



Bayesian Model Averaging over Directed Acyclic Graphs With Implications for Prediction in Structural Equation Modeling

David Kaplan

Department of Educational Psychology



April 13th, 2015

University of Nebraska-Lincoln

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- This work was conducted in collaboration with Danielle Lee.
- The research reported in this paper was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305D110001 to The University of Wisconsin - Madison. The opinions expressed are those of the authors and do not necessarily represent views of the Institute or the U.S. Department of Education.
- I would like to thank Fabrizio Sanchez for early contributions to the R code used in this project.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The distinctive feature that separates Bayesian statistical inference from its frequentist counterpart is its focus on describing and modeling all forms of uncertainty.
- The primary focus of uncertainty within a Bayesian modeling exercise concerns prior knowledge about model parameters.
- In the Bayesian framework, all unknowns are assumed to be random, described by probability distributions.
- Parameters constitute the central focus of statistical modeling, and because they are, by definition, unknown, Bayesian inference encodes background knowledge about parameters in the form of the prior distribution.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- Within the Bayesian framework, parameters are not the only unknown elements.
- The Bayesian framework recognizes that models themselves possess uncertainty insofar as a particular model is typically chosen based on prior knowledge of the problem at hand, and the variables that have been used in previously specified models.
- This form of uncertainty often goes unnoticed.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- Quoting Hoeting et al. (1999)

“Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions that are more risky than one thinks they are.”(pg. 382)

- An internally consistent Bayesian framework for structural equation modeling and estimation must also account for model uncertainty.
- The current approach to addressing the problem of uncertainty lies in the method of *Bayesian model averaging* (BMA).



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- This talk examines Bayesian model averaging as a means of improving the predictive performance of structural equation models.
- We expand the framework of Madigan and his colleagues by considering a structural equation model as a special case of a directed acyclic graph.
- We then provide an algorithm that searches the model space for sub-models that satisfy the conditions of Occam's razor and Occam's window and obtains a weighted average of the sub-models using posterior model probabilities as weights.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The organization of this talk is as follows.
 - 1 Brief overview Bayesian SEM.
 - 2 Outline the method of Bayesian model averaging and Occam's window following closely the work of Madigan and his colleagues (e.g. Madigan and Raftery 1994, Raftery et al., 1997).
 - 3 Description of our algorithm for searching the space of possible sub-models of a general structural equation model.
 - 4 Description of simulation study and case study
 - 5 Results and conclusion



Bayesian SEM

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- A structural equation model among observed variables can be specified as follows. Let

$$\mathbf{y} = \boldsymbol{\alpha} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta} \quad (1)$$

- It is notationally convenient to arrange the model parameters into sets with common conjugate distributions as .

$$\boldsymbol{\theta}_{norm} = \{\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Gamma}\} \quad (2)$$

$$\boldsymbol{\theta}_{IW} = \{\boldsymbol{\Phi}, \boldsymbol{\Psi}\} \quad (3)$$



- Formally, we write

$$\boldsymbol{\theta}_{norm} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega}), \quad (4)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are the mean and variance hyperparameters, respectively, of the normal prior.

- For blocks of variances and covariances in Ξ and Ψ , we assume that the prior distribution is inverse-Wishart, i.e.

$$\boldsymbol{\theta}_{IW} \sim IW(\mathbf{R}, \delta), \quad (5)$$

where \mathbf{R} is a positive definite scale matrix, and $\delta > q - 1$, where q is the number of observed variables.

- Different choices for \mathbf{R} and δ will yield different degrees of “informativeness” for the inverse-Wishart distribution.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- Once priors are elicited for all model parameters, MCMC is used to obtain estimates of their posterior distributions.
- Commercial software for conducting Bayesian structural equation modeling include AMOS and Mplus.
- Open source programs such as JAGS, rjags, and OpenBUGS can be used to specify Bayesian SEMs.
- Most software default to non-informative (uniform or highly diffused) priors.



