Multilevel Analysis of Grouped and Longitudinal Data

A Comparison of Multilevel Regression and Structural Equation Models

Joop Hox

Utrecht University

Utrecht Seminar Longitudinal research in work & health psychology, April 27-28, 2000

- grouped and longitudinal data as hierarchical data structures

- multilevel regression model

  - for grouped data
  - for longitudinal data

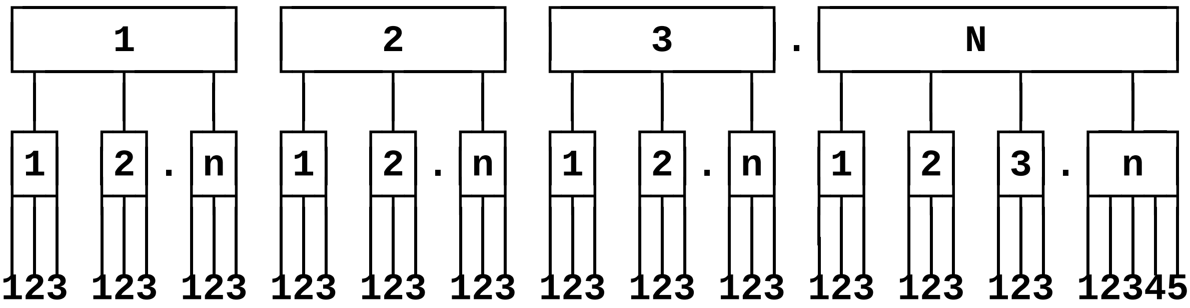
- multilevel structural equations model

  - for grouped and longitudinal data

- multilevel regression & latent growth curve models for longitudinal data

  - latent growth curve model
  - comparison with multilevel models

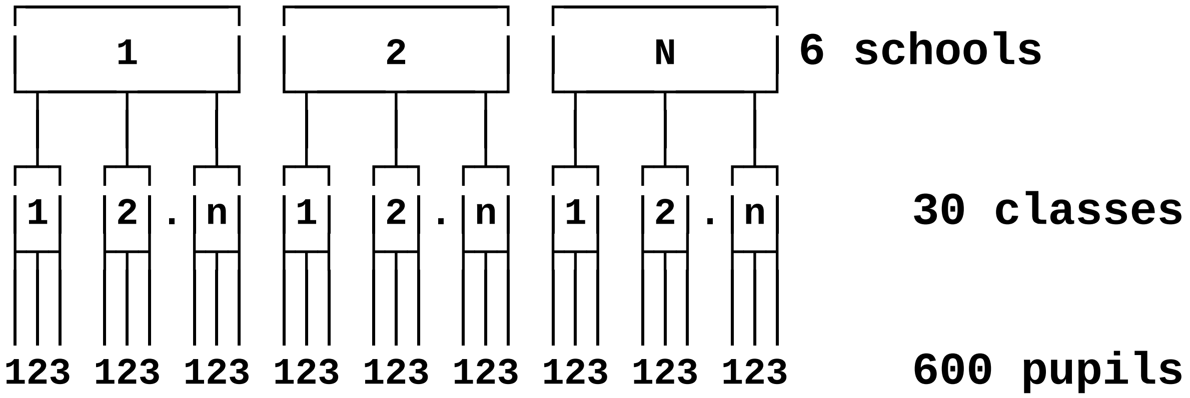
Hierarchical Data Structure



A three-level data structure. The groups at the different levels may all have different sizes. Variables may be defined at all available levels.

Examples:	education	time series	survey
Level 3	schools	organizations	regions
Level 2	classes	respondents	clusters
Level 1	pupils	occasions	respondents

Problems with Hierarchical Data Structures



Suppose we have a dependent variable at the lowest level (level 1), for example the pupils' school career. We want to predict this variable using explanatory variables from all three levels, for example pupils' SES, class size, school organization.

- How?
- What is the proper sample size?
- What if the effect of class size is not the same in different schools?

