SEM with Nonnormal Continuous Variables

Definitions and Distinctions

First, it is important to distinguish between continuous and categorical variables. Continuous variables are variables measured on a ratio or interval scale, such as temperature, height, or income in dollars. In practice, social scientists usually treat ordinal variables with many categories, such as 5- or 7-point (or more) Likert-type scales of agreement, as "continuous." There is evidence to suggest that treating such variables is not likely to result in much practical impact on structural equation modeling results when distributions are relatively normal (e.g., Babakus, Ferguson, & Jöreskog, 1987; Dolan, 1994; Hutchinson & Olmos, 1998; Johnson & Creech, 1983; Rhemtulla, Brosseau-Laird, & Salvalei, 2012).[1] Because the maximum likelihood (ML) analysis approach most commonly employed for SEM assumes multivariate normality, special data analytic techniques for nonnormal continuous variables should be used if any of the continuous variables in the model are nonnormal. Use of unadjusted ML with nonnormal continuous variables will negatively impact fit and will tend to underestimate standard errors for parameter tests (i.e., Type I errors). The parameter estimates (e.g., loadings, correlations, predictive paths) are largely unaffected (Finch, West, & MacKinnon, 1997).

In contrast, we will consider categorical variables to be nominal variables with two (i.e., binary, dichotomous) or more categories or ordinal variables with fewer than five values. Examples might include categorical variables, such as genetic sex, dead vs. alive, heart attack vs. no heart attack, and race and ethnicity categories, or ordinal variables with few response options, such as "never," "sometimes," or "always" and "poor," "good," or "excellent." The present handout is concerned with nonnormal continuous variables, but see the handout "SEM with Categorical Variables" from this class for references and more information on categorical variables.

Detection of Multivariate Nonnormality

Descriptive analysis is a good first step before embarking on any type of inferential analytic approach in general. One can begin by examining univariate distributions and skew and kurtosis. West, Finch, & Curran (1995) suggest that skewness > 2 and kurtosis > 7 are indicative of moderate or serious nonnormality, but these should be considered loose guidelines because they are univariate measures. ML assumes multivariate normality, which involves the joint distributions of all the variables taken together. If the univariate distributions are nonnormal, then the multivariate distribution will be nonnormal, but one can also have multivariate nonnormality even when all the individual variables are normally distributed (although severe multivariate nonnormality is probably less likely in practice when all of the univariate distributions are normal). Tests of multivariate normality are not widely implemented in structural equation modeling packages and procedures (although EQS and LISREL are two exceptions). Mardia's multivariate skewness and kurtosis tests are distributed normally (z-test) in very large samples, so could be evaluated against a t, z, or chi-square distribution (although remember that significance does not necessarily indicate large magnitude).[2] Multivariate kurtosis is generally a greater concern than skewness, although the two measurements tend to be related. EQS also provides a "normalized estimate" of Mardia's kappa. Bentler and Wu (2002) suggest that a normalized estimate greater than 3 will lead to chi-square and standard error biases. It is good to know your data, but a *determination of whether or not your data are multivariate normal is not needed to decide on which estimation approach you should use*, as will be explained below.

Scaled Chi-Square and Robust Standard Errors

With multivariate nonnormality, it is relatively easy to see concerning levels of bias with mild to moderate departures from normality.[3] The effect of violating the assumption of nonnormality is that *chi-square is too large* (so too many models are rejected) and *standard errors are too small* (so significance tests of path coefficients will result in Type I error; e.g., Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992). The

[1] This is consistent with evidence that dependent variables with ordinal variables with 5 to 7 or more values also produce unbiased and efficient results with ordinary least squares regression and ANOVA models (see the handout from my Univariate Statistics course called "Levels of Measurement and Choosing the Correct Statistical Test" for references and details).
[2] See Lawrence DeCarlo's (1997) macros for SPSS and SAS to calculate Mardia's multivariate skewness and kurtosis estimates and test them for significance (available at http://www.columbia.edu/~ld208/) and the mardiaTest function in the R MVN package.
[3] In terms of the significance tests, the normality is not relevant for exogenous measured variables (I mean that they serve as measured variable predictors rather than indicators for a latent exogenous variable, in which they are still dependent variables in the loadings), as there should be no impact the standard errors for predictive relationships as is the case with regression analysis.

