Unit 3: Multilevel Panel and Growth Models IV. Latent Growth Models and Multilevel-SEM Syntheses

PS2701-2019

Longitudinal Analysis

Weeks 11-12

Professor Steven Finkel

IV. Latent Growth (SEM) Models

Growth Models in SEM

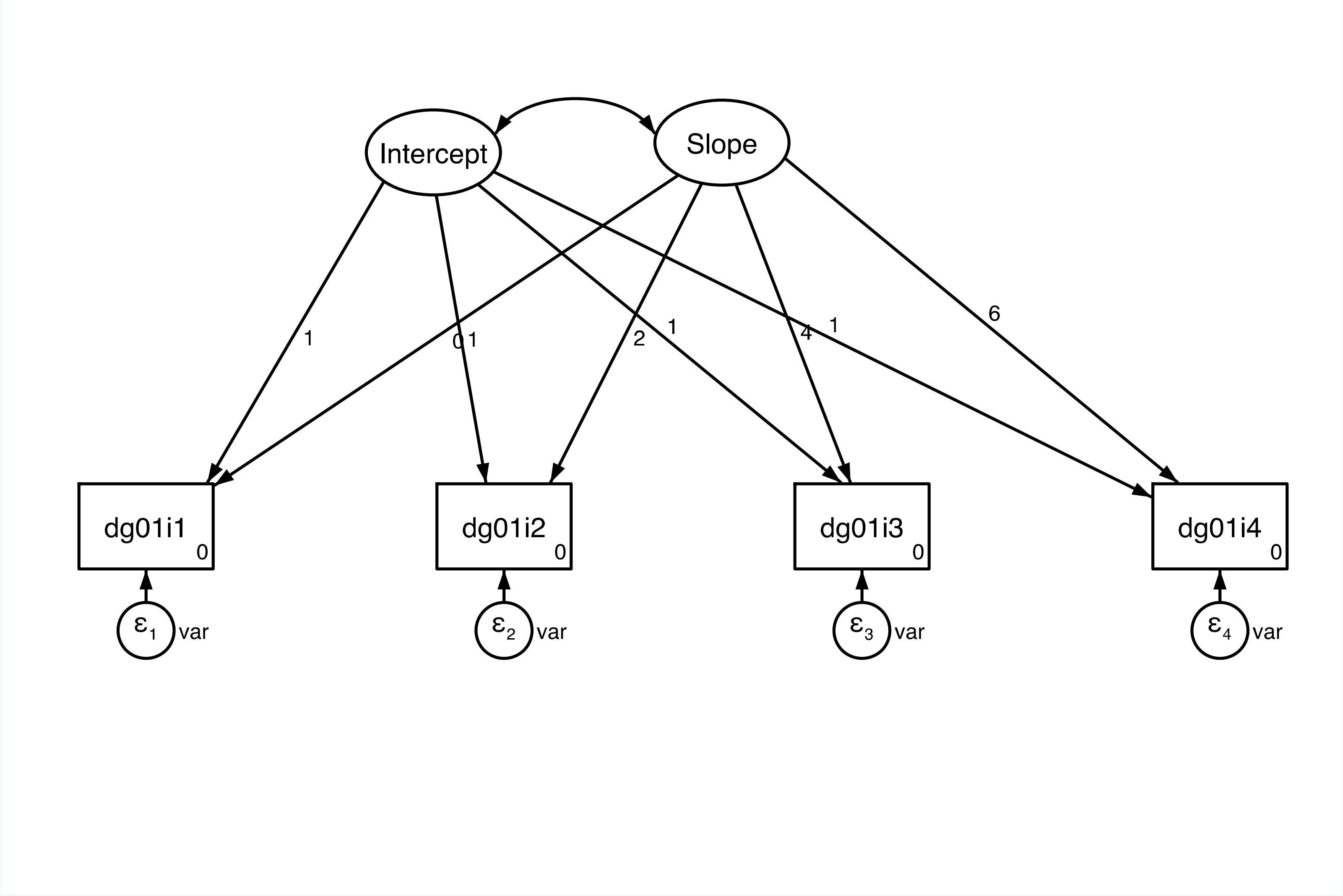
- Over the past fifteen years, models have been developed for incorporating the longitudinal growth perspective into SEMs
- Actually, this was one of the first ways that SEM panel analysis was integrated with other traditions, in this case the multilevel modeling framework
- Advantages of SEM approach:
  - flexibility in modeling and testing (via χ^2 differences, e.g.) of alternative specifications, error structures, etc.;
  - Ability to incorporate multiple indicators of the X and Y variables into the analysis to control for measurement error;
  - Ability to incorporate reciprocal linkages or associations between growth processes in different domains – i.e. it will be possible to have growth curves of two different variables and see how growth in one influences growth in another over time

