Estimating Multilevel Linear Models as Structural Equation Models

Daniel J. Bauer

North Carolina State University

Patrick J. Curran

University of North Carolina Chapel Hill

Presentation at the

2002 Meeting of the Psychometric Society

June 21-June 23, 2002

Outline

- **Describe Motivation**
- **Introduce Multilevel Linear Model (MLM)**
- **Show that MLM can be estimated as SEM**
- **Show that we can extend MLM within SEM**

Motivation

Stengths/Limitations of MLMs

- **Optimal For**
  - **Obtaining correct SEs for nested data**
  - **Estimating & predicting random effects**
- **Difficult For**
  - **Estimating measurement models**
  - **Obtaining explicit tests of mediation**

Strengths/Limitations of SEM

- **Opposite of above**

Goals are to combine the strengths of the two models and bridge modeling traditions

A 2-Level MLM w/ L1 Covariates

Level 1 Model

$$y_{ij} = \pi_{0j} + \sum_{p=1}^{P} \pi_{pj} x_{p_{ij}} + r_{ij}$$

where

$$r_{ij} \sim MVN(\mathbf{0}, \mathbf{\Sigma}_{r_j})$$

Level 2 Model

$$\pi_{0j} = \beta_{00} + u_{0j}$$

$$\pi_{1j} = \beta_{10} + u_{1j}$$

$$\vdots$$

$$\pi_{Pj} = \beta_{P0} + u_{Pj}$$

where

$$\pi_{0j}, \pi_{1j}, \cdots, \pi_{0P} \sim MVN(\boldsymbol{\beta}, \mathbf{T})$$

A 2-Level MLM w/ L1 Covariates

Reduced-Form Equation

$$y_{ij} = \underbrace{\beta_{00} + \sum_{p=1}^{P} \beta_{p0} x_{p_{ij}}}_{\text{Fixed Coefficients}} + \underbrace{u_{00} + \sum_{p=1}^{P} u_{pj} x_{p_{ij}}}_{\text{Random Coefficients}} + r_{ij}$$

This is a special case of the linear mixed-effects model of Laird & Ware (1982)

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{Z}_j \mathbf{u}_j + \mathbf{r}_j$$

where

$\mathbf{X}_j$ **is the design matrix for the fixed effects** $\boldsymbol{\beta}$

$\mathbf{Z}_j$ **is the design matrix for the random effects** $\mathbf{u}_j$

implying that

$$\mathbf{y}_j \sim MVN(\mathbf{X}_j \boldsymbol{\beta}, \mathbf{Z}_j \mathbf{T} \mathbf{Z}_j' + \boldsymbol{\Sigma}_{r_j})$$

From MLM to SEM

In our case, $\mathbf{X}_j = \mathbf{Z}_j$, so

$$\mathbf{y}_j \sim MVN(\mathbf{X}_j \boldsymbol{\beta}, \mathbf{X}_j \mathbf{T} \mathbf{X}_j' + \boldsymbol{\Sigma}_{r_j})$$

Further*, if the design is balanced* then $\mathbf{X}_j = \mathbf{X}$

and

$$\mathbf{y}_j \sim MVN(\mathbf{X} \boldsymbol{\beta}, \mathbf{X} \mathbf{T} \mathbf{X}' + \boldsymbol{\Sigma}_r)$$

This is the same structure as a CFA model where

$$\mathrm{X} = \Lambda$$
$$\beta = \kappa$$
$$\mathrm{T} = \Phi$$
$$\Sigma_r = \Theta_\delta$$

