metaSEM: An R Package for Meta-Analysis Using Structural Equation Modeling

Mike W.-L. Cheung
National University of Singapore

July 28, 2012

Abstract

The **metaSEM** package provides functions to conducting univariate and multivariate meta-analysis using a structural equation modeling approach via the **OpenMx** package. It also implemented the two-stage structural equation modeling (TSSEM) approach to conducting fixed- and random-effects meta-analytic structural equation modeling (MASEM) on correlation or covariance matrices. This paper outlines the basic theories of its implementations. Examples are used to illustrate the procedures.

Keywords:~meta-analysis, structural equation modeling, meta-analytic structural equation modeling, **metaSEM**, R.

1. Introduction

Meta-analysis is a popular technique to synthesizing research findings in social, behavioral, educational and medical sciences (e.g., Borenstein, Hedges, Higgins, and Rothstein 2009; Hedges and Olkin 1985; Hunter and Schmidt 2004; Whitehead 2002). There are several standalone packages for meta-analysis, e.g., Comprehensive Meta-Analysis and RevMan. Many standard statistical packages, for instance, SPSS (Lipsey and Wilson 2000), SAS (Arthur, Bennett, and Huffcutt 2001) and STATA (Sterne 2009), have macros or packages to fitting some meta-analytic models. There are also several packages to conducting meta-analysis in R, for instance, **meta** (Schwarzer 2012), **rmeta** (Lumley 2009), **mvmeta** (Gasparrini 2012), **metaLik** (Guolo and Varin 2012) and **metafor** (Viechtbauer 2010).

The **metaSEM** package is yet another R package to conducting univariate and multivariate meta-analysis. It formulates meta-analytic models as structural equation models (Cheung 2008, 2011b) via the OpenMx package (Boker, Neale, Maes, Wilde, Spiegel, Brick, Spies, Estabrook, Kenny, Bates, Mehta, and Fox 2011). It also implemented the two-stage structural equation modeling (TSSEM) approach (Cheung and Chan 2005b, 2009) to conducting fixed- and random-effects meta-analytic structural equation modeling (MASEM) on correlation or covariance matrices. The main functions in this package are:

- `meta()` and `reml()`: `meta()` fits univariate and multivariate meta-analysis with maximum likelihood (ML) estimation method while `reml()` estimates the variance components of the random-effects with restricted (residual) maximum likelihood (REML) estimation method. Mixed-effects meta-analysis can be fitted by using study characteristics as predictors. Equality constraints on the intercepts, regression coefficients and variance components can be imposed.

- `meta3()` and `reml3()`: They fit three-level meta-analysis by considering cluster effect. `meta3()` fits the three-level meta-analysis with ML estimation method while `reml3()` estimates the variance components with REML estimation method.

- `tssem1()` and `tssem2()`: `tssem1()` fits the first stage analysis of TSSEM by pooling correlation or covariance matrices with either a fixed- or random-effets model. `tssem2()`, which is a wrapper of `wls()`, fits the second stage analysis of TSSEM by fitting structural models on the pooled correlation/covariance matrix.

- `wls()`: It fits a correlation or covariance structure analysis with weighted least squares (WLS) estimation method.

Besides reporting Wald confidence intervals (CIs) based on z statistic, likelihood-based CIs on the parameter estimates may also be requested (Cheung 2009a; Neale and Miller 1997). Some generic functions, such as `anova()`, `coef()`, `vcov()`, `print()`, `summary()` and `plot()`, have also been implemented.

This paper was based on the **metaSEM** package version 0.7-1 and the **OpenMx** package version 1.2.4-2063. The paper is organized as follows. The next section presents basic ideas on structural equation models and how they are linked to meta-analytic models. Basic theory on the TSSEM are then presented. Examples are used to illustrate these procedures. How to install the package is finally discussed.

2. Structural Equation Modeling Based Meta-Analysis

In this section, basic structural equation models are introduced. Univariate and multivariate meta-analysis are formulated as special cases of SEM (see Cheung 2008, 2011b).

2.1. Structural equation model

Structural equation modeling is a multivariate technique to fitting and testing hypothesized models. Let $\mathbf{y}$ be a $p \times 1$ vector of the sample data where p is the number of variables. It is hypothesized that the model for the first and second moments are $\mu = \mu(\theta)$ and $\mathbf{\Sigma} = \mathbf{\Sigma}(\theta)$, respectively, where θ is a vector of parameters.

The -2*log-likelihood of the ith case is:

$$-2 * logL_i(\theta; \mathbf{y}_i)_{\mathrm{ML}} = p_i * log(2\pi) + log|\mathbf{\Sigma_i}(\theta)| + (\mathbf{y}_i - \mu_{\mathbf{i}}(\theta))' \mathbf{\Sigma_i}(\theta)^{-1} (\mathbf{y}_i - \mu_{\mathbf{i}}(\theta)), \quad (1)$$

where p_i is the number of filtered variables in the ith case, $\mu_\mathbf{i}(\theta)$ and $\mathbf{\Sigma_i}(\theta)$ are the model implied mean vector and the model implied covariance matrix for the ith case, respectively. Since there is a subscript i in these quantities, the model implied mean vector and covariance matrix may vary across cases. In order words, this model handles incomplete data automatically by selecting only the complete variables in the log-likelihood function.

To obtain the parameter estimates, we may take the sum of the -2*log-likelihood over all cases and minimize it. This is known as the full information maximum likelihood (FIML or simply ML) estimation method. Iterative methods are used to obtain the solutions. If it is convergent, the asymptotic covariance matrix (thus the standard errors) of the parameter estimates may be obtained from the inverse of the Hessian matrix. The parameter estimates divided by their standard errors (*SE*s) follow a z distribution under the null hypothesis. Likelihood ratio statistic may also be used to compare nested models.

2.2. Univariate fixed-effects model

Let us assume that there is only one effect size y_i in the ith study. y_i can be any effect size, such as odds ratio, raw mean difference, standardized mean difference, correlation coefficient and its Fisher's z transformed score. When the sample sizes are reasonably large, y_i can be assumed to be normally distributed with a variance of v_i (see e.g., Borenstein *et ˜al.* 2009, for the formulas of common effect sizes). The univariate fixed-effects model for the ith study is:

$$y_i = \beta_{\text{fixed}} + e_i, \tag{2}$$

where β_{fixed} is the common effect under a fixed-effects model and $\text{var}(e_i) = v_i$ is the known sampling variance.

To fit the unvariate fixed-effects meta-analysis in SEM, we may use the following model:

$$\mu_i(\theta) = \beta_{\text{fixed}} \tag{3}$$

and

$$\mathbf{\Sigma}_i(\theta) = v_i \tag{4}$$

Since v_i is known, the only parameter in the SEM is β_{fixed} .

2.3. Univariate random-effects model

A random-effects model allows studies to have their own study specific effect. The model for the ith study is:

$$y_i = \beta_{\text{random}} + u_i + e_i. \tag{5}$$

where β_{random} is the average effect under a random-effects model and $\text{var}(u_i) = \tau^2$ is the heterogeneity variance that has to be estimated. To fit the unvariate random-effects meta-analysis in SEM, we may use the following model:

$$\mu_i(\theta) = \beta_{\text{random}} \tag{6}$$

and

$$\boldsymbol{\Sigma}_i(\theta) = \tau^2 + v_i \tag{7}$$

v_i and $\tau^2 + v_i$ are known as the conditional and the unconditional variances in the literature of meta-analysis, respectively. In this model we have to estimate both β_{random} and τ^2.

Check the formulas

Quantifying heterogeneity Although the Q statistic (Cochran 1954) may be used to test the null hypothesis of homogeneity of effect sizes, it does not indicate the degree of heterogeneity. One popular index to quantifying the heterogeneity of effect sizes is the I^2 proposed by Higgins and Thompson (2002). The general formula is:

$$I^2 = \frac{\hat{\tau^2}}{\hat{\tau^2} + \tilde{v}} \tag{8}$$

where $\tilde{v}$ is the *typical* within-study sampling variance.

Takkouche, Cadarso-Suárez, and Spiegelman (1999) suggested to use the harmonic mean of v_i, i.e.,

$$\tilde{v}_{HM} = \frac{1}{\sum_{i=1}^{k} v_i}. \tag{9}$$

Higgins and Thompson (2002) preferred to define the *typical* within-study sampling variance based on the Q statistic:

$$\tilde{v}_Q = \frac{(k-1)\sum_{i=1}^{k} v_i}{(\sum_{i=1}^{k} v_i)^2 - \sum_{i=1}^{k} v_i^2}. \tag{10}$$

One of the main reasons of using this *typical* within-study sampling variance is that the I^2 can be simplied as

$$I^2_Q = Q - (k-1)/Q. \tag{11}$$

Besides these two estimators, Xiong, Miller, and Morris (2010) also discussed an estimate on I^2 that is based on the arithmetic mean, that is,

$$\tilde{v}_{AM} = \sum_{i=1}^{k} v_i/k. \tag{12}$$

2.4. Univariate mixed-effects model

The mixed-effects meta-analysis extends the random-effects meta-analysis by using study characteristics as predictors. Assuming that $\mathbf{x}_i$ is a $(m+1) \times 1$ vector of predictors including a constant of 1 where m is the number predictors in the ith study, the mixed-effects model is:

$$y_i = \beta' \mathbf{x}_i + u_i + e_i, \tag{13}$$

where β is a vector of regression coefficients including the intercept. To fit the univariate mixed-effects meta-analysis in SEM, we may use the following model for the conditional mean and variance:

$$\mu_i(\theta|\mathbf{x}_i) = \beta'\mathbf{x}_i \tag{14}$$

and

$$\mathbf{\Sigma}_i(\theta|\mathbf{x}_i) = \tau^2 + v_i. \tag{15}$$

Since $\mathbf{x}_i$ is specified via definition variables (Boker *et˜al.* 2011), the means and covariance matrix of $\mathbf{x}$ are not estimated. $\mathbf{x}$ is treated as a design matrix rather than random variables. This approach is consistent with conventional meta-analysis.

Explained variance The percentage of explained variance by the inclusion of predictors can be calculated by comparing the $\hat{\tau}_0^2$ without predictor and the $\hat{\tau}_1^2$ with predictors (Raudenbush 2009):

$$R^2 = \frac{\hat{\tau}_0^2 - \hat{\tau}_1^2}{\hat{\tau}_0^2}. \tag{16}$$

By definition, R^2 is non-negative. When the calculated R^2 is negative, it is truncated to zero.

2.5. Multivariate mixed-effects model

Let us assume that there are p effect sizes with m predictors in k studies. The model for the multivariate effect sizes in the ith study is:

$$\mathbf{y}_i = \mathbf{B}\mathbf{x}_i + \mathbf{u}_i + \mathbf{e}_i, \tag{17}$$

where $\mathbf{y}_i$ is a $p \times 1$ effect sizes, $\mathbf{B}$ is a $p \times (m+1)$ regression coefficients including the intercepts, $\mathbf{x}_i$ is a $(m+1)\times 1$ predictors including 1 in the first column, $\mathbf{u}_i$ is a $p\times 1$ study specific random effects, and $\mathbf{e}_i$ is a $p \times 1$ sampling error. We assume that $\text{var}(\mathbf{e}_i) = \mathbf{V}_i$ is known and given in the ith study and $\text{var}(\mathbf{u}_i) = \mathbf{T}^2$ is the variance component of the between-study heterogeneity that has to be estimated.

The -2*log-likelihood of the above model is:

$$-2*logL_i(\mathbf{B}, \mathbf{T}^2; \mathbf{y}_i)_{\text{ML}} = p_i*log(2\pi) + log|\mathbf{T}^2+\mathbf{V}_i| + (\mathbf{y}_i - \mathbf{B}\mathbf{x}_i)'(\mathbf{T}^2+\mathbf{V}_i)^{-1}(\mathbf{y}_i - \mathbf{B}\mathbf{x}_i), \tag{18}$$

where p_i is the number of effect sizes in the ith study.

