ_{bj} are normally distributed zero mean independent variables with full variance/covariance matrix Σ_w and Σ_b respectively. For categorical variables the variance of ϵ_{wpij} is fixed to 1 and the mean parameter μ_{pb} is fixed to 0 for identification purposes.

The estimation of this model is a two stage limited-information estimation. In the first stage we estimate the $p-$th univariate model using the two-level maximum-likelihood method as in Muthen and Asparouhov (2006), i.e., the parameters Π_{wpq}, $q = 1, ...Q_1$; Π_{bpq}, $q = 1, ...Q_2$; μ_{bp}, t_{pk}, Σ_{wpp} and Σ_{bpp}. Performing all P univariate estimations we obtain all estimates for all parameters in the above model except for the off diagonal estimates of Σ_w and Σ_b. In the second stage we estimate every pair of bivariate models by fixing the univariate parameters to their first stage estimates. Thus for the bivariate model of variables p_1 and p_2 we need to estimate just two parameters $\Sigma_{wp_1p_2}$ and $\Sigma_{bp_1p_2}$. This second stage estimation is performed again by multilevel ML estimation described in Muthen and Asparouhov (2006), facilitating the maximization of the bivariate likelihood described in Olsson (1979) and Olsson et alt. (1982). The univariate and bivariate estimation uses numerical integration for each variable that is not normally distributed. For example the univariate estimation of a categorical variable uses one dimensional numerical integration, while the univariate estimation of a normal variable uses zero dimensions and follows the algorithm described in Raudenbush & Bryk (2002). The bivariate estimation of two normal variables uses zero dimensions of numerical integration. The bivariate estimation of two categorical variables uses two dimensional numerical integration. The bivariate integration of a normal variable and a categorical variable uses one dimensional integration.

The asymptotic covariance of the the first and second stage estimates of the parameters in model (8-11) are computed as in Muthen & Satorra (1995) for single level models and are based on the first derivatives of the first stage likelihood for the univariate parameters, the first derivatives of the second stage likelihood for the bivariate parameters as well as the first derivatives of the second stage likelihood for the univariate parameters to account for the dependence of the second stage estimates on the first stage estimates. Let's denote the vector of all parameter estimates of the unrestricted model as s and their asymptotic covariance by G. Arguments as in Muthen & Satorra (1995) show that these estimates are consistent.

4. Estimating the Structural Model

In this section we describe the estimation of the structural model (1-7). This estimation is based on the Muthen (1984) method. Notice that the structural model (1-7) can be viewed as a restricted model nested within the unrestricted (8-11). That is because model (1-7) implies the following model, beginning with the categorical variable model

$$y_{pij} = k \Leftrightarrow t^*_{pk} < y^*_{pij} < t^*_{pk+1}. \qquad (12)$$

$$y_{wij} = \Pi^*_w x_{wij} + \epsilon_{wij} \qquad (13)$$

$$y_{bj} = \mu^*_b + \Pi^*_b x_{bj} + \epsilon_{bj}. \qquad (14)$$

where the variance/covariance matrix ϵ_{wij} and ϵ_{bj} are Σ^*_w and Σ^*_b respectively. The unstandardized estimates implied by (1-7) are

$$\Sigma^{**}_w = \Lambda_w (I - B_w)^{-1} \Psi_w (I - B_w)^{-1\ T} \Lambda^T_w + \Theta_w \qquad (15)$$

$$\Sigma^{**}_b = \Lambda_b (I - B_b)^{-1} \Psi_b (I - B_b)^{-1\ T} \Lambda^T_b + \Theta_b \qquad (16)$$

$$\Pi^{**}_w = \Lambda_w (I - B_w)^{-1} \Gamma_w \qquad (17)$$

$$\Pi^{**}_b = \Lambda_b (I - B_b)^{-1} \Gamma_b \qquad (18)$$

$$\mu^{**}_b = \nu_b + \Lambda_b (I - B_b)^{-1} \alpha_b \qquad (19)$$

where I is the identity matrix. Let Δ_w be a diagonal matrix of dimension P with 1 on diagonal if the p-th variable is not categorical and $1/\sqrt{\Sigma^{**}_{wpp}}$ if the variable is categorical. Let δ_b be a P dimensional vector with the $p-$th entry 0 if the $p-$th variable is not categorical, and μ^{**}_{bp} if the $p-$th variable is categorical.

