16

SMALL SAMPLE SOLUTIONS FOR STRUCTURAL EQUATION MODELING

Yves Rosseel

DEPARTMENT OF DATA ANALYSIS, GHENT UNIVERSITY, GHENT, BELGIUM

Introduction

Structural equation modeling (SEM) is a statistical modeling procedure that is used in the social and behavioral sciences to study the relationships among latent variables (Bollen, 1989). Usually, a structural equation model can be divided into two parts: The measurement part relates the latent variables to a set of observed variables or indicators, and the structural part represents the hypothesized relationships among these latent variables. A typical model is presented in Figure 16.1.

SEM has a bad reputation when it comes to sample size requirements, which is likely due to a combination of factors. First, structural equation models can become quite large, involving many (observed and latent) variables. As a result, many parameters must be estimated, and a reasonable amount of data is needed to obtain good-quality estimates for those parameters. Second, the statistical machinery behind (traditional) SEM is based on large sample theory, which implies that good performance (of both point estimation and inference) is only guaranteed when the sample size is large enough. Third, some simulation studies in the SEM literature have suggested that huge sample sizes are needed in order to yield trustworthy results. And although these findings were only relevant for specific settings (one infamous example is the so-called "Asymptotically Distribution Free" or ADF estimation method), these studies fueled the now conventional wisdom that SEM can only be used if the sample size is reasonably large (say, $n > 500$) or even very large ($n > 2000$).

For many reasons, however, small sample sizes are simply a reality. When this is the case, many applied researchers hesitate to use SEM and instead employ suboptimal procedures, such as regression or path analysis based on sum scores. Unfortunately, these procedures often lead to biased results and misinformed

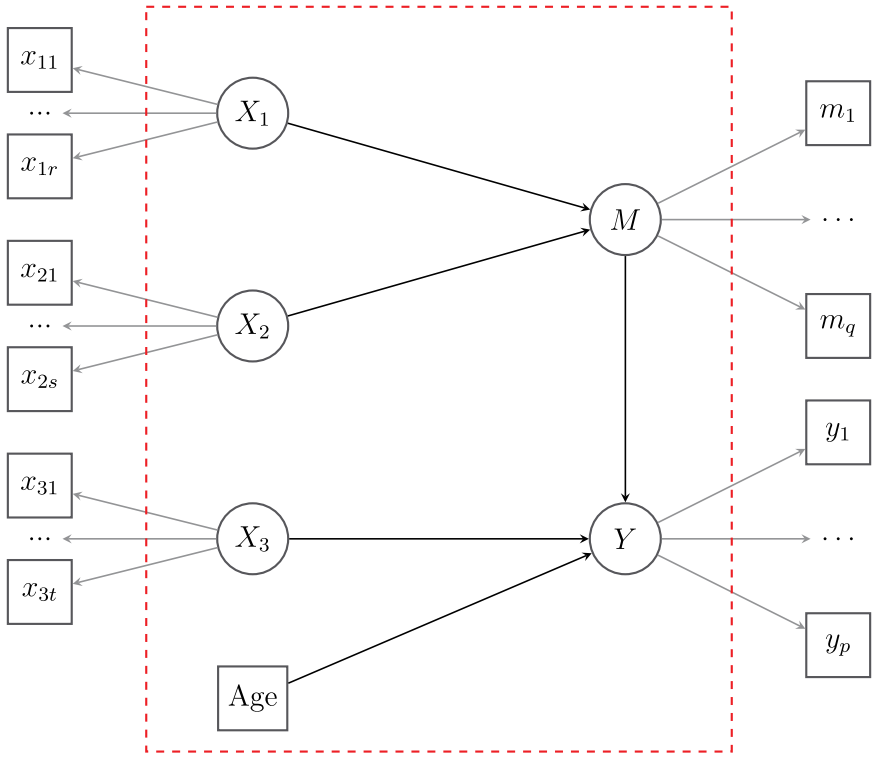

FIGURE 16.1 A typical structural equation model with a structural part (within the dashed box) and multiple measurement models. Age is an observed variable, but X_1, X_2, X_3, M, and Y are latent variables. Each latent variable is measured by a set of observed indicators. For example, X_1 is measured by a set of r indicators: x_{11}, x_{12}, x_{13}, …, x_{1r}