In applied research, different studies may report different effect sizes, that is, p_i may vary across studies. The above -2*log-likelihood may handle missing effect sizes by using different dimenions of the elements in the above equation. It is expected that there is no missing data in $\mathbf{x}_i$. When there are missing data in $\mathbf{x}_i$, the whole study will be deleted before the analysis.

The I^2 and R^2 in univariate meta-analysis are also calculated for each effect size in multivariate meta-analysis.

2.6. Restricted Maximum Likelihood (REML) Estimation Method

Since both the fixed- and random-effects are estimated simultaneously, it is well-known that $\hat{\mathbf{T}}^2_{\mathrm{ML}}$ based on the ML estimation in 18 is negatively biased. It is because it does not take the uncertainty in estimating $\hat{\mathbf{B}}_{\mathrm{ML}}$ into account. If the unbiasness of the variance component is crucial in the research questions, we may obtain the variance component $\hat{\mathbf{T}}^2_{\mathrm{REML}}$ based on the REML estimation method (Cheung 2011a; Harville 1977; Patterson and Thompson 1971).

The -2*log-likelihood of the model is:

$$-2{*}logL_i(\mathbf{T}^2;\mathbf{y}_i)_{\mathrm{REML}} = p_i{*}log(2\pi)+log|\mathbf{T}^2+\mathbf{V}_i|+(\mathbf{y}_i-\alpha\mathbf{X}_i)'(\mathbf{T}^2+\mathbf{V}_i)^{-1}(\mathbf{y}_i-\alpha\mathbf{X}_i)+|\mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i|, \tag{19}$$

where $\alpha = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, and $\mathbf{X}$, $\mathbf{V}$ and $\mathbf{y}$ are the stacked matrices and vector of the correspondent matrices and vectors.

Since the fixed effects $\mathbf{B}$ is not involved in the above -2*log-likelihood function, it can be calculated in a second stage.

2.7. Three-level meta-analysis

Observed effect sizes may be related or dependent. For example, effect sizes reported by the same research team may be more similar when comparing to effect sizes reported by other research teams. Effect sizes reported by studies from the same country may be more similar when comparing to studies across countries. If the degree of dependence is known, multivariate meta-analysis introduced before may be applied. When the degree of dependence is unknown, a three-level meta-analytic model may be used (e.g., Konstantopoulos 2011). The model is:

$$y_{ij} = \beta'\mathbf{x}_{ij} + u_{(2)ij} + u_{(3)j} + e_{ij}, \tag{20}$$

where y_{ij} is the effect size for the ith study at the jth cluster, β is a vector of regression coefficients including the intercept, $\mathbf{x}_{ij}$ is the predictors including 1 in the first element for the ith study at the jth cluster, $u_{(2)ij}$ and $u_{(3)j}$ are the random-effects at level-2 and level-3, respectively, and e_{ij} is the known sampling variance of the effect size.

To fit the three-level meta-analytic model in SEM, we may use the following model for the conditional mean and variance:

$$\mu_{ij}(\theta|\mathbf{x}_{ij}) = \beta'\mathbf{x}_{ij} \tag{21}$$

and

$$\boldsymbol{\Sigma}_{ij}(\theta|\mathbf{x}_{ij}) = \tau^2_{(2)} + \tau^2_{(3)} + v_{ij}. \tag{22}$$

where $\tau^2_{(2)} = \mathrm{var}(u_{(2)ij})$ and $\tau^2_{(3)} = \mathrm{var}(u_{(3)j})$ are the heterogeneity at level-2 and level-3, respectively. Similar to that listed in Eq. 19, $\tau^2_{(2)}$ and $\tau^2_{(3)}$ may also be estimated with REML estimation method. This approach was implemented in `reml3()`.

Quantifying heterogeneity and explained variance Similar to the I^2 defined in Eq. 8 for random-effects meta-analysis, we may define the degree of heterogeneity for three-level meta-analysis in level 2 and level 3 as:

$$I^2_{(2)} = \frac{\hat{\tau}^2_{(2)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{v}} \tag{23}$$

and

$$I^2_{(3)} = \frac{\hat{\tau}^2_{(3)}}{\hat{\tau}^2_{(2)} + \hat{\tau}^2_{(3)} + \tilde{v}} \tag{24}$$

where $\tilde{v}$ is the *typical* within-study sampling variance defined in Eqs. 9, 10 and 12.

When there are predictors, we may calculate the R^2 for the level 2 and level 3 similar to that defined in Eq. 16:

$$R^2_{(2)} = \frac{\hat{\tau}^2_{0(2)} - \hat{\tau}^2_{1(2)}}{\hat{\tau}^2_{0(2)}}. \tag{25}$$

and

$$R^2_{(3)} = \frac{\hat{\tau}^2_{0(3)} - \hat{\tau}^2_{1(3)}}{\hat{\tau}^2_{0(3)}}. \tag{26}$$

When the estimates are negative, they are truncated to zero.

2.8. Examples

Three example data sets are used to demonstrate the procedures of fitting univariate and multivariate meta-analyses. The first data set was taken from Becker (1983) who reported 10 studies on sex differences in conformity using the fictitious norm group paradigm. *di* and *vi* are the standardized mean difference and its sampling variance, respectively. Becker hypothesized that the logarithm of the number of items (*items*) predicted the effect size.

The second data set is adapted from Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, and Colditz (1998). They summarized five published trials comparing surgical and non-surgical treatments for medium-severity periodontal disease, one year after treatment. Publication year *pub_year* was hypothesized as a predictor.

The third data set was reported by Konstantopoulos (2011) that was based on Cooper, Valentine, Charlton, and Melson (2003). It described fifty-six effect sizes clustered in 11 districts. Year of publication was used as a predictor.

Univariate random-effects model The function `meta()` is used to conduct the analyses. The arguments `y` and `v` are used to specify the effect sizes and its sampling variances (and covariances for multivariate meta-analysis), respectively. By default, a random-effects meta-analysis is fitted. After running the analysis, `summary()` is used to extract the results. The estimated fixed- and random-effects are represented by the `Intercept1` to `Intercept`p and by the `Tau2_1_1` to `Tau2_`p`_`p parameters where p is the number of effect sizes, especially. `coef()` and `vcov()` may be used to extract the parameter estimates and their asymptotic sampling covariance matrix, respectively.

From the following analyses, the Q statistic ($df = 9$) is 30.65, $p < .001$. The estimated heterogeneity variance is 0.0774 while the I^2 based on the Q statistic is .6718. The average effect size with its 95% Wald confidence interval (CI) based on the random-effects model is 0.1747 (-0.0475, 0.3970).

```
R> ## Load the library
R> library(metaSEM)
R> ## Show the first few studies of the data set
R> head(Becker83)

  study    di   vi percentage items
1     1 -0.33 0.03         25     2
2     2  0.07 0.03         25     2
3     3 -0.30 0.02         50     2
4     4  0.35 0.02        100    38
5     5  0.69 0.07        100    30
6     6  0.81 0.22        100    45

R> ## Random-effects meta-analysis with ML
R> summary( random1 <- meta(y=di, v=vi, data=Becker83) )

Running Meta analysis with ML

Call:
meta(y = di, v = vi, data = Becker83)

95% confidence intervals: z statistic approximation
Coefficients:
            Estimate Std.Error    lbound    ubound z value
Intercept1  0.174734  0.113378 -0.047482  0.396950  1.5412
Tau2_1_1    0.077376  0.054108 -0.028674  0.183426  1.4300
           Pr(>|z|)
Intercept1   0.1233
Tau2_1_1     0.1527

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Heterogeneity indices (based on the estimated Tau2):
                             Estimate
```

```
Intercept1: I2 (Q statistic)   0.6718

Number of studies (or clusters): 10
Number of observed statistics: 10
Number of estimated parameters: 2
Degrees of freedom: 8
-2 log likelihood: 7.928307
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Univariate mixed-effects model We may include a predictor to conduct a mixed-effects meta-analysis. The argument `x` is used to specify the predictors. If there are more than one predictor, `cbind()` is used to specify them. The estimated regression coefficients are represented by the `Slope`i_j parameter where i and j represent the ith effect size and jth predictor. The following analysis indicates that $log(items)$ is a significant predictor with the estimated regression coefficient and its 95% CI of 0.2109 (0.1225, 0.2992) with $R^2 = 1$.

```
R> ## Mixed-effects meta-analysis with "log(items)" as the predictor
R> summary( meta(y=di, v=vi, x=log(items), data=Becker83) )

Running Meta analysis with ML

Call:
meta(y = di, v = vi, x = log(items), data = Becker83)

95% confidence intervals: z statistic approximation
Coefficients:
              Estimate   Std.Error      lbound      ubound
Intercept1 -3.2015e-01  1.0981e-01 -5.3539e-01 -1.0492e-01
Slope1_1    2.1088e-01  4.5084e-02  1.2251e-01  2.9924e-01
Tau2_1_1    1.0000e-10  2.0095e-02 -3.9386e-02  3.9386e-02
           z value  Pr(>|z|)
Intercept1 -2.9154  0.003552 **
Slope1_1    4.6774 2.905e-06 ***
Tau2_1_1    0.0000  1.000000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Explained variances (R2):
                           Estimate
```

```
y1: Tau2 (no predictor)      0.0774
y1: Tau2 (with predictors)   0.0000
y1: R2                       1.0000

Number of studies (or clusters): 10
Number of observed statistics: 10
Number of estimated parameters: 3
Degrees of freedom: 7
-2 log likelihood: -4.208024
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Univariate fixed-effects model Mathematically, fixed-effects meta-analysis is a special case of the random-effects meta-analysis by fixing the heterogeneity variance of the random-effects at 0. The argument `RE.constraints`, which expects a matrix as input, is used to constrain the variance component of the random effects.

```
R> ## Fixed-effects meta-analysis
R> summary( meta(y=di, v=vi, data=Becker83,
                RE.constraints=matrix(0, ncol=1, nrow=1)) )

Running Meta analysis with ML

Call:
meta(y = di, v = vi, data = Becker83, RE.constraints = matrix(0,
    ncol = 1, nrow = 1))

95% confidence intervals: z statistic approximation
Coefficients:
            Estimate Std.Error    lbound    ubound z value
Intercept1  0.100640  0.060510 -0.017957  0.219237  1.6632
           Pr(>|z|)
Intercept1  0.09627 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Heterogeneity indices (based on the estimated Tau2):
                             Estimate
Intercept1: I2 (Q statistic)        0

Number of studies (or clusters): 10
```

```
Number of observed statistics: 10
Number of estimated parameters: 1
Degrees of freedom: 9
-2 log likelihood: 17.86043
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Multivariate random-effects model Multivariate meta-analysis can be fitted by specifying the multivariate effect sizes and its sampling covariance matrix in the arguments `y` and `v` with `cbind()`, respectively. Only the lower triangle of the sampling covariance matrix arranged by the column major is used in `v`. For example, if $\mathbf{V} = \begin{bmatrix} V_{11} & & \\ V_{21} & V_{22} & \\ V_{31} & V_{32} & V_{33} \end{bmatrix}$, we may use `v=cbind(V11,V21,V31,V22,V32,V33)` in `meta()`. The Q statistic ($df = 8$) of the following example is 128.2267, $p < .001$. The estimated variance component is $\begin{bmatrix} 0.0070 & \\ 0.0095 & 0.02614 \end{bmatrix}$. The I^2 based on the Q statistic for PD and AL are .6021 and .9250, respectively. The pooled effect sizes with their 95% Wald CIs based on the random-effects model for PD and AL are 0.3448 (0.2397, 0.4500) and -0.3379 (-0.4972, -0.1787), respectively.