Traditional Approaches

- Disaggregate all school and class level variables to the individual pupil level
- Perform analyses (ANOVA, multiple regression, etc.) with all explanatory variables

Some improvements:

- Express explanatory variables as deviances from their class means. thus: pupil SES is split up into two variables: the class mean and the individual pupil's deviation from that mean. this makes it possible to separate class effects and pupil effects more clearly
- Add cross-level interactions. this makes it possible to model different effects for different classes/schools

Problems with Ordinary (OLS) Multiple Regression on Hierarchical Data

- **Multiple Regression assumes**
  - **independent error terms**
  - **equal variances of errors for all observations (homoscedastic errors)**
  - **normal distribution for errors**
  - **linearity**
  - **no measurement error in predictors**

- **However, with hierarchical data**
  - **errors are not independent**
  - **errors have different variances (heteroscedastic errors)**

Observations from the same groups are more similar than observations from different groups, due to: selection processes, shared group history, contextual group effects. The amount of this within-group correlation is usually indicated by the *intraclass correlation*.

- **ordinary statistical tests are *not***

at all **robust against these violations**

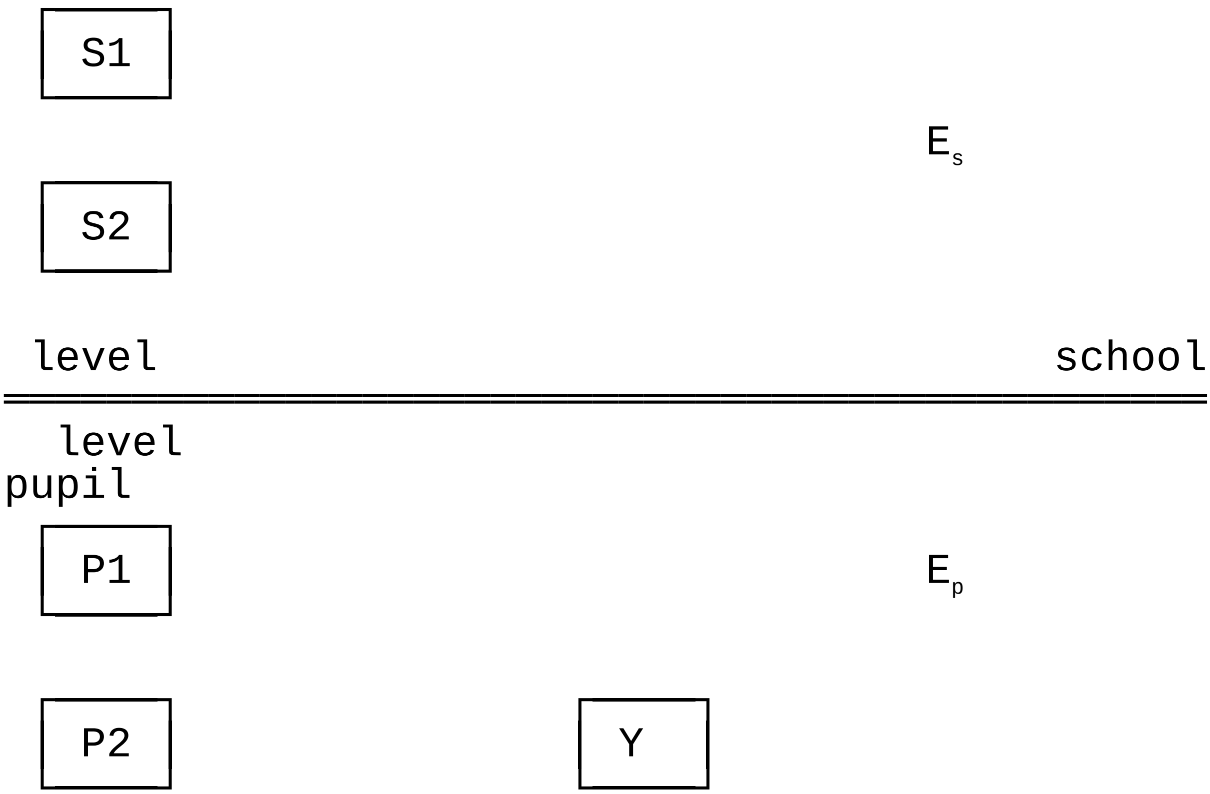
OLS Techniques on Multilevel Data

Actual type I error in t-test; individuals in cells are interrelated, nominal alpha level is 0.05

