Bayesian Multi-group Comparison–Asymptotic Measurement Invariance

Authors: Zhang Lijin, Song Qiongya, Pan Junhao, Pan Junhao

Date: 2019-03-01T00:00:00+00:00

Abstract

Measurement invariance of assessment instruments constitutes a prerequisite for conducting multi-group comparisons within the latent variable framework. The Bayesian approximate measurement invariance methodology leverages the advantageous properties of Bayesian principles by providing appropriate prior distributions for cross-group parameter differences, thereby relaxing the stringent constraints imposed by traditional multi-group confirmatory factor analysis approaches. This approach simultaneously circumvents common issues associated with conventional methods, including excessively poor model fit, cumbersome modification procedures, and inflated Type I error rates, thus exhibiting substantial practical utility. This article provides a comprehensive overview of the principles and advantages of approximate measurement invariance, and demonstrates its specific analytical implementation in Mplus software through an empirical example.

Full Text

Bayesian Multiple-Group Analysis: Approximate Measurement Invariance

ZHANG Lijin[1], SONG Qiongya[1], PAN Junhao[1]*

[1] Department of Psychology, Sun Yat-sen University, Guangzhou 510006, China

Abstract: Measurement invariance (MI) is a prerequisite for meaningful multi-group comparisons within the latent variable framework. The Bayesian approximate measurement invariance method leverages the advantageous properties of Bayesian estimation by providing appropriate prior distributions for cross-group differences in parameters, thereby relaxing the stringent restrictions imposed by traditional multiple-group confirmatory factor analysis. This approach simultaneously avoids problems commonly encountered in conventional methods, such as poor model fit, cumbersome modification processes, and inflated Type I error rates, making it highly valuable for applied research. This article summarizes and introduces the principles and advantages of the approximate measurement invariance method, and demonstrates its specific analytical procedure in Mplus software through an empirical example.

Keywords: Bayesian method; multiple-group confirmatory factor analysis; approximate measurement invariance

1 Introduction

In social science fields such as psychology and education, testing for group differences has long received widespread attention, including investigations of cognitive ability differences across age groups (Ouyang, Tian, Xin, & Zhan, 2016) and behavioral and attitudinal differences across cultural backgrounds (Shorey, Allan, Cohen, Fite, Stuart, & Temple, 2018). In such research, the variables of interest are often abstract constructs that cannot be directly observed but must be reflected through directly observable variables such as scale items. These unobservable variables are termed latent variables (e.g., attitudes, values, abilities), while the scale items used to reflect them are called observed variables.

When conducting multi-group comparisons within the latent variable framework, measurement invariance (MI) of the measurement instrument serves as a prerequisite for valid group comparisons. Measurement invariance means that the construct being measured does not change across different populations; that is, participants from different groups exhibit the same response pattern to identical items (Mellenbergh, 1989; van de Schoot, Kluytmans, Tummers, Lugtig, Hox, & Muthén, 2013). Structural equation modeling, grounded in the latent variable framework, provides a practical approach for testing measurement invariance. Structural equation modeling typically comprises two components: a measurement model and a structural model. The measurement model primarily reflects the relationship between observed and latent variables, usually taking the form of a confirmatory factor analysis (CFA) model. Measurement invariance testing can thus be implemented through the measurement model component of structural equation modeling.

Multiple structural equation modeling approaches have been developed for testing measurement invariance, including multiple-group confirmatory factor analysis (e.g., Raju, Laffitte, & Byrne, 2002; Stark, Chernyshenko, & Drasgow, 2006), multilevel confirmatory factor analysis (Jak, Oort, & Dolan, 2013), multilevel factor mixture models (Kim, Cao, Wang, & Nguyen, 2017), the alignment method (Asparouhov & Muthén, 2014), and the extended alignment method (Marsh et al., 2018). Among these, multiple-group confirmatory factor analysis (multiple-group CFA) is the most commonly used method (Meade & Lautenschlager, 2004; Raju, Laffitte, & Byrne, 2002). This paper will focus on an in-depth discussion of traditional multiple-group CFA methods and Bayesian multiple-group CFA methods.