A Classic Case: The Growth Model

Multilevel Linear Growth Model

Level 1: $y_{ti} = \pi_{0i} + \pi_{1i} x_{ti} + r_{ti}$

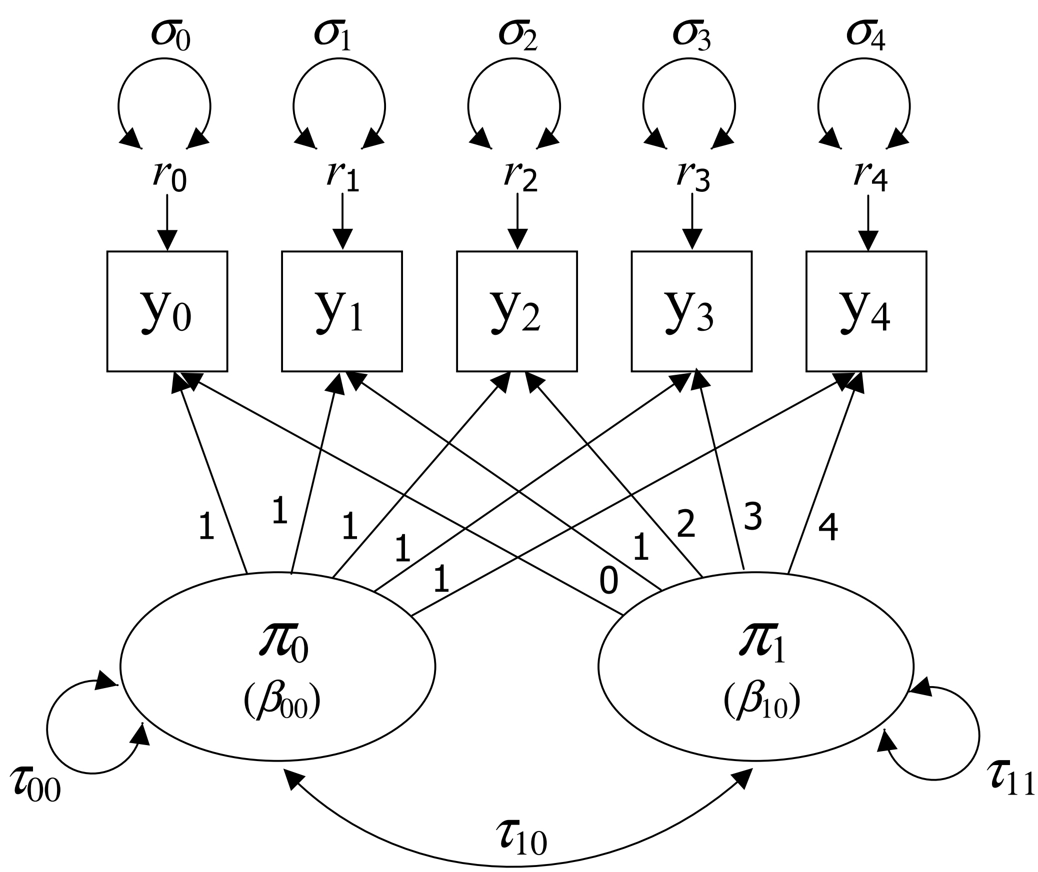
$$\Sigma_r = DIAG(\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4)$$

Level 2: $\pi_{0i} = \beta_{00} + u_{0i}$

$\pi_{1i} = \beta_{10} + u_{1i}$

$$T = \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix}$$

Linear Latent Curve Model in SEM



$X_i = X = \Lambda$ because assuming balanced design.
Random coefficients are represented as latent variables.

Balanced Data

Example

3 male & 3 female students per school to evaluate effect of sex on language ability

