Journal of Management
Vol. XX No. X, Month XXXX 1–17
DOI: 10.1177/0149206315591075

Editorial Commentary

Bayesian Structural Equation Modeling With Cross-Loadings and Residual Covariances: Comments on Stromeyer et al.

Tihomir Asparouhov
Bengt Muthén
Muthén & Muthén
Alexandre J. S. Morin
Australian Catholic University

A recent article in the Journal of Management *gives a critique of a Bayesian approach to factor analysis proposed in* Psychological Methods*. This commentary responds to the authors' critique by clarifying key issues, especially the use of priors for residual covariances. A discussion is also presented of cross-loadings and model selection tools. Simulated data are used to illustrate the ideas. A reanalysis of the example used by the authors reveals a superior model overlooked by the authors.*

Keywords: *statistical methods; macro topics; measure development; research methods; psychometrics; structural equation modeling (SEM); Bayesian methods*

Consider a researcher who, drawing on theory, formulates a measurement model; designs a measurement instrument; collects data; estimates a typical confirmatory factor analysis (CFA) model with zero cross-loadings and residual covariances; and finds that the fit statistics indicate a poor fit between the data and the model. What should the researcher do in this

Acknowledgments: Muthén and Asparouhov are employed by Muthén & Muthén, which distributes the Mplus program used in the article.

Corresponding author: Tihomir Asparouhov, Muthén & Muthén, 3463 Stoner Avenue, Los Angeles, CA 90066, USA.

E-mail: tihomir@statmodel.com

arguably common situation? From a practical perspective, traditional CFA approaches limit the researcher's ability to investigate cross-loading and residual covariance parameters. It is well known that a stepwise relaxation of parameters using model modification indices both abandons the original theoretical hypothesis and risks leading to the wrong model. Freeing all cross-loadings from their fixed zero status leads to traditional applications of exploratory factor analysis (EFA) with rotations, where the focus has typically been on the number of factors rather than the hypothesized location of important versus ignorable factor loadings. Freeing all residual covariances uses up all degrees of freedom so that the model is no longer identified. A new approach using Bayesian structural equation modeling (BSEM) resolves these issues as described in Muthén and Asparouhov (2012).

In essence, the focus of this approach is not only to test the model but to generate ideas about possible model modifications that can yield a better-fitting model. Thus, the goal of BSEM is not to confirm or disconfirm the CFA model but to provide the means to evaluate the difference between the hypothesized CFA model and the data. In doing so, the BSEM approach may identify possible cross-loadings, or residual covariances, present in the population model that may have been otherwise missed by the CFA approach and that may be meaningful. As we discuss below, the alternative view that these cross-loadings or residual covariance should be avoided, which stems from problems with the way these have traditionally been included in a model, appears neither realistic nor in line with the principles of classical test theory or recent statistical studies.

In a special issue of *Journal of Management* devoted to Bayesian estimation (Zyphur & Oswald, 2015), Stromeyer, Miller, Sriramachandramurthy, and DeMartino (2015: 508) provide a discussion of BSEM, stating,

> Given the promise we see in the BSEM technique to enable management scholars to push the boundaries of theory testing, we offer the following series of recommendations to guide the future use of the BSEM technique with regards to measurement model development.

The Stromeyer et al. (2015) recommendations, however, are flawed in that they are based on criticisms of BSEM that the current article shows are not justified. Muthén and Asparouhov (2012) discussed two key factor analysis application areas for BSEM: cross-loadings and residual covariances. Stromeyer et al. have objections in both of these areas. In their critique of BSEM with cross-loadings, they argue that this can obscure measurement problems and lead to misestimated factor correlations. On the basis of our experience and our simulation studies, we dismiss these objections. Their strongest critique concerns the residual covariance application of BSEM. Their Recommendation 6 (pp. 514-515) states three reasons for why this technique should be avoided and concludes that it should not be used until guidelines become available. In this paper, we present our disagreement with this view, addressing the Stromeyer et al. concerns by providing a detailed account of the proper use of BSEM with residual covariances. It is shown that this technique in fact provides unique insights not obtained by other factor analysis approaches.

The current paper also presents our disagreement with Recommendation 5 of Stromeyer et al. (2015: 513-514), where it is stated that Bayesian information criterion (BIC) should be given preference over discrepancy information criterion (DIC) for model selection involving BSEM. Our view is that only DIC properly accounts for the BSEM parameters with small-variance priors.

In the following pages, we address each of these three issues in turn (cross-loadings, residual covariances, DIC vs. BIC). Illustrations are provided using several simulated data examples. Finally, we apply the BSEM technique thus illustrated to a reanalysis of the Stromeyer et al. (2015) example. This reanalysis leads to an alternative model that is overlooked by using ordinary factor analysis or BSEM with only cross-loadings.

Cross-Loadings in BSEM

First we want to point out that the Stromeyer et al. (2015) criticism of the cross-loadings has nothing to do with the BSEM method per se but concerns cross-loadings in general. The authors do not have a problem with how cross-loadings are found and estimated with BSEM; rather, they argue against the general use of cross-loadings, thereby dismissing several other well-accepted modeling techniques, such as EFA and exploratory structural equation modeling (Asparouhov & Muthén, 2009). Stromeyer et al. argue that modeling cross-loadings is akin to "modeling noise." We disagree. Psychometric indicators are seldom perfectly pure construct indicators. Even completely reliable ratings of insomnia or physiological measures of sleep patterns are likely to present significant levels of true score (i.e., valid) associations with multiple constructs, such as burnout, depression, stress, drug abuse, and so on. Morin, Arens, and Marsh (in press) note,

> Remember that, according to the reflective logic of factor analyses, the factors are specified as influencing the indicators, rather than the reverse. Thus, small cross-loadings should be seen as reflecting the influence of the factor on the construct-relevant part of the indicators, rather than the indicators having an impact on the nature of the factor itself.