1.1 Multiple-Group Confirmatory Factor Analysis

Multiple-group CFA modeling requires establishing a separate confirmatory factor analysis model for each group and comparing differences in response patterns across groups (Kim et al., 2017; Rutkowski & Syetina, 2014). Testing measurement invariance in multiple-group CFA involves sequentially imposing cross-group equality constraints on measurement model parameters, including: (1) Configural invariance, which requires identical factor structures across groups (i.e., scale items load onto the same latent variables in each group); (2) Metric invariance (also called weak measurement invariance), which specifies equal factor loadings across groups, meaning that a one-unit change in the latent variable produces equivalent change in observed variables across groups; (3) Scalar invariance (also called strong measurement invariance), which requires equal item intercepts across groups, indicating that different groups share the same measurement reference point. Only when both units and reference points are equivalent across groups (i.e., loading and intercept invariance) can meaningful cross-group comparisons of factor means be conducted; otherwise, biased estimates may result (Schmitt & Kuljanin, 2008); and (4) Error-variance invariance (also called strict measurement invariance), which specifies equal error variances across groups, meaning that cross-group differences in observed variable score variance fully reflect cross-group differences in factor score variance (Rutkowski & Svetina, 2014).

Multiple-group CFA requires model fit comparison after each measurement invariance constraint is imposed to test whether the constrained model shows significantly worse fit. Only when current invariance constraints are satisfied can subsequent invariance tests proceed. Nested model comparisons can employ chi-square difference tests, which assess whether model fit deteriorates significantly after imposing constraints based on changes in chi-square values and degrees of freedom. However, because chi-square difference tests are sensitive to sample size, researchers have proposed more effective model comparison criteria, such as ΔCFI 0.01 and ΔRMSEA 0.03 (Little & Card, 2013; Rutkowski & Svetina, 2014).

The multiple-group CFA method is conceptually straightforward and easy to master, leading to its widespread application. When multiple-group CFA models satisfy cross-group intercept invariance, researchers can further test structural invariance, including cross-group comparisons of factor means and factor variances/covariances (Little & Card, 2013). Currently popular structural equation modeling software supports multiple-group CFA modeling, facilitating researchers' application of this method.

1.2 Limitations of Multiple-Group Confirmatory Factor Analysis

Despite its advantages and widespread use, multiple-group CFA often imposes overly stringent restrictions on cross-group invariance in applied research, requiring strict equality of loadings and intercepts across groups, which makes constrained models more likely to be rejected (Asparouhov & Muthén, 2014; Kim et al., 2017). Marsh et al. (2018) noted that scalar invariance models are almost never satisfied in applied research, with this problem becoming more severe when the number of groups is large, substantially impacting subsequent analyses (Rutkowski & Svetina, 2014).

When loading or intercept invariance is not satisfied in applied research, researchers typically employ post-hoc modification methods, combining theoretical considerations with modification indices to free parameters that violate cross-group invariance, thereby establishing partial measurement invariance models. However, this approach has several limitations: (1) Post-hoc modification based on modification indices requires model fit comparison for each freed parameter, making the modification process extremely cumbersome when many parameters need adjustment; (2) Parameters identified as violating measurement invariance through modification indices may result from sampling variability (Asparouhov & Muthén, 2014), and parameter freeing is also subject to researcher subjectivity, increasing Type I error rates and reducing replicability; and (3) Conducting group comparisons based on such partial measurement invariance models can lead to biased estimates of cross-group factor mean differences (Marsh et al., 2018).

These limitations of multiple-group CFA pose substantial obstacles for applied research, with problems becoming more severe when the number of groups or violating parameters is large. Consequently, methodologists have continuously explored and developed more accurate and applicable multi-group comparison methods. The Bayesian approximate measurement invariance method proposed by Asparouhov and Muthén (2013) effectively addresses problems arising from overly strict cross-group restrictions in traditional methods. Unfortunately, due to its relatively recent development, applied researchers are unfamiliar with this modeling approach and its software implementation. This article provides a comprehensive introduction and summary of this method, explaining its principles and advantages, and demonstrates how to apply approximate measurement invariance in empirical research through example analysis, comparing it with traditional multiple-group CFA methods.

2.1 Methodological Principles

In structural equation modeling, besides traditional frequentist estimation methods (e.g., maximum likelihood estimation), Bayesian estimation methods have become increasingly popular in recent years (Wang, Deng, & Bi, 2017; Zhang, Lu, Wei, & Pan, 2018; van de Schoot et al., 2017). Compared to traditional methods that often encounter problems such as model non-identification and biased parameter estimates with small samples, non-normal data, and complex models, Bayesian structural equation modeling is more tolerant of extreme model or data conditions (Li, 2011).

