Alternative Estimation Methods

ML

Remember that the usual approach to estimating fit and coefficients in SEM is the maximum likelihood (ML) approach. ML uses derivatives to minimize the following fit function:

$$F_{\mathrm{ML}} = \log\left|\mathbf{\Sigma}(\boldsymbol{\theta})\right| + tr\left(\mathbf{S}\mathbf{\Sigma}^{-1}(\boldsymbol{\theta})\right) - \log\left|\mathbf{S}\right| - (p+q)$$

The ML estimator assumes that the variables in the model are (conditionally) multivariate normal (i.e., the joint distribution of the variables is distributed normally).[1]

GLS

Generalized least squares is an alternative fitting function. The GLS fit function also minimizes the discrepancy between $\mathbf{S}$ and $\mathbf{\Sigma}$, but uses a weight matrix for the residuals, designated W.

$$F_{\mathrm{GLS}} = \left(\frac{1}{2}\right) tr\left(\left\{\left[\mathbf{S} - \mathbf{\Sigma}(\boldsymbol{\theta})\mathbf{W}^{-1}\right]\right\}^{2}\right)$$

Notice that this is a much simpler function (e.g., no logs), and it is clear that the discrepancy between the obtained covariance matrix and the covariance matrix implied by the model ($\mathbf{S} - \mathbf{\Sigma}$) is minimized after weighting it by $\mathbf{W}$. Although any $\mathbf{W}$ can be chosen for the weight matrix, most commonly, the inverse of the covariance matrix, $\mathbf{S}$, is used in SEM packages. F_{GLS} is asymptotically equivalent to F_{ML}, meaning that as sample sizes increase, they are approximately equal. F_{GLS} is based on the same assumptions as F_{ML} and would be used under the same conditions. It is thought to perform less well, however, in small samples, so F_{ML} is usually chosen instead of F_{GLS}. The simplicity of the function, however, means that other weight matrices could be used in an attempt to correct for violations of distributional assumptions.

ADF/WLS/AGLS

The asymptotic distribution free function for SEM is given by Browne (1984). It is described as arbitrary generalized least squares (AGLS) by Bentler in the EQS package and weighted least squares (WLS) by Joreskog and Sorbom in LISREL (and the related approach described below used by Mplus and lavaan). The main advantage of the ADF estimator is that it does not require multivariate normality. The ADF estimator is based on the F_{GLS}, except a different $\mathbf{W}$ is chosen. It can be written in a general form that encompasses GLS, ML, and ULS (not discussed here) where the difference depends on the choice of $\mathbf{W}$:

$$F_{\mathrm{ADF}} = F_{\mathrm{AGLS}} = F_{\mathrm{WLS}} = (\mathbf{s} - \boldsymbol{\sigma})' \mathbf{W}^{-1} (\mathbf{s} - \boldsymbol{\sigma})$$

$\mathbf{W}$ used in F_{ADF} is based on a covariance of all of the elements of the covariance matrix, $\mathbf{S}$. That is, a covariance matrix is constructed that estimates the covariances between each s_{ij} element of $\mathbf{S}$, and is therefore a $½[v(v+1)]$ by $½[v(v+1)]$ matrix, with v as the number of observed variables. It is important to realize that the "covariances of covariances" are related to kurtosis estimates (so called "fourth-order moments").[2] So, the GLS fit function is weighted by variances and kurtosis in attempt to correct for violations of the normality assumption. Another way of saying this is that when the data are normal, the ADF estimator reduces to GLS because there is no kurtosis. The large weight matrix causes serious practical difficulties when there is a large number of variables in the model (e.g., more that 20 or so), and computer packages (e.g., EQS) do not allow estimation unless the number of cases is equal or greater than number of elements in the weight matrix (i.e., $½[v(v+1)]$ times $½[v(v+1)]$ divided by 2). Simulation studies suggest that chi-square values are severely overestimated with small samples and that sample sizes of about 5000 are necessary for good estimates. A study by Olsson, Foss, Troye, and Howell (2000) suggests