n per cell	intraclass correlation .00	.01	.05	.20	.40	.80
10	.05	.06	.11	.28	.46	.75
25	.05	.08	.19	.46	.63	.84
50	.05	.11	.30	.59	.74	.89
100	.05	.17	.43	.70	.81	.92

(Barcikowski, *JES*, 1981)

Graphical Representation of a Simple Multilevel Regression Model



- **Features**
  - dependent variable Y at pupil level
  - explanatory variables P at pupil level
  - error term E_p at pupil level
  - explanatory variables S at school level
  - error term Es at school level

Multilevel Regression Model: Lowest Level

- Ordinary regression, one explanatory X:

  $$Y_i = \beta_0 + \beta_1 X_i + e_i$$

  - β_0 the intercept
  - β_1 the regression slope
  - e_i the residual error term

- Multilevel regression, lowest level:

  $$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$$

  - β_{0j} the intercept
  - β_{1j} the regression slope
  - e_{ij} the residual error term

- Subscript *j* for groups. We assume for each group a different intercept β_{0j} and a different slope β_{1j}

- **Errors $e_{ij} \approx N(0,\sigma^2)$ (or $\approx N(0,\sigma_j^2)$)**

Multilevel Regression Model: Second Level

- **At the lowest level** $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$

- **Regression coefficients β_j vary across groups. This variation is predicted by group level explanatory variables. With one group level Z the model for the ß is:**

  $$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$

  $$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}$$

  - γ_{00} **and** γ_{01} **are the intercept and slope to predict** $ß_{0j}$ **from** Z_j
  - u_{0j} **is the residual error term in the equation for** $ß_{0j}$
  - γ_{10} **and** γ_{11} **are the intercept and slope to predict** $ß_{1j}$ **from** Z_j
  - u_{1j} **is the residual error term in the equation for** $ß_{1j}$

Single Equation Version

- At the lowest (individual) level we have

  $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$

- and at the second (group) level

  $\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$
  $\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}$

- Substitution and rearranging terms gives

  $Y_{ij}= \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_jX_{ij} +$

  $u_{1j} X_{ij} + u_{0j} + e_{ij}$

- $\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_jX_{ij}$ are the fixed coefficients in the fixed (deterministic) part
- $u_{0j} + u_{1j} X_{ij} + e_{ij}$ are the random error terms in the random (stochastic) part
- Other salient features:
  - *cross-level* interaction Z_jX_{ij}
  - error term $u_{1j} X_{ij}$ implies

heteroscedasticity

Assumptions

- The model is

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$$

- Assumptions
  - individual level errors e_{ij} are independent, $\approx N(0, \sigma^2)$
  - group level errors $u_{.j}$ are independent, $\approx N(0, \Sigma)$
  - group level errors $u_{.j}$ are independent from the e_{ij}
  - plus usual assumptions of multiple regression analysis: linear relationships, explanatory variables measured without error

Interpretation

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} j + u_{1j} X_{ij} + u_{0j} + e_{ij}$$

- Fixed part resembles simple regression with intercept γ_{00}, and regression slopes γ_{10}, γ_{01}, and γ_{11}. The regression coefficients in the fixed part are interpreted as raw regression coefficients, same as in ordinary multiple regression

- The error term is more complicated:
  $u_{1j}\ X_{ij} + u_{0j} + e_{ij}$

- There are several error variances
  σ^2 variance of lower level errors e_{ij}
  σ^2_{u0} variance of the higher level errors u_{0j} (=variance of intercepts $ß_{0j}$)
  σ^2_{u1} variance of the higher level errors u_{1j} (=variance of the slopes $ß_{1j}$)

