STRUCTURAL EQUATION MODELING: A MULTIDISCIPLINARY JOURNAL
https://doi.org/10.1080/10705511.2021.1878896

Expanding the Bayesian structural equation, multilevel and mixture models to logit, negative-binomial, and nominal variables

Tihomir Asparouhov and Bengt Muthén

Mplus

ABSTRACT

Recent work on the Polya-Gamma distribution provides a breakthrough for the Bayesian modeling of logit, count, and nominal variables. We describe how the methodology is incorporated in the Mplus modeling framework and illustrate it with several examples: logistic latent growth models, multilevel IRT, multilevel time-series models for count data, multilevel nominal regression, and nominal factor analysis.

KEYWORDS

Polya-Gamma distribution; Bayesian estimation; Negative-binomial SEM; Nominal SEM

Introduction

The Bayesian estimation of the structural equation modeling framework implemented in Mplus is described in Asparouhov and Muthén (2010a). The framework includes mixture models as well as multilevel models. The observed endogenous variables can be continuous normally distributed outcomes as well as categorical outcomes based on the probit link function. In this note we describe how the Mplus framework is extended to include these new types of variables: count variables based on the Poisson and the negative-binomial distributions, nominal variables as well as binary variables based on the logit link function. This expanded framework is similar to the ML estimation framework implemented in Mplus, see Muthén and Asparouhov (2007). The model is largely unchanged, but we can now utilize the more efficient Bayesian estimation. Numerical integration is typically required in the ML estimation of latent variable models with non-normal outcomes. Therefore, the number of latent variables and random effects in the model cannot exceed 3 or 4 because the computational speed grows exponentially with the number of latent variables. In the Bayesian estimation, the computational speed grows linearly with the number of latent variables, and generally, an unlimited number of latent variables can be used without substantially increasing the computational time.

Categorical variables in the Asparouhov and Muthén (2010a) framework, based on the probit link function, utilize the conceptualization of an underlying latent variable. For each categorical variable Y, there is a latent variable Y^* that is cut according to certain threshold parameters to obtain the categorized observed variable Y. The existence of such an underlying latent variable is immediately provided by the probit link function. The latent variable construct is very important in the Bayesian estimation. Any structural model that can be formulated for Y, can also be formulated as a linear model for Y^*. For this linear model, all conditional distributions (for structural parameters, latent factors, random effects, and missing data) in the MCMC estimation are explicit. Provided with conjugate priors, the MCMC estimation is fast and efficient.

For other types of variables such as count, logit-categorical, or nominal, a conceptualization of such underlying latent variable had not been found until very recently. Fox's (2010) work on item response modeling, for example, in the absence of underlying latent variables, utilizes the Metropolis-Hastings algorithm as a part of the MCMC algorithm. Such an approach tends to be more computationally demanding, less efficient than explicit conditional distributions, more difficult to implement in a generalized framework, and more prone to slow mixing estimation. Groundbreaking recent work by Pillow and Scott (2012) and Polson et al. (2013) yields a underlying latent variable methodology for negative-binomial and Poisson count variables, logit-binary variables and nominal variables. The approach utilizes the Polya-Gamma (PG) distribution and is uniquely suited for structural, multilevel and mixture models. With the PG method, regression parameters with conjugate normal priors can be updated in the MCMC estimation from explicit conditional normal distributions. Similarly, normally distributed latent variables or random effects (random slopes, loadings or intercepts) can be updated from explicit conditional normal distributions. These explicit conditional distributions produce highly efficient MCMC estimation. Kim et al. (2018) compare the Polya-Gamma based estimation method to other Bayesian methods in the context of structural equation models with logit-binary variables and found that the Polya-Gamma method yields superior performance.

For Mixture models, the categorical latent class variables are generally modeled as nominal variables. Using the PG methodology, we can now also extend the Mplus Bayesian mixture framework to include latent class variable regression on other variables. Within the MCMC estimation, where the latent class variable is imputed in each iteration, the latent class regression is simply a nominal variable regression and the PG methodology applies.

CONTACT Tihomir Asparouhov tihomir@statmodel.com 3463 Stoner Ave. Los Angeles, CA 90066.

We are indebted to Mårten Schultzberg for sharing his explorations of the Polya-Gamma methodology and to Alberto Maydeu-Olivares for helpful discussions on the nominal factor analysis model.

Missing data on count/nominal/categorical variables are typically easy to resolve in likelihood based methods (Bayes and ML). However, missing data on the predictors/covariates of such variables is not. In the ML estimation for example, such missing data lead to additional dimensions of numerical integration. Even for a simple logit regression model, the dimensions of numerical integration could easily become substantial depending on the amount of missing data in the covariates. In this situation the PG methodology can be utilized as well. The missing values can be imputed in the MCMC estimation from explicit conditional distributions. This also extends to the mixture modeling situation where the latent class variable has missing predictors. Furthermore, the method can be applied to the three-stage estimation, see Asparouhov and Muthén (2014a), where the final step in the estimation is conducted with the Bayes estimator instead of the ML estimator. An illustration of that approach is provided in Asparouhov and Muthén (2020a).

There is one drawback of the PG methodology that makes the Bayesian estimation slower, however. If categorical data are modeled with the probit link function, the underlying normal variable has a conditional distribution that varies linearly with the model covariates. The model parameters remain the same across individuals, which makes matrix manipulations efficient. This does not apply to the PG methodology. The underlying variable for count/nominal/logit variables has a conditional distribution that varies across observations. Matrix manipulation must be performed separately for each observation. This affects the computational speed of the estimation but it does not affect the generality of the methodology or the mixing efficiency.

In the next section we describe the PG methodology and how it is implemented in Mplus 8.5. We then illustrate the Bayesian estimation with several simulation studies.

Bayesian estimation for models with logit, count and negative-binomial variables

We begin by providing a formal definition for the PG distribution. A random variable W has a Polya-Gamma distribution with parameters b and c, i.e. $W \sim PG(b,c)$, if W is obtained as the following infinite sum

$$W = \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{G_k}{(k-0.5)^2+(c/(2\pi))^2} \tag{1}$$

where G_k are independent gamma random variables with distribution $Ga(b,1)$. It is not possible to simplify the above infinite sum, but in most cases the sum can be approximated as a finite sum. For example, the first 500 terms of the above sum provide a good approximation for most values of b and c. The $PG(b,c)$ distribution takes only positive values and has mean

$$E(W) = (b/(2c))\tanh(c/2) \tag{2}$$

and variance

$$Var(W) = (b/(4c^3))(\sinh(c)-c)/(\cosh(c/2))^2. \tag{3}$$

For large values of b and c (>200) the PG distribution can be approximated by a normal distribution. Also, for large values of c, the variance of the distribution goes to 0 and so the distribution becomes approximately equal to a constant. Note also that the parameter b is always positive, while the parameter c can be positive and negative and the PG distribution is independent of the sign of c.

To be able to use this distribution in the MCMC estimation, an efficient method for generating PG random variables is needed. Multiple such methods have been proposed in Windle et al. (2014). Mplus utilizes four of these methods: finite sum approximation, normal approximation, saddle point approximation, and Devroye's approximation. Depending on the parameters b and c, one of the four methods is used. Such a hybrid approach is designed to find an optimal compromise between the speed of the computation and the quality of the approximation.

Next we describe how the PG distribution is used to facilitate the Bayesian estimation for logistic, negative-binomial, and nominal regressions. Note that the PG distribution is not used to model a dependent variable. It is used only to construct a normally distributed underlying latent variable from an observed logit, count, or nominal variable.

Logistic regression

Consider the logistic regression for a binary outcome Y (with outcome values of 0 and 1) given by the equation

$$P(Y=0) = \frac{1}{1+Exp(\beta X)} \tag{4}$$

where X represents a set of covariates and β represents a set of parameters that are to be estimated.

The underlying variable Y^* in this case is constructed in two step. In the first step we generate

$$W \sim PG(1,\beta X). \tag{5}$$

In the second step we compute Y^* as follows:

$$Y^* = \frac{Y-0.5}{\sqrt{W}}. \tag{6}$$

The logistic regression model (4) for Y implies the following linear regression model for Y^*

$$Y^* = \beta\sqrt{W}X + \varepsilon, \tag{7}$$

where ε is a standard normal random variable. This linear regression can be used to estimate the regression coefficients β. Note that the predictor variables in the linear regression are $\sqrt{W}X$ rather than the original covariates X used in the logistic regression. Once the underlying latent variable Y^* is generated, the logistic regression equation for Y is replaced by the linear regression for Y^*. Thus any structural/multivariate/multilevel model involving the logistic regression for Y is transformed into a structural/multivariate/multilevel model for Y^*. Such models are then estimated as in Asparouhov and Muthén (2010a).

To be more specific, the MCMC estimation proceeds as follows. In each MCMC iteration, we generate W and compute Y^*. We then update any structural parameters, random effects, latent variables and missing data, using the linear model for Y^*. Note again that the covariates in the logistic equation are changed to $\sqrt{W}X$. If those same covariates are used as predictors in another equation, they are unchanged for that other equation or are changed according to a different PG deviate. Also note that if the covariates X must be updated due to missing values or if the covariates include latent variables, the scale coefficient $\sqrt{W}$ is then attached to the regression parameters, i.e., in the Gibbs sampler step that updates X, the coefficients β in the logistic regression are replaced with the coefficients $\beta\sqrt{W}$ in the linear regression of Y^*.

In the above description, we did not use the intercept in the logistic regression. The intercept is a special case of a regression parameter for a covariate that is the constant 1. In the Mplus implementation, however, the intercept is actually used and is called a threshold τ. In line with the probit regression, the actual Mplus parameterization for the logistic regression is as follows

$$P(Y=0)=\frac{Exp(\tau-\beta X)}{1+Exp(\tau-\beta X)}=\frac{1}{1+Exp(-\tau+\beta X)}. \quad (8)$$

Thus the logistic regression intercept is equivalent to the threshold parameter in Mplus but is with the opposite sign.

