Structural equation modeling with clustered data:

Means, motives, and opportunities

Kris Preacher
University of Kansas
March 6, 2009

"There are more multilevel SEMs than are dreamt of in your philosophy, Horatio."

—Rod McDonald, SMEP 2008

Structural equation modeling (SEM)

Throughout, I'll emphasize:

- The **means** by which MSEM can be employed in practice.
- The **motives** for using MSEM rather than simpler methods.
- The modeling **opportunities** made possible by MSEM that cannot be exploited within more restrictive modeling frameworks like SEM and MLM.

Structural equation modeling (SEM)

SEM is a statistical modeling framework that:

- represents hypotheses as a system of simultaneous linear equations.
- permits the estimation of direct and indirect regression effects.
- tests the entire model as a global hypothesis.
- separates error variance from reliable variance, automatically correcting effects for attenuation due to unreliability.

SEM is hugely popular, and is now used routinely in virtually every subdiscipline in psychology.

Most other methods used in psychology are special cases of SEM.

Structural equation modeling (SEM)

One way to represent SEM in matrix equation form is the "LISREL" model:

Data model: $\mathbf{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$

Structural model: $\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta}$

...together imply:

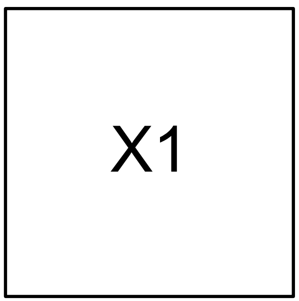
Covariance structure:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\left(\mathbf{I} - \mathbf{B}\right)^{-1}\left(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi}\right)\left(\mathbf{I} - \mathbf{B}\right)^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

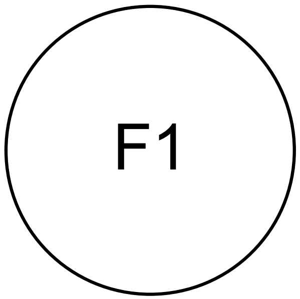
Structural equation modeling (SEM)

SEM can also be represented in diagram form:

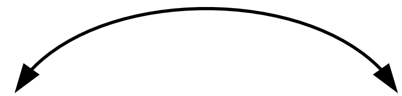
Measured variable:



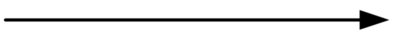
Latent variable:



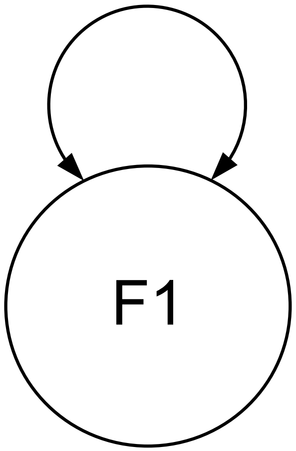
Covariance ("sling"):



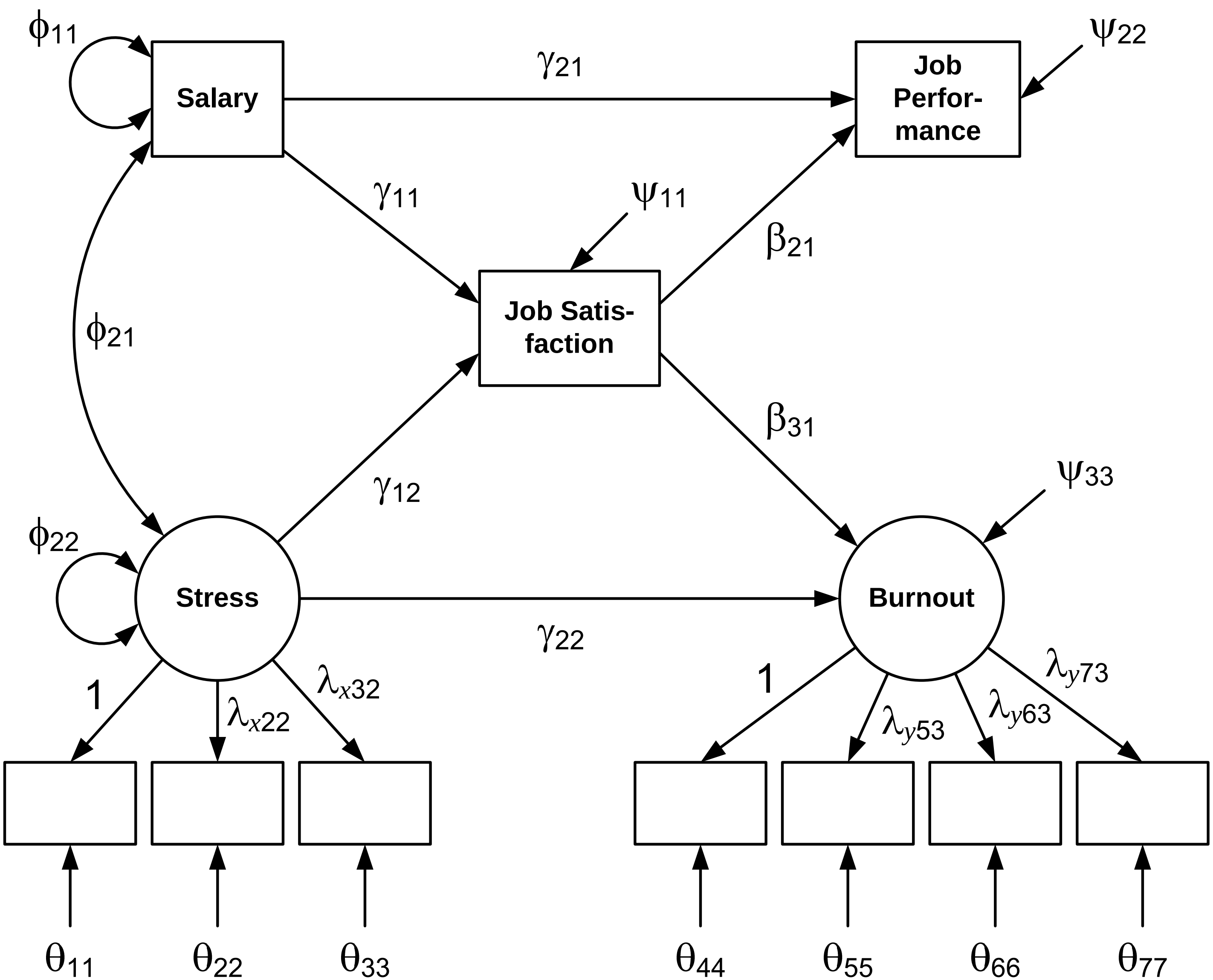
Directed path ("arrow"):



Variance:



Structural equation modeling (SEM)



SEM assumptions

In order to trust the results of SEM (parameter estimates, standard errors, model fit), certain assumptions have to be met to a reasonable degree:

- Errors should be normally distributed with means = 0.
- Errors should be uncorrelated with each other and with predictors.
- Error variances should be finite and homoscedastic.
- Effects should be linear (unless special steps are taken).
- The model should be properly specified (i.e., all the constraints should be appropriate).
- Observations should be **independent** (i.e., knowing one case's data vector should tell us nothing about another case's data vector).

Clustered data

Clustered data violate the assumption of independence.

What are clustered data?

Clustered data commonly occur when some cases are linked in ways that enhance the probability of similar responses.

Examples:

"Level 1"		**"Level 2"**
students		classrooms
corporations		nations
spouses	nested within	couples
voters		districts
repeated measures		individuals

Clustered data

The consequences of violating independence typically include **biased standard errors** (SEs).

When bias exists in SEs, it is usually **downward** bias, leading to spuriously inflated significance.

Structural equation modeling (SEM)

In other words, if we are going to use ordinary SEM with clustered data, we cannot trust:

- standard errors
- tests of parameter estimates
- model fit indices

Applying ordinary SEM to clustered data would be a **serious mistake**.

Multilevel modeling (MLM)

A standard and popular strategy for handling clustered data is multilevel modeling (MLM). This is one common way to express a multilevel model:

Level-1 expression...

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + e_{ij}$$

Multilevel modeling (MLM)

A standard and popular strategy for handling clustered data is multilevel modeling (MLM). This is one common way to express a multilevel model:

Level-1 expression...

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + e_{ij}$$

Level-2 expression...

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_{1j} + \gamma_{12} w_{2j} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} w_{1j} + \gamma_{22} w_{2j} + u_{2j}$$

Multilevel modeling (MLM)

A standard and popular strategy for handling clustered data is multilevel modeling (MLM). This is one common way to express a multilevel model:

Level-1 expression...

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + e_{ij}$$

Level-2 expression...