Estimation

- Specialized software estimates parameter values, standard errors, deviance
  (HLM, MLn, Varcl, Mixreg, MLA, also in BMDP and SAS)
  - different likelihood functions (e.g., FML, RML)
  - different estimation methods (e.g., EM, Fisher scoring, IGLS)
  - different features (e.g., graphics, macro's)
  - different model extensions (e.g., meta analysis, logistic models)
  - residuals at several levels
  - nested models can be tested using the deviance

Some Programs

HLM . user friendly interface
- FML+RML estimation, constraints
- meta-analysis extension
- 3 levels

MLn . powerful & *very* flexible
- data manipulation, graphics, macros
- FML+RML estimation, constraints
- unlimited number of levels

Varcl
- FML estimation, constraints
- extensions for non-normal data
- 3 levels

Mixreg
- FML estimation
- allows autocorrelation structure
- binary/ordinal data in MIXOR
- 2 levels
- freeware

MLA . **FML estimation, constraints**

- **bootstrapping options**
- **2 levels**
- **batch operation**

Some Special Models
The Intercept-Only Model
(empty model, null model)

- **Intercept-Only model: contains only intercept and corresponding error terms**

  $$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

  - **null model**

  - **compute intra class correlation ρ:**

    $$\rho = \sigma^2_{u0} / (\sigma^2_{u0} + \sigma^2)$$

  - **ρ is a *population estimate* of the variance explained by the grouping structure**

The Fixed Model

- The Fixed Model: intercepts vary across groups, slopes are fixed

  $$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{0j} + e_{ij}$$

  - fixed slope: explanatory variable has same effect in all groups
    (= parallel slopes)
  - similar to ANCOVA with random grouping factor (identical if equal group sizes and RML estimation)

- Since this model estimates distinct intercept variances for each level, it is also called a variance component model

The Random Coefficient Model

- Random Coefficient Model: both intercept and slopes vary across groups

  $$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$$

  - for each regression slope there is a 2nd level error term $u_{.j}$

- The full multilevel model adds explanatory variables at the higher level(s), and cross-level interactions

  $$Y_{ij}=\gamma_{00}+\gamma_{10}X_{ij}+\gamma_{01}Z_j+\gamma_{11}Z_jX_{ij}+u_{1j}X_{ij}+u_{0j}+e_{ij}$$

Model Selection & Exploration Strategy

1: *Intercept-only model:* **intra class correlation; deviance is used as baseline**

2: *Fixed model:* **test slopes for significance. Test the improvement of the model using the difference of the deviance of this model and the previous (intercept-only) model)**

3: *Random coefficient model*: **test if any slope has a significant variance component. (use the chi-square test on the deviances to test for improvement)**

4: **Test higher level explanatory variables, inspect if they explain between group variation (intercept variance)**

5: **Test cross-level interactions between explanatory group level variables and individual level explanatory variables that had**

significant slope variation in step 3, inspect if they explain slope variances

Example

Data: 28 classes with 428 pupils,
dependent variable loneliness at end of school year
explanatory variables loneliness at start,
pupil sex, class size, teacher experience (centered)

Expl Model for lonely posttest

var	int. only	fixedRC		+expl. var
const	.03 (.11)	.04 (.09)	.06 (.09)	.02 (.08)
lon pre		.66 (.04)	.66 (.04)	.66 (.04)
sex (girl)		-.16 (.06)	-.18 (.08)	-.17 (.08)
class size				-.05 (.02)
teach.exp.				-.03 (.01)
interac t.e.*sex				-.03 (.01)
σ^2	.72	.41	.39	.39
σ^2_{const}	.31	.19	.19	.14
σ^2_{sex}	$(\rho=.30)$		.08	.05
deviance	1128	894	889	877

Multilevel Regression Model for Longitudinal Data

- lowest level: occasions; second level: individuals; (higher levels: groups)

- important conceptual distinction:
  - fixed occasion model, all individuals same number and times of measurement
  - growth curve model, number and times of measurement arbitrary