From a statistical perspective, traditional frequentist methods treat unknown parameters as constants, whereas Bayesian methods treat them as random variables, combining sample data with prior information to obtain posterior distributions of unknown parameters (Wang et al., 2017). Bayesian methods can therefore flexibly incorporate prior information (e.g., from previous research or pilot studies), and informative priors can lead to more accurate parameter estimates (Yuan & MacKinnon, 2009). Researchers' confidence in prior accuracy is reflected in the strength of prior information, that is, the variance of the prior distribution. Smaller prior variance exerts greater influence on the posterior distribution of unknown parameters. For example, parameters constrained to 0 in structural equation models can be viewed as receiving a prior distribution with mean 0 and variance 0.

Given these advantageous properties of Bayesian methods in structural equation modeling, increasing numbers of methodologists and applied researchers have devoted efforts to developing and applying Bayesian structural equation models (van de Schoot et al., 2017), leading to a series of new modeling approaches. For instance, traditional structural equation modeling requires theoretically assuming no cross-loadings or correlated measurement errors, strictly constraining these parameters to 0—restrictions that are often difficult to satisfy in practice (Muthén & Asparouhov, 2012). Based on the Bayesian capability to incorporate prior information, Muthén and Asparouhov (2012) creatively proposed that providing normal prior distributions with mean 0 and extremely small variance for cross-loadings and measurement error correlations allows these parameters to vary slightly around 0, relaxing strict model restrictions. Significant cross-loadings or measurement error correlations discovered after relaxing restrictions can be treated as nuisance parameters resulting from sampling variability that require no theoretical explanation. This idea of relaxing model restrictions has received extensive research and extension since its proposal, with researchers developing numerous optimization methods based on this concept (e.g., Lu, Chow, & Loken, 2016; Pan, Ip, & Dubé, 2017). Muthén and Asparouhov (2013) further applied this approach to multi-group structural equation modeling, developing the approximate measurement invariance method.

In traditional multiple-group CFA modeling, we sequentially impose strict equality constraints on cross-group loadings and intercepts based on the configural invariance model. However, as previously noted, such restrictions are almost never satisfied in applied research (Marsh et al., 2018). Although partial measurement invariance can be adopted by freeing some violating parameters, this approach has numerous limitations. In approximate measurement invariance analysis, providing normal prior distributions with mean 0 and extremely small variance for cross-group differences in loadings and intercepts relaxes the strict restrictions of traditional methods, allowing small cross-group differences in these parameters. Muthén and Asparouhov (2013) noted that the approximate measurement invariance method essentially assumes the existence of many small, random cross-group differences that may cancel each other out, thus not affecting tests of structural invariance. This method can achieve accurate estimation of structural equation model parameters while permitting small cross-group differences.

Based on the idea of relaxing model restrictions and the advantageous properties of Bayesian methods, approximate measurement invariance offers several advantages over traditional methods: (1) It avoids the problem of poor model fit caused by overly strict restrictions in traditional methods; (2) It relaxes restrictions on all parameters in a single estimation without requiring multiple post-hoc modifications, thereby avoiding higher Type I error rates and biased parameter estimates resulting from repeated use of the same data (Draper, 1995); (3) It can serve as a tool for identifying parameters that violate measurement invariance (Muthén & Asparouhov, 2013), as it can simultaneously detect all non-invariant parameters; and (4) It can test the degree of invariance violation by providing different small-variance priors (Cieciuch, Davidov, Algesheimer, & Schmidt, 2017).

Simulation studies have also shown that when loadings or intercepts exhibit small cross-group differences, the Bayesian approximate measurement invariance method performs well in both model fit and parameter estimation (Kim et al., 2017; van de Schoot et al., 2013). Moreover, because the approximate measurement invariance method is based on the Bayesian structural equation modeling framework, it offers greater flexibility and more effective model fit evaluation (Kim et al., 2017; Schmitt & Kuljanin, 2008).

2.2 Analysis Steps

Since its proposal, the approximate measurement invariance method has received extensive development and application (De Bondt & Van Petegem, 2015; van de Schoot et al., 2013). However, due to its relatively recent development, specific analytical procedures and evaluation criteria have not yet been standardized. Based on application studies and simulation research on approximate measurement invariance (Fong & Ho, 2014; Kim et al., 2017; van de Schoot et al., 2013), this article summarizes a more complete and comprehensive analytical procedure.