```
R> ## Show the data set
R> Berkey98

  trial pub_year no_of_patients   PD    AL var_PD cov_PD_AL
1     1     1983             14 0.47 -0.32 0.0075    0.0030
2     2     1982             15 0.20 -0.60 0.0057    0.0009
3     3     1979             78 0.40 -0.12 0.0021    0.0007
4     4     1987             89 0.26 -0.31 0.0029    0.0009
5     5     1988             16 0.56 -0.39 0.0148    0.0072
  var_AL
1 0.0077
2 0.0008
3 0.0014
4 0.0015
5 0.0304

R> ## Multivariate meta-analysis with a random-effects model
R> summary( meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL),
                data=Berkey98) )

Running Meta analysis with ML

Call:
```

```
meta(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    data = Berkey98)

95% confidence intervals: z statistic approximation
Coefficients:
             Estimate  Std.Error     lbound     ubound
Intercept1  0.3448392  0.0536312  0.2397239  0.4499544
Intercept2 -0.3379381  0.0812479 -0.4971812 -0.1786951
Tau2_1_1    0.0070020  0.0090497 -0.0107351  0.0247391
Tau2_2_1    0.0094607  0.0099698 -0.0100797  0.0290010
Tau2_2_2    0.0261445  0.0177409 -0.0086270  0.0609161
           z value  Pr(>|z|)
Intercept1  6.4298 1.278e-10 ***
Intercept2 -4.1593 3.192e-05 ***
Tau2_1_1    0.7737    0.4391
Tau2_2_1    0.9489    0.3427
Tau2_2_2    1.4737    0.1406
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 128.2267
Degrees of freedom of the Q statistic: 8
P value of the Q statistic: 0
Heterogeneity indices (based on the estimated Tau2):
                             Estimate
Intercept1: I2 (Q statistic)   0.6021
Intercept2: I2 (Q statistic)   0.9250

Number of studies (or clusters): 5
Number of observed statistics: 10
Number of estimated parameters: 5
Degrees of freedom: 5
-2 log likelihood: -11.68131
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Multivariate mixed-effects model As an illustration, we use *pub_year* as a predictor. To make the intercept more interpretable, we center the publication year at 1979, the first record of publication year in the data set. The estimated regression coefficients and their 95% CIs on PD and AL are 0.0064 (-0.2050, 0.2177) and -0.0706 (-0.3883, 0.2471), respectively. The R^2 in predicting PD and AL are .0000 and .0433, respectively. The likelihood ratio statistic on testing both regression coefficient is $\chi^2(df = 2) = 0.3273, p = .8490$. Thus, the null hypothesis of both regression coefficients are zero is not rejected.

Sometimes, we may want to test the equality of the regression coefficients and see if they are differnt. We may impose the equality constraint on the regression coefficients by using the same label in the argument `coef.constraints`. The average regression coefficient is 0.0017 (-0.1991, 0.2025). The likelihood ratio statistic on testing the equality of the regression coefficients is $\chi^2(df = 1) = 0.3270, p = .5674$. There is no evidence that one regression coefficient is stronger than the other.

```
R> ## Multivariate meta-analysis with "publication year-1979" as a predictor
R> mult2 <- meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL), data=Berkey98,
               x=scale(pub_year, center=1979), model.name="No constraint")

Running No constraint

R> summary(mult2)

Call:
meta(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    x = scale(pub_year, center = 1979), data = Berkey98, model.name = "No constraint")

95% confidence intervals: z statistic approximation
Coefficients:
             Estimate  Std.Error     lbound     ubound
Intercept1  0.3440001  0.0857659  0.1759021  0.5120982
Slope1_1    0.0063540  0.1078235 -0.2049762  0.2176842
Intercept2 -0.2918175  0.1312796 -0.5491208 -0.0345141
Slope2_1   -0.0705888  0.1620965 -0.3882921  0.2471146
Tau2_1_1    0.0080405  0.0101206 -0.0117955  0.0278766
Tau2_2_1    0.0093413  0.0105515 -0.0113392  0.0300218
Tau2_2_2    0.0250135  0.0170788 -0.0084603  0.0584873
           z value  Pr(>|z|)
Intercept1  4.0109 6.048e-05 ***
Slope1_1    0.0589   0.95301
Intercept2 -2.2229   0.02622 *
Slope2_1   -0.4355   0.66322
Tau2_1_1    0.7945   0.42692
Tau2_2_1    0.8853   0.37599
Tau2_2_2    1.4646   0.14303
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 128.2267
Degrees of freedom of the Q statistic: 8
P value of the Q statistic: 0
```

```
Explained variances (R2):
                            Estimate
y1: Tau2 (no predictor)       0.0070
y1: Tau2 (with predictors)    0.0080
y1: R2                        0.0000
y2: Tau2 (no predictor)       0.0261
y2: Tau2 (with predictors)    0.0250
y2: R2                        0.0433

Number of studies (or clusters): 5
Number of observed statistics: 10
Number of estimated parameters: 7
Degrees of freedom: 3
-2 log likelihood: -12.00859
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

```
R> ## Multivariate meta-analysis with both regression coefficients fixed at 0
R> mult0 <- meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL), data=Berkey98,
                x=scale(pub_year, center=1979),
                model.name="Both regression coefficients fixed at 0",
                coef.constraints=matrix(c("0", "0"), nrow=2))
```

```
Running Both regression coefficients fixed at 0
```

```
R> summary(mult0)
```

```
Call:
meta(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    x = scale(pub_year, center = 1979), data = Berkey98, coef.constraints = matrix(c("
        "0"), nrow = 2), model.name = "Both regression coefficients fixed at 0")

95% confidence intervals: z statistic approximation
Coefficients:
             Estimate  Std.Error     lbound     ubound
Intercept1  0.3448392  0.0536312  0.2397239  0.4499544
Intercept2 -0.3379381  0.0812479 -0.4971812 -0.1786951
Tau2_1_1    0.0070020  0.0090497 -0.0107351  0.0247391
Tau2_2_1    0.0094607  0.0099698 -0.0100797  0.0290010
Tau2_2_2    0.0261445  0.0177409 -0.0086270  0.0609161
           z value  Pr(>|z|)
Intercept1  6.4298 1.278e-10 ***
Intercept2 -4.1593 3.192e-05 ***
Tau2_1_1    0.7737    0.4391
```

```
Tau2_2_1    0.9489    0.3427
Tau2_2_2    1.4737    0.1406
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 128.2267
Degrees of freedom of the Q statistic: 8
P value of the Q statistic: 0
Explained variances (R2):
                           Estimate
y1: Tau2 (no predictor)      0.0070
y1: Tau2 (with predictors)   0.0070
y1: R2                       0.0000
y2: Tau2 (no predictor)      0.0261
y2: Tau2 (with predictors)   0.0261
y2: R2                       0.0000

Number of studies (or clusters): 5
Number of observed statistics: 10
Number of estimated parameters: 5
Degrees of freedom: 5
-2 log likelihood: -11.68131
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

```
R> ## Likelihood ratio test on testing both regression coefficients are 0
R> anova(mult2, mult0)
```

```
           base                              comparison ep
1 No constraint                                    <NA>  7
2 No constraint Both regression coefficients fixed at 0  5
   minus2LL df       AIC    diffLL diffdf         p
1 -12.00859  3 -18.00859        NA     NA        NA
2 -11.68131  5 -21.68131 0.3272789      2 0.8490481
```

```
R> ## Multivariate meta-analysis with an equality constraint on the slopes
R> mult3 <- meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL), data=Berkey98,
               x=scale(pub_year, center=1979), model.name="With equality constraint",
               coef.constraints=matrix(c("0.3*Equal_Slope", "0.3*Equal_Slope"), nrow=2))
```

```
Running With equality constraint
```

```
R> summary(mult3)
```

```
Call:
meta(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    x = scale(pub_year, center = 1979), data = Berkey98, coef.constraints = matrix(c("
        "0.3*Equal_Slope"), nrow = 2), model.name = "With equality constraint")

95% confidence intervals: z statistic approximation
Coefficients:
              Estimate  Std.Error     lbound     ubound
Intercept1   0.3437612  0.0849829  0.1771979  0.5103245
Equal_Slope  0.0016748  0.1024443 -0.1991123  0.2024619
Intercept2  -0.3390010  0.1041005 -0.5430344 -0.1349677
Tau2_1_1     0.0070474  0.0094638 -0.0115013  0.0255962
Tau2_2_1     0.0095165  0.0105668 -0.0111940  0.0302269
Tau2_2_2     0.0261979  0.0180773 -0.0092330  0.0616288
            z value  Pr(>|z|)
Intercept1   4.0451 5.231e-05 ***
Equal_Slope  0.0163  0.986956
Intercept2  -3.2565  0.001128 **
Tau2_1_1     0.7447  0.456471
Tau2_2_1     0.9006  0.367800
Tau2_2_2     1.4492  0.147278
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 128.2267
Degrees of freedom of the Q statistic: 8
P value of the Q statistic: 0
Explained variances (R2):
                           Estimate
y1: Tau2 (no predictor)      0.0070
y1: Tau2 (with predictors)   0.0070
y1: R2                       0.0000
y2: Tau2 (no predictor)      0.0261
y2: Tau2 (with predictors)   0.0262
y2: R2                       0.0000

Number of studies (or clusters): 5
Number of observed statistics: 10
Number of estimated parameters: 6
Degrees of freedom: 4
-2 log likelihood: -11.68158
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Likelihood ratio test on the equality of regression coefficients
```

```
R> anova(mult2, mult3)

           base               comparison ep  minus2LL df
1 No constraint                     <NA>  7 -12.00859  3
2 No constraint With equality constraint  6 -11.68158  4
        AIC    diffLL diffdf         p
1 -18.00859        NA     NA        NA
2 -19.68158 0.3270107      1 0.5674246
```

Multivariate fixed-effects model A multivariate fixed-effects meta-analysis can be conducted by fixing the variance component at a zero matrix. The following code illustrates the syntax.

```
R> ## Multivariate meta-analysis with a fixed-effects model
R> summary( meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL),
                data=Berkey98, RE.constraints=matrix(0, nrow=2, ncol=2)) )

Running Meta analysis with ML

Call:
meta(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    data = Berkey98, RE.constraints = matrix(0, nrow = 2, ncol = 2))

95% confidence intervals: z statistic approximation
Coefficients:
            Estimate Std.Error    lbound    ubound z value
Intercept1  0.307219  0.028575  0.251212  0.363225  10.751
Intercept2 -0.394377  0.018649 -0.430929 -0.357825 -21.147
            Pr(>|z|)
Intercept1 < 2.2e-16 ***
Intercept2 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Q statistic on homogeneity of effect sizes: 128.2267
Degrees of freedom of the Q statistic: 8
P value of the Q statistic: 0
Heterogeneity indices (based on the estimated Tau2):
                             Estimate
Intercept1: I2 (Q statistic)        0
Intercept2: I2 (Q statistic)        0

Number of studies (or clusters): 5
```

```
Number of observed statistics: 10
Number of estimated parameters: 2
Degrees of freedom: 8
-2 log likelihood: 90.88326
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

REML The `reml()` function may be used to estimate the variance component with the REML estimation method. It should be noted that it does not estimate the fixed-effects. The fixed-effects estimates can be obtained via the `meta()` function by specifying the estimated variance component from `reml()` as fixed values in the `RE.constraints` argument. This approach is consistent with the idea of REML that removes the fixed-effects parameter when estimating the variance component.

