Bayesian Structural Equation Modeling and Its Research Status

Authors: Lijin Zhang, Lu Jiaqi, Wei Xiayan, Pan Junhao, Pan Junhao

Date: 2018-12-27T00:00:00+00:00

Abstract

In psychological research, Structural Equation Modeling (SEM) is widely employed to test causal effects among latent variables, with estimation methods comprising two categories: frequentist approaches (e.g., maximum likelihood estimation) and Bayesian approaches. In recent years, due to the popularity of Bayesian statistics and its advantages in structural equation modeling—such as ease of handling small samples, missing data, and complex models—Bayesian structural equation modeling has developed rapidly; however, its application in the domestic field of psychology remains insufficient. This paper primarily introduces the methodological foundations and advantageous characteristics of Bayesian structural equation modeling, as well as several commonly employed Bayesian structural equation models and their current application status, aiming to present a new research tool for applied researchers.

Full Text

Bayesian Structural Equation Modeling and Its Current Research Status

ZHANG Lijin[1]; LU Jiaqi[1]; WEI Xiayan[1]; PAN Junhao[1]*
[1] Department of Psychology, Sun Yat-sen University, Guangzhou, 510006, China

Abstract: Structural Equation Modeling (SEM) is widely used in psychological research to examine causal effects among latent variables. Model estimation can be performed using either frequentist approaches (e.g., maximum likelihood estimation) or Bayesian methods. In recent years, due to the popularity of Bayesian statistics and its advantages in handling small samples, missing data, and complex models within SEM, Bayesian Structural Equation Modeling (BSEM) has developed rapidly. However, its application in domestic psychology research remains insufficient. This paper primarily introduces the methodological foundations and desirable properties of BSEM, as well as several commonly used

Bayesian structural equation models and their current applications, aiming to provide applied researchers with a new research tool.

Keywords: structural equation modeling; Bayesian estimation; maximum likelihood estimation

1. Introduction

In psychological, behavioral, and other social science research, investigators often study abstract constructs, traits, or factors such as intelligence, personality, attitudes, and abilities through latent variables (LVs) that cannot be directly observed. A latent variable typically corresponds to several related, directly observable manifest variables. Latent variables can be viewed as abstractions and generalizations of their corresponding manifest variables, while manifest variables serve as reflective indicators of specific latent variables. Structural Equation Modeling (SEM) is widely recognized as one of the most powerful modern statistical methods for analyzing relationships among latent variables and represents one of the most commonly used analytical approaches in psychology, education, and behavioral sciences (Hou, Wen, & Cheng, 2004; Lee & Song, 2012; Wang, 2014). A key reason for SEM' s popularity lies in its flexibility—many models, including confirmatory factor analysis models, mediation models, and latent growth curve models, can be expressed within the SEM framework.

Traditional SEM comprises two components: a measurement model and a structural model. The measurement model links manifest variables to their corresponding latent variables, typically taking the form of a Confirmatory Factor Analysis (CFA) model, which is primarily used to examine the factor structure of latent variables. It is defined as follows:

$$y_i = \mu + \Lambda\omega_i + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{1}$$

where $y_i(p \times 1)$ represents the observed values of the i-th subject on p related manifest variables, $\mu(p \times 1)$ is the intercept term, $\Lambda(p \times q)$ is the factor loading matrix reflecting the relationship between manifest variables y_i and latent variables $\omega_i(q \times 1)$, and $\varepsilon_i(p \times 1)$ is the measurement error term for manifest variables, which follows a $N[0, \Psi_\varepsilon]$ distribution. The model includes the following assumptions: no correlation exists between measurement errors and latent variables; the variance-covariance matrix of measurement errors Ψ_ε is a diagonal matrix, meaning no correlation among measurement errors. Traditional measurement models also typically assume that the number of latent variables and the relationships between latent and manifest variables are known, and that each manifest variable loads on only one latent variable without cross-loadings.

The structural model is primarily used to analyze the "causal effects" among latent variables. It is defined as follows: Let $\omega_i = (\eta_i^T, \xi_i^T)^T$, where $\eta_i(q_1 \times 1)$ represents endogenous latent variables and $\xi_i(q_2 \times 1)$ represents exogenous latent variables:

$$\eta_i = \Pi\eta_i + \Gamma\xi_i + \delta_i = \Lambda_\omega\omega_i + \delta_i \tag{2}$$

where $\Lambda_\omega = (\Pi, \Gamma)$, $\Pi(q_1 \times q_1)$ and $\Gamma(q_1 \times q_2)$ are path coefficient matrices. Π reflects effects among endogenous latent variables, while Γ reflects effects of exogenous latent variables on endogenous latent variables. ξ_i follows a $N[0, \Phi_\xi]$ distribution, and $\delta_i(q_1 \times 1)$ is the residual term, which follows a $N[0, \Psi_\delta]$ distribution.

The analytical steps for structural equation modeling typically include model specification and identification, model fit assessment, model modification, and parameter estimation. Compared with traditional path analysis methods, structural equation modeling accounts for measurement errors in manifest variables, providing more accurate estimates of relationships among latent variables (Li, 2011).

2. Bayesian Structural Equation Modeling

The primary estimation methods for structural equation models include frequentist approaches (e.g., maximum likelihood estimation) and Bayesian methods. Although frequentist methods are currently more widely applied, research on Bayesian structural equation modeling has grown steadily in recent years due to the popularity of Bayesian methods and their numerous advantages in statistical modeling. In particular, since 2012, the number of applied studies using Bayesian structural equation modeling has increased substantially (Van de Schoot et al., 2017).

The essential difference between Bayesian and frequentist approaches lies in their treatment of unknown parameters: frequentist methods treat unknown parameters as fixed constants and estimate population parameters from sample parameters, whereas Bayesian methods treat unknown parameters as random variables, with the goal of obtaining the posterior distribution of unknown parameters. When analyzing SEM using Bayesian methods, researchers can specify prior distributions for unknown parameters or latent variables based on theory or previous research findings. When precise prior information is unavailable, non-informative prior distributions (e.g., uniform distributions) or vague informative prior distributions (e.g., normal distributions with large variances) can be specified. According to Bayes' theorem, combining prior distributions with the data likelihood function yields the posterior distribution of unknown parameters and latent variables. Markov Chain Monte Carlo (MCMC) algorithms (e.g., Gibbs sampling and Metropolis-Hastings algorithms) are then used to iteratively draw a large number of samples from the posterior distribution. These samples are used to estimate the mean, credible intervals, and other statistics of the posterior distribution for statistical inference (Li, 2011).

