Taking into account sampling variability of model selection indices: A parametric bootstrap approach

Sunthud Pornprasertmanit, Wei Wu, Todd D. Little

University of Kansas

Poster presented at the Tenth Annual Society of Multivariate Experimental Psychology Graduate Student Pre-conference, Vancouver, Canada on October 11, 2012.

Author Note

Sunthud Pornprasertmanit, Wei Wu, and Todd D. Little, Department of Psychology and Center for Research Methods and Data Analysis, University of Kansas.

We thank Kristopher J. Preacher and Edgar C. Merkle for useful comments. We appreciate help from the members of the KU Model Evaluation Workgroup, Center for Research Methods and Data Analysis, University of Kansas, who provided feedback and helpful suggestions.

Partial support for this project was provided by grant NSF 1053160 (Wei Wu & Todd D. Little, co-PIs) and the Center for Research Methods and Data Analysis at the University of Kansas (Todd D. Little, director). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

Correspondence should be directed to Sunthud Pornprasertmanit, Department of Psychology and Center for Research Methods and Data Analysis, University of Kansas, Lawrence, KS 66044. Email: psunthud@ku.edu

Abstract

AIC and BIC rankings are commonly used in model selection. When the population model is not one of the competing models, this approach leads to selection of a misspecified model and the model selected varies by sample size (Preacher & Merkle, 2012). We propose a parametric bootstrap approach for model selection. A simulation study shows that the proposed approach rejects all candidate models when the population model is not under consideration and picks the population model when it is under consideration.

Taking into account sampling variability of model selection indices:

A parametric bootstrap approach

In structural equation modeling, researchers often need to compare two or more competing models. For nonnested models, the decision is often made based on the ranking of the Akaike information criterion (AIC) or Bayesian information criterion (BIC). When the set of competing models does not include the population model, Preacher and Merkle (2012) demonstrated that the model selected based on the ranking of BIC varied over repeated sampling and the problem did not diminish with increased sample size. We propose an approach to take into account this sampling variability through parametric bootstrap, which is a extension of Millsap's (2010) approach to model selection.

Parametric Bootstrap

The proposed bootstrap approach involves the following steps, using BIC as an example (the same procedure applies to AIC).

1. Fit the competing models (say A and B) to the observed data and save the observed difference in BIC (ΔBIC).
2. Create a set of multiply simulated data (e.g., 1,000) based on the parameter estimates from each of the models.
3. Fit Models A and B to each simulated dataset and save the ΔBIC. This step results in sampling distributions of ΔBIC for model A (ΔBIC_A) and ΔBIC for model B (ΔBIC_B).
4. The observed ΔBIC is then compared to both distributions. If the observed ΔBIC lies in the range (given a priori alpha level) of *one of* the distributions, then either Model A or B is favored. On the other hand, if the observed ΔBIC lies in the range of both

distributions (suggesting they are equally good) or outside the range of *both* distributions (suggesting they are equally bad), a selection cannot be determined (Jones & Tukey, 2000). This approach can easily accommodate more than two competing models.

See Figure 1 for the example of bootstrap distributions and Table 1 for a summary of decision rules.

Table 1. Decision rules of the parametric bootstrap approach.

Hypotheses	Model A is favored	Model A is not favored
Model B is favored	Two models are equally favored	Model B is favored
Model B is not favored	Model A is favored	Neither model is favored

