Two Decades of Structural Equation Modeling in China: Methodological Research and Model Development in the 21st Century

Authors: Wang Yang, Wen Zhonglin, Li Wei, Fang Jie, Wen Zhonglin

Date: 2022-03-08T00:05:16+00:00

Abstract

During the first two decades of the 21st century, domestic methodological research on Structural Equation Modeling (SEM) primarily encompassed five themes: model development, parameter estimation, model evaluation, measurement invariance, and special data processing, with particularly substantial achievements in model development (i.e., various SEM variants). For each theme, the development and achievements of methodological research were systematically summarized based on a brief introduction of background knowledge. Finally, the progress of international methodological research on SEM and future research directions were also discussed.

Full Text

Methodological Research and Model Development of Structural Equation Models in China's Mainland in the 20 Years of the New Century

WANG Yang[1], **WEN Zhonglin**[2], **LI Wei**[3], **FANG Jie**[4]

[1] School of Public Administration, Guangdong University of Finance, Guangzhou 510521, China
[2] School of Psychology/Center for Studies of Psychological Application, South China Normal University, Guangzhou 510631, China
[3] School of Education Science and Technology, Northwest Minzu University, Lanzhou 730124, China
[4] Institute of Guangdong Economy and Social Development & Department of Applied Psychology, Guangdong University of Finance & Economics, Guangzhou 510320, China

Abstract

In the first two decades of the twenty-first century, methodological research on structural equation models (SEM) in China' s mainland has primarily involved five themes: model development, parameter estimation, model evaluation, measurement invariance, and special data processing, with particularly abundant achievements in model development (i.e., various SEM variations). For each theme, we systematically summarize the development and achievements of methodological research based on a brief overview of background knowledge. Finally, we also discuss recent progress in foreign methodological research on SEM and future research directions.

Keywords: structural equation model; model development; parameter estimation; model evaluation; measurement invariance

---

Structural equation model (SEM) represents a generalization of regression models and offers numerous advantages that regression models lack. SEM can simultaneously handle multiple independent and dependent variables, meeting the increasingly complex demands of theoretical models in social science research. It can analyze both observed and latent variables, aligning with the inherent nature of variables in social science research. SEM allows independent variables to contain measurement error, yielding more precise parameter estimates, and provides a rich array of fit indices for model evaluation. These advantages have established SEM as an important statistical method in social science research. The earliest domestic SEM methodological research in China can be traced to Zhang Jianping' s (1993) review article on SEM. The publication of the first domestic SEM monograph by Hou Jietai et al. (2004) greatly promoted the dissemination and application of SEM in China.

Since the beginning of the new century, domestic SEM methodological research in China has made substantial progress and produced abundant achievements. Using the full-text database of China National Knowledge Infrastructure (CNKI, https://www.cnki.net/) as the data source, with publication years set from 2001 to 2020, and keywords including structural equation, latent variable, hidden variable, structural model, measurement model, confirmatory factor analysis, confirmatory factor analysis, linear structural relationship, covariance structure, and covariance matrix, we screened and obtained 192 SEM papers published in journals. The distribution by discipline and publication year is shown in Table 1. Articles focusing on introductory SEM knowledge, current status of disciplinary applications, or primarily application-oriented were excluded. Received date: 2021-12-28. This research was supported by the National Natural Science Foundation of China (32171091), Guangdong Provincial Philosophy and Social Science Planning Project (Youth Project) (GD21YXL04), National Social Science Fund Projects (17BTJ035, 19BMZ080), Gansu Provincial Education Science Planning Project (GS[2021]GHB1777), Guangdong Provincial Regular Institutions of Higher Education Innovation Team Project (Humanities and Social Sciences) (2019WCXTD005), and Guangdong Provincial Education Science Planning Project (2020GXJK342). Corresponding author: Wen Zhonglin, E-mail: wenzl@scnu.edu.cn.

Table 1 Frequency distribution of SEM methodological papers published in domestic journals by discipline from 2001 to 2020 (sorted by number of publications)

2001-2005 | 2006-2010 | 2011-2015 | 2016-2020

Note: Other disciplines include systems science (7 papers), management (6), mathematics (4), physical education (3), informatics (3), education (3), sociology (2), biology (1), computer science (1), and linguistics (1). Comprehensive journals are primarily university journals. The figure includes 29 papers on reliability calculation based on SEM and mediation-moderation methods, which are reviewed in separate articles (see Fang et al., 2022; Wen, Fang, Xie et al., 2022; Wen, Fang, Chen et al., 2022) and are only counted here without appearing in the main text.

In terms of disciplines, 13 different fields have published SEM methodological research, with psychology publishing the most, followed by medicine and statistics. Regarding publication years, there were 18 papers from 2001-2005, 50 from 2006-2010, 68 from 2011-2015, and 56 from 2016-2020. SEM methodological research has generally shown an upward trend, with 2011-2015 being the peak period. This trend is consistent with the development of psychometric methodology research in China (Wen et al., 2021).

Articles were classified according to the SEM research themes they addressed. Themes discussed in 10 or more articles were grouped into categories; otherwise, they were classified as "other." Based on this, domestic SEM methodological research papers can be divided into five themes: model development (69 papers), parameter estimation (38 papers), model evaluation (17 papers), measurement invariance (15 papers), special data processing (10 papers), and other topics (43 papers). This classification differs slightly from that of Wen et al. (2021) because this article is not limited to papers published in psychology journals. This paper summarizes the progress of domestic SEM methodological research in the 20 years of the new century (2001-2020) and identifies frontier topics in this field by comparing them with recent foreign SEM methodological studies.

2. Development of Structural Equation Models

Structural equation models consist of measurement models and structural models. The measurement model reflects the relationship between latent variables and their indicators; when used alone, it becomes confirmatory factor analysis (CFA). The structural model reflects the influence relationships among (latent) variables. If the latent variables in the structural model are replaced by mean or total scores of measurement indicators for analysis, it becomes path analysis. In recent years, SEM has developed numerous new variations. In terms of measurement models, these mainly include bifactor models, exploratory structural equation modeling, specially designed measurement models (random intercept factor analysis model, fixed-links model, and Thurstone model), and formative measurement models. In terms of structural models, the primary development is the actor-partner interdependence model. For full models (i.e., complete SEM containing both measurement and structural models), the main development involves SEM with item parceling (i.e., item aggregation). Additionally, developments in SEM for group heterogeneity research and longitudinal research are also noteworthy.

2.1.1 Bifactor Model

In traditional CFA, a questionnaire item has non-zero loading on only one factor. The bifactor model allows questionnaire items to additionally load on a global factor based on a general multi-factor CFA model (Gu et al., 2014). The global factor can be a trait factor (which can be used to explore and verify higher-order factor structures, calculate CFA-based reliability, and analyze relationships between various factors and criterion variables) or a method factor (which can be used to test common method bias).

There is a nested relationship between bifactor models and higher-order factor models. Any higher-order factor model can be converted into a bifactor model, but only bifactor models satisfying proportional constraints (i.e., the ratio of global factor loadings to local factor loadings is constant within each dimension) can be converted into higher-order factor models (Gu et al., 2014). Simulation studies have found that for special bifactor models satisfying proportional constraints, their predictive accuracy for latent criterion variables is inferior to that of higher-order factor models (Xu et al., 2017). However, in more general cases that do not satisfy proportional constraints, using bifactor models for predictive validity analysis yields better model fit, statistical power, and accuracy of validity coefficient estimation than higher-order factor models (Wen et al., 2019).

2.1.2 Exploratory Structural Equation Modeling

Traditional CFA models are typical independent cluster models where cross-factor loadings of questionnaire items are fixed at zero, which may artificially inflate factor correlations and often prevents CFA from fitting factor structures obtained from exploratory factor analysis. Exploratory structural equation modeling (ESEM) effectively overcomes these limitations. ESEM allows cross-factor loadings to be non-zero based on CFA, more realistically reflecting factor structures while achieving better model fit. Mai and Wen (2013) introduced the principles of ESEM in detail, compared its similarities and differences with exploratory factor analysis and CFA, and provided recommendations for using ESEM. Notably, if Bayesian estimation with special priors is used for parameter estimation, not only can cross-factor loading restrictions be relaxed like in ESEM, but residual correlation restrictions can also be more flexibly relaxed (Muthén & Asparouhov, 2012).

2.1.3 Specially Designed Measurement Models

Random Intercept Factor Analysis Model. The random intercept factor analysis model adds a latent intercept factor to the general CFA model. The latent intercept factor varies between subjects but not between items (i.e., the latent intercept factor loadings are fixed as constants across all items), thereby reflecting certain stable traits of subjects (such as social desirability or acquiescence bias) and can be used to explain and control for item wording effects. Research has found that compared with bifactor models, the random intercept factor analysis model helps increase the proportion of trait variance in total questionnaire score variance, providing better structural validity for questionnaires under conditions where item wording effects exist, with trait variance exceeding method variance (Wei et al., 2016).

Fixed-Links Model. The fixed-links model is specifically designed for experimental research, with the main function of separating target factors from irrelevant factors in experiments to make measurement of experimental target concepts more accurate. Latent variables consist of one target factor and several non-target factors. Factor loadings are fixed, and cross-factor loadings are allowed. Target factor loadings are specified based on existing theoretical knowledge and experience, while non-target factor loadings are uniformly set as a constant. Model evaluation relies not only on fit indices but also on target factor variance. Significant target factor variance indicates that the psychological process represented by the latent variable is necessary for task completion (Ren et al., 2017).

Thurstone Model. Paired comparison tasks and ranking tasks are common in social science research. Examples of paired comparison tasks include choosing the preferred face from two presented face pictures, while ranking tasks include ordering three face pictures by preference. Data generated from such tasks are ordinal and do not satisfy the basic assumptions of classical test theory (Wang et al., 2014), making them better analyzed with specialized models such as the Thurstone model. In this model, latent variables represent options in comparison or ranking tasks, while measurement indicators represent subjects' preference choices for the option versus other options. If subjects prefer the current option, the factor loading is fixed at 1; otherwise, it is -1. For ranking tasks, measurement indicator residuals are set to zero. The advantage of this model is that it can simultaneously obtain detailed information about mean differences among task options and individual differences among subjects, detect subtle option differences, and avoid social desirability effects (Song & Liu, 2016).

2.1.4 Formative Measurement Model

The formative measurement model (FM) represents a special form of measurement model. The main differences between FM and traditional measurement models (also called reflective measurement models, RM) are (Jia & Bao, 2009): (1) In RM, factors influence measurement indicators, whereas in FM, the opposite occurs—factors are constructed from measurement indicators; (2) RM requires high internal consistency among measurement indicators that are interchangeable, while FM measurement indicators can be uncorrelated or even negatively correlated. For FM, what matters more is that measurement indicators cover all aspects of the construct, and RM reliability, validity, and fit evaluation indices are often not applicable to FM; (3) In RM, error terms exist at the measurement indicator level, while in FM, error terms exist at the latent variable level; (4) RM is suitable for confirmatory research, emphasizing the fit between theoretical models and actual data and the accuracy of parameter estimation, whereas FM is suitable for exploratory research, focusing more on the predictive effectiveness of measurement indicators on latent variables.