parameter estimate values are generally not affected by nonnormality (Finch et al., 1997). The scaled chi-square and "robust" standard errors corrections to ML estimation is a method suggested by Satorra and Bentler (1988; 1994).[4] It appears to be a good general approach to dealing with nonnormality (Hu, Bentler, & Kano, 1992; Curran, West, & Finch, 1996; see Finney & Stefano, 2013 for a more complete review). Adjustments are made to the chi-square (and to relative fit indices in some packages, such as Mplus, lavaan, and EQS) and standard errors based on a weight matrix derived from an estimate of multivariate kurtosis (the parameter estimates themselves are not altered). Mplus and lavaan refer to the Satorra-Bentler corrections as MLM (for "maximum likelihood mean-adjusted" chi-square; invoked with `estimator=MLM`).[5] In the output, the chi-square and the standard errors reported use these adjustments when MLM is used. This approach is used when there are no missing values, and `estimator=MLR` is used when there are missing values (see below). Mplus also prints this kurtosis adjustment, referred to as the "scaling correction factor" (scf; or as d in Finney & DiStefano, 2013). The scaling correction factor is the standard chi-square divided by the scaled chi-square. The ratio is derived from a multivariate kurtosis estimate used to adjust the chi-square and standard errors. When data are multivariate normal, this scaling correction factor is 1.0, and there is no adjustment to the standard ML chi-square. Greater multivariate kurtosis leads to a higher scaling correction factor value (e.g., 1.6 suggests the ML chi-square is approximately 60% higher than the scaled chi-square).

Depending on the complexity of the model and the severity of the problem, sample sizes of greater than 250 may be needed (Hu & Bentler, 1999; Yu & Muthén, 2002). For smaller samples, there is a potential danger of overcorrection with this method, and it may be wise to test the model with both the unadjusted ML and the Satorra-Bentler corrections to be sure the conclusions do not differ. Otherwise, it may make sense to use these adjustments routinely to avoid incorrect conclusions about fit and biases in significance tests.

Bootstrapping is an increasingly popular and promising approach in a number of contexts, and this resampling method can also be used to correct fit and standard errors for nonnormality in SEM, but it seems that more work is needed to understand how well it performs under various conditions (e.g., specific bootstrap approach, sample sizes needed). The simulation work that has been done (Fouladi, 1998; Hancock & Nevitt,1999; Nevitt & Hancock, 2001; see Hancock & Liu, 2012 for a more recent discussion) suggests that, in terms of bias, a standard "naïve" bootstrap seems to work at least as well as robust adjustments to standard errors. However, the Nevitt and Hancock (2001) results suggest that standard errors may be erratic for sample size of 200 or less and samples of 500 to 1,000 may be necessary to overcome this problem. The complexity of the model should be taken into account as their simulations were based on a moderately complex factor model (i.e., smaller sample sizes may be acceptable for simpler models). A separate bootstrapping can be done for chi-square fit estimates, known as the Bollen-Stine bootstrap approach. The Bollen-Stine chi-square approach seems to adequately control Type I error but there is some cost to power (Nevitt & Hancock, 2001). Bootstrapping approaches have now been incorporated in most major SEM packages, and it is the only approach to nonnormal continuous variables available in Amos.

Alternative Fit Indices

Relatively little simulation work on alternative fit indices (e.g., RMSEA, IFI, CFI) derived from robust approaches to nonnormal continuous variables (Satorra-Bentler robust approach or bootstrapping) is currently available, but the Satorra-Bentler scale chi-square appears to outperform the maximum unadjusted likelihood chi-square when data are nonnormal (Curran et al., 1996). Thus far, studies suggest that at least some alternative fit indices (TLI, CFI, RMSEA) using standard cutoffs (Hu & Bentler, 1999) also perform fairly well with the robust approach as long as the Satorra-Bentler scale chi-square for the null is also used to compute alternative fit indices and sample size is reasonably large (N = 250 or larger; Nevitt & Hancock, 2000; Maydeu-Olivares, Shi, & Rosseel, 2018; Yu & Muthén, 2002). Findings from Maydeu-Olivares et al. (2018) suggest that the RMSEA may have some erratic values, especially for small sample sizes, and the SRMR was largely accurate across the conditions they examined. The user should use some caution, because programs do not always recalculate incremental fit indices such as the CFI, TLI, or the IFI

[4] The robust standard errors are the same as those proposed by Huber, White, and Eiker and those used with multilevel models and generalized estimating equations (Eiker, 1963; Huber, 1967; Liang & Zeger, 1986; White, 1982).
[5] In LISREL and EQS, the Satorra-Bentler corrections are referred to as ML Robust.

using the scaled chi-square for the tested model or the null model (I know that Mplus and EQS do use the scaled chi-squares in their calculation). Relative fit indices will likely be problematic when scaling corrections to the null model are not used (Hu & Bentler, 1999).