Bayesian Model Building

- The frequentist and Bayesian goals of model building are the same.
- First, a researcher will specify an initial model relying on a lesser or greater degree of prior theoretical knowledge.
- Second, this model will be fit to data obtained from a sample from some relevant population.
- Third, an evaluation of the quality of the model will be undertaken, examining where the model might deviate from the data, as well as assessing any possible model violations. At this point, model respecification may come into play.
- Finally, depending on the goals of the research, the “best model” will be chosen for some purpose.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- Despite these similarities there are important differences - particularly in the model specification stage.
- For Bayesians, the first phase of modeling building requires the specification of a full probability model for the data and the parameters of the model, where the latter requires the specification of the prior distribution.
- The notion of model fit, therefore, implies that the full probability model fits the data.
- Lack of model fit may well be due to incorrect specification of likelihood, the prior distribution, or both.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- Another difference between the Bayesian and frequentist goals of model building relates to the justification for choosing a particular model among a set of competing models.
- Model building and model choice in the frequentist domain is based primarily on choosing the model that best fits the data.
- This has certainly been the key motivation for model building, respecification, and model choice in the context of structural equation modeling (Kaplan 2009).
- In the Bayesian domain, the choice among a set of competing models is based on which model provides the best posterior predictions.
- That is, the choice among a set of competing models should be based on which model will best predict what actually happened.



Bayesian Model Averaging

- Consider a quantity of interest such as a future observation.
- Following Madigan and Raftery (1994), we will denote this quantity as Υ .
- Next, consider a set of competing models M_k , $k = 1, 2, \dots, K$ that are not necessarily nested.
- The posterior distribution of Υ given data y can be written as

$$p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y). \quad (6)$$

where $p(M_k|y)$ is the posterior probability of model M_k written as

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{l=1}^K p(y|M_l)p(M_l)}, \quad l \neq k. \quad (7)$$

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The interesting feature of equation (6) is that $p(M_k|y)$ will likely be different for different models.
- The term $p(y|M_k)$ can be expressed as an integrated likelihood

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k, \quad (8)$$

where $p(\theta_k|M_k)$ is the prior distribution of θ_k under model M_k (Raftery et al., 1997).

- Thus, BMA provides an approach for combining models specified by researchers.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The advantage of BMA has been discussed in Madigan and Raftery (1994), who showed that BMA provides better predictive performance than that of a single model based on a log-score rule.
- We show that a Bayesian model averaged structural model (*BMA-SEM*) also provides better prediction of the endogenous variable of interest than any single model, including the initially specified model.
- We show this through predictive coverage as well as the log-score rule.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- BMA is difficult to implement.
 - 1 The number of terms in $p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y)$ can be quite large and the corresponding integrals are hard to compute.
 - 2 Eliciting $p(M_k)$ may not be straightforward. The uniform prior $1/M$ is often used.
 - 3 Choosing the class of models to average over is also challenging.
- The problem of reducing the overall number of models that one could incorporate in the summation has led to a solution based on the notion of *Occam's window* (Madigan and Raftery, 1994).



Occam's Window

- To motivate the idea behind Occam's window, consider the problem of finding the best subset of predictors in a linear regression model.
- Following Raftery et al., (1997) we consider an initially large number of predictors, but perhaps the goal is to find a subset that provides accurate predictions.
- As noted in the earlier quote by Hoeting et al. (1999), the concern in drawing inferences from a single "best" model is that the choice of a single set of predictors ignores uncertainty in model selection.
- Occam's window (Madigan and Raftery, 1994) provides such an approach for Bayesian model averaging by reducing the subset of models under consideration.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



- The Occam's window algorithm proceeds in two steps (Raftery et al., 1997).
- In the first step, models are eliminated if they predict the data less well than the model that provides the best predictions.
- Consider a set of models M_k , $k = 1 \dots K$, and a cutoff value C chosen in advance by the analyst.

$$A' = \left\{ M_k : \frac{\max_l \{p(M_l|y)\}}{p(M_k|y)} \leq C \right\}. \quad (9)$$