conclusions. Perhaps a better strategy would be to keep the spirit of SEM but to also look for solutions to handle the small sample problem. In this chapter, I will describe some of these solutions. For readers seeking guidance on choosing an appropriate sample size for their study, I suggest reading Muthén and Muthén (2002) or Wolf, Harrington, Clark, and Miller (2013).

The remainder of this chapter is organized into three sections: First, I discuss some issues that may arise with small sample sizes in SEM. Next, I present four alternative estimation approaches that may be used (instead of traditional SEM) when the sample size is small. Finally, I describe some small sample corrections for test statistics and standard errors.

Some issues with small samples sizes in SEM

Consider a fairly large model similar to the model in Figure 16.1. If all observed variables are continuous, the default estimator in most (if not all) SEM software packages is maximum likelihood. Usually, the maximum likelihood estimator is

a good choice because it features many desirable statistical properties. In addition, the maximum likelihood approach can be adapted to handle missing data (under the assumption that data are missing at random), and so-called "robust" standard errors and test statistics have been developed to deal with non-normal data and mis-specified models.

However, if the sample size is rather small (say, $n < 200$), then several problems may arise; this has been well documented in the literature (Bentler & Yuan, 1999; Boomsma, 1985; Nevitt & Hancock, 2004). First, the model may not converge, which means that the optimizer (the algorithm trying to find the values for the model parameters that maximize the likelihood of the data) has failed to find a solution that satisfies one or more convergence criteria. On rare occasions, the optimizer is simply mistaken. In this case, changing the convergence criteria, switching to another optimization algorithm, or providing better starting values may solve the problem. But if the sample size is small, it may very well be that the data set does not contain enough information to find a unique solution for the model.

A second problem may be that the model converged but resulted in a non-admissible solution. This means that some parameters are out of range. The most common example is a negative variance. Another example is a correlation value that exceeds 1 (in absolute value). It is important to realize that some estimation approaches (both frequentist and Bayesian) may—by design—never produce out-of-range solutions. Although this may seem like a desirable feature, it merely masks potential problems with either the model or the data. It is important that users notice negative variances (or other out-of-range parameters; Savalei & Kolenikov, 2008). Negative variances are often harmless, but they can be a symptom of structural misspecification. Several ways to test for structural misspecification are discussed in Kolenikov and Bollen (2012).

A third problem relates to the fact that maximum likelihood is a large sample technique. This implies that working with small sample sizes may lead to biased point estimates, standard errors that are too small, confidence intervals that are not wide enough, and p-values for hypothesis tests that cannot be trusted.

Possible solutions for point estimation

In this section, I briefly describe four alternative approaches to estimate parameters in an SEM framework with small sample sizes. The purpose of this section is not to give an exhaustive overview of all possible solutions, or to compare them under different settings, but to briefly introduce these solutions because they are not widely known among applied users. I limit myself to frequentist methods and solutions that are available in free and open-source software.

Penalized likelihood estimation

Penalized likelihood methods (or regularization methods) have been developed in the (statistical) machine-learning literature and are particularly useful

when the sample size is small—compared to the number of variables in the model (Hastie, Tibshirani, & Friedman, 2009). Penalized likelihood methods are similar to ordinary likelihood methods (like maximum likelihood estimation) but include an additional penalty term to control for the complexity of the model. The penalty term can be formulated to incorporate prior knowledge about the parameters or to discourage parameter values that are less realistic (e.g., far from zero). Two popular penalty terms are the l_2 or ridge penalty, and the l_1 or lasso (least absolute shrinkage and selection operator) penalty (Tibshirani, 1996).