It follows that these small cross-loadings do not taint the constructs by adding "noise" but rather allow them to be estimated using all of the relevant information present at the indicator level.

Stromeyer et al. (2015) go on to ask why scholars would want to develop instruments that include cross-loadings. We do not argue that researchers should not aspire to develop indicators that are as close to perfect as possible. Although "pure" indicators of a single construct may exist, we surmise that such indicators remain at best a convenient fiction and that, in practice, most indicators will present both some level of random noise and also some level of construct-relevant association with other constructs (Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013; Marsh, Morin, Parker, & Kaur, 2014; Morin et al., in press; Sass & Schmitt, 2010; Schmitt & Sass, 2011). Relying on a measurement model that makes these associations explicit through cross-loadings is clearly a better way forward than relying on methods where these cross-loadings are swept under the rug. Thurstone (1947) proposed his "simple structure" principles to ensure conceptual clarity in the assessed constructs and as a set of rules to guide factor rotation, not as a set of guidelines to determine whether a factor solution was meaningful or not. In a typical BSEM cross-loading model, a limited number of midsize cross-loadings will be added to the CFA model, and all remaining cross-loadings will be near zero. Thus the BSEM cross-loading model does not compromise the conceptual clarity of the constructs.

Stromeyer et al. (2015) take issue with the argument that the exclusion of cross-loadings will result in "inflated" factor correlations. Recent statistical literature on measurement models including cross-loadings (Marsh et al., 2009, 2010, 2014; Morin et al., in press) show that factor correlations tend to be upwardly biased when true cross-loadings are constrained to be zero. This argument is based not on an opinion but on the results from simulation studies (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011) and studies based on simulated data (Marsh et al., 2013; Morin et al., in press). The advantage of such studies is that the true population model is known and thus could be used as an objective benchmark against which to compare the results from the estimated models. Thus, such studies are able to provide guidelines that are based on empirical facts rather than opinions, such as those expressed in Stromeyer et al. In addition, simulation studies can be used to show that adding small variance prior cross-loadings in a BSEM model does not lead to "artificially reduced common factor covariance," contrary to what Stromeyer et al. claim to be a logical conclusion.

What the above-mentioned studies showed was that even when small and substantively meaningless cross-loadings are present in the population model but ignored in CFA models, the factor correlations will tend to be substantially biased. Interestingly, these studies also show that when the population model meets the independent cluster assumptions inherent in CFA, relying on models allowing for the estimation of cross-loadings will nevertheless result in unbiased estimates of factor correlations notwithstanding the loss in parsimony associated with these models. Overall, these studies clearly show that the inclusion of cross-loadings is neither logically flawed nor logically incorrect but rather empirically supported by statistical research. Going back to the flawed argument that cross-loadings "taint" the nature of the constructs, these results rather show that it is the exclusion of these cross-loadings that modifies the meaning of the constructs.

The key to successful use of cross-loadings in a BSEM application is determining the small variance of the cross-loadings prior that will allow the nonzero cross-loadings to be estimated while keeping the rest near zero. This should be done as is described in the next section for the residual covariances, beginning with a very small prior variance that makes the BSEM and the CFA model identical and slowly increasing the variance until improvements in model fit diminish and/or quality of model identification diminishes as measured by the rate of convergence in the Bayesian iterations. Appendix A gives a summary of the steps recommended for BSEM with both cross-loadings and residual covariances.

Residual Covariances in BSEM

Stromeyer et al. (2015) point out that BSEM cross-loadings analysis is much easier to estimate and interpret than BSEM residual covariance analysis. We argue that this should not be the case and that both approaches use the same basic idea, which is to explicitly model some otherwise unmodeled source of influence on the indicators used within a specific measurement model. Whereas cross-loadings model meaningful associations between items and nontarget factors, residual covariances model shared sources of influence on the indicators that are unrelated to the factors, such as wording effects (e.g., negatively worded items or items with parallel wording), context (e.g., items referring to similar contexts, such as work vs. family), and so on. In more technical terms, both approaches involve adding to the model a set of potentially misspecified parameters with small priors around zero. These parameters are neither completely fixed to zero nor completely free but are instead approximately fixed to zero. In BSEM, we preserve the hypothesized SEM model while allowing the data to drive away from zero some of these additional parameters when evidence in the data exists. Despite the fact that the priors for the BSEM residual covariance analysis are more advanced than for the BSEM cross-loadings analysis due to its multivariate nature, simple guidelines are available and clarified below.

The BSEM method with residual covariances uses small informative priors for the residual covariance parameters in CFA to estimate a full residual variance-covariance matrix as part of the CFA model estimation. With this method, the residual covariance matrix is unconstrained by the model but is constrained by the prior. We consider a simple CFA model for a factor η and a loading matrix Λ:

$$Y = \nu + \Lambda\eta + \varepsilon, \tag{1}$$

where

$$\mathrm{Var}(\varepsilon) = \theta \tag{2}$$

is a diagonal residual covariance matrix. We consider situations where the CFA model does not fit the data well according to the posterior predictive p value (PPP; Muthén & Asparouhov, 2012) and illustrate what can be learned from a BSEM analysis with unconstrained residual covariances.