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_{1j} + \gamma_{12} w_{2j} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} w_{1j} + \gamma_{22} w_{2j} + u_{2j}$$

Combined (reduced) form...

$$y_{ij} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + \gamma_{10} x_{1ij} + \gamma_{11} x_{1ij} w_{1j} + \gamma_{12} x_{1ij} w_{2j} + \gamma_{20} x_{2ij} + \gamma_{21} x_{2ij} w_{1j} + \gamma_{22} x_{2ij} w_{2j} + u_{0j} + u_{1j} x_{1ij} + u_{2j} x_{2ij} + e_{ij}$$

Multilevel modeling (MLM)

MLM has several advantages:

- Partitions variance in the dependent variable.
- Returns ML estimates of intercepts, slopes, and (co)variance components.
- Flexibly handles predictor variables at any level of measurement.
- Handles an arbitrary number of "nesting" levels.

Multilevel modeling (MLM)

MLM has **severe limitations**, however:

- Multivariate models are possible, but extremely difficult to specify (data management nightmare).
- Latent variables are (nearly) impossible to include, and require all factor loadings = 1. So measurement error is a problem.
- Model fit indices are not available (only a few selection criteria).
- Level-2 outcome variables are not an option.*
- Level-1 and level-2 effects are often **conflated**. If steps are taken to "unconflate" them, bias is introduced.*

Multilevel modeling (MLM)

Because MLM **cannot accommodate Level-2 outcomes**, three problematic strategies have been used:

Two-step analyses

Ad hoc combination of results from OLS and MLM analyses

Unclear whether parameters from OLS and MLM can be combined

Aggregation to Level-2

Severe reduction in sample size → loss of power

Discounts within-unit variation

Increases risk of ecological and atomistic fallacies

Gives equal weight to small groups and large groups

May not fairly represent group-level constructs

Disaggregation to Level-1

Ignores clustering altogether

Fails to separate Within and Between effects

Spuriously inflates power for testing effects

Multilevel modeling (MLM)

Conflation of "Between" and "Within" effects:

One way to understand the **conflation** of Between and Within effects:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + e_{ij}$$

$$y_{ij} = \gamma_{00} + \gamma_{10} \left(x_{ij} - \bar{x}_{.j} + \bar{x}_{.j} \right) + u_{0j} + e_{ij}$$

$$y_{ij} = \gamma_{00} + \gamma_{10} \left(x_{ij} - \bar{x}_{.j} \right) + \gamma_{10} \bar{x}_{.j} + u_{0j} + e_{ij}$$

implicit constraint on the Within and Between effects of X

Multilevel modeling (MLM)

Conflation of "Between" and "Within" effects:

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{.j}\right) + \gamma_{10}\bar{x}_{.j} + u_{0j} + e_{ij}$$

constraint

We could simply estimate **separate** Between and Within effects and *test* whether they are different:

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{.j}\right) + \gamma_{01}\bar{x}_{.j} + u_{0j} + e_{ij}$$

Within effect

Between effect

Multilevel modeling (MLM)

Conflation of "Between" and "Within" effects:

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{.j}\right) + \gamma_{10}\bar{x}_{.j} + u_{0j} + e_{ij}$$

constraint

We could simply estimate **separate** Between and Within effects and *test* whether they are different:

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{.j}\right) + \gamma_{01}\bar{x}_{.j} + u_{0j} + e_{ij}$$

Within effect

Between effect

...but the Between effect will be **biased**.

Multilevel modeling (MLM)

Applying MLM to test complex stuctural hypotheses can lead to extreme difficulty and serious misspecification.

We simply cannot do some things within the MLM framework.

Separation of Between and Within effects using MLM leads to biased Between effects.

What is one to do??

Multilevel SEM

Multilevel structural equation modeling (MSEM): a collection of methods designed to permit SEM (or SEM-like) modeling with clustered / multilevel data.

At least **17** methods have been suggested to combine SEM and MLM.

Some are better than others, and all have advantages and disadvantages.

Most use either SEM or MLM as a "base method" and attempt to expand it.

Dimensions along which the methods differ:

- speed / efficiency
- generality / flexibility of the model
- ease of implementation
- software availability

Multilevel SEM

We won't cover all these methods.

Schmidt's (1969) method

Many of these methods trace their origin to the dissertation of Bill Schmidt, University of Chicago (now MSU).

Schmidt developed a ML approach for fitting SEMs to two-level data in 1969.

Schmidt's (1969) method

Schmidt's method involves decomposing observed scores into independent **Between** and (pooled) **Within** components.

$$\mathbf{y} = \mathbf{y}_B + \mathbf{y}_W$$

Covariance matrices:

$$\begin{aligned} \operatorname{var}(\mathbf{y}) &= \operatorname{var}(\mathbf{y}_B + \mathbf{y}_W) \\ &= \operatorname{var}(\mathbf{y}_B) + \operatorname{var}(\mathbf{y}_W) \quad \rightarrow \quad \boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W \end{aligned}$$

$\boldsymbol{\Sigma}_B$ and $\boldsymbol{\Sigma}_W$ are estimated with $\mathbf{S}_B$ and $\mathbf{S}_W$ and separately analyzed.

We might want to fit one SEM to the person-level covariances (treating people as cases) and a different SEM to the cluster-level covariances (treating clusters as cases).

Schmidt's (1969) method

How do we obtain the $\mathbf{S}_B$ and $\mathbf{S}_W$ matrices to use as input?

Schmidt (1969) computed estimates of $\mathbf{S}_B$ and $\mathbf{S}_W$ directly from raw data. Muthén (1989, 1990) and Goldstein & McDonald (1988) adopt this method.

$$\mathbf{S}_W = \frac{1}{N-J} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \left(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{.j}\right)\left(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{.j}\right)'$$

$$\mathbf{S}_B = \frac{1}{J-1} \sum_{j=1}^{J} n_j \left(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..}\right)\left(\overline{\mathbf{y}}_{.j} - \overline{\mathbf{y}}_{..}\right)'$$

(from Kaplan, 2009)

Schmidt's (1969) method

Cronbach (1976) and Härnqvist (1978) further suggested scaling these $\mathbf{S}_B$ and $\mathbf{S}_W$ matrices by the products of the SDs of the total scores.

This results in two orthogonal covariance matrices that sum to the total sample $\mathbf{R}$.

These matrices, in turn, can be submitted to further analyses.

Schmidt's (1969) method

Limitations of Schmidt's method:

- There is no clear way to introduce strictly group-level variables.
- The method cannot accommodate partially missing data because covariance matrices are used as input.
- The Between and pooled Within matrices are not independent. Separate analysis fails to recognize this fact.

Schmidt's (1969) method

Goldstein (1987, 1995) introduced a way to obtain the $\mathbf{S}_B$ and $\mathbf{S}_W$ matrices by "tricking" MLM software into running a multivariate model.

The idea is to put all DVs into one column of the data matrix, then "dummy-code" the predictor set and include the dummies as random "predictors" (intercepts).

This addresses problems of unequal cluster sizes and missing data, but the method is difficult to implement and the $\mathbf{S}_B$ and $\mathbf{S}_W$ matrices are themselves estimates. This estimation error is not addressed in subsequent modeling (Hox, 2000b, 2002; Hox & Maas, 2004).

j	Y	M
1	0.57	0.11
1	1.21	2.11
⋮	⋮	⋮
1	0.23	0.73
2	-1.15	-0.36
2	-3.72	-2.97
⋮	⋮	⋮
2	-3.86	-2.56

Restructuring the data

j	Z	S_Y	S_M
1	0.57	1	0
1	0.11	0	1
1	1.21	1	0
1	2.11	0	1
⋮	⋮	⋮	⋮
1	0.23	1	0
1	0.73	0	1
2	-1.15	1	0
2	-0.36	0	1
2	-3.72	1	0
2	-2.97	0	1
⋮	⋮	⋮	⋮
2	-3.86	1	0
2	-2.56	0	1

Muthén's method

Two influential lines of work were inspired by Schmidt's (1969) dissertation—one led by Bengt Muthén and one led by Rod McDonald and Harvey Goldstein.

Muthén camp
(Muthén, 1989, 1990, 1991, 1994; Muthén & Satorra, 1995)

McDonald / Goldstein camp
(Goldstein & McDonald, 1988, McDonald, 1993, 1994, McDonald & Goldstein, 1989)

The method is implemented in several versions of Mplus, which also provides a robust quasi-likelihood estimator (MUML) that assumes balanced group size.

Muthén's method

Versions of this general model are described in detail by McDonald and Goldstein (1989), McDonald (1993), and Muthén (1990).

Also discussed by **many** other authors (e.g., Bovaird, 2007; Cheung & Au, 2005; Duncan, Duncan, & Alpert, 1997; Heck, 2001; Heck & Thomas, 2000; Hox, 1995, 2002; Hox & Maas, 2001, 2004; Kamata, Bauer, & Miyazaki, 2008; Kaplan, 1998, 2000, 2009; Kaplan & Elliott, 1997a, 1997b; Peugh, 2006; Selig, Card, & Little, 2008; Stapleton, 2006).

Raudenbush (1995) extended the general method to the case of unbalanced Level-1 sample sizes by considering the data "complete" but with missing cases, employing an E-M algorithm by using an SEM program like EQS in tandem with separate E-step Fortran program.