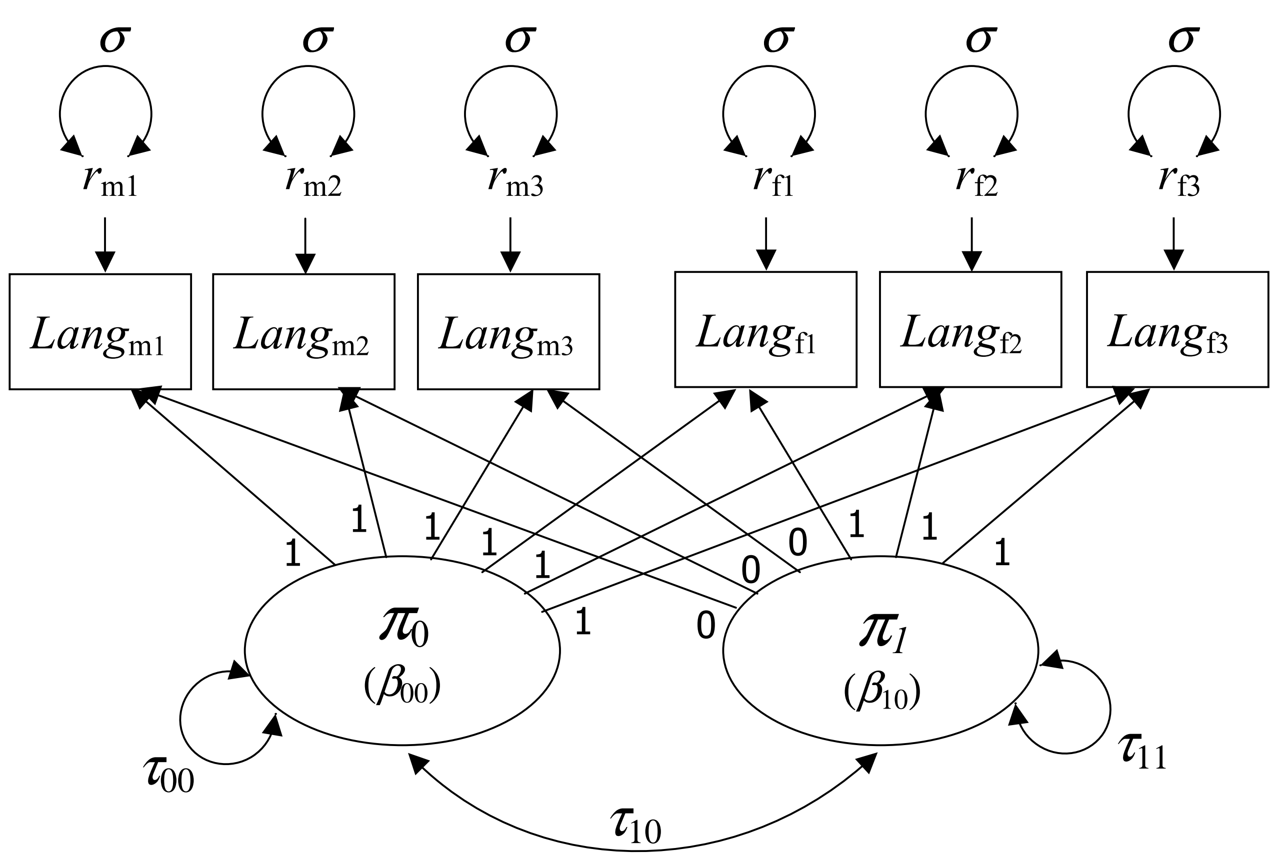
Multilevel Linear Model

Level 1: $lang_{ij} = \pi_{0j} + \pi_{1j} sex_{ij} + r_{ij}$ $\qquad \Sigma_r = \sigma I$

Level 2: $\pi_{0j} = \beta_{00} + u_{0j}$

$\pi_{1i} = \beta_{10} + u_{1j}$ $\qquad T = \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix}$