Model:

$$Y_{tj} = \beta_{0j}+\beta_{1j}T_{tj}+\beta_{2j}X_{tj}+e_{tj}$$

and

$$\beta_{0j} = \gamma_{00}+\gamma_{01}Z_j+u_{0j}$$
$$\beta_{1j} = \gamma_{10}+\gamma_{11}Z_j+u_{1j}$$
$$\beta_{2j} = \gamma_{20}+\gamma_{21}Z_j+u_{2j}$$

thus

$$Y_{tj} = \gamma_{00}+\gamma_{10}T_{tj}+\gamma_{20}X_{tj}+\gamma_{01}Z_j+\gamma_{11}Z_jT_{tj}+\gamma_{21}Z_jX_{tj}+$$
$$u_{1j}T_{tj}+u_{2j}X_{tj}+u_{0j}+e_{tj}$$

Advantages of Multilevel Regression Model for Longitudinal Data

Why? Because:

- lowest level models specific growth curves
- number of repeated measures and their spacing may vary across persons
- covariances between repeated measures can be modeled
- identical to Manova approach if groups are balanced (and estimation is RML)
- simple to add higher levels (Bryk & Raudenbush, 1992, p133)
- simple to add time varying and/or time invariant covariates

However:

- in standard multilevel model error covariances are identical between all time points
  - time dependent error structure possible in Mixreg and MLn

Example Data Set from Rogosa and Saner

- 200 cases
- 1 dependent variable Y
- 5 equidistant time points (*t* = 0,1,2,3,4)
- 1 time invariant (person level) covariate Z

- Artifical data generated to follow linear growth model, with measurement error in the dependent variable
- 7% Data points randomly deleted (MCAR)

- Model:

$$Y_{tj} = \gamma_{00}+\gamma_{10}T_{tj}+\gamma_{01}Z_j+\gamma_{11}Z_jT_{tj}+u_{1j}T_{tj}+u_{0j}+e_{tj}$$

(JEBS, 1995)

Multilevel Regression Results (Z centered)

missing values data set: 14 Z missing results in 14 cases deleted; analysis on 186 cases
(866 observations left of 1200 = 73% efficiency)

Model	(1)	(2)	(3)	(4)	(5)
interc	54 (.67)	44 (.6)	44 (.5)	44 (.6)	44 (.5)
time		5 (.1)	5 (.2)	5 (.2)	5 (.1)
Z				3 (.2)	2 (.2)
Z*time				.6 (.1)	
σ^2_ε	83 (5)	23 (1)	12 (.8)	13 (.8)	12 (.8)
σ^2_{int}	42 (6)	54 (6)	50 (6)	50 (6)	41 (5)
σ^2_{time}			4 (.6)	4 (.6)	3 (.4)
σ_{i*t}			-3 (1)	-11 (2)	-7 (1)
r_{i*t}			-.19	-.74	-.66
deviance	6509	5625	5452	5320	5249

compare

`explained variance'	σ^2_{ε}	**(1) (5)**	**73%**
	σ^2_{int}	**(4) (5)**	**19%**
	σ^2_{time}	**(4) (5)**	**39%**

Multilevel Regression, Comments

- Fixed regression coefficients estimated well, even in small samples

- With small number of groups (individuals in longitudinal data) low power for group and cross-level interaction effects

- With small number of groups low accuracy for estimation and tests of higher level variance components

- Higher level sample size: at least 20, preferably 50, if variance components are important preferably 100

Multilevel Structural Equation Models

- **Data from *N* individuals, divided into *G* groups, in p-variate vector Y_{ig}**
  (i for individuals, i=1..N; g for groups, g=1..G)

- **Decompose Y_{ig} into a between groups component Y_B = $_{.g}$ and a within groups component Y_W = Y_{ig} - $_{.g}$.**
  (replace observed *Total* score Y_T=Y_{ig} by Y_B (disaggregated group mean) and Y_W (individual deviation from group mean)

