Chapter 4
Issues in Meta-Analytic Structural Equation Modeling

Abstract This chapter provides short overviews of unresolved issues in MASEM. The first part of this chapter describes software that can be used to conduct MASEM using TSSEM, the GLS-method and the univariate method. The metaSEM-package is very useful for MASEM. Analyses using this package are shown in the last two chapters of this book. The second issue is about the use of different fit-indices to evaluate the homogeneity of correlation matrices at Stage 1 of TSSEM. The third issue is about handling missing correlations in specific studies. The basic approach is to delete a variable that is associated with a missing correlation, but more efficient methods are possible. The last issue is about a recent adaptation to the existing MASEM approach that may have advantages for handling heterogeneity. The adaptation involves a Stage 2 analysis based on a multigroup model.

Keywords Meta-analytic structural equation modeling · Software · MetaSEM · OpenMx · Fit-indices · Maximum likelihood · Missing correlations

4.1 Software to Conduct MASEM

In principle, all structural equation modeling software can be used to perform meta-analytic structural equation modeling. However, it may involve some complex programming to set up the right model. The easiest way to perform TSSEM is to use the dedicated R-package metaSEM (Cheung 2015a). It requires some basic knowledge of the R-program (see below), but the package itself is quite user friendly. It includes functions to fit the fixed effects Stage 1 model, the random effects Stage 1 model, and to fit the Stage 2 model to the pooled correlation matrix from Stage 1. The package includes several convenient functions to read in the data and to extract parts of the output. It also includes all functions to do standard meta-analysis. Cheung (2015b) gives an overview of the many possibilities with the metaSEM-package.

S. Jak, *Meta-Analytic Structural Equation Modelling*,
SpringerBriefs in Research Synthesis and Meta-Analysis,
DOI 10.1007/978-3-319-27174-3_4

Fixed effects MASEM based on the GLS approach can also be performed using the metaSEM-package by constraining the random effects to be zero in the random effects function, but the function uses maximum likelihood estimation. I added an example of the original GLS-approach using R on my website (http://suzannejak.nl/masem).

As the multivariate methods are found to perform better than the univariate methods (see Chap. 2), it is not recommended to perform MASEM using the univariate methods. If one still wants to use them, one could in principle use any meta-analysis program to pool the correlation coefficients in Stage 1, and use any structural equation modeling program to fit the Stage 2 model. In order to pool the correlation coefficients, the R-packages 'metafor' (Viechtbauer 2010) and 'metaSEM' (Cheung 2015a) are very useful. David Wilson (Lipsey and Wilson 2001) has written macros for SPSS, SAS, and STATA to carry out univariate meta-analysis. The macro's are available from his website: (http://mason.gmu.edu/~dwilsonb/ma.html). Several other commercial software programs exist. See Bax et al. (2007) for a comparison of several programs.

For Stage 2 you need a SEM-program. Freely available software packages to conduct structural equation modeling are the R-packages Lavaan (Rosseel 2012) and OpenMx (Boker et al. 2011). In addition there are commercial programs such as Mplus (Muthén and Muthén 2012) and Lisrel (Jöreskog and Sörbom 1996). For the Stage 2 analysis with WLS-estimation, OpenMx and Lisrel are most suitable, as Mplus and Lavaan cannot read in the weight matrix in addition to the pooled correlation matrix.

The freely available programs are packages in R. Therefore, in order to conduct MASEM it is very convenient to be familiar with the R-program. R is a free software environment for statistical computing and graphics. Learning R may be a bit daunting in the beginning, but soon will pay back the effort. To get started with R, several manuals can be found under the contributed documentation on www.r-project.org. For example, these two documents provide a short overview of R (and explain how to install R), and will provide you with enough R-knowledge to be able to use the metaSEM package.

– Marthews, D. (2014). The friendly beginners' R course. http://cran.r-project.org/other-docs.html. Accessed 08 Jan 2015.
– Paradis, E. (2005). R for Beginners. http://cran.r-project.org/other-docs.html. Accessed 08 Jan 2015.