The specific steps for analyzing SEM using Bayesian methods include:

(1) **Model specification and providing prior information for unknown parameters:** Let $k = 1, \ldots, p$, $h = 1, \ldots, q_1$. Conjugate prior distributions for different unknown parameters in SEM can be specified as follows (Li, 2011):

$$\mu \sim N(\mu_0, \mathbf{H}_{\mu 0}), \quad \Lambda_k \sim N(\Lambda_{0k}, \mathbf{H}_{0k}), \quad \Lambda_{\omega h} \sim N(\Lambda_{0\omega h}, \mathbf{H}_{0\omega h}), \quad \Phi \sim \text{Wishart}(\mathbf{R}_0, \rho_0) \tag{3}$$

where Λ_k^T is the k-th row of the loading matrix Λ, $\Lambda_{\omega h}^T$ is the h-th row of the path coefficient matrix Λ_ω, and $\mu_0, \Lambda_{0k}, \Lambda_{0\omega h}, \rho_0$ and positive definite matrices $\mathbf{R}_0, \mathbf{H}_{\mu 0}, \mathbf{H}_{0k}, \mathbf{H}_{0\omega h}$ are hyperparameter values specified based on theory or previous research findings, reflecting prior information and researchers' confidence in its accuracy.

(2) **Setting MCMC algorithm iteration numbers and assessing convergence before model fit evaluation and parameter estimation.** Convergence can be assessed through autocorrelation plots, trace plots, and potential scale reduction factors (see Wang et al., 2017 for details).

(3) **Overall model-data fit can be evaluated through Posterior Predictive Checking.** This yields the Posterior Predictive p-value (PPp-value), which differs from p-values in hypothesis testing. It represents the proportion of iterations in MCMC algorithms where a test statistic based on the theoretical model exceeds the test statistic from the sample data. Therefore, a PPp-value around 0.5 (close to the random probability of 1/2) indicates good model fit. Posterior predictive checking also provides a 95% confidence interval for the difference in test statistics between sample data and model-generated data. When the lower bound of this interval is negative and 0 falls near the center of the interval, the model fits well (Muthén & Asparouhov, 2012).

(4) **To avoid the influence of subjectivity in prior information, researchers can conduct Sensitivity Analysis (Greenland, 2001) to examine whether estimation results remain stable under different prior specifications, thereby enhancing result reliability.**

Compared with traditional methods, Bayesian structural equation modeling offers numerous advantages: (1) Bayesian sampling-based methods are less reliant on large-sample asymptotic theory than frequentist methods, thus performing well even with small samples (Muthén & Asparouhov, 2012); (2) Bayesian SEM analysis is based on raw observed values rather than covariance matrices, making it easier to handle complex model and data situations such as missing data and nonlinear relationships among latent variables, where traditional methods often encounter model identification problems (Li, 2011); (3) Bayesian methods can provide more effective statistics for model fit assessment, model comparison, and parameter estimation (Pan, Ip, & Dubé, 2017); (4) Bayesian methods can flexibly incorporate prior information, such as pilot studies and previous research findings, into model estimation, and effective prior information can yield more accurate parameter estimates (Yuan & MacKinnon, 2009; Zhang, Hamagami, Wang, & Nesselroade, 2007).

Despite these advantages that can better meet applied researchers' needs in empirical studies—such as handling complex models and small sample problems—Bayesian structural equation modeling remains underutilized in domestic psychology research. This paper introduces several commonly used Bayesian structural equation models and their application progress, including confirmatory factor analysis, structural equation models, mediation models, latent growth curve models, multiple-group models, and multilevel models, aiming to introduce new research tools to applied researchers and promote their application in domestic psychology research.

3.1.1 Bayesian Confirmatory Factor Analysis

Confirmatory factor analysis is commonly used to verify relationships between manifest and latent variables based on theoretical hypotheses. In traditional CFA models, the local independence assumption requires that manifest variables be uncorrelated after conditioning on latent variable values, meaning all off-diagonal elements in the measurement error variance-covariance matrix Ψ_ε are constrained to 0. Additionally, traditional CFA models generally do not permit cross-loadings. However, in practical research, these strict constraints often lead to poor model fit or even model rejection (Muthén & Asparouhov, 2012). Some researchers have noted that restrictions imposed by traditional methods are overly strict and even unnecessary, as they can easily reject models that actually fit the data well in large samples (Lu, Chow, & Loken, 2016; Marsh et al., 2009; Muthén & Asparouhov, 2012). Moreover, studies have found that imposing excessive constraints in actual data analyses can reduce the accuracy of unknown parameter estimates (Asparouhov & Muthén, 2009; Hsu, Troncoso Skidmore, Li, & Thompson, 2014).

In traditional methods, to address problems caused by these restrictions, researchers often combine theoretical considerations with modification index recommendations (Modification Index; Sörbom, 1989) to add cross-loadings or correlations among measurement errors to the model. However, this modification index-based approach has several limitations: (1) the modification process is time-consuming and tedious when many parameters require modification, as they must be adjusted individually; (2) it can easily lead to model overfitting and reduce generalizability (Maccallum, Roznowski, & Necowitz, 1992); (3) it is difficult to find a globally optimal model (Chou & Bentler, 1990); and (4) it can increase Type I error rates (Draper, 1995).

Muthén and Asparouhov (2012) innovatively proposed a Bayesian confirmatory factor analysis model that combines exploratory and confirmatory approaches, relaxing restrictions on measurement error correlations or cross-loadings. In traditional methods, cross-loadings and measurement error correlations are strictly constrained to 0, both for model parsimony and because freely estimating these parameters can lead to model non-identification. However, Muthén and Asparouhov's (2012) method maintains model identifiability while relaxing these restrictions by providing cross-loadings with a normal prior distribution with mean 0 and extremely small variance, or by providing appropriate inverse Wishart distributions for error term matrices. Simulation studies have shown that when relaxing restrictions on cross-loadings or measurement error correlations, this method yields fewer significant cross-loadings or measurement error correlations than the modification index approach, and satisfactory model fit can be achieved in a single analysis, whereas traditional methods typically require multiple modifications.

However, Muthén and Asparouhov' s (2012) method, while relaxing model restrictions, can also produce numerous non-zero cross-loadings or measurement error correlations (Lu, Chow, & Loken, 2016), resulting in overly complex factor loading or error term matrices. Consequently, models are prone to overfitting, complicating the interpretation and replication of research findings.

Lu et al. (2016) noted that Muthén and Asparouhov' s (2012) method essentially applies Bayesian Ridge regularization to CFA models. To address the aforementioned problems, Lu et al. (2016) introduced another Bayesian regularization method: by providing spike-and-slab prior distributions for the loading matrix, important cross-loadings are retained while other weak cross-loadings are shrunk toward zero. This approach avoids the model overfitting that Ridge regularization may cause and prevents excessive shrinkage of important cross-loadings.

Pan et al. (2017) introduced the covariance Lasso (Least absolute shrinkage and selection operator) regularization method for error term variance-covariance matrices in CFA models. By estimating a sparse error covariance matrix, this method relaxes restrictions on measurement error correlations while shrinking weak, unimportant measurement error correlations toward zero, avoiding model overfitting or non-positive definite error term matrices caused by excessive measurement error correlations. Their empirical research found that allowing "a small number" of measurement error correlations satisfied both model parsimony and fit.