Although FM is not as popular as RM (Jia & Bao, 2009), it has received some attention in China, with studies introducing its basic principles and characteristics (Wang et al., 2013; Wang & Li, 2011; Ye & Li, 2014). Wang et al. (2011) simulation study showed that misspecifying FM as RM may cause bias in path coefficient estimation and increased Type I and Type II error rates, recommending model elaboration and model decomposition methods to avoid model misspecification.

2.2 Development of Structural Models

Social science research often focuses on dyadic data, such as data from both spouses, teachers and students, and superiors and subordinates on the same variable. Such data often lack independence (Li & Huang, 2010). To avoid potential inflation of Type I and Type II errors, specialized statistical methods are needed. The actor-partner interdependence model (APIM) is a structural model specifically designed to analyze relationships between paired variables. Taking the influence of empathy ability (independent variable) of teachers and students on their perceived social support (dependent variable) as an example, the APIM path specification includes four components: (1) the effect of student/teacher empathy on their own perceived social support, i.e., actor effects; (2) the effect of student/teacher empathy on the other' s perceived social support, i.e., partner effects; (3) the correlation between student and teacher empathy, which controls for the influence of the other independent variable when analyzing the effect of one independent variable; and (4) the correlation between residuals of dependent variables for students and teachers, which controls for other sources of dependence between dependent variables beyond independent variables (Liu & Wu, 2017). APIM can analyze the magnitude and direction of actor and partner effects and determine which is more dominant. For details on APIM principles (Li & Huang, 2010), analysis procedures (Liu & Wu, 2017), model variations (e.g., APIM with mediators and moderators; Chen et al., 2020; Liu & Wu, 2017), and software operations (e.g., Mplus and SPSS; Chen et al., 2020; He et al., 2018), please refer to relevant methodological literature.

2.3 Development of Full Models

Item parceling can be considered a special method for constructing full models. This approach aggregates original test items into item parcels, sacrificing the reliability of measurement model analysis but improving parameter estimation and model fit for structural models. Domestic research on item parceling involves two aspects. First is the introduction and demonstration of parceling techniques and strategies. For example, Bian et al. (2007) introduced the basic logic, advantages and disadvantages, and specific methods of item parceling in detail. Wu and Wen (2011) further refined parceling methods and provided operational procedures based on this foundation. For unidimensional scales, the former researchers recommended random parceling (i.e., parceling without following any pattern) from an economical and practical perspective (Bian et al., 2007); the latter researchers recommended the balanced method (i.e., first sorting items by factor loading magnitude, then distributing items to each parcel in an "S" shape) from the perspective of maximizing fit improvement (Wu & Wen, 2011). For multidimensional scales, both groups of researchers recommended the internal consistency method (i.e., aggregating all items within each dimension into one item parcel; Bian et al., 2007; Wu & Wen, 2011), as this method preserves the multidimensional structure of latent variables. Wang et al. (2014) recommended equal division parceling by dimension for multidimensional scales in full models and demonstrated through examples that this method can simplify models, improve the power of path coefficient tests, and achieve ideal fit.

The second aspect involves comparisons between item parceling and other model forms. For instance, Yang et al. (2010) used an application example to compare item parceling, path analysis, and unpacked full models, finding that item parceling yielded better model fit indices, while the unpacked full model produced higher R^2 values than item parceling and path analysis.

2.4 Structural Equation Models for Group Heterogeneity Research

Exploring unobservable group heterogeneity is of interest to many studies, which has led to numerous SEM-based heterogeneity analysis methods. Liu and Liu (2015) and Li et al. (2015) summarized such methods, mainly including latent class/profile models, factor mixture models, and multilevel latent class models.

2.4.1 Latent Class/Profile Models Latent class/profile models classify subjects based on latent trait scores and represent the application of cluster analysis ideas in SEM. If measurement indicators are categorical variables, it is a latent class model (LCM); if they are continuous variables, it is a latent profile model (LPM).

Domestic research on LCM/LPM involves three aspects. First is the introduction of basic principles and analysis procedures of LCM/LPM. For example, Zhang et al. (2010) introduced the statistical principles, analysis process, and application in psychological research of LCM. Guo et al. (2009), Meng et al. (2010), and Zeng et al. (2013) demonstrated the analysis process of LCM using simulation study data and application cases. Yin et al. (2020) introduced the basic principles, steps, and application in organizational behavior research of LPM.

Second is research on LCM/LPM with covariates or analysis of relationships between latent classes and covariates. For example, Wang and Bi (2018) summarized analysis methods for LCM with covariates (i.e., regression mixture models) and provided Mplus syntax templates. The academic community generally recommends using the LTB method (Lanza et al., 2013) for categorical outcome variables and the BCH method (Bolck et al., 2004) or robust three-step method for continuous outcome variables, with the latter also applicable when covariates are predictor variables (Wang & Bi, 2018). Zhang et al. (2017) introduced methods for subsequent analysis of LPM (i.e., analyzing relationships between latent classes and antecedent/consequent variables after classification) and pointed out through simulation studies that the inclusive classification analysis method (i.e., including variables needed for subsequent analysis as covariates during latent profile classification) provides better parameter estimation effects in subsequent analysis, combining accuracy and robustness. Further research found that when subsequent analysis variables include outcome variables and product terms of independent variables and latent class variables, subsequent parameter estimation effects are better. For more specific analysis procedures, see Zhang et al. (2019).

Third involves examining the classification effectiveness of LCM/LPM through simulation studies, including comparisons between LCM/LPM and other person-centered methods and the impact of different data and model conditions on classification results. For example, Ma et al. (2014) found through simulation comparisons that, except for special cases with only two latent classes and extremely unbalanced sample sizes across classes, LCM classification accuracy is comparable to that of K-means clustering and mixed Rasch models. Zhao et al. (2013) pointed out through simulation comparisons that LPM classification accuracy is higher than hierarchical clustering. Wang et al. (2017) examined the effects of number of classes, class separation, sample size, and number of measurement indicators on Entropy classification accuracy in LPM through simulation studies.

2.4.2 Factor Mixture Model The factor mixture model integrates CFA and LCM within a single model and can be viewed as an LCM using latent variables from CFA as classification indicators, or as a CFA model considering group heterogeneity. Chen et al. (2015) introduced the basic principles, main advantages, application directions, and analysis steps of factor mixture models. Li et al. (2020) compared the performance of latent class factor models (a variation of factor mixture models assuming group heterogeneity exists only in latent means) and LCM under different sample sizes and factor correlation conditions. Results showed that the former had better model fit and classification effects than the latter, with more parsimonious and easily identifiable models.

2.4.3 Multilevel Latent Class Model Multilevel LCM is specifically designed for multilevel structured data and can classify the same measurement indicators at both individual (Level 1) and organizational (Level 2) levels. Zhang et al. (2013) introduced the basic principles of multilevel latent class models, demonstrated analysis procedures using elementary school English ability tests as examples, and compared the effectiveness differences between multilevel LCM and general LCM.

2.5 Structural Equation Models for Longitudinal Research

Longitudinal research is a design that repeatedly measures the same research subjects and variables multiple times, processes and analyzes sequential data to understand variable development trends, relationships between variables, and individual differences. Domestic SEM methods in longitudinal research mainly involve models describing development trends and differences (latent growth model, piecewise growth model, latent class growth model, growth mixture model, piecewise growth mixture model, and latent transition model) and models exploring mutual influences between variables (cross-lagged model).

2.5.1 Models Describing Development Trends and Differences **Latent Growth Model.** The latent growth model (LGM) uses observed values of variables at different time points as measurement indicators, with an intercept factor reflecting subjects' baseline trait levels (factor loadings fixed at 1) and several slope factors reflecting linear or non-linear change trends of traits. It can simultaneously explore individual differences and development trends of psychological traits. Numerous methodological articles have introduced the basic principles, common variations, software operations (e.g., Mplus and SAS), and advantages and disadvantages of LGM (Li et al., 2012; Song & Wu, 2017; Su & Xu, 2017; Xu et al., 2007). Li et al. (2014) compared LGM and multilevel models from perspectives of mathematical form, prerequisite assumptions, data format, parameter estimation, and modeling flexibility and complexity, pointing out that multilevel modeling is simpler and more direct, while LGM is more flexible, can relax restrictions on equal measurement errors, freely estimate each measurement error, and provides better parameter estimation precision.

Piecewise Growth Model. LGM assumes that individual developmental trajectories are always continuous, ignoring possible stage characteristics and turning points in development (e.g., slow growth in early stages and fast growth in later stages). Piecewise growth models (PGM) allow growth curves to have different developmental stages. Liu et al. (2013) explored differences in parameter estimation effectiveness between PGMs defined under SEM and multilevel model frameworks and the consequences of incorrectly specifying PGM as non-stage models through simulation studies.

Latent Class Growth Model and Growth Mixture Model. LGM assumes that latent variables have the same developmental trajectory across individuals, ignoring possible heterogeneity. Latent class growth models (LCGM) and growth mixture models (GMM) combine LCM and LGM to classify individuals according to different trait development trends. The former assumes no individual differences within the same class, while the latter has no such restriction (Li et al., 2015), with the former being a special case of the latter (Wang et al., 2014; Xiao et al., 2020). For detailed introductions to LCGM, see Lü and Zhao (2018) or Wang et al. (2014); for GMM, see Liu (2007), Wang et al. (2014), Xiao et al. (2020), or Yu et al. (2018).

Piecewise Growth Mixture Model. The piecewise growth mixture model (PGMM) combines PGM and GMM, allowing developmental trajectories to be both stage-based and group-heterogeneous. Wang et al. (2017) described the basic principles, common model forms, parameter estimation methods and influencing factors, sample size requirements, fit evaluation indices, application status, and future research directions of PGMM in detail. Liu et al. (2014) used simulation studies to examine the effects of latent class separation and developmental model shape on model selection and parameter estimation of PGMM.

Latent Transition Model. The latent transition model (LTM) extends LCM to longitudinal research, not only exploring possible latent classes at each time point and allowing latent classes to change but also examining the probability of individuals transitioning from one class to others. Wang et al. (2015) and Huang (2019) introduced the statistical principles of LTM and demonstrated analysis procedures and result interpretation using adolescent impulsive behavior and English reading comprehension tests as examples. Huang (2018) introduced the theoretical foundation, transition mechanism, model characteristics, application status, and development prospects of LTM based on mixed item response theory.

2.5.2 Models Exploring Mutual Influences Between Variables The cross-lagged model analyzes mutual influences between multiple variables across time to explore causal relationships. The model focuses on several effects (Liu et al., 2022): (1) autoregressive effects, i.e., the influence of a variable' s previous measurement on its subsequent measurement, reflecting cross-time stability of variables (test-retest reliability); (2) cross-lagged effects, i.e., the effect of variable A' s previous measurement on variable B' s subsequent measurement while controlling for variable B' s previous measurement, and the effect of variable B' s previous measurement on variable A' s subsequent measurement while controlling for variable A' s previous measurement. The causal order of A and B is determined by whether the predictive effect of the prior causal variable on the subsequent outcome variable (represented by standardized path coefficients) is significantly stronger than the predictive effect of the prior outcome variable on the subsequent causal variable (Zhou et al., 2020). This method can better satisfy questionnaire research requirements for causal inference regarding temporal precedence and control of irrelevant variables (Wen, 2017).