Other methods

A number of other creative approaches have been suggested:

- Bauer (2003); Curran (2003)
- McArdle & Hamagami (1996)
- Mehta & Neale (2005)
- Rovine & Molenaar (1998, 2000, 2001)

Discussions of these and other models can be found in Bovaird (2007) and Selig, Card, and Little (2008).

Random slopes

Ideally, MSEM should be able to accommodate random slopes—i.e., allow for the possibility that effects vary across clusters.

The inability of the two-matrix approaches to model random slopes has been noted as a **significant hindrance** to exploiting the full potential of MSEM (e.g., Kaplan, 2000; Lee, 2007).

Overcoming this disadvantage of MSEM, recent advances have made the inclusion of random slopes possible.

To date, at least **five methods** exist for applying MSEM with random slopes, but most have limitations that mitigate their generality.

Random slopes: Chou, Bentler, & Pentz's method

Chou et al. (1998; also Hanushek, 1974) describe a **two-stage slopes-as-outcomes approach**.

It involves first estimating the parameters of a structural equation model **within each Level-2 unit**, then using the resulting estimates within each unit as input for a second-stage Level-2 analysis.

Advantages:

- Permits Level-2 variability in any parameter (including slopes)

Limitations:

- Biased estimation of group-level variability in random coefficients
- No correction for unbalanced group sizes
- Non-simultaneous estimation of the Between and Within components of a model
- Unstable or unusable with groups below a certain size

Random slopes: Raudenbush's method

Raudenbush et al. (1991) and Raudenbush and Sampson (1999) describe an MLM-based MSEM method that adds a restricted measurement model to an otherwise ordinary application of MLM.

Advantages:

- Allows MLM with latent variables
- Permits random slopes for structural coefficients
- Easily handles missing data, unbalanced n_j, and multiple levels
- Ordinary MLM software can be used for estimation

Limitations:

- Fit indices are not available
- Factor loadings in the measurement model are required to be equal
- Complex models still very difficult to specify

Random slopes: Ansari and Jedidi's method

Ansari, Jedidi, and Jagpal (2000), Jedidi and Ansari (2001), and Ansari, Jedidi, and Dube (2002) propose a **Bayesian method** that can accommodate random covariance structures.

Advantages:

- Flexible incorporation of prior information
- Estimation of individual/group-specific estimates of model parameters (and functions of those parameters)
- No need to rely on asymptotic theory
- Accommodates unbalanced cluster sizes
- Can accommodate random slopes (<u>and covariances</u>)

Limitations:

- The method is exceedingly difficult to implement
- No software can accommodate it (except maybe WinBUGS)

Random slopes: The GLLAMM method

Rabe-Hesketh, Skrondal, and Pickles (2004a) and Skrondal and Rabe-Hesketh (2004) describe a very general multilevel latent variable modeling framework.

Advantages:

- Able to accommodate complex models and data types
- Can accommodate random slopes
- Implemented in the Stata add-on package GLLAMM
- Can handle more than two levels of nesting

Limitations:

- Uses numerical integration even when not necessary, significantly increasing computation time, often rendering complex models computationally intractable (Bauer, 2003; Kamata et al., 2008)
- GLLAMM cannot accommodate random slopes for latent covariates, which limits what can be tested

Random slopes: Muthén & Asparouhov's method

Muthén and Asparouhov (2008) describe a very general model for two-level SEM that builds on the earlier work of Muthén (1989, 1990, 1991, 1994; Muthén & Satorra, 1995) and Goldstein and McDonald (1988; McDonald & Goldstein, 1989; McDonald, 1993, 1994).

The model is implemented in Mplus.

Versions 5 and later of Mplus implement ML estimation via an accelerated E-M algorithm described by Lee and Poon (1998), and incorporate further improvements and refinements, some of which are described by Lee and Tsang (1999), Bentler and Liang (2003, 2008), Liang and Bentler (2004), and Asparouhov and Muthén (2007).

Random slopes: Muthén & Asparouhov's method

Advantages:

- Can accommodate missing data, unbalanced n_j, and random slopes
- MLR estimator does not require the assumption of normality
- Yields robust estimates of asymptotic covariances of parameter estimates and χ^2
- More computationally efficient than previous methods
- Implemented in Mplus
- Can separate Between and Within effects
- Can handle continuous, count, ordinal/binary, etc. data, mixtures
- There are many special cases

Limitations:

- Handles only two levels of nesting.

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{X}_i + \boldsymbol{\varepsilon}_i$$ ← **measurement model**

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables
$\boldsymbol{\varepsilon}_i$ ($p \times 1$) and $\boldsymbol{\zeta}$ ($m \times 1$) are vectors of error terms
$\boldsymbol{\eta}_i$ ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

$\boldsymbol{\nu}$ ($p \times 1$) and $\boldsymbol{\alpha}$ ($m \times 1$) are vectors of intercepts
$\boldsymbol{\Lambda}$ ($p \times m$) is a matrix of factor loadings
$\mathbf{K}$ ($p \times q$) and $\boldsymbol{\Gamma}$ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)
$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \nu + \Lambda\eta_i + \mathbf{K}\mathbf{X}_i + \varepsilon_i \quad \longleftarrow \text{measurement model}$$

$$\eta_i = \alpha + \mathbf{B}\eta_i + \Gamma\mathbf{X}_i + \zeta_i \quad \longleftarrow \text{structural model}$$

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables

ε_i ($p \times 1$) and ζ ($m \times 1$) are vectors of error terms

η_i ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

ν ($p \times 1$) and α ($m \times 1$) are vectors of intercepts

Λ ($p \times m$) is a matrix of factor loadings

$\mathbf{K}$ ($p \times q$) and Γ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)

$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \nu + \Lambda\eta_i + \mathbf{K}\mathbf{X}_i + \varepsilon_i \quad \longleftarrow \text{measurement model}$$

$$\eta_i = \alpha + \mathbf{B}\eta_i + \Gamma\mathbf{X}_i + \zeta_i \quad \longleftarrow \text{structural model}$$

Variable vectors

- $\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables
- ε_i ($p \times 1$) and ζ ($m \times 1$) are vectors of error terms
- η_i ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

- ν ($p \times 1$) and α ($m \times 1$) are vectors of intercepts
- Λ ($p \times m$) is a matrix of factor loadings
- $\mathbf{K}$ ($p \times q$) and Γ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)
- $\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \nu + \Lambda \eta_i + \mathbf{K}\mathbf{X}_i + \varepsilon_i$$ ← **measurement model**

$$\eta_i = \alpha + \mathbf{B}\eta_i + \Gamma\mathbf{X}_i + \zeta_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables
ε_i ($p \times 1$) and ζ ($m \times 1$) are vectors of error terms
η_i ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

ν ($p \times 1$) and α ($m \times 1$) are vectors of intercepts
Λ ($p \times m$) is a matrix of factor loadings
$\mathbf{K}$ ($p \times q$) and Γ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)
$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \nu + \Lambda\eta_i + \mathbf{K}\mathbf{X}_i + \varepsilon_i$$ ← **measurement model**

$$\eta_i = \alpha + \mathbf{B}\eta_i + \Gamma\mathbf{X}_i + \zeta_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables
ε_i ($p \times 1$) and ζ ($m \times 1$) are vectors of error terms
η_i ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

ν ($p \times 1$) and α ($m \times 1$) are vectors of intercepts
Λ ($p \times m$) is a matrix of factor loadings
$\mathbf{K}$ ($p \times q$) and Γ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)
$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \nu + \Lambda\eta_i + \mathbf{K}\mathbf{X}_i + \varepsilon_i$$ ← **measurement model**

$$\eta_i = \alpha + \mathbf{B}\eta_i + \Gamma\mathbf{X}_i + \zeta_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables

ε_i ($p \times 1$) and ζ ($m \times 1$) are vectors of error terms

η_i ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

ν ($p \times 1$) and α ($m \times 1$) are vectors of intercepts

Λ ($p \times m$) is a matrix of factor loadings

$\mathbf{K}$ ($p \times q$) and Γ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)

$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{X}_i + \boldsymbol{\varepsilon}_i$$ ← **measurement model**

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables

$\boldsymbol{\varepsilon}_i$ ($p \times 1$) and $\boldsymbol{\zeta}$ ($m \times 1$) are vectors of error terms

$\boldsymbol{\eta}_i$ ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

$\boldsymbol{\nu}$ ($p \times 1$) and $\boldsymbol{\alpha}$ ($m \times 1$) are vectors of intercepts

$\boldsymbol{\Lambda}$ ($p \times m$) is a matrix of factor loadings

$\mathbf{K}$ ($p \times q$) and $\boldsymbol{\Gamma}$ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)

$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The single-level SEM can be expressed, for typical individual i...

$$\mathbf{Y}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{X}_i + \boldsymbol{\varepsilon}_i$$ ← **measurement model**

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i$$ ← **structural model**

Variable vectors

$\mathbf{Y}_i$ ($p \times 1$) is a vector of dependent variables

$\boldsymbol{\varepsilon}_i$ ($p \times 1$) and $\boldsymbol{\zeta}$ ($m \times 1$) are vectors of error terms

$\boldsymbol{\eta}_i$ ($m \times 1$) is a vector of random effects (a.k.a. factors, latent variables)

Parameter matrices

$\boldsymbol{\nu}$ ($p \times 1$) and $\boldsymbol{\alpha}$ ($m \times 1$) are vectors of intercepts

$\boldsymbol{\Lambda}$ ($p \times m$) is a matrix of factor loadings

$\mathbf{K}$ ($p \times q$) and $\boldsymbol{\Gamma}$ ($m \times q$) are matrices of slopes for covariates in $\mathbf{X}_i$ ($q \times 1$)

$\mathbf{B}$ ($m \times m$) is a matrix of structural regression slopes

Muthén & Asparouhov's method

The <u>multilevel</u> SEM simply adds j subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← Level-1 measurement model

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← Level-1 structural model

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← Level-2 structural model

Elements of j-subscripted parameter matrices (i.e., intercepts, slopes, and loadings that vary over clusters) are included in the new vector $\boldsymbol{\eta}_j$ ($r \times 1$). These r random effects have a structural model at Level-2.