To illustrate how this penalization works, imagine a univariate regression model with a large number of predictors. Without any penalization, all the regression coefficients are computed in the usual way. However, the ridge penalty term will shrink (all) the coefficients towards zero, whereas the lasso penalty will additionally shrink tiny coefficients all the way to zero. In the latter approach, only "strong" predictors (for which there is strong support in the data) survive, while "weak" predictors that can hardly be distinguished from noise are eliminated. In general, adding penalty terms leads to models that are less complex, and this is particularly beneficial if the sample size is small. Alternative penalty terms have been proposed to overcome some of the limitations of the ridge and lasso penalties. Two recent penalties are smoothly clipped absolute deviation (SCAD; Fan & Li, 2001) and minimax concave penalty (MCP; Zhang, 2010). Interestingly, penalized likelihood methods are closely related to Bayesian estimation methods. In particular, ridge and lasso penalties correspond to Gaussian and Laplace priors, respectively, whereas both SCAD and MCP correspond to certain improper priors (Huang, Chen, & Weng, 2017).

Although these penalization approaches have been around for a few decades, they have only recently been applied to SEM (Jacobucci, Grimm, & McArdle, 2016; see also Huang et al., 2017). Fortunately, we now have access to several free and open-source `R` packages that have implemented these methods for SEM. Two examples are the `regsem` package (Jacobucci, Grimm, Brandmaier, Serang, & Kievit, 2018) and the `lslx` package (Huang & Hu, 2018).

A disadvantage of these penalized methods is that the user needs to indicate which parameters require penalization, and how much. In an exploratory analysis, it may be useful and even advantageous to penalize parameters towards zero if little support for them can be found in the data. However, SEM is usually a confirmatory approach, and the user needs to ensure that all parameters that are initially postulated in the model are not removed by the penalization.

Model-implied instrumental variables

Bollen (1996) proposed an alternative estimation approach for SEM based on model-implied instrumental variables in combination with two-stage least squares (MIIV-2SLS). In this approach, the model is translated to a set of (regression) equations. Next, each latent variable in these equations is replaced with its marker

indicator (usually the first indicator, where the factor loading is fixed to unity and the intercept is fixed to zero) minus its residual error term. The resulting equations no longer contain any latent variables but have a more complex error structure. Importantly, ordinary least squares estimation is no longer suitable for solving these equations because some predictors are now correlated with the error term in the equation. This is where the instrumental variables (also called instruments) come into play. For each equation, a set of instrumental variables must be found. An instrumental variable must be uncorrelated with the error term of the equation but strongly correlated with the problematic predictor. Usually, instrumental variables are sought outside the model, but in Bollen's approach, the instrumental variables are selected from the observed variables that are part of the model. Several (automated) procedures to find these instrumental variables within the model have been developed. Once the instruments are selected, an estimation procedure is needed to estimate all the coefficients of the equations. Econometricians developed a popular method to accomplish this called two-stage least squares (2SLS).

A major motivation for MIIV-2SLS is that it is robust: It does not rely on normality and is less likely to spread bias (which may arise from structural misspecifications) in one part of the model to other parts of the model (Bollen, 2018). Another attractive feature of MIIV-2SLS is that it is noniterative. Meaning, there cannot be any convergence issues, and MIIV-2SLS may provide a reasonable solution for models where the maximum likelihood estimator fails to converge.

An important aspect of the model-implied instrumental variables approach is the (optimal) selection of instruments when there are a large number of instruments to choose from. Bollen, Kirby, Curran, Paxton, and Chen (2007) found that with small sample sizes (e.g., $n < 100$) it was best not to use a large number of instruments because it led to greater bias. More research is needed to evaluate the performance of this estimator in settings where the sample size is (very) small. The MIIV-2SLS approach is available in the `R` package `MIIVsem` (Fisher, Bollen, Gates, & Rönkkö, 2017).