3. Parameter Estimation Methods for Structural Equation Models

SEM parameter estimation is based on analysis of covariance structures. Let Σ( ) and S represent the covariance matrix derived from the theoretical model and the sample covariance matrix, respectively, with being the parameter vector. The fitting function F[S, Σ( )] represents the distance between Σ( ) and S. The parameter estimation process involves finding the estimate of that minimizes F[S, Σ( )]. Different methods for constructing the fitting function produce different parameter estimation methods. The most commonly used method is maximum likelihood estimation (ML). For non-normal data, robust maximum likelihood estimation (MLR) or weighted least squares with mean and variance adjusted (WLSMV) are typically used. The former provides more accurate estimation of factor correlations and parameter standard errors, while the latter provides more accurate estimation of factor loadings (Li, 2016) and is more suitable for data with fewer response categories.

Domestic research on SEM parameter estimation methods mainly involves two aspects: introduction of methods, primarily including partial least squares (PLS) and Bayesian methods; and comparison of parameter estimation methods.

3.1 Partial Least Squares Method

Traditional parameter estimation methods solve SEM by minimizing the fitting function, emphasizing parameter estimation precision. In contrast, PLS solves SEM by minimizing residual variance, emphasizing the predictive precision of predictor variables on outcome variables in equations. This characteristic aligns well with the main purpose of formative modeling (pursuing maximum explanatory power of measurement indicators on factors). Therefore, PLS is often used to analyze formative models, typically using specialized software such as SmartPLS, semPLS, or WarpPLS. Compared with traditional parameter estimation methods, PLS advantages mainly include: (1) better suitability for small samples and non-normal data; (2) suitability for complex models (i.e., models with high ratios of variables to sample size); (3) suitability for formative model analysis; (4) suitability for SEM research exploring the effects of multiple predictor variables (Luo, 2020).

Domestic introductions to PLS mainly involve two aspects. First is the introduction and evaluation of PLS. For example, Zhu and Liu (2005) and Ning et al. (2007) introduced the parameter estimation process of PLS in detail. Sun and Yang (2009) summarized and discussed three important issues in PLS-based SEM: how to select measurement models, how to use bootstrap methods to estimate and test parameters, and how to evaluate models. Liu et al. (2005) discussed the geometric meaning of the PLS algorithm. Ning and Liu (2007) examined the effectiveness of PLS in estimating SEM parameters through simulation studies, finding that this method underestimates structural coefficients and overestimates factor loadings. When sample size is large, results are basically credible.

Second is the extension of PLS in specific models and data. For example, Lin et al. (2006) introduced how to extend the PLS algorithm for two latent variables to multiple latent variables. Cheng and Yi (2016) introduced the basic principles and advantages of using PLS to estimate second-order factor models. Zhao (2011) discussed the advantages and disadvantages of different weight estimation algorithm patterns when using PLS to estimate second-order factor models. Wang et al. (2020) introduced quantile effect estimation methods for PLS path models (general SEM predicts the mean of dependent variables using independent variables, while quantile effects refer to predicting percentiles of dependent variables using independent variables). Meng and Wang (2009) and Li and Yue (2017) introduced how to apply PLS-based path analysis and PLS-based SEM to compositional data (i.e., data ranging between 0-1 with sum equal to 1). Ren and Wang (2010) introduced and recommended using fuzzy PLS for SEM modeling of data with uncertainty (e.g., middle options in Likert scales often indicate uncertainty). Application examples showed that this method provides more accurate parameter estimates and better model fit than PLS without fuzzy processing.

3.2 Bayesian Methods

Bayesian statistical analysis is a process that incorporates existing experience and knowledge about parameters to be estimated (i.e., prior information) into parameter estimation. The parameter estimation process when using Bayesian methods for SEM is as follows: (1) First, specify the theoretical model, which is no different from ordinary SEM. (2) Set prior distribution parameters (also called hyperparameters) for all unknown parameters. Parameters of research interest are generally target factor loadings and path coefficients, typically assumed to follow normal distributions requiring specification of mean and variance parameters, which can refer to existing research, especially meta-analyses (Yan & Mao et al., 2018). It should be emphasized that although many software programs can provide default prior distribution parameters (non-informative priors, equivalent to using only the Bayesian estimation framework without utilizing prior information), the core advantage of Bayesian methods is using prior information to aid parameter estimation. Research has found that using non-informative priors is not stronger than ML, and Bayesian estimation based on non-informative priors may even cause serious instability and biased estimation results with small samples (Smid & Winter, 2020). Therefore, when using Bayesian methods, informative priors should be used as much as possible. (3) Use Gibbs sampling of Markov chain Monte Carlo (MCMC) methods to obtain posterior distributions of parameters (Zhang, 2009). Whether estimation results have converged can be determined using potential scale reduction, trace plots, and effective sample size (for specific interpretation methods, see Wang, Deng, & Bi, 2017 and Smid & Winter, 2020). After convergence, central tendency measures of posterior distributions (e.g., mean or median) can be used as point estimates of model parameters. Meanwhile, with posterior distributions, Bayesian credible intervals can be obtained.

Compared with traditional frequency theory-based SEM parameter estimation methods, the main advantages of Bayesian methods are: (1) better estimation effects under small sample conditions, especially when effective prior information can be provided; (2) faster computation speed (Wang, Deng, & Bi, 2017); (3) when models are complex or the ratio of parameters to sample size is high, traditional parameter estimation often encounters convergence problems, while Bayesian methods often converge to appropriate solutions (Liang & Yang, 2016; Wang, Deng, & Bi, 2017); (4) Bayesian methods are less affected by nonnormality (Yan, Li, et al., 2018); (5) models that are not identifiable by traditional methods (e.g., allowing all cross-factor loadings and residual correlations to be freely estimated would exhaust degrees of freedom and make the model unidentifiable) may still be identifiable by incorporating prior information and using MCMC methods (Yan, Li, et al., 2018); (6) credible intervals obtained through Bayesian methods are more intuitively interpretable than traditional confidence intervals.

Wang, Deng, and Bi (2017) and Yan, Li, et al. (2018) introduced the basic concepts of Bayesian SEM (BSEM) and demonstrated analysis procedures and result interpretation with examples. Zhang et al. (2019) analyzed the application advantages and current status of Bayesian methods in different SEM variations (e.g., ordinary measurement models, latent mediation models, latent growth models, multi-group SEM, and multilevel SEM) and introduced model evaluation and available software for BSEM. Qin et al. (2020) introduced how to use SAS software to call OpenBUGS programs for more efficient implementation of BSEM.

3.3 Other Parameter Estimation Methods

In addition to PLS and Bayesian methods, Wu (2012) suggested using the Tikhonov regularization method to modify ML parameter estimation. Simulation studies showed this helps improve convergence rates and speed, reduce improper solutions, and decrease estimation bias. For dichotomous and ordinal data, some researchers suggested using polychoric correlation matrices to estimate model parameters (Wu & Zu, 2010) and provided LISREL example syntax (Wang et al., 2012; Wu et al., 2012). Parameter estimates based on polychoric correlation matrices have smaller bias than those based on Pearson correlation matrices (Zhou et al., 2013). Tong et al. (2009) introduced constrained least squares solutions for SEM and extended them to higher-order measurement models and multi-group SEM (Tong, Zou, et al., 2009). This method can improve parameter estimation convergence rates and speed and obtain unique solutions.

3.4 Comparison of Parameter Estimation Methods

Some SEM parameter estimation method comparison studies focus on the performance of weighted least squares (WLS), diagonal weighted least squares (DWLS), and generalized least squares (GLS) (Jiao, Wang, et al., 2015; Jiao, Wang, et al., 2015; Wu, 2010). However, these methods are not widely applied, and literature has long pointed out that their overall performance is not superior to ML (Hou et al., 2004).

More noteworthy are comparisons between currently popular estimation methods and new methods. For example, Liu, Luo, et al. (2012) simulated and compared the precision of MLR, WLSMV, and MCMC methods in estimating dichotomous data measurement models. Results showed that all three methods provided good estimation precision for factor loadings and other parameters, with MLR and WLSMV slightly outperforming MCMC, and WLSMV also having the advantage of faster computation. Tian and Fu (2004, 2005) simulated and compared the performance of ML and Bayesian methods in SEM parameter estimation, finding that Bayesian estimation precision was slightly better than ML but not substantially. Liang and Yang (2016) simulated and compared MLR and non-informative Bayesian methods in measurement models. Results showed: (1) For ability to identify misspecified models, Bayesian methods were stronger under non-normal conditions but weaker than MLR under normal conditions. (2) Bayesian methods had far stronger convergence ability than MLR, with this advantage being particularly pronounced in complex models (e.g., bifactor models). Yan and Mao et al. (2018) compared ML and Bayesian methods in small sample latent variable modeling through an application example. Although both produced similar parameter estimates for path coefficients and factor loadings, the former produced abnormal solutions such as negative variances.

Huo (2006), Li (2012), and Zhang (2007) compared PLS with traditional parameter estimation methods (e.g., ML, WLS, DWLS, and GLS) from theoretical aspects such as parameter estimation purpose, basic principles, and prerequisite conditions. Zhang (2015) conducted simulation comparisons between the two, finding that PLS parameter estimation had stronger stability but lower sensitivity to misspecified models. Liu and Chen (2007) proposed using the Generalized Maximum Entropy method to estimate SEM parameters and simulated and compared parameter estimation bias between Generalized Maximum Entropy and PLS, finding that the mean squared error of parameters obtained by Generalized Maximum Entropy was always smaller than that of PLS.

4. Structural Equation Model Evaluation

Interpretation of SEM parameters is based on the assumption that the hypothesized model fits the actual data well. Model fit is primarily evaluated through fit indices, involving methodological issues including proposal of new fit indices, critical values of fit indices, selection of fit indices, and model evaluation criteria beyond fit indices. Additionally, comparison and selection among multiple models are also important aspects of model evaluation.

4.1 Development of Fit Indices

Since SEM became popular, researchers have proposed over 40 fit indices. Wen et al. (2004) discussed properties that good fit indices should have (not systematically affected by sample size, penalizing complex models, and being sensitive to misspecified models). It is now generally accepted that CFI (comparative fit index), TLI (Tucker-Lewis index), RMSEA (root mean square error of approximation), and SRMR (standardized root mean square residual) are fit indices with good statistical properties (Wen & Liu, 2020).

In recent years, some studies have attempted to propose new fit indices. For example, Wang et al. (2018) proposed a corrected GFI (goodness-of-fit index, CGFI) to address the defects of GFI being systematically affected by sample size and not penalizing complex models. Another new fit index is the equivalence-testing-based fit index proposed by Yuan and Chan (2016; Wang et al., 2020). The basic idea is to address the logical problem of traditional ^{2} (taking the null hypothesis of perfect model fit as the hypothesis to be proved) by setting a new null hypothesis (model misfit greater than a tolerable small positive number) and alternative hypothesis (model misfit not greater than tolerable misfit). Based on this, fit indices RMSEAt and CFIt are proposed. Unlike traditional RMSEA and CFI, RMSEAt and CFIt have inferential statistical properties. Taking RMSEAt as an example, it represents that the size of model misfit does not exceed RMSEAt, and the probability of making this inference error does not exceed the significance level α.