Since the introduction of this relaxed restriction approach, applied research using Bayesian CFA for data analysis has increased substantially. Leveraging Bayesian methods' excellent properties in handling complex models, Golay, Reverte, Rossier, Favez, and Lecerf (2013) reanalyzed the four-factor structure of the Wechsler Intelligence Scale, testing both second-order and bifactor models, with results showing that Bayesian methods performed better than maximum likelihood estimation in both model identification and estimation. Falkenström et al. (2015) found that when examining the structural validity of the patient version of the Working Alliance Inventory, maximum likelihood estimation indicated poor model fit, but Bayesian methods that relaxed restrictions on measurement error correlations showed good model-data fit. Additionally, because

Bayesian methods provide more accurate parameter estimates in small samples (Muthén & Asparouhov, 2012), Crenshaw, Christensen, Baucom, Epstein, and Baucom (2016) used Bayesian CFA to revise the Communication Patterns Questionnaire in a clinical small-sample study (52 subjects, 18 unknown parameters including 9 factor loadings and 9 measurement error variances). Furthermore, since Bayesian methods estimate factor scores more accurately than traditional maximum likelihood methods in small samples (Muthén & Asparouhov, 2012), Alessandri and De Pascalis (2017) used Bayesian CFA to estimate factor scores for the "life orientation" factor in 51 subjects in an EEG experimental study, which were then used in subsequent analyses of relationships among factors.

3.1.2 Bayesian Structural Equation Models

Building upon measurement models, structural models are used to examine "causal effects" among latent variables unaffected by measurement error. As described in Section 2 of this paper, when estimating structural equation models using Bayesian methods, appropriate prior distributions can be specified for unknown parameters and latent variables, which are then combined with data to obtain their posterior distributions for statistical inference.

With relaxed restrictions on measurement models, estimation of unknown parameters in structural models also becomes more accurate. Pan et al. (2017) employed Bayesian Lasso methods to estimate error covariance matrices when establishing measurement models and found that, compared with traditional methods, Bayesian Lasso produced smaller estimation bias for path coefficients in structural models.

Moreover, in structural model estimation, Bayesian methods can provide prior distributions for all path coefficients. Similar to Bayesian confirmatory factor analysis, Muthén and Asparouhov (2012) suggested that path coefficients originally fixed at 0 could be relaxed by providing them with normal prior distributions with mean 0 and extremely small variance, enabling simultaneous model exploration and validation in a single estimation.

To facilitate researchers' application of this method, we provide a detailed description of the implementation steps for using informative priors in both measurement and structural models simultaneously, using Muthén and Asparouhov' s (2012) Bayesian structural equation model as an example. This study reanalyzed Kaplan' s (2009) longitudinal research, which collected data from 6,677 public school tenth-grade students. The established model is shown in Figure 1 [Figure 1: see original paper]. The model included the following hypotheses: students' sixth-grade science achievement and family socioeconomic status influence their science achievement through tenth-grade science achievement; teachers' science certification level influences students' science achievement through student engagement, which affects perceived curriculum challenge, ultimately impacting science achievement through tenth-grade science achievement.

Figure 1. Structural equation model of science achievement (adapted from

Kaplan (2009))

The researchers found poor model fit using traditional maximum likelihood estimation, with the hypothesized model being rejected (RMSEA = 0.081, CFI = 0.844). Modification indices suggested numerous parameters requiring modification. Under non-informative priors, Bayesian estimation yielded similar results, with posterior predictive checking intervals not including zero, indicating poor model fit (95% CI [1644, 1720]).

Based on the previously described approach of relaxing model restrictions, after standardizing the data, Muthén and Asparouhov (2012) provided normal prior distributions with mean 0 and variance 0.01 for 11 parameters originally fixed at 0 in the structural model, allowing 95% of the variation in these parameters to fall within (-0.2, 0.2). Since the data were standardized, these prior distributions theoretically covered the range of potentially omitted effects well. The researchers also provided the same prior distributions for 18 path coefficients from three variables (sixth-grade science achievement, family socioeconomic status, and teacher science certification level) to six observed indicators. In traditional methods, freely estimating these path coefficients would imply differential item functioning across the six observed indicators. Here, however, the researchers merely relaxed strict restrictions by providing normal prior distributions with mean 0 and extremely small variance.

Additionally, the researchers relaxed restrictions on cross-loadings and measurement error correlations for the six observed indicators. Cross-loadings received normal prior distributions with mean 0 and variance 0.01, while error term matrices received inverse Wishart distributions IW(I, 30). The corresponding Mplus code for the model can be found in the supplementary materials of Muthén and Asparouhov (2012).

Results showed improved model fit compared to the model without relaxed restrictions, with posterior predictive checking intervals including 0 (95% CI [-24, 44]) and a posterior predictive p-value of 0.276. The analysis also identified six significant structural model parameters and 11 significant measurement error correlations, though their estimates were all small[1]. Muthén and Asparouhov (2012) noted that the relaxed model would be non-identifiable using traditional maximum likelihood estimation, but this could be avoided within the Bayesian framework. They also noted that convergence rates decreased after relaxing model restrictions, requiring more iterations and longer computation times to obtain model estimates. Interested readers may refer to Muthén and Asparouhov (2012) for further discussion of these issues.

Furthermore, an increasing number of applied researchers have begun adopting this approach in empirical studies. Scherer, Siddiq, and Teo (2015), based on conceptual overlap among the latent variables studied, argued that cross-loadings must be considered in modeling, but traditional methods encounter model identification or estimation problems when adding too many cross-loadings. Drawing on Bayesian structural equation modeling principles, Scherer et al. (2015) used Bayesian methods to relax measurement model restrictions and subsequently established structural models to examine relationships between teachers' use of information technology and teacher characteristics. Salarzadeh, Moghavvemi, Wan, Babashamsi, and Arashi (2017) similarly employed this method to study factors influencing students' intention to use e-learning platforms, comparing maximum likelihood and Bayesian estimation results and finding that Bayesian estimation yielded more accurate model estimates with smaller root mean square errors and absolute mean errors.

3.1.3 Bayesian Mediation Models

Mediation analysis plays a crucial role in psychology, as it can explain the mechanisms through which independent variables affect dependent variables, facilitating the validation of existing theories and the construction of new ones (Luo & Jiang, 2014). When testing mediation effects, the key is to examine the significance of indirect effects—that is, whether the independent variable "significantly" affects the dependent variable through the mediator. Before Yuan and MacKinnon' s (2009) Bayesian mediation modeling approach, most mediation analyses in psychological research were conducted within the frequentist framework (Nuijten, Wetzels, Matzke, Dolan, & Wagenmakers, 2015).

Currently, commonly used methods for testing indirect effect significance in the frequentist framework include the Sobel test and Bootstrap method. The Sobel test directly examines the significance of the product of direct effect coefficients ($\mathrm{H}: ab = 0$; where a and b are the two direct effect coefficients), but this method requires the assumption that the product of coefficients follows a normal distribution; otherwise, estimates will be biased. This assumption is often difficult to satisfy in actual data analysis, but the Bootstrap method avoids this problem by constructing interval estimates (Wen & Ye, 2014).

Mediation analysis within the Bayesian framework uses MCMC methods for parameter estimation based on samples drawn from posterior distributions. After obtaining posterior distributions of direct effect coefficients, various functional forms of parameters can be easily estimated, such as the product of two direct effect coefficients, and credible intervals for indirect effects can be readily constructed (Yuan & MacKinnon, 2009). Therefore, Bayesian methods easily handle complex mediation models, such as Bayesian serial mediation models (Tofighi & Mackinnon, 2016) and moderated mediation models (Wang & Preacher, 2015), whereas traditional methods often encounter model identification and estimation problems with complex models (Kenny, Korchmaros, & Bolger, 2003).