Equivalent SEM



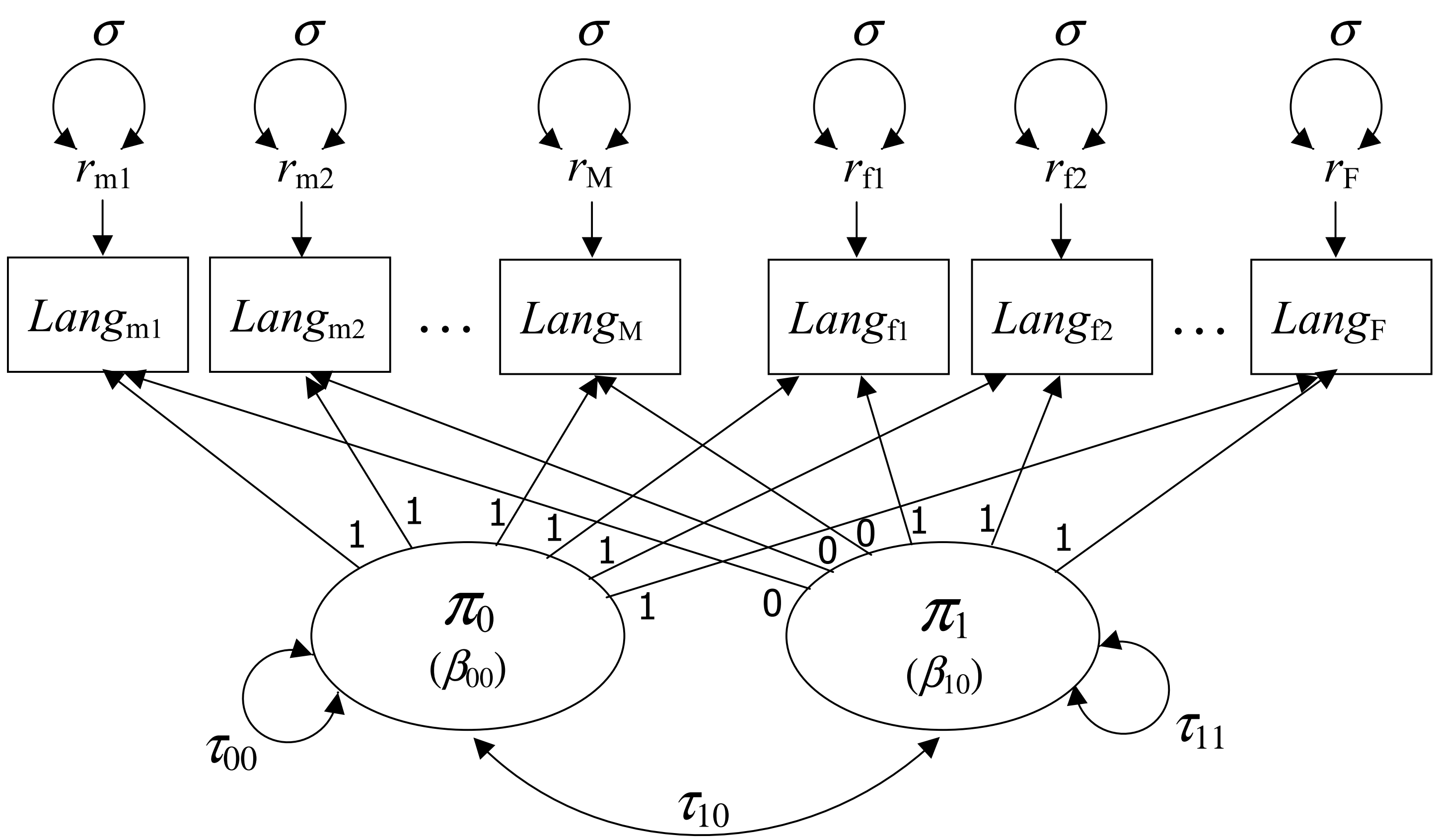
Order of the 3 males, 3 females w/in units j is arbitrary

Strategies for Imbalanced Data

Treat as missing

- **Construct complete-data** $\hat{\Sigma}(\theta), \hat{\mu}(\theta)$
- **Compare each** $\mathbf{y}_j$ **to submatrices** $\hat{\Sigma}(\theta)_j, \hat{\mu}(\theta)_j$

Example: **M = max # male students & F = max # female students in any given school.**