Additionally, for the recently popular BSEM, there are specialized fit evaluation indices: (1) Posterior predictive p-value (ppp). It reflects the gap between the observed data fit function and the sample data fit function based on posterior distribution (Liang & Yang, 2016). Values approximating 0.5 indicate good model fit, while values close to 0 or 1 indicate poor fit. (2) Bayes factors. They can be roughly understood as the ratio of support strength for two competing models from current data. Empirically, a Bayes factor greater than 10 for two competing models indicates strong evidence supporting the model represented by the numerator; a Bayes factor less than 1/10 indicates strong evidence supporting the model represented by the denominator (Hu et al., 2018); a Bayes factor between 1/3 and 3 indicates similar support strength for both models (Zhang et al., 2019). (3) Lv. Compared with Bayes factors, this statistic has the advantages of smaller computational load and less dependence on prior information. Smaller Lv values indicate better model fit. Li et al.' s series of papers introduced the statistical principles of Lv in detail and applied it to model selection for SEM with ordinal variables and missing data and two-level SEM (Li & Wang, 2011, 2012; Li & Yang, 2014). (4) Bayesian information criterion (BIC) and deviance information criterion (DIC). Both indices are only used for model comparison, with smaller values indicating better model fit.

Additionally, when data do not follow a normal distribution, some studies propose using the Satorra-Bentler corrected ^{2} to evaluate model fit (Liu et al., 2013). Jin and Liang (2005) introduced fit evaluation indices suitable for PLS-based SEM, including factor communality, R^2 (predictive effect of exogenous latent variables on endogenous latent variables), and redundancy (average variance of endogenous latent variable measurement indicators explained by exogenous latent variables).

4.2 Critical Values of Fit Indices

For commonly used fit indices CFI, TLI, RMSEA, and SRMR, it is generally believed that when CFI and TLI are not less than 0.9 (Bentler & Bonett, 1980) and RMSEA and SRMR are not greater than 0.08 (Browne & Cudeck, 1992), the model is acceptable. However, some researchers recommend stricter criteria: CFI and TLI not less than 0.95 and RMSEA and SRMR not greater than 0.05 (Hu & Bentler, 1999). This stricter criterion is now commonly used as the standard for excellent model fit. Such critical standards are only empirically based conventional judgments.

Some researchers have designed special true models and misspecified models for simulation studies to determine optimal fit critical values under different conditions based on the sum of Type I and Type II error rates when different critical values are adopted for each fit index (Guo et al., 2007, 2008). However, the gap between true and misspecified models is complex and diverse, and optimal fit critical values change with this gap. Literature has pointed out that this practice of determining fixed critical values through simulation studies is inappropriate (Wen & Hou, 2008; Marsh et al., 2004).

4.3 Selection of Fit Indices

Given the numerous available fit indices, consideration must be given to which indices to report. Wang et al. (2010), through summarizing and analyzing existing fit index performance evaluation studies, concluded that TLI and RMSEA are the most trustworthy fit indices, although CFI, RNI, Mc, SRMR, and ^{2}/df also have some reference value. Some researchers, by analyzing fit index formulas, found that except for extreme cases where the theoretical model is not superior to the independence model (a model with only observed variables that are uncorrelated), CFI is always greater than or equal to TLI (Wen & Liang, 2015). Therefore, when TLI is acceptable, CFI is redundant. Moreover, CFI does not penalize complex models (Wen et al., 2004), so when several models all fit the data well, it cannot help researchers select a more parsimonious model. Thus, researchers can choose to report TLI, RMSEA, and SRMR.

4.4 Other Model Evaluation Criteria

Applied researchers often treat fit indices as the most important or even the only model evaluation criterion. Many articles have criticized this practice. On the one hand, each index only evaluates fit from a specific perspective and has inherent limitations (Wang et al., 2020). On the other hand, fit indices are generally affected by factors other than model fit degree, such as sample size, data distribution, factor loadings, and parameter estimation methods (Wang et al., 2020; Wen et al., 2008; Shi & Maydeu-Olivares, 2020). Therefore, it is necessary to refer to other criteria when evaluating model fit, such as parameter estimation evaluation, overall fit, internal fit, and cross-validation (Hou et al., 2004; Wen et al., 2004, 2008; Zheng & Wu, 2014).

First, examine whether the parameter estimation process converges normally. Identification or convergence problems often indicate unreasonable model specification. Then check four aspects: (1) Are model parameter signs appropriate and statistically significant? Nonsignificant parameters should be considered for modification (Hou et al., 2004). (2) Is R^2 sufficiently large? For measurement models, too small R^2 indicates low loadings, meaning low item reliability (Wen et al., 2008). Many studies have found that fit indices tend to support models with lower reliability (e.g., Greiff & Heene, 2017), so it is necessary to balance fit indices and reliability. (3) Are there abnormal elements in the residual matrix (Wen et al., 2008)? Too large absolute residuals indicate obvious gaps between theoretical models and data. (4) Which paths or loadings have large modification indices (Hou et al., 2004)? Too large modification indices mean that arrows in the model may point to wrong positions. Analyzing these parameter estimation-related evaluation indicators before examining fit indices can improve the power to detect misspecified models (Wen et al., 2008; Zheng & Wu, 2014).

After parameter estimation evaluation, fit indices are examined. Fit indices evaluate overall model fit. In addition, internal model fit can be examined, i.e., evaluating whether each latent variable is appropriately set from the model' s internal quality (Zheng & Wu, 2014). This mainly includes: (1) checking measurement tool reliability, where CFA can be used to calculate composite reliability; (2) checking measurement tool validity, including content validity, criterion-related validity, and construct validity.

Additionally, composite validity (i.e., cross-validation) can be examined by splitting data into two parts, using Sample 1 (calibration sample) to estimate parameters, then assigning these parameters to Sample 2 and examining its fit. Differences between results from the two samples can also be compared (Zheng & Wu, 2014). The ideal situation is that both samples show good and similar fit.

4.5 Strategies for Model Comparison and Selection

For the same dataset, more than one model may fit well, requiring comparison of multiple models' fit to select the optimal model. Liu et al. (2007) introduced concepts and characteristics of nested models and comparison and selection methods for nested and non-nested models. For nested models, they recommended comparing five models: independent null model Mn, saturated model Ma, the theoretically interesting model Mt, and two secondarily interesting theoretical models Mc and Mu. First, use Ma’ s ^{2} (minimum ^{2}) and Mn’ s degrees of freedom (maximum degrees of freedom) for ^{2} testing. If statistically significant, all models are unacceptable; if not significant, find the optimal model through ^{2} difference tests among models. For non-nested models, it is generally recommended to compare the expected cross-validation index (ECVI) and Akaike information criterion (AIC), with smaller values indicating better models. Luo and Zhang (2006) used creative thinking tests as examples to demonstrate how to compare and select optimal CFA models based on ^{2} difference tests and model parsimony. Additionally, Guo (2005)···

5. Multi-Group Measurement Models (Measurement Invariance)

Measurement invariance refers to SEM having the same structure and parameter values across different groups or time points, generally involving testing model form, factor loadings, intercepts, factor and error variance-covariance, and latent mean invariance (latent mean comparison involves mean structure models, which are special forms of SEM; Hou et al., 2004). Most domestic methodological literature on measurement invariance focuses on introducing various models to be tested, testing procedures, and model evaluation criteria (e.g., Bai & Chen, 2004; Liu, 2005; Liu & Wu, 2005; Liu & Yuan, 2015; Wei & Zheng, 2015; Wu & Zhang, 2011; Wu et al., 2009; Xu, 2010; Zhang et al., 2012; Zhao, 2007; Zheng et al., 2014). Additionally, two research directions on measurement invariance have received attention. The first is how to implement measurement invariance analysis in specific models or data. For example, Zheng et al. (2011) introduced measurement invariance analysis methods for second-order factor models. The main difference from traditional measurement invariance analysis is that all invariance constraints must be set separately at both first-order and second-order levels. Li and Liu (2011) recommended using WLSMV for parameter estimation and the corrected chi-square difference test (DIFFTEST) based on WLSMV to compare nested models when conducting measurement invariance analysis for ordinal data. Their simulation studies showed that WLSMV provided accurate estimation of factor loadings and threshold parameters, and DIFFTEST had acceptable Type I and Type II error rates (Li & Liu, 2011), with performance not inferior to item response theory-based methods (Liu, Li, et al., 2012).

The second aspect is that many constrained models in measurement invariance analysis are too strict and difficult to achieve (Wen et al., 2019). Some studies specifically discuss this issue. For example, when configural invariance and loading invariance hold but intercept invariance is not satisfied, if researchers still want to compare latent means, the projection method can be used (Wang et al., 2020; Deng & Yuan, 2016). This method decomposes each group’ s observed variable means into two orthogonal components: common scores (representing latent means) and specific factors. Cross-group invariance analysis of these two components does not depend on intercept terms, thus allowing latent mean comparison to bypass the traditional prerequisite of intercept invariance.

If only configural invariance holds but subsequent invariance constrained models fit poorly, the alignment method is recommended. This method constructs a loss function reflecting differences in intercepts and loadings across groups and obtains loading and intercept estimates that minimize the loss function. At this time, corresponding parameters across groups in the constrained model are not completely equal but sufficiently close, and this model has the same fit as the configural invariance model. The alignment method can be considered an approximate invariance model. For specific principles, application examples, and Mplus syntax, see Wen et al. (2019). Additionally, recent researchers have recommended using BSEM to analyze measurement invariance. By setting prior distributions with mean 0 and extremely small variance for parameters to be tested, cross-group invariance restrictions can be relaxed to achieve approximate invariance analysis (Song et al., 2021).

6. Special Data Processing in Structural Equation Models

Special data processing here mainly includes missing data, non-continuous data, non-normal data, and latent variable scores.

6.1 Missing Data Issues in SEM

When too much missing data exists in SEM modeling, model estimation may encounter problems (e.g., non-positive definite covariance matrices; Lin et al., 2010). Recommending good missing data imputation methods and comparing the effects of different methods are important research tasks. In terms of method recommendations, multiple imputation (MI) and full information maximum likelihood (FIML) are currently the most respected missing data processing methods (Wang & Deng, 2016). MI imputes missing data multiple times, analyzes each imputed complete dataset to obtain multiple estimates of target parameters, and finally summarizes multiple estimates to obtain final parameter estimates. FIML does not replace missing values but uses iterative estimation based on information from non-missing data. MI advantages include fully considering data uncertainty and more flexibly handling mixed data containing both continuous and non-continuous variables (Mansolf et al., 2020); disadvantages include more complex and time-consuming analysis processes (Ye et al., 2014). The main advantage of FIML is operational simplicity (Wang & Deng, 2016), while disadvantages include sometimes encountering inflated Type I error rates and model convergence problems (Mansolf et al., 2020).