Moreover, Bayesian mediation models do not require the assumption that the product of coefficients follows a normal distribution. Previous research has found that, except under non-informative prior conditions where Bayesian and traditional methods produce similar results, Bayesian methods demonstrate higher statistical power and more accurate parameter estimation under informative and vague informative prior conditions (MacKinnon, Lockwood, & Williams, 2004; Tofighi & MacKinnon, 2011; Tofighi & Mackinnon, 2016), and perform better in small samples (Miočević, MacKinnon, & Levy, 2017).

Due to the importance of mediation analysis and the excellent properties of Bayesian mediation models, an increasing number of studies have adopted Bayesian methods to test mediation effects. Given that Bayesian analysis does not rely on normal distribution assumptions for parameters, Shuck, Zigarmi, and Owen (2015) used Bayesian multiple mediation models to test the mediating effects of work engagement, employee dedication, and work passion in the relationship between basic psychological needs and work intentions. Due to small sample sizes, Zeman, Dallaire, Folk, and Thrash (2017) similarly used this method to test the mediating effect of children' s emotion regulation in the relationship between incarceration risk experiences, environmental risk, and childhood mental disorders. Additionally, combining Bayesian mediation modeling with Bayesian CFA methods that relax measurement model restrictions, Jacobson, Lord, and Newman (2017) verified that anxiety affects depressive symptoms through the mediation of emotional intelligence.

3.2 Bayesian Latent Growth Curve Models

Longitudinal research, also known as tracking research, involves long-term follow-up observations of research subjects with repeated measurements of relevant variables, enabling observation of relatively complete developmental processes and identification of key turning points in development. Compared with cross-sectional research, the greatest advantage of longitudinal research design is its ability to reasonably infer causal relationships among variables, making it an important research method in psychology (Liu & Meng, 2003).

Currently, among structural equation modeling approaches for analyzing longitudinal data, Latent Growth Curve Models (LGCM; Bollen & Curran, 2006) are widely applied. These models can analyze not only overall average change trends but also individual differences in developmental trajectories and how these differences are influenced by predictor variables.

LGCM is suitable for longitudinal research data observed at several fixed time points (Liu & Meng, 2003). A simple unconditional linear latent growth curve model is defined as follows (Wang, Wang, & Jiang, 2011):

$$y_{ti} = \eta_{0i} + \lambda_t \eta_{1i} + \varepsilon_{ti} \tag{4}$$

$$\eta_{0i} = \eta_0 + \zeta_{0i} \tag{5}$$

$$\eta_{1i} = \eta_1 + \zeta_{1i} \tag{6}$$

where y_{ti} is the observed variable data for individual i at time point t, η_{0i} and η_{1i} are random coefficients describing characteristics of change over time (also called growth factors), λ_t is the time score, and ε_{ti} is the composite error term at time point t, representing random measurement error and individual-specific time effects for subject i. In equations (5) and (6), η_0 and η_1 reflect the average initial level and average rate of change of the variable, respectively; thus η_{0i} and η_{1i} are called latent intercept and latent slope growth factors. The variances of ζ_{0i} and ζ_{1i} reflect between-individual variation among research subjects.

Additionally, researchers can incorporate predictor variables into the model to predict these two latent growth factors. If the developmental trajectory is nonlinear, nonlinear trajectories can be examined by introducing second-order or higher-order functions of time scores or by specifying free time scores (Meredith & Tisak, 1990).

Bayesian methods offer many advantages over traditional estimation methods when estimating latent growth curve models. First, longitudinal data situations are often complex, with unequal measurement intervals, nonlinear developmental trajectories, and non-normal data distributions, sometimes requiring complex data transformations. Bayesian estimation relies on raw observed values rather than variance-covariance matrices, making it easier to handle such data transformations (Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin, 2014). Second, traditional methods often encounter estimation problems or even model identification issues when handling complex LGCMs, whereas Bayesian methods can estimate complex models more effectively (Zhang, Hamagami, Wang, & Nesselroade, 2007). Third, missing data are common in longitudinal research (Ye, Tang, Zhang, & Cao, 2014), and Bayesian methods perform better in handling missing data (Li, 2011). Additionally, Bayesian methods perform better in small-sample situations (Gelman, Carlin, Stern, & Rubin, 2003). Zhang et al. (2007) found that Bayesian methods could estimate latent growth curve models with as few as 20 subjects measured at four different time points, making Bayesian methods extremely valuable for application.

In empirical research, Winans-Mitrik et al. (2014) noted that Bayesian methods provide more stable inferences than traditional frequentist methods that rely on hypothesis testing, better meeting clinical assessment requirements. They used Bayesian LGCM to verify the promoting effect of an aphasia intervention program on language comprehension, finding reliable and sustained therapeutic effects after four measurements of subjects' aphasia severity. Additionally, because Bayesian methods perform better with small samples and non-normal data (Gelman et al., 2014), Maier, Bohlmann, and Palacios (2016) used Bayesian LGCM to analyze cross-language associations in children' s language development, finding that the interaction between English and Spanish vocabulary expression abilities affected the growth of English vocabulary expression ability. We use this study as an example to detail the modeling steps for Bayesian LGCM.

The study' s subjects were 177 preschool children from low- to middle-income families in the Los Angeles area. Researchers followed up with them at three time points: the beginning of the fall semester, four months later, and six months later. At each follow-up, children' s English and Spanish vocabulary expression and reception abilities were measured through language tests and behavioral coding by coders, and multiple covariates were assessed, including children' s demographic variables, teacher and classroom characteristics, and classroom and home language environments.

Researchers handled missing values using Bayesian methods, treating missing observations as unknown values to be estimated. In modeling analyses, unconditional latent growth curve models were first established separately for four outcome variables: English vocabulary reception ability, English vocabulary expression ability, Spanish vocabulary reception ability, and Spanish vocabulary expression ability (Figure 2 [Figure 2: see original paper]).

The model constructed two growth factors for each variable, with the covariance between intercept and slope factors freely estimated. Each factor had measurements at three time points as their observed indicators. For slope factors, factor loadings at the three time points were fixed at 0, 0.4, and 1 according to linear relationships based on measurement intervals. Residuals for observed indicators at the three time points were allowed to be freely estimated. The means of intercept and slope growth factors reflected the overall average initial values and rates of change for the variables, while factor variances reflected between-individual differences. In model estimation, researchers used Bayesian estimation in Mplus 6.11 with default non-informative prior distributions. After establishing unconditional models, researchers further incorporated covariates into the models to predict intercept and slope factors, establishing conditional latent growth curve models.