The metaSEM-package uses OpenMx in the background to fit all models. OpenMx is a package in R that can be used for structural equation modeling. OpenMx is very flexible, because the user can use all possibilities of the R-programming environment. This makes OpenMx a suitable program to use in the specification of meta-analytic structural equation models. Because for the MASEM researcher it may be useful to understand OpenMx, I included annotated examples of fitting a path model and a factor model in OpenMx in Appendices B and D.

4.2 Fit-Indices in TSSEM

The chi-square measure of fit can be used in Stage 1 to test the homogeneity of correlation matrices across samples. The chi-square test has as the null hypothesis that the model holds exactly in the population, so all differences between the observed and population matrices are due to sampling. In structural equation modeling it is common to look at measures of approximate or relative fit as well. The Root Mean Squared Error or Approximation (RMSEA, Steiger and Lind 1980) for example, is a measure of approximate fit. The RMSEA is based on the idea that models are approximations to reality and do not have to reflect reality perfectly (MacCallum 2003). If a researcher uses the RMSEA to evaluate the fit of a Stage 1 model in MASEM, he or she implicitly assumes that homogeneity does not have to hold exactly but only approximately. However, it is unclear how much deviation from homogeneity is acceptable when fitting the Stage 2 model under a fixed effects model. At some point, the parameters in the Stage 2 model will become biased and confidence intervals may become too small. Research using simulated data, varying for example the amount and type of heterogeneity (heterogeneity in one or all correlation coefficients), would be needed to evaluate the RMSEA values that are associated with unacceptable heterogeneity.

The CFI is based on a comparison of the fit of the specified model with the fit of the independence model, which is a model in which all variables are assumed to be independent. The CFI strongly depends on the size of the observed correlations. The lower the observed correlations, the better the independence model will fit the data, the lower CFI will be. Because the size of the correlations should not play a role in evaluating heterogeneity, I expect that the CFI is not very useful to evaluate the homogeneity of correlation coefficients in MASEM.

The Standardized Root Mean Squared Residual (SRMSR) is based on the differences between the observed and model implied correlation coefficients. Larger differences between the correlation coefficients will lead to a larger SRMSR, so the SRMSR seems to be useful to evaluate homogeneity at Stage 1. However, just as with the RMSEA, simulation research is needed to evaluate the critical SRMSR values associated with unacceptable heterogeneity.

4.3 Missing Correlations in TSSEM

In fixed effects two-stage SEM, it is no problem when some studies do not include all relevant variables. The missing variables will just be filtered out in the analysis. It is a problem if there are missing correlations for variables that are included in the study. Ideally, researchers always report the correlations between all variables in their study. However, often not all correlations between the research variables are given in a paper. Sometimes, the missing correlations can be derived from other statistics the authors do provide, such as regression coefficients. This is not always possible, for example when two variables are both outcome variables in regression analyses. In the random effects Stage 1 analysis, missing correlations are not a problem, but in the fixed effects analysis they are. As a consequence, for each missing correlation, one of the two variables associated with the correlation has to be treated as missing. Preferably, one would delete the variable with the least remaining correlations with other variables.

Methods to handle missing correlation coefficients in TSSEM more efficiently have been proposed by Jak et al. (2013) and Cheung (2014). Both methods are based on the idea of fixing the missing correlations at some appropriate value (a value that does not lead to a non-positive definite correlation matrix), for example at zero, and estimating an extra parameter for each missing correlation. This way, the fixed values for the missing correlations do not affect the results, and all correlations that are present are used in the analyses. These methods are not implemented in the metaSEM package yet. So, in order to use these methods one will have to specify the needed models in OpenMx directly, or use the program to generate syntax to conduct fixed effects TSSEM with Lisrel (Cheung 2009). A possible problem with this approach is that the fit of the independence model may not be appropriate anymore due to the fixed zeros in the observed correlation matrices (Cheung 2015b). The fit of the independence model is used when calculating some fit-indices, like the CFI. However, the problem of the missing correlations plays a role in Stage 1 of the analysis, and as discussed earlier, the CFI may not be the most appropriate fit measure to evaluate the homogeneity of correlation matrices.