In terms of method comparisons, Yang and Cao (2012) simulated and compared the effects of full Bayesian methods (treating missing data as unknown parameters and estimating model parameters and missing values by simulating joint posterior distributions of all variables and missing values) and partial Bayesian methods (equivalent to ignoring missing values in Bayesian estimation) in handling missing data in LGM. Results showed that when missing proportions exceeded 50%, the former had significantly smaller mean squared errors than the latter, i.e., higher parameter estimation precision; when missing proportions were small, both methods performed similarly. Chen and Liu (2015) simulated and compared ML and the Diggle-Kenward selection model in handling non-random missing data in LGM, with the latter generally performing better. Deng et al. (2018) used empirical data to compare listwise deletion, expectation maximization algorithm, MI, and FIML in handling missing data in SEM, finding that MI and FIML obtained better model fit, expectation maximization algorithm obtained the smallest parameter standard errors, and listwise deletion obtained the largest parameter standard errors. Wang and Deng (2016) used simulation studies to explore the role of auxiliary variables when using FIML to handle missing data in SEM, generally finding that including auxiliary variables helps obtain more reliable parameter estimates.

6.2 Non-Continuous and Non-Normal Data Issues in SEM

Questionnaire data used in social science research are often ordinal. Directly treating them as continuous variables in modeling may reduce parameter estimation precision and model fit. Gao (2012) proposed a continuous processing procedure for ordinal data and demonstrated through application examples that this method improves model fit.

Fang and Huang (2010) introduced common SEM modeling methods for non-normal data, recommending the use of the Bollen-Stine Bootstrap method to correct ^{2} test results and provided Amos operation demonstrations.

6.3 Latent Variable Score Issues in SEM

Zhang et al. (2005) introduced the estimation principles of latent variable scores in SEM and demonstrated latent variable score calculation using LISREL. Liu and Liu (2017) introduced weighted factor score calculation methods for global and local factors in bifactor models. Simulation studies showed that weighted factor scores are closest to true scores and have the highest reliability compared with other methods of synthesizing total test scores and dimension scores. Additionally, Zhang et al. (2012) introduced methods for determining latent variable data types.

7. Other Topics

In addition to the five major themes discussed above, domestic SEM methodological research includes other noteworthy topics, including how to incorporate traditional statistical and measurement methods into the SEM framework, discussions on error correlation issues, improvements to SEM modeling steps, SEM power analysis, etc.

Given SEM's many advantages and its compatibility with other statistical methods, some researchers have introduced methods combining traditional statistical and measurement methods with SEM to improve analysis accuracy and modeling flexibility. Examples include reliability calculation based on CFA models (see Wen, Fang, Chen, et al., 2022), various mediation models based on SEM (see Wen, Fang, Xie, et al., 2022) and moderation models (see Fang et al., 2022), multilevel models based on SEM (Bi, 2019; Fang et al., 2011; Zhang et al., 2006; Zhang et al., 2008), meta-analysis based on SEM (Gui et al., 2016; Qian et al., 2015), time series analysis based on SEM (Zhu, 2016), artificial neural network models based on SEM (Yan et al., 2019; Zhao & Wan, 2003), item factor analysis based on SEM (Wu & Tu, 2013), indicator systems based on SEM (i.e., hierarchical measurement indicator systems with weights established for abstract concepts; Jia, 2011; Si et al., 2014; Tian, 2007; Wang & Fu, 2004; Yu, 2020; Zhang & Wang, 2008), fuzzy comprehensive evaluation algorithms based on SEM (Zhuang & Liu, 2013), hidden Markov models based on SEM (Wang et al., 2018; Xia et al., 2016), and computer adaptive testing based on bifactor models (Liu et al., 2019; Mao et al., 2019; Mao et al., 2018).

Usually, measurement error correlations should not be specified in SEM without sufficient justification. When systematic correlation sources exist among error terms, error correlations can be specified. For example, it is reasonable to specify errors of items measured by the same method and errors of repeated measurements of the same item as correlated. Additionally, when the model has multiple large error correlation modification indices, error correlations can be explained by adding latent common method factors (Hu et al., 2018).

To address the problem that inappropriate model specification in SEM modeling processes may reduce model fit, Chen (2004) proposed improving SEM modeling steps, mainly adding steps to identify and eliminate inappropriate questionnaire items (items with low correlation with measured variables or excessive correlation with other items) and variables with excessively strong correlations, and re-evaluating measurement model and structural model fit.

An (2016) summarized common methods for power analysis in SEM and demonstrated how to conduct power analysis using Mplus with teacher-student relationship questionnaires as examples. Wang and Zhang (2007) introduced backward prediction algorithms for SEM, i.e., how to predict future relationships between variables based on current SEM, and supported the prediction effectiveness of this algorithm through simulation studies. Shan and Zhang (2020) derived formulas for calculating average causal effects when dependent variables and covariates (irrelevant variables that simultaneously affect independent and dependent variables) are latent variables while independent variables are observed variables. Jia and Liu (2008) compared similarities and differences between SEM and simultaneous equation models from aspects of variables, samples, data, parameter estimation, and model interpretation.

8. Discussion and Future Directions

In the 20 years of the new century, SEM has received increasing attention and application in social sciences, which has also driven the development of SEM methodological research. In China, more than ten different disciplines have contributed to SEM methodological research, producing abundant results across five themes. Within one year after 2020, just in the direction of SEM in longitudinal research, multiple new articles were added (Fang et al., in press; Gao et al., 2021; Liu, 2021; Wen & Zhu, 2021; Yuan et al., 2021; Zheng et al., 2021). Specific introductions can be found in Liu et al. (2022) and will not be repeated here. In other directions, such as measurement invariance, there are also new methodological developments (Song et al., 2021). To better understand current SEM research, we also introduce some newer foreign SEM methodological studies, which reveal directions worth exploring and expanding for domestic methodologists in the future.

8.1 Expansion of Original Themes

In terms of model development, we briefly introduce bifactor ESEM and SEM trees. Although bifactor models and ESEM each overcome some important limitations of traditional CFA, they also have shortcomings. The former ignores the fact that cross-factor loadings are widespread, which may overestimate factor correlations; the latter ignores the possibility of higher-order factors, which can lead to overestimated cross-factor loadings. Bifactor ESEM, which combines bifactor models and ESEM, well compensates for the limitations of using the two models independently (Morin et al., 2016).

SEM trees combine SEM and decision trees, allowing classification of a certain outcome variable or its change trajectory (e.g., children' s reading ability development trajectory) based on predictor variables selected by researchers (e.g., children' s motor skills, learning styles, and life knowledge). When there are many valuable predictor variables, predictor variables generally have interaction effects, or there is insufficient prior knowledge about the number of classifications, SEM trees are considered a good alternative to finite mixture models (i.e., LCM, LPM, LCGM, GMM, and PGMM) (Jacobucci et al., 2017).

In terms of parameter estimation methods, we briefly introduce research progress on parameter estimation accuracy when models are misspecified and evaluation of parameter estimation uncertainty. Lai and Zhang (2017) found through simulation comparison studies that for CFA models, ML can provide reliable parameter point estimates even when models are seriously misspecified; but for full SEM, when misspecification is large, various parameter estimation methods have relatively large biases. The fungible parameter estimates proposed by Pek and Wu (2018) evaluate the sensitivity of SEM parameter estimation to sampling variation unrelated to the parameters.

In terms of model evaluation, we briefly introduce some new fit indices (including local fit tests and BSEM fit indices) and new developments in critical values.

Traditional fit indices are only used to evaluate overall model fit. The local fit test proposed by Thoemmes et al. (2018) provides separate fit evaluations for different parts of the model, helping to locate sources of model misspecification and can still be used even when models are not identified or do not converge.

With the popularity of BSEM, how to evaluate its goodness of fit has become an important issue. Fit evaluation indices used in BSEM are not familiar traditional fit indices to researchers, and most are only used for model comparison (e.g., Bayes factors and deviance information criterion). Garnier-Villarreal and Jorgensen (2020) constructed seven new fit indices based on BSEM, such as RMSEA, CFI, and TLI, by replacing ^{2} with parameter posterior means.

Regarding critical value determination, McNeish and Wolf (2021) proposed a dynamic fit index critical value based on data simulation technology. This critical value considers various model and data characteristic factors affecting model fit and can effectively reject misspecified models.

In terms of measurement invariance analysis, by defining new differential item functioning (DIF) or viewing invariance issues as clustering or moderation problems of model parameters, cross-group measurement invariance can be interpreted from new different perspectives (Bauer, 2017; De Roover et al., 2020; Schulze & Pohl, 2021).

Additionally, there are expansions of SEM for special data, involving non-continuous data SEM modeling methods such as logistic latent growth models and nominal variable factor analysis (Asparouhov & Muthén, 2021), and research related to small sample data modeling (Jiang & Yuan, 2017; Smid & Winter, 2020).

8.2 Opening New Topics

New topics in the SEM field continue to emerge, such as SEM for exploratory purposes, SEM for experimental research, use of instrumental variables in SEM, and development of SEM software packages.

The vast majority of SEM applications are confirmatory in nature. However, for large-scale studies with massive data, it may be necessary to explore which of multiple antecedent variables have practical effects without prior hypotheses. This exploratory perspective on SEM can be implemented using regularization methods. Regularized SEM adds a penalty term to the fitting function or sets special priors for parameters (e.g., small-variance cross-factor loading priors) to shrink small coefficients to zero, thereby serving as variable or path screening (Jacobucci et al., 2016; Lu et al., 2016; Muthén & Asparouhov, 2012; Pan et al., 2017).

In recent years, many foreign methodological articles have combined traditional statistical methods in experimental research with latent variable modeling ideas, forming latent variable modeling methods based on experimental research. For example, Breitsohl (2019) compared between-subjects ANOVA with two SEM-based methods: structured-means-modeling and multiple-indicator multiple-cause models. Both SEM methods represent dependent variables as latent variables, with the difference being that structured-means-modeling simultaneously models dependent variables corresponding to each experimental treatment and compares latent means, while multiple-indicator multiple-cause models directly establish regression of latent dependent variables on manipulation variables. Another example is the latent repeated measures analysis of variance proposed by Langenberg et al. (2020), which replaces single-indicator observed outcome variables with multi-indicator latent variables, improving the power of main effects and interaction effects, relaxing assumptions about missing data and residual structures, and verifying whether strong invariance is satisfied through measurement invariance analysis.

Instrumental variables are variables that researchers are not interested in but can explain endogeneity of predictor variables (i.e., predictor variables correlate with model residuals) and are unrelated to model residuals. By using instrumental variables, predictor variables are decomposed into exogenous parts unrelated to residuals and endogenous parts correlated with residuals, and only the exogenous part is used to estimate path coefficients of interest. This can solve endogeneity problems, obtain more accurate estimates of model coefficients, and thereby enhance SEM' s causal inference ability (Maydeu-Olivares et al., 2020).

Although there are many specialized SEM software programs (e.g., Amos, EQS, LISREL, and Mplus), traditional software also has some limitations, and many newly proposed frontier methods cannot be quickly incorporated into traditional software. At this time, software packages developed by researchers themselves play an important role in improving traditional software limitations and promoting new method applications. Many methodological articles are dedicated to introducing newly developed SEM software packages. For example, Gonzales (2021), Rosseel (2012), and Igolkina and Meshcheryakov (2020) introduced comprehensive SEM software packages JMP Pro, lavaan, and semopy, respectively. Jiang et al. (2017) introduced the equivalence testing software package equaltestMI, and Zhang et al. (2021) introduced the CFA model modification software package blcfa.