Results showed that the four unconditional models for English vocabulary reception ability, English vocabulary expression ability, Spanish vocabulary reception ability, and Spanish vocabulary expression ability all fit well, with PPp values of 0.56 (95% CI [-13.51, 10.46]), 0.26 (95% CI [-8.38, 16.85]), 0.53 (95% CI [-11.33, 9.66]), and 0.48 (95% CI [-12.67, 11.81]), respectively. All slope factor estimates were greater than 0, indicating that all four vocabulary abilities increased over time. Significant between-individual differences were observed in both initial levels and rates of change across the four abilities (i.e., variances of intercept and slope factors were significantly different from 0).

Conditional models with covariates showed PPp values of 0.31 (95% CI [-26.96, 52.53]), 0.21 (95% CI [-23.84, 58.95]), 0.29 (95% CI [-32.40, 51.26]), and 0.34 (95% CI [-28.35, 52.86]), respectively, all indicating good model fit. Parameter estimation results showed that age positively predicted intercept factors in all four conditional models, indicating that older preschool children tended to have higher initial vocabulary levels. English exposure levels in the home also positively predicted children' s English vocabulary abilities but negatively affected initial Spanish vocabulary abilities. Additionally, due to Bayesian methods' advantages in estimating factor scores, researchers obtained each subject' s values on intercept and slope factors under Bayesian LGCM to further explore interconnections between English and Spanish vocabulary abilities, though detailed discussion is omitted here due to space limitations.

Figure 2. Unconditional latent growth curve model (using English vocabulary reception ability as an example)

The above example briefly introduces the application of Bayesian methods in latent growth curve models. Currently, in most studies employing Bayesian LGCM, researchers simply change the estimation method to Bayesian estimation and often provide software-default non-informative priors for parameter estimation. Future research could examine the impact of appropriate prior distributions on parameter estimation in latent growth curve models to further leverage Bayesian methods' advantages in model fit and parameter estimation.

3.3 Bayesian Multiple-Group Structural Equation Models

In empirical research, researchers often divide populations into multiple groups based on certain variables, such as gender groups. In such cases, data typically consist of few groups with many observations per group, where within-group observations are independent—i.e., multiple-group data. In structural equation modeling of multiple-group data, researchers can study similarities and differences among models across groups by establishing group-specific models. Multiple-group structural equation modeling focuses on testing various invariance assumptions across groups (Li, 2011).

Invariance assumption testing must first be conducted in measurement models, i.e., measurement invariance assumptions. Testing measurement invariance can be performed through a series of nested steps, sequentially including configural invariance (same factor structure across groups, meaning scale items load on the same latent variables across groups), metric invariance (equal item loadings across groups), scalar invariance (equal item intercepts across groups), and error variance/covariance invariance (equal item error variances across groups) (Vandenberg & Lance, 2000). Measurement invariance forms the basis for comparing latent variable-level parameters across groups, i.e., testing structural invariance, which includes factor variance/covariance invariance and factor mean structure invariance (Wang et al., 2011). When models do not meet cross-group scalar invariance, differences in factor scores across groups typically cannot clearly reflect true differences at the latent variable level (Schmitt & Kuljanin, 2008).

When testing invariance, traditional methods first require establishing a baseline model for each group, then constructing a configural model that integrates the baseline models across groups. Corresponding parameters (e.g., loadings, intercepts) are then constrained to be invariant across groups in the configural model. Model fit changes are used to determine whether parameters violate cross-group invariance. If model fit changes significantly, parameters must be modified according to modification index recommendations (Wang et al., 2011). When the proportion of parameters violating invariance is small relative to the number of parameters constrained to be invariant, such modifications have minimal impact on parameter estimation accuracy. However, when many groups are involved, parameters violating invariance are typically numerous. In such cases, model modifications can easily lead to biased parameter estimates. Additionally, traditional methods require modifying parameters individually according to modification indices, resulting in a tedious modeling process, complex models, and often poor model fit and biased parameter estimates (Asparouhov & Muthén, 2014).

Bayesian methods handle situations with many groups effectively. They relax strict cross-group invariance restrictions by providing appropriate prior distributions for parameter differences across groups (e.g., normal distributions with mean 0 and extremely small variance), allowing small cross-group differences to exist to some extent. This avoids problems of poor model fit caused by overly strict restrictions while providing more accurate parameter estimates (Asparouhov & Muthén, 2014). Additionally, Bayesian methods perform better with small-sample multiple-group data (Kim et al., 2017).

Considering that traditional methods impose excessive restrictions that can lead to poor model fit, Fong (2014) combined Bayesian multiple-group modeling with Bayesian CFA methods that relax measurement model restrictions to test cross-group invariance of Hospital Anxiety and Depression scores between male and female groups, finding that female anxiety levels were significantly higher than male levels.

De Bondt and Van Petegem (2015), building on this approach and following Asparouhov and Muthén' s (2014) recommendations, relaxed strict parameter invariance restrictions by providing prior distributions for parameter differences across groups to test gender invariance of the Over-Excitability Questionnaire. They found that multiple-group models rejected under traditional methods showed good fit after relaxing strict parameter restrictions, satisfying measurement invariance. We use this study as an example to detail the implementation steps for Bayesian multiple-group structural equation modeling.

De Bondt and Van Petegem (2015) used Bayesian multiple-group structural equation modeling to test gender invariance of the Over-Excitability Questionnaire. The study collected data from 516 college students who completed the online questionnaire, which includes five dimensions (psychomotor, sensual, intellectual, imaginational, and emotional) with 50 items.

The researchers first established measurement models for male and female groups separately, using both traditional frequentist and Bayesian estimation methods. In the Bayesian approach, restrictions on measurement error correlations and cross-loadings were relaxed: cross-loadings received normal prior distributions with mean 0 and variance 0.01, while observed indicator error covariance matrices received inverse Wishart prior distributions IW(I, 56). Results showed that corresponding models for both genders fit poorly under traditional methods (CFI values < 0.8), while measurement models for both genders fit well under Bayesian methods, with model PPp values of 0.905 and

0.767, respectively, and 95% intervals including zero, satisfying cross-group configural invariance.

Building on this foundation, the researchers established five multiple-group models to compare cross-group invariance across the five dimensions of the Over-Excitability Questionnaire between males and females. As a reference group, factor means and variances in the male group were fixed at 0 and 1, respectively, while factor means and variances in the female group and covariances between male and female groups received non-informative prior distributions—normal priors with mean 0 and extremely large variance—allowing free estimation to compare factor mean differences across groups. Furthermore, the researchers relaxed strict restrictions not only on measurement error correlations and cross-loadings but also on loading invariance and scalar invariance by providing normal prior distributions with mean 0 and variance 0.01 for group differences in factor loadings and intercepts. The corresponding Mplus code for the model can be found in the supplementary materials of De Bondt and Van Petegem (2015).

Multiple-group model results were as follows: the Over-Excitability Questionnaire met cross-group scalar invariance across all five dimensions, meaning item loadings and intercepts were equivalent across male and female groups, with good model fit (intellectual dimension: PPp = 0.540; imaginational dimension: PPp = 0.392; emotional dimension: PPp = 0.500; sensual dimension: PPp = 0.598; psychomotor dimension: PPp = 0.518). By comparing factor mean differences across groups, female factor means were significantly higher than male means in emotional and sensual dimensions, while male factor means were significantly higher than female means in the psychomotor dimension. No cross-gender differences in factor means were found in other dimensions.