2.2.1 Configural Invariance

Similar to traditional methods, establishing an approximate measurement invariance model first requires building a baseline model—the configural invariance model. Only when model fit is adequate can we proceed to impose loading and intercept invariance constraints. Adequate model fit is indicated by a posterior predictive p-value (PPp) around 0.5 and a posterior predictive interval that includes 0. The criterion of PPp around 0.5 is somewhat subjective; simulation studies suggest using a cutoff of 0.1 (Meghan & Zhang, 2018; Muthén & Asparouhov, 2012), where a PPp greater than 0.1 with a posterior predictive interval including 0 indicates adequate model fit.

Additionally, in testing configural invariance, we can follow Muthén and Asparouhov' s (2012) approach of relaxing strict restrictions on cross-loadings or measurement error correlations (Fong & Ho, 2014) by providing normal prior distributions with mean 0 and extremely small variance for these parameters.

2.2.2 Loading and Intercept Invariance

Based on configural invariance, traditional methods require sequentially imposing loading and intercept invariance constraints. This is primarily because when traditional methods encounter significant model fit deterioration (i.e., violation of cross-group invariance constraints), researchers must use post-hoc modification methods to free violating parameters one by one. Simultaneously imposing loading and intercept equality constraints would introduce too many potential sources of model misfit, complicating the identification of violating parameters (Little & Card, 2013).

The approximate measurement invariance method can relax strict equality restrictions on all loadings and intercepts simultaneously in a single estimation by providing normal prior distributions with mean 0 and extremely small variance for cross-group differences. Therefore, in approximate measurement invariance analysis, loading and intercept invariance constraints can be imposed simultaneously—that is, providing normal prior distributions with mean 0 and small variance (a variance of 0.01 is recommended, which allows 95% of cross-group differences to fall within (-0.2, 0.2) when data are standardized) for cross-group differences in both loadings and intercepts (Muthén & Asparouhov, 2013). Poor model fit at this stage indicates that the data do not satisfy invariance assumptions, while adequate model fit requires further sensitivity analysis.

2.2.3 Sensitivity Analysis

Both van de Schoot et al. (2013) and Kim et al. (2017) recommend conducting sensitivity analysis (Greenland, 2001) when using approximate measurement invariance to avoid researcher subjectivity in selecting prior variance.

Specifically, researchers should first determine a predetermined prior variance, with 0.01 recommended (Kim et al., 2017). After imposing cross-group invariance constraints and providing N(0, 0.01) priors for loading and intercept differences, we obtain model-data fit indices such as the PPp value, Deviance Information Criterion (DIC), and Bayesian Information Criterion (BIC) provided in Mplus. This step constitutes the loading and intercept invariance stage described above.

When the PPp value indicates adequate model fit, we then provide prior distributions with larger variance for cross-group differences, typically N(0, 0.05) (Kim et al., 2017; van de Schoot et al., 2013). Model fit is compared between models with these two prior specifications: if the model with N(0, 0.01) priors shows better fit or no significant difference, we can conclude that the model satisfies cross-group loading and intercept invariance; if the model with N(0, 0.01) priors shows significantly worse fit, this indicates that imposing cross-group invariance constraints is incorrect and causes noticeable model deterioration, meaning the model does not satisfy cross-group invariance (Kim et al., 2017).

For model comparison, Kim et al. (2017) recommend using the DIC criterion because BIC will always favor the model with prior variance of 0.05, likely because the larger variance allows more substantial cross-group differences and typically identifies more significantly non-invariant parameters. In contrast, DIC penalizes overly complex models and aligns more closely with the concept of Bayesian deviance (Kaplan & Depaoli, 2012), making it more commonly used in Bayesian estimation. In practice, when the absolute difference in DIC between two models exceeds 7, there is sufficient evidence to conclude that the model with smaller DIC is superior; otherwise, the two models are not significantly different (Li, 2011; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002).

With loading and intercept invariance established, researchers can proceed to compare factor means across groups to investigate differences in the construct across multiple groups. Although approximate measurement invariance currently meets the needs of most applied research, no appropriate prior distribution has been developed to relax strict restrictions on cross-group equality of measurement error variances/covariances. Consequently, approximate measurement invariance cannot yet support tests of error variance/covariance invariance. We hope that with the vigorous development of Bayesian methods, better solutions to this limitation will emerge in the future.

2.3 Partial Approximate Measurement Invariance

The Bayesian approximate measurement invariance method assumes numerous small group differences exist in loading and intercept parameters, and performs well in parameter estimation under such conditions. However, if some loadings or intercepts exhibit large cross-group differences, the small-variance prior N(0, 0.01) may severely compress the true group differences, leading to biased estimates of cross-group mean differences (Muthén & Asparouhov, 2013). Therefore, researchers recommend adopting partial approximate measurement invariance to freely estimate these parameters after identifying significant violations, achieving more accurate cross-group factor mean comparisons (Muthén & Asparouhov, 2013; van de Schoot et al., 2013).