8.3 Conclusion

Regarding frontier SEM methods, most work in domestic journals belongs to tracking, introduction, or commentary and integration, lacking in-depth research on the statistical properties of these methods. On the one hand, understanding of connections between methods is insufficient. Few studies attempt to clarify mathematical relationships between competing methods (e.g., Fang et al., in press; Wen & Liang, 2015), such as which methods are approximate or even equivalent, or how different methods can be transformed through certain changes. Such analyses are not uncommon abroad (Serang et al., 2019; Usami et al., 2015, 2019; Yuan & Deng, 2021). Such analyses help deepen researchers' comprehensive understanding and mastery of methods. On the other hand, there are also few simulation studies comparing methods. For example, foreign simulation studies on fit indices have never stopped, including studies exploring the impact of various non-fit factors (e.g., parameter estimation methods and reliability) on fit index estimation (McNeish et al., 2018; Shi & Maydeu-Olivares, 2020) and studies comparing new fit indices with popular fit indices (Counsell et al., 2020; Garnier-Villarreal & Jorgensen, 2020).

Despite these shortcomings, domestic SEM methodological achievements in the past 20 years have still provided strong support for improving the level of domestic quantitative research. It is believed that with the mutual promotion between methodological research and application demands, there will be more high-quality SEM methodological literature in the future.

References

An, M. (2016). Monte Carlo method for sample size estimation in structural equation models. *Injury Medicine (Electronic Edition)*, *5*(4), 45–49.

Bai, X., & Chen, Y. (2004). The concept of measurement equivalence and its determination conditions. *Advances in Psychological Science*, *12*(2), 231–239.

Bi, X. (2019). Measurement of urban community social capital based on multilevel confirmatory factor analysis—Case study and related methodological review. *Sociological Studies*, *34*(6), 213–237.

Bian, R., Che, H., & Yang, H. (2007). Application of item parceling in structural equation models. *Advances in Psychological Science*, *15*(3), 567–576.

Chen, M. (2004). Improvement of structural equation modeling methods and application in CRM empirical research. *Science Research Management*, *25*(2), 70–75.

Chen, N., & Liu, H. (2015). Handling non-random missing data in growth models based on growth models: Selection models and maximum likelihood methods. *Psychological Science*, *38*(2), 463–471.

Chen, S., Yang, Q., Qiu, J., Fan, X., He, J., Fan, X., & Hao, C. (2020). Application and SPSS implementation of actor-partner interdependence moderation models for paired data in public health. *Modern Preventive Medicine*, *47*(20), 11–17.

Chen, Y., Wen, Z., & Gu, H. (2015). Factor mixture model: Integration of latent class analysis and factor analysis. *Advances in Psychological Science*, *23*(3), 415–423.

Cheng, H., & Yi, D. (2016). Research on partial least squares-second-order factor model in comprehensive variable construction problems. *Modern Management Science*, *(2)*, 18–21.

Deng, J., Chen, Y., & Guan, Y. (2018). Comparison of missing data imputation methods based on structural equation models. *Journal of Mathematical Medicine*, *31*(2), 159-161.

Fang, J., Qiu, H., & Zhang, M. (2011). Contextual effect analysis based on multilevel structural equation models—Comparison with multilevel linear models. *Advances in Psychological Science*, *19*(2), 284-292.

Fang, J., Wen, Z., Ouyang, J., & Cai, B. (2022). Methodological research on moderation effects in China. *Advances in Psychological Science*, *XX*(X), ???-???.

Fang, J., Wen, Z., & Huang, G. (in press). Exploring longitudinal relationships: Tracking models based on cross-lagged structures. *Psychological Science*.

Fang, M., & Huang, Z. (2010). Handling non-normal data in structural equation models. *Chinese Journal of Health Statistics*, *27*(1), 84-87.

Fu, J., & Tian, X. (2004). Discussion on maximum likelihood method and Bayes method in structural equation model analysis (English). *Journal of Applied Statistics and Management*, *23*(6), 53-58.

Gao, J. (2012). Handling methods for ordinal data in structural equation models. *Statistics and Decision*, *(18)*, 19-21.

Gao, W., Li, W., Lin, W., Weng, Q., Wang, Y., & Yang, S. (2021). Application of latent change score models in longitudinal research in organizational behavior. *China Human Resources Development*, *38*(11), 6-25.

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, *33*(5), 313-317.

Gu, H., Wen, Z., & Fang, J. (2014). Bifactor model: A new perspective for multidimensional construct measurement. *Psychological Science*, *37*(4), 973-979.

Gui, Y., Zhang, C., Xu, C., Peng, L., Zuo, H., Yang, Y., Niu, Y. (2016). Application of structural equation model in meta-analysis—Fixed effects model. *Chinese Journal of Evidence-Based Medicine*, *16*(2), 229-234.

Guo, Q., Li, F., Chen, X., Wang, W., & Meng, Q. (2008). Performance of fit indices under different conditions and selection of critical values. *Acta Psychologica Sinica*, *40*(1), 109-118.

Guo, Q., Wang, W., Chen, X., & Han, D. (2007). Model fit judgment in confirmatory factor analysis. *Psychological Exploration*, *27*(4), 83-87.

Guo, X., Pei, L., & Zhang, Y. (2009). Latent class model and data simulation analysis. *Journal of Mathematical Medicine*, *22*(6), 631-635.

Guo, Y. (2005). Calculation of Bayes factors for general nonlinear structural equation models. *Journal of Suzhou University (Natural Science Edition)*, *21*(4), 24–28.

He, J., Fan, X., & Hao, C. (2018). Implementation of actor-partner interdependence models for paired data in MPLUS. *Modern Preventive Medicine*, *45*(3), 528–531.

Hou, J., Wen, Z., & Cheng, Z. (2004). *Structural Equation Model and Its Application*. Educational Science Publishing House.

Hu, C., Kong, X., & Peng, K. (2018). Bayes factors and their implementation in JASP. *Advances in Psychological Science*, *26*(6), 951–965.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Hu, P., Lu, H., & Ma, Z. (2018). Feasibility and conditions of allowing error correlations in confirmatory factor analysis. *Statistics and Decision*, *(19)*, 37–41.

Huang, M. (2018). Research progress on latent transition models based on mixed IRT. *Examination Research*, *(66)*, 102–110.

Huang, M. (2019). Application of latent transition model in educational testing —Taking English reading comprehension test as an example. *Journal of Puyang Vocational and Technical College*, *32*(4), 91–95.

Huo, Y. (2006). Analysis and comparison of LISREL and PLS path modeling principles. *Statistics and Decision*, *(20)*, 19–21.

Jia, X. (2011). Comparability issues in structural equation model evaluation systems. *Journal of Applied Statistics and Management*, *30*(2), 246–253.

Jia, X., & Liu, L. (2008). Comparison between structural equation model and simultaneous equation model. *Journal of Applied Statistics and Management*, *27*(3), 439–446.

Jia, Y., & Bao, G. (2009). Frontier analysis of formative measurement models in structural equation models. *Foreign Economics and Management*, *31*(5), 52–59.

Jiao, X., Wang, C., Wang, D., & Liu, Y. (2015). Comparison of GLS and WLS for measurement model structural equation models. *Chinese Journal of Public Health*, *31*(1), 7–9.

Jiao, X., Wang, D., Wang, C., & Liu, Y. (2015). Performance comparison of GLS and WLS for structural equation models. *Chinese Journal of Public Health*, *31*(9), 1195–1198.

Jin, Y., & Liang, Y. (2005). Fit indices of partial least squares (PLS) method and its application in satisfaction research. *Journal of Applied Statistics and Management*, *24*(2), 40–44.

Li, C., & Liu, H. (2011). Simulation study on measurement invariance testing for ordinal data and influencing factors. *Psychological Science*, *34*(6), 1482-1487.

Li, G., Zhao, L., Deng, K., Zhang, Y., Gong, X., & Gao, Y. (2020). Comparison and application of latent class and latent class factor analysis in heterogeneous group classification. *Journal of Guangdong Pharmaceutical University*, *36*(1), 112-117.

Li, L., Gao, Y., Zhang, M., & Zhang, Y. (2012). Latent growth curve model and its application. *Chinese Journal of Health Statistics*, *34*(5), 713-716.

Li, L., Zhao, L., Zhou, S., Zhang, M., Gao, Y., & Zhang, Y. (2015). Latent variable analysis methods for group heterogeneity research. *Chinese Journal of Health Statistics*, *32*(4), 711-715.

Li, L., Zhou, S., Zhang, M., Zhang, Y., & Gao, Y. (2014). Application and comparison of multilevel models and latent growth curve models in longitudinal data analysis. *Chinese Journal of Epidemiology*, *35*(6), 741-744.

Li, S., & Yue, L. (2017). Multivariate structural equation model based on compositional data. *Journal of Henan Polytechnic University (Natural Science Edition)*, *36*(3), 407-411.

Li, X. (2012). Analysis and comparison of LISREL and PLS modeling methods. *Science and Technology Management Research*, *(20)*, 230-233.

Li, Y., & Huang, F. (2010). Actor-partner interdependence model (APIM) for dyadic data analysis. *Advances in Psychological Science*, *18*(8), 1321-1328.

Li, Y., & Wang, X. (2011). Model selection for structural equation models with ordinal variables. *Statistics and Decision*, *(14)*, 15-18.

Li, Y., & Wang, X. (2012). Model selection problem for structural equation models with missing data. *Journal of Applied Statistics and Management*, *31*(6), 1021-1027.

Li, Y., & Yang, A. (2014). Bayesian model selection for two-level structural equation models. *Statistics and Decision*, *(16)*, 4-9.

Liang, X., & Yang, Y. (2016). Confirmatory factor analysis when structural and distributional assumptions are violated: Comparison of robust maximum likelihood estimation and Bayesian estimation (English). *Psychological Science*, *39*(5), 1256-1267.

Lin, X., Kong, D., Fu, H., & Ding, Y. (2010). Research progress on structural equation models. *China Medical Innovation*, *7*(5), 174-176.

Lin, S., Wu, B., Ning, L., & He, T. (2006). An iterative method for PLS algorithm based on multiple latent variables. *Journal of Systems Engineering*, *21*(4), 446-450.

Liu, C., & Wu, X. (2017). Paired patterns of actor-partner interdependence and their testing. *Psychological Development and Education*, *33*(1), 105–112.

Liu, H., & Liu, Y. (2015). Expansion of latent variable scales and research prospects. *Statistics and Decision*, *(6)*, 8–12.

Liu, H., Li, C., Zhang, P., & Luo, F. (2012). Testing method for measurement equivalence of categorical data and comparison: Testing group differences in item threshold (difficulty) parameters. *Acta Psychologica Sinica*, *44*(8), 1124–1136.

Liu, H., Luo, F., Wang, Y., & Zhang, Y. (2012). Estimation of multidimensional test item parameters: Comparison of SEM and MIRT methods. *Acta Psychologica Sinica*, *44*(1), 121–132.

Liu, H. (2007). How to describe differences in development trends: Latent growth mixture model. *Advances in Psychological Science*, *15*(3), 539–544.

Liu, J., Xu, Y., & Liu, H. (2013). Application of non-normal confirmatory factor analysis in gene overall effect. *Bioinformatics*, *11*(3), 192–195.

