SEM MATLAB toolbox – Usage Notes

To start GUI

i. Point working directory of MATLAB to 'SEM toolbox'

ii. Type 'sem_gui' at MATLAB command line to bring up GUI

Model specification

i. Enter the number of observed variables and latent variables in the text boxes. Failure to enter a number after clicking on a text box will produce an error message.

ii. In the symmetric and asymmetric matrices, first 'x' rows correspond to latent variables and the remaining 'y' rows correspond to observed variables, where 'x' and 'y' are the number of latent and observed variables respectively.

iii. Free parameters should be specified as 'inf', fixed parameters can take any value. A fixed parameter of 0 indicates 'no relationship' between those two variables.

iv. All diagonal elements of symmetric matrix should be specified to a positive value or 'inf', since these represent the variances of the latent and observed variables. Failure to do so will result in an error message.

v. All bi-directional connections should be specified in the symmetric matrices on BOTH sides of diagonal. Failure to match connections on both sides of diagonal will produce an error message.

vi. All uni-directional connections should be specified in asymmetric matrix. Any uni-directional connections should NOT be reflected on other side of diagonal, else an error message will result. If connection from variable 'x' to variable 'y' should be specified, the cell (y,x) should be selected – NOT (x,y).

vii. All factor loadings, since they are uni-directional, should be specified in the asymmetric matrix.

Model estimation

i. The 'Load' button is used to load data. The file extension can be '.txt', '.mat' or any other file format that can be imported into MATLAB. The data file should be an 'm x n' matrix, where 'm' is the number of subjects and 'n' is the number of observed variables. When the file is selected for loading, a message box will appear stating that the variable 'data' already exists on the workspace and asking if you would like to overwrite it – click 'yes'.

ii. Before estimating parameters, check if multivariate normality assumption of observed variables is met. Warning is produced if assumption is violated.

iii. Three different discrepancy functions to measure difference between 'model-implied covariance' matrix and 'data covariance matrix' are given – ML, GLS and ULS. ML is by

far the most reliable and is recommended. **Normal and Bootstrap Standard errors are calculated ONLY for ML.**

iv. After pressing 'Calculate' under the Parameter estimation option, a text box appears with initial guess of all parameters as 1. This is a safe initial guess but can be modified if you have good reason to do so.

v. The order of parameters are as follows – the parameters from symmetric matrix are first, followed by those from asymmetric matrix. For both these matrices, the parameters are listed columnwise i.e. all parameters from column 1, then all parameters from column 2 etc. For symmetric matrix, only bottom triangle of matrix is read, because, matrix is same on both sides of diagonal – order of parameters will have to take this into account.

vi. **Once parameters are estimated, one MUST check if model is identified.** If model is not identified, parameters of model are not reliable. Error message is produced if model is not identified.

vii. Once model identification is done, standard errors and CIs can be estimated using an approximate or exact method. Approximate method will take less than a minute, Exact method will be in the order of hours – around 10 hours or more. There is an appreciable difference in accuracy between estimates of approximate and exact methods, so 'exact' SEs and CIs should be estimated if possible.

Model evaluation

i. Overall model fit can be assessed using the 'Goodness-of-fit' option. This calculates 6 different measures of model fit. For chi-square and RMSEA, p-value (2^{nd} row) and 90% CIs (lower bound, 2^{nd} row and upper bound, 3^{rd} row) are also provided respectively.

ii. Standardised residuals, showing difference between model-implied covariance matrix and data-covariance matrix, can be visualised. Labels of the observed variables can be input to aid interpretation.

iii. Lagrange multipliers and Wald indices are yet to be coded.

Other

i. Reset button can be pressed to start over. Possible reasons to press reset button are: to correct a mistake in model specification, to specify a new model etc.

ii. All relevant variables (model specification, model parameters, standard errors, GOF measures etc.) are in the workspace, so can be saved for future reference.

iii. **The toolbox only allows one SEM model to be analysed at a time.**

iv. **Functions are provided to carry out entire fitting procedure: sem_fit() carries out fitting and estimates approximate standard errors, sem_fit_bt() carries out fitting and estimates exact standard errors.**

v. The file 'data.txt' is a sample dataset and can be used to fit a model in Section 5.11 in the document 'SEM models.pdf'. The results are in 'results.txt'.