Two-step estimation

In the two-step estimation approach, a strict distinction is made between the measurement part and the structural (regression) part of the model, and estimation proceeds in two steps. In the first step, all the measurement models are fitted one by one. In the second step, the full model is fitted, including the structural part, but the parameters of the measurement models are kept fixed to the values found in the first step. The main motivation for the two-step approach is to separate the measurement model(s) from the structural part during estimation so that they cannot influence each other. In the traditional maximum likelihood framework, all parameters are fitted simultaneously. As a result, misspecifications in the structural model may affect the estimated factor loadings of one or more measurement models, and this may lead to interpretation problems for the latent variables (Burt, 1976; see also Anderson & Gerbing, 1988).

The two-step approach received renewed attention in the latent class literature (Bakk & Kuha, 2018; Bakk, Oberski, & Vermunt, 2014) and was recently implemented within the `R` package `lavaan` (Rosseel, 2012); see also Chapter 17 (Smid & Rosseel). For very large models with many latent variables, it may be expected that fewer convergence problems arise because the model is estimated in parts. Encountering convergence issues in the first step allows the researcher to identify the problematic measurement model(s). If convergence issues only occur in the second step, it becomes clear that the problem lies in the structural part of the model.

Factor score regression

The simple but powerful idea of factor score regression is to replace all latent variables with factor scores. Similar to the two-step method, each measurement model is fitted one at a time. Next, factor scores are computed for all the latent variables in the usual way. Once the latent variables have been replaced by their factor scores, all variables are observed. In a final step, the structural part of the model is estimated. This estimation often consists of a regression analysis or a path analysis (as in Figure 16.1). The name "factor score regression" refers to both scenarios.

If used naïvely, factor score regression will likely lead to substantial bias in the estimated parameters of the structural part, even when the sample size is very large (Skrondal & Laake, 2001), because the factor scores have been treated as if they were observed without measurement error. Fortunately, there are various ways to correct for this bias. For example, Croon (2002) devised a correction that removes the bias, and several studies have shown that this method works remarkably well (Devlieger, Mayer, & Rosseel, 2016; Devlieger & Rosseel, 2017). Croon's method works as follows: First, the variance–covariance matrix of the factor scores is computed. Based on the information from the measurement models, the elements of the variance–covariance matrix are corrected in order to approximate the model-implied variances and covariances of the latent variables. This corrected variance–covariance matrix then forms the input of a regular regression or path analysis.

Similar to the two-step method, factor score regression (combined with Croon's correction) may be a useful alternative for fairly large models (with many measurement models) in combination with a relatively small sample size. In addition, it is possible to fit the measurement models using a noniterative estimator (for an example based on instrumental variables, see Hägglund, 1982). For the structural model, depending on whether it is recursive, a single-stage or a two-stage least squares estimator can be used. In short, this method can be made fully noniterative, which would avoid any convergence issues (Takane & Hwang, 2018). Still, Croon's correction may produce a variance–covariance matrix (for the variables belonging to the structural part) that is not positive definite—particularly if the measurement error is substantial. Therefore, Croon's correction is not entirely free

from estimation problems. In this case, the only solution may be to create a sum score for each latent variable and to estimate a model where each latent variable has a single indicator (the sum score) with its reliability fixed to a realistic value that is provided by the user (Savalei, 2018).

Discussion

All the methods I have described in this section have advantages and disadvantages. The penalized likelihood approach is perhaps the only method that was specifically designed to handle (very) small sample sizes. The other three methods use a divide-and-conquer approach; they break down the full model into smaller parts and estimate the parameters of each part in turn. Apart from reducing the complexity and being less vulnerable to convergence issues, the latter three methods have the advantage of being good at localizing the problematic parts within a large model.

At the time of writing, it is not clear which method is universally superior when the sample size is (very) small. Instead of picking one method, I would recommend that applied users try all of them. Each method may provide additional insights. If the final results agree across multiple methods, then your confidence in your findings will increase. However, if the results diverge, this may indicate that the sample size is simply too small, given the complexity of the model.