The PG methodology does not extend to nonbinary ordered polytomous variables with the logit link function. The Mplus implementation allows for the simultaneous modeling of binary variables with the logit link function and nonbinary ordered polytomous variables with the probit link function. To use such modeling the option LINK = PROBIT LOGIT must be specified in the ANALYSIS command. The natural extension of the PG methodology to nonbinary categorical variables leads to nominal variables rather than ordinal. We discuss the PG methodology for nominal variable further below.

In a multivariate model where a binary variables Y is used to generate an underlying latent variables Y^*, a natural question arises regarding the possible correlation between Y^* and other variables in the model. Such correlation can easily be estimated, however, the correlation does not naturally translate into some kind of an association between the observed binary variables Y and the other variables. In the Mplus implementation, such correlations are not pursued at this time. The main tool for correlating binary variables in this logit based modeling framework is to use normally distributed latent variables that predict the binary variables, as in item response theory, for example. The same logic applies to structural parameters where the underlying latent variable Y^* is used as a predictor for another variable. While it is possible to estimate such a structural parameter within the MCMC estimation, the parameters would not translate easily into an interpretable model for the observed variable Y and thus are not included in this modeling framework. The latent variable Y^* must only be used as a dependent variable. These observations apply also to the rest of the PG-based types of variables: negative-binomial, Poisson, and nominal variables.

Negative-binomial regression

First, we discuss the properties of the negative-binomial distribution and then we will extend that discussion to the negative-binomial regression model.

Suppose that a variable Y has a negative-binomial distribution. The probability distribution function for Y is given by

$$P(Y=k)=\binom{k+r-1}{r-1}p^r(1-p)^k, \quad (9)$$

where r and p are the parameters of the distribution. The interpretation of this distribution is that Y represents the number of successes in a sequence of Bernoulli (binary) trials before the occurrence of the $r-$ th failure, where the parameter p represents the failure probability. Such an interpretation requires that r is an integer parameter but in fact the negative-binomial distribution is defined for any positive value r using the notation that

$$\binom{k+r-1}{r-1}=\frac{\Gamma(k+r)}{\Gamma(k+1)\Gamma(r)}=\prod_{i=1}^{k}\frac{r-1+i}{i}, \quad (10)$$

where Γ is the Gamma function. The mean of Y is

$$\mu=E(Y)=\frac{r(1-p)}{p} \quad (11)$$

and the variance is

$$v=Var(Y)=\frac{r(1-p)}{p^2}. \quad (12)$$

Instead of using the parameter r, the Mplus implementation uses the parameter α

$$\alpha=\frac{1}{r} \quad (13)$$

which is called the dispersion parameter. It can be shown that

$$p=\frac{1}{1+\alpha\mu} \quad (14)$$

$$v=\mu+\alpha\mu^2. \quad (15)$$

The above equation has a symbolic meaning as it explicates how the negative-binomial distribution differs from the Poisson distribution. For the Poisson distribution, the variance is equal to the mean. This is often viewed as an impractical limitation, which would rarely be satisfied in real data. When $\alpha>0$, the variance of the negative-binomial distribution is bigger than the mean and the term $\alpha\mu^2$ represents the difference between the variance and the mean. It should be noted, however, that the Poisson distribution limitation does not extend to Poisson regression in the following sense. In a Poisson regression, conditional on all predictors, the variance and the mean are equal. However, the marginal/unconditional mean would not be equal to the marginal/unconditional variance. Regardless, the negative-binomial distribution yields a more flexible modeling framework for the count variables than the Poisson distribution. If the dispersion parameter α converges to 0 and $(1-p)/\alpha$

converges to λ, the negative-binomial distribution with parameters α and p converges to the Poisson distribution with parameter λ.

Next we discuss how the negative-binomial distribution is used to construct regression models. Several such methods/parameterizations are described in Hilbe (2011). The first parameterization that we describe here is the NB-2 parameterization. In this parameterization, the relationship between a count variable Y and a set of predictors X is defined by the following equation

$$p = \frac{1}{1 + \alpha Exp(\beta X)}, \tag{16}$$

where β is a vector of regression parameters. The intercept is not included in this discussion again as the intercept is simply the special case of a regression parameter for the constant covariate of 1. From (14) and (16) we derive that

$$\mu = E(Y) = Exp(\beta X). \tag{17}$$

We denote this model by $Y \sim NB2(\beta X, \alpha)$. Conditional on a set of covariates X, the variable Y has a negative binomial distribution with probability function given in (9), where p is computed as in (16) and r is computed as in (13). When α converges to 0, $(1-p)/\alpha$ converges to $\lambda = Exp(\beta X)$ and thus the $NB2(\beta X, 0)$ model is the same as the standard Poisson regression model, usually denoted by $Po(\beta X)$.

Mplus implements also a different negative-binomial regression parameterization which we refer to as the PG parameterization. In this alternative parameterization, equations (16–17) are replaced by

$$p = \frac{1}{1 + Exp(\beta X)} \tag{18}$$

$$\mu = E(Y) = \frac{Exp(\beta X)}{\alpha}. \tag{19}$$

We denote this model by $Y \sim NBPG(\beta X, \alpha)$. Conditional on a set of covariates X, the variable Y has a negative binomial distribution with probability function given in (9), where p is computed as in (18) and r is computed as in (13).

The difference between the two parameterizations is quite simple and it is only in the intercept of the regression model. If the intercept of the NB-2 model is denoted by α_{NB2} and the intercept of the NBPG model is denoted by α_{PG}, the two parameters are related by the following equation:

$$\alpha_{NB2} = \alpha_{PG} - \log(\alpha) \tag{20}$$

where α is the dispersion parameter. The change in the parameterization has no effect on any other parameter, i.e. regression coefficients or the dispersion parameter.

In the Mplus software, the NB-2 parameterization is also implemented with the ML estimator, while the PG parameterization is only available with the Bayesian estimator. A variable can be specified as an NB-2 variable using the option COUNT = Y(nb). A variable can be specified as PG variable using the option COUNT = Y(nbpg). A Poisson distribution variable can be specified simply as COUNT = Y. The Bayesian implementation for the Poisson distribution is simply a negative-binomial variable, using the NB-2 parameterization, with the dispersion parameter α fixed to 0.01, i.e. this implementation is an approximation to the Poisson distribution. The PG methodology does not directly extend to the Poisson distribution and it must be approximated that way as a negative-binomial distribution. The dispersion parameter cannot be fixed directly to 0 as that will cause numerical problems in the estimation; see equation (24) below, where the log of the dispersion parameter is evaluated.

The original work by Polson et al. (2013) uses the PG parameterization of the negative-binomial distribution. This is one reason why we implemented this parameterization in Mplus as well. The NB-2 parameterization was also implemented as this would be considered the most commonly used parameterization, due to the simplicity of (17). The MCMC algorithms for the two parameterizations are very similar, with the exception of the updating of the dispersion parameter, which we discuss below. Simulation studies revealed that the mixing quality of the Bayesian estimation is not the same between the two parameterizations for models that involve constraints on the intercept parameters, i.e. where the intercept parameters are held fixed to a particular value, such as in growth models, or where intercept parameters are held equal to other parameters. In addition, models involving intercept constraints are not equivalent between the two parameterizations. Intercept constraints in the NB2 parameterization translate into a different set of constraints for the NBPG parameterization. The NB-2 parameterization appears to have slightly worse mixing performance as compared to the PG parameterization but in most situations the difference would be ignorable.

The Bayesian estimation of the negative-binomial regression

The Bayesian estimation of the negative-binomial regression is very similar to that of the logistic regression. Each negative-binomial variable Y is augmented with an underlying latent variable Y^* and the negative-binomial regression model is transformed to a linear regression model for Y^*. At that point, as in the logistic regression case, the model is estimated with the standard methodology.

First we describe the construction of Y^* for the PG parameterization. The first step is to generate W as

$$W \sim PG(r + Y, \beta X). \tag{21}$$

In the second step, Y^* is computed as follows:

$$Y^* = \frac{Y - r}{2\sqrt{W}}. \tag{22}$$

The negative-binomial regression for Y implies a linear regression model for Y^*

$$Y^* = \beta\sqrt{W}X + \varepsilon, \tag{23}$$

where ε is a standard normal random variable.

The construction of the underlying Y^* in the NB-2 parameterization is as follows. In the first step we generate W as

$$W \sim PG(r + Y, \beta X + \log(\alpha)). \tag{24}$$

In the second step Y^* is computed as follows:

$$Y^* = \frac{Y - r}{2\sqrt{W}} - \log(\alpha)\sqrt{W}. \quad (25)$$

Again, the negative-binomial regression for Y implies a linear regression model for Y^*

$$Y^* = \beta\sqrt{W}X + \varepsilon, \quad (26)$$

where ε is a standard normal random variable.

As in the logistic regression, in each MCMC iteration a new value for W is generated based on the current model parameters, Y^* is computed, the covariates X are multiplied by $\sqrt{W}$ and the parameters β are estimated using the linear regression model for Y^*.

Estimating the dispersion parameter

The negative-binomial regression has one additional parameter to be estimated: the dispersion parameter. Here we follow the approach described in Zhou and Carin (2015) and Neelon (2019) based on the *Chinese restaurant table* (CRT) distribution. We begin with the PG parameterization. Conditional on all model parameters, we augment the data with the following variable:

$$A = \sum_{i=1}^{n}\sum_{j=1}^{Y_i} A_{ij} \quad (27)$$

where n is the sample size, Y_i is the value of the negative-binomial variable for the $i-$th observation, and A_{ij} is a binary 0/1 variable with the following distribution:

$$P(A_{ij} = 1) = \frac{r}{r + j - 1}, \quad (28)$$

where $r = 1/\alpha$. If the dispersion parameter is given the conjugate Inverse-Gamma prior $IG(a, b)$, the conditional distribution of α given all model parameters and data, including the above augmented variable A, is the following Inverse-Gamma distribution:

$$[\alpha|*] \sim IG(a + A, b - \sum_{i=1}^{n}\log(p_i)) \quad (29)$$

where p_i is the probability given in equation (18) computed for the $i-$th observation.