New vectors and matrices

$\boldsymbol{\mu}$ ($r \times 1$) is a vector of intercepts of cluster-level equations
$\boldsymbol{\beta}$ ($r \times r$) is a matrix of slopes for random effects
$\boldsymbol{\gamma}$ ($r \times s$) is a matrix of slopes for cluster-level covariates in $\mathbf{X}_j$ ($s \times 1$)
$\boldsymbol{\zeta}$ ($p \times 1$) is a vector of cluster-level residuals

Muthén & Asparouhov's method

The multilevel SEM simply adds j subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\gamma}\mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

Elements of j-subscripted parameter matrices (i.e., intercepts, slopes, and loadings that vary over clusters) are included in the new vector $\boldsymbol{\eta}_j$ ($r \times 1$). These r random effects have a structural model at Level-2.

New vectors and matrices

$\boldsymbol{\mu}$ ($r \times 1$) is a vector of intercepts of cluster-level equations
$\boldsymbol{\beta}$ ($r \times r$) is a matrix of slopes for random effects
$\boldsymbol{\gamma}$ ($r \times s$) is a matrix of slopes for cluster-level covariates in $\mathbf{X}_j$ ($s \times 1$)
$\boldsymbol{\zeta}$ ($p \times 1$) is a vector of cluster-level residuals

Muthén & Asparouhov's method

The multilevel SEM simply adds j subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

Elements of j-subscripted parameter matrices (i.e., intercepts, slopes, and loadings that vary over clusters) are included in the new vector $\boldsymbol{\eta}_j$ ($r \times 1$). These r random effects have a structural model at Level-2.

New vectors and matrices

$\boldsymbol{\mu}$ ($r \times 1$) is a vector of intercepts of cluster-level equations
$\boldsymbol{\beta}$ ($r \times r$) is a matrix of slopes for random effects
$\boldsymbol{\gamma}$ ($r \times s$) is a matrix of slopes for cluster-level covariates in $\mathbf{X}_j$ ($s \times 1$)
$\boldsymbol{\zeta}$ ($p \times 1$) is a vector of cluster-level residuals

Muthén & Asparouhov's method

The multilevel SEM simply adds j subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

Elements of j-subscripted parameter matrices (i.e., intercepts, slopes, and loadings that vary over clusters) are included in the new vector $\boldsymbol{\eta}_j$ $(r \times 1)$. These r random effects have a structural model at Level-2.

New vectors and matrices

$\boldsymbol{\mu}$ $(r \times 1)$ is a vector of intercepts of cluster-level equations

$\boldsymbol{\beta}$ $(r \times r)$ is a matrix of slopes for random effects

$\boldsymbol{\gamma}$ $(r \times s)$ is a matrix of slopes for cluster-level covariates in $\mathbf{X}_j$ $(s \times 1)$

$\boldsymbol{\zeta}$ $(p \times 1)$ is a vector of cluster-level residuals

Muthén & Asparouhov's method

The multilevel SEM simply adds j subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

Elements of j-subscripted parameter matrices (i.e., intercepts, slopes, and loadings that vary over clusters) are included in the new vector $\boldsymbol{\eta}_j$ ($r \times 1$). These r random effects have a structural model at Level-2.

New vectors and matrices

$\boldsymbol{\mu}$ ($r \times 1$) is a vector of intercepts of cluster-level equations
$\boldsymbol{\beta}$ ($r \times r$) is a matrix of slopes for random effects
$\boldsymbol{\gamma}$ ($r \times s$) is a matrix of slopes for cluster-level covariates in $\mathbf{X}_j$ ($s \times 1$)
$\boldsymbol{\zeta}$ ($p \times 1$) is a vector of cluster-level residuals

Muthén & Asparouhov's method

The multilevel SEM simply adds *j* subscripts to matrices to denote that they may vary across clusters:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

Furthermore,

$$\boldsymbol{\varepsilon}_{ij} \sim MVN\left(\mathbf{0}, \boldsymbol{\Theta}\right)$$

$$\boldsymbol{\zeta}_{ij} \sim MVN\left(\mathbf{0}, \boldsymbol{\Psi}\right)$$

$$\boldsymbol{\zeta}_j \sim MVN\left(\mathbf{0}, \boldsymbol{\psi}\right)$$

Muthén & Asparouhov's method

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

To clarify the role of η_j, in a fully-random model it would be: →

$$\boldsymbol{\eta}_j = \begin{bmatrix} \text{vec}\{\boldsymbol{\nu}_j\} \\ \text{vec}\{\boldsymbol{\Lambda}_j\} \\ \text{vec}\{\mathbf{K}_j\} \\ \text{vec}\{\boldsymbol{\alpha}_j\} \\ \text{vec}\{\mathbf{B}_j\} \\ \text{vec}\{\boldsymbol{\Gamma}_j\} \end{bmatrix}$$

Notice that the general model can accommodate ***random slopes***, an extremely valuable ability absent in most other MSEM approaches.

Muthén & Asparouhov's method

$$\mathbf{Y}_{ij} = \nu_j + \Lambda_j \eta_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \varepsilon_{ij}$$ ← Level-1 measurement model

$$\eta_{ij} = \alpha_j + \mathbf{B}_j \eta_{ij} + \Gamma_j \mathbf{X}_{ij} + \zeta_{ij}$$ ← Level-1 structural model

$$\eta_j = \mu + \beta \eta_j + \gamma \mathbf{X}_j + \zeta_j$$ ← Level-2 structural model

This general model contains as special cases:

- **Multilevel models**
- **Structural equation models**
  - ❑ exploratory and confirmatory factor analysis
  - ❑ path analysis
  - ❑ regression models

...and therefore

- **All models in the SEM, MLM, or regression frameworks**

Multilevel SEM

Special case: **Single-level path analysis**

Path analysis may be specified by fixing the "grey" matrices to zero:

$$\mathbf{Y}_{ij} = \nu_j + \mathbf{\Lambda}_j \eta_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \varepsilon_{ij}$$ ← **Level-1 measurement model**

$$\eta_{ij} = \alpha_j + \mathbf{B}_j \eta_{ij} + \mathbf{\Gamma}_j \mathbf{X}_{ij} + \zeta_{ij}$$ ← **Level-1 structural model**

$$\eta_j = \mu + \beta \eta_j + \gamma \mathbf{X}_j + \zeta_j$$ ← **Level-2 structural model**

and setting:

$$\mathbf{\Lambda}_j = \mathbf{I} \qquad \varepsilon_{ij} \sim MVN(\mathbf{0}, \mathbf{0})$$

$$\alpha_j = \alpha \qquad \zeta_{ij} \sim MVN(\mathbf{0}, \mathbf{\Psi})$$

$$\mathbf{B}_j = \mathbf{B} \qquad \zeta_j \sim MVN(\mathbf{0}, \mathbf{0})$$

$$\mathbf{\Gamma}_j = \mathbf{\Gamma}$$

Multilevel SEM

Special case: **Single-level SEM**

Ordinary SEM may be specified by fixing the "grey" matrices to zero:

$$\mathbf{Y}_{ij} = \boldsymbol{\nu}_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

and setting:

$$\boldsymbol{\Lambda}_j = \boldsymbol{\Lambda} \qquad \boldsymbol{\Gamma}_j = \boldsymbol{\Gamma} \qquad \boldsymbol{\varepsilon}_{ij} \sim MVN(\mathbf{0}, \boldsymbol{\Theta})$$

$$\boldsymbol{\alpha}_j = \boldsymbol{\alpha} \qquad \boldsymbol{\nu}_j = \boldsymbol{\nu} \qquad \boldsymbol{\zeta}_{ij} \sim MVN(\mathbf{0}, \boldsymbol{\Psi})$$

$$\mathbf{B}_j = \mathbf{B} \qquad \mathbf{K}_j = \mathbf{K} \qquad \boldsymbol{\zeta}_j \sim MVN(\mathbf{0}, \mathbf{0})$$