In traditional multiple-group CFA, partial measurement invariance refers to freeing parameters that violate cross-group invariance when strict loading or intercept invariance models are not satisfied, thereby establishing partial measurement invariance models. Structural invariance can be further tested based on such models (Schmitt & Kuljanin, 2008), and when the number of violating parameters is relatively small, partial measurement invariance has minimal impact on the accuracy of structural model parameter estimation (Little & Card, 2013).

Similarly, in Bayesian approximate measurement invariance, providing N(0, 0.01) priors for cross-group differences can detect all parameters with cross-group differences significantly different from 0 (i.e., non-invariant parameters). By removing the previously provided N(0, 0.01) priors for these parameters and adopting Mplus' s default non-informative priors (i.e., freely estimating them), we can establish partial approximate measurement invariance models (Muthén & Asparouhov, 2013). Compared to traditional methods, Bayesian partial approximate measurement invariance can detect all violating parameters in a single estimation, dramatically reducing the workload of post-hoc modification methods and avoiding researcher subjectivity in selecting parameters for modification. Simulation studies also show that Bayesian partial approximate measurement invariance yields more accurate parameter estimates (van de Schoot et al., 2013).

Nevertheless, Bayesian partial approximate measurement invariance models share the same two main problems as traditional partial measurement invariance models: (1) Parameters requiring freeing may result from sampling variability; and (2) The impact of freeing violating parameters on structural model parameter estimation remains uncertain (Schmitt & Kuljanin, 2008). Therefore, researchers should also integrate theoretical hypotheses to establish more theoretically meaningful partial approximate measurement invariance models. We also hope future simulation studies will investigate the impact of partial measurement invariance on structural model parameter estimation to provide more accurate guidance for applied researchers.

3 Example: Bayesian Approximate Measurement Invariance Analysis in Structural Equation Modeling

To examine gender differences in responses to a social support scale, this study employs both traditional multiple-group CFA and Bayesian approximate measurement invariance methods to demonstrate the specific analytical steps of the Bayesian approach and compare it with multiple-group CFA. Data from 353 undergraduate students (mean age = 19.55, SD = 1.45, range: 17-24), including 89 males, were analyzed using Mplus 8.0. The Mplus code for Bayesian approximate measurement invariance is provided in the Appendix.

The study used the College Student Social Support Scale developed by Ye and Dai (2008), which consists of 17 items measuring three dimensions: subjective support (5 items), objective support (6 items), and support utilization (6 items), rated on a 5-point Likert scale. Sample items include "Most classmates care about me" and "When facing difficult choices, I actively seek help from others." The internal consistency coefficient for the entire scale was 0.928 in this study, with subscale coefficients of 0.892, 0.818, and 0.881, indicating good reliability.

Previous research indicates that internal consistency coefficients assume independence of measurement errors among observed variables; otherwise, biased reliability estimates may result (Raykov, 2001). Composite reliability, which allows for correlated measurement errors and different item loadings, is more appropriate for latent variables (Xu, 2008). According to Fornell and Larcker (1981), composite reliability above 0.6 indicates adequate scale reliability. In this study, subscale composite reliabilities were 0.894, 0.819, and 0.882, demonstrating good composite reliability.

3.1 Multiple-Group Confirmatory Factor Analysis

Based on skewness and kurtosis analyses, item skewness ranged from -1.107 to -1.43 and kurtosis ranged from -0.705 to 1.534. According to conventional criteria, absolute skewness > 2 and absolute kurtosis > 7 indicate non-normal distributions (West, Finch, & Curran, 1995). Thus, all items in this study approximately followed normal distributions, and with no missing data, maximum likelihood (ML) estimation could be used. In multiple-group CFA modeling, we sequentially established configural invariance, loading invariance, and intercept invariance models. Following Little and Card' s (2013) recommendation, $|\Delta\text{CFI}| < 0.01$ indicates no meaningful difference in model fit.