- Equation (9) compares the model with the largest posterior model probability, $\max_l \{p(M_l|y)\}$, to a given model $p(M_k|y)$.
- If the ratio in equation (9) is less than or equal to a chosen value C , then it is discarded from the set of models to be included in the model averaging.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- In the second step, models are discarded from consideration if they receive less support from the data than simpler sub-models.

$$B = \left\{ M_k : \exists M_l \in A', M_l \subset M_k, \frac{p(M_l|y)}{p(M_k|y)} > 1 \right\}. \quad (10)$$

- Equation (10) states that there exists a model M_l within the set A' and where M_l is simpler than M_k .
- If the simpler model receives more support from the data than the more complex model, then it is included in the set B .
- Notice that the second step corresponds to the principal of Occam's razor (Madigan and Raftery, 1994)



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

Occam's
Window

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- With step 1 and step 2, the problem of Bayesian model averaging is simplified by replacing equation (6) with

$$p(\Upsilon|y, \mathcal{A}) = \sum_{M_k \in \mathcal{A}} p(\Upsilon|M_k, y)p(M_k|y, \mathcal{A}), \quad (11)$$

where \mathcal{A} is the relative complement of A' and B . That is, the models under consideration for Bayesian model averaging are those that are in A' but not in B .



The BMA-SEM Method

- The general steps of our BMA-SEM method are as follows
 - 1 Specify an initial model of interest recognizing that this may not be the model that generated the data.
 - 2 Implement the up and down algorithm to reduce the space of models to a reasonable size,
 - 3 Obtain the posterior model probabilities for each model,
 - 4 Obtain the weighted average of structural parameters over each model, weighted by the posterior model probabilities.
 - 5 Compare predictive performance of the BMA-SEM to the initially specified SEM by computing the reduced form of the models and calculating the log-score or the predictive coverage.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Up and Down
Algorithm

Model
Averaging

Predictive
Performance

Simulation
Study

Case Study

Conclusion



Up and Down Algorithm

- We implement Occam's window by considering path diagrams as DAGs and use an approach by Madigan and Raftery (1994) we refer to as the *Up and Down Algorithm*.
 - 1 A sub-model or super-model of the initial model is obtained by dropping (adding) an edge (path) from the DAG.
 - 2 Occam's window is employed to select the sub-models which predict as well as the best model.
 - 3 The difference between the BICs of two models is used to compute the log posterior odds with non-informative priors on the models.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Up and Down
Algorithm

Model
Averaging
Predictive
Performance

Simulation
Study

Case Study

Conclusion



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Up and Down
Algorithm

Model
Averaging
Predictive
Performance

Simulation
Study

Case Study

Conclusion

24 / 41

- 4 For each comparison, if the difference in BICs between a model and its sub-model is larger than $\log(C)$, then the sub-model is preferred.
- 5 If the difference between the BICs is smaller than $-\log(C)$, then the model is preferred and all sub-models discarded because all remaining sub-models will have a smaller posterior model probabilities based on monotone test statistics (Gabriel, 1969).
- 6 If the difference of BICs is between $-\log(C)$ and $\log(C)$, then both models are preferred.
- 7 With this subset of models, we again calculate the difference between each BIC and the minimum BIC in the subset and selected sub models as an accepted set of models for BMA if the corresponding difference of BICs was less than $2 \times \log(C)$.



Model Averaging

- A set of J ($j = 1, 2, \dots, J$) possible structural equation models in \mathcal{A} is chosen through the algorithm.
- For each model in this set, we calculate posterior model probabilities as

$$p(M_j|y) = \frac{\exp(-.5 \times \Delta BIC)}{\sum_{l=1}^J \exp(-.5 \times \Delta BIC)}, \quad l \neq j \quad (12)$$

where ΔBIC is the difference between the BIC of M_j and the maximum of the BICs of all the models in the set.