3.4 Bayesian Multilevel Structural Equation Models

Cluster sampling is a commonly used sampling method in psychological research. It involves dividing a population into clusters (groups) based on certain criteria, randomly selecting some clusters, and using some individuals within those clusters as the sample. Data generated by this sampling method have a multilevel (nested) structure—for example, student-level data nested within classroom-level data when sampling students from different classrooms. The longitudinal data mentioned above can also be considered a type of multilevel data, where different measurement time points are nested within the individual level.

For such data structures, traditional analytical methods would ignore the nested structure information and violate independence assumptions, as variables within the same cluster may be correlated. Multilevel models can decompose variance across levels and explore relationships among variables at different levels. In latent variable modeling, Multilevel Structural Equation Modeling (MSEM) allows models to have different variances and covariances from between-group and within-group levels and can handle nested data with more than two levels

(Rabe-hesketh, Skrondal, & Pickles, 2004).

However, due to the complexity of multilevel data structures and modeling, traditional maximum likelihood estimation often encounters problems with parameter estimation, model convergence, and fit. Moreover, as the average sample size per level decreases, parameter estimation accuracy and model convergence rates decline (Hox, Maas, & Brinkhuis, 2010; Li & Beretvas, 2013; Meuleman & Billiet, 2009; Preacher, Zhang, & Zyphur, 2011). Additionally, the Intraclass Correlation Coefficient (ICC)—the ratio of between-group variability to total variability—affects parameter estimation quality. In MSEM analysis, when low ICC coincides with small average sample sizes per level, non-positive definite covariance matrices and incorrect parameter estimates can easily occur, such as negative residual variances (Depaoli & Clifton, 2015; Li & Beretvas, 2013).

Compared with maximum likelihood estimation, Bayesian estimation performs better in multilevel structural equation modeling: First, Bayesian methods provide more accurate parameter estimates (Asparouhov & Muthén, 2010; Baldwin & Fellingham, 2013). Second, providing certain prior information for parameters can prevent problems such as negative variance estimates or model non-convergence (Chung, Rabe-Hesketh, Dorie, Gelman, & Liu, 2013). Third, Bayesian methods more easily handle complex models, such as three-level models with ordered categorical variables (Asparouhov & Muthén, 2012). Finally, the most common problem in MSEM is small sample sizes at the second level, and Bayesian methods can still provide accurate parameter estimates in small-sample situations (Asparouhov & Muthén, 2010; Baldwin & Fellingham, 2013).

In empirical research, Johnson, Schoot, Delmar, and Crano (2015) used Bayesian multilevel structural equation modeling to explore the mutual influence and promotion between two interpersonal processes in group cooperation—discussion and argumentation. Tamminen, Gaudreau, Mcewen, and Crocker (2016) established Bayesian MSEM to test factors influencing athletes' enjoyment levels, avoiding problems with model non-convergence common in traditional methods, and found that athletes' personal emotion regulation strategies, team climate at the group level, and peer relationships were all related to enjoyment levels.

To facilitate researchers' application of this method, we use Prem, Scheel, Weigelt, Hoffmann, and Korunka' s (2018) study as an example to demonstrate how to apply this method to actual data analysis. Prem et al. (2018) used Bayesian multilevel structural equation modeling to explore the mediating mechanisms through which job characteristics influence workplace procrastination. The study collected data from 110 employees who completed questionnaires three times daily for 12 days.

The questionnaire included four variables: the independent variable (job characteristics) comprising three dimensions (time pressure, problem solving, and planning and decision-making); two mediating variables (cognitive appraisal processes and self-regulatory effort), where cognitive appraisal processes included challenge appraisal and hindrance appraisal dimensions; the dependent variable (workplace procrastination); and control variables (sleep quality and occupational self-efficacy). Within-person and between-person level hypotheses were consistent, as shown in Figure 3 [Figure 3: see original paper]: Hypothesis 1 proposed that job characteristics reduce workplace procrastination by increasing challenge appraisal, thereby reducing self-regulatory effort; Hypothesis 2 proposed that job characteristics increase workplace procrastination by increasing hindrance appraisal, thereby increasing self-regulatory effort.

Due to the nested data structure, the researchers used multilevel structural equation modeling for analysis, decomposing variable variance into between-person and within-person components. Compared with other multilevel mediation analyses, structural equation modeling produces less bias. Additionally, since indirect effects at the within-person level often exhibit skewed distributions, Bayesian estimation can better test indirect effects (Preacher, Zyphur, & Zhang, 2010).

The researchers used Mplus 8.0 for Bayesian multilevel structural equation modeling analysis, with results indicating good model fit (PPp = 0.460). Parameter estimation results showed that at the within-person level, all three dimensions of job characteristics had negative sequential indirect effects on workplace procrastination—that is, they affected workplace procrastination by increasing challenge appraisal and reducing self-regulatory effort (time pressure: indirect effect = -0.004, 95% CI [-0.009, -0.000]; problem solving: indirect effect = -0.011, 95% CI [-0.023, -0.002]; planning and decision-making: indirect effect = -0.004, 95% CI [-0.009, -0.000]). Additionally, time pressure had a positive sequential indirect effect on workplace procrastination—that is, it affected workplace procrastination by increasing hindrance appraisal and increasing self-regulatory effort (indirect effect = 0.006, 95% CI [0.002, 0.012]). At the between-person level, however, these sequential indirect effects did not hold. Therefore, this study revealed a close relationship between job characteristics and workplace procrastination, primarily realized through cognitive appraisal and self-regulatory effort at the within-person level.

Figure 3. Conceptual model diagram of the study

4. Model Evaluation and Fit Indices

Traditional model fit indices are not suitable for Bayesian estimation. Therefore, the following indices are needed for effective model assessment when using Bayesian methods.

4.1 Posterior Predictive p-value

Model-data fit can be assessed through posterior predictive checking (Gelman, Meng, Stern, & Rubin, 1996). Posterior predictive checking compares differences between actual data and data generated by the hypothesized model, enabling evaluation of model fit to actual data. Posterior predictive checking yields the

Posterior Predictive p-value (PPp-value). Unlike p-values in hypothesis testing, the PPp-value represents the proportion of iterations in MCMC algorithms where a test statistic based on the theoretical model exceeds the test statistic from the sample data. Therefore, a PPp-value around 0.5 (close to the random probability of 1/2) indicates good model fit. Posterior predictive checking also provides a 95% confidence interval for differences in test statistics between sample data and model-generated data. When the lower bound of this interval is negative and 0 falls near the center of the interval, the model fits well (Muthén & Asparouhov, 2012).

4.2 Bayes Factor

The Bayes Factor is an important statistic for model comparison and can be used to compare non-nested models. It compares the probabilities of supporting two models, M_0 and M_1, given a dataset. This comparison does not rely on hypothesis testing and does not tend to support the alternative hypothesis M_1 even with large sample sizes (Li, 2011). Kass and Raftery (1995) proposed criteria for interpreting Bayes Factors. For example, a Bayes Factor between 1 and 3 indicates similar support for both models from the data, in which case the principle of "parsimony" should be considered in model selection or other indices should be combined for judgment.