Thus the standardized estimates implied by (1-7) are

$$\Sigma^*_w = \Delta_w \Sigma^{**}_w \Delta_w \qquad (20)$$

$$\Sigma^*_b = \Delta_w \Sigma^{**}_b \Delta_w \qquad (21)$$

$$\Pi^*_w = \Delta_w \Pi^{**}_w \qquad (22)$$

$$\Pi^*_b = \Delta_w \Pi^{**}_b \qquad (23)$$

$$\mu^*_b = \Delta_w (\mu^{**}_b - \delta_b) \qquad (24)$$

$$t^*_{pk} = \Delta_w (\tau_{pk} - \delta_b). \qquad (25)$$

The difference between the standardized and the unstandardized is that for categorical variables we have standardized the variance on the within level to 1 and the mean on the between level 0. This is needed so that the structural and the unrestricted models are compared on the same scale. Let s^* be a vector of all standardized estimates Σ^*_w, Σ^*_b, Π^*_w, Π^*_b, μ^*_b, t^*_{pk} in the same order as the unrestricted parameters are placed in s. Let W be the weight matrix with the same dimension as the vector s. Define the fit function as

$$F = (s - s^*) W (s - s^*)^T \qquad (26)$$

Minimizing the fit function with respect to the parameters of model (1-7) is the last stage of the estimation process. The weighted least square estimates are the parameter estimates that minimize F. Asymptotic covariance for these estimates are obtained as in Muthen & Satorra (1995) as well as a chi-square test comparing the restricted model to the unrestricted model.

Using a different weight matrix W we obtain different estimators. If W is the identity matrix we get the ULS estimator. If $W = G^{-1}$ we get the WLS estimator. Frequently the unrestricted model has a large number of parameters and the size of G is larger than the number of clusters in the sample. This leads to singular G matrix and thus the WLS estimator would be undefined. In such cases we can set $W = G_0^{-1}$ where G_0 is the same as G for all diagonal entries and it's zero for all off-diagonal entries. In this case we get the diagonal WLS estimator, which in Mplus is given by the the WLSM and WLSMV estimators, the difference between the two is how the chi-square is computed. For the WSLM estimator the chi-square is computed using a first order correction of the fit function F while the WSLMV uses second order correction. Further discussion on the advantages of the different choices of W is available in Muthen et alt. (1997) for single level models, however all results apply for the two level model as well.

5. Simulation Study

In this section we conduct a simple simulation study to examine the performance of the WLSM estimator and to compare this estimator to the ML estimator. We use a confirmatory factor analysis model with 2 factors on the within and the between level. There are 6 dependent variables in the model, the first 3 are measurements for the first factor on the between and the within level and the last 3 variables are measurements for the second factor on

Table 1: Two-level factor analysis model with categorical variables.

parameter	true value	WLSM bias	ML bias	WLSM coverage	ML coverage	Efficiency ratio
λ_{w2}	1.0	3%	2%	97%	100%	1.14
ψ_{w12}	0.4	2%	-14%	97%	89%	0.89
ψ_{w11}	0.7	2%	-23%	94%	75%	0.71
λ_{b2}	1.0	5%	4%	96%	94%	0.96
ψ_{b12}	0.2	-1%	-22%	94%	81%	0.91
ψ_{b11}	0.4	1%	-31%	93%	57%	0.77
τ_{11}	-0.3	-3%	-6%	96%	87%	1.17
τ_{12}	0.4	-1%	-14%	96%	81%	1.00
τ_{13}	1.2	0%	-11%	95%	55%	0.71
τ_{14}	1.8	0%	-10%	98%	47%	0.56
θ_{b1}	0.2	-2%	-55%	97%	32%	0.66

both levels as well. We conduct the two simulation studies, one with ordered polytomous with 5 categories and one with continuous variables. The model is described by the following equations. For $p = 1, ..., 3$

$$y_{wpij} = \lambda_{wp}\eta_{w1ij} + \varepsilon_{wpij}$$

$$y_{bpj} = \lambda_{bp}\eta_{b1j} + \varepsilon_{bpj}.$$

For $p = 4, ..., 6$

$$y_{wpij} = \lambda_{wp}\eta_{w2ij} + \varepsilon_{wpij}$$

$$y_{bpj} = \nu_p + \lambda_{bp}\eta_{b2j} + \varepsilon_{bpj}.$$

All loading parameters λ_{wp} and λ_{bp} are 1, the first and the fourth are fixed to 1 for identification purposes during the estimation. The variance/covariance for the within level factors Ψ_w is given by $\psi_{w11} = 0.7$, $\psi_{w22} = 0.6$ and $\psi_{w12} = 0.4$. The variance/covariance for the between level factors Ψ_b is given by $\psi_{b11} = 0.4$, $\psi_{b22} = 0.3$ and $\psi_{b12} = 0.2$. The residual variances are $\theta_{wp11} = 1$ and $\theta_{bp11} = 0.2$. The mean parameters for the continuous variables are $\nu_p = 1$. For categorical variables the thresholds are given by $\tau_{p1} = -0.3$, $\tau_{p2} = 0.4$, $\tau_{p3} = 1.2$ and $\tau_{p4} = 1.8$, thus the distribution of the categorical variables is skewed towards the lower categories. We generate 100 samples according to the above model using 100 clusters of size 10 and analyze the data using both the ML and the WLSM both implemented in Mplus. For the model with categorical data the estimation requires 8 dimensional numerical integration. We used monte-carlo integration with 500 integration points. To improve the convergence rates we also used non-adaptive integration as well as the Cholesky parameterization. For WLSM estimator we used rectangular integration with 10 integration points.

