Empirically Corrected Test Statistics for Structural Equation Modeling with Large p and Small N

Survey data typically contain many variables. Structural equation modeling (SEM) is commonly used in analyzing such data. The most widely used statistic for evaluating the adequacy of a SEM model is the likelihood ratio statistic $T_{ml}$ and a rescaled statistic $T_{rml}$. When the number of observations (N) is large and the number of items or variables (p) is small, these two statistics perform reasonably well for normally distributed data or with certain violation to normality. However, in practice, p can be rather large while N is always limited due to not having enough participants. Even with normally distributed data and relatively large N, empirical results show that $T_{ml}$ and $T_{rml}$ rejected the correct model close to 100% when p is relatively large. Better statistics are needed for SEM with large p and/or small N. Parallel to Bartlett correction, we propose an empirical approach to correct $T_{ml}$ and $T_{rml}$ so that the mean of the resulting statistics approximately equals the degrees of freedom of the nominal chi-square distribution. Results show that empirically corrected statistics follow the nominal chi-square distribution much more closely than previously proposed corrections to $T_{ml}$ and $T_{rml}$, and they control type I errors reasonably well whenever N is greater than 50 or 2p.