- The posterior model probabilities are then used as weights when obtaining the posterior means of parameters across all the models in the set.

$$E(\theta_i|y, M_j) = \sum_{M_j \in \mathcal{A}} \hat{\theta}_i p(M_j|y). \quad (13)$$



Predictive Performance

- To compare the predictive performance based on a BMA-SEM to the predictive performance based on the initially specified structural equation model it is convenient to transform the structural form of the model to its reduced form

$$(\mathbf{I} - \mathbf{B})\mathbf{y} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}. \quad (14)$$

If $(\mathbf{I} - \mathbf{B})$ is non-singular then equation can be written as

$$\mathbf{y} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\zeta} \quad (15)$$

$$= \boldsymbol{\Pi}_0 + \boldsymbol{\Pi}_1\mathbf{x} + \boldsymbol{\zeta}^*. \quad (16)$$

where $\boldsymbol{\Pi}_0$ is the vector of reduced form intercepts, $\boldsymbol{\Pi}_1$ the vector of reduced form slopes, and $\boldsymbol{\zeta}^*$ the vector of reduced form disturbances with variance matrix, $\boldsymbol{\Psi}^*$.



- To gauge predictive performance we follow Hoeting, et al., (1999).
 - 1 Randomly divide the data set into “model averaging” data and “prediction” data.
 - 2 Fit a single Bayesian SEM and BMA-SEM to the model averaging data.
 - 3 Convert the structural form of each model to their reduced forms.
 - 4 Predict the final dependent variable in the reduced form for the prediction data with the result of the reduced form of the SEM and BMA-SEM from the model averaging data.
 - 5 Compare their predictive performance based on a 90 percent predictive coverage interval.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Up and Down
Algorithm
Model
Averaging

Predictive
Performance

Simulation
Study

Case Study

Conclusion

28 / 41

- We can also convert the predictive coverage percentage into a *log score*.
- The theory of BMA states that the model averaged coefficients will attain better prediction than any other model based on the log-score rule.
- Scoring rules yield a measure of the accuracy of probabilistic predictions.
- The goal is to maximize the log-score.
- The log of the predictive performance is the negative of *surprisal*, a measure often used in Bayesian inference.



Simulation Study

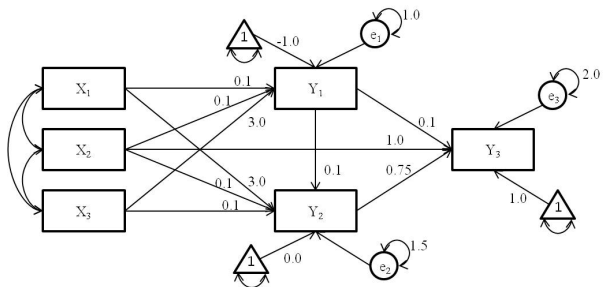


Figure 1: Initial path model.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- For comparison purposes, a single Bayesian SEM (BSEM) based on the original model was fit to the model averaging data.
- We used conjugate but highly diffused priors for all parameters.
- Posterior distributions of the parameters were obtained using the “rjags” program in R (Plummer, 2015).



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The "coda" package (Plummer, 2015) was also used for post-processing of diagnostics and summaries.
- The algorithm was set to produce 5,000 burn-in iterations, with 495,000 post-burn-in draws and a thinning interval of 50, from two chains starting at different locations of the posterior distribution.
- Inspection of diagnostics showed that all parameters stabilized to their respective posterior distributions.



Results of simulation study

Table 1: Selected models by BMA for the simulation data

Parameter	M_1	M_2	M_3
$y_1 \sim x_1$	•	•	•
$y_2 \sim x_1$	•	•	•
$y_3 \sim x_1$			
$y_1 \sim x_2$	•	•	•
$y_2 \sim x_2$	•	•	•
$y_3 \sim x_2$	•	•	•
$y_1 \sim x_3$	•	•	•
$y_2 \sim x_3$	•	•	
$y_3 \sim x_3$			
$y_2 \sim y_1$		•	•
$y_3 \sim y_1$	•	•	•
$y_3 \sim y_2$	•	•	•
BIC	45229.95	45233.29	45235.78
PMP	0.804	0.152	0.044