Multilevel SEM

Special case: **Factor analysis**

Ordinary SEM may be specified by fixing the "grey" matrices to zero:

$$\mathbf{Y}_{ij} = \nu_j + \Lambda_j \eta_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \varepsilon_{ij}$$ ← **Level-1 measurement model**

$$\eta_{ij} = \alpha_j + \mathbf{B}_j \eta_{ij} + \Gamma_j \mathbf{X}_{ij} + \zeta_{ij}$$ ← **Level-1 structural model**

$$\eta_j = \mu + \beta \eta_j + \gamma \mathbf{X}_j + \zeta_j$$ ← **Level-2 structural model**

and setting:

$$\Lambda_j = \Lambda \qquad \varepsilon_{ij} \sim MVN(\mathbf{0}, \Theta)$$

$$\alpha_j = \alpha \qquad \zeta_{ij} \sim MVN(\mathbf{0}, \Psi)$$

$$\zeta_j \sim MVN(\mathbf{0}, \mathbf{0})$$

Multilevel SEM

Special case: **Traditional multilevel model**

Traditional MLM may be specified by fixing the "grey" matrices to zero:

$$\mathbf{Y}_{ij} = \nu_j + \boldsymbol{\Lambda}_j \boldsymbol{\eta}_{ij} + \mathbf{K}_j \mathbf{X}_{ij} + \boldsymbol{\varepsilon}_{ij}$$ ← **Level-1 measurement model**

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\alpha}_j + \mathbf{B}_j \boldsymbol{\eta}_{ij} + \boldsymbol{\Gamma}_j \mathbf{X}_{ij} + \boldsymbol{\zeta}_{ij}$$ ← **Level-1 structural model**

$$\boldsymbol{\eta}_j = \boldsymbol{\mu} + \boldsymbol{\beta} \boldsymbol{\eta}_j + \boldsymbol{\gamma} \mathbf{X}_j + \boldsymbol{\zeta}_j$$ ← **Level-2 structural model**

and setting: $\boldsymbol{\Lambda}_j = \mathbf{I}$

$$\boldsymbol{\varepsilon}_{ij} \sim MVN\left(\mathbf{0}, \mathbf{0}\right)$$

$$\boldsymbol{\zeta}_{ij} \sim MVN\left(\mathbf{0}, \boldsymbol{\Psi}\right)$$

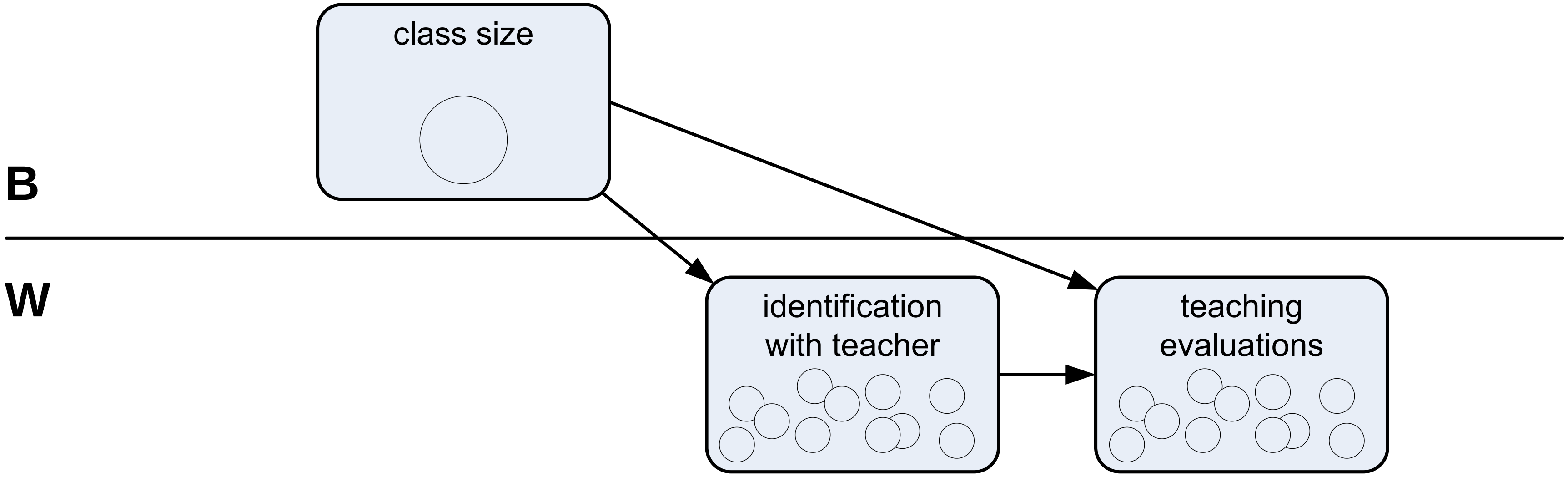
$$\boldsymbol{\zeta}_j \sim MVN\left(\mathbf{0}, \boldsymbol{\psi}\right)$$

Multilevel SEM

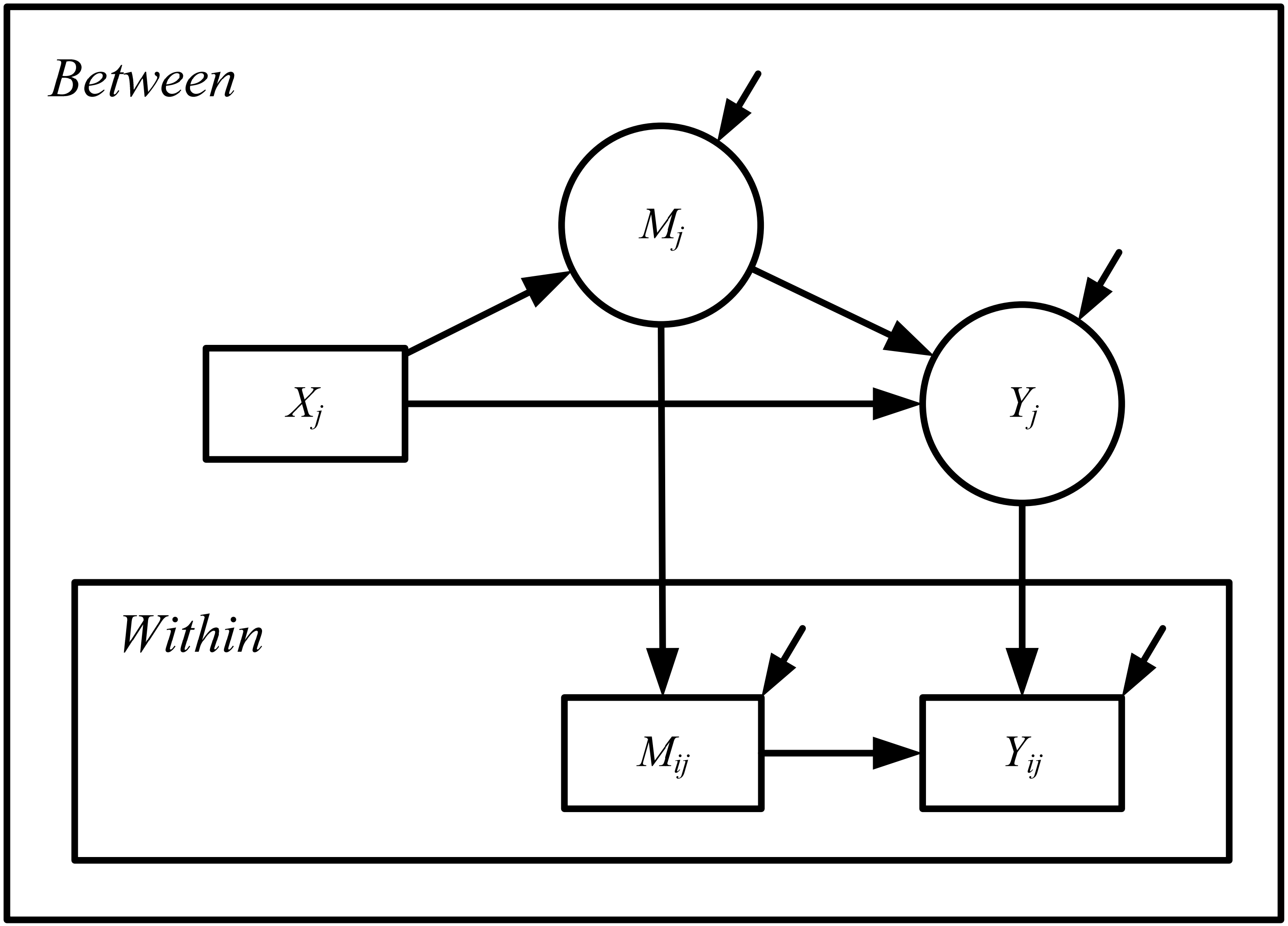
Some models MSEM can allow researchers to fit...