For the BSEM method, we set a prior for the θ matrix as the inverse Wishart prior

$$\theta \sim \mathrm{IW}\left(dD, d\right), \tag{3}$$

where d is the degrees of freedom of the distribution and D is the diagonal matrix equal to the CFA estimate of θ. This choice of prior is motivated as follows, drawing on Muthén and Asparouhov (2012: 335). The prior mean for θ is

$$\frac{d}{d-p-1}D, \tag{4}$$

where p is the number of observed variables. The prior mode for θ is

$$\frac{d}{d+p+1}D. \tag{5}$$

Thus for all off-diagonal elements of θ, the prior mean and the prior mode are both zero. For the diagonal elements of θ and sufficiently large d, both the prior mode and the prior mean will be close to D. As d increases, the prior variance for all parameters converges to zero; see formulas (A15) and (A17) in Muthén and Asparouhov (2012). Setting the d parameter to a large value is equivalent to analyzing the CFA model with residual variance parameters fixed to the CFA estimates D and residual covariance parameters fixed to zero. Thus the estimated BSEM model for large d would be equivalent to the CFA model, and it would yield PPP = 0 when the CFA model has PPP = 0. We conclude that the BSEM model would be rejected when d is sufficiently large.

Stromeyer et al. (2015) state that BSEM models with unconstrained residual covariances will yield outstanding model fit regardless of what model is specified. This statement is incorrect for three reasons. First, the authors ignore the basic asymptotic result that as d is sufficiently large, the PPP for the BSEM model will become the same as the PPP for the CFA model, that is, the BSEM model with large d will reject the model in those cases where the CFA model is rejected. Second, in their own data analysis, the authors fail to investigate prior sensitivity as recommended in Muthén and Asparouhov (2012). If these authors had explored inverse Wishart priors with different degrees of freedom, they would have inevitably obtained

BSEM models with large degrees-of-freedom parameter *d* and PPP = 0. Just as in the cross-loading BSEM analysis, very small prior variances for the cross-loadings are necessary as part of the model estimation; so is the BSEM residual covariance analysis with large degrees-of-freedom parameter *d*. The idea behind BSEM is simple. Attach a small-variance prior to all possible additions to the CFA model and let the data determine if such additions are necessary. The basic BSEM logic converts a fixed-to-zero parameter to an approximately fixed-to-zero parameter by specifying a small-variance prior. In that regard, Stromeyer et al. misrepresent the BSEM logic and the basic BSEM implementation. Third, the authors misrepresent the primary goal of estimating a BSEM model, which is not to confirm or disconfirm the CFA model but to evaluate the sources of the differences between the hypothesized CFA model and the data. The interpretation of that difference and the actual conclusions of the BSEM analysis are entirely up to the substantive researcher.

Gradually lowering the degrees-of-freedom parameter *d* in the BSEM model would yield a more flexible model where the residual covariances are no longer severely constrained to zero. For a sufficiently low degrees-of-freedom parameter, the BSEM model will essentially estimate a completely unconstrained variance-covariance matrix. Therefore, such a model would yield a high PPP value, and the model would not be rejected. This happens because the prior restrictions are minimal, and thus the BSEM model is sufficiently close to the unidentified and unrestricted model without the prior restrictions, not because the BSEM residual covariance approach is flawed.

In this section, we illustrate the proper usage of BSEM with several examples. For each example, a CFA model is estimated and is rejected based on PPP = 0. The CFA analysis is then followed up with a BSEM analysis. Here we assume that any cross-loading BSEM analysis has already been completed, and we focus only on the residual covariance. With a sufficiently low degrees-of-freedom parameter *d*, the PPP for the BSEM analysis is guaranteed to be greater than 0.05. We conduct BSEM sensitivity analysis by varying the degrees-of-freedom parameter *d* with the primary goal of determining the largest *d* that yields a PPP value greater than 0.05.

We consider this model to be the BSEM model of interest. By definition, this model is not rejected by the data and can be considered to be the BSEM model closest to the CFA model that fits well enough. This model resolves all of the CFA model misfits. At the same time, the larger the degrees-of-freedom parameter *d* is, the stricter the prior, and thus the CFA model modifications will be the smallest possible. This model is also better identified than models with lower *d* parameters and yields faster convergence. We summarize the BSEM sensitivity analysis in several steps.

Step 1. Select a starting value for *d*. Possible starting values are $d = 100$ (for $N \approx 500$), $d = 1{,}000$ (for $N \approx 5{,}000$), or $d = 10{,}000$ (for $N \approx 50{,}000$).

Step 2. Estimate the BSEM model with the initial or current *d* value with three possible outcomes.

- (a) Fast convergence (similar to CFA) and PPP > 0.05: You can use that BSEM model.
- (b) Slow or no convergence: Increase *d* and repeat Step 2 with new *d*.
- (c) Fast convergence but PPP < 0.05: Decrease *d* and repeat Step 2 with new *d*.

In the above process, the rate of increase or decrease of the *d* values can be selected in an ad hoc manner. For example, if a starting value of $d = 100$ is used, depending on the result, the next value of *d* can be 50 or 150. The above iterative process should not take more than five iterations, ideally. Note also that following the above process, we technically do not arrive at the maximum d values for which PPP > 0.05. We arrive at a d value that yields fast convergence, which can be interpreted as the model is sufficiently identified, and for which PPP > 0.05, which can be interpreted as the model fits sufficiently well as compared to the CFA model. These two characteristics of the BSEM model should be sufficient to make inference on the needed model modifications. If the inference is not clear, one possibility is to consider higher values of d that still yield PPP > 0.05. This will tend to reduce the values of the estimated residual covariance and perhaps make the inference easier.

Also note that in the construction of the inverse Wishart prior, it is important to use the exact D matrix from the Bayes CFA estimation when the degrees-of-freedom parameter d is large. If d is small, one can use an approximate and rounded D matrix because the prior is not strong and unlikely to yield misfit due to mismatch in the residual variances. For large d, however, if D is not set as in the CFA model, the PPP may reject the BSEM model due to a mismatch in the residual variance rather than the residual covariance.