In the configural invariance model, one pair of residual correlations was found in each group. After freely estimating these, the configural invariance model showed adequate fit (Table 1). The loading invariance model did not satisfy cross-group loading equality (CFI = 0.917, $|\Delta\text{CFI}| = 0.04$). Using post-hoc modification, we freed loading equality constraints for two items[1], establishing a partial loading invariance model that did not differ significantly from the configural model (CFI = 0.920, $|\Delta\text{CFI}| = 0.01$). Testing intercept invariance on this partial loading invariance model revealed violation of intercept invariance (CFI = 0.909, $|\Delta\text{CFI}| = 0.11$). After freeing intercept equality constraints for three items[2], the partial intercept invariance model showed no significant difference from the partial loading invariance model (CFI = 0.919, $|\Delta\text{CFI}| = 0.01$).

Based on the partial intercept invariance model, comparisons of factor mean differences revealed that females scored significantly higher than males on all three dimensions–subjective support, objective support, and support utilization (standardized mean differences: dif = 0.389, p = 0.002; dif = 0.566, $p < 0.001$; dif = 0.410, p = 0.003).

Table 1 Multiple-Group Confirmatory Factor Analysis Model Fit and Comparison

Model	RMSEA	90% CI of RMSEA	Pass?
Configural invariance	[0.071, 0.091]		
Loading invariance	[0.071, 0.090]		
Partial loading invariance	[0.070, 0.089]		
Intercept invariance	[0.073, 0.091]		
Partial intercept invariance	[0.069, 0.088]		

[1] The freed items were two items from the support utilization dimension: "When I encounter trouble, I usually actively seek help from others" and "When I have worries, I actively confide in classmates and friends."

[2] The freed items included two items from the objective support dimension: "In difficult times, I can rely on family or relatives" and "I often receive care and support from family and relatives" ; and one item from the support utilization dimension: "When I have worries, I actively confide in family and relatives."

3.2 Bayesian Approximate Measurement Invariance

Following Muthén and Asparouhov' s (2012) recommendation, we provided an inverse Wishart distribution IW(I, df) for the error term matrix, where I is the identity matrix and df = p + 6 (df = 23 in this example), allowing 95% of variation in each residual covariance parameter to fall within (-0.2, 0.2). The configural invariance model showed adequate fit (Table 2).

Providing N(0, 0.01) prior distributions for cross-group differences in loadings and intercepts yielded adequate model fit for the loading and intercept invariance model. Sensitivity analysis using N(0, 0.05) priors for these differences showed no meaningful difference in model fit between the two prior specifications ($|\Delta DIC| = 2.252$) (Table 2), indicating that the data satisfied cross-group loading and intercept invariance. Additionally, after relaxing strict restrictions on cross-group differences, the "DIFFERENCE OUTPUT" identified one intercept parameter[3] that violated cross-group invariance. Freely estimating this parameter established a partial approximate measurement invariance model (Table 2).

Table 2 Bayesian Approximate Measurement Invariance Analysis–Model Fit and Comparison

Model	95% C.I.[1]	ΔDIC[2]
Configural invariance	[-86.681, 62.040]	
Loading & intercept invariance (prior variance 0.01)	[-90.916, 60.739]	
Loading & intercept invariance (prior variance 0.05)	[-90.851, 57.596]	2.252
Partial measurement invariance	[-93.554, 58.890]	

[1] 95% C.I. represents the confidence interval for the difference between observed and replicated chi-square values.
[2] ΔDIC represents the DIC difference between models with prior variances of 0.01 and 0.05 in the sensitivity analysis.

[3] The violating parameter was the intercept for the item "When I have worries, I actively confide in family and relatives."

Based on the partial approximate measurement invariance model, comparisons of factor means also revealed that females scored significantly higher than males on subjective support, objective support, and support utilization (standardized mean differences: dif = 0.394, $p = 0.004$; dif = 0.443, $p < 0.001$; dif = 0.559, $p < 0.001$). Consistent with findings from Muthén and Asparouhov (2012) and van de Schoot et al. (2013), the approximate measurement invariance method produced factor mean difference estimates similar to those from traditional methods.

3.3 Method Comparison

The above analysis demonstrates that traditional multiple-group CFA imposes overly strict restrictions on models, with more cumbersome model comparison and modification processes. Furthermore, existing comparison criteria in multiple-group CFA remain controversial (Little & Card, 2013) and are susceptible to researcher subjectivity. When loading or intercept invariance models are not satisfied, post-hoc modification methods identify more violating parameters, and model modification may involve researcher bias in selecting parameters for adjustment. Additionally, more simulation studies are needed to explore how partial measurement invariance affects subsequent factor mean difference estimation.