4.3 Bayesian Information Criterion

Like the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC; Schwarz, 1978) considers model complexity when evaluating model fit, penalizing complex models, with BIC imposing stronger penalties on complex models than AIC. Smaller BIC values indicate better model fit. However, BIC is a relative fit index and cannot evaluate the significance of differences in fit between two models.

4.4 Deviance Information Criterion

Like BIC, the Deviance Information Criterion (DIC) is used to compare competing models (Spiegelhalter, Best, Carlin, & Van der Linde, 2002). Smaller DIC values indicate better model fit. However, because DIC aligns more closely with the concept of Bayesian Deviance (Kaplan & Depaoli, 2012), it is more commonly applied. In contrast, BIC is more often used for model comparison in frequentist methods.

5. Software Introduction

Although Bayesian estimation of SEM offers many advantages, its application has lagged, partly because it appears to require researchers to have a strong foundation in Bayesian statistics (Muthén & Asparouhov, 2012). In reality, current software capable of analyzing SEM using Bayesian methods can not only meet most researchers' needs but is also easy to learn and use. The main commonly used software includes the following, with all except Mplus being free and open-source:

(1) **Mplus** (Muthén & Muthén, 1998–2017) is currently the most popular latent variable analysis software. Its programming language is simple and easy to learn. Due to its many default settings, researchers generally only need to change the estimation method from maximum likelihood to Bayesian estimation and provide corresponding prior information. If no prior information is provided, Mplus typically provides non-informative prior distributions by default. However, Mplus 8.0 cannot estimate some special models, such as latent moderation models (i.e., models where the moderator is a latent variable). When researchers set the estimation method to Bayesian, Mplus 8.0 can provide PPp-values, DIC, and BIC values as model fit and evaluation indices.

(2) **WinBUGS** (Windows version of Bayesian Inference Using Gibbs Sampling; Lunn et al., 2000) is a software package specifically designed for Bayesian statistical inference. Compared with Mplus, it requires researchers to have a higher level of Bayesian statistical knowledge. Some default estimation settings in Mplus need to be specified by researchers themselves in WinBUGS, but it is powerful and can flexibly estimate complex models. WinBUGS can provide DIC values as model fit and evaluation indices.

(3) **Stan** (Stan Development Team, 2014) can estimate complex BSEMs (such as latent moderation models, multilevel models, etc.) and can interface with the most popular data analysis languages (e.g., R, Matlab, Python).

(4) The **blavaan** package in R software (Merkle & Rosseel, 2015) cannot yet estimate some special models, such as latent moderation models or models with ordered categorical variables. However, users can export JAGS (Just Another Gibbs Sampler; Plummer, 2005) code to estimate complex and special models as needed. Additionally, Mplus software can be interfaced with blavaan through the mplus2lavaan() function. Blavaan can provide PPp-values and DIC values as model fit and evaluation indices.

Among the applied studies introduced in this paper, most used Mplus software for Bayesian structural equation modeling (e.g., Crenshaw et al., 2016; Falkenström et al., 2015; Golay et al., 2013; Zeman et al., 2017), while two studies used R and JAGS software for modeling (Praetorius et al., 2017; Winans-Mitrik et al., 2014). Researchers can choose appropriate analysis software according to their modeling needs.

6. Discussion

Bayesian methods have irreplaceable advantages in structural equation modeling, demonstrating better performance in model identification and fit, parameter estimation, handling complex models, and small-sample situations. The method has also developed rapidly in recent years (Van de Schoot et al., 2017). Based on Bayesian methods' characteristics of incorporating prior information and not relying on variance-covariance matrices, new modeling approaches have emerged, such as relaxing strict restrictions on traditional measurement models (Lu et al., 2016; Muthén & Asparouhov, 2012; Pan et al., 2017) and allowing small cross-group differences in multiple-group models (Muthén & Asparouhov, 2013). These modeling approaches are difficult to implement using traditional methods.

However, as of June 2018, searches in the China National Knowledge Infrastructure (CNKI) database revealed no applied studies using Bayesian structural equation modeling in domestic psychology journals. We hope this paper can provide domestic psychology researchers with new ideas for empirical research, enabling them to handle small samples and model non-identification situations under traditional methods with ease. Furthermore, the dominant position of frequentist methods has resulted in insufficient understanding of Bayesian methods among psychology researchers. Wang et al. (2017) also noted that academic understanding of Bayesian methods is quite limited, which may cause applied researchers to worry about mastering this new research tool. In reality, for researchers with structural equation modeling experience, using Mplus software for Bayesian estimation is very easy to master. We hope this introduction can dispel this misconception.

Although Bayesian methods offer many advantages, potential dangers include the latent influence of priors, misinterpretation of Bayesian characteristics and results, and incorrect reporting of Bayesian results. To address these potential dangers, Depaoli and van de Schoot (2015) proposed the WAMBS checklist to avoid misuse of Bayesian methods. Researchers can refer to this checklist to avoid errors in result reporting. As applied and methodological research using Bayesian methods to estimate structural equation models continues to increase (Van de Schoot et al., 2017), we believe that modeling methods, usage techniques, and reporting standards will become increasingly refined.

References

Hou, J., Wen, Z., & Cheng, Z. (2004). *Structural equation modeling and its applications*. Beijing: Educational Science Publishing House.

Li, X. (2011). *Structural equation modeling: Bayesian methods* (J. Cai, J. Pan, & Y. Zhou, Trans.). Beijing: Higher Education Press.

Liu, H., & Meng, Q. (2003). Longitudinal data analysis methods. *Advances in Psychological Science*, 11(5), 586–592.

Luo, S., & Jiang, Y. (2014). *Questionnaire survey research methods in management*. Chongqing University Press.

Wang, J., Wang, X., & Jiang, B. (2011). *Structural equation modeling: Methods* *and applications*. Beijing: Higher Education Press.

Wang, M. (2014). *Latent variable modeling and Mplus application: Basic volume*. Chongqing University Press.

Wang, M., Deng, Q., & Bi, X. (2017). Bayesian methods for latent variable modeling. *Advances in Psychological Science*, 25(10), 1682–1695.

Wen, Z., & Ye, B. (2014). Mediation effect analysis: Methods and model development. *Advances in Psychological Science*, 22(5), 731–745.

Ye, S., Tang, W., Zhang, M., & Cao, W. (2014). Analysis of missing data handling methods and their application status in longitudinal research. *Advances in Psychological Science*, 22(12), 1985–1994.

Alessandri, G., & De Pascalis, V. (2017). Double dissociation between the neural correlates of the general and specific factors of the life orientation test-revised. *Cognitive, Affective, & Behavioral Neuroscience*, 17(5), 917–931.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438.

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. (Technical Report). Retrieved June, 17, 2014.

Asparouhov, T., & Muthén, B. (2012, July). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. Retrieved from http://www.statmodel.com/download/NCME12.pdf.

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508.

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, 18(2), 151–164.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley-Interscience.

Chou, C. P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, Lagrange multiplier, and Wald tests. *Multivariate Behavioral Research*, 25(1), 115–136.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A non-degenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 78(4), 685–709.