Liu, J., Peng, S., & Tu, D. (2019). Optimization design and effect verification of CAT testing under bifactor models. *Journal of Jiangxi Normal University (Natural Science Edition)*, *43*(2), 128–134.

Liu, J., & Chen, Y. (2007). GME method for structural equation model parameter estimation. *Journal of National University of Defense Technology*, *29*(1), 116–121.

Liu, Y., & Yuan, C. (2015). Overview of measurement invariance research and theoretical framework. *Chinese Journal of Nursing*, *50*(1), 110–116.

Liu, Y., & Wu, W. (2005). Psychological measurement balance research and examples. *Psychological Science*, *28*(1), 170–174.

Liu, Y. (2005). Analysis of measurement balance issues in comparative research. *Journal of Applied Statistics and Management*, *24*(3), 25–31.

Liu, X., Ma, R., Luo, Y., Li, Z., Zhang, C., & Zhang, Y. (2013). Application of non-normal confirmatory factor analysis in gene overall effect. *Bioinformatics*, *11*(3), 192–195.

Liu, X., & Liu, H. (2017). Synthesis methods for total test scores and dimension scores based on bifactor models. *Acta Psychologica Sinica*, *49*(9), 1234–1246.

Liu, Y., Du, H., Fang, J., & Wen, Z. (2022). Research on longitudinal data analysis methods and model development in China. *Advances in Psychological Science*, *XX*(X), ???–???.

Liu, Y., Luo, F., & Liu, H. (2014). Influencing factors of piecewise growth mixture models: Distance and shape. *Acta Psychologica Sinica*, *46*(9), 1400–1412.

Liu, Y., Zhao, Q., & Liu, H. (2013). Method comparison for piecewise growth models. *Psychological Exploration*, *33*(5), 415-422.

Luo, F., & Zhang, H. (2006). Experimental research using confirmatory factor analysis to test multidimensionality of tests. *Statistical Research*, *23*(4), 76-79.

Luo, L. (2020). Application of PLS-SEM multivariate statistical analysis in event audience research. *Journal of Shanghai University of Sport*, *44*(11), 86-94.

Lü, Y., & Zhao, R. (2018). Group-based trajectory model—A new method for intervention research. *Psychological Exploration*, *38*(1), 91-96.

Ma, W., Bian, Y., Guo, W., & Xie, M. (2014). Comparison of K-means, latent class model, and mixed Rasch model. *Psychological Exploration*, *34*(5), 454-458.

Mai, Y., & Wen, Z. (2013). Exploratory structural equation modeling (ESEM): Integration of EFA and CFA. *Advances in Psychological Science*, *21*(5), 934-939.

Mao, X., Liu, H., & Tang, Q. (2019). Comparison of polytomous item selection strategies in bifactor model MCAT. *Psychological Science*, *42*(1), 187-193.

Mao, X., Xia, M., & Xin, T. (2018). Full-information item bifactor analysis: Model, parameter estimation, and application. *Advances in Psychological Science*, *26*(2), 325-336.

Meng, C., Wu, J., Li, Y., Zhou, Y., Li, N., Zhang, Y., & Zhao, R. (2010). Principle of latent class analysis and its application in cluster analysis. *Chinese Journal of Health Statistics*, *27*(3), 237-239.

Meng, J., & Wang, H. (2009). Log-contrast partial least squares path analysis model for multivariate compositional data. *Journal of Applied Statistics and Management*, *28*(3), 436-442.

Ning, L., & Liu, J. (2007). Simulation data analysis of PLS algorithm for two latent variables. *Statistics and Decision*, *(16)*, 159-160.

Ning, L., Liu, J., Wu, B., & He, T. (2007). Research on iterative algorithms for structural equation models. *Journal of Systems Engineering*, *22*(1), 84-87.

Qian, L., Chen, B., Yao, N., Yang, J., & Li, J. (2015). Application of structural equation model-based meta-analysis in the relationship between depression and social support. *Chinese Journal of Health Statistics*, *32*(1), 63-65.

Qin, Z., Yan, X., Shen, Y., Xiao, J., He, S., & Ren, W. (2020). Implementation of OpenBUGS for Bayesian analysis of structural equation models in SAS macro programs. *Chinese Journal of Health Statistics*, *37*(3), 475-480.

Ren, H., & Wang, S. (2010). Algorithm solution for fuzzy PLS-structural equation model with two latent variables. *Statistics and Decision*, *(7)*, 47-49.

Ren, X., Wang, T., & Schweizer, K. (2017). Fixed-links model and its application in cognitive psychology research. *Advances in Psychological Science*, *25*(10), 1675–1681.

Shan, N., & Zhang, X. (2020). Causal inference with measurement error covariates and unobservable response variables. *Journal of Jilin Normal University (Natural Science Edition)*, *41*(1), 62–66.

Si, J., Xiao, H., & Jiang, Y. (2014). Problems and countermeasures in the application of structural equation models in comprehensive evaluation. *Modern Management Science*, *(11)*, 40–42.

Song, Q., & Wu, Y. (2017). Longitudinal data latent growth curve model and its implementation in Mplus. *Chinese Journal of Epidemiology*, *38*(8), 1135–1139.

Song, X., & Liu, H. (2016). Handling paired comparison data and ranking data: Model analysis methods. *Journal of Beijing Normal University (Natural Science Edition)*, *52*(4), 525–531.

Song, Q., Zhang, L., & Pan, J. (2021). Bayesian multi-group comparison—Approximate measurement invariance. *Psychological Exploration*, *41*(1), 69–75.

Su, R., & Xu, M. (2017). Application of latent growth models in sports. *Statistics and Information Forum*, *32*(2), 91–95.

Sun, J., & Yang, X. (2009). Several issues to note in PLS path model application. *Statistical Education*, *(11)*, 3–10.

Tian, F. (2007). Constructing indicator systems using structural equation models. *Journal of Anhui University (Philosophy and Social Sciences Edition)*, *31*(6), 92–95.

Tian, X., & Fu, J. (2004). Discussion on statistical methods of structural equation models and comparison. *Journal of Suzhou University (Natural Science Edition)*, *21*(4), 80–85.

Tong, Q., Liu, T., & Tong, H. (2009). Constrained least squares solution and deterministic algorithm for structural equation models. *Journal of Numerical Computation and Computer Applications*, *30*(3), 171–180.

Tong, Q., Zou, X., Xiong, L., & Tong, H. (2009). Constrained least squares solution for multilevel and multi-object structural equation models. *Journal of Wuhan University of Technology (Information and Management Engineering Edition)*, *31*(6), 865–869.

Wang, B., Zhang, M., Zhang, J., & Hu, J. (2015). Individual stage development described by transition matrix: Latent transition model. *Psychological Research*, *8*(4), 36–43.

Wang, C., Wang, D., Zhao, X., Fang, Q., & Liu, Y. (2010). Application and comparison of fit indices in structural equation models. *Modern Preventive*

Medicine, *37*(1), 7-9.

Wang, H., Han, H., Cai, S., Liang, Q., Wu, X., Wang, B., Zhang, Z., & Liu, S. (2012). Structural equation model for categorical variables and its application. *Chinese Journal of Health Statistics*, *29*(4), 522–524.

Wang, H., & Fu, L. (2004). Application of PLS path model in constructing comprehensive evaluation indices. *Systems Engineering—Theory and Practice*, *24*(10), 80-85.

Wang, H., & Zhang, Y. (2007). Predictive modeling methods for structural equation models. *Journal of Beijing University of Aeronautics and Astronautics*, *33*(4), 477–480.

Wang, J., Tang, W., Zhang, M., Zhang, W., & Guo, K. (2017). Methods and research status of piecewise growth mixture models. *Advances in Psychological Science*, *25*(10), 1696–1704.

Wang, K., Chen, F., Tan, M., & Chen, P. (2018). A new corrected fit index for evaluating structural equation model fit. *Chinese Journal of Health Statistics*, *35*(3), 349–354.

Wang, K., Liu, H., & Jiang, C. (2018). Hidden Markov structural equation model and its Bayesian estimation. *Journal of Applied Statistics and Management*, *37*(2), 272-279.

Wang, M., & Bi, X. (2018). Regression mixture model: Methodological progress and software implementation. *Advances in Psychological Science*, *26*(12), 2272–2280.

Wang, M., Bi, X., & Ye, H. (2014). Growth mixture model: Analyzing development trends of different categories of individuals. *Sociological Studies*, *29*(4), 220–241.

Wang, M., & Deng, Q. (2016). Structural equation modeling with missing data: Role of auxiliary variables in full information maximum likelihood estimation. *Acta Psychologica Sinica*, *48*(11), 1489–1498.

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

Wang, M., Deng, Q., Bi, X., Ye, H., & Yang, W. (2017). Performance of classification accuracy index Entropy in latent profile analysis: A Monte Carlo simulation study. *Acta Psychologica Sinica*, *49*(11), 1473–1482.

Wang, N., Ge, S., Wang, Z., & Wu, J. (2013). Differences between reflective and formative measurement models: Empirical research based on TAM. *Systems Engineering—Theory and Practice*, *33*(12), 3127-3138.

Wang, N., Zhong, W., & Mei, S. (2011). Measurement model misspecification and simulation analysis in Chinese management research. *Journal of Industrial Engineering and Engineering Management*, *25*(2), 44–49.

Wang, R., Liu, K., Yan, Z., Wang, Q. (2014). Simplification method of structural equation models and application in human-computer behavior research. *Mathematics in Practice and Theory*, *44*(3), 92–99.

Wang, S., Luo, F., & Liu, H. (2014). Traditional scoring and IRT scoring models for forced-choice personality tests. *Advances in Psychological Science*, *22*(3), 549–557.

Wang, X., Li, X., & Shao, J. (2011). Formative measurement model: A new perspective for structural equation models. *Advances in Psychological Science*, *19*(2), 293–300.

Wang, Y., Wen, Z., & Fu, Y. (2020). Equivalence testing—A new perspective for structural equation model evaluation and measurement invariance analysis. *Advances in Psychological Science*, *28*(11), 1961–1969.

Wang, Y., Wen, Z., Liu, H., Shen, J., Tan, Y., Li, D., Ma, Y. (2021). Review of psychometric methods research in China in the 20 years of the new century. *Advances in Psychological Science*, *29*(8), 1331–1344.

Wang, Y., Wen, Z., Xie, J., & Ouyang, J. (2022). Review of mediation effect methodological research in China. *Advances in Psychological Science*, *XX*(X), ???–???.

Wang, Y., Wen, Z., Chen, H., Ye, B., & Cai, B. (2022). Reliability research in China in the 20 years of the new century. *Advances in Psychological Science*, *XX*(X), ???–???.

Wang, Z., Tian, M., & Hou, Z. (2020). Quantile effect measurement based on partial least squares path models. *Journal of Systems Science and Mathematical Sciences*, *40*(4), 715–727.

Wei, J., Zhang, C., Zhao, Y., & Zhang, J. (2016). Application of random intercept factor analysis model in controlling item wording effects. *Psychological Science*, *39*(4), 1005–1010.

Wei, X., & Zheng, G. (2015). Overview of measurement invariance research. *Systems Engineering*, *33*(3), 64–71.