Multilevel SEM

Conceptual model...

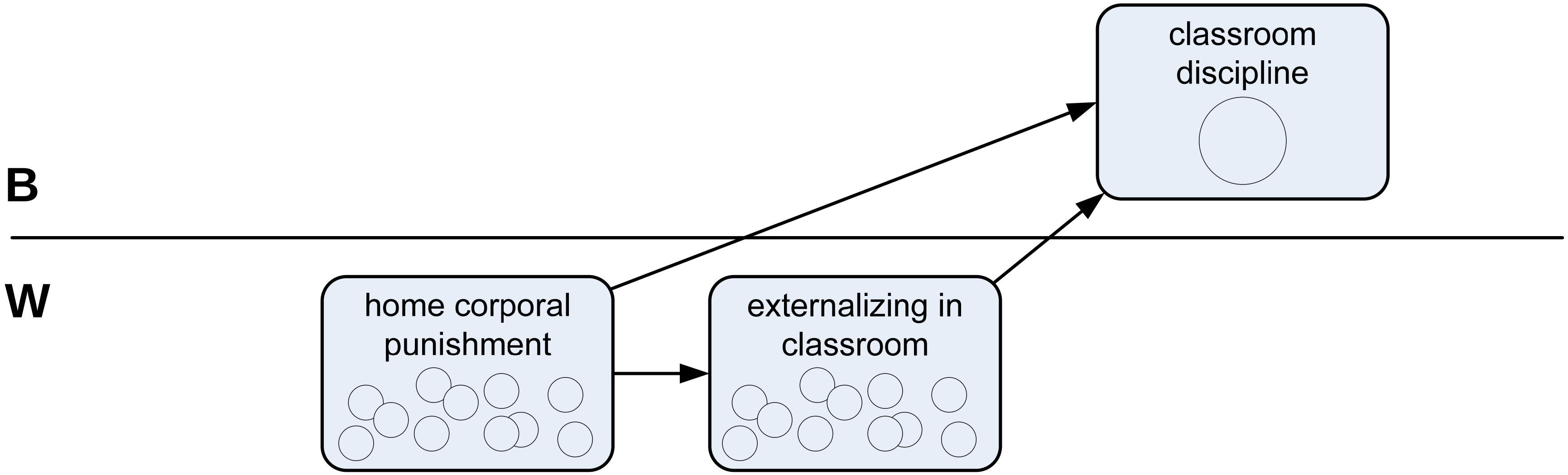


Actual model...



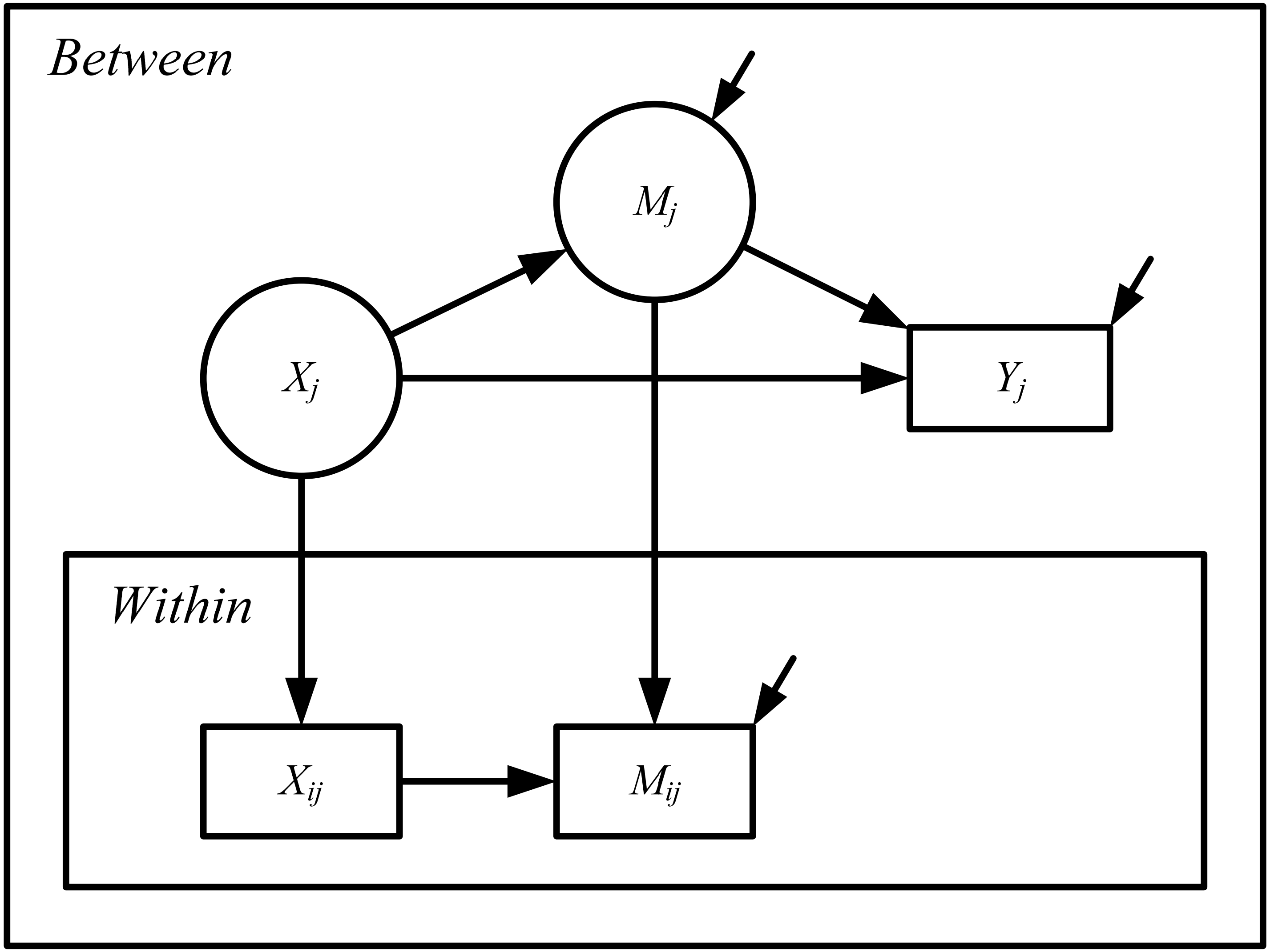
Multilevel SEM

Conceptual model...