The BSEM model can help us discover the places where the CFA model fails. Whether or not the CFA model is modified based on BSEM discoveries is a separate issue. This decision should depend on what the BSEM analysis determines and the possible substantive interpretations. Below, we illustrate five different possible outcomes from the BSEM analysis. In some situations, good model modifications can be discovered. In others, BSEM would suggest that the CFA misfit is due to small and unimportant residual correlations and the main model should stay unchanged.

The advantage of using BSEM is that the model modifications are based on the original CFA model and the BSEM analysis includes the original CFA loadings pattern. For example, evaluating the differences between the model-estimated and observed variance-covariance matrix does not provide that kind of embedding of the original model and would simply point out which observed covariances are not matched by the model rather than how the model should be modified.

Large Isolated Residual Covariances

In this section, we show that BSEM can pinpoint a few large residual correlations that are responsible for the misfit (such as when two items have strictly parallel wording). Such large correlations can be included and interpreted in a modified CFA model. Consider a simulated factor analysis example with six indicators and a single factor. All means are set to zero. All loadings and residual variances are set to 1. In the data generation, we include a residual correlation of 0.50 between the first two indicators. The sample size for this simulated example is 500. The CFA analysis without any residual correlations is rejected with PPP = 0. The BSEM model with d = 100 is not rejected and satisfies the goal of the BSEM sensitivity analysis. To set up the prior for the residual covariances in BSEM, we set D to be the identity matrix as all CFA residual variance estimates are close to 1. The estimate for the residual correlation between the first two variables is 0.30 and statistically significant. All remaining residual correlations are smaller than 0.11 and not significant. Thus the BSEM analysis suggests that a residual correlation is the main reason for the CFA model misfit. This correlation can be included in the CFA model to improve the model fit, the accuracy of the parameter estimates, and the accuracy of the factor score estimates.

Missing Factor

In this section, we show that BSEM can point to a missing factor in the CFA analysis (such as when one tries to estimate a single organizational commitment factor when in fact items measuring commitment to the organization and the workgroup are included). We generate a data set based on a two-factor model where each factor has three indicators. Parameters are set as in the previous section. Estimating a CFA model with one factor yields a model rejection. The BSEM model with $d = 250$ is not rejected and satisfies the goal of the BSEM sensitivity analysis. All residual correlations are estimated to be between 0.25 and 0.38 by absolute value, and all are statistically significant. Some correlations are positive and some are negative. Clearly this situation is easy to distinguish from the situation presented in the previous section. The fact that all correlations are misfitted should be interpreted as evidence that the one-factor model is an insufficient representation of the data, and the possibility for more factors should be explored.

Extra Factor: Two Factors Instead of One

In this section, we show that BSEM can suggest reducing the number of factors by combining two highly correlated factors into one (such as when one tries to estimate multiple job satisfaction factors when in fact all items tap into the same overarching construct). We simulate a factor analysis example with six indicators and a single factor. In the data generation, we also include a residual correlation of 0.7 between the first two indicators. The data are analyzed with a CFA model with two factors with three indicators each. The CFA model is rejected. The correlation between the two factors is estimated to be 0.8, which is high but not sufficiently so to consider combining them. The standard error for the factor correlation is 0.06. The corresponding BSEM analysis with $d = 50$ is not rejected. The correlation between the two factors is 0.98 with a standard error of 0.05. This is clearly enough evidence to consider combining the two factors into one. In addition, the BSEM analysis points to a residual correlation between the first two indicators. That residual correlation is estimated to be 0.41 and is statistically significant. All other residual correlation parameters are smaller than 0.19 and can be considered ignorable. Using this BSEM analysis, we can conclude that the two factors should be combined as one and that there is one isolated residual correlation.

Extra Factor: Unstable Factor That Can Be Replaced by Residual Correlations

In this section, we show that BSEM can suggest reducing the number of factors by eliminating unstable factors and replacing them with isolated residual correlations (such as when the presence of positively and negatively worded items suggests the existence of two factors when in fact a single construct is assessed but methodological artifacts need to be controlled for). We simulate a factor analysis example with six indicators but only the first three have loadings of 1 and the remaining three have loadings of 0. Two residual correlations are included in the data generation $\Theta_{34} = 0.4$ and $\Theta_{56} = 0.7$. The sample size is 200. We analyze the data with a two-factor CFA model. The first factor is measured by the first three indicators. The second factor is measured by Indicators 3 through 6. Thus the CFA model includes a cross-loading from the second factor to the third indicator. The CFA is rejected with PPP = 0. However, all main loadings are significant. The cross-loading is not significant. The BSEM model with $d = 50$ is not rejected. In the BSEM model, the structure for the first factor appears to be unchanged, but the structure for the second factor is changed completely. Now the loadings for the last two indicators are small and insignificant, while the cross-loading is significant. This indicates that the second factor is poorly defined and is constructed simply to account for some isolated residual correlations. The estimate for Θ_{56} is 0.55 and statistically significant, while all other covariances are not. The remaining structure of the second factor is now based only on one main loading (Indicator 4) and one cross-loading (Indicator 3), which also suggests that the factor is replaceable by the residual correlation Θ_{34}. The BSEM analysis identified a poorly defined factor and identified two residual correlations that can be added to the model instead of that factor, arriving again at the true model that generated the data.