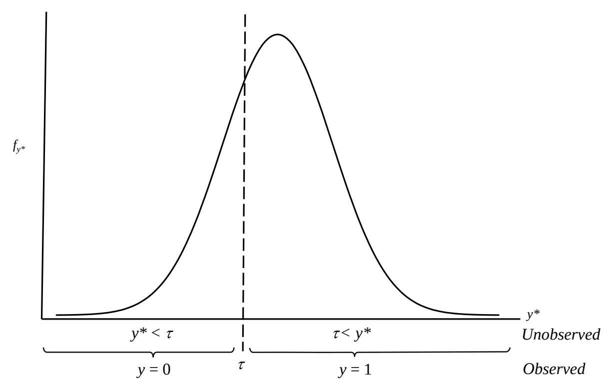
---

[1] The conditional portion of this assumption is that the distribution of the y variable that is of importance is the residual distribution. That is, if after accounting for the predictors of y, the distribution is normal, then the assumption is met. It also should be noted here that, like regression, there is not an assumption about the distribution of the predictor/independent variable, only the dependent variable.

[2] A raw form equation for kurtosis is $\left[\sum\left(Y-\bar{Y}\right)^{4}/N\right]/s^{4}$, where the deviations from the mean are raised to the fourth power.

that ADF estimation performs poorly when the model is misspecified. Combined with the limitation of variables, this is usually seen as an unattractive approach when nonnormality exists.

WLS for Categorical Variables
The ADF estimator is not very practical as a general estimation approach in its original form, but it has been implemented with considerable success with categorical (binary and ordinal) variables in modified form. Models with categorical variables are always considered to be in violation of the normality assumption and, thus, the usual F_{ML} estimator is not recommended. In the context of the categorical variable estimation, the ADF estimator is most often referred to these days as WLS. The modified approach that has developed is a multiple-step estimation involving polychoric correlations as input to create the asymptotic covariance matrix used for weighting in the WLS estimation. The idea behind the method is that categorical variables can be conceived as having an underlying continuous unobserved variable, called y^*. y^* is estimated by *polychoric* correlations which correct for loss of information when Pearson correlations are used due to cruder categorization of a continuous variable (see Olsson, 1979; MacCallum, Zhang, Preacher, & Rucker, 2002). *Tetrachoric* correlations are a special case of polychoric correlations involving only binary variables, and *polyserial* correlations are those involving the correlation between a binary and a continuous variable. Often all three types are referred to more generally as polychoric correlations. The concept of y^* is the same as that invoked to conceptualize probit analysis. The variable y^* is a true value that is not observed but leads to the observed response of y, which is binary or ordinal. The value of y^* can be thought of as a propensity to respond 0 or 1 on the y variable in the case that y is binary, for example. The figure below is an analogue representation of the idea:



In a sense, the WLS estimation is performed on the estimated y^* variables in this two-step estimation process. LISREL requires the user to implement this process explicitly in separate steps, but other software programs, such as Mplus, lavaan, and EQS, allow the two-step process to be handled automatically (as long as raw data are available). The approach requires an inversion of the full weight matrix, which can become cumbersome when there are many variables. Estimation using this approach performs reasonably well statistically in a number of circumstances, but can be improved upon.

DWLS
Muthén (1993) suggested a modification of this general categorical variable approach, known as diagonally weighted least squares (DWLS) estimation or a "limited information" approach. The DWLS approach uses the WLS estimator with polychoric correlations as input to create the asymptotic covariance matrix. The approach is computationally more practical because it avoids inversion of the large weight matrix (using something called “Taylor expansion”). The method seems to perform better statistically as well (Rhemtulla, Brosseau-Laird, & Savalei, 2012), performing better than the full WLS for small samples. The approach is typically paired with robust estimation adjustments (sometimes called the "sandwich" estimator) that improves standard error, chi-square, and fit indices. In Mplus (and lavaan, and sometimes more generally in the literature), the DWLS with adjustment is referred to as WLSM or WLSMV, depending on whether just means or means and variances are used in the adjustment process. The robust DWLS methods seem to work well in many conditions, including smaller samples and with nonnormal data (e.g., Rhumtulla et al., 2012). EQS (Bentler & Wu, 2002) uses an alternative robust method described as a "partitioned maximum likelihood" approach, obtaining estimates in separate steps depending on the types of variables involved.