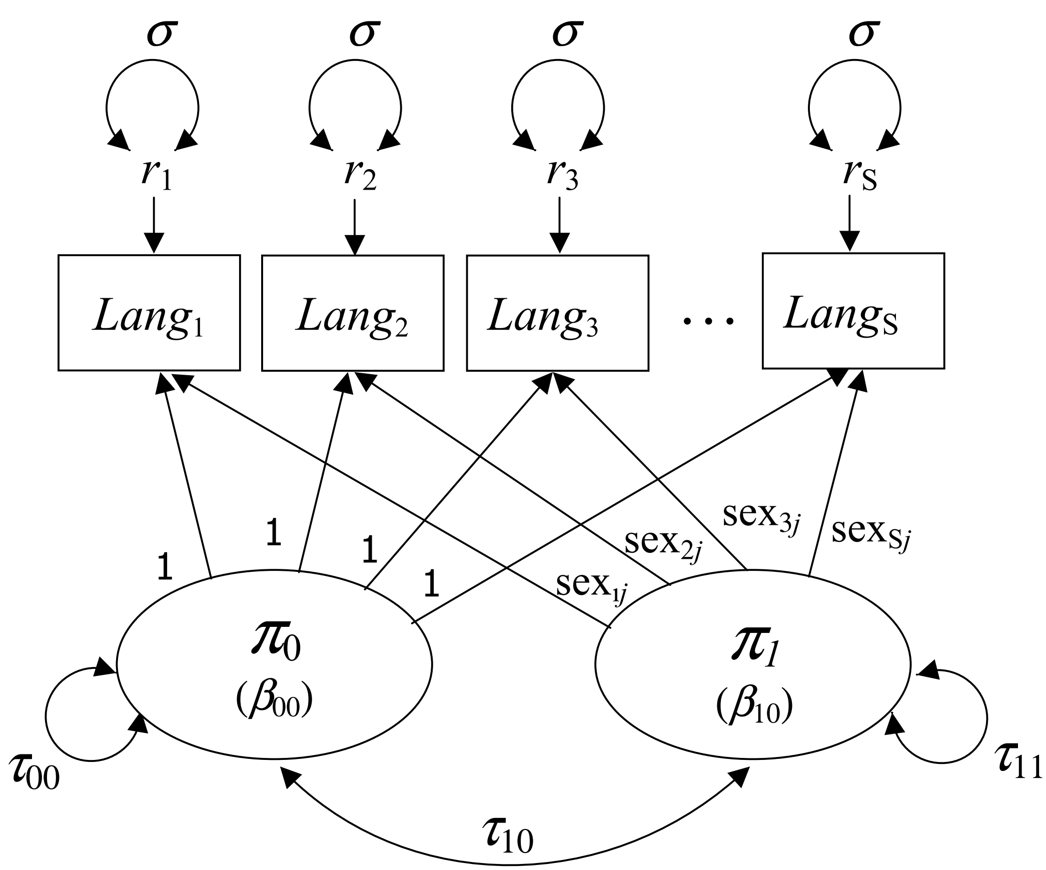


Strategies for Imbalanced Data

Compute $\hat{\Sigma}(\theta)_j, \hat{\mu}(\theta)_j$ ***directly from*** Λ_j

- **Truer to multilevel approach:** $\Lambda_j = \mathbf{X}_j$
- $\mathbf{X}_j$ **referred to as *definition variables* for** Λ_j
- **Due to Neale**

Example: **S = max # students in any given school**



What to do if you're imbalanced?

Both approaches provide computationally equivalent results but

- **Strategy 1 is better for few discrete covariates & complex residual structures.**
- **Strategy 2 is better for continuous covariates (highly imbalanced data) & homogeneity of error variance.**

Adding Higher-Level Predictors

Adding Level 2 Covariates

Problem is $\mathbf{X}_j \neq \mathbf{Z}_j$ but one Λ_j

Rovine & Molenaar Solution:

- **Fixed effects factors have means, no variance**
- **Random effects factors have variance, no means**
- **Define $\Lambda_j = BLOCK(\mathbf{X}_j, \mathbf{Z}_j)$**
- **True to mixed-effects model, non-intuitive.**

Alternative Solution:

- **Extends approach used w/ latent curve models**
- **L2 predictors are 'fixed X' covariates**
  - **Effects contained in Γ**
- **Computationally equivalent to R & M Solution**