- **Note that Y_B and Y_W are orthogonal and additive: Y_T=Y_B+Y_W**

Multilevel Structural Models

- Decompose the population

  $\Sigma_T = \Sigma_B + \Sigma_W$

- Σ_B and Σ_W are modeled by separate models for the between groups and within groups structure

- Unfortunately, we cannot simply use S_B as an estimate of Σ_B, and S_W for Σ_W

Multilevel Structural Models

- Σ_W is estimated by the pooled within groups covariance matrix S_{PW}

$$S_{PW} = \frac{\sum^{G} \sum^{n} (Y_{ig} - {}_{g})(Y_{ig} - {}_{g})\tilde{}}{N - G}$$

- We can estimate the population within group structure by constructing and testing a model for S_{PW}

Multilevel Structural Models

- The sample between groups covariance matrix S_B is

$$S_B = \frac{\sum^{G} n_g (\ \ - \ \ _g)(\ \ - \ \ _g)^{\sim}}{G}$$

- This estimates

$$S_B = \Sigma_W + c\Sigma_B$$

- *c* is a scaling factor for the group size

- Thus, for S_B specify two models: one for the within groups structure and one for the between groups structure

Multilevel Structural Models

- Use the multigroup option of conventional covariance structure software to model S_{PW} and S_B (based on N-G and G observations) simultaneously

- The model for Σ_W must be specified for both S_{PW} and S_B, with equality restrictions between both `groups'

- The model for Σ_B is specified for S_B, with the scale factor *c* built into the model

- Strictly this applies only if all groups have the same size (balanced case)

- If group sizes differ, use average group size given by

$$C = \frac{N^2 - \sum^{G} n_g^2}{N\,(G-1)}$$

Multilevel CFA Example

- **Data are scores on 6 intelligence measures of 187 children from 37 families** (word list, cards, matrices, figures, animals, and occupations)

- **Decompose IQ measures:**

Means, variances and ICC for family data

Measure	Total Var.	Family Var.	Individual Var.	ICC
Word list	15.2	7.5	7.7	.37
Cards	28.5	13.7	14.8	.35
Matrices	16.4	5.2	11.1	.15
Figures	21.2	6.8	14.4	.16
Animals	22.8	8.5	14.4	.22
Occupat.	21.4	9.1	12.3	.28

Within Families Model

- S_{PW} based on 150 observations, S_B on 37
- CFA on S_{PW} for 2 factor model: χ^2=7.21, df=8, p=.51. (1 factor model: χ^2=44.87, df=9, p=.00)
- For S_B same model with equality restrictions across both 'groups' *plus* family level model
- Benchmark models for family level:

Family level benchmark models

Model	χ^2	df	p
Null model	125.4	29	.00
Independence model	52.5	23	.00
Saturated model	7.2	8	.51

Between Families Model

- **One factor model fits well (χ^2=21.3, df=17, p=.21)**
- **Thus we end with a 2 factor model for within families covariation and 1 factor model for between families covariation**

Multilevel CFA, Path Diagram

Multilevel CFA, Standardized Results

	Individual		Family
	I	II	I
Word list	.30*	-	.84*
Cards	.52	-	.78
Matrices	.70	-	1.00*
Figures	-	.30	.58
Animals	-	.70	.86
Occup.	-	.48	$.33^{ns}$

Correlation between individual factors: 0.22^{ns}; * = fixed; ns = not significant

Multilevel Structural Models, Comments

- **Higher level sample size should be sufficient: at least 50, preferably 100**

- **We need S_{PW} and S_B: can be computed by standard software or BW (Muthén), Split2 (Hox), Streams (Gustafsson)**

- **In path models: often there are higher level variables that don't exist at lower levels (e.g.class size)**
  - **Lisrel: put pseude values (1,0) in S_{PW}, phantom variables in model, adjust df**
  - **Eqs, Amos: don't worry**

- **Setups are non-standard**
  - **Lisrel: set AD=OFF**
  - **Simplis: don't use**
  - **Take a look at the STREAMS preprocessor**