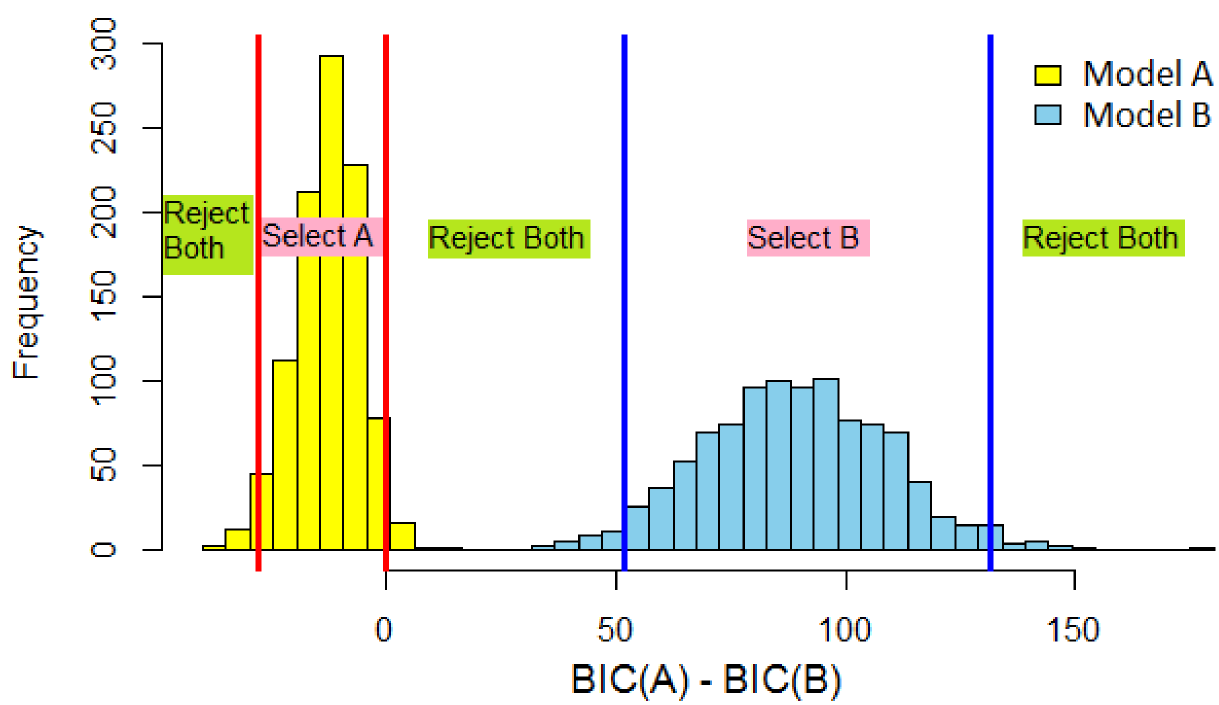

Figure 1. Sampling Distribution of ΔBIC under competing models

Table 2. The proportion of each order of the preference models from 1,000 replications in each sample size using the BIC method[a].

N	A > B > C	A > C > B	B > A > C	B > C > A	C > A > B	C > B > A
80	**.500**	**.485**	.005	.000	.010	.000
120	**.430**	**.558**	.003	.000	.008	.001
200	**.323**	**.659**	.002	.000	.016	.000
500	**.173**	**.793**	.000	.000	.034	.000
1500	.029	**.854**	.000	.000	**.117**	.000
5000	.000	**.736**	.000	.000	**.264**	.000

Note. Proportion exceeding .05 are in boldface. The preferred models have lower BIC values.
[a] The AIC result is very similar to the BIC result thus is not reported here.

Simulation Study

We conducted a simulation study to compare the performance of the proposed approach to the traditional AIC and BIC ranking approach. Following Preacher and Merkle (2012), we created data from Model D and fitted the data to Models A, B, and C (see Figure 2). We examined six sample size conditions: 80, 120, 200, 500, 1500, and 5000. The proposed approach is implemented in the *simsem* package version 0.3-5 in R (Pornprasertmanit, Miller, & Schoemann, 2012; see https://simsem.org for the example code and the package).

Results

The results for BIC were presented here. The results for AIC were similar. Using the BIC ranking, the model selection result varied across the replications within each sample size, which is consistent with Preacher and Merkle (2012; see Table 2).

The bootstrap method reveals that the fit of the competing models is too similar to be differentiated. Selection cannot be determined in a dominant proportion of samples and this proportion increased with increased sample size (see Table 3).

Table 3. The proportion of each order of the preference models from 1,000 replications in each sample size using the parametric bootstrap method[b].

N	A > B > C	A > C > B	B > A > C	B > C > A	C > A > B	C > B > A	(B, C) > A	(A, C) > B	(A, B) > C	A > (B, C)	B > (A, C)	C > (A, B)	(A, B, C)
80	**.063**	**.110**	.018	.010	**.061**	.014	.014	**.119**	**.067**	**.120**	.021	**.073**	**.310**
120	**.083**	**.185**	.029	.009	**.095**	.023	.016	**.138**	**.058**	**.110**	.014	**.077**	**.163**
200	.047	**.131**	.013	.001	**.127**	.013	.023	**.184**	.031	**.155**	.002	.026	**.247**
500	.000	.002	.000	.000	.005	.000	.000	.100	.001	.037	.001	.003	**.851**
1500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	**1.000**
5000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	**1.000**

Note. Proportions exceeding .05 are in boldface. The models in the same parenthesis mean that the selection cannot be determined among those models. [b] The results using AIC or BIC are the same.

In addition, if the population Model D is included, the parametric bootstrap performed as well as the traditional ranking approaches in selecting the population model when $N >= 120$ (see Table 4).

Conclusion

The current study shows that when the population model is not under consideration, the parametric bootstrap approach is superior to the BIC ranking approach given that it accounts for sampling variability and provides an option to discern indistinguishable models. When the population model is included, the bootstrap approach performed as well as the ranking approach except for small *n*. Given that the population model is very likely excluded in the competing models in practice, the parameter bootstrap approach seems more beneficial than the traditional approach.

More studies are to be conducted to evaluate the performance of the bootstrap approach for more variety of models and to fully reveal its benefits over the traditional ranking approach.

Table 4. The simulation result from both the BIC ranking and parametric bootstrap approach given Model D is under consideration.

N	BIC Ranking				Parametric Bootstrap				
	A	B	C	D	A	B	C	D	None
80	.029	.000	.000	**.971**	.001	.001	.003	**.882**	**.113**
120	.006	.000	.000	**.994**	.000	.001	.001	**.981**	.017
200	.000	.000	.000	**1.000**	.000	.000	.000	**.998**	.002
500	.000	.000	.000	**1.000**	.000	.000	.000	**1.000**	.000
1500	.000	.000	.000	**1.000**	.000	.000	.000	**1.000**	.000
5000	.000	.000	.000	**1.000**	.000	.000	.000	**1.000**	.000

Note. Proportions exceeding .05 are in boldface.

References

Jones, L. V., & Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods, 5*, 411-414.

Millsap, R. E. (September, 2010). *A simulation paradigm for evaluating "approximate fit" in latent variable modeling*. Paper presented at the conference "Current topics in the Theory and Application of Latent Variable Models: A Conference Honoring the Scientific Contributions of Michael W. Browne", Ohio State University, Columbus , OH.

Pornprasertmanit, S., Miller, P. J., & Schoemann, A. M. *simsem: simulated structural equation modeling version 0.3-5* [Computer Software]. Available at the Comprehensive R Archive Network.

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*, 1-14.