In contrast, Bayesian approximate measurement invariance is straightforward to implement. By relaxing strict restrictions on group differences, it avoids the poor model fit common in traditional methods, and sensitivity analysis helps researchers avoid subjectivity in prior specification, yielding more reliable results.

4 Summary and Outlook

Due to its advantageous properties, Bayesian approximate measurement invariance has gained increasing popularity among applied researchers. For example, Cieciuch et al. (2017) applied Bayesian approximate measurement invariance to European Social Survey data, using the 21-item Portrait Value Questionnaire (PVQ-21) to measure personal worldviews and compare cross-national differences. Comparing traditional multiple-group CFA with Bayesian approximate measurement invariance across 15 countries with six measurement occasions, they found that the Bayesian method enabled more countries to satisfy measurement invariance, facilitating subsequent factor mean comparisons. This method has also been applied to test gender invariance of anxiety and depression factors (Fong & Ho, 2014) and to examine attitude differences in intimate relationships between partners (Chiorri, Day, & Malmberg, 2014). We hope this introduction helps researchers master this method more quickly to overcome obstacles encountered in practical data analysis.

However, as a modeling tool, Bayesian approximate measurement invariance inevitably has certain applicability scopes and limitations. First, the method is more computationally time-consuming and less efficient with many groups. Second, how to handle situations where loading and intercept invariance cannot be satisfied remains an open question. Researchers have proposed solutions such as partial approximate measurement invariance (Muthén & Asparouhov, 2013) and the alignment method (Marsh et al., 2018), but these approaches still require development. Finally, although approximate measurement invariance can test loading and intercept invariance and be used for subsequent factor mean comparisons, it currently cannot test error variance/covariance invariance.

Besides Bayesian approximate measurement invariance, numerous other methods exist for testing measurement invariance across groups, including multilevel confirmatory factor analysis, multilevel factor mixture models, the alignment method, and extended alignment methods. While these methods have unique advantages, their applicability scopes are relatively narrow and they have many limitations. For instance, multilevel analysis requires many groups and large group differences in parameters, making it less broadly applicable than Bayesian approximate measurement invariance (Kim et al., 2017). The alignment and extended alignment methods only require configural invariance and do not involve subsequent model fit comparisons, which is disadvantageous for identifying violating parameters (Kim et al., 2017; Marsh et al., 2018). Discussion of the relative merits and applicability scopes of existing methods remains to be explored, and such comparisons could be extended to structural invariance, including tests of factor mean and factor variance/covariance invariance.

References

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

Ouyang, X., Tian, W., Xin, T., & Zhan, P. (2016). Measurement invariance analysis of longitudinal data within the IRT framework: An example of cognitive ability tests for 4- to 5-year-old children. *Psychological Science*, 3, 606-613.

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

Xu, W. (2008). Application of structural equation modeling in reliability testing. *Statistics & Information Forum*, 23(7), 9-13.

Ye, Y., & Dai, X. (2008). Development of the College Student Social Support Scale. *Chinese Journal of Clinical Psychology*, 16(5), 465-468.

Zhang, L., Lu, J., Wei, X., & Pan, J. (2018). Bayesian structural equation modeling and its research status. *Advances in Psychological Science*. doi:10.12074/201812.00865

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment.

Structural Equation Modeling: A Multidisciplinary Journal, 21(4), 495-508.

Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2017). Testing for approximate measurement invariance of human values in the European social survey. *Sociological Methods & Research*, 47(4), 665-686.

Chiorri, C., Day, T., & Malmberg, L. E. (2014). An approximate measurement invariance approach to within-couple relationship quality. *Frontiers in Psychology*, 5, 983.

De Bondt, N., & Van Petegem, P. (2015). Psychometric evaluation of the Overexcitability Questionnaire-Two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, 6, 1963.

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

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

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.

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

Jak, S., Oort, F. J., & Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 265-282.

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

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

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

Little, T. D., & Card, N. A. (2013). *Longitudinal structural equation modeling*. The Guilford Press.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., & Muthén, B., et al. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, 23(3), 524-545.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(1), 60-72.

Meghan K. Cain & Zhang Z. Y. (2018). Fit for a Bayesian: An evaluation of PPp and DIC for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 1-12.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143.

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

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

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

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87(3), 517-29.

Raykov, T. (2001). Bias of coefficient alpha fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, 25(1), 69-76.

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational & Psychological Measurement*, 74(1), 31-57.