Small Residual Correlations

In this section, we show that BSEM can suggest that the failure of the CFA model may be due simply to small residual correlations. If these correlations are small, one option is to retain the original CFA model and treat it as a good approximation to the data. Here, we generate data from a one-factor model with six indicators, adding the following residual correlations $\Theta_{12} = \Theta_{14} = 0.05$ and $\Theta_{13} = \Theta_{35} = -0.05$. The sample size is 500. The CFA one-factor model is rejected, while the BSEM model with $d = 100$ is not rejected. The BSEM residual correlations are all less than 0.1 and not statistically significant. These BSEM results are quite different from those in the previous sections and can lead us to the conclusion that no major model change is needed and the misfit in the CFA analysis is due to minor differences between the model and the data.

BSEM Residual Covariances Conclusion

The BSEM model points out which residual correlations are not accounted for by the CFA model, which factors have stable structures, and which do not. That information can be used for modifying the CFA model to obtain a better-fitting model. One key issue in considering the BSEM results is which covariances/correlations should be freed as candidates to be included in the CFA model. In making this decision, we should consider the statistical significance as computed by the BSEM model estimation as well as the substantive significance, meaning whether or not the estimated value is perceived to be sufficiently large to be of substantive importance.

Two of the possible outcomes are clear. If a correlation is substantively and statistically significant, it should be included in the CFA. If a correlation is substantively and statistically insignificant, it should not be included and should be fixed to zero. There is some ambivalence in the remaining two situations, and the decision is somewhat subjective. We provide some general guidance in that regard. If a correlation is substantively significant but not statistically significant, one can increase the degrees-of-freedom parameter d in the BSEM estimation. If that parameter is not sufficiently large, the estimated standard errors in the BSEM estimation may be artificially too large due to poor model identification. Thus, increasing d may align the substantive and statistical significance. On the other hand, if a parameter is substantively insignificant but statistically significant, we may choose to treat that parameter as approximately zero and interpret the CFA model as a sufficiently good approximation for the data. In addition, consider the fact that the estimation of the residual correlation parameters is tied up with the estimation of the factor model, that is, the estimates of the residual correlation parameters are not independent. If only one correlation is misfitted by the CFA model, the BSEM model may show more than one statistically significant correlation involving the same variables due to the fact that the model will attempt to compensate the misfit and thereby affect the remaining correlations related to those indicators. In such situations, focusing on the biggest correlations in the BSEM model can be used as a good strategy. For example, if two correlations are statistically significant within the same set of factor indicators but only one is substantively significant, we can choose to free only that one in the CFA model and hope that the resulting factor model adjustment will reduce the value of the remaining one. In this somewhat ad hoc process, the PPP value should be used to evaluate the CFA model fit.

The Use of DIC Versus BIC in BSEM Analysis

Stromeyer et al. (2015) discuss the use of DIC and BIC to choose among models in BSEM analysis. Because of the special nature of the BSEM parameters, we think DIC is more appropriate than BIC for comparing BSEM to CFA models. The model complexity penalty for DIC is based on the estimated number of parameters, while the penalty for BIC is based on the actual number of parameters. Thus BIC can unnecessarily penalize the BSEM model by counting small-variance prior parameters as actual parameters and thereby overshadowing information provided by BSEM. For this reason, we find that the recommendation given in Stromeyer et al. (2015: 513-514) to place more weight on BIC than DIC is misguided, and we illustrate this with a simulation study.

For both BSEM and CFA models, DIC is computed as follows. In each Markov chain Monte Carlo (MCMC) iteration, we compute the deviance for the current parameters:

$$D(\Theta) = -2\log[p(Y \mid \Theta)], \tag{6}$$

where $p(Y \mid \Theta)$ represents the likelihood of the observed data given the current parameters Θ. We then compute the effective number of parameters p_D:

$$p_D = \bar{D} - D\left(\bar{\Theta}\right), \tag{7}$$

where $\bar{D}$ represents the average deviance across all MCMC iterations and $\bar{\Theta}$ represents the average of the parameters across the MCMC iterations. Finally, DIC is computed as

$$\mathrm{DIC} = \bar{D} + p_D. \tag{8}$$

The computation of BIC in Mplus is performed as follows:

$$\mathrm{BIC} = D\left(\bar{\Theta}\right) + k\log\left(N\right), \tag{9}$$

where k is the number of model parameters and N is the sample size.[1]

Table 1
DIC Versus BIC Comparison for BSEM

Model	BIC	DIC
CFA	6267	6172
BSEM	6460	6144

Note: DIC = discrepancy information criterion; BIC = Bayesian information criterion; BSEM = Bayesian structural equation modeling; CFA = confirmatory factor analysis.

For BIC and DIC, a low value is preferred. This is obtained for high log-likelihood/low deviance and low model complexity penalty. The complexity penalty for DIC is the estimated number of parameters p_D, and for BIC, it is $k\log(N)$. For BIC, the model complexity penalty increases with sample size and uses the actual number of parameters k rather than the estimated number of parameters as in DIC.

Consider the following simulated example that illustrates the advantage of DIC over BIC for BSEM residual covariance analysis. We generate data according to a factor analysis model with one factor and 10 indicators. All factor loadings and residual variances are set to 1, and the means are set to 0. For the data generation model, we also include a residual correlation of 0.5 for two of the indicators. The sample size for this simulation example is 200. We analyze the data with the one-factor CFA model and the corresponding BSEM model with $d = 200$. Table 1 contains the results of this simulation. Using BIC for model selection yields the incorrect conclusion to prefer the CFA model with no residual correlations in the model. Using DIC for model selection yields the correct conclusion to prefer the BSEM model including one large residual correlation. In this example, PPP also agrees with DIC and the CFA model is rejected by PPP, while the BSEM model is not. In this example, it is interesting to note that the complexity penalty for the BIC uses $k = 75$ true parameters, while for the DIC, the estimated number of parameters is $p_D =$ 40. The DIC recognizes that most of the extra residual covariance parameters are nothing more than parameters nearly fixed to zero, and it is not penalizing for those parameters.