Small sample inference for SEM

In this section, I will assume that a point estimation using maximum likelihood estimation, for example, was successful and resulted in an admissible solution. The next step is to note the inference part of the model: the chi-square test statistic (for overall goodness of fit), the standard errors of the individual model parameters, and the corresponding confidence intervals and/or hypothesis tests. Many authors have documented that when the sample size is small, the chi-square test leads to inflated Type I errors even under ideal circumstances (i.e., correctly specified model, normal data; see Nevitt & Hancock, 2004 and references therein). Similarly, standard errors are often attenuated (too small), and confidence intervals are not wide enough. In the next two subsections, I will briefly discuss a few attempts to tackle these small sample inference issues in SEM.

Improving the chi-square test statistic

Several corrections have been suggested to improve the performance of the chi-square test statistic, such as the Bartlett correction (Bartlett, 1937, 1954). Variations of this correction exist, but the one most studied (Nevitt & Hancock, 2004; see also Fouladi, 2000; Savalei, 2010) is a simplified correction proposed by Bartlett (1950) in the context of exploratory factor analysis. A more general

correction is called the Swain correction (Swain, 1975), and even more corrections were described by Yuan, Tian, and Yanagihara (2015).

Herzog, Boomsma, and Reinecke (2007) and Herzog and Boomsma (2009) compared these corrections and concluded that the Swain correction worked best; however, they only looked at complete and normal data. Shi, Lee, and Terry (2018) also compared the various corrections in more realistic settings and concluded that the so-called "empirically" corrected test statistic proposed by Yuan et al. (2015) generally yielded the best performance—particularly when fitting large structural equation models with many observed variables. Still, they warn that when the number of variables (P) in a model increases, the sample size (n) also needs to increase in order to control Type I error. They suggest that, roughly speaking, n should be larger than P^2 (Shi et al., 2018, p. 39). Finally, Jiang and Yuan (2017) proposed four new corrected test statistics aiming to improve model evaluation in nonnormally distributed data with small sample sizes. The results were promising, but they conclude:

> To our knowledge, there does not exist a test statistic that performs well universally. As indicated by our results, the overall model evaluation with small n is rather challenging in SEM. Although the new statistics allow more reliable model evaluation than existing ones under conditions of nonnormally distributed data at small n, their performances are still far from universally optimum.
>
> *(Jiang & Yuan, 2017, p. 493)*

To evaluate models when the sample size is small, perhaps the chi-square test should be abandoned altogether and alternatives approaches should be explored. One approach is to consider confidence intervals and tests of close fit based on the standardized root mean square residual (SRMR; Maydeu-Olivares, 2017; Maydeu-Olivares, Shi, & Rosseel, 2018). Although these tests were not constructed with small sample sizes in mind, they seem to work well even when n = 100 (the smallest sample size considered in Maydeu-Olivares et al., 2018) and the model is not too large. These tests have been implemented as part of the `lavResiduals()` function of the `lavaan` package.

Yet another approach is to consider local fit measures. This can be based on evaluating just a subpart of the model. For example, one measurement model at a time, as in the two-step and factor score regression approaches, or one equation at a time, as is done using the Sargan test in Bollen's model-implied instrumental variables approach. A different set of local fit measures is based on graphical criteria such as d-separation or trek-separation (Thoemmes, Rosseel, & Textor, 2018).

Better standard errors and confidence intervals

Literature on the performance of standard errors in SEM is limited (Yuan & Bentler, 1997; Yuan & Hayashi, 2006) and is mostly concerned with the effect of

nonnormality or model misspecifications on the quality of the standard errors. Small sample sizes, indeed, were not the focus of these studies. But in general, it is well known that if large sample theory is used to construct analytic expressions in order to compute standard errors, they may perform poorly in small samples.