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

Shorey, R. C., Allan, N. P., Cohen, J. R., Fite, P. J., Stuart, G. L., & Temple, J. R. (2018). Testing the factor structure and measurement invariance of the Conflict in Adolescent Dating Relationship Inventory. *Psychological Assessment*. Advance online publication. http://dx.doi.org/10.1037/pas0000678

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91(6), 1292-1306.

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4, 770.

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

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage Publications, Inc.

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

Appendix: Mplus Code for Bayesian Approximate Measurement Invariance

1. Configural Invariance Model

```
TITLE: Configural Model
DATA: FILE = socialsupport.dat;
VARIABLE:
    NAMES = Gender SS1-SS17;
    USEV = SS1-SS17;
    KNOWNCLASS IS c (Gender=1-2); !Divide into two groups by gender, 1 = male, 2 = female
    CLASSES IS c(2);    !Number of groups = 2
ANALYSIS:
    MODEL=CONFIGURAL;    !Configural invariance model, allow cross-group loadings and interc
    TYPE=MIXTURE;
    ESTIMATOR = BAYES;
    PROC = 2;
    BITERATIONS = 100000(10000); !Minimum iterations = 10000, maximum iterations = 100000, s
MODEL: %OVERALL%
    F1 BY SS1-SS5*;      !*Free the constraint that the first observed indicator's loading e
    F2 BY SS6-SS11*;
    F3 BY SS12-SS17*;
    F1-F3@1;    !Use fixed variance method for model identification
    SS1-SS17(P#_1-P#_17); !Define measurement error parameters, P#_1 refers to the first mea
    SS1-SS17 WITH SS1-SS17(P#_18-P#_153);    !Define measurement error correlation parameter
MODEL PRIORS:    !Provide prior distributions for parameters
    DO(1,2)P#_1-P#_17~IW(1,23);
    DO(1,2)P#_18-P#_153~IW(0,23); !DO(1,2) represents looping from 1 to 2
    !Provide IW(I, 23) distribution for both groups' error term matrices, relaxing strict re
OUTPUT:    TECH1    TECH8    STDY    SVALUES;
PLOT: TYPE= PLOT2;
```

2. Loading and Intercept Invariance Model

```
TITLE: Scalar Model
DATA: FILE = socialsupport.dat
VARIABLE:
    NAMES = Gender SS1-SS17;
    USEV = SS1-SS17;
    KNOWNCLASS IS c (Gender=1-2);
    CLASSES IS c(2);
ANALYSIS:
    MODEL=ALLFREE;    !Freely estimate all parameters for both groups
    TYPE=MIXTURE;
    ESTIMATOR = BAYES;
    PROC = 2;
    BITERATIONS = 100000(10000);
MODEL: %OVERALL%
    F1 BY SS1-SS5*(lam#_1-lam#_5);
    F2 BY SS6-SS11*(lam#_6-lam#_11);
    F3 BY SS12-SS17*(lam#_12-lam#_17);
    [SS1-SS17](nu#_1-nu#_17);
    SS1-SS17(P#_1-P#_17);
    SS1-SS17 WITH SS1-SS17 (P#_18-P#_153);
%c#1%      !Set group 1 as reference group, fix factor variances to 1 and means to 0
    F1-F3@1;
    [F1-F3@0];
%c#2%      !Freely estimate factor means and variances in group 2
    F1-F3;
    [F1-F3];
MODEL PRIORS:
    DO(1,2)P#_1-P#_17~IW(1,23);
    DO(1,2)P#_18-P#_153~IW(0,23);
    DO(1,17)DIFF(lam1_#-lam2_#)~N(0,0.01);
    DO(1,17)DIFF(nu1_#-nu2_#)~N(0,0.01); !Provide N(0, 0.01) prior for loading and intercept
    !DIFF(lam1_#-lam2_#) refers to the difference between the #th loading parameter in group
OUTPUT:    TECH1    TECH8 STDY    SVALUES;
PLOT: TYPE= PLOT2;
```

3. Partial Approximate Measurement Invariance Model

```
MODEL PRIORS:
    DO(1,2)P#_1-P#_17~IW(1,23);
    DO(1,2)P#_18-P#_153~IW(0,23);
    DO(1,17)DIFF(lam1_#-lam2_#)~N(0,0.01);
    DO(1,7)DIFF(nu1_#-nu2_#)~N(0,0.01);
    DO(9,17)DIFF(nu1_#-nu2_#)~N(0,0.01); !Mplus will provide non-informative priors to freel
    !Remaining model specifications are identical to the loading and intercept invariance mo
```