The BIC is asymptotically guaranteed to select the correct model (Schwarz, 1978). In the above example, it is key that the sample size is small ($n = 200$). In this case, the deviance gains of the BSEM analysis are overpowered by the inflated penalty of BIC, and that leads to the incorrect conclusion when using BIC. If the sample size is larger, both BIC and DIC would pick the BSEM model because the deviance component of the information criteria becomes dominant. Note, however, that for any sample size, we can construct a model with a sufficiently large number of indicators where the BIC will fail. This is because the inflation of the penalty in BIC will be larger for examples with more indicators, and thus one can expect such problems to occur even with larger sample sizes. Thus we recommend using DIC for comparing BSEM models with other BSEM or CFA models. Comparing CFA models estimated by Bayes, BIC may still be preferable.

Application to Stromeyer et al. Example

This section applies the BSEM method with residual covariances to the five-factor model for the 19 items analyzed by Stromeyer et al. (2015). The resulting BSEM model is compared to the BSEM model with cross-loadings of Stromeyer et al. The two CFA models suggested by the BSEM models are compared as well. Data from both of the randomly drawn Samples 1 and 2 are considered (n = 500 in both samples). Sample 1 data are analyzed first, using Sample 2 for comparison and cross-validation.

As a first step, the IW(dD, d) residual covariance prior was chosen with D as the residual variances of the Bayes five-factor CFA model shown in Table 2 of Stromeyer et al. (2015) and d set to 100. This BSEM model maintains the cross-loadings used in Stromeyer et al.'s Table 3, choosing prior variances of 0.01 in line with Muthén and Asparouhov (2012). (Stromeyer et al. used prior variances of 0.02.) Because this BSEM analysis gives a PPP much greater than 0.05, the next step increases d to 200, resulting in PPP = 0.129. Given that this PPP value is reasonably close to the stipulated threshold of 0.05, no further increase of d is made. An interesting finding from this BSEM analysis is that the two factors Plan and Marshal are as highly correlated as 0.95 (the Sample 2 correlation is 0.93). This suggests combining the Plan and Marshal factors, resulting in a four-factor model for which the same BSEM priors are applied.[2]

Table 2 shows the factor loading and factor correlation estimates from the four-factor BSEM model using Sample 1 data. It is seen that two cross-loadings are significant and larger than 0.2, the indicator S3 loading on Plan-Marshal and the indicator P3 loading on Implement Finance. Only two residual correlations (not shown) are significant and greater than 0.2, S3 with P1 and P2 with P3. The Mplus input for this BSEM analysis is shown in Appendix B.

The top of Table 3 shows fit statistics for the Stromeyer et al. (2015) BSEM model (labeled *SMSD five-factor*, using the last name initials of the four authors) and our five-factor and four-factor models (labeled *AMM five-factor/four-factor*). These models are compared using DIC as proposed in the previous section. Table 3 shows that the AMM five-factor model has a better DIC value than the SMSD five-factor model. This advantage also holds for Sample 2. The AMM four-factor model has a slightly higher (worse) DIC value than the AMM five-factor model but is preferable due to avoiding the high factor correlation. The AMM four-factor model maintains a PPP value greater than 0.05.

The bottom part of Table 3 shows fit results for the CFA models suggested by the BSEM analyses, referred to as BSEM-based CFA. The SMSD five-factor CFA model is that of Table 4 in Stromeyer et al. (2015), where three cross-loadings have been added. The AMM four-factor CFA model adds the two cross-loadings and two residual covariances discussed in connection with Table 2. Note that for these models, only diffuse priors are used; that is, BSEM-type informative priors are not used. Because of this, the comparison between models is carried out using BIC, a more common criterion for CFA models. BIC favors the AMM four-factor model for both Sample 1 and Sample 2. The same conclusion is reached when using maximum-likelihood estimation.

Table 7 of Stromeyer et al. (2015) shows the results of cross-validation comparing model estimates for Sample 1 and Sample 2. The strictest invariance across the samples is the "strong" case (scalar invariance) where factor indicator intercepts and loadings are held equal across samples while the factor means, factor variances, factor covariances, and indicator residual variances are allowed to vary. Here we also add the more relevant full invariance case where all parameters are held equal across samples. BIC is again considered given that BSEM analysis is not used. The BIC values for the SMSD five-factor model are for the *strong* case and the full invariance case 41453 and 41275, respectively. These BIC values are higher (worse) than those of the corresponding AMM four-factor values, 41355 and 41189, respectively.

Table 2
Estimates From Four-Factor BSEM With Cross-Loading Prior Variance 0.01 and Residual Covariance Priors With d = 200 (Sample 1)

Indicator	Search	Plan-Marshal	Implement People	Implement Finance
S1	**.93†**	−.04	−.04	−.02
S2	**.98†**	−.07	.02	−.00
S3	**.42†**	.30	.07	.02
P1	.37	**.45†**	.01	.00
P2	.02	**.64†**	−.05	.15
P3	−.05	**.65†**	−.09	.24
P4	−.06	**.83†**	−.09	−.02
M1	−.02	**.89†**	.02	−.12
M2	−.08	**.75†**	.09	−.09
M3	−.02	**.70†**	.11	−.05
IP1	−.04	−.07	**.95†**	.01
IP2	−.09	.06	**.85†**	.02
IP3	−.06	−.06	**.93†**	.00
IP4	.11	.01	**.70†**	.05
IP5	.05	.06	**.76†**	−.06
IP6	.03	−.05	**.76†**	−.01
IF1	.01	−.03	.05	**.91†**
IF2	.02	−.02	−.02	**.99†**
IF3	−.04	.04	.03	**.79†**
Latent variable correlation matrix				
Search	—			
Plan-marshal	.65 †	—		
Implement people	.46 †	.64†	—	
Implement finance	.29 †	.55†	.44†	—

Note: Factor loadings in bold were freely estimated using diffuse priors. Daggers indicate 95% credibility interval does not contain zero. This measurement scale is developed by McGee, Peterson, Mueller, and Sequeira (2009).