Latent Growth Curve Model, Path Diagram

- **Note that the latent growth curve model is a *fixed occasions* model**

Example Data Set from Rogosa and Saner

- 200 cases
- 1 dependent variable Y
- 5 equidistant time points (t = 0,1,2,3,4)
- 1 time invariant (person level) covariate Z

- Artifical data follow linear growth model
- 7% Data points randomly deleted (MCAR)

Path diagram:

Latent Growth Model Results

(Z centered, AMOS, ML, missing data set, analysis on 1119 data points, efficiency=93% (multilevel regression was 73%)

Model	(1)	(2)
interc	44.2 (.55)	44.2 (.50)
time	4.9 (.16)	4.9 (.14)
Z		1.5 (.24)
Z*time		.58 (.07)
mean σ^2_ε	12.4	12.4
σ^2_{int}	52.8 (6.18)	42.0 (5.19)
σ^2_{time}	4.1 (.56)	2.6 (.42)
σ_{i*t}	-3.3 (1.40)	-7.1 (1.27)
r_{i*t}	-.23	-.68
χ^2	14.4	31.4
df	10	13
p	.17	.00
TLI	1.00	.99

SMC on interc .20
on time .37

Comparing Estimates, Missing Data, Z Centered

pop.	exp. var.	multilevel regression	latent growth model
44	interc.	44.2 (.52)	44.2 (.50)
5	time	4.9 (.14)	4.9 (.14)
1.4	Z	1.4 (.24)	1.5 (.24)
.67	Z*time	.61 (.07)	.58 (.07)
12	σ^2_{ε}	12.4 (.79)	8.7-14.9
47	σ^2_{int}	40.8 (5.16)	42.0 (5.19)
3.2	σ^2_{time}	2.5 (.41)	2.6 (.42)
-.73	r_{it}	-.66	-.68

extra's:	nice plots	
	robust tests (MLn)	robust tests (EQS);
	missings (Amos, Mx	missings (MLn?)
	reliabilities (HLM)	SMC for Y

Conclusions

- Multilevel Regression Models (MRM) and Latent Growth Models (LGM) perform highly similarly. With data sets and models where both may be used, there are no strong reasons to prefer one over the other

- MRM are more elegant than LGM if we have a large number of unequally spaced data points. At the limit, LGM may become impossible

- MRM are better if we want to include higher levels. Possible in LGM, but difficult

- LGM can incorporate complex path models, MRM not

- LGM can incorporate multiple group models, MRM not

Example with Real Data: Goldstein Adolescent Growth Data

- Repeated measures of heights of adolescent children
- 110 boys, 436 measurements at ages 11-16, plus a measure of their adult height
- Measures: height in cm, bone width (age-standardized)

Characteristics that make this data a challenge:

- Dependent variable adult height at second level, difficult in multilevel regression
- Number of repeated measures and their spacing varies across persons, difficult in latent growth modeling

Goldstein's (1995) Multivariate Repeated Measures Model

$$Y_{tj} = \delta_{tj}^{(1)}\left(\gamma_0^{(1)} + \sum_{h=1}^{5}\gamma_h^{(1)} T_{tj}^{h} + u_{0j}^{(1)} + u_{1j}^{(1)} + e_{tj}^{(1)}\right) +$$

$$\delta_{tj}^{(2)}\left(\gamma_0^{(2)} + u_{0j}^{(2)} + e_{tj}^{(2)}\right) +$$

$$\delta_{j}^{(3)}\left(\gamma_0^{(3)} + u_{0j}^{(3)}\right)$$

δ is a dummy indicating the measure:

$\delta^{(1)}$ = 1 if Y is adolescent height, 0 otherwise
$\delta^{(2)}$ = 1 if Y is bone age, 0 otherwise
$\delta^{(3)}$ = 1 if Y is adult height, 0 otherwise