Note. PMP; posterior model probability.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Table 2: Comparison of the result of BMA to the result of the BSEM for the simulation data

Parameter	BMA			EAP	BSEM		
	mean(βy)	SD(βy)	P(βy)%		SD	95%	PPI
y1~x1	0.101	0.020	100.0	0.100	0.020	0.061	0.139
y2~x1	2.987	0.025	100.0	2.983	0.025	2.934	3.032
y3~x1	0.000	0.000	0.0	-			
y1~x2	0.082	0.020	100.0	0.082	0.020	0.043	0.122
y2~x2	0.121	0.025	100.0	0.118	0.025	0.101	0.167
y3~x2	1.037	0.029	100.0	1.037	0.029	0.980	1.096
y1~x3	2.983	0.020	100.0	2.983	0.020	2.944	3.022
y2~x3	0.363	0.103	95.6	0.249	0.077	0.094	0.401
y3~x3	0.000	0.000	0.0	-			
y2~y1	0.014	0.032	19.6	0.052	0.025	0.003	0.101
y3~y1	0.111	0.009	100.0	0.111	0.009	0.093	0.129
y3~y2	0.748	0.009	100.0	0.748	0.009	0.731	0.765

Note. EAP; expected a posteriori, SD; posterior standard deviation, PPI; posterior probability interval.



Table 3: Comparison of the predictive performance for simulation study

Method	Simulation %	(N=2,500) Log score
BMA	71.00	-.3425
Best model from BMA	70.44	-.3504
Initial Model (BSEM)	61.08	-.4930

Note. %, percent of predictive coverage, Log score; log score of predictive coverage.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Case Study

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- For the case study, we use data from PISA 2009 to estimate a model relating reading proficiency to a set of background and reading strategy variables.
- The sample was collected from PISA-eligible students in the United States, and the sample size was 5,053.
- The sample was split into a model averaging set ($N = 2,526$) and a predictive testing set ($N = 2,527$).



- Exogenous variables are Gender (male=0, female=1), immigrant status (Immigr), and a measure of the economic, social, and cultural status of the student (ESCS).
- Mediating variables are three measures of student reading strategies: memorization strategies (MEMOR), elaboration strategies (ELAB), and control strategies (CSTRAT).
- The first plausible value of the reading assessment was used as the final outcome variable.
- Also for comparison purposes, a BSEM was specified for the PISA example with diffused conjugate priors on all model parameters.



Case Study

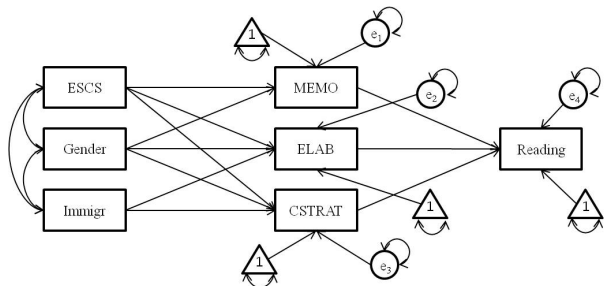


Figure 2: Initial path model.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Results of case study

Table 4: Selected models by BMA for the PISA data

Parameter	M_1	M_2	M_3	M_4
MEMO~ESCS	•		•	•
ELAB~ESCS	•	•	•	•
CSTRAT~ESCS	•	•	•	•
Reading~ESCS	•	•	•	•
MEMO~Gender	•	•	•	•
ELAB~Gender	•	•	•	
CSTRAT~Gender	•	•	•	•
Reading~Gender	•	•	•	•
MEMO~Immigr	•		•	•
ELAB~Immigr				
CSTRAT~Immigr			•	
Reading~Immigr				
ELAB~MEMO	•	•	•	•
CSTRAT~MEMO	•	•	•	•
Reading~MEMO	•	•	•	•
CSTRAT~ELAB	•	•	•	•
Reading~ELAB	•	•	•	•
Reading~CSTRAT	•	•	•	•
BIC	39461.68	39464.55	39464.74	39467.63
PMP	0.664	0.158	0.144	0.034