4.4 The ML-Approach to MASEM

A recent alternative to estimating the Stage 2 model in the two-stage approach is to use a maximum likelihood (ML) approach (Oort and Jak 2015). In this approach, multigroup analysis is used for all models. The test of homogeneity of correlation matrices (Stage 1) is identical to TSSEM. The difference lies in fitting the structural model. In the ML-approach, a common $\mathbf{R}_{\text{MODEL}}$ is fitted to the observed matrices or all studies, where $\mathbf{R}_{\text{MODEL}}$ may have the structure of any structural equation model. For example, if one would fit a factor model in Stage 2, the model for each study i would be:

$$\begin{aligned} &\boldsymbol{\Sigma}_i = \mathbf{D}_i\left(\mathbf{X}_i\mathbf{R}_{\text{MODEL}}\mathbf{X}_i^{\text{T}}\right)\mathbf{D}_i, \\ &\text{with} \\ &\mathbf{R}_{\text{MODEL}} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}. \end{aligned} \tag{4.1}$$

Here, $\mathbf{D}_i$ and $\mathbf{X}_i$ are the diagonal and selection matrices defined in Chap. 2, $\boldsymbol{\Lambda}$ is a matrix of factor loadings, $\boldsymbol{\Phi}$ is a matrix with factor variances and covariances, and $\boldsymbol{\Theta}$ is a matrix with residual variances (and covariances). Because $\mathbf{R}_{\text{MODEL}}$ is a restriction of $\mathbf{R}$ in the Stage 1 model, the difference between the associated chi-square values has a chi-square distribution itself with degrees of freedom equal to the difference in the numbers of free parameters in $\mathbf{R}$ and $\mathbf{R}_{\text{MODEL}}$. Oort and

Jak (2015) used simulated data to show that using maximum likelihood estimation in both stages of meta-analysis through SEM leads to almost identical results as using WLS-estimation in Stage 2 of the analysis. The differences in estimation bias, power rate and Type 1 error rates were not consistent and hardly noticeable.

There are some fundamental and practical differences which may guide a researcher's choice between the two methods. Advantages of the ML procedure are that the same estimation method is used at both stages, and that the Stage 1 and Stage 2 models are nested. The ML-procedure may also provide more flexibility in the application of equality constraints across studies in the structural model. In principle, some Stage 2 parameters could be set equal across a subset of studies, another parameter could be set equal across another subset of studies and other parameters could be freely estimated in all studies. Disadvantage of the ML-approach are that it is currently limited to fixed effects models, and that no readily available software package to apply the method exists. The WLS-procedure has practical advantages. In the WLS procedure, the Stage 2 model is not a multi-group model, so that estimation convergence is much faster than in the ML-approach. The necessity to calculate a weight matrix (the inverse of the matrix of asymptotic variances and covariances of the pooled correlation coefficients) may count as a disadvantage of the WLS method, but fortunately the readily available R package metaSEM takes this burden off the user's hands. As a result, the WLS-approach may actually be easier to take than the ML-approach.

References

Bax, L., Yu, L.-M., Ikeda, N., & Moons, K. G. M. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology, 7*, 40. doi:10.1186/1471

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., & Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76*(2), 306–317.

Cheung, M. W.-L. (2009). TSSEM: A LISREL syntax generator for two-stage structural equation modeling (Version 1.11) [Computer software and manual].

Cheung, M. W.-L. (2014). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46*, 29–40.

Cheung, M. W.-L. (2015a). MetaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*, 1521. doi:10.3389/fpsyg.2014.01521

Cheung, M. W.-L. (2015b). *Meta-analysis: A structural equation modeling approach*. Chichester, UK: Wiley.

Jak, S., Oort, F. J., Roorda, D. L., & Koomen, H. M. Y. (2013). Meta-analytic structural equation modelling with missing correlations. *Netherlands Journal of Psychology, 67*(4), 132–139.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: Users' reference guide*. Chicago: Scientific Software International.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models. *Multivariate Behavioral Research, 38*, 113–139. doi:10.1207/S15327906MBR3801_5

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Oort, F. J., & Jak, S. (2015). Maximum likelihood estimation in meta-analytic structural equation modeling. 10.6084/m9.figshare.1292826

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

Steiger, J. H., & Lind, J. M. (1980). *Statistically-based tests for the number of common factors*. Paper presented at Psychometric Society Meeting, Iowa City, Iowa.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.