When assumptions underlying analytic standard errors are not met, it is often suggested to use a resampling approach instead. One popular method is the bootstrap (Efron & Tibshirani, 1993): A bootstrap sample is generated (either by randomly sampling rows from the original data set with replacement or by simulating a new data set under the model), and a new set of parameters is estimated for this bootstrap sample. This is repeated a large number of times (say, 1,000), and the standard deviation of a parameter across all replicated bootstrap samples is used as an estimate of the standard error for that parameter. Unfortunately, despite many other advantages (Chernick, 2007), the bootstrap does not appear to be a reliable solution when the sample size is (very) small (Yung & Bentler, 1996). Hence, it may be worthwhile to explore better analytical solutions after all. Small sample corrections for (robust) standard errors have been developed in econometrics (MacKinnon & White, 1985) and have recently been adapted to the SEM context (Dudgeon, Barendse, & Rosseel, 2018). The preliminary results are encouraging, but this is still a work in progress. At the time of writing, this technology is not yet available in SEM software.

Conclusion

In this chapter, we discussed several problems in the context of SEM when the sample size is small and standard (maximum likelihood) estimation methods are used, such as nonconvergence, non-admissible solutions, bias, poorly performing test statistics, and inaccurate standard errors and confidence intervals. As potential solutions to attain better point estimates (or a solution at all), we briefly discussed four alternative estimation approaches: penalized likelihood estimation, model-implied instrumental variables, two-step estimation, and factor score regression. Only the first method was specifically designed to handle small samples. The latter approaches were developed with other concerns in mind, but they may be viable alternatives for estimation when the sample size is small.

For the inference part, I discussed various attempts to improve the performance of the chi-square test statistic for evaluating global fit in the presence of small samples. For the standard errors, I underlined that bootstrapping may not be the solution we are looking for. Unfortunately, to attain better standard errors (and confidence intervals) in the small sample setting, we may need to wait until new technology is available. Admittedly, my selection of topics in this chapter is somewhat biased. I certainly did not present all the solutions that have been proposed in the literature, but interested readers may consult Deng, Yang, and Marcoulides (2018) for an alternative perspective.

In this chapter, I focused on frequentist solutions, but a Bayesian approach is presented in Chapter 17 of this book. A major advantage of the Bayesian approach is that it does not rely on large sample asymptotics. This implies, for example, that getting correct standard errors and credible (confidence) intervals is not an issue in the Bayesian framework. On the other hand, model evaluation in a Bayesian framework requires a new set of skills; this may be intimidating for those who are unfamiliar with the Bayesian framework. The same is true for specifying priors. Choosing adequate priors is an essential ingredient of Bayesian estimation, and you should be prepared to critically reflect on this before you proceed. If you are unwilling to specify any priors, and you rely on software defaults, then you should probably avoid Bayesian SEM altogether (McNeish, 2016). If, on the other hand, you fully endorse the Bayesian approach, and you have a priori knowledge that can be encoded in informative priors, the Bayesian approach is an excellent choice.

A last closing comment: If the sample size is (very) small, it may be that the data simply do not contain enough information to answer the research questions. In that case, one should not expect miracles from statistical technology. Small samples sizes have limitations, and we should accept them.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411.

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*(4), 871–892.

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis*, *22*(4), 520–540.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A*, *160*(901), 268–282.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, *3*(2), 77–85.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B (methodological)*, 16, 296–298.

Bentler, P. M., & Yuan, K. H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*(2), 181–197.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A. (1996). An alternative two stage least squares (2sls) estimator for latent variable. *Psychometrika*, *61*(1), 109–121.

Bollen, K. A. (2018). Model implied instrumental variables (MIIVs): An alternative orientation to structural equation modeling. *Multivariate Behavioral Research*, 54(1), 1–16. doi:10.1080/00273171.2018.1483224.

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007). Latent variable models under misspecification: Two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods & Research*, *36*(1), 48–86.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229–242.

Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, *5*(1), 3–52.

Chernick, M. R. (2007). *Bootstrap methods: A guide for practitioners and researchers* (2nd ed.). Hoboken, NJ: Wiley.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Mahwah, NJ: Lawrence Erlbaum.

Deng, L., Yang, M., & Marcoulides, K. M. (2018). Structural equation modeling with many variables: A systematic review of issues and developments. *Frontiers in Psychology*, *9*, n.p.