```
R> ## Random-effects meta-analysis with REML
R> summary( VarComp <- reml(y=di, v=vi, data=Becker83) )

Running Variance component with REML

Call:
reml(y = di, v = vi, data = Becker83)

95% confidence intervals: z statistic approximation
Coefficients:
          Estimate Std.Error    lbound    ubound z value
Tau2_1_1  0.091445  0.064228 -0.034439  0.217329  1.4238
         Pr(>|z|)
Tau2_1_1   0.1545

Number of studies (or clusters): 10
Number of observed statistics: 9
Number of estimated parameters: 1
Degrees of freedom: 8
-2 log likelihood: -6.110579
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Extract the variance component
R> VarComp_REML <- matrix( coef(VarComp), ncol=1, nrow=1 )
R> ## Meta-analysis by treating the variance component as fixed
R> summary( meta(y=di, v=vi, data=Becker83, RE.constraints=VarComp_REML) )

Running Meta analysis with ML

Call:
```

```
meta(y = di, v = vi, data = Becker83, RE.constraints = VarComp_REML)

95% confidence intervals: z statistic approximation
Coefficients:
            Estimate Std.Error    lbound    ubound z value
Intercept1  0.180189  0.117535 -0.050176  0.410555  1.5331
           Pr(>|z|)
Intercept1   0.1253

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Heterogeneity indices (based on the estimated Tau2):
                             Estimate
Intercept1: I2 (Q statistic)   0.7075

Number of studies (or clusters): 10
Number of observed statistics: 10
Number of estimated parameters: 1
Degrees of freedom: 9
-2 log likelihood: 7.986161
OpenMx status1: 1 ("0" and "1": considered fine; other values indicate problems)

R> ## Estimate variance components with REML
R> summary( reml(y=di, v=vi, x=log(items), data=Becker83) )

Running Variance component with REML

Call:
reml(y = di, v = vi, x = log(items), data = Becker83)

95% confidence intervals: z statistic approximation
Coefficients:
           Estimate  Std.Error     lbound     ubound
Tau2_1_1  0.0052656  0.0212014 -0.0362884  0.0468196
         z value Pr(>|z|)
Tau2_1_1  0.2484   0.8039

Number of studies (or clusters): 10
Number of observed statistics: 8
Number of estimated parameters: 1
Degrees of freedom: 7
-2 log likelihood: -10.84567
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)
```

```
R> ## Estimate variance components with REML
R> summary( reml(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL), data=Berkey98) )
```

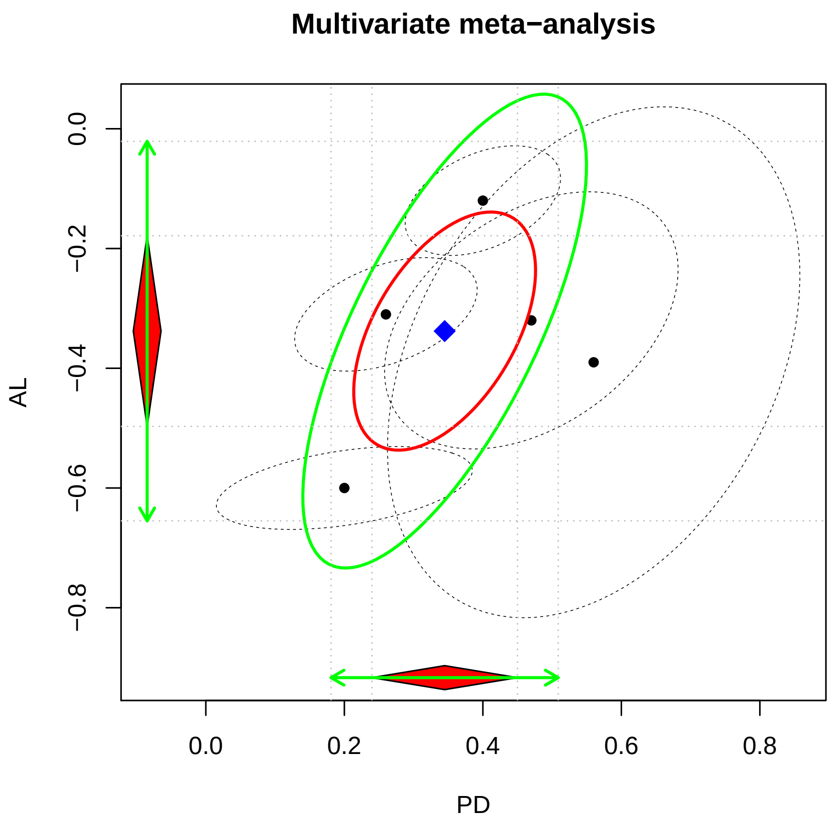
```
Running Variance component with REML

Call:
reml(y = cbind(PD, AL), v = cbind(var_PD, cov_PD_AL, var_AL),
    data = Berkey98)

95% confidence intervals: z statistic approximation
Coefficients:
          Estimate Std.Error    lbound    ubound z value
Tau2_1_1  0.011733  0.013645 -0.015011  0.038477  0.8599
Tau2_2_1  0.011916  0.014416 -0.016340  0.040172  0.8266
Tau2_2_2  0.032651  0.024402 -0.015176  0.080479  1.3380
         Pr(>|z|)
Tau2_1_1   0.3899
Tau2_2_1   0.4085
Tau2_2_2   0.1809

Number of studies (or clusters): 5
Number of observed statistics: 8
Number of estimated parameters: 3
Degrees of freedom: 5
-2 log likelihood: -18.86768
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)
```

Plots of multivariate effect sizes If multivariate meta-analysis is conducted, pairwise plots on the pooled effect sizes and their confidence ellipses can be obtained via the `plot()` function. By default, 95% confidence intervals on the average effect sizes and confidence ellipses on the random effects are plotted. For example,

```
R> Berkey98.fit <- meta(y=cbind(PD, AL), v=cbind(var_PD, cov_PD_AL, var_AL),
                        data=Berkey98)
```

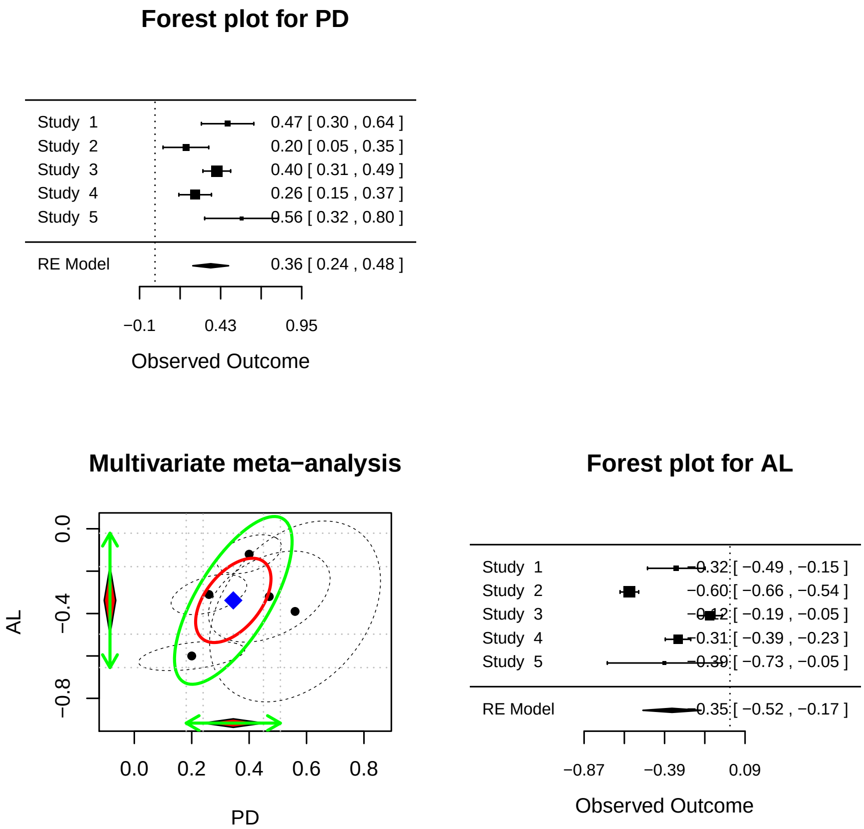
```
Running Meta analysis with ML
```

```
R> plot(Berkey98.fit, main="Multivariate meta-analysis", axis.label=c("PD", "AL"))
```



By combining with the forest plots from the **metafor** package, we may combine the univariate and multivariate natures of the effect sizes in a single figure. This will be very useful for multivariate meta-analysis.

```
R> ## Load the metafor package to display forest plots
R> library(metafor)
R> plot(Berkey98.fit, diag.panel=TRUE, main="Multivariate meta-analysis",
        axis.label=c("PD", "AL"))
R> forest( rma(yi=PD, vi=var_PD, data=Berkey98) )
R> title("Forest plot for PD")
R> forest( rma(yi=AL, vi=var_AL, data=Berkey98) )
R> title("Forest plot for AL")
```



Three-level meta-analysis The `meta3()` function may be used to fit three-level meta-analytic models. It is assumed that effect sizes within `cluster` are dependent. The Q statistic ($df = 55$) of the following example is 578.86, $p < .001$. The I^2 at level-2 and level-3 are .3440 and .6043, respectively. The estimated coefficient of $Year$ of publication is 0.0051, $p = .5518$. The R^2 at level-2 and level-3 are .0000 and .0221, respectively.

```
R> ## ML estimation method
R> ## No predictor
R> summary( meta3(y=y, v=v, cluster=District, data=Cooper03) )

Running Meta analysis with ML

Call:
meta3(y = y, v = v, cluster = District, data = Cooper03)

95% confidence intervals: z statistic approximation
```

```
Coefficients:
            Estimate  Std.Error     lbound     ubound
Intercept  0.1844554  0.0805411  0.0265977  0.3423131
Tau2_2     0.0328648  0.0111397  0.0110314  0.0546982
Tau2_3     0.0577384  0.0307423 -0.0025154  0.1179921
          z value Pr(>|z|)
Intercept  2.2902 0.022010 *
Tau2_2     2.9502 0.003175 **
Tau2_3     1.8781 0.060362 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Q statistic on homogeneity of effect sizes: 578.864
Degrees of freedom of the Q statistic: 55
P value of the Q statistic: 0
Heterogeneity indices (based on the estimated Tau2):
                   Estimate
I2_2 (Q statistic)   0.3440
I2_3 (Q statistic)   0.6043

Number of studies (or clusters): 11
Number of observed statistics: 56
Number of estimated parameters: 3
Degrees of freedom: 53
-2 log likelihood: 16.78987
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Year as a predictor
R> summary( meta3(y=y, v=v, cluster=District, x=scale(Year, scale=FALSE), data=Cooper03) )

Running Meta analysis with ML

Call:
meta3(y = y, v = v, cluster = District, x = scale(Year, scale = FALSE),
    data = Cooper03)

95% confidence intervals: z statistic approximation
Coefficients:
            Estimate  Std.Error     lbound     ubound
Slope_1    0.0050737  0.0085266 -0.0116382  0.0217856
Intercept  0.1780268  0.0805219  0.0202067  0.3358469
Tau2_2     0.0329390  0.0111620  0.0110618  0.0548162
Tau2_3     0.0564628  0.0300330 -0.0024007  0.1153264
          z value Pr(>|z|)
```

```
Slope_1    0.5950 0.551814
Intercept  2.2109 0.027042 *
Tau2_2     2.9510 0.003168 **
Tau2_3     1.8800 0.060104 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 578.864
Degrees of freedom of the Q statistic: 55
P value of the Q statistic: 0
Explained variances (R2):
                         Estimate
Tau2_2 (no predictor)      0.0329
Tau2_2 (with predictors)   0.0329
R2_2 (level-2)             0.0000
Tau2_3 (no predictor)      0.0577
Tau2_3 (with predictors)   0.0565
R2_3 (level-3)             0.0221

Number of studies (or clusters): 11
Number of observed statistics: 56
Number of estimated parameters: 4
Degrees of freedom: 52
-2 log likelihood: 16.43629
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## REML estimation method
R> ## No predictor; with LBCI
R> summary( reml3(y=y, v=v, cluster=District, data=Cooper03, intervals.type="LB") )

Running Variance component with REML

Call:
reml3(y = y, v = v, cluster = District, data = Cooper03, intervals.type = "LB")

95% confidence intervals: Likelihood-based statistic
Coefficients:
       Estimate Std.Error   lbound   ubound z value
Tau2_2 0.032737  0.011092 0.016264 0.062842  2.9513
Tau2_3 0.065062  0.035510 0.022234 0.207846  1.8322
       Pr(>|z|)
Tau2_2 0.003164 **
Tau2_3 0.066921 .
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of studies (or clusters): 56
Number of observed statistics: 55
Number of estimated parameters: 2
Degrees of freedom: 53
-2 log likelihood: -81.14044
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Year as a predictor
R> summary( reml3(y=y, v=v, cluster=District, x=scale(Year, scale=FALSE), data=Cooper03) )