Crenshaw, A. O., Christensen, A., Baucom, D. H., Epstein, N. B., & Baucom, B. R. W. (2016). Revised scoring and improved reliability for the communication patterns questionnaire. *Psychological Assessment*, 29(7), 913–924.

De Bondt, N., & Van Petegem, P. (2015). Psychometric evaluation of the over-excitability questionnaire-two applying Bayesian structural equation modeling

(BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, 6, 1963.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, 22(3), 327–351.

Depaoli, S., & van de Schoot, R. (2015). Improving transparency and replication in Bayesian statistics: The wambs-checklist. *Psychological Methods*, 22(2), 240–261.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 45–97.

Falkenström, F., Hatcher, R. L., & Holmqvist, R. (2015). Confirmatory factor analysis of the patient version of the working alliance inventory-short form revised. *Assessment*, 22(5), 581–593.

Fong, T. C., & Ho, R. T. (2014). Testing gender invariance of the hospital anxiety and depression scale using the classical approach and Bayesian approach. *Quality of Life Research*, 23(5), 1421–1426.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis*. London: Chapman and Hall/CRC.

Gelman, A., Meng, X. L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4), 733–760.

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment*, 25(2), 496–508.

Greenland, S. (2001). Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Analysis*, 21(4), 579–583.

Hox, J. J., Maas, C. J. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64, 157–170.

Hsu, H.-Y., Troncoso Skidmore, S., Li, Y., & Thompson, B. (2014). Forced zero cross-loading misspecifications in measurement component of structural equation models: Beware of even "small" misspecifications. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 10(4), 138–152.

Jacobson, N. C., Lord, K. A., & Newman, M. G. (2017). Perceived emotional social support in bereaved spouses mediates the relationship between anxiety and depression. *Journal of Affective Disorders*, 211, 83-91.

Johnson, A. R., Schoot, R. V. D., Delmar, F., & Crano, W. D. (2015). Social influence interpretation of interpersonal processes and team performance over time using Bayesian model selection. *Journal of Management*, 41(2), 574-606.

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions* (2nd ed.). Newbury Park, CA: Sage.

Kaplan, D. (2014). *Bayesian statistics for the social sciences.* New York, NY: Guilford.

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650-673).

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8(2), 115-128.

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 524-544.

Lee, S. Y., & Song, X. Y. (2012). *Basic and advanced Bayesian structural equation modeling*. John Wiley & Sons.

Li, X., & Beretvas, S. N. (2013). Sample size limits for estimating upper level mediation models using multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 241-264.

Lu, Z. H., Chow, S. M., & Loken, E. (2016). Bayesian factor analysis as a variable-selection problem: Alternative priors and consequences. *Multivariate Behavioral Research*, 51(4), 519-539.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325-337.

Maccallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111(3), 490-504.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99-128.

Maier, M. F., Bohlmann, N. L., & Palacios, N. A. (2016). Cross-language associations in the development of preschoolers' receptive and expressive vocabulary. *Early Childhood Research Quarterly*, 36, 49-63.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students'evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122.

Merkle, E. C., & Rosseel, Y. (2015). Blavaan: Bayesian structural equation models via parameter expansion. *Statistics*, 58(6), 129–138.

Meuleman, B., & Billiet, J. (2009). A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods*, 3(1), 45–58.

Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling: A Multidisciplinary Journal*, (2), 1–18.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335.

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis: Mplus Web Note 17. http://www.statmodel.com/examples/webnotes/webnote17.pdf.

Muthén, L. K., & Muthén, B. O. (1998–2017). *Mplus user' s guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén.

Nuijten, M. B., Wetzels, R., Matzke, D., Dolan, C. V., & Wagenmakers, E. J. (2015). A default Bayesian hypothesis test for mediation. *Behavior Research Methods*, 47(1), 85–97.

Pan, J., Ip, E. H., & Dubé, L. (2017). An alternative to post hoc model modification in confirmatory factor analysis: The Bayesian lasso. *Psychological Methods*, 22(4), 687–704.

Plummer, M. (2005). JAGS: Just Another Gibbs Sampler. *Astrophysics Source Code Library*.

Praetorius, A. K., Koch, T., Scheunpflug, A., Zeinz, H., & Dresel, M. (2017). Identifying determinants of teachers' judgment (in) accuracy regarding students' school-related motivations using a Bayesian cross-classified multi-level model. *Learning and Instruction*, 52, 148–160.

Prem, R., Scheel, T. E., Weigelt, O., Hoffmann, K., & Korunka, C. (2018). Procrastination in daily working life: A diary study on within-person processes that link work characteristics to workplace procrastination. *Frontiers in Psychology*, 9, 1087. doi: 10.3389/fpsyg.2018.01087.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209-233. doi: 10.1037/a0020141.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(2), 161-182.

Rabe-hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167-190.

Salarzadeh, H. J., Moghavvemi, S., Wan, C. M. R., Babashamsi, P., & Arashi, M. (2017). Testing students'e-learning via facebook through Bayesian structural equation modeling. *Plos One*, 12(9), e0182311.

Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education*, 88, 202-214.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.

Shuck, B., Zigarmi, D., & Owen, J. (2015). Psychological needs, engagement, and work intentions: A Bayesian multi-measurement mediation approach and implications for HRD. *European Journal of Training and Development*, 39(1), 2-21.

Sörbom, D. (1989). Model modification. *Psychometrika*, 54, 371-384.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64, 583-616.

Stan Development Team. (2014). *Stan modeling language: Users guide and reference manual*, Version 2.2.0.

Tamminen, K. A., Gaudreau, P., Mcewen, C. E., & Crocker, P. R. (2016). Interpersonal emotion regulation among adolescent athletes: A Bayesian multilevel model predicting sport enjoyment and commitment. *Journal of Sport and Exercise Psychology*, 38(6), 541-555.

Tofighi, D., & MacKinnon, D. P. (2011). R mediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, 43(3), 692-700.

Tofighi, D., & Mackinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2), 194-205.

Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217-239.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.

Wang, L., & Preacher, K. J. (2015). Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2), 249-263.

Winans-Mitrik, R. L., Hula, W. D., Dickey, M. W., Schumacher, J. G., Swoyer, B., & Doyle, P. J. (2014). Description of an intensive residential aphasia treatment program: Rationale, clinical processes, and outcomes. *American Journal of Speech-Language Pathology*, 23(2), 330-342.

Yuan, Y., & Mackinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, 14(4), 301-322.

Zeman, J. L., Dallaire, D. H., Folk, J. B., & Thrash, T. M. (2017). Maternal incarceration, children' s psychological adjustment, and the mediating role of emotion regulation. *Journal of Abnormal Child Psychology*, 46(2), 223-238.

Zhang, Z., Hamagami, F., Wang, L. L., Nesselroade, J. R., & Grimm, K. J. (2007). Bayesian analysis of longitudinal data using growth curve models. *International Journal of Behavioral Development*, 31(4), 374-383.

---

[1] Researchers treated significant measurement error correlations as nuisance parameters only.

Note: Figure translations are in progress. See original paper for figures.