The Satorra-Bentler Scaled Chi-square and Standard Errors

Satorra and Bentler suggested that multivariate kurtosis estimates be used to “scale” or correct the chi-square value and standard errors (Satorra & Bentler, 1988; Satorra & Bentler, 1994). Chi-square is usually inflated with nonnormal samples and standard errors are usually too small (although depending on whether the distribution is platykurtic or leptokurtic, the direction of bias can differ). This approach is used for continuous nonnormal variables and appears to do fairly well with small samples (200-500 cases; Curran, West, & Finch, 1996). This is the same robust adjustment used for in conjunction with DWLS for categorical variables.

Bootstrapping

Bootstrap is another approach to problems with nonnormality (but is not typically recommended for binary and ordinal variables with few categories). In the bootstrap approach, a large number of samples (usually 500 or 1000 are recommended) are drawn from your data. The samples are drawn *with replacement,* so that the same cases may be drawn into the same bootstrap sample. These repeated samples create a mini sampling distribution, and based on the central limit theorem, it should have desirable distributional characteristics. There are a number of variations on bootstrapping with SEM, including “naïve” bootstrap, bias correction, and bias corrected accelerated (but see Bollen & Stine, 1993; Yung & Bentler, 1996). The bootstrap samples are used to calculate new standard errors ("naïve" bootstrap) and can be used to correct the chi-square for fit (Bollen-Stine bootstrap). The z-tests or “critical ratio” uses the bootstrap standard errors, and are considered “approximate” significance tests. Evidence by Nevitt and Hancock (2001) suggest an original sample size of 500 or greater may be needed for stable bootstrap estimation.

Other Estimators

Other possible estimators include two-stage least squares (2SLS), three-stage least squares (3SLS), ordinary least squares (OLS), and unweighted least squares (ULS). Most of these approaches are seldom used, because they provide poor estimation (e.g., ULS) or because they have not been very thoroughly investigated (e.g., 3SLS). 2SLS has received more attention in statistical papers because it does not rely on normality assumptions and is one approach to moderator tests in SEM (more on this topic later; see Bollen & Biesanz, 2002; Bollen & Paxton, 1998).

References and Further Reading

Bollen, K.A. (2001). Two-Stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications. Ch 7, pp 199-138 in R. Cudeck, S. du Toit and D. Sorbom (eds) Structural Equation Modeling: Present and Future: A Festschrift in honor of Karl Joreskog, Scientific Software International: Lincolnwood.

Bollen, K. A., & Biesanz, J. C. (2002). A note on a two-stage least squares estimator for higher-order factor analyses. *Sociological Methods & Research, 30*, 568-579.

Bollen, K. A., & Paxton, P. (1998). Two-stage least squares estimation of interaction effects. In R. E. S. G. A. Marcoulides (Eds.), Interaction and nonlinear effects in structural equation modeling (pp. 125-151). Mahwah, NJ, Lawrence Erlbaum Associates.

Bollen, K.A., & Stine, R.A. (1993). Bootstrapping goodness-of-fit measures in structural equation modeling. In K.A. Bollen & J.S. Long (Eds.), Testing structural equation models. Newbury Park: Sage.

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.

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.

MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotimization of quantitative variables. Psychological Methods, 7, 19-40.

Muthén , B.O. (1993). Goodness of fit with categorical and other nonnormal variables. In K.A. Bollen, & J.S. Long (eds.), *Testing structural equation Models* (pp. 205-234). Newbury Park, CA: Sage.

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.

Olsson, u. H. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psycltometrika, 44*, 442-460.

Olsson, U.H., Foss, T., Troye, S. V., & Roy D. Howell (2000). The Performance of ML, GLS and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. Structural Equation Modeling, 7 (4), 557-595.

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.

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. Issues and Techniques (pp.125-157). Mahwah, NJ: Lawrence Erlbaum.