Running Variance component with REML

Call:
reml3(y = y, v = v, cluster = District, x = scale(Year, scale = FALSE),
    data = Cooper03)

95% confidence intervals: z statistic approximation
Coefficients:
         Estimate  Std.Error     lbound     ubound z value
Tau2_2  0.0326502  0.0110529  0.0109870  0.0543134  2.9540
Tau2_3  0.0722656  0.0405349 -0.0071813  0.1517125  1.7828
       Pr(>|z|)
Tau2_2 0.003137 **
Tau2_3 0.074619 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Number of studies (or clusters): 56
Number of observed statistics: 54
Number of estimated parameters: 2
Degrees of freedom: 52
-2 log likelihood: -72.09405
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)
```

3. Meta-analytic structural equation modeling

MASEM combines the idea of meta-analysis and SEM by pooling correlation or covariance matrices and testing structural equation models on the pooled correlation or covariance matrix. There are two stages in conducting a MASEM. In the first stage the correlation or covariance matrices are pooled together. In the second stage, the pooled correlation or covariance matrix is used to fit structural equation models.

Cheung and Chan (2005b, 2009) proposed a two-stage structural equation modeling (TSSEM) based on a fixed-effects model. The `metaSEM` package implemented the TSSEM approach. Moreover, the TSSEM approach has been extended to the random-effects model by using a multivariate meta-analysis (Cheung 2011b) in the first stage analysis. Regardless of whether a fixed- or random-effects model is used, the `tssem2()` function will handle this automatically. In other words, parameter estimates, standard errors and goodness-of-fit indices in the stage 2 analysis has already taken the stage 1 model into account.

3.1. Stage 1 analysis

The objective of the stage 1 analysis is to obtain a pooled correlation or covariance matrix. Under the fixed-effects model, it is assumed that all population correlation or covariance matrices are the same while there are study specific correlation or covariance matrices under the random-effects model.

Fixed-effects model The population correlation or covariance matirx in the ith study can be decomposed as:

$$\Sigma_i(\theta) = \mathbf{D}_i \mathbf{P}_i \mathbf{D}_i \tag{27}$$

where $\Sigma_i(\theta)$ is the model implied covariance matrix, $\mathbf{D}_i$ is the diagonal matrix of standard deviations, and $\mathbf{P}_i$ is the correlation matrix. Under the assumption of homogeneity of correlation matrices, we may obtain a common correlation matrix by imposing the constraint $P = P_1 = P_2 = \ldots = P_k$ where D_i may vary across studies. If we want to obtain a common covariance matrix under the assumption of homogeneity of covariance matrices, we may also add the constraint $D = D_1 = D_2 = \ldots = D_k$.

Random-effects model When a random-effects model is used, the correlation or covariance matrices are treated as vectors of multivariate effect sizes. The model for the ith sample correlation vector $\mathbf{r}_i$ is:

$$\mathbf{r}_i = \rho_{\text{random}} + \mathbf{u}_i + \mathbf{e}_i \tag{28}$$

where ρ_{random} is the average correlation vector, $\mathbf{T}^2 = \mathbf{u}_i$ is the variance component of the random effects, and $V_i = \mathsf{Var}(\mathbf{e}_i)$ is the conditional sampling covariance matrix. V_i is estimated by the `asyCov()` internally. Multivariate meta-analysis as listed in Eqs. 17 and 18 are used to conduct the stage 1 analysis.

3.2. Stage 2 analysis

After the stage 1 analysis based on either a fixed- or a random-effects model, a vector of pooled correlation or covariance matrix $\bar{\mathbf{r}}$ and its asymptotic covariance matrix $\mathbf{V}$ are available. A correlational structural model $\rho(\hat{\gamma})$ is fitted with weighted least squares

(WLS) estimation method by minimizing the following fit function (Cheung and Chan 2005b, 2009):

$$F(\hat{\gamma}) = (\bar{\mathbf{r}} - \rho(\hat{\gamma}))^T \mathbf{V}^{-1} (\bar{\mathbf{r}} - \rho(\hat{\gamma})). \tag{29}$$

Likelihood-ratio statistics and various goodness-of-fit indices may be used to judge whether the proposed structural model is appropriate while *SE*s may be used to test the significance of the parameter estimates.

3.3. Examples

An example from Cheung (2009b) is used to illustrate the procedures. In this example, Digman (1997) reported a second-order factor analysis on a five-factor model with 14 studies. He proposed that there were two second-order factors for the five-factor model: an alpha factor for agreeableness, conscientiousness, and emotional stability, and a beta factor for extroversion and intellect.

Fixed-effects model The `tssem1()` function is used to pool the correlation matrices with a fixed-effects model in the first stage by specifying `method='FEM'` in the argument. `tssem2()` is then used to fit a factor analytic model on the pooled correlation matrix with the inverse of its asymptotic covariance matrix as the weight matrix. When a correlation structure is fitted, it is necessary to ensure that the diagonals of the model implied matrix are all ones. We may impose this constraint by specifying `diag.constraints=TRUE`.

The fit indices for testing the homogeneity of the correlation matrices in Stage 1 analysis are $\chi^2(df = 130, N = 4,496) = 1,499.73, p < .001$, CFI=0.6825, TLI=0.6581, SRMR=0.1750 and RMSEA=0.1812. This indicates that it is not reasonable to assume that the correlation matrices are homogenous. Sub-group analysis or random-effects model that will be illustrated later are more appropriate. As an exercise, we continute to fit the stage 2 model. The fit indices for fitting the structural model in Stage 2 are $\chi^2(df = 4, N = 4496) = 65.06, p < .001$, CFI=0.9802, TLI=0.9506, SRMR=0.0284 and RMSEA=0.0583. Although the goodness-of-fit indices look good, we should be cautious in interpreting them because of the poor goodness-of-fit indices in Stage 1 analysis.

```
R> ## Show the first 2 studies in Digman97
R> head(Digman97$data, n=2)

$`Digman 1 (1994)`
       A     C   ES     E    I
A   1.00  0.62 0.41 -0.48 0.00
C   0.62  1.00 0.59 -0.10 0.35
ES  0.41  0.59 1.00  0.27 0.41
E  -0.48 -0.10 0.27  1.00 0.37
I   0.00  0.35 0.41  0.37 1.00

$`Digman 2 (1994)`
```

```
       A    C   ES     E     I
A   1.00 0.39 0.53 -0.30 -0.05
C   0.39 1.00 0.59  0.07  0.44
ES  0.53 0.59 1.00  0.09  0.22
E  -0.30 0.07 0.09  1.00  0.45
I  -0.05 0.44 0.22  0.45  1.00
```

```
R> ## Show the first 2 sample sizes in Digman97
R> head(Digman97$n, n=2)
```

```
[1] 102 149
```

```
R> ## Stage 1 analysis
R> fixed1 <- tssem1(Digman97$data, Digman97$n, method="FEM")
```

```
Running TSSEM1 Analysis of Correlation Matrix
```

```
R> summary(fixed1)
```

```
Call:
tssem1FEM(my.df = my.df, n = n, cor.analysis = cor.analysis,
    model.name = model.name, cluster = cluster, suppressWarnings = suppressWarnings)

Coefficients:
       Estimate Std.Error z value  Pr(>|z|)
S[1,2] 0.363116  0.013390 27.1182 < 2.2e-16 ***
S[1,3] 0.390176  0.012903 30.2402 < 2.2e-16 ***
S[1,4] 0.103751  0.015070  6.8848 5.786e-12 ***
S[1,5] 0.092246  0.015071  6.1208 9.309e-10 ***
S[2,3] 0.415999  0.012539 33.1756 < 2.2e-16 ***
S[2,4] 0.135208  0.014799  9.1365 < 2.2e-16 ***
S[2,5] 0.141213  0.014890  9.4836 < 2.2e-16 ***
S[3,4] 0.244505  0.014175 17.2494 < 2.2e-16 ***
S[3,5] 0.138167  0.014857  9.2995 < 2.2e-16 ***
S[4,5] 0.424514  0.012395 34.2477 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Goodness-of-fit indices:
                                   Value
Sample size                    4496.0000
Chi-square of target model     1499.7340
DF of target model              130.0000
```

```
p value of target model           0.0000
Chi-square of independent model 4454.5995
DF of independent model         140.0000
RMSEA                             0.1812
SRMR                              0.1750
TLI                               0.6581
CFI                               0.6825
AIC                            1239.7340
BIC                             406.3114
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Model specification with RAM formulation
R> ## S matrix
R> Phi <- matrix(c(1,"0.3*cor","0.3*cor",1), ncol=2, nrow=2)
R> S1 <- bdiagMat(list(diag(c("0.2*e1","0.2*e2","0.2*e3","0.2*e4","0.2*e5")), Phi))
R> ## Not necessary but useful in understanding the model
R> dimnames(S1) <- list(c("A","C","ES","E","I","Alpha","Beta"),
                        c("A","C","ES","E","I","Alpha","Beta"))
R> S1

      A        C        ES       E        I
A     "0.2*e1" "0"      "0"      "0"      "0"
C     "0"      "0.2*e2" "0"      "0"      "0"
ES    "0"      "0"      "0.2*e3" "0"      "0"
E     "0"      "0"      "0"      "0.2*e4" "0"
I     "0"      "0"      "0"      "0"      "0.2*e5"
Alpha "0"      "0"      "0"      "0"      "0"
Beta  "0"      "0"      "0"      "0"      "0"
      Alpha     Beta
A     "0"       "0"
C     "0"       "0"
ES    "0"       "0"
E     "0"       "0"
I     "0"       "0"
Alpha "1"       "0.3*cor"
Beta  "0.3*cor" "1"

R> ## Convert it into mxMatrix class
R> S1 <- as.mxMatrix(S1)
R> ## A matrix
R> Lambda <-  matrix(c(".3*Alpha_A",".3*Alpha_C",".3*Alpha_ES",
                     rep(0,5),".3*Beta_E",".3*Beta_I"), ncol=2, nrow=5)
R> A1 <- rbind( cbind(matrix(0,ncol=5,nrow=5), Lambda),
```

```
               matrix(0, ncol=7, nrow=2) )
R> ## Not necessary but useful in understanding the model
R> dimnames(A1) <- list(c("A","C","ES","E","I","Alpha","Beta"),
                       c("A","C","ES","E","I","Alpha","Beta"))
R> A1

      A   C   ES  E   I   Alpha         Beta
A     "0" "0" "0" "0" "0" ".3*Alpha_A"  "0"
C     "0" "0" "0" "0" "0" ".3*Alpha_C"  "0"
ES    "0" "0" "0" "0" "0" ".3*Alpha_ES" "0"
E     "0" "0" "0" "0" "0" "0"           ".3*Beta_E"
I     "0" "0" "0" "0" "0" "0"           ".3*Beta_I"
Alpha "0" "0" "0" "0" "0" "0"           "0"
Beta  "0" "0" "0" "0" "0" "0"           "0"

R> A1 <- as.mxMatrix(A1)
R> F1 <- create.Fmatrix(c(1,1,1,1,1,0,0), name="F1")
R> ## Stage 2 analysis
R> fixed2 <- tssem2(fixed1, Amatrix=A1, Smatrix=S1, Fmatrix=F1, diag.constraints=TRUE,
                  model.name="TSSEM2 Digman97")

Running TSSEM2 Digman97

R> summary(fixed2)

Call:
wls(Cov = tssem1.obj$pooledS, asyCov = tssem1.obj$acovS, n = tssem1.obj$total.n,
    Amatrix = Amatrix, Smatrix = Smatrix, Fmatrix = Fmatrix,
    diag.constraints = diag.constraints, cor.analysis = cor.analysis,
    intervals.type = intervals.type, model.name = model.name,
    suppressWarnings = suppressWarnings)

