Bootstrap Computational Problems

Tihomir Asparouhov & Bengt Muthén

February 24, 2021

This note discusses situations when computational problems occur with the Bootstrap estimation of the standard errors. The following problems with the Bootstrap standard errors can be encountered:

1. Some bootstrap draws might not have completed. One can see this in the output in the following section
Number of bootstrap draws
Requested 500
Completed 400

2. Large discrepancy between non-bootstrap and bootstrap standard errors. This can be observed if the model is run with the bootstrap and without the bootstrap command and the standard error differ in large magnitude, for example the bootstrap standard errors are 10 times larger.

3. The point estimate for a particular parameter is outside of the 95% bootstrap confidence interval. Most acutely this occurs for standardized estimates and R-squared estimates.

While the manifestation of these problems is different, the source is usually the same: the data is not diverse enough due to many repeating observations or simply small sample size. Similar observations in the data do not contribute additional information to the estimation process. Every bootstrap draw is less diverse than the original data and if you start with low diversity you can get critically low on the diversity and ultimately in the information contained in the data/bootstrap draw. This creates bias in the following way. In principle, the idea behind bootstrap is that by randomly selecting observations from the original data, we can construct samples that are more or less similar to the original data as judged by any quantity. The one quantity that bootstrap is having difficulty with is diversity in the sample. If the original data has precisely 30 patterns in the data, every bootstrap draw can have only less or equal to 30 patterns, it can never have "more than 30 patterns", i.e., the population of bootstrap draws is biased downwards with respect to the number of patterns in the data. If that quantity has a substantial impact on the estimation, one would categorically have to dismiss the bootstrap procedure altogether.

Lack of diversity in the data can happen for the entire data set or it can happen for a particular subset of variables. For example, when the estimation of a correlation in the original data is heavily dependent on only a few observations, if those observations are not selected in the bootstrap draw, the bootstrap sample will produce very different estimates from the original sample and will not be helpful in the construction of accurate standard errors. Generally this applies to any data with influential outliers. Bootstrap draws that exclude the outliers will have very different parameter estimates and may produce large standard errors.

Here are specific recommendations to address each of the three problems listed above.

1. This is not necessarily a problem at all. In fact it can be viewed as a benefit. We want bootstrap draws that are similar to the original sample. Bootstrap draws that did not converge are not similar to the original sample and it is best to have such draws excluded. If at least 50% of the bootstrap replications converged and there are no other issues with the bootstrap, this issue can be ignored. If there are other issues with the bootstrap, it might be necessary to exclude even more bootstrap draws. This can be achieved by sharpening the convergence criterions MCONVERGENCE and CONVERGENCE. If a bootstrap draw is not of the same quality as the original sample, sharpening the convergence criteria can lead to excluding more bootstrap draws from the final accumulated result which may result in more accurate final result.

In some situations it may be useful to examine further the convergence issues for a particular set of bootstrap draws. The first step here is to determine which bootstrap draws did not converge. This information is reported in TECH9. The next step is to save the bootstrap sample draws using the command SAVEDATA: SAVE=BOOTSTRAP; FILE IS bootreps*.dat. This is a feature introduced in Mplus Version 8.6. The bootstrap draw data sets are saved as bootrepsN.dat where N is the order of the draw, i.e., the first bootstrap sample is saved as bootreps1.dat, etc. At this point individual bootstrap samples can be estimated separately. The error messages can be evaluated and estimation options can be tweaked to improve the chances of successful estimation.

2. Changing the estimator may resolve this problem (especially if using WLSMV, switching to ML or ULSMV has a good chance of resolving the problem). Removing weak spots in the model can resolve the problem. For example: factors with two indicators, categories with very few observations, loadings that have large SE (such may switch signs between the different bootstrap draws). Outlier analysis is recommended in this situation as well.

It maybe useful in this situation to examine the entire bootstrap distribution for a particular parameter. One way to do that is using the Mplus plot option: PLOT: TYPE IS PLOT3, which will allow you to examine the actual bootstrap distribution. Alternatively, the full bootstrap distribution for each parameter can be saved as a data file using the command SAVEDATA: SAVE=BOOTSTRAP; RESULTS = file.dat;

3. This situation probably can not be fixed and the bootstrap method should be replaced by non-bootstrap standard errors. The Bayes estimator provides non-symmetric confidence intervals and can be used as an alternative to the bootstrap method. The full bootstrap distribution can be saved as a data file using the command SAVEDATA: SAVE=BOOTSTRAP; RESULTS = file.dat, although further insights are unlikely to resolve the problem.