In conclusion, our four-factor model performs better than the proposed five-factor model of Stromeyer et al. (2015) in both its BSEM version and its CFA version, and it also cross-validates better. Our four-factor model would not be entertained unless a residual covariance BSEM model is analyzed. The high correlation between the Plan and Marshal factors is not seen in the cross-loading BSEM analysis of Stromeyer et al. An interesting aside is that the high factor correlation is also not seen in EFA. EFA with five factors shows the CFA loading pattern that is hypothesized, and no high factor correlations are obtained. EFA is, however, limited in that it does not allow residual covariances. Both the five- and four-factor EFA BIC values are worse than those of our four-factor model. All in all, this implies that BSEM analysis with residual covariances is uniquely positioned to uncover a model alternative that would otherwise be overlooked. It enables "thinking outside the box" of regular EFA and CFA as well as cross-loading BSEM. Our four-factor model alternative performs better in statistical terms than the five-factor model preferred by Stromeyer et al. Whether or not this four-factor alternative is palatable from a subject matter point of view is a different matter, but it is a useful discovery nevertheless. If the researcher's theories strongly suggest a five-factor

Table 3
BSEM and BSEM-Based CFA Fit Statistics

Model	Sample	PPP	No. Par.	No. Est. Par.	DIC	BIC
BSEM						
SMSD five-factor cross-loading BSEM	1	0.00	143	113	20371	21032
	2	0.00	143	113	20346	21006
AMM five-factor cross-loading, residual covariance BSEM	1	0.13	314	156	20157	21792
	2	0.15	314	160	20179	21806
AMM four-factor cross-loading, residual covariance BSEM	1	0.12	291	161	20163	21648
	2	0.09	291	161	20185	21671
BSEM-based CFA						
SMSD five-factor	1	0.00	70	71	20449	20742
	2	0.00	70	71	20536	20828
AMM four-factor	1	0.00	68	67	20424	20712
	2	0.00	68	67	20492	20779

Note: BSEM = Bayesian structural equation modeling; CFA = confirmatory factor analysis; SMSD = BSEM model estimated by Stromeyer, Miller, Sriramachandramurthy, and DeMartino (2015; labeled using the last name initials of the authors); AMM = BSEM model estimated in this study (labeled using the last name initials of the authors); PPP = posterior predictive *p* value; No. Par. = number of free parameters; No. Est. Par. = estimated number of parameters; BIC = Bayesian information criterion; DIC = discrepancy information criterion.

model, efforts should be made to investigate other item formulations in the instrument development so that the four-factor alternative is not competitive.

Conclusion

This paper has shown that the BSEM technique is a valuable tool for a thorough factor analysis of a measurement instrument. The technique provides unique insights not obtained by other types of factor analysis. The paper counters the criticism leveled against BSEM by Stromeyer et al. (2015). A reanalysis of the example used by the authors uses BSEM with residual covariances to reveal a superior model overlooked by the authors. We hope that the guidelines we have presented will stimulate more use of BSEM.

Appendix A

Bayesian Structural Equation Modeling (BSEM) Analysis Steps

This appendix summarizes BSEM analysis steps (for details, see Muthén & Asparouhov, 2012).

1. Choose a uniform metric for the variables, for example, by standardizing the variables when the model is scale free (no parameter equalities across variables), so that scale issues do not interfere with prior settings (see Muthén & Asparouhov, 2012).
2. Specify cross-loadings with a prior variance that is small enough that the Mplus output 95% confidence interval for the difference between the observed and the replicated chi-square

values is the same as for the original confirmatory factor analysis (CFA) model without cross-loadings (with variable on a scale with variance 1, a prior variance of 0.001 may be suitable).

3. Increase the cross-loadings prior variance by a factor of 10 and then in smaller steps, such as 5. Multiple runs should be done, at least five. Then monitor:
   a. Speed of convergence (number of iterations); a too-large prior variance leads to nonconvergence due to nonidentification.
   b. Ninety-five percent confidence interval for the difference between the observed and the replicated chi-square values.
4. From the runs in Step 3, make a subjective selection based on Criteria a and b, where for
   a. the number of iterations to convergence will increase as prior variance increases.
   b. the confidence interval limits will decrease as prior variance increases, with diminishing returns when stabilizing.

Subjectively select a prior variance for the cross-loadings that gets the majority of the decrease in Criterion b without sacrificing much on Criterion a.

5. Compute *D* of the residual covariance inverse Wishart prior using the residual variance estimates from the CFA model or the cross-loading model.
6. Conduct a sensitivity analysis described in the manuscript in the Residual Covariances section.
7. Conclude on the model: Possible model modifications and decision on a BSEM-based CFA model versus the BSEM model as the final model.

Appendix B

Mplus Bayesian Structural Equation Modeling (BSEM) Input

This appendix gives the Mplus input for the BSEM model with cross-loadings and residual covariances presented in Table 2.