95% confidence intervals: Likelihood-based statistic
Coefficients:
             Estimate Std.Error  lbound  ubound z value
Amatrix[1,6]  0.56258        NA 0.53242 0.59286      NA
Amatrix[2,6]  0.60512        NA 0.57509 0.63532      NA
Amatrix[3,6]  0.71913        NA 0.68862 0.75031      NA
Amatrix[4,7]  0.78200        NA 0.71911 0.85587      NA
Amatrix[5,7]  0.55089        NA 0.49939 0.60231      NA
Smatrix[1,1]  0.68351        NA 0.64852 0.71653      NA
Smatrix[2,2]  0.63382        NA 0.59636 0.66927      NA
Smatrix[3,3]  0.48285        NA 0.43702 0.52580      NA
```

```
Smatrix[4,4]  0.38847        NA 0.26741 0.48290      NA
Smatrix[5,5]  0.69653        NA 0.63721 0.75061      NA
Smatrix[7,6]  0.36261        NA 0.31843 0.40650      NA
             Pr(>|z|)
Amatrix[1,6]       NA
Amatrix[2,6]       NA
Amatrix[3,6]       NA
Amatrix[4,7]       NA
Amatrix[5,7]       NA
Smatrix[1,1]       NA
Smatrix[2,2]       NA
Smatrix[3,3]       NA
Smatrix[4,4]       NA
Smatrix[5,5]       NA
Smatrix[7,6]       NA

Goodness-of-fit indices:
                                               Value
Sample size                                4496.0000
Chi-square of target model                   65.0605
DF of target model                            4.0000
p value of target model                       0.0000
Number of constraints imposed on "Smatrix"    5.0000
DF manually adjusted                          0.0000
Chi-square of independent model            3101.3209
DF of independent model                      10.0000
RMSEA                                         0.0583
SRMR                                          0.0284
TLI                                           0.9506
CFI                                           0.9802
AIC                                          57.0605
BIC                                          31.4167
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Fixed-effects model with sub-group analysis Studies may not share the same population correlation matrix. If the studies can be grouped into various subgroups, we may pool the correlation matrices separately for each subgroup (Cheung and Chan 2005a). This is similar to the subgroup analysis in conventional meta-analysis (Hedges and Olkin 1985). For example, Digman (1997) groups the 14 studies into several groups according to their sample characteristics. These include children, adolescents, young adults, and mature adults. This variable is stored in the variable `Digman97$cluster`. We may further group these studies into younger participants versus older participants. Separate fixed-effects analysis may be applied into these two groups.

The following R code may be used to conduct the analysis. Users have to supply the cluster (a vector of labels) to the `cluster` argument in `tssem1()`. The correlation or covariance matrices will be pooled separately for each cluster. The structural models will also be fitted separately for each cluster.

```
R> #### Create a variable for different samples
R> #### Younger participants: Children and Adolescents
R> #### Older participants: others
R> sample <- ifelse(Digman97$cluster %in% c("Children","Adolescents"),
                    yes="Younger participants", no="Older participants")
R> #### Show the sample
R> sample

 [1] "Younger participants" "Younger participants"
 [3] "Younger participants" "Younger participants"
 [5] "Younger participants" "Older participants"
 [7] "Older participants"   "Older participants"
 [9] "Older participants"   "Older participants"
[11] "Older participants"   "Older participants"
[13] "Older participants"   "Older participants"

R> ## Stage 1 analysis with clusters
R> fixed1.cluster <- tssem1(Digman97$data, Digman97$n, method="FEM",
                            cluster=sample)

Running TSSEM1 Analysis of Correlation Matrix
Running TSSEM1 Analysis of Correlation Matrix

R> #### Please note that the estimates for the younger participants are problematic.
R> summary(fixed1.cluster)

$`Older participants`

Call:
tssem1FEM(my.df = data.cluster[[i]], n = n.cluster[[i]], cor.analysis = cor.analysis,
    model.name = model.name, suppressWarnings = suppressWarnings)

Coefficients:
       Estimate Std.Error z value  Pr(>|z|)
S[1,2] 0.297484  0.015455 19.2483 < 2.2e-16 ***
S[1,3] 0.370088  0.014552 25.4315 < 2.2e-16 ***
S[1,4] 0.137688  0.016423  8.3836 < 2.2e-16 ***
S[1,5] 0.097971  0.016744  5.8510 4.888e-09 ***
```

```
S[2,3] 0.393709  0.014163 27.7975 < 2.2e-16 ***
S[2,4] 0.182984  0.016075 11.3829 < 2.2e-16 ***
S[2,5] 0.092664  0.016664  5.5608 2.685e-08 ***
S[3,4] 0.260756  0.015573 16.7440 < 2.2e-16 ***
S[3,5] 0.096063  0.016594  5.7889 7.083e-09 ***
S[4,5] 0.411753  0.013917 29.5853 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Goodness-of-fit indices:
                                    Value
Sample size                     3658.0000
Chi-square of target model       823.8769
DF of target model                80.0000
p value of target model            0.0000
Chi-square of independent model 2992.9294
DF of independent model           90.0000
RMSEA                              0.1513
SRMR                               0.1528
TLI                                0.7117
CFI                                0.7437
AIC                              663.8769
BIC                              167.5032
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)

$`Younger participants`

Call:
tssem1FEM(my.df = data.cluster[[i]], n = n.cluster[[i]], cor.analysis = cor.analysis,
    model.name = model.name, suppressWarnings = suppressWarnings)

Coefficients:
        Estimate Std.Error z value  Pr(>|z|)
S[1,2]  0.604396  0.022189 27.2389 < 2.2e-16 ***
S[1,3]  0.465441  0.027579 16.8767 < 2.2e-16 ***
S[1,4] -0.030869  0.036048 -0.8563   0.39181
S[1,5]  0.061581  0.034650  1.7772   0.07554 .
S[2,3]  0.501309  0.026431 18.9666 < 2.2e-16 ***
S[2,4] -0.060834  0.034660 -1.7551   0.07923 .
S[2,5]  0.321019  0.031157 10.3033 < 2.2e-16 ***
S[3,4]  0.175422  0.033776  5.1937 2.062e-07 ***
S[3,5]  0.305214  0.031680  9.6344 < 2.2e-16 ***
S[4,5]  0.478573  0.026966 17.7473 < 2.2e-16 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Goodness-of-fit indices:
                                      Value
Sample size                        838.0000
Chi-square of target model         344.1826
DF of target model                  40.0000
p value of target model              0.0000
Chi-square of independent model   1461.6701
DF of independent model             50.0000
RMSEA                                0.2131
SRMR                                 0.1508
TLI                                  0.7307
CFI                                  0.7845
AIC                                264.1826
BIC                                 74.9419
OpenMx status: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Stage 2 analysis with clusters
R> fixed2.cluster <- tssem2(fixed1.cluster, Amatrix=A1, Smatrix=S1, Fmatrix=F1,
                          diag.constraints=TRUE, intervals.type="LB")

Running TSSEM2 (Fixed Effects Model) Analysis of Correlation Structure
Running TSSEM2 (Fixed Effects Model) Analysis of Correlation Structure

R> summary(fixed2.cluster)

$`Older participants`

Call:
wls(Cov = tssem1.obj$pooledS, asyCov = tssem1.obj$acovS, n = tssem1.obj$total.n,
    Amatrix = Amatrix, Smatrix = Smatrix, Fmatrix = Fmatrix,
    diag.constraints = diag.constraints, cor.analysis = cor.analysis,
    intervals.type = intervals.type, model.name = model.name,
    suppressWarnings = suppressWarnings)

95% confidence intervals: Likelihood-based statistic
Coefficients:
              Estimate Std.Error    lbound    ubound
Amatrix[1,6]  0.512525        NA  0.476852  0.548377
Amatrix[2,6]  0.550049        NA  0.514877  0.585392
Amatrix[3,6]  0.732091        NA  0.695584  0.770073
```

```
Amatrix[4,7]  0.967544        NA  0.868045  1.109618
Amatrix[5,7]  0.430460        NA  0.369165  0.486893
Smatrix[1,1]  0.737318        NA  0.699282  0.772612
Smatrix[2,2]  0.697447        NA  0.657314  0.734902
Smatrix[3,3]  0.464043        NA  0.406978  0.516169
Smatrix[4,4]  0.063858        NA -0.231965  0.246579
Smatrix[5,5]  0.814705        NA  0.762935  0.863710
Smatrix[7,6]  0.349090        NA  0.292052  0.403236
             z value Pr(>|z|)
Amatrix[1,6]      NA       NA
Amatrix[2,6]      NA       NA
Amatrix[3,6]      NA       NA
Amatrix[4,7]      NA       NA
Amatrix[5,7]      NA       NA
Smatrix[1,1]      NA       NA
Smatrix[2,2]      NA       NA
Smatrix[3,3]      NA       NA
Smatrix[4,4]      NA       NA
Smatrix[5,5]      NA       NA
Smatrix[7,6]      NA       NA

Goodness-of-fit indices:
                                              Value
Sample size                               3658.0000
Chi-square of target model                  21.9237
DF of target model                           4.0000
p value of target model                      0.0002
Number of constraints imposed on "Smatrix"   5.0000
DF manually adjusted                         0.0000
Chi-square of independent model           2267.2438
DF of independent model                     10.0000
RMSEA                                        0.0350
SRMR                                         0.0160
TLI                                          0.9801
CFI                                          0.9921
AIC                                         13.9237
BIC                                        -10.8950
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

$`Younger participants`

Call:
wls(Cov = tssem1.obj$pooledS, asyCov = tssem1.obj$acovS, n = tssem1.obj$total.n,
```

```
    Amatrix = Amatrix, Smatrix = Smatrix, Fmatrix = Fmatrix,
    diag.constraints = diag.constraints, cor.analysis = cor.analysis,
    intervals.type = intervals.type, model.name = model.name,
    suppressWarnings = suppressWarnings)

95% confidence intervals: Likelihood-based statistic
Coefficients:
                Estimate Std.Error      lbound      ubound
Amatrix[1,6]  7.4764e-01        NA  7.0068e-01  7.9470e-01
Amatrix[2,6]  9.1196e-01        NA  8.7319e-01  9.5151e-01
Amatrix[3,6]  6.7721e-01        NA  6.2601e-01  7.2752e-01
Amatrix[4,7]  1.5236e-01        NA  1.4099e-02  3.4171e-01
Amatrix[5,7]  3.2896e+00        NA  1.5251e+00  2.3344e+02
Smatrix[1,1]  4.4104e-01        NA  3.6842e-01  5.0906e-01
Smatrix[2,2]  1.6833e-01        NA  9.4595e-02  2.3756e-01
Smatrix[3,3]  5.4138e-01        NA  4.7070e-01  6.0812e-01
Smatrix[4,4]  9.7679e-01        NA  8.8327e-01  9.9937e-01
Smatrix[5,5] -9.8213e+00        NA -9.9502e+05 -1.3113e+00
Smatrix[7,6]  1.1708e-01        NA  1.1658e-02  2.7501e-01
             z value Pr(>|z|)
Amatrix[1,6]      NA       NA
Amatrix[2,6]      NA       NA
Amatrix[3,6]      NA       NA
Amatrix[4,7]      NA       NA
Amatrix[5,7]      NA       NA
Smatrix[1,1]      NA       NA
Smatrix[2,2]      NA       NA
Smatrix[3,3]      NA       NA
Smatrix[4,4]      NA       NA
Smatrix[5,5]      NA       NA
Smatrix[7,6]      NA       NA

Goodness-of-fit indices:
                                                Value
Sample size                                  838.0000
Chi-square of target model                   144.8651
DF of target model                             4.0000
p value of target model                        0.0000
Number of constraints imposed on "Smatrix"     5.0000
DF manually adjusted                           0.0000
Chi-square of independent model             2469.4301
DF of independent model                       10.0000
RMSEA                                          0.2051
```

```
SRMR                                             0.1051
TLI                                              0.8568
CFI                                              0.9427
AIC                                            136.8651
BIC                                            117.9410
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

Random-effects model TSSEM using a random-effects model may be requested by specifying the `method='REM'` argument in `tssem1()`. By default (`RE.diag.only=FALSE`), a positive definite covariance matrix among the random-effects is used. For practical reasons, e.g., there are not enough studies, it may not be feasible to estimate the full covariance matrix of the random effects. A diagonal matrix on the random-effects may be specified by using `RE.diag.only=TRUE`.

The fit indices for fitting the structural model in Stage 2 are $\chi^2(df = 4, N = 4,496) = 8.51, p < .001$, CFI=0.9911, TLI=0.9776, SRMR=0.0463 and RMSEA=0.0158. This indicates that the model fits the data quite well.