Note. PMP, posterior model probability.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Table 5: Comparison of the result of BMA to the result of the BSEM for the PISA data

Parameter	BMA			EAP	BSEM		
	mean(βy)	SD(βy)	P(βy)%		SD	95%	PPI
MEMO~ESCS	0.069	0.037	84.2	0.057	0.024	0.011	0.103
ELAB~ESCS	0.096	0.021	100.0	0.140	0.025	0.091	0.188
CSTRAT~ESCS	0.178	0.017	100.0	0.275	0.024	0.228	0.322
Reading~ESCS	0.358	0.018	100.0	-	-	-	-
MEMO~Gender	0.276	0.043	100.0	0.274	0.043	0.190	0.359
ELAB~Gender	-0.144	0.048	96.6	-0.021	0.044	-0.108	0.066
CSTRAT~Gender	0.182	0.031	100.0	0.297	0.043	0.213	0.380
Reading~Gender	0.235	0.032	100.0	-	-	-	-
MEMO~Immigr	0.170	0.090	84.2	-	-	-	-
ELAB~Immigr	0.000	0.000	0.0	0.142	0.058	0.029	0.257
CSTRAT~Immigr	0.013	0.034	14.4	0.229	0.057	0.119	0.341
Reading~Immigr	0.000	0.000	0.0	-	-	-	-
ELAB~MEMO	0.458	0.018	100.0	-	-	-	-
CSTRAT~MEMO	0.443	0.016	100.0	-	-	-	-
Reading~MEMO	-0.221	0.019	100.0	-0.251	0.020	-0.291	-0.211
CSTRAT~ELAB	0.371	0.015	100.0	-	-	-	-
Reading~ELAB	-0.131	0.018	100.0	-0.149	0.019	-0.187	-0.112
Reading~CSTRAT	0.338	0.021	100.0	0.442	0.022	0.399	0.484

Note. EAP, expected a posterior; SD, posterior standard deviation; PPI, posterior probability interval.



Table 6: Comparison of the predictive performance for case study

Method	PISA %	(N=2,527) Log score
BMA	90.34	-.1015
Best model from BMA	90.30	-.1020
BSEM	88.44	-.1228

Note. %, percent of predictive coverage; Log score, log score of predictive coverage.

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion



Conclusion

Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- For decades, the focus of attention in SEM has been on goodness-of-fit.
- This focus has led to a proliferation of alternative fit indices that are designed to mitigate the known sensitivities of the LR test.
- This focus on goodness-of-fit is understandable but has detracted from developing models that could be used beyond the immediate investigation.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- The question of using a model for some purpose beyond the immediate investigation leads us to consider the accuracy of a model's predictions.
- The issue of predictive accuracy is a central feature of Bayesian statistics - arguably more central than goodness-of-fit.
- Indeed, the BIC and DIC get us close to choosing models based on considering predictive accuracy.
- If the goal of model building is one of predictive accuracy, then attachment to one's specific model is of less importance.
- Thus, we are less concerned about the fit of our model and more concerned about finding a model that will predict well.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- In the Bayesian world, BMA is known to yield models that perform better than any given sub-model on the criteria of predictive accuracy.
- This is due to the fact that not all models are equally good, as measured by their posterior model probabilities – yet all models contain some important information.
- By combining models, while at the same time accounting for model uncertainty, we obtain a “stronger” model in terms of predictive accuracy.
- Popular notions of the idea of model averaging include Nate Silver’s “538” website and his book “The Signal and the Noise” (2012).



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

- We show that Bayesian model averaging can be applied to structural equation models.
- We show that the theory of BMA works in the SEM domain and obtains a model yields better predictions than any given sub-model.
- As always, the full benefit of our BMA-SEM approach will rest on its application to practical problems where prediction is of high priority.



Introduction

Bayesian
Structural
Equation
Modeling

Bayesian
Model
Building

Bayesian
Model
Averaging

The
BMA-SEM
Method

Simulation
Study

Case Study

Conclusion

THANK YOU