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*(5), 741–770.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM. *Methodology*, *13*, 31–38.

Dudgeon, P., Barendse, M., & Rosseel, Y. (2018). *Leverage-based confidence intervals for structural equation modelling*. Paper presented at the Modern Modeling Methods (M3) conference, Storrs, CT.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap* (Vol. 57). Boca Raton, FL: Chapman & Hall/CRC.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fisher, Z., Bollen, K. A., Gates, K., & Rönkkö, M. (2017). MIIVsem: Model implied instrumental variable (MIIV) estimation of structural equation models [Computer software manual] (Version R package 0.5.2). Retrieved from https://CRAN.R-project.org/package=MIIVsem.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate nonnormality. *Structural Equation Modeling: A Multidisciplinary Journal*, 7(3), 356–410.

Hägglund, G. (1982). Factor analysis by instrumental variables methods. *Psychometrika*, *47*(2), 209–222.

Hastie, T., Tibshirani, R. J., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(1), 1–27.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 361–390.

Huang, P.-H., Chen, H., & Weng, L.-H. (2017). A penalized likelihood method for structural equation modeling. *Psychometrika*, *82*(2), 329–354.

Huang, P.-H., & Hu, W.-H. (2018). lslx: Semi-confirmatory structural equation modeling via penalized likelihood [Computer software manual] (Version R package 0.6.4). Retrieved from https://CRAN.R-project.org/package=lslx.

Jacobucci, R., Grimm, K. J., Brandmaier, A. M., Serang, S., & Kievit, R. A. (2018). regsem: Regularized structural equation modeling [Computer software manual] (Version R package 1.2-0). Retrieved from https://CRAN.R-project.org/package=regsem.

Jacobucci, R., Grimm, K. J., & McArdle, J. J. (2016). Regularized structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(4), 555–566.

Jiang, G., & Yuan, K.-H. (2017). Four new corrected statistics for SEM with small samples and nonnormally distributed data. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 479–494.

Kolenikov, S., & Bollen, K. A. (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification?. *Sociological Methods & Research*, *41*(1), 124–167.

MacKinnon, J., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558.

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 389–402.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(4), 599–620.

Nevitt, J., & Hancock, G. R. (2004). Evaluating small sample approaches for model test statistics in structural equation modeling. *Multivariate Behavioral Research*, *39*(3), 439–478.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Savalei, V. (2010). Small sample statistics for incomplete non-normal data: Extensions of complete data formulae and a Monte Carlo comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 241–264.

Savalei, V. (2018). A comparison of several approaches for controlling measurement error in small samples. *Psychological Methods*. Advance online publication. doi: http://dx.doi.org/10.1037/met0000181.

Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, *13*(2), 150–170.

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 21–40.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575.

Swain, A. J. (1975). *Analysis of parametric structures for variance matrices* (Unpublished doctoral dissertation), University of Adelaide, Department of Statistics.

Takane, Y., & Hwang, H. (2018). Comparisons among several consistent estimators of structural equation models. *Behaviormetrika*, *45*(1), 157–188.

Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods*, *23*(1), 27.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*, *58*(1), 267–288.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, *73*(6), 913–934.

Yuan, K.-H., & Bentler, P. M. (1997). Improving parameter tests in covariance structure analysis. *Computational Statistics & Data Analysis*, *26*(2), 177–198.

Yuan, K.-H., & Hayashi, K. (2006). Standard errors in covariance structure models: Asymptotics versus bootstrap. *British Journal of Mathematical and Statistical Psychology*, *59*(2), 397–417.

Yuan, K.-H., Tian, Y., & Yanagihara, H. (2015). Empirical correction to the likelihood ratio statistic for structural equation modeling with many variables. *Psychometrika*, *80*(2), 379–405.

Yung, Y.-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 195–226). Mahwah, NJ: Lawrence Erlbaum.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, *38*(2), 894–942.