```
R> random1 <- tssem1(Digman97$data, Digman97$n, method="REM", RE.diag.only=TRUE)

Running TSSEM1 (Random Effects Model) Analysis of Correlation Matrix

R> summary(random1)

Call:
meta(y = ES, v = acovR, RE.constraints = diag(x = paste(RE.startvalues,
    "*Tau2_", 1:no.es, "_", 1:no.es, sep = ""), nrow = no.es,
    ncol = no.es), I2 = I2, model.name = model.name)

95% confidence intervals: z statistic approximation
Coefficients:
               Estimate   Std.Error      lbound      ubound
Intercept1   3.9465e-01  5.4223e-02  2.8837e-01  5.0092e-01
Intercept2   4.4009e-01  4.1258e-02  3.5923e-01  5.2096e-01
Intercept3   5.4542e-02  6.1716e-02 -6.6418e-02  1.7550e-01
Intercept4   9.8668e-02  4.6219e-02  8.0811e-03  1.8925e-01
Intercept5   4.2966e-01  4.0156e-02  3.5096e-01  5.0836e-01
Intercept6   1.2851e-01  4.0816e-02  4.8514e-02  2.0851e-01
Intercept7   2.0526e-01  4.9591e-02  1.0806e-01  3.0245e-01
Intercept8   2.3994e-01  3.1924e-02  1.7737e-01  3.0250e-01
Intercept9   1.8910e-01  4.3014e-02  1.0480e-01  2.7341e-01
Intercept10  4.4413e-01  3.2547e-02  3.8034e-01  5.0792e-01
Tau2_1_1     3.7207e-02  1.5000e-02  7.8079e-03  6.6607e-02
Tau2_2_2     2.0305e-02  8.4348e-03  3.7735e-03  3.6837e-02
```

```
Tau2_3_3     4.8219e-02  1.9723e-02  9.5631e-03  8.6876e-02
Tau2_4_4     2.4610e-02  1.0624e-02  3.7872e-03  4.5433e-02
Tau2_5_5     1.8725e-02  8.2474e-03  2.5602e-03  3.4889e-02
Tau2_6_6     1.8256e-02  8.7889e-03  1.0302e-03  3.5482e-02
Tau2_7_7     2.9424e-02  1.2263e-02  5.3893e-03  5.3458e-02
Tau2_8_8     9.6511e-03  4.8824e-03  8.1728e-05  1.9220e-02
Tau2_9_9     2.0933e-02  9.1280e-03  3.0430e-03  3.8824e-02
Tau2_10_10   1.1151e-02  5.0467e-03  1.2592e-03  2.1042e-02
            z value  Pr(>|z|)
Intercept1   7.2782 3.384e-13 ***
Intercept2  10.6668 < 2.2e-16 ***
Intercept3   0.8838  0.376824
Intercept4   2.1348  0.032777 *
Intercept5  10.6999 < 2.2e-16 ***
Intercept6   3.1486  0.001641 **
Intercept7   4.1390 3.488e-05 ***
Intercept8   7.5159 5.662e-14 ***
Intercept9   4.3963 1.101e-05 ***
Intercept10 13.6460 < 2.2e-16 ***
Tau2_1_1     2.4805  0.013120 *
Tau2_2_2     2.4073  0.016069 *
Tau2_3_3     2.4448  0.014492 *
Tau2_4_4     2.3164  0.020535 *
Tau2_5_5     2.2704  0.023184 *
Tau2_6_6     2.0772  0.037785 *
Tau2_7_7     2.3995  0.016420 *
Tau2_8_8     1.9767  0.048075 *
Tau2_9_9     2.2933  0.021829 *
Tau2_10_10   2.2095  0.027142 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 2381.007
Degrees of freedom of the Q statistic: 130
P value of the Q statistic: 0
Heterogeneity indices (based on the estimated Tau2):
                              Estimate
Intercept1: I2 (Q statistic)    0.9487
Intercept2: I2 (Q statistic)    0.9082
Intercept3: I2 (Q statistic)    0.9414
Intercept4: I2 (Q statistic)    0.8894
Intercept5: I2 (Q statistic)    0.9005
Intercept6: I2 (Q statistic)    0.8537
```

```
Intercept7: I2 (Q statistic)    0.9093
Intercept8: I2 (Q statistic)    0.7714
Intercept9: I2 (Q statistic)    0.8746
Intercept10: I2 (Q statistic)   0.8431

Number of studies (or clusters): 14
Number of observed statistics: 140
Number of estimated parameters: 20
Degrees of freedom: 120
-2 log likelihood: -110.8453
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

R> random2 <- tssem2(random1, Amatrix=A1, Smatrix=S1, Fmatrix=F1, diag.constraints=TRUE, int

Running TSSEM2 (Random Effects Model) Analysis of Correlation Structure

R> summary(random2)

Call:
wls(Cov = pooledS, asyCov = asyCov, n = tssem1.obj$total.n, Amatrix = Amatrix,
    Smatrix = Smatrix, Fmatrix = Fmatrix, diag.constraints = diag.constraints,
    cor.analysis = cor.analysis, intervals.type = intervals.type,
    model.name = model.name, suppressWarnings = suppressWarnings)

95% confidence intervals: Likelihood-based statistic
Coefficients:
             Estimate Std.Error  lbound  ubound z value
Amatrix[1,6]  0.57255        NA 0.47370 0.67690      NA
Amatrix[2,6]  0.59010        NA 0.49047 0.69487      NA
Amatrix[3,6]  0.77046        NA 0.65994 0.90431      NA
Amatrix[4,7]  0.69340        NA 0.56258 0.87183      NA
Amatrix[5,7]  0.64011        NA 0.50833 0.78644      NA
Smatrix[1,1]  0.67218        NA 0.54176 0.77561      NA
Smatrix[2,2]  0.65178        NA 0.51711 0.75944      NA
Smatrix[3,3]  0.40640        NA 0.18193 0.56452      NA
Smatrix[4,4]  0.51919        NA 0.23937 0.68353      NA
Smatrix[5,5]  0.59026        NA 0.38131 0.74161      NA
Smatrix[7,6]  0.39366        NA 0.30236 0.49030      NA
             Pr(>|z|)
Amatrix[1,6]       NA
Amatrix[2,6]       NA
Amatrix[3,6]       NA
Amatrix[4,7]       NA
```

```
Amatrix[5,7]       NA
Smatrix[1,1]       NA
Smatrix[2,2]       NA
Smatrix[3,3]       NA
Smatrix[4,4]       NA
Smatrix[5,5]       NA
Smatrix[7,6]       NA

Goodness-of-fit indices:
                                               Value
Sample size                                4496.0000
Chi-square of target model                    8.5119
DF of target model                            4.0000
p value of target model                       0.0745
Number of constraints imposed on "Smatrix"    5.0000
DF manually adjusted                          0.0000
Chi-square of independent model             514.5591
DF of independent model                      10.0000
RMSEA                                         0.0158
SRMR                                          0.0463
TLI                                           0.9776
CFI                                           0.9911
AIC                                           0.5119
BIC                                         -25.1319
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

4. Other Useful Functions

4.1. Analysis of Correlation/Covariance Structure with Weighted Least Squares

The `wls()` function may be used to fit a correlation/covariance structure with weighted least squares (WLS) estimation method. The following example fits a one-factor CFA model on the correlation matrix with WLS estimation method. It should be noted that the only off-diagonal elements are used when a correlation structure is fitted.

```
R> #### Analysis of correlation structure
R> R1 <- matrix(c(1.00, 0.22, 0.24, 0.18,
                  0.22, 1.00, 0.30, 0.22,
                  0.24, 0.30, 1.00, 0.24,
                  0.18, 0.22, 0.24, 1.00), ncol=4, nrow=4)
R> n <- 1000
```

```
R> acovR1 <- asyCov(R1, n)
R> ## One-factor CFA model
R> (A1 <- cbind(matrix(0, nrow=5, ncol=4),
               matrix(c("0.2*a1","0.2*a2","0.2*a3","0.2*a4",0),
               ncol=1)))

     [,1] [,2] [,3] [,4] [,5]
[1,] "0"  "0"  "0"  "0"  "0.2*a1"
[2,] "0"  "0"  "0"  "0"  "0.2*a2"
[3,] "0"  "0"  "0"  "0"  "0.2*a3"
[4,] "0"  "0"  "0"  "0"  "0.2*a4"
[5,] "0"  "0"  "0"  "0"  "0"

R> A1 <- as.mxMatrix(A1)
R> (S1 <- diag(c("0.2*e1","0.2*e2","0.2*e3","0.2*e4",1)))

     [,1]     [,2]     [,3]     [,4]     [,5]
[1,] "0.2*e1" "0"      "0"      "0"      "0"
[2,] "0"      "0.2*e2" "0"      "0"      "0"
[3,] "0"      "0"      "0.2*e3" "0"      "0"
[4,] "0"      "0"      "0"      "0.2*e4" "0"
[5,] "0"      "0"      "0"      "0"      "1"

R> S1 <- as.mxMatrix(S1)
R> ## The first 4 variables are observed while the last one is latent.
R> (F1 <- create.Fmatrix(c(1,1,1,1,0), name="F1"))

FullMatrix 'F1'

@labels: No labels assigned.

@values
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    0    1    0    0    0
[3,]    0    0    1    0    0
[4,]    0    0    0    1    0

@free: No free parameters.

@lbound: No lower bounds assigned.

@ubound: No upper bounds assigned.
```

```
R> wls.fit1 <- wls(Cov=R1, asyCov=acovR1, n=n, Fmatrix=F1, Smatrix=S1, Amatrix=A1,
                   cor.analysis=TRUE, diag.constraints=TRUE, intervals="LB")

Running WLS Analysis of Correlation Structure

R> summary(wls.fit1)

Call:
wls(Cov = R1, asyCov = acovR1, n = n, Amatrix = A1, Smatrix = S1,
    Fmatrix = F1, diag.constraints = TRUE, cor.analysis = TRUE,
    intervals.type = "LB")

95% confidence intervals: Likelihood-based statistic
Coefficients:
             Estimate Std.Error  lbound  ubound z value
Amatrix[1,5]  0.42159        NA 0.34435 0.50017      NA
Amatrix[2,5]  0.52376        NA 0.44603 0.60472      NA
Amatrix[3,5]  0.57092        NA 0.49204 0.65446      NA
Amatrix[4,5]  0.42159        NA 0.34435 0.50017      NA
Smatrix[1,1]  0.82226        NA 0.74983 0.88141      NA
Smatrix[2,2]  0.72567        NA 0.63430 0.80105      NA
Smatrix[3,3]  0.67405        NA 0.57165 0.75790      NA
Smatrix[4,4]  0.82226        NA 0.74990 0.88149      NA
             Pr(>|z|)
Amatrix[1,5]       NA
Amatrix[2,5]       NA
Amatrix[3,5]       NA
Amatrix[4,5]       NA
Smatrix[1,1]       NA
Smatrix[2,2]       NA
Smatrix[3,3]       NA
Smatrix[4,4]       NA

Goodness-of-fit indices:
                                              Value
Sample size                               1000.0000
Chi-square of target model                   0.0134
DF of target model                           2.0000
p value of target model                      0.9933
Number of constraints imposed on "Smatrix"   4.0000
DF manually adjusted                         0.0000
Chi-square of independent model            207.8671
DF of independent model                      6.0000
```

```
RMSEA                                              0.0000
SRMR                                               0.0012
TLI                                                1.0295
CFI                                                1.0000
AIC                                               -3.9866
BIC                                              -13.8021
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

4.2. Likelihood-based Confidence Intervals

Most CIs are based on the estimated standard errors. These are known as Wald CIs. Wald CIs are symmetric around the estimates. The Wald CIs might be outside of the meaningful boundaries, for example, a negative lower limit for the variance or larger than 1 for a correlation coefficient. A preferable approach is to construct the CIs based on the likelihood. This is known as the likelihood based CI (Cheung 2009a; Neale and Miller 1997). Likelihood based CIs on the parameter estimates can be required by specifying `intervals.type='LB'` argument. This is especially useful in constructing confidence intervals for the variance components.