For the NB-2 parameterization, the estimation is slightly more complex. The data augmentation for the variable A remains the same. However, the probability p_i computed in (16) for the NB-2 parameterization clearly contains the value of α. Thus, the above explicit conditional distribution computation breaks down. The RHS in equation (29) should not contain α. To work around this problem, we internally use α_{PG} as the model parameter instead of α_{NB2}. If α_{PG} is the model parameter, then p_i no longer contains α and is simply computed from the rest of the model parameters. This internal reparameterization, however, implies certain limitations on the structural model that can be estimated with the CRT dispersion updating and the NB-2 parameterization. Here we list these limitations. The intercept in the negative-binomial regression cannot be a fixed parameter since that parameter is needed for the internal reparameterization. The intercept in the negative-binomial regression cannot be held equal to other parameters because that kind of equality constraint will not hold in the internal reparameterization. The intercept in the negative-binomial regression cannot have an informative prior as that will also break down in the internal reparameterization.

When these model limitations apply, we use a different updating procedure for the dispersion parameter. Namely, we use the Metropolis-Hastings (MH) method, also discussed in Neelon (2019). The Mplus implementation of the MH method is as follows. At each MCMC iteration, a random deviate ε is drawn from the normal distribution $N(0, \nu_0)$. If the current dispersion parameter is α, the proposal dispersion parameter α^* is computed as follows:

$$\alpha^* = \alpha Exp(\varepsilon). \quad (30)$$

The proposal parameter is then accepted with probability

$$\min(1, \frac{P_0(\alpha^*)\prod_{i=1}^{n} P(Y_i|\alpha^*)}{P_0(\alpha)\prod_{i=1}^{n} P(Y_i|\alpha)}), \quad (31)$$

where P_0 is the prior density function for the dispersion parameter, n is the sample size, Y_i is the $i-$th observed value and $P(Y_i|\alpha)$ and $P(Y_i|\alpha^*)$ are computed using (9) and (16).

The initial value for the proposal variance ν_0 is set to 0.01 and it is adjusted every 50 iterations to maintain approximate rate of acceptance of the new parameters around 50%. This variance can be adjusted only in the burnin period which by default is set to the first 1000 MCMC iterations. The proposal variance ν_0 does not change beyond the burnin period to preserve the validity of the MH method. Subsequently, more iterations are conducted to achieve convergence and to obtain the posterior distribution of the parameters. In addition, the first 50 iterations in this process are computed using the conjugate CRT method discussed above, which tends to give good starting values for the estimation.

The MH method can be used for any model estimation; however, the conjugate CRT method yields a more efficient mixing approach. In the CRT method, every MCMC iteration provides a new draw from the posterior distribution of the dispersion parameter. In contrast, the MH method provides a new draw every other iteration on average. Thus, Mplus uses the CRT method, unless the model includes intercept constraints, as discussed above, in which case we use the less efficient MH method.

Model extensions

In some practical applications, count data may contain an excessive amount of 0 values. In such cases, it is desirable to model the 0 category separately from the count distribution. One such simple approach is described in Kang et al. (2020) and is based on two-part modeling. A count variable Y with an excessive amount of zeros can be modeled as a set of two variables: a 0/1 binary variable U which serves as an indicator for when Y is positive, and a count variable $Y_0 = Y - 1$, which can be modeled as a Poisson or a negative-binomial variable. When $Y = 0$, the binary variable $U = 0$ and the count variable Y_0 is missing. When $Y > 0$, the binary variable $U = 1$ and the count variable

$Y_0 = Y - 1$. This modeling approach allows the 0 category to be modeled independently from the rest of the count categories and that includes a separate probit or logit regression for the binary variable U. The Bayesian framework described above can be used for this two-part modeling approach.

Several alternative models are implemented in Mplus with the maximum-likelihood estimation such as the negative binomial hurdle model, the zero-inflated Poisson model and the zero-inflated negative-binomial model, see Hilbe (2011). These models are currently not implemented in Mplus with the Bayesian estimation, however, the PG methodology can be used for the zero-inflated models as well, see Neelon (2019).

Nominal regression

Suppose that a nominal variable Y takes k unordered values and that X is a vector of predictors. The nominal regression model is described by the following equation:

$$P(Y = j) = \frac{Exp(\beta_j X)}{\sum_{i=1}^{k} Exp(\beta_i X)}, \tag{32}$$

where for identification purposes $\beta_k = 0$ and β_j for $j = 1, \ldots, k-1$ are vectors of regression parameters. If $k = 2$, the nominal regression model is the same as the logistic regression model for a binary variable. The Bayesian estimation of the nominal regression parameters can be obtained by estimating the regression parameters for a sequence of $k-1$ binary logistic regressions. To estimate β_1, we form a new binary variable

$$Z = \begin{cases} 1, & \text{if } Y = 1 \\ 0, & \text{if } Y > 1 \end{cases}. \tag{33}$$

Given (32), we get that

$$P(Z = 0) = \frac{1}{1 + Exp(\beta_1 X + A)} \tag{34}$$

where

$$A = -log(\sum_{i=2}^{k} Exp(\beta_i X)) \tag{35}$$

does not depend on β_1. Equation ((34)) is a binary logistic regression for Z and can be estimated with the PG method, i.e. we can obtain the conditional distribution of $[\beta_1|\beta_2, \ldots, \beta_{k-1}]$ needed in the Gibbs sampler. Similarly, we can obtain the conditional distribution $[\beta_2|\beta_1, \beta_3 \ldots, \beta_{k-1}]$ by forming a new binary variable and logistic regression based on the second category of the nominal variable. The process is repeated $k-1$ times.

The above estimation method does not allow for the simultaneous estimation of $\beta_1, \beta_2, \ldots, \beta_{k-1}$. Therefore the Bayesian estimation method cannot accommodate constraints that involve all of the nominal regression parameters. In general, this appears to be a minor limitation. Equality constraints modeling among the nominal regression parameters is rare. However, the limitation is important in the general SEM model. Two situations in particular are hindered by the sequential estimation of the model parameters. The first one is when one of the predictors has missing values. The second situation is when one of the predictors is a continuous latent variable. To update the latent variable (or equivalently the missing predictor), the regression parameters and the predictors switch roles. The regression parameters become the predictors and the predictors are estimated as regression parameters that are equal across the categories. The situation creates equality constraints across the regression parameters because the predictor is the same in all categories. This problem must be resolved with a different method. The MH method can be utilized here as well. Consider the updating of a latent variable η which is a predictor for a set of nominal variable $Y_1, \ldots, Y_m$. First we compute the conditional distribution of η given all other variables except the nominal variables. This can be computed as in Asparouhov and Muthén (2010a) and the result is a normal distribution with mean μ and variance σ^2. If the current value of the latent variable is represented by η, the new proposal value η^* is computed as

$$\eta^* = \eta + \varepsilon \tag{36}$$

where ε is drawn from a normal distribution $N(0, v_0)$. The proposal value is accepted with probability

$$\min(1, Exp(((\eta - \mu)^2 - (\eta^* - \mu)^2)/(2\sigma^2)) \frac{\prod_{i=1}^{m} P(Y_i|\eta^*)}{\prod_{i=1}^{m} P(Y_i|\eta)}). \tag{37}$$

Here, the $N(\mu, \sigma^2)$ distribution essentially takes the role of the prior. The probabilities $P(Y_i|\eta^*)$ and $P(Y_i|\eta)$ are computed as in (32). The same proposal distribution variance v_0 is used for all observations in the sample. That variance is adjusted in every MCMC iteration during the burnin phase to maintain a rejection/acceptance ratio near 50% across the entire population.

The second situation where the MH method is needed is the case when a normally distributed predictor of nominal variables has missing values. The computation in that case is similar.

Note here that the MH method is not needed when the nominal variable is binary. The problem with equalities across the categories occurs only for nominal variables with more than two categories. In the Mplus implementation it is best to specify a binary nominal variable as a logit binary. This way the more efficient estimation with the explicit conditional distributions is utilized.

Missing data

There are two distinct types of missing data that must be addressed. Missing data on the dependent nominal/count/logit variables and missing data for predictors of nominal/count/logit variables.

First we consider the missing data on the dependent variables. There are two different variations in the Bayesian estimation. In the first approach, the missing data is imputed within each MCMC iteration from the current model estimates. Thereafter, the imputed values are used the same way as the observed values and are essentially used to

update model parameters, missing data, etc. The second approach, which is equivalent to the first approach, does not impute missing values. Instead when a quantity must be updated conditional on the dependent variable, we only condition on the actual observed values. The first approach has the advantage of somewhat simpler computations and it also provides imputed values for the missing dependent variables if such are needed. The second approach has the advantage that it mixes more efficiently as it has fewer unknown quantities to update and is computationally faster. Mplus implements the second approach. When multiple imputations are requested, however, the values are imputed as in the first approach but those values are not used in the estimation.

In the current Mplus 8.5 implementation, missing values of predictors of nominal/count/logit variables are allowed when the predictor is normally distributed. The estimation in the nominal cases was discussed earlier. The estimation for the count and logit case is as follows. From the observed count or logit value Y, we generate the underlying variable Y^*, using the current imputed values for the missing predictors. The missing predictor for Y is a missing predictor for Y^*. In that case, the conditional distribution for the missing predictor (conditional on Y^*) is obtained from the multivariate normal distribution (for Y^* and the missing predictor) and is an explicit normal distribution. The missing predictor is updated from this conditional normal distribution.

Mixture modeling

In Asparouhov and Muthén (2010a), three different Bayesian estimation methods are described for mixture models. The methods differ in how the latent class variable is updated: as a separate group, together with the latent continuous variables, or together with the underlying continuous variables when the model includes categorical variables. All three approaches extend naturally to the nominal/count/logit variables because of the construction of the underlying latent variables described above. Not all methods are available for all models and not all methods are equally efficient. Mplus automatically selects the most efficient among the available methods, although it is possible for a specific method to be requested for a particular estimation via the ANALYSIS option CGENERATION.