TITLE:	BSEM with cross loadings and residual covariances;
DATA:	FILE = ESE_Ordered_Text.txt;
VARIABLE:	NAMES = rank group rand id s1 s2 s3 p1 p2 p3 p4
	m1 m2 m3 ip1 ip2 ip3 ip4 ip5 ip6 if1 if2 if3;
	USEVARIABLES = s1 s2 s3 p1 p2 p3 p4
	m1 m2 m3 ip1 ip2 ip3 ip4 ip5 ip6 if1 if2 if3;
	USEOBSERVATION = (group EQ 1);
DEFINE:	STANDARDIZE s1 s2 s3 p1 p2 p3 p4
	m1 m2 m3 ip1 ip2 ip3 ip4 ip5 ip6 if1 if2 if3;
ANALYSIS:	ESTIMATOR = BAYES;
	!Activates the Bayesian estimator
	CHAINS = 2;
	!Specifies using two Markov Chains for conducting the analysis
	PROCESSORS = 2;
	!For use in multi-core systems, assigns two processors with one per chain
	FBITERATIONS = 15000;
	!Specifying that the minimum number of Markov Chain iterations is 15000

(continued)

Appendix B (continued)

MODEL:	search BY s1* s2 s3
	p1 p2 p3 p4 m1 m2 m3 ip1 ip2 ip3 ip4 ip5 ip6 if1 if2 if3 (xload1-xload16);
	plan BY p1* p2 p3 p4 m1 m2 m3
	s1 s2 s3 ip1 ip2 ip3 ip4 ip5 ip6 if1 if2 if3 (xload17-xload28);
	imp_ppl BY ip1* ip2 ip3 ip4 ip5 ip6
	s1 s2 s3 p1 p2 p3 p4 m1 m2 m3 if1 if2 if3 (xload48-xload60);
	imp_fn BY if1* if2 if3
	s1 s2 s3 p1 p2 p3 p4 m1 m2 m3 ip1 ip2 ip3 ip4 ip5 ip6 (xload61-xload76);
	search WITH plan imp_ppl imp_fn;
	plan WITH imp_ppl imp_fn;
	imp_ppl WITH imp_fn;
	serch@1;
	plan@1;
	imp_ppl@1;
	imp_fn@1;
	s1-if3 (pload1-pload19);
	s1-if3 WITH s1-if3 (cload1-cload171);
MODEL PRIORS:	xload1-xload76~N(0,0.01);
	pload1~ IW(43.2,200);
	pload2~IW(22.8,200);
	pload3~IW(114.8,200);
	pload4~IW(98.4,200);
	pload5~IW(78.6,200);
	pload6~IW(85.0,200);
	pload7~IW(101.2,200);
	pload8~IW(56.0,200);
	pload9~IW(92,200);
	pload10~IW(79.2,200);
	pload11~IW(45.6,200);
	pload12~IW(57,200);
	pload13~IW(56,200);
	pload14~IW(80,200);
	pload15~IW(76,200);
	pload16~IW(92.6,200);
	pload17~IW(33.6,200);
	pload18~IW(14.8,200);
	pload19~IW(67.4,200);
	cload1-cload171~IW(0,200);
OUTPUT:	STAND(STDYX);
	RESIDUAL TECH1 TECH8 STDY;
PLOT:	TYPE = PLOT2;

Notes

1. In principle, in the computation of Bayesian information criterion (BIC), the deviance should be evaluated at the maximum-likelihood parameter estimates rather than the posterior mean; however, asymptotically the two are the same and we use the above computation as an approximation for the actual BIC value. The exact BIC value cannot be estimated naturally within the Markov chain Monte Carlo estimation.

2. An interesting aside is the relationship between prior choices and size of factor correlations in the five-factor BSEM analysis. Holding the inverse Wishart prior constant and varying the prior variances for the cross-loadings

as 0.005, 0.01, and 0.02, the Plan-Marshal factor correlation is 0.99, 0.95, and 0.78 (Sample 2 values are 0.99, 0.93, and 0.81). A similar degree of variation in factor correlations does not appear when using only cross-loadings and not residual covariances. A conjecture is that with stricter cross-loading priors, the use of residual covariances gives decreased cross-loadings, and that increases the factor correlation because more indicator correlations have to go through the factors.

References

Asparouhov, T., & Muthén, B. 2009. Exploratory structural equation modeling. *Structural Equation Modeling*, 16: 397-438.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. 2010. A new look at the Big-Five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22: 471-491.

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. 2013. Why item parcels are (almost) never appropriate: Two wrongs do not make a right. Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, 18: 257-284.

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. 2014. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analyses. *Annual Review of Clinical Psychology*, 10: 85-110.

Marsh, H. W., Muthén, B., Asparouhov, A., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. 2009. Exploratory structural equation modeling, integrating CFA and EFA: Students' evaluations of university teaching. *Structural Equation Modeling*, 16: 439-476.

McGee, J. E., Peterson, M., Mueller, S. L., & Sequeira, J. M. 2009. Entrepreneurial self-efficacy: Refining the measure. *Entrepreneurship Theory and Practice*, 33: 965-988.

Morin, A. J. S., Arens, A. K., & Marsh, H. W. in press. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*.

Muthén, B., & Asparouhov, T. 2012. Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, 17: 313-335.

Sass, D. A., & Schmitt, T. A. 2010. A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, 45: 73-103.

Schmitt, T. A., & Sass, D. A. 2011. Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational & Psychological Measurement*, 71: 95-113.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*, 6: 461-464.

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. 2015. The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*, 41: 491-520.

Thurstone, L. L. 1947. *Multiple factor analysis*. Chicago: University of Chicago.

Zyphur, M. J., & Oswald, F. L. 2015. Bayesian estimation and inference: A user's guide. *Journal of Management*, 41: 390-420.