```
R> ## Random-effects meta-analysis with ML
R> summary( meta(y=di, v=vi, data=Becker83, intervals.type="LB") )

Running Meta analysis with ML

Call:
meta(y = di, v = vi, data = Becker83, intervals.type = "LB")

95% confidence intervals: Likelihood-based statistic
Coefficients:
            Estimate Std.Error    lbound    ubound z value
Intercept1  0.174734  0.113378 -0.052165  0.437627  1.5412
Tau2_1_1    0.077376  0.054108  0.015124  0.302999  1.4300
           Pr(>|z|)
Intercept1   0.1233
Tau2_1_1     0.1527

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Heterogeneity indices (I2) and their 95% likelihood-based CIs:
                              lbound Estimate ubound
Intercept1: I2 (Q statistic) 0.28410  0.67182 0.8888
```

```
Number of studies (or clusters): 10
Number of observed statistics: 10
Number of estimated parameters: 2
Degrees of freedom: 8
-2 log likelihood: 7.928307
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)

R> ## Mixed-effects meta-analysis with "log(items)" as a predictor
R> summary( meta(y=di, v=vi, x=log(items), data=Becker83, intervals.type="LB") )

Running Meta analysis with ML

Call:
meta(y = di, v = vi, x = log(items), data = Becker83, intervals.type = "LB")

95% confidence intervals: Likelihood-based statistic
Coefficients:
              Estimate   Std.Error      lbound      ubound
Intercept1 -3.2015e-01  1.0981e-01 -5.4408e-01 -7.7598e-02
Slope1_1    2.1088e-01  4.5084e-02  1.1838e-01  3.0789e-01
Tau2_1_1    1.0000e-10  2.0095e-02  1.0000e-10  5.7947e-02
           z value  Pr(>|z|)
Intercept1 -2.9154  0.003552 **
Slope1_1    4.6774 2.905e-06 ***
Tau2_1_1    0.0000  1.000000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Q statistic on homogeneity of effect sizes: 30.64949
Degrees of freedom of the Q statistic: 9
P value of the Q statistic: 0.0003399239
Explained variances (R2):
                           Estimate
y1: Tau2 (no predictor)      0.0774
y1: Tau2 (with predictors)   0.0000
y1: R2                       1.0000

Number of studies (or clusters): 10
Number of observed statistics: 10
Number of estimated parameters: 3
Degrees of freedom: 7
-2 log likelihood: -4.208024
OpenMx status1: 0 ("0" and "1": considered fine; other values indicate problems)
```

4.3. Reading External Data Files

Data sets are most likely stored externally. metaSEM reads three types of data formats. The first type is full correlation/covariance matrices, for example, fullmat.dat is the same as the built-in data set Cheung09. Missing values are represented by NA (the default option). Suppose you have saved it at d:\fullmat.dat, you may read it by using the following command in R:

```
my.df <- readFullMat(file="d:/fullmat.dat")
```

The second type is lower triangle correlation/covariance matrices, for example, lowertriangle.dat. Missing values are represented by the strings NA. Suppose you have saved it at d:\lowertriangle.dat, you may read it by using the following command in R:

```
my.df <- readLowTriMat(file = "d:/lowertriangle.dat", no.var = 9, na.strings="NA")
```

The third type is vectors of correlation/covariance elements based on column vectorization. One row represents one study, for example, stackvec.dat. Suppose you have saved it at d:\stackvec.dat, you may read it by using the following R command:

```
my.df <- readStackVec(file="d:/stackvec.dat")
```

5. Installation

First of all, you need R to run it. Since metaSEM uses OpenMx as the workhorse, OpenMx has to be installed. To install OpenMx, run the following command inside an R session:

```
install.packages('OpenMx', repos='http://openmx.psyc.virginia.edu/packages/')
```

See http://openmx.psyc.virginia.edu/installing-openmx for the details on how to install OpenMx. Moreover, metaSEM also depends on the ellipse package that can be installed by the following command inside an R session:

```
install.packages('ellipse')
```

5.1. Windows platform

Download the Windows binary of metaSEM. If the file is saved at d:\. Run the following command inside an R session:

```
install.packages(pkgs="d:/metaSEM_0.7-1.zip", repos=NULL)
```

Please note that d:\ in Windows is represented by either d:/ or d:\\ in R.

5.2. Linux and Mac OS X platform

Download the source package of `metaSEM`. Run the following command (as Root) inside an R session:

```
install.packages(pkgs="metaSEM_0.7-1.zip", repos=NULL, type="source")
```

6. Acknowledgements

This package cannot be written without `R` and `OpenMx`. Contributions by the R Development Core Team and the OpenMx Core Development Team are highly appreciated.

References

Arthur W, Bennett W, Huffcutt AI (2001). *Conducting meta-analysis using SAS*. Routledge. ISBN 9780805838091.

Becker BJ (1983). "Influence again: A comparison of methods for meta-analysis." In *Paper presented at the annual meeting of the American Educational Research Association*. Montreal.

Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA (1998). "Meta-analysis of multiple outcomes by regression with random effects." *Statistics in Medicine*, **17**(22), 2537–2550. doi:10.1002/(SICI)1097-0258(19981130)17:22<2537::AID-SIM953>3.0.CO;2-C.

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011). "OpenMx: An Open Source Extended Structural Equation Modeling Framework." *Psychometrika*, pp. 1–12. URL http://dx.doi.org/10.1007/s11336-010-9200-6.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009). *Introduction to meta-analysis*. John Wiley & Sons, Chichester, West Sussex, U.K. ; Hoboken. ISBN 0470057246.

Cheung MWL (2008). "A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling." *Psychological Methods*, **13**(3), 182–202. doi:10.1037/a0013163.

Cheung MWL (2009a). "Constructing approximate confidence intervals for parameters with structural equation models." *Structural Equation Modeling: A Multidisciplinary Journal*, **16**(2), 267–294. doi:10.1080/10705510902751291.

Cheung MWL (2009b). "TSSEM: A LISREL syntax generator for two-stage structural equation modeling (Version 1.11)." URL `http://courses.nus.edu.sg/course/psycwlm/internet/tssem.zip`.

Cheung MWL (2011a). "Implementing restricted maximum likelihood (REML) estimation in structural equation models." *Structural Equation Modeling: A Multidisciplinary Journal*. Manuscript accepted for publication.

Cheung MWL (2011b). "Multivariate meta-analysis as structural equation models." *Structural Equation Modeling: A Multidisciplinary Journal*. Manuscript accepted for publication.

Cheung MWL, Chan W (2005a). "Classifying correlation matrices into relatively homogeneous subgroups: a cluster analytic approach." *Educational and Psychological Measurement*, **65**(6), 954–979. `doi:10.1177/0013164404273946`.

Cheung MWL, Chan W (2005b). "Meta-analytic structural equation modeling: a two-stage approach." *Psychological Methods*, **10**(1), 40–64. `doi:10.1037/1082-989X.10.1.40`.

Cheung MWL, Chan W (2009). "A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling." *Structural Equation Modeling: A Multidisciplinary Journal*, **16**(1), 28–53. `doi:10.1080/10705510802561295`.

Cochran W (1954). "The combination of estimates from different experiments." *Biometrics*, **10**(1), 101–129. ISSN 0006-341X.

Cooper H, Valentine JC, Charlton K, Melson A (2003). "The effects of modified school calendars on student achievement and on school and community attitudes." *Review of Educational Research*, **73**(1), 1 –52. `doi:10.3102/00346543073001001`. URL `http://rer.sagepub.com/content/73/1/1.abstract`.

Digman JM (1997). "Higher-order factors of the Big Five." *Journal of Personality and Social Psychology*, **73**(6), 1246–1256. `doi:10.1037/0022-3514.73.6.1246`.

Gasparrini A (2012). *mvmeta: multivariate meta-analysis and meta-regression.* R package version 0.2.4, URL `http://CRAN.R-project.org/package=mvmeta`.

Guolo A, Varin C (2012). *metaLik: Likelihood inference in meta-analysis and meta-regression models.* R package version 0.3, URL `http://CRAN.R-project.org/package=metaLik`.

Harville DA (1977). "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems." *Journal of the American Statistical Association*, **72**(358), 320–338. `doi:10.2307/2286796`.

Hedges LV, Olkin I (1985). *Statistical methods for meta-analysis.* Academic Press, Orlando, FL.

Higgins JPT, Thompson SG (2002). "Quantifying heterogeneity in a meta-analysis." *Statistics in Medicine*, **21**(11), 1539–1558. ISSN 1097-0258. doi:10.1002/sim.1186. URL http://onlinelibrary.wiley.com.libproxy1.nus.edu.sg/doi/10.1002/sim.1186/abstract.

Hunter JE, Schmidt FL (2004). *Methods of meta-analysis: correcting error and bias in research findings*. Sage. ISBN 9781412904797.

Konstantopoulos S (2011). "Fixed effects and variance components estimation in three-level meta-analysis." *Research Synthesis Methods*, **2**(1), 61–76. ISSN 1759-2887. doi:10.1002/jrsm.35. URL http://onlinelibrary.wiley.com/doi/10.1002/jrsm.35/abstract.

Lipsey MW, Wilson D (2000). *Practical Meta-Analysis*. 1 edition. Sage Publications, Inc. ISBN 0761921680.

Lumley T (2009). *rmeta: Meta-analysis*. R package version 2.16, URL http://CRAN.R-project.org/package=rmeta.

Neale MC, Miller MB (1997). "The use of likelihood-based confidence intervals in genetic models." *Behavior Genetics*, **27**(2), 113–120. doi:10.1023/A:1025681223921.

Patterson HD, Thompson R (1971). "Recovery of inter-block information when block sizes are unequal." *Biometrika*, **58**(3), 545 –554. doi:10.1093/biomet/58.3.545.

Raudenbush SW (2009). "Analyzing effect sizes: random effects models." In HM˜Cooper, LV˜Hedges, JC˜Valentine (eds.), *The handbook of research synthesis and meta-analysis*, 2 edition, p. 295–315. Russell Sage Foundation, New York.

Schwarzer G (2012). *meta: Meta-Analysis with R*. R package version 2.1-0, URL http://CRAN.R-project.org/package=meta.

Sterne J (2009). *Meta-Analysis: An Updated Collection from the Stata Journal*. 1 edition. Stata Press. ISBN 1597180491.

Takkouche B, Cadarso-Suárez C, Spiegelman D (1999). "Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis." *American Journal of Epidemiology*, **150**(2), 206 –215. URL http://aje.oxfordjournals.org/content/150/2/206.abstract.

Viechtbauer W (2010). "Conducting meta-analyses in R with the metafor package." *Journal of Statistical Software*, **36**(3), 1–48. ISSN 1548-7660. URL http://www.jstatsoft.org/v36/i03/.

Whitehead A (2002). *Meta-analysis of controlled clinical trials*. John Wiley & Sons, Ltd, Chichester, UK. ISBN 0471983705.

Xiong C, Miller JP, Morris JC (2010). "Measuring study-specific heterogeneity in meta-analysis: application to an antecedent biomarker study of alzheimer's disease." *Statistics in Biopharmaceutical Research*, **2**(3), 300–309. ISSN 1946-6315. doi:10.1198/sbr.2009.0067. URL http://pubs.amstat.org/doi/abs/10.1198/sbr.2009.0067.

Affiliation:

Mike W.-L. Cheung
Department of Psychology
Faculty of Arts and Social Sciences
National University of Singapore
Singapore 117570
E-mail: mikewlcheung@nus.edu.sg
URL: http://courses.nus.edu.sg/course/psycwlm/internet/metaSEM/