Priors

In this section we discuss the choice of priors for the model parameters as well as the default settings in Mplus. For all regression parameters we use normal priors which are the conjugate priors for the above computation. The dispersion parameter prior is the inverse gamma prior which is also a conjugate prior. The default prior for the negative-binomial regression parameters is the improper and uninformative prior $N(0,\infty)$. For the logit and nominal regression parameters, the default prior is set to the weakly informative prior $N(0,5)$. This weakly informative prior prevents estimation problems that sometimes occur when there is a categorical predictor and the contingency table of the dependent variable and the predictor contains empty cells. The weakly informative prior is also in line with the default priors used in Mplus for the probit regression parameters. The default prior for the dispersion parameter is $IG(-1,0)$, which is an improper and uninformative prior with a constant density function over the interval $(0,\infty)$.

Model fit

At this time model fit tools are fairly limited in Mplus 8.5. Significance of parameters can be evaluated through the credibility intervals and hypothesis involving multiple parameters can be evaluated via the Bayesian Wald test implemented in Mplus MODEL TEST, see Asparouhov and Muthén (2020b). For a pair of nested models, the significance of the additional parameters in the more general model can be evaluated with the Bayesian Wald test. Model estimated distribution tables are also computed in Mplus and can be obtained with the RESIDUAL option of the OUTPUT command. These distribution tables can be compared to the observed distribution tables to evaluate fit.

The general model – SEM, multilevel, and mixture models

In this section, we described the general model implemented in the Bayesian framework of Mplus and show how nominal/count/logit variables are incorporated in that model.

We begin with the single level SEM model. Let Y be a vector of p observed continuous dependent variables, η be a vector of m latent variables, and X be a vector of q independent observed variables. The structural equation model for these variables is given by the following equations

$$Y = \nu + \Lambda\eta + KX + \varepsilon \tag{38}$$

$$\eta = \alpha + B\eta + \Gamma X + \zeta. \tag{39}$$

To include categorical variables via the probit link function in the above model, we construct the underlying continuous variable as usual. For each categorical variable Y_j in the model, taking the values from 1 to k, we assume that there is a normally distributed latent variable Y_j^* and threshold parameters $\tau_{1j},\ldots,\tau_{k-1j}$ such that

$$Y_j = t \Leftrightarrow \tau_{t-1j} \le Y_j^* < \tau_{tj}, \tag{40}$$

where $\tau_{0j} = -\infty$ and $\tau_{kj} = \infty$. This construction converts a categorical variable Y_j into an unobserved continuous variable Y_j^*. To include categorical variables in the above SEM model, we use Y_j^* in (38) instead of Y_j. For identification purposes the residual variance of Y_j^* is fixed to 1. This model represents the Bayesian SEM model implemented in Mplus Version 8.4.

Next we describe the inclusion of the nominal/count/logit variables in the model. Suppose that N is a nominal variable taking values from $1,\ldots,k$. The SEM model extend to the nominal variable as follows

$$P(N=j) = \frac{Exp(\nu_{N,j} + \beta_{N,j}Y + \Lambda_{N,j}\eta + \Gamma_{N,j}X)}{\sum_{i=1}^{k} Exp(\nu_{N,i} + \beta_{N,i}Y + \Lambda_{N,i}\eta + \Gamma_{N,i}X)}, \tag{41}$$

where for identification purposes the parameters $\nu_{N,j}, \beta_{N,j}, \Lambda_{N,j}, \Gamma_{N,j}$ are fixed to 0 if $j = k$.

The model for a binary logistic link variable L is a special case of the nominal variable model ($k = 2$) and reduces down to

$$P(L = 1) = \frac{1}{1 + Exp(\nu_L + \beta_L Y + \Lambda_L \eta + \Gamma_L X)}. \quad (42)$$

The general SEM model extends to a negative-binomial variable P as follows. The distribution of P is as in equation (9) where

$$p = \frac{1}{1 + \alpha Exp(\nu_P + \beta_P Y + \Lambda_P \eta + \Gamma_P X)} \quad (43)$$

if the variable is based on the NB-2 parameterization and

$$p = \frac{1}{1 + Exp(\nu_P + \beta_P Y + \Lambda_P \eta + \Gamma_P X)} \quad (44)$$

if the variable is based on the PG parameterization.

Note that in the above model the nominal/count/logit variables do not have residual correlations the way continuous (observed and latent) and Probit-categorical variables do, i.e., through the covariances of ε and ζ. The main way to model nonindependence between such variables is through their common predictors. In particular, using common latent variables in the model leads to conditional nonindependence the same way residual correlations do. Note also that the model as specified above does not allow for the nominal/count/logit variables to be mediators, i.e. predictors for another variable in the model. This to some extent can also be remedied by specifying a latent continuous variable behind the variable and then the latent continuous variable can be used as a proxy mediator. Such an approach is often used with continuous observed variables where in addition, the residual variance is fixed to zero so that the continuous observed variable becomes identical to the latent variable. With nominal/count/logit variables, we cannot fix the "residual" to zero as such a residual does not exist. Therefore a latent variable proxy is equivalent to a latent factor measured by a single observed variable, which is usually an unidentified model. Further model restrictions must be devised to resolve this problem. Note also that the underlying continuous variable for nominal/count/logit variables is not in the model. This is in contrast to the Probit-categorical variables where the underlying continous variable can be used as a predictor for other variables. The underlying continuous variable for nominal/count/logit variables is simply a tool that is used in the Bayesian estimation, but it bares no special meaning or interpretation in the actual SEM model.

Multilevel and mixture model extensions

The extension of the SEM model to a two-level model is done as usual. Every regression or intercept parameter can be a cluster specific normally distributed random effect. On the between level, all random effects can be regressed on each other and other variables or be correlated with each other. The extension of the SEM model to a general mixture model is also done as usual. A latent categorical variable C is included in the model and every model parameter can be class-specific, i.e., the SEM model is different for every latent class. The latent class variable is treated as a nominal variable and can be regressed on other variables. If k is the number of classes in the model, the latent class distribution is given by

$$P(C = j|X) = \frac{Exp(\nu_{C,j} + \Gamma_{C,j} X)}{\sum_{i=1}^{k} Exp(\nu_{C,i} + \Gamma_{C,i} X)}. \quad (45)$$

Further discussion on the multilevel and mixture model extensions is available in Asparouhov and Muthén (2010a) and Asparouhov and Muthén (2019).

Examples

In this section we illustrate the Bayesian estimation with several simulation studies. All of the Mplus scripts used for these analyses for are available online.[1]

Logistic latent growth model

A quadratic logistic growth model is given by the following equation

$$P(Y_t = 1) = \frac{1}{1 + Exp(i + st + qt^2)}, \quad (46)$$

where i, s and q are normally distributed random effects

$$\begin{pmatrix} i \\ s \\ q \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \right). \quad (47)$$

We generate 8 equally spaced binary observations at times t, starting at −1.4 and ending at 1.4. In practical applications, the times of observations are often set to integer values. With eight observations the times can be set to 0,1, …, 7 or 1,2, …, 8. Using such alternative time scores yields equivalent models, however, there are certain implications that concern the Bayesian estimation. If the values of t are larger such as 7 or 8 (as compared to 1.4), the value of t^2 will be as large as 49 or 64. This leads to a dramatic reduction in the scale of the quadratic term q and the variances of q could become very small. The implications of that are twofold. First, the small variance of q may lead to slow convergence, worse mixing and larger model parameter autocorrelations across the MCMC iterations. Second, the quadratic effect may erroneously be deemed to be substantively minor. It is preferable to have the time scale be set so that the times of observations are approximately on a standard normal scale. With our choice of time scale, the mean of the time scores is 0 and the variance is 0.96, i.e. approximately standardized.

We conduct the simulation study using 100 replications for several different sample sizes $N = 100, 200, 500, 1000$. The results are presented in Table 1. Most parameters estimates appear to be unbiased and the coverage is near the nominal levels. For smaller sample sizes, however, the variance of the

[1] http://statmodel.com/download/lcn.zip.

Table 1. Quadratic logistic growth model: absolute bias(coverage).

Parameter	True Value	N = 100	N = 200	N = 500	N = 1000
μ_1	−.3	.00(.95)	.00(.96)	.00(.99)	.00(.98)
μ_2	.2	.01(.97)	.00(.99)	.00(.93)	.00(.95)
μ_3	.25	.00(.93)	.00(.95)	.00(.94)	.00(.95)
σ_{11}	.5	.01(.98)	.00(.98)	.01(.94)	.00(.98)
σ_{12}	.1	.02(.92)	.00(.95)	.01(.93)	.01(.93)
σ_{13}	.05	.02(.98)	.02(.99)	.00(.98)	.00(.92)
σ_{22}	.3	.08(.96)	.05(.95)	.01(.95)	.00(.96)
σ_{23}	.05	.01(1.00)	.01(1.00)	.01(.95)	.00(.95)
σ_{33}	.2	.14(.96)	.10(.93)	.04(.94)	.02(.91)

quadratic random effect has some bias. This bias occurs also with the ML estimator. For optimal performance, the model needs medium or higher sample sizes. For small sample sizes, the results should be interpreted cautiously. A more thorough simulation study would involve varying the number of time points in the growth model. Having fewer time points would likely require even bigger sample sizes. Note also that the sample size requirements depend on the growth curve. A linear growth model would require less time points and smaller sample sizes than a quadratic curve model.

The above model can be estimated in two other completely different ways. First, the model can be estimated using the probit link function and then the mean parameter estimates can be multiplied by $D = 1.749$ while the variance covariance estimates are multiplied by D^2. The value of D is chosen as in Camilli (2017). This approach provides an approximation to the logistic growth model and is based on the fact that the logit and the probit distribution functions are quite similar. In most applications, this approximation works sufficiently well. The advantage of the probit-based model estimation is that it is faster when using Bayesian methods.

The second estimation alternative is the multilevel approach. The above model can be estimated as a two-level model where each cluster has eight observations while t and t^2 are treated as predictors for the binary variable. In this setup, s and q are simply the random coefficients for the predictors while i is the random intercept. In the Mplus language the variable i is also the between part of the binary variable.

To illustrate these alternative methods, we generate one data set using the above model with sample size $N = 2000$ and we estimate the model using the three different approaches. The results are reported in Table 2. All three approaches use Bayesian estimation and some of the small variation in the results can be attributed to randomization. We used 10,000 MCMC iterations in this estimation to ensure that such variation is minimal. The logit growth model and the two-level logit model yield almost identical results as expected because the two models are the same. The probit growth model also yields similar results but some larger differences are visible in the estimates of the variances of the random effects.