Multivariate Repeated Measures Model

$$Y_{tj} = \delta_{tj}^{(1)} \left(\gamma_0^{(1)} + \sum_{h=1}^{5} \gamma_h^{(1)} T_{tj}^{h} + u_{0j}^{(1)} + u_{1j}^{(1)} + e_{tj}^{(1)} \right) +$$

$$\delta_{tj}^{(2)} \left(\gamma_0^{(2)} + u_{0j}^{(2)} + e_{tj}^{(2)} \right) +$$

$$\delta_{j}^{(3)} \left(\gamma_0^{(3)} + u_{0j}^{(3)} \right)$$

- **Adolescent height: a 5th degree polynomial for age, time-dependent errors, and intercept and linear slope variance at the 2nd (person) level**

- **Standardized bone weight: intercept, linear slope for age, time dependent errors, and intercept variance at the 2nd level**

- **Adult height: intercept and 2nd level variance**

Results Multivariate Multilevel Regression on Adolescent Growth Data (age centered at 13)

Fixed part	Estimate (SE)
adult height:	
intercept	174.6
adolescent height	
intercept	153
age1	6.91 (.2)
age2	.43 (.09)
age3	-.14 (.03)
age4	-.03 (.01)
age5	.03 (.03)
bone age	
intercept	.21 (.09)
age1	.03 (.03)

Random part:

adult height	62.5			
height intercept	49.5	54.5		
age1 slope	1.11	1.14	2.5	
bone intercept	.57	3.00	.02	.85

Problem with Multivariate Multilevel Regression: How to Predict Adult Height

Solution 1: use multiple regression on covariances in random part (2 step procedure)

predictor	b (SE)	beta	NB SE's
height intercept	1.1 (.05)	1.0	assume
age1 slope	.0 (.20)	-.01	N=110
bone intercept	-3.2 (.38)	-.36	

Multiple correlation: R^2=.83

Solution 2: Goldstein (1995) suggests to predict adult height using the estimated residuals from MLn. Advantage: all information in the data is used. Disadvantage: difficult to interpret.

Adolescent Growth as Latent Curve Model

- Data must be translated from multilevel structure into flat data file with repeated measures as variables
  - Problem: measures between 11-16 yrs imply 6 measures, on average each boy has 4 measures, thus 33% of measures is missing
  - Solution: use modern ML methods to estimate model in the presence of missing data

Path Diagram for Latent Curve Model of Adolescent Growth

Results Latent Growth Model on Adolescent Growth Data

Estimates for paths to adult height: R^2=.9

predictor	b (SE)	beta
height intercept	1.21 (.05)	1.25
age1 slope	-3.71 (.85)	-.31
bone intercept	-4.34 (.47)	-.48

Results multiple regression analysis: R^2=.83

predictor	b (SE)	beta	
height intercept	1.08 (.05)	1.01	(SE
age1 slope	-.03 (.20)	-.01	N=
bone intercept	-3.16 (.38)	-.36	110)

More about the Latent Growth Model for Adolescent Height

- χ^2_{77}=702 p=.00 TLI=.84
- Model predicts well, but fits poorly
- From MLn results we expect a higher degree polynomial; but SEM becomes unstable after fitting a quadratic trend
- Most of the mis-fit relates to poor modeling of bone-size and height trends:
  - The bone size measure appears poorly age-standardized (both means and variances vary over the years
  - Height trend needs higher polynomials that cannot be fitted
- The functional sample size for these data is N≈55, which explains estimation problems

- **Part of the mis-fit relates to non-normality of the data. There are correction formula's for this, but these need full data**

Conclusions

- Multilevel Regression Models (MRM) and Latent Growth Models (LGM) perform highly similarly. When both are applicable, there are no strong preferences

- MRM are more elegant than LGM if we have a large number of unequally spaced data points. At the limit, LGM may become impossible (the adolescent growth data are close to that limit)

- MRM are better if we want to include higher levels

- LGM can incorporate complex path models, MRM not or with difficulty (2-step modeling on covariance matrix at higher levels)

- LGM can incorporate multiple group models, MRM not or with difficulty (using interactions with dummy coded groups)