- So what is known as **Latent Curve**, or **Latent Growth Models** provide a very flexible way of approaching multilevel longitudinal (growth) processes.
- The way it proceeds is surprisingly simple: we specify the two random effects in the growth model (the random intercept and random slope for time) as ***latent variables*** within the SEM framework.
- These latent variables are then linked to the observed outcome over time via **the measurement error portion of the SEM model(!!!) with fixed constraints that model their appropriate theoretical linkages**
- Then, as in all SEM analysis, we estimate the parameters that minimize the difference between the observed variances and covariances of the X and Y variables and the variance-covariance matrix that is implied by the model, and finally assess the model (if it is overidentifed) in terms of how well it can "reproduce" the variances and covariances of the observed variables.
-

Example: Growth in Democracy, 4 waves, 1996-2002

- Use wide-form data at first, then we will show long-form equivalent via "GSEM" later
- We model each Polity outcome as being predicted by two latent variables called "Intercept" and "Slope"
- So, at Level 1 of the hierarchy, we have a *T*-Indicator Y-measurement model with two latent variables, one corresponding to the intercept of the growth trajectory and one corresponding to the slope.

Latent Growth Curve, No Level 2 Predictors



- Can you see that the **parameters** of the multilevel growth model are actually **latent variables** from the perspective of the SEM approach? That is, INTERCEPT for each case is a latent η variable with 4 indicators, Y1, Y2, Y3, and Y4, and SLOPE is also a latent η variable with 4 indicators Y1, Y2, Y3, and Y4.
- So the outcome variables in the multilevel growth framework are treated as if they are imperfect measures of the INTERCEPT and SLOPE.
- Weird, but true!
- The means of the latent variables give you the "fixed effect" associated with the intercept and slope; their variance gives you the variance component

- Look again at Level 1 Multilevel Growth Equation from earlier PPTs:

$$(1) \qquad Y_{it} = \beta_{0i} + \beta_{1i} Time_i + \varepsilon_{it}$$

- INTERCEPT is actually β_{0i} and it is "causing" Y_{it} with a regression coefficient of "1" in each wave
- SLOPE is actually β_{1i} and it is "causing" Y_{it} with a regression coefficient of "0" at time 1, a regression coefficient of "2" at time 2, a regression coefficient of "4" at time 3, and so forth? (These would be the values of "Time" for those observations)
- Writing the equation for all waves gives:

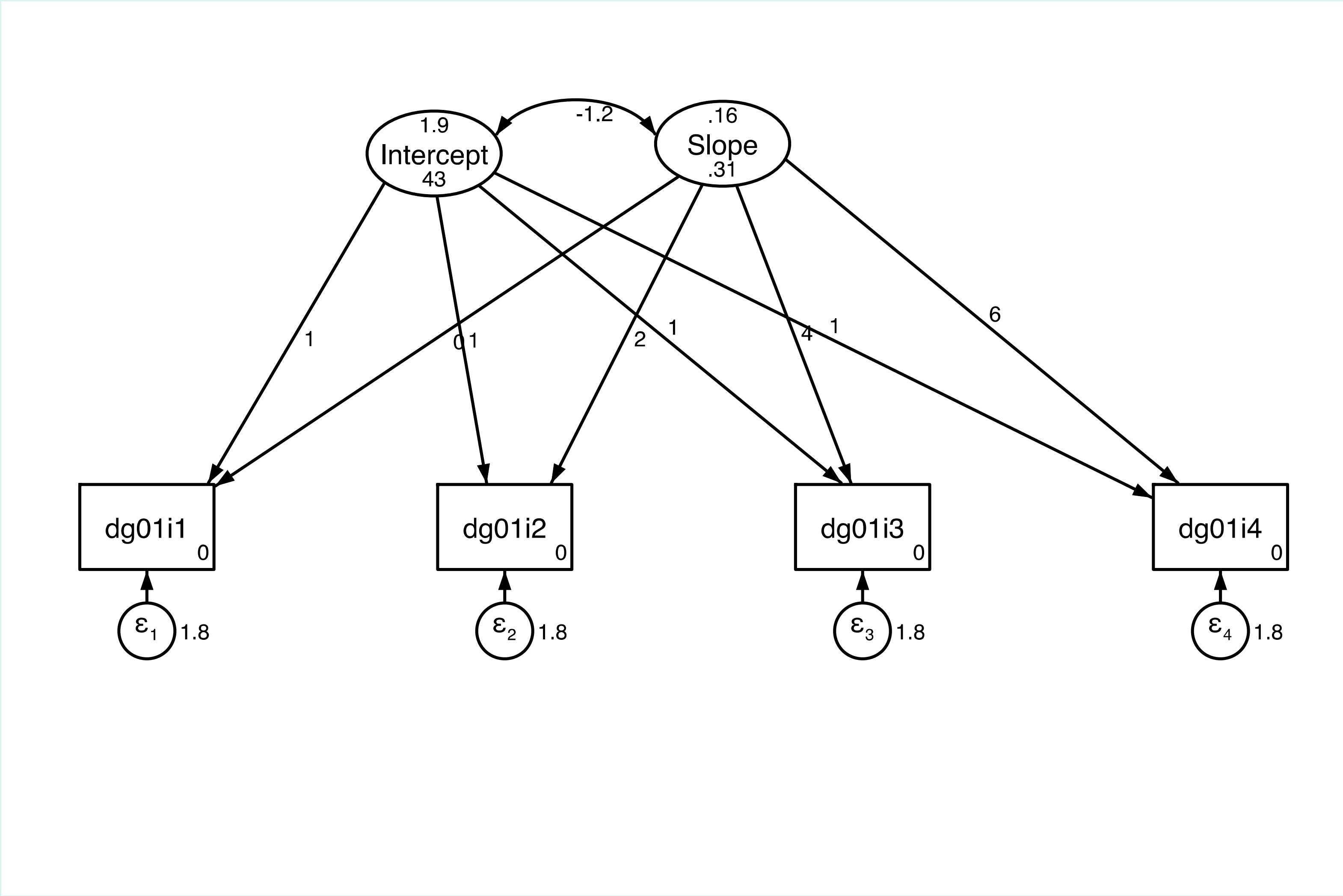
$$Y_{1i} = \beta_{0i} * 1 + \beta_{1i} * 0 + \varepsilon_{1i}$$

$$Y_{2i} = \beta_{0i} * 1 + \beta_{1i} * 2 + \varepsilon_{2i}$$

$$Y_{3i} = \beta_{0i} * 1 + \beta_{1i} * 4 + \varepsilon_{3i}$$

$$Y_{4i} = \beta_{0i} * 1 + \beta_{1i} * 6 + \varepsilon_{4i}$$

Stata SEM Results: Latent Growth Model



Additional Tests

- Examine GOF statistics to assess model fit
- Evaluate Modification Indices to see where the model could be improved
  - Relaxation of equal error variances (heteroskedasticity in the time-specific errors)
  - Relaxation of no autocorrelation assumption
- Add Level 2 explanatory variables
- Add Level 1 time-varying covariates

Growth Model with Level 2 Explanatory Variables