Table 2. Alternative estimation for quadratic logistic growth model: estimate (standard error).

Parameter	True Value	Logit Growth	Two-level Logit	Probit Growth
μ_1	−.3	−.31(.03)	−.31(.03)	−.34(.03)
μ_2	.2	.15(.02)	.15(.02)	.16(.03)
μ_3	.25	.23(.03)	.23(.03)	.24(.03)
σ_{11}	.5	.48(.07)	.48(.07)	.60(.07)
σ_{12}	.1	.09(.03)	.09(.03)	.10(.04)
σ_{13}	.05	.08(.04)	.07(.04)	.04(.04)
σ_{22}	.3	.27(.04)	.28(.04)	.34(.04)
σ_{23}	.05	.01(.03)	.01(.04)	.01(.04)
σ_{33}	.2	.19(.05)	.20(.05)	.27(.05)

Multilevel IRT

Multilevel IRT models have been discussed in Fox (2010), Asparouhov and Muthén (2016), and Muthén and Asparouhov (2018). With the Bayesian implementation of the logit link for binary variables, we can now estimate these models in the logit scale, instead of probit, which is traditionally used with the IRT models.

Consider the IRT model where P binary items measure a single latent variable. Observations are nested within groups, such as countries or regions, and all IRT model parameters are allowed to vary across the groups. This setup can be viewed as multiple-group IRT model or as multilevel IRT model. In multiple-group IRT models, the group specific parameters are treated as fixed, i.e., nonrandom parameters. In a multilevel IRT model, the parameters are treated as group specific random effects. Suppose that Y_{pij} is the p-th binary item for individual i in cluster or group j. The multilevel IRT model is given by the following equations:

$$P(Y_{pij} = 1) = \frac{1}{1 + Exp(\tau_{pj} - \lambda_{pj}\eta_{ij})} \tag{48}$$

$$\eta_{ij} = \eta_{wij} + \eta_{bj} \tag{49}$$

$$\eta_{wij} \sim N(0, Exp(\psi_j)) \tag{50}$$

$$\psi_j \sim N(0, \psi_w) \tag{51}$$

$$\eta_{bj} \sim N(0, \psi_b) \tag{52}$$

$$\tau_{pj} \sim N(\tau_p, \nu_p) \tag{53}$$

$$\lambda_{pj} \sim N(\lambda_p, w_p). \tag{54}$$

The difficulty parameter τ_{pj} varies across clusters and its mean across all the groups is the average difficulty parameter τ_p. The discrimination parameter λ_{pj} varies across clusters and its mean across all the groups is the average difficulty discrimination λ_p. The factor mean in group j is η_{bj}, i.e., we can estimate a group specific factor mean. Also $\eta_{wij} = \eta_{ij} - \eta_{bj}$ is the individual-specific factor deviation from the group specific mean η_{bj}. The factor variance in group j is $Exp(\psi_j)$, i.e., we can estimate a group specific factor variance through the random effects $\psi_j = \log(Var(\eta_{wij}|j))$. In a standard IRT model, the factor mean is fixed to 0 and the factor variance is fixed to 1 for identification purposes. These identification restrictions are now replaced by fixing the mean of the random effect η_{bj} to 0 and by fixing the mean of the random effect ψ_j to 0, respectively. The above model is able to estimate group specific factor mean and variance without having to assume metric or scalar

measurement invariance. The model has $4P + 2$ parameters: τ_p, λ_p, ν_p, w_p, for $p = 1, \ldots, P$ as well as ψ_w and ψ_b. The number of random effects in the model is 2P+2, which are given in equations (51–54).

We conduct a simulation study to illustrate the quality of the Bayesian estimation using the following parameter values $\tau_p = .3$, $\lambda_p = 1.4$, $\nu_p = w_p = \psi_w = \psi_b = 0.1$. Using $P = 8$ indicators, we generate 100 samples with 200 groups of size 30. The results of the simulation study for some of the parameters are reported in Table 3 and show minimal bias and coverage near the nominal level.

In practical applications, the above model should be followed up by further analysis. Random effects with insignificant variance estimates should be eliminated from the model. If the credibility interval of the variance of a random effect is very close to zero, it should be considered insignificant and the random effect should be replaced by a nonrandom group invariant parameter. The reduced model flexibility will actually improve the accuracy for all other model parameters.

It is important to note here that all the parameters in the above model are between level parameters. This implies that the sample size used in the estimation of the parameters is the number of groups. With very few groups, this model would not be feasible. For example, with less than 30 groups, the model would be difficult to estimate and even if it is possible to estimate it, the random effect variance estimates would depend on the priors. In addition, the standard errors are likely to be large and most if not all random effect variances would not be significant. When the number of groups is small, the alignment methodology described in Asparouhov and Muthén (2014b) can accomplish the same multiple group flexibility without relying on random effects.

Alternatively, with a small number of groups, one can pursue somewhat less-flexible multilevel IRT models along the line of traditional multilevel SEM models. One such formulation is as follows

$$P(Y_{pij} = 1) = \frac{1}{1 + Exp(\tau_{pj} - \lambda_{wp}\eta_{wij} - \lambda_{bp}\eta_{bj})} \quad (55)$$

$$\eta_{wij} \sim N(0, 1) \quad (56)$$

$$\eta_{bj} \sim N(0, 1) \quad (57)$$

$$\tau_{pj} \sim N(\tau_p, \nu_p). \quad (58)$$

This model does not include random loadings and random factor variances. The between factor η_{bj} can be interpreted as a cluster specific mean but only if the between and within level loadings are proportional. More specifically, if $\lambda_{bp} = s\lambda_{wp}$ for a proportionality scale parameter s, then $\lambda_{wp}\eta_{wij} + \lambda_{bp}\eta_{bj} = \lambda_{wp}(\eta_{wij} + s\eta_{bj})$. The last expression allows us to interpret $s\eta_{bj}$ as the cluster specific factor mean. The proportionality of the loadings can be tested with MODEL TEST in Mplus. If the proportionality is rejected, allowing for different within and between level loadings becomes an advantage over the multilevel IRT model given in (48–54).

Next we conduct a simulation study to evaluate the performance of model (55–58) for small samples. We use the following parameter values for this simulation study $\lambda_{wp} = 1.4$, $\lambda_{bp} = 0.5$, $\tau_p = 0.3$ and $\nu_p = 0.1$. Again using $P = 8$, we generate 100 samples for several combinations of number of groups and group sizes. In this simulation study, we also illustrate the effect of weakly informative priors for the random effect variance parameters ν_p. The Mplus default prior is the improper prior with density function of 1 on the interval from 0 to infinity, which is specified as $IG(-1, 0)$. As an alternative, we also estimate the model using two different weakly informative priors for ν_p. The first prior is $IG(3, .4)$, which has a mode at the true value of .1 and standard deviation of 0.2. The second prior is IG(3,1), which has a mode at 0.25 and standard deviation of 0.5.

The results of the simulation study for some of the parameters are reported in Table 4. Some bias can be found in the estimates and the bias generally appears to decrease depending on how the sample size is increased. If we increase the groups sizes, the within level loadings improve. If we increase the number of groups, the between level loadings improve. If the number of group increases, the bias will disappear even if the group sizes stay small. Such a result is in line with asymptotic theory, which guarantees unbiased model estimates as the number of groups increases, regardless of the size of the groups.

The weakly informative prior has a fairly substantial effect on the random effect variance parameter estimates for the sample with 40 groups. The results illustrate that the estimates can be improved with such priors. However, if the prior is not set well, the bias can actually become worse. This is exactly what happened with the $IG(3, 1)$ prior. The effect of the prior is intuitively easy to understand. If the noninformative prior yields a posterior distribution with standard deviation of 0.2 and we add a weakly informative prior with the same standard deviation, the estimate of the variance would then become approximately the average of the noninformative prior estimate and the median of the weakly informative prior. The underlying issue here is that little information is extracted from the data regarding the variance of the random threshold. Large standard errors for these parameters, relative to the size of the variance, essentially results in nonsignificant variance

Table 3. Two-level IRT.

Parameter	Absolute bias(Coverage)
τ_1	.01(.92)
λ_1	.03(.92)
ν_1	.01(.96)
w_1	.00(.95)
ψ_w	.01(.93)
ψ_b	.00(.95)

Table 4. Two-level IRT, multilevel SEM style: Absolute bias (Coverage).

Groups	40	40	40	40	80	200
Group sizes	15	15	15	30	15	15
ν_p priors	–	IG(3,0.4)	IG(3,1)	–	–	-
τ_1	.01(.98)	.00(.96)	.01(.99)	.02(.98)	.00(.95)	.00(.96)
λ_{w1}	.09(.93)	.08(.93)	.10(.91)	.02(.97)	.06(.94)	.02(.94)
λ_{b1}	.01(.98)	.03(.94)	.01(.96)	.04(.98)	.02(.93)	.01(.96)
ν_1	.10(.94)	.04(1.00)	.15(.07)	.04(.96)	.05(.92)	.02(.92)

Table 5. Simple two-level IRT: Absolute bias (Coverage).

Groups	40	20	10
Group sizes	15	20	30
τ_1	.02(.93)	.01(.93)	.03(.98)
λ_1	.03(.92)	.06(.92)	.09(.91)
v	.02(.98)	.04(.93)	.10(.90)

Table 6. Multilevel autoregressive model for count data, N = 500, T = 10.

Parameter	True Value	Absolute bias(Coverage)
v	.3	.01(.92)
ρ	.5	.01(.87)
α	.4	.02(.88)
v_1	.2	.00(.93)
v_2	.3	.02(.87)

parameters. In fact, in the particular setting of the first column of Table 4, if we fix the v_p parameters to zero, we get the same estimates for the remaining parameters. This also confirms that at this sample size and level of threshold noninvariance, the random effects τ_{pj} are not essential.