Tables 1 and 2 show the results of the simulation study for the model with categorical and continuous variables. The table contains the results only for a representative set of parameters for compactness. For the categorical factor analysis model both estimators used approximately 1 min for each replication on a 2 processor 3GHz computer. For the continuous factor analysis both estimators takes 1 second or less per replication because numerical integration is not used. The convergence rates for the WLSM estimator are 100% for both categorical and continuous factor analysis models. The convergence rates for the ML estimator are 100% for the continuous factor analysis models however they are only 47% for the categorical factor analysis model. In fact under different parameterizations and integration settings the convergence rates for the ML estimator are even lower, while those for WLSM estimator appear to be independent of the parameterization or integration settings and remain at 100%. For each parameter we present the relative bias computed as follows. If θ is the true parameter value and if $\bar{\hat{\theta}}$ is the average of the θ estimate across all replications the relative bias is computed by

$$\frac{\bar{\hat{\theta}} - \theta}{\theta}.$$

The coverage is the percentage of time the 95% confidence interval obtained by the estimator contains the true value. The efficiency ratio is computed as follows. Let $\hat{\theta}_{wi}$ be the θ estimate for the WLSM estimator in the $i-$th replication and let $\hat{\theta}_{mi}$ be the θ estimate for the ML estimator in the $i-$th replication. The efficiency ratio is

$$\sqrt{\frac{\sum_i(\theta_{wi} - \theta)^2}{\sum_i(\theta_{mi} - \theta)^2}}$$

and it simply shows how much more variable the estimates of WLSM are than those of ML. Theoretically the ML estimator is the most efficient, implying efficiency ratio of at least 1, however due to the numerical integration the estimates obtain by the ML method are simply approximations and the true ML estimates are not available when we consider the categorical factor analysis.

The results of the categorical factor analysis presented in Table 1 show that not only the WLSM outperforms the ML method in terms of convergence and robustness

Table 2: Two-level factor analysis model with continuous variables.

parameter	true value	WLSM bias	ML bias	WLSM coverage	ML coverage	Efficiency ratio
λ_{w2}	1.0	2%	1%	95%	97%	1.07
ψ_{w12}	0.4	-1%	-1%	96%	96%	1.03
ψ_{w11}	0.7	-1%	-1%	97%	98%	1.02
θ_{w1}	1.0	0%	0%	94%	96%	1.05
λ_{b2}	1.0	4%	3%	96%	96%	1.16
ψ_{b12}	0.2	-3%	-3%	93%	95%	1.05
ψ_{b11}	0.4	-2%	-2%	92%	95%	1.07
ν_1	1	0%	0%	96%	94%	1.00
θ_{b1}	0.2	-3%	-3%	97%	32%	1.23

of the parameterization but also in the quality of the estimation. The WLSM estimates are more efficient and less biased than the ML estimates. The standard error WLSM estimates outperformed the the standard error ML estimates in terms of coverage as well. The bias of the WLSM is nearly non-existent, the coverage of the standard errors is very close to the nominal 95% value. The chi-square statistic for the WLSM estimator comparing the factor analysis model to the unrestricted two-level model rejected the structured model only 6% of the time, i.e., the correct model was accepted 94% of the time.

The results of the continuous factor analysis model are presented in Table 2. In this case the ML estimates are computed precisely because numerical integration is not used. Both estimators show no bias and virtually perfect coverage of the confidence intervals. The efficiency loss of the WLSM estimator is negligible for all parameters except the loading parameter on the between level and the residual variance parameter on the between level, where the efficiency is slightly bigger but still within reasonable limits. This simulation shows that that the WLSM estimator is not only a computationally feasible alternative but it is also nearly the optimal estimator.

REFERENCES

Hedeker, D. & Gibbons, R. D. (1996), "MIXREG: a computer program for mixed-effects regression analysis with autocorrelated errors," *Computer Methods and Programs in Biomedicine*, **49**, 229-252.

Muthen, B. (1984), "A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators," *Psychometrika*, **49**, 115-132.

Muthen, B. & Asparouhov, T. (2007), "Growth mixture analysis: Models with non-Gaussian random effects," Forthcoming in *Advances in Longitudinal Data Analysis* eds. Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G., Chapman & Hall/CRC Press.

Muthen, B. (1994), "Multilevel covariance structure analysis," *Sociological Methods & Research*, **22**, 376-398.

Muthen, B. & Asparouhov, T. (2002), "Latent Variable Analysis With Categorical Outcomes: Multiple-Group And Growth Modeling In Mplus," *Mplus Web Note No.4*, http://statmodel.com/.

Muthen, B. & Satorra, A. (1995), "Technical aspects of Muthen's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model," *Psychometrika*, **60**, 489-503.

Muthen, B., du Toit, S.H.C. & Spisic, D. (1997), "Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes," Accepted for publication in *Psychometrika*.

Olsson, U. (1979), "Maximum likelihood estimation of the polychoric correlation coefficient," *Psychometrika*, **44**, 443-460.

Olsson, U., Drasgow, F., & Dorans, N.J. (1982), "The polyserial correlation coefficient," *Psychometrika*, **47**, 337-347.

Raudenbush, S. & Bryk, A. (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods. Second Edition.* Thousand Oaks. Sage Publications, Inc.