Both solutions can be extended to 3+ Level models

Expanding the Model: A New Approach to Multilevel CFA

Adding a measurement model for item level outcomes

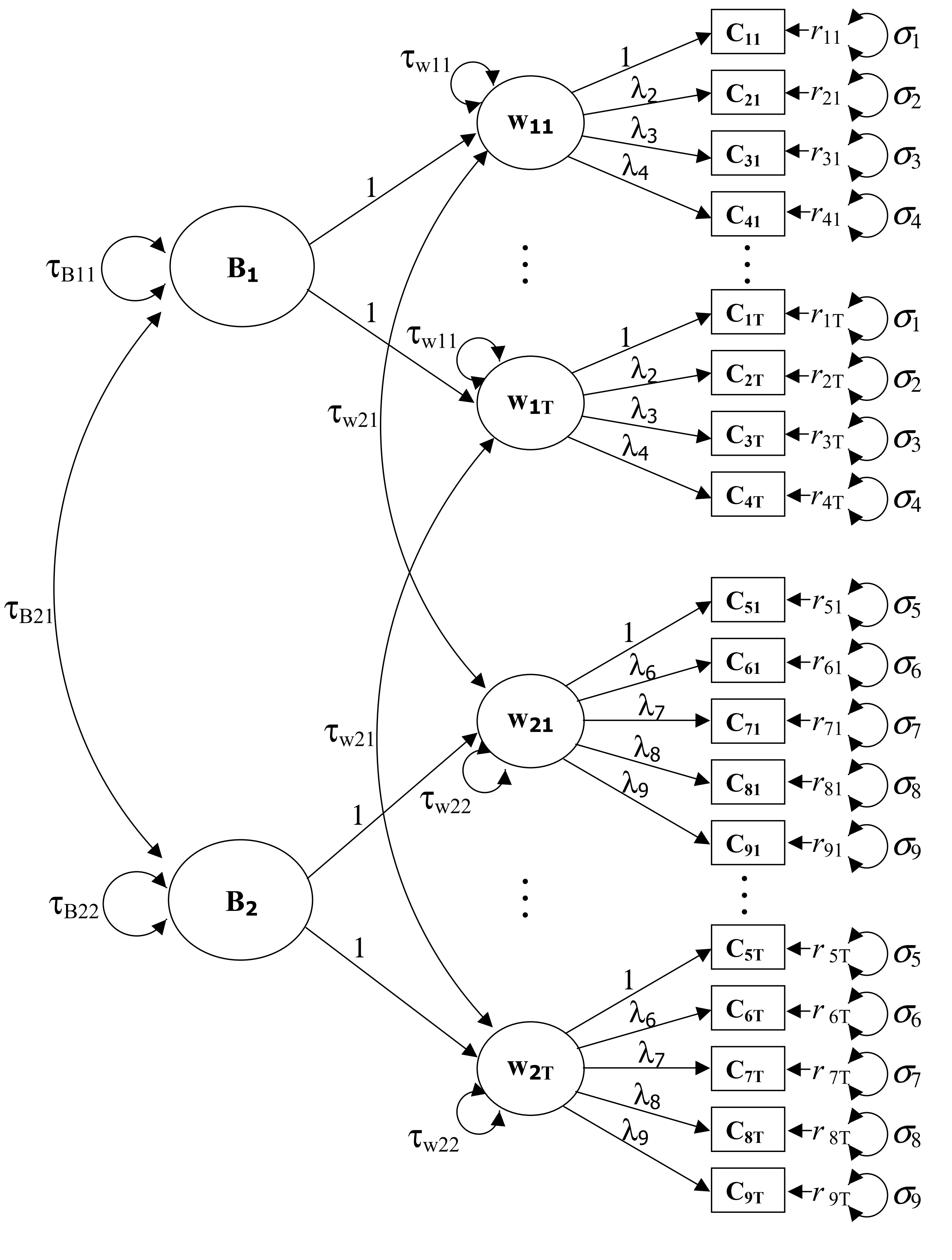
Example:

Data from High-School & Beyond: Teacher Survey

- **456 schools; 10,365 teachers**
  - **Imbalanced: # teachers/school ranges from 1 to 30**
  - **Let max # teachers = T = 30**
- **9 item measure of teacher perceptions of control**
  - **4 items on control of school policy**
  - **5 items on control of classroom teaching/planning**
  - **6 point Likert scales; Centered at mean**

Estimating 2-Factor Model

High-School & Beyond 2-Factor Model



Empirical Validation

Comparing SEM and MLM estimates

Parameter	Multilevel CFA		PROC MIXED	
λ_1	1.0		1.0	Loadings fixed to values from multilevel CFA (Cannot be estimated directly)
λ_2	.9932	(.02342)	.9932	
λ_3	1.1492	(.02508)	1.1492	
λ_4	1.2867	(.02615)	1.2867	
λ_5	1.0		1.0	
λ_6	.9803	(.01746)	.9803	
λ_7	.5444	(.01088)	.5444	
λ_8	.6080	(.01532)	.6080	
λ_9	.4269	(.01047)	.4269	
τ_{w11}	.5114	(.01848)	.5114	(.01132)
τ_{w22}	.6384	(.01970)	.6384	(.01292)
τ_{w21}	.2637	(.01011)	.2637	(.00885)
τ_{B11}	**.2029**	(.01726)	**.2027**	(.01647)
τ_{B22}	.1611	(.01426)	.1611	(.01379)
τ_{B21}	.1153	(.01240)	.1153	(.01232)
σ_1	1.2579	(.02160)	1.2579	(.02084)
σ_2	1.4890	(.02471)	1.4890	(.02413)
σ_3	1.3828	(.02481)	1.3828	(.02367)
σ_4	1.0047	(.02214)	1.0047	(.02015)
σ_5	.9614	(.01778)	.9614	(.01686)
σ_6	.5799	(.01295)	.5799	(.01176)
σ_7	.3675	(.00636)	.3675	(.00615)
σ_8	1.0119	(.01566)	1.0119	(.01548)
σ_9	.4610	(.00718)	.4610	(.00711)