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

Sunthud Pornprasertmanit

Wei Wu

Todd D. Little

Department of Psychology and Center for Research Methods and Data Analysis, University of Kansas

In structural equation modeling, researchers often compare two or more competing models to determine which model provides a better fit to the data. For comparison of nonnested models, the decision is often made based on the ranking of the Akaike information criterion (AIC) or Bayesian information criterion (BIC). When the competing models do 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. They thus argued that to make accurate decision on model selection, this sampling variability should be accounted for. The current presentation proposed an approach to take into account this sampling variability through parametric bootstrap, which is a generalization of Millsap's (2010)'s approach to model selection.

The proposed approach involves the following steps: 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). The observed ΔBIC is then compared to both distributions to make inference on model selection. If the observed ΔBIC lies only in a 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 or outside the range of *both* distributions, suggesting that the two models are equally good or bad, the selection cannot be determined (Jones & Tukey, 2000). This approach can easily accommodate more than two competing models.

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 analyzed the data by Model A, B, and C (see Figure 1). We examined six sample size conditions: 80, 120, 200, 500, 1500, and 5000. The proposed approach is implemented using the `simsem` package in R (Pornprasertmanit, Miller, & Schoemann, 2012; see https://github.com/simsem/simsem/wiki).

Using the BIC and AIC rankings, the model selection results varied among the replications within each sample size, which is consistent with Preacher and Merkle (2012) (Table 1). The bootstrap method revealed that the reason behind it was that the fit of the competing models was too similar to be differentiated (selection cannot be determined in a large proportion of samples and this proportion increased with increased sample size) (see Table 2). In addition, if the population Model D is included, parametric bootstrap performed as well as AIC and BIC rankings in picking up the population model

Note: this abstract is the proposal submitted for the Tenth Annual Society of Multivariate Experimental Psychology Graduate Student Pre-conference, Vancouver, Canada. Latest updated: July, 1, 2012.

when $N \geq 120$. Therefore, the parametric bootstrap method is superior to the BIC ranking approach because it accounts for sampling variability and provides an option to discern indistinguishable models.

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 0.2-7* [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.

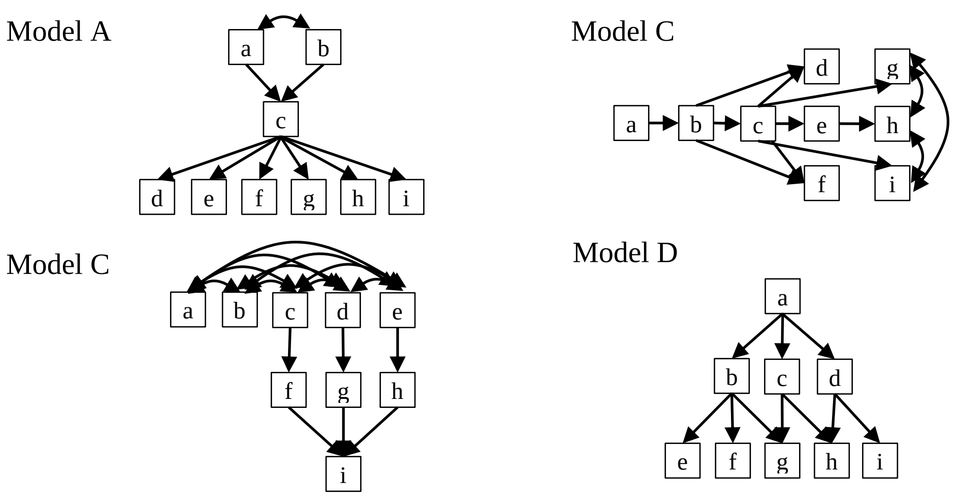

Figure 1. Model A, B, C were used in model comparison and Model D was used in data generation. This figure is modified from the Figure 1 in Preacher and Merkle (2012).

Table 1. 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.

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

Note: this abstract is the proposal submitted for the Tenth Annual Society of Multivariate Experimental Psychology Graduate Student Pre-conference, Vancouver, Canada. Latest updated: July, 1, 2012.

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.

Note: this abstract is the proposal submitted for the Tenth Annual Society of Multivariate Experimental Psychology Graduate Student Pre-conference, Vancouver, Canada. Latest updated: July, 1, 2012.