Wen, C., Wu, W., & Lin, G. (2018). Alignment—A new multi-group analysis method. *Advances in Psychological Science*, *27*(1), 1–8.

Wen, C., & Zhu, H. (2021). Random intercept latent transition analysis (RI-LTA)—Separating individual self-transition and inter-individual differences. *Advances in Psychological Science*, *29*(10), 1773–1782.

Wen, H., & Liang, Y. (2015). Essence of common fit index tests in structural equation models. *Psychological Science*, *38*(4), 987–994.

Wen, Z., & Hou, J. (2008). Critical values for testing: How large must the difference be to distinguish?—Comment on "Performance of fit indices under different conditions and selection of critical values" . *Acta Psychologica Sinica*, *40*(1), 119–124.

Wen, Z., Hou, J., & Marsh, H. W. (2004). Structural equation model testing: Fit indices and chi-square criteria. *Acta Psychologica Sinica*, *36*(2), 186–194.

Wen, Z., & Liu, H. (2020). *Mediation and Moderation Effects: Methods and Applications*. Educational Science Publishing House.

Wen, Z., Tang, D., & Gu, H. (2019). General simulation comparison of bifactor models and higher-order factor models from a predictive perspective. *Acta Psychologica Sinica*, *51*(3), 383–391.

Wen, Z., Fang, J., Chen, H., Ye, B., & Cai, B. (2022). Reliability research in China in the 20 years of the new century. *Advances in Psychological Science*, *XX*(X), ???–???.

Wen, Z., Fang, J., Xie, J., & Ouyang, J. (2022). Review of mediation effect methodological research in China. *Advances in Psychological Science*, *XX*(X), ???–???.

Wen, Z., Fang, J., Shen, J., Tan, Y., Li, D., & Ma, Y. (2021). Review of psychometric methods research in China in the 20 years of the new century. *Advances in Psychological Science*, *29*(8), 1331–1344.

Wu, R. (2010). Monte Carlo simulation of estimation convergence problems in structural equation models. *Statistics and Decision*, *(6)*, 32–34.

Wu, R. (2012). Parameter estimation method for structural equation models based on Tikhonov regularization. *Statistics and Decision*, *(20)*, 23–25.

Wu, R., & Tu, D. (2013). Item factor analysis: SEM-based and IRT-based methods. *Psychology and Behavior*, *11*(1), 124–131.

Wu, R., & Zu, J. (2010). Estimation and application of polychoric correlation coefficients. *Statistics and Decision*, *(3)*, 25–28.

Wu, S., & Zhang, Y. (2011). Structural equation model equivalence testing and its application in group comparison. *Chinese Journal of Health Statistics*, *28*(3), 237–239.

Wu, S., Zhang, Y., Zhang, K., Xu, Y., & Sun, N. (2009). Application of mean structure models in depression case-control clinical research. *Chinese Journal of Health Statistics*, *26*(4), 352–354.

Wu, Y., Liu, Y., Ling, W., & Lu, H. (2012). Discussion on structural modeling for dichotomous data—Taking neurotic personality affecting teacher job burnout as an example. *Journal of Guangzhou University (Natural Science Edition)*, *11*(4), 98–104.

Wu, Y., & Wen, Z. (2011). Item parceling strategies in structural equation modeling. *Advances in Psychological Science*, *19*(12), 1859–1867.

Xia, Y., Chen, G., & Liu, Y. (2016). Robust inference for hidden Markov latent variable models based on multivariate t-distribution. *Journal of Systems Science and Mathematical Sciences*, *36*(10), 1783–1803.

Xiao, J., Ye, L., & Fang, Y. (2020). Application progress of growth mixture models in health trajectory research. *Chinese Journal of Health Statistics*, *37*(4), 637–640.

Xu, B., Chen, B., & Chen, Q. (2007). Application of structural equation models in repeated measurement data. *Modern Preventive Medicine*, *34*(20), 3889–3891.

Xu, H. (2010). Application of structural equation model multi-group analysis in applied linguistics—Demonstration with Amos 17.0. *Foreign Language Education in China*, *3*(1), 59–67.

Xu, S., Yu, Z., & Li, Y. (2017). Simulation comparison of bifactor models and higher-order models from a predictive perspective. *Acta Psychologica Sinica*, *49*(8), 1103–1112.

Yan, B., Chu, X., & Zhang, L. (2019). User perception modeling method combining structural equation model and artificial neural network. *Journal of Shanghai Jiao Tong University*, *53*(7), 830–837.

Yan, N., Li, Y., Li, Y., Guo, L., & Mao, Z. (2018). Application of Bayesian structural equation model in sports and exercise psychology. *Journal of Beijing Sport University*, *41*(9), 75–82.

Yan, N., Mao, Z., Li, Y., Li, Y., & Guo, L. (2018). Small sample latent variable modeling: Application of Bayesian estimation. *China Sport Science and Technology*, *54*(6), 52–58.

Yang, T., Yang, H., & Wang, X. (2010). Comparison of structural equation models: Focusing on user acceptance of information systems. *Journal of Intelligence*, *29*(3), 15–19.

Yang, L., & Cao, Y. (2012). Comparison of two longitudinal missing data processing methods under Bayesian theoretical framework—Taking latent variable growth curve model as an example. *Journal of Jiangxi Normal University (Natural Science Edition)*, *36*(5), 461–465.

Ye, H., & Li, M. (2014). Reflective or formative? Parallel or hierarchical? Model construction and testing of PSM. *Psychological Exploration*, *34*(3), 265–271.

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

Yin, K., Peng, J., & Zhang, J. (2020). Application of latent profile analysis in organizational behavior. *Advances in Psychological Science*, *28*(7), 1056–1070.

Yu, J., Chen, X., Gao, Y., Zhang, Y., Chen, B., Kong, X., Yang, S., & Li, L. (2018). Application of latent growth mixture model in medical research. *Chinese*

Journal of Health Statistics, *35*(4), 496–499.

Yu, L. (2020). Research on first-level indicator measurement methods in academic evaluation—Structural equation dimension reduction method. *Journal of Information Resources Management*, *10*(5), 76–84.

Yuan, S., Cao, W., Zhang, M., Wu, S., Wei, X. (2021). Toward more precise causal analysis: New developments in cross-lagged models. *China Human Resources Development*, *38*(2), 23-41.

Zeng, X., Xiao, L., & Zhang, Y. (2013). Principle and case analysis of latent class analysis. *Chinese Journal of Health Statistics*, *30*(6), 815–817.

Zhang, J. (1993). A new statistical method and research approach—Review of structural equation modeling. *Acta Psychologica Sinica*, *25*(1), 93–101.

Zhang, J., Jiao, C., & Zhang, M. (2010). Application of latent class analysis technology in psychological research. *Advances in Psychological Science*, *18*(12), 1991–1998.

Zhang, J., Zhang, M., Jiao, C., & Wang, L. (2013). Application of multilevel latent class model in educational evaluation—Taking English academic ability test as an example. *Educational Research and Experiment*, *(3)*, 78–84.

Zhang, J., Zhang, M., & Li, G. (2017). Subsequent analysis of latent profile models—Bias improvement of inclusive classification analysis method. *Psychological Exploration*, *37*(5), 434–440.

Zhang, J., Zhang, M., & Geng, S. (2012). Discussion on characteristics of latent variable space in psychological research. *Psychological Exploration*, *32*(5), 404–409.

Zhang, J., Zhang, M., Jiao, C., & Wang, L. (2019). Application of inclusive classification analysis method in subsequent multiple regression of latent profile models. *Psychological Exploration*, *39*(1), 40–46.

Zhang, L., Pan, J., Dubé, L., & Ip, E. H. (2021). blcfa: An R package for Bayesian model modification in confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 649–658.

Zhang, L., Yang, J., & Xiang, A. (2012). Application research of multi-group analysis in medical psychology. *Chinese Journal of Health Statistics*, *29*(1), 105-109.

Zhang, Y., Liu, G., Lu, L., Zheng, J., Xu, X., & Yao, H. (2006). Application of multilevel CFA model in construct validity evaluation. *Chinese Journal of Health Statistics*, *23*(1), 24–26.

Zhang, Y., Liu, G., & Xu, X. (2008). Multilevel structural equation model and its application. *Chinese Journal of Health Statistics*, *25*(2), 120–123.

Zhang, Y., Liu, G., Zheng, J., & Xu, X. (2005). Latent variable scores of confirmatory factor analysis models and their application. *Modern Preventive*

Medicine, *32*(4), 285–286.

Zhang, Y. (2009). Gibbs sampling and Bayesian estimation for structural equation models. *Statistics and Decision*, *(6)*, 23–25.

Zhang, Y., & Wang, H. (2008). Application of structural equation models in system comprehensive evaluation indices. *Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition)*, *21*(1), 10–12.

Zhao, B. (2007). Testing measurement equivalence with Amos. *Chinese Journal of Health Statistics*, *24*(6), 659–661.

Zhao, H., & Wan, D. (2003). Comparison between structural equation model and artificial neural network model. *Systems Engineering—Theory Methodology Applications*, *12*(3), 262–269.

Zhao, L., Li, L., Zhou, S., Zhang, Y., & Gao, Y. (2013). Simulation comparison of latent profile analysis and hierarchical clustering. *Journal of Guangdong Pharmaceutical College*, *29*(2), 206–209.

Zhao, P. (2011). Discussion on weight estimation algorithm patterns in second-order PLS-PM models. *Statistics and Decision*, *(13)*, 4–7.

Zheng, G., Wei, X., & Yin, X. (2014). New developments in measurement invariance research. *Management Modernization*, *34*(5), 123–125.

Zheng, S., Zhang, L., Qiao, X., & Pan, J. (2021). Intensive tracking data analysis: Models and applications. *Advances in Psychological Science*, *29*(11), 1–14.

Zheng, W., & Wu, W. (2014). Structural equation model fit evaluation: Overall fit, internal fit, and cross-validation. *Psychological Exploration*, *34*(1), 75–79.

Zheng, X., Gu, H., & Zhao, B. (2011). Measurement equivalence testing for second-order factor models—Taking college students' online altruistic behavior scale as an example. *Psychological Science*, *34*(5), 1195–1200.

Zhou, G., Fan, B., Wang, C., You, D., Liu, Y., Xue, F., Chen, W., & Zhang, T. (2020). Application of cross-lagged path analysis in studying causal temporal relationships between variables. *Chinese Journal of Health Statistics*, *37*(6), 813–817.

Zhou, Y., Ou, C., Zhao, Z., Peng, C., Lu, M., & Lu, J. (2013). Application of structural equation model based on polychoric correlation coefficient in studying influencing factors of college students' self-learning expectations. *Chinese Journal of Health Statistics*, *30*(4), 529–531.

Zhu, L., & Liu, L. (2005). A simple method for linear structural equation parameter estimation. *Applied Probability and Statistics*, *21*(2), 161–168.

Zhu, M. (2016). Improving ARMA model parameter estimation based on structural equation models. *Software Guide*, *15*(9), 6–9.

Zhuang, W., & Liu, Z. (2013). Improvement and system implementation of a fuzzy comprehensive evaluation algorithm based on structural equation models. *Statistics and Decision*, *(12)*, 24–27.

Note: The reference list continues with English-language references that were already in proper format in the original text. These have been preserved exactly as they appeared.

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