Next, we consider an even simpler multilevel IRT model. In this model the thresholds and the loadings are group invariant as well a the factor variance. Only the factor mean is allowed to vary across groups. The model is described as follows

$$P(Y_{pij}=1)=\frac{1}{1+Exp(\tau_p-\lambda_p\eta_{ij})} \tag{59}$$

$$\eta_{ij}=\eta_{wij}+\eta_{bj} \tag{60}$$

$$\eta_{wij}\sim N(0,1) \tag{61}$$

$$\eta_{bj}\sim N(0,v). \tag{62}$$

Because of its simplicity, the model can be estimated with very small samples fairly well. Using the earlier setup with $P=8$, $\tau_p=0.3$, $\lambda_p=1.4$, and $v=0.2$, we conduct a simulation study with varying number of groups and group sizes across 100 replications. The results for some of the parameters are reported in Table 5.

Multilevel autoregressive model for count data

Polson et al. (2013) describe an autoregressive model for count data. The model is based on single time-series data using the PG negative-binomial parameterization. Here we describe a multilevel version for that model where the data consists of multiple time-series count data for a group of individuals observed across time. We use the NB-2 and Poisson parameterizations for this illustration.

Suppose that Y_{it} is a count observation for individual i at time t. We consider the following autoregressive model:

$$Y_{it}\sim NB2(v_i+\eta_{it},\alpha) \tag{63}$$

$$v_i\sim N(v,v_1) \tag{64}$$

$$\eta_{it}=\rho\eta_{i,t-1}+\varepsilon_{it},\ \text{ for } t=2,\dots,T \tag{65}$$

$$\varepsilon_{it}\sim N(0,v_2). \tag{66}$$

The model has a total of five parameters: v, ρ, α, v_1, and v_2, and can be viewed as an adaptation of the DSEM methodology described in Asparouhov et al. (2017) to count data. The main attribute of this modeling approach is the ability to separates the correlation due to observations nested within person (modeled via parameter v_2) from the correlation that is due to observations taken in proximity of time (modeled via parameter ρ).

To evaluate the performance of the Bayesian estimation, we generate and analyze 100 data sets with $N=500$ individuals and $T=10$ observations. The results of the simulation study are reported in Table 6 and show that the estimator performs well.

The autoregressive coefficient in this model is time invariant. It is possible to estimate a model with time-specific autocorrelation coefficients, however, such a model is somewhat poorly identified and it is not recommended. The model is estimated in Mplus as a single level multivariate model. Full DSEM model flexibility as in Asparouhov et al. (2017) is currently not implemented. This has two implications. First, models with $T>50$ are going to be very slow to estimate. Second, the autocorrelation coefficient ρ cannot be subject specific. The above model also does not perform very well with smaller values of T. Using $T=5$ for example, yields larger biases and lower coverage.

In practical applications, especially when T is not large, it is recommended that an additional parameter is estimated for the variance of the initial starting value $\eta_{i,1}$. Assuming a stationary time-series model for $\eta_{i,t}$, the variance $Var(\eta_{i,t})=v_2/(1-\rho^2)$. At the first time point, however, $\eta_{i,1}$ is not predicted by another variable and thus the residual variance for ε_{i1} should be set to $v_2/(1-\rho^2)$. This would involve a complex parameter constraint, however. A simpler alternative is to just estimate the first residual variance as a separate model parameter, i.e. constrain the residual variance of $\eta_{i,t}$ to be time invariant only for times $2,\dots,T$.

The above model performs better with the Poisson distribution instead of the negative-binomial distribution, particularly for smaller sample sizes. That is because the autoregressive model is imposed on the latent variables and the variances of the latent variables are somewhat confounded with the dispersion parameters. Recall that a Poisson distribution where the mean parameter is a Gamma distributed random effect is a negative-binomial distribution, i.e. a random effect essentially introduces and models overdispersion. In the autoregressive model, the latent variables are primarily identified through the across-time correlations between the variables, i.e. the autoregressive process is truly identified by the across time correlations rather than by fitting the overdispersion in the variables. When the time series is short (T is small), however, the autoregressive information is weak (due to small T the model has difficulty in distinguishing between correlation that is due to observations nested within person and due to observations taken in proximity of time[2]) and could become conflated with the overdispersion.

[2]If $T=2$, there is only one correlation $Corr(Y_{i1},Y_{i2})$ which cannot identify both v_2 and ρ. If $T=3$, the model implied correlation based on v_2 is only marginally different from the model implied correlation based on ρ and to distinguish between the two a large sample size N is needed.

To illustrate this point we conduct a simulation study with $T = 5$ and $N = 100$ for auto-regressive Poisson model and autoregressive negative-binomial model. The autoregressive Poisson model is described as the autoregressive negative-binomial model given above with $\alpha = 0$, i.e.,

$$Y_{it} \sim Po(\nu_i + \eta_{it}) \tag{67}$$

$$\nu_i \sim N(\nu, \nu_1) \tag{68}$$

$$\eta_{it} = \rho\eta_{i,t-1} + \varepsilon_{it}, \text{ for } t = 2, \ldots, T \tag{69}$$

$$\varepsilon_{it} \sim N(0, \nu_2). \tag{70}$$

We generate 100 replications for each of the two distribution types and analyze the data with the same distribution type. The results of the simulation study are reported in Table 7. The Bayesian estimation for the Poisson autoregressive model performs well, while for the negative-binomial it does not. For small T, the Poisson autoregressive model is a well identified model that can be used in practice. The negative-binomial model for small T is somewhat poorly identified and cannot be recommended for practical applications. In the negative-binomial model, the dispersion parameter estimates are close to 0. This implies that the Poisson autoregressive model provides sufficient fit for the data even when it is generated by the negative-binomial autoregressive model when T is small.

The above model can be incorporated into a cross-lagged panel model (RI-CLPM) as in Hamaker et al. (2015). In particular, having a second time series, for example, with continuous items, could actually improve the identifiability of the model. That is because the cross-lagged relations can contribute to the measurement of the latent factors η_{it} used in the count time series.

Multilevel nominal regression

In this section we consider the nominal regression model in two-level settings. In particular, we illustrate the concepts of latent centering for the covariates, see Asparouhov and Muthén (2019), contextual effect, as well as random regression coefficients and intercepts.

Let N_{ij} be a nominal variable for individual i in cluster j and X_{ij} be the corresponding predictor. Let K be the number of unordered categories for the nominal variable. The two-level nominal regression model is given by the following equations:

$$X_{ij} = X_{w,ij} + X_{b,j} \tag{71}$$

Table 7. Multilevel autoregressive model for count data, N = 100, T = 5: Absolute bias (Coverage).

Parameter	True Value	Poisson	Negative-binomial
ν	.3	.00(.92)	.14(.58)
ρ	.5	.06(.94)	.26(.65)
α	.4	–	.31(.44)
ν_1	.2	.02(.88)	.03(.93)
ν_2	.3	.02(.90)	.28(.50)
convergence rate		100%	72%

$$X_{w,ij} \sim N(0, \psi_w) \tag{72}$$

$$X_{b,j} \sim N(\mu, \psi_b) \tag{73}$$

$$P(Y_{ij} = k) = \frac{Exp(\alpha_{jk} + \beta_{jk}X_{w,ij} + \gamma_k X_{b,j})}{\sum_{i=1}^{K} Exp(\alpha_{jk} + \beta_{jk}X_{w,ij} + \gamma_k X_{b,j})} \tag{74}$$

$$\alpha_{jk} \sim N(\alpha_k, \theta_k) \tag{75}$$

$$\beta_{jk} \sim N(\beta_k, \sigma_k). \tag{76}$$

The parameters α_{jK}, β_{jK} and γ_K are all fixed to zero for identification purposes. In this model, $X_{b,j}$ is the mean of X_{ij} in cluster j. This mean is an unobserved latent variable. In principle, it is possible to use the average of X_{ij} in cluster j as the mean. However, it has been shown that such an approach yields biased estimates due to not accounting for the measurement error in that average, see Lüdtke et al. (2008), Asparouhov and Muthén (2006), and Asparouhov and Muthén (2019). The variable $X_{w,ij}$ is the group centered covariate. In the above model, the effect of the covariate has two separate effects: the effect of the cluster mean $X_{b,j}$ as well as the effect of the group centered covariate $X_{w,ij}$. The difference between these two effects is called the contextual effect, see Lüdtke et al. (2008). If the covariate X_{ij} is not decomposed as the centered portion and the centering portion, and is used directly in the nominal regression, we would estimate an "uninterpretable blend" between the two different effects, see Raudenbush and Bryk (2002). Since $X_{w,ij}$ varies within cluster, we can estimate a cluster specific random effect β_{jk}. Similarly, the intercept in the nominal regression can be estimated as a random effect α_{jk}. The effect of $X_{b,j}$ cannot be cluster specific. The model has a total of $2K - 1$ random effects: α_{jk}, β_{jk} and $X_{b,j}$. Correlations between these random effects can be added to the above model with one exception. The correlation between α_{jk} and $X_{b,j}$ would not be identified because it is confounded with the parameters γ_k. If we want to estimate the correlation between these two random effects we must remove γ_k from the model, otherwise the model would not be identified.

We conduct a simulation study to evaluate the performance of the Bayesian estimation for the above model. Using $K = 3$, we generate 100 samples with 200 groups of size 30. In this simulation study we include the parameter $\rho = Cov(\alpha_{1j}, \alpha_{2j})$. The results, reported in Table 8, show that the bias is minimal and the coverage is near the nominal levels.

Model (71–76) has a large number of random effects and is expected to require a fairly large sample for optimal performance. For smaller samples, simpler two-level nominal regression should be explored. One such model is the model where the random slope variance σ_k is fixed to 0, i.e. the random effect for $X_{w,ij}$ is replaced by a nonrandom regression. We conduct a simulation study to evaluate the performance of the Bayesian estimation for that model using different number of clusters C and cluster sizes L. The results of this simulation are reported in Table 9. The bias is minimal and the coverage is near the nominal levels. Slightly larger bias is visible for the variance of

Table 8. Two-level nominal regression.