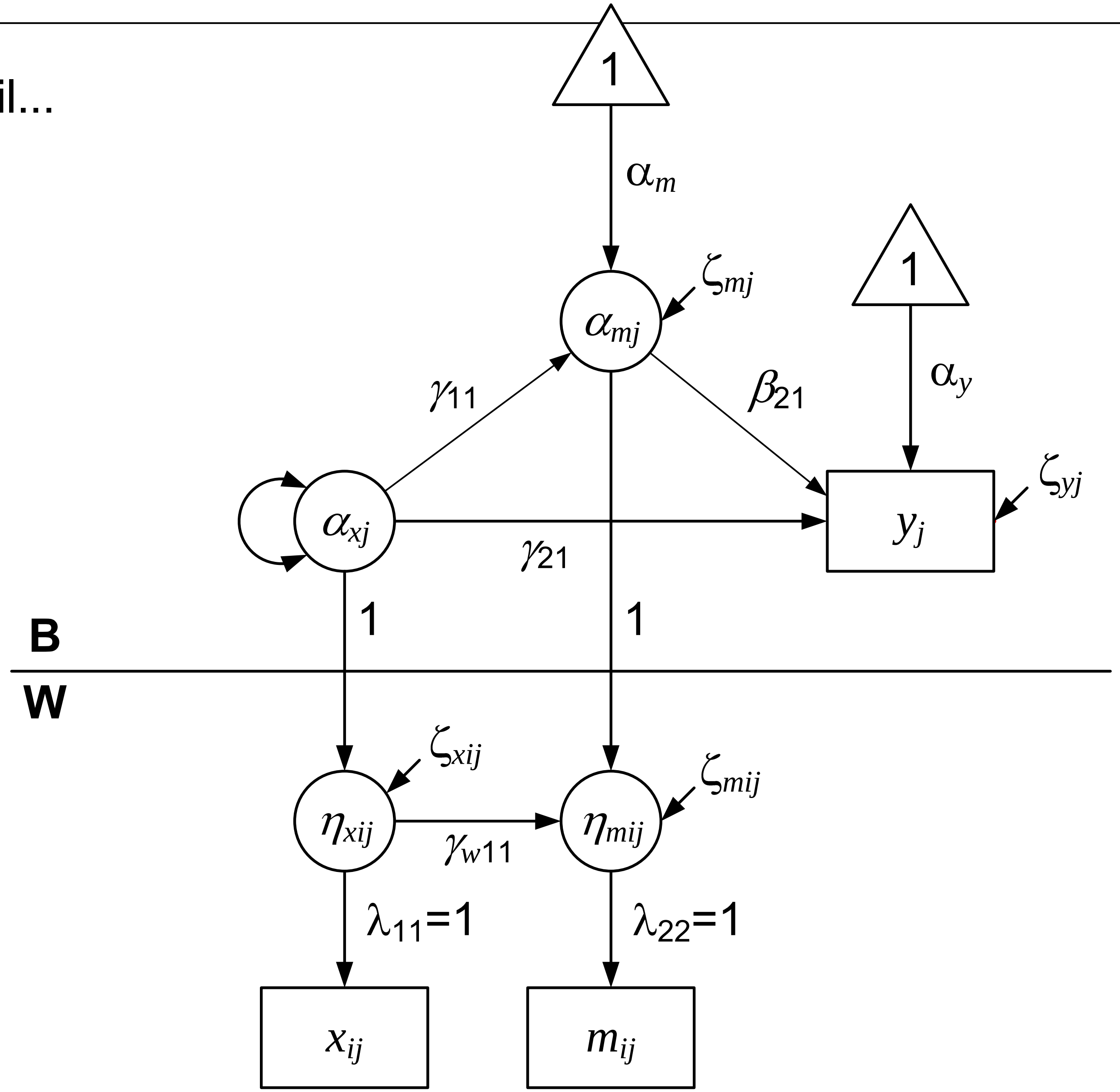
Multilevel SEM

Actual model...



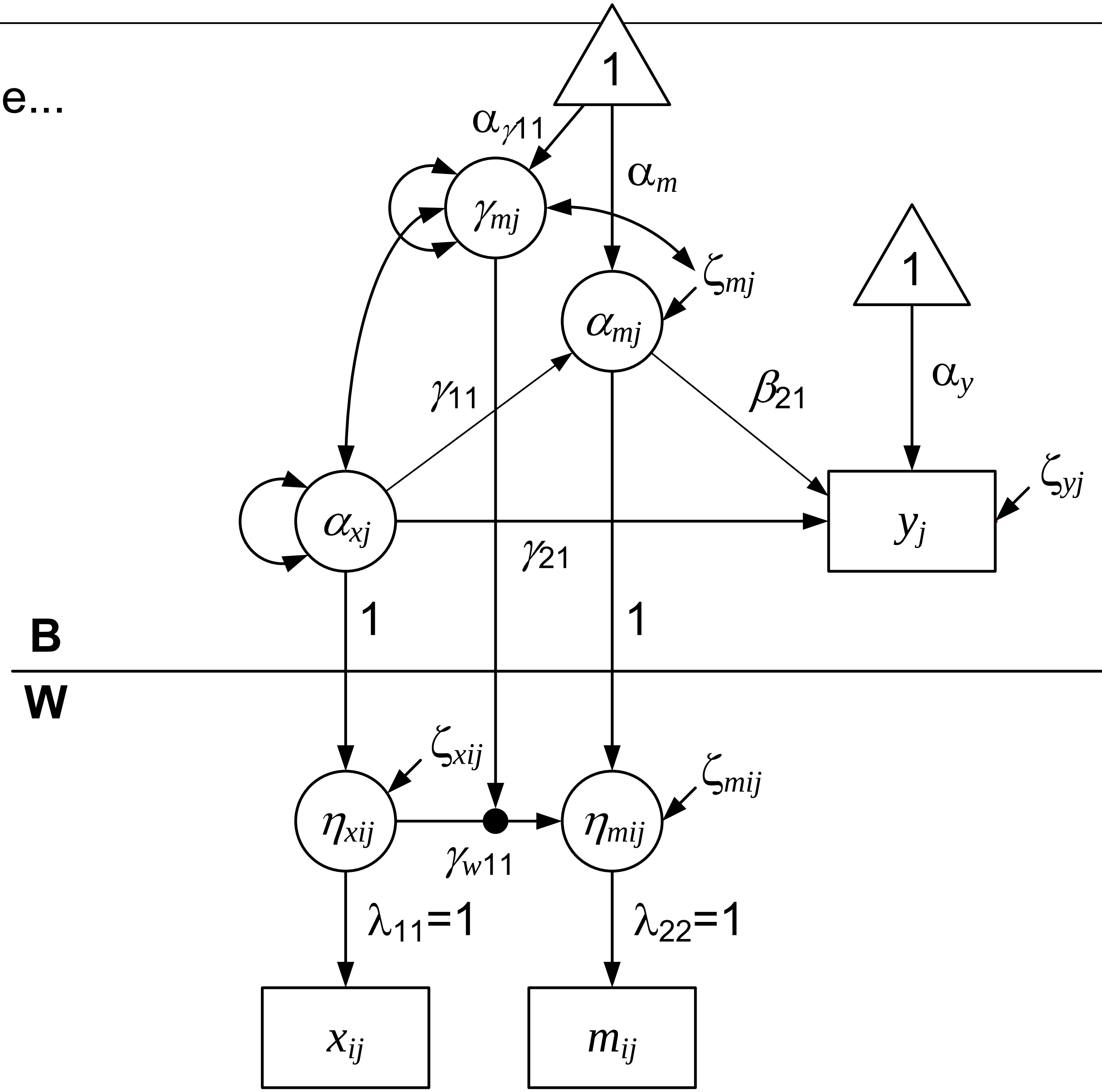
Multilevel SEM

In a little more detail...

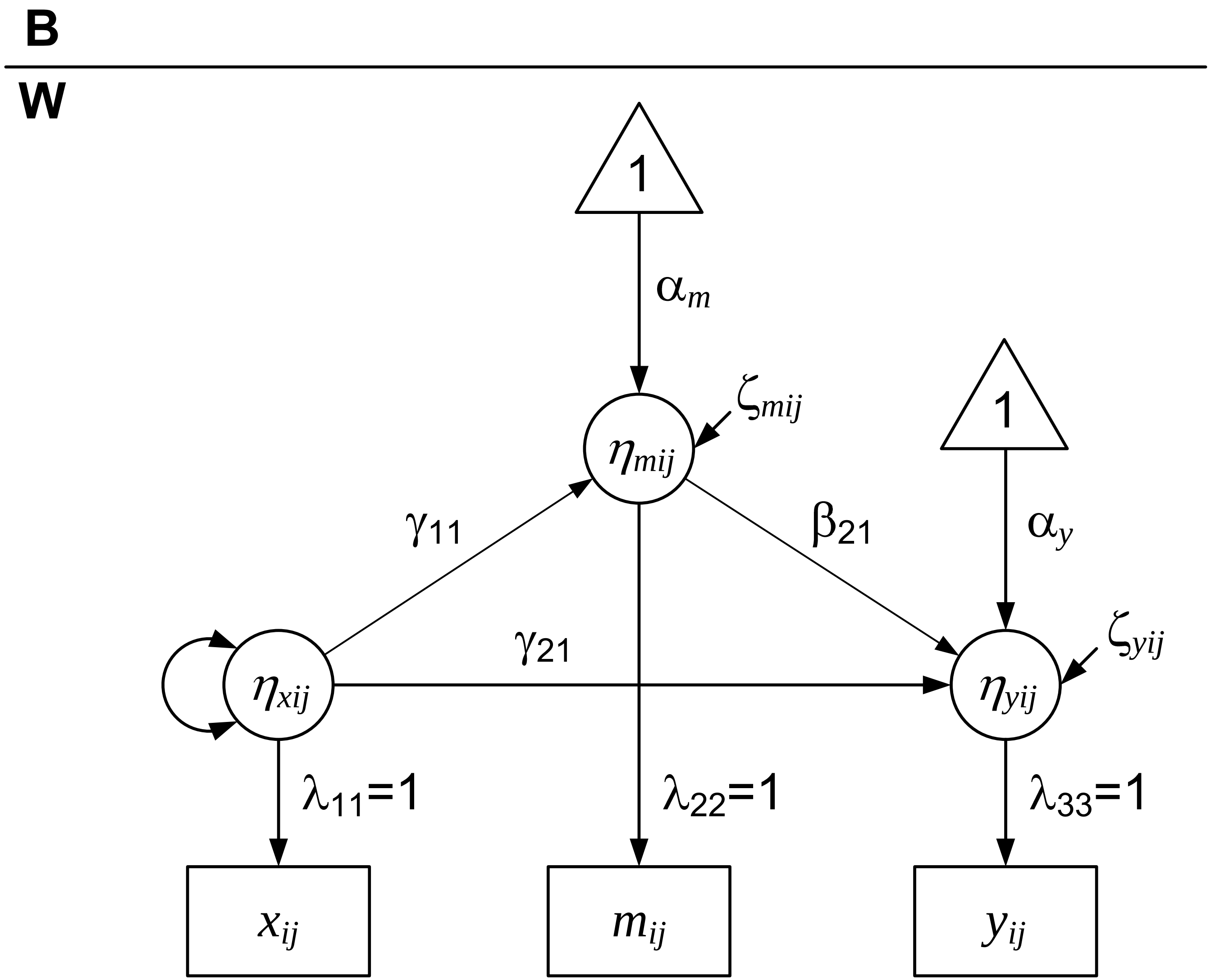


Multilevel SEM

With a random slope...