Nested Tests

Nested tests (likelihood ratio test) require special attention for robust estimation. The scaling correction factor (*scf*) must be used to weight the difference (Satorra, 2000; Satorra & Bentler, 2001). The following formula gives the adjustment to the difference in chi-square which can be used for significance testing:

$$\Delta\chi^2_{SB} = \frac{\chi^2_{M0} - \chi^2_{M1}}{\left(df_{M0}scf_{M0} - df_{M1}scf_{M1}\right) / df_{M0} - df_{M0}}$$

The *scf* is equal to the ratio of traditional ML chi-square to the Satorra-Bentler scale chi-square for the model, or $scf = \chi^2_{ML} / \chi^2_{SB}$. This computation of the weighted difference test can produce negative values and Satorra and Bentler (2010) have proposed an alternative computation. Mplus provides correct automated nested tests with the DIFFTEST command that can be used for robust methods (as well as categorical estimation methods; see Asparouhov & Muthén, 2013; Bryant & Satorra, 2012; Satorra & Bentler, 2010).

Missing Data Estimation with Nonnormal Continuous Variables

For nonnormal continuous data where some data are missing, a variation on the full maximum likelihood method can be used (Yuan & Bentler, 2006). In Mplus and lavaan, this is obtained with `estimator = MLR`.

References

Asparouhov, T. and Muthen, B. (2013). Computing the strictly positive Satorra-Bentler chi-square test in Mplus. Unpublished manuscript. https://www.statmodel.com/examples/webnotes/SB5.pdf
Babakus, E., Fergnson, C. E. & Joreskog, K G, (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 24*, 2228.
Bentler, P.M., & Wu, E.J.C. (2002). *EQS for Windows user's guide.* Encino, CA. Multivariate Software, Inc.
Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Psychometrika, 66*, 507–514.
Curran, P. J., West, S. G, & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.
DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2*, 292-307.
Dolan, C. V. (1994), Factor analysis of variables with 2, 3, 5, and 7 response categories: A comparison of categorical variable estimators using simulated data. *British journal of Mathematical and Statistical Psychology, 47*, 309-326.
Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," The Annals of Mathematical Statistics, 34, 447–456
Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal, 4*(2), 87-107.
Finney, S.J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course, 2nd Edition* (pp. 439–492). Charlotte, NC: Information Age Publishing.
Fouladi, R.T. (1998, April). *Covariance structure analysis techniques under conditions of multivariate normality and non-normality—modified and bootstrap based test statistics.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
Hancock, G. R., & Liu, M. (2012). Bootstrapping standard errors and data-model fit statistics in structural equation modeling. In R. H. Hoyle (Ed.), Handbook of structural equation modeling (p. 296–306). The Guilford Press.
Hancock, G.R., & Nevitt, J. (1999). Bootstrapping and the identification of exogenous latent variables within structural equation models. *Structural Equation Modeling*, *6*, 394-399.
Hutchinson, S, R., & Olmos, A (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *5,* 344-364.
Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6(1*), 1-55.
Hu, L., Bentler, P.M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.
Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*
Johnson, D.R., & Creech, J.C. (1983) Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review, 48*, 398-407.
Lee, S-Y. (2007). *Structural Equation Modeling: A Bayesian Approach.* New York: Wiley.
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.
Nevitt, J., & Hancock, G.R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling, 8,* 353-377.
Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*, 354-373.
Satorra, A., & Bentler, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 Proceedings of the Business and Economic Statistics Section of the American Statistical Association*, 308-313.
Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye and C.C. Clogg (eds.), *Latent Variable Analysis: Applications to Developmental Research* (pp. 399-419). Newbury Park: Sage.
Satorra, A. & Bentler, P.M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75,* 243-248. doi:10.1007/s11336-009-9135-y
Yuan, K. H., & Bentler, P. M. (2006). Asymptotic robustness of standard errors in multilevel structural equation models. *Journal of Multivariate Analysis, 97*(5), 1121-1141.
West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), Structural equation modeling: Concepts, issues, and applications (p. 56–75). Sage Publications.
White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity". Econometrica, 48 (4): 817–838.
Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*