Parameter	True Value	Absolute bias(Coverage)
α_1	1	.01(.95)
α_2	.5	.00(.95)
β_1	.4	.01(.95)
β_2	−.4	.01(.96)
γ_1	.6	.01(.96)
γ_2	.2	.01(.96)
θ_1	.5	.05(.94)
θ_2	.5	.05(.92)
σ_1	.2	.02(.95)
σ_2	.2	.01(.97)
ρ	.2	.03(.97)
ψ_w	1	.00(.96)
ψ_b	.3	.03(.86)
μ	0	.00(.96)

Table 9. Two-level nominal regression with non random slopes: Absolute bias (Coverage).

Parameter	True Value	$C = 100, L = 15$	$C = 200, L = 20$
α_1	1	.02(.98)	.01(.93)
α_2	.5	.02(.94)	.00(.95)
β_1	.4	.01(.91)	.00(.91)
β_2	−.4	.00(.95)	.00(.96)
γ_1	.6	.01(.93)	.04 (.96)
γ_2	.2	.02(.97)	.00(.92)
θ_1	.5	.10(.96)	.03(.91)
θ_2	.5	.06(.96)	.03(.94)
ρ	.2	.04(.92)	.02(.93)
ψ_w	1	.00(.95)	.01(.95)
ψ_b	.3	.01(.91)	.01(.90)
μ	0	.01(.94)	.01(.95)

the random intercept when the number of clusters is 100. As for other multilevel models with smaller number of clusters, weakly informative priors can be used to reduce that bias.

Nominal factor analysis

Revuelta et al. (2020) discusses factor analysis models where all the factor indicators are nominal variable. Such models can be estimated with the ML estimator using numerical integration as long as the number of factors is not large. Models with up to three factors can be estimated fairly well with the ML estimator. However, for models with more than three factors, the ML estimation will be very slow and will be prone to convergence problems. The Bayesian estimator can be used as an alternative to the ML estimator, because it will not be limited by the number of factors. In this section we evaluate the performance of the Bayesian estimator for such models and provide some insights into this fairly novel modeling technique.

Using nominal factor indicators, instead of ordinal, is necessary for those situations when the outcome cannot be regarded as ordinal. Nominal variables are also a useful modeling alternative when the proportional odds ratio, assumed with the ordinal logistic model, does not hold. The nominal model, however, has some drawbacks as well and it should not be used routinely instead of ordinal models. One drawback is that the nominal model is less parsimonious. Many more parameters will be estimated with the nominal model, and very likely, some loss of power will occur. A nominal model will likely require bigger sample size than an ordinal model. Currently, there is no simple statistical tool that can be used to determine whether a variable should be treated as ordinal or as nominal. This is particularly the case in factor analysis where the predictors is not observed and the proportionality of the odds ratio cannot be easily evaluated. Potentially, one can lean into the substantive interpretation of the indicator and assume that the substantive interpretation is the correct choice. Such an approach, however, has some limitations as well. First, the substantive interpretation can be ambiguous and both nominal and ordinal may be viable options. Second, the substantive interpretation does not necessarily have to match with the best statistical model. This issue is particularly acute for small samples, where lack of power is bound to interfere with an attempt to make a rigorous choice. The Bayesian estimation of the nominal factor analysis model has one advantage over the ML estimation in this regard. Plausible values can be imputed for the factor, which can subsequently be used to study the nominal vs. ordinal nature of the variable in separate analysis.

Another practical aspect in the nominal vs. ordinal dilemma is the fact that the categories can be ordered in $L!$ different ways where L is the number of categories. This means that the comparison between the nominal model and ordinal model is essentially a comparison between the nominal model and $L!$ ordinal models. If the latent factor is imputed, log-likelihood comparison could potentially be used to guide an informed choice. Note, however, that if there is only one predictor for the categorical variable, the best ordering for the categories of the nominal variable is easy to obtain. Suppose that we have a nominal variable N regressed on a covariate X and the regression coefficients are $\beta_1, \ldots, \beta_{L-1}, \beta_L = 0$. To find the best ordering of the categories most suitable for an ordinal logistic regression of N on X, we have to order the categories so that the regression coefficients β_i become a monotonic sequence. Increasing or decreasing sequence works equally well. This is because in a logistic regression with a positive coefficient, higher values of the predictor implies an increase in the likelihood for the higher values of the categorical variable. In the nominal regression, such a relationship between the categorical variable and the covariate exist only when the regression coefficients are monotonic. To be more specific, if a set of data is generated from a logistic regression and is then analyzed with a nominal regression, the regression coefficients will appear in a monotonic order. Conversely, if we have a categorical variable that is treated as nominal, and we consider the question of whether or not the variable can be treated as ordinal variable, the best fitting model would be obtained when the categories are ordered so that the nominal regression coefficients are monotonic.

In the presence of multiple covariates, if the same category ordering yields monotonic regression coefficients for every covariate, then clearly that ordering would be optimal. However, it is certainly possible that the nominal regression coefficients of one covariate yields ordering that is different from the ordering implied by another covariate. This could be a clear sign that ordinal regression is not appropriate. However, the ordering of the coefficients is subject to these coefficients standard errors and the mismatch in the ordering of the coefficients may not be statistically significant. In such

situations, comparing the log-likelihoods among the $L!$ different ordinal logistic regressions may be necessary.

On the opposite spectrum of this discussion is the proportional odds assumption underlying the ordinal logistic model. Suppose that a categorical variable U is a measurement indicator for a factor η through an ordinal logistic regression. The model implies that

$$\log(P(U \leq j)/P(U>j)) = \alpha_j + \lambda\eta. \tag{77}$$

The proportional odds assumption refers to the fact that the coefficient λ is independent of j, i.e it is the same across all $j = 1, \ldots, L-1$, see Agresti (1990). If this assumption does not hold, the ordinal logistic measurement model should be replaced with a nominal measurement model.

Here we describe three different ways to test the proportional odds assumption for a latent variable η. The first method is as follows. Using the ML estimator, the factor analysis model is estimated as well as the factor scores. At this point, we can simply use Brant's (1990) test for the proportional odds assumption. This will be obtained in Mplus by estimating the logistic regression of U on the factor score with the ML estimator. The drawback of this method is that the uncertainty/measurement error in the factor score is not accounted for, which can result in underestimation of the p value in Brant's test. It would be preferable, to impute the factor, and then use Brant's test using the plausible values. However, currently Brant's test is not available in Mplus for multiple imputed data.

The second and the third method we describe here can be used with multiple imputed data. To obtain plausible values for the factor, the factor model must be estimated with the Bayesian estimator, using the probit link function for all ordinal indicators. Currently, the logit link function is not available in the Mplus Bayesian framework for nonbinary variables and thus we must switch to the probit link function for those indicators. It is unlikely that such a switch will create estimation issues for the purpose of testing the proportional odds assumption. Given the plausible values for the factor, the second method proceeds as follows. We define $L-1$ binary variables W_j for $j = 1, \ldots L-1$ as follows

$$W_j = \begin{cases} 0, & \text{if } U \leq j \\ 1, & \text{if } U>j. \end{cases} \tag{78}$$

The next step involves simultaneously estimating the $L-1$ logistic regressions of W_j on the plausible values η, using the multiple imputed data for η and the ML estimator. In this estimation, we include Wald's test for the hypothesis that all $L-1$ regression coefficients are equal. For the validity of Wald's test for multiple imputed data see Asparouhov and Muthén (2010b). The Wald's test in Mplus is specified using the MODEL TEST command and is our second method for testing the proportional odds assumption.

The third method also uses the multiple imputed data for the factor and Wald's test. For this method, we estimate the following nonproportional odds ratio logistic regression model

$$P(U = j) = \frac{1}{1 + Exp(-\tau_j + \beta\eta)} - \frac{1}{1 + Exp(-\tau_{j-1} + \beta\eta)}, \tag{79}$$

where as usual $\tau_0 = -\infty$, $\tau_L = \infty$. To make this model into a nonproportional odds ratio model, we introduce the following additional model constraints. For $j = 2, \ldots, L-1$

$$\tau_j = \alpha_j + \beta_j\eta. \tag{80}$$

In the ML estimation of Mplus, such model constraints are introduced using the CONSTRAINT option of the VARIABLE command where η is specified. In addition, the parameters α_j and β_j are declared as NEW parameters in the MODEL CONSTRAINT command, where also equation (80) is specified. Finally, the proportional odds ratio test is obtained using Wald's test on the hypothesis that $\beta_j = 0$ for $j = 2, \ldots, L-1$.

Next we illustrate the Bayesian estimation of the nominal factor analysis model with a simulation study. The model we consider has four nominal variables $N_1, \ldots, N_4$ and a covariate X. The latent factor η is measured by the four indicators and is predicted by the covariate. Let L_i denote the number of categories for the nominal variable N_i. In this example, we set $L_1 = L_2 = 3$, $L_3 = 2$, and $L_4 = 4$. The model is described by the following equations:

$$P(N_i = j) = \frac{Exp(\alpha_{ij} + \lambda_{ij}\eta)}{\sum_{j=1}^{L_i} Exp(\alpha_{ij} + \lambda_{ij}\eta)} \tag{81}$$

$$\eta = \beta X + \varepsilon. \tag{82}$$

For identification purposes, the intercept in (82) is fixed to zero and ε is assumed to have a standard normal distribution, i.e. the residual variance of the factor is fixed to 1. In addition, both α_{ij} and λ_{ij} are fixed to 0 when $j = L_i$. We generate 100 data sets with sample size N = 1000 and N = 2000 and the following parameter values $\alpha_{11} = .5$, $\alpha_{12} = -.5$, $\alpha_{21} = .7$, $\alpha_{22} = 0$, $\alpha_{31} = -1$, $\alpha_{41} = .4$, $\alpha_{42} = -.2$, $\alpha_{43} = .2$, $\lambda_{11} = 1$, $\lambda_{12} = .7$, $\lambda_{21} = .3$, $\lambda_{22} = .5$, $\lambda_{31} = -.5$, $\lambda_{41} = .6$, $\lambda_{42} = .3$, $\lambda_{43} = -.5$, $\beta = .4$. The results for a subset of the parameters are reported in Table 10. Some small biases are visible for some of the loadings parameter in the $N = 1000$ case and some rather large MSE values can be seen for those parameters. Increasing the sample size to $N = 2000$ appears to resolve both. The reduction in the MSE is dramatic. This indicates that the Bayesian estimation of the nominal factor analysis model may need larger samples for optimal performance. In fact, the ML estimator appear to perform better for smaller sample sizes. With the alternative factor analysis parameterization, where a loading is fixed and the factor variance is estimated, the Bayesian estimation appears to perform even worse

Table 10. Nominal factor analysis with 4 indicators: Absolute bias/Coverage/MSE.