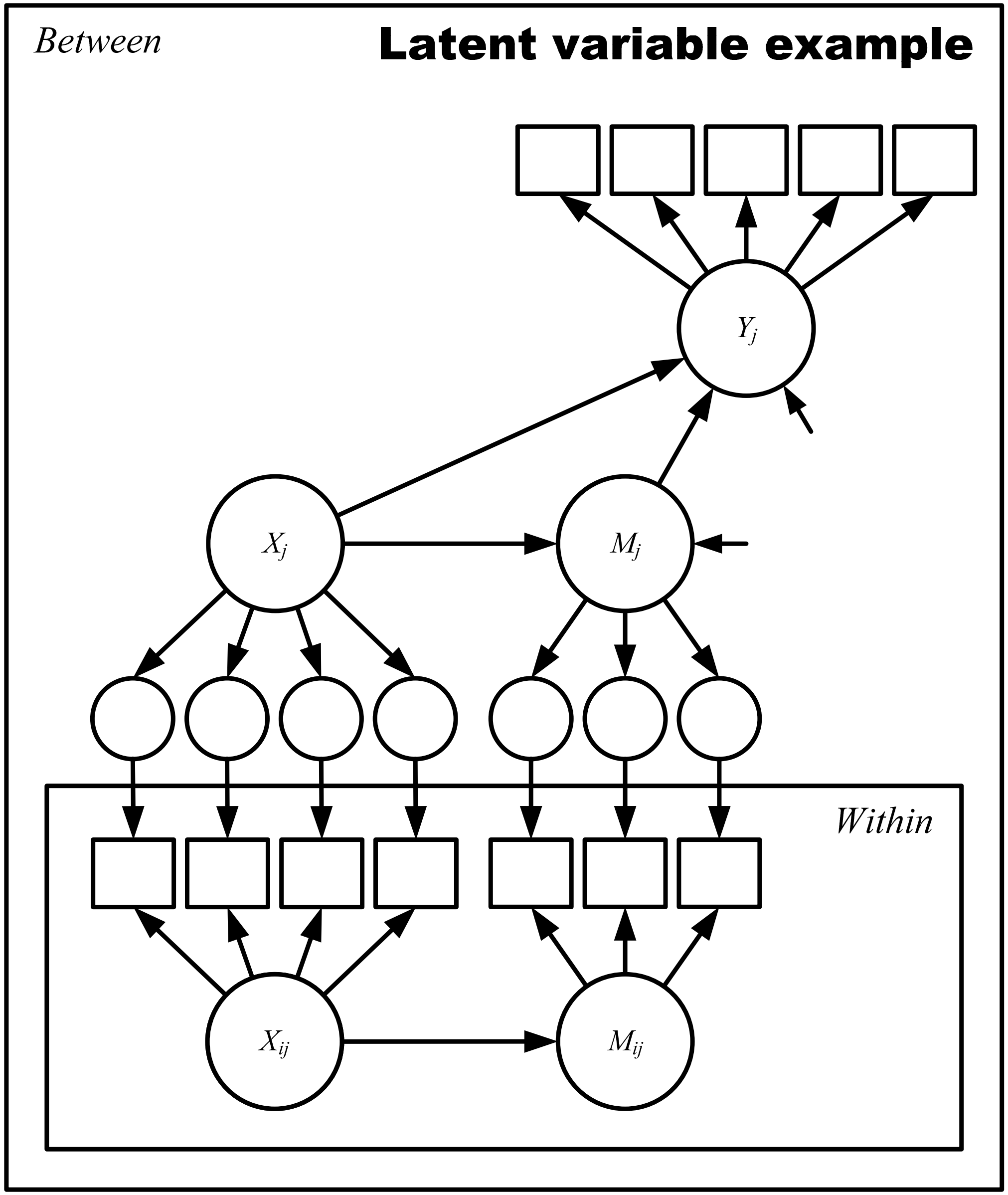


A single-level path analysis:



Multilevel SEM

Latent variables!



Empirical example (from the MSALT data)

Global self-esteem was computed as a composite of 5 items.

Performance in math was assessed by the single item “Compared to other students in this class, how well is this student performing in math?” (1 = “near the bottom of this class;” 5 = “one of the best in the class”).

Talent and **effort** were assessed with the items “How much natural mathematical talent does this student have?” (1 = “very little;” 7 = “a lot”) and “How hard does this student try in math?” (1 = “does not try at all;” 7 = “tries very hard”), respectively.

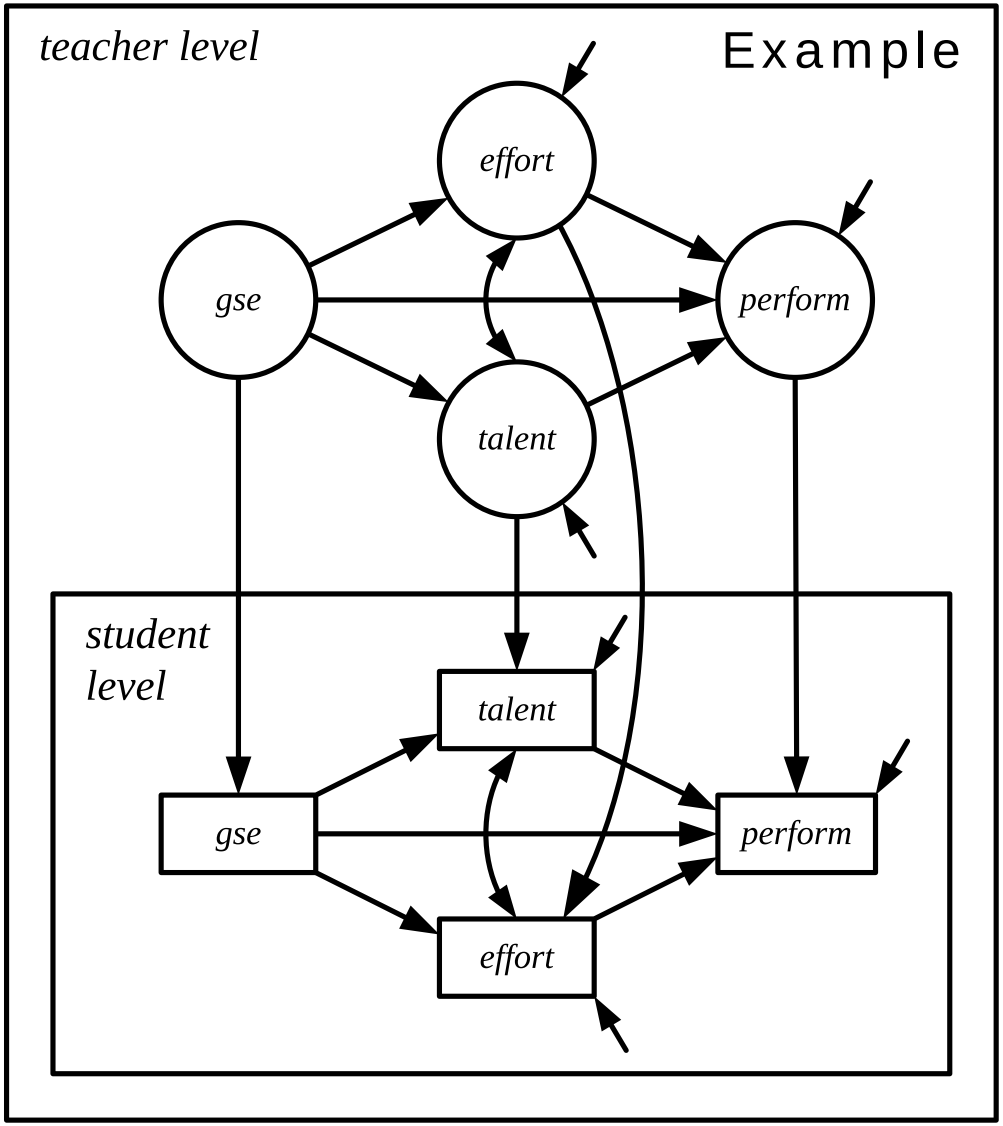
ICCs were: gse (.03), talent (.11), effort (.13), and perform (.04).

Empirical example (from MSALT data)

All four variables were assessed at the student level.

Intercepts are random, slopes are fixed.

The MSEM approach permits investigation of multiple mediators simultaneously, at both levels.



Relevant issues

Sample size...

- Little is known about sample size in MSEM. Related work suggests that having a large number of clusters (J) is more important than have large clusters (n_j).
- Larger samples are required when ICC is low.

Significance testing...

- Confidence intervals are recommended over significance testing.
- CIs can be obtained using a **parametric bootstrap** procedure; often more appropriate than z-tests based on standard errors.

Software...

- Mplus
- WinBUGS
- GLLAMM (Stata)
- Mx