Parameter	N = 1000	N = 2000	N = 1000 + Z
α_{11}	.01/.93/.01	.00/.94/.00	.00/.96/.01
α_{21}	.01/.94/.01	.01/.90/.00	.00/.93/.01
α_{31}	.00/.93/.01	.00/.95/.00	.01/.93/.01
α_{41}	.00/.96/.02	.00/.93/.01	.01/.91/.01
λ_{11}	.06/.92/.22	.01/.95/.03	.03/.91/.02
λ_{21}	.00/.94/.02	.01/.93/.01	.00/.97/.01
λ_{31}	.02/.94/.01	.00/.98/.01	.01/.94/.01
λ_{41}	.08/.89/.15	.02/.91/.03	.00/.94/.02
β	.03/.91/.01	.02/.93/.00	.00/.98/.00

Table 11. Nominal factor analysis with 10 indicators: Absolute bias (Coverage).

Parameter	N = 1000	N = 500	N = 300
α_{11}	.00(.96)	.00(.96)	.02(.95)
α_{12}	.00(.97)	.00(.94)	.00(.93)
λ_{11}	.02(.91)	.03(.96)	.07(.89)
λ_{12}	.02(.90)	.01(.89)	.01(.92)

in terms of convergence rates, quality of mixing, as well as bias and MSE.

For comparison purposes, we also conduct the following simulation study. To the above model with 4 nominal indicators, we add one additional continuous factor indicator Z. We set the mean of the indicator to 0, the factor loading to 1 and the residual variance to 1. We generate and analyze 100 data sets with $N = 1000$ observations. The results of this simulation study, also shown in Table 10, indicate that the biases in the parameters are resolved as well as the large MSE. The additional continuous factor indicator appears to have stabilized the Bayesian estimation. This simulation study also suggests that a factor analysis that includes not just nominal indicators, but a mixture of different types of indicators, may be the most practical modeling approach, particularly when the sample sizes are small or moderate. In fact, the estimation of the nominal factor analysis can be improved by also adding ordinal indicators, i.e. indicators that help with measuring the latent factor but are less demanding than nominal variables in terms of the number of additional parameters that are to be estimated. This further emphasizes the need to improve our understanding of when a categorical variable should be specified as nominal or as ordinal variable. The estimation is also improved by the presence of factor predictors, which in a way also improve the uncertainty in the measurement model.

The model estimation can also be improved by simply adding more nominal indicators. To illustrate this, we conduct a simulation study with 10 nominal indicators each with three categories, measuring 1 latent factor (and no covariates). We generate and analyze 100 samples with $N = 1000$, 500 and 300 observations, using the following parameter values: $\lambda_{i1} = 1$, $\lambda_{i2} = 0.5$ and $\alpha_{ij} = 0$. The results for a subset of the parameters are reported in Table 11. The bias is minimal and the coverage is near the nominal levels. Note, however, that in practical situations, some of the nominal indicators may carry very little information. If some of the categories are rare, the data will contain very little information about those category-specific parameters. In turn, those parameters will have large posterior distributions that likely will lead to slow convergence. To avoid such convergence issues it may be necessary to add stronger weakly informative priors for the problematic parameters. For example, the Mplus default prior of $N(0, 5)$ could be replaced by a prior of $N(0, 1)$. Having indicators with rare categories will naturally lead to poor estimation for these category-specific parameters, however, this will not compromise the estimation of the rest of the model parameters.

Conclusion

The Bayesian estimation described here for structural, multilevel and mixture models with logit, count and nominal variables provides a valuable alternative to the ML estimation which is often limited by the number of latent variables that can be included in the model. Many models that are computationally intractable with the ML estimation are now feasible with the Bayesian estimation. The methodology can be further combined with the BSEM technique described in Muthén and Asparouhov (2012) to enhance model fit and explore model modifications. The underlying latent variable technique facilitated by the PG methodology has expanded our ability to structurally model these new types of variables.

Some questions and challenges remain and clearly there are many opportunities for further methodological research. In this article, the underlying latent variables Y^* are treated only as an auxiliary estimation technique. Clearly, however, these variables contain information that could be used for model testing and modifications. Substantial residual correlations between Y^* and other model variables would indicate the need for modeling such correlations via additional latent variables. Note, however, that the Y^* correlations would not be identical to the model implied correlations obtained by the introduction of additional latent variables. The relationship between these two types of correlations must be explored further. The potential to include Y^* correlations in the structural model should not be ruled out completely as well, and it may be possible to do so in the future. Similarly, the question remain regarding how Y^* can be included as a moderator in the structural model. Furthermore, a discrepancy between the model estimated variance covariance for Y^* and the average sample variance covariance for Y^* (across the MCMC iterations) can be interpreted as model deficiency or the lack of it as evidence for well fitting model. This could potentially lead to a posterior predictive p value model test evaluation similar to what is available for SEM models with continuous outcomes. Other model fit evaluation techniques, based on the contingency tables for example, should be studied further as well.

In this expanded modeling framework, it is fairly easy to incorporate a large number of latent variables and random effects. Further research is needed to enhance our ability to test the statistical significance of these latent variables and random effects. The current most used methodology based on the credibility intervals of the variance parameters appears to work well only for large sample sizes. Further practical methodological development is clearly needed to address this issue.

Finally, we note again that currently there is no simple adaptation of the PG methodology for the case of ordinal logistic modeling. One such attempt is described in Montesinos-Lopez et al. (2015). However, the level of complexity appears to be prohibitive in terms of adapting that approach in a generalized framework such as the one implemented in Mplus.

References

Agresti, A. (1990). *Categorical data analysis*. John Wiley & Sons.

Asparouhov, T., Hamaker, E., & Muthén, B. (2017). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 359–388. https://doi.org/10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2006). *Constructing covariates in multilevel regression*. Mplus Web Notes: No. 11. Muthén & Muthén. http://www.statmodel.com/download/webnotes/webnote11.pdf

Asparouhov, T., & Muthén, B. (2010a). *Bayesian analysis using Mplus: Technical implementation*. Technical appendix. Muthén & Muthén. http://www.statmodel.com/download/Bayes3.pdf

Asparouhov, T., & Muthén, B. (2010b). *Chi-square statistics with multiple imputation (Mplus technical appendices: Version 2)*. Muthén & Muthén. http://www.statmodel.com/download/MI7.pdf

Asparouhov, T., & Muthén, B. (2014a). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 329–341. https://doi.org/10.1080/10705511.2014.915181

Asparouhov, T., & Muthén, B. (2014b). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 495–508. https://doi.org/10.1080/10705511.2014.919210

Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in realworld applications* (pp. 163–192). Information Age.

Asparouhov, T., & Muthén, B. (2019). Latent variable centering of predictors and mediators in multilevel and time-series models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*, 119–142. https://doi.org/10.1080/10705511.2018.1511375

Asparouhov, T., & Muthén, B. (2020a). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary second model*. Mplus Webnote No.21. Muthén & Muthén. https://www.statmodel.com/examples/webnotes/webnote21.pdf

Asparouhov, T., & Muthén, B. (2020b). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–14. https://doi.org/10.1080/10705511.2020.1764360

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*, 1171–1178. https://doi.org/10.2307/2532457

Camilli, G. (2017). The scaling constant D in item response theory. *Open Journal of Statistics*, *7*, 780–785. https://doi.org/10.4236/ojs.2017.75055

Fox, J. P. (2010). *Bayesian item response theory*. Springer.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge University Press.

Kang, T., Gaskins, J., Levy, S., & Datta, S. (2020). A longitudinal Bayesian mixed effects model with hurdle Conway-Maxwell-Poisson distribution. *Statistics in Medicine*. https://doi.org/10.1002/sim.8844

Kim, S., Lu, Z., & Cohen, A. (2018). An improved estimation using polya-gamma augmentation for Bayesian structural equation models with dichotomous variables. *Measurement: Interdisciplinary Research and Perspectives*, *16*, 1536–6367. https://doi.org/10.1080/15366367.2018.1437303

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*, 203–229. https://doi.org/10.1037/a0012869

Montesinos-Lopez, O. A., Montesinos-Lopez, A., Crossa, J., Burgue˜no, J., & Eskridge, K. (2015). Genomic-enabled prediction of ordinal data with Bayesian logistic ordinal regression. *Genes, Genomes, Genetics*, *5*, 2113–2126. https://doi.org/10.1534/g3.115.021154

Muthén, B., & Asparouhov, T. (2007). Growth mixture analysis: Models with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Advances in longitudinal data analysis*(pp. 143–165). Chapman & Hall/CRC Press.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, *17*, 313–335. https://doi.org/10.1037/a0026802

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups. *Sociological Methods & Research*, *47*, 637–664. https://doi.org/10.1177/0049124117701488

Neelon, B. (2019). Bayesian zero-inflated negative binomial regression based on polya gamma mixtures. *Bayesian Analysis*, *14*, 829–855. https://doi.org/10.1214/18-BA1132

Pillow, J., & Scott, J. (2012). Fully Bayesian inference for neural models with negative-binomial spiking. In P. Bartlett, F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 25, pp. 1907–1915). MIT Press.

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using polya-gamma latent variables. *Journal of the American Statistical Association*, *108*, 1339–1349. https://doi.org/10.1080/01621459.2013.829001

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Sage Publications.

Revuelta, J., Maydeu-Olivares, A., & Ximénez, C. (2020). Factor analysis for nominal (first choice) data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*, 781–797. https://doi.org/10.1080/10705511.2019.1668276

Windle, J., Polson, N. G., & Scott, J. G. (2014). *Sampling polya-gamma random variates: Alternate and approximate techniques*. arXiv preprint arXiv:1405.0506. https://arxiv.org/abs/1405.0506

Zhou, M., & Carin, L. (2015). Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*, 307–320. https://doi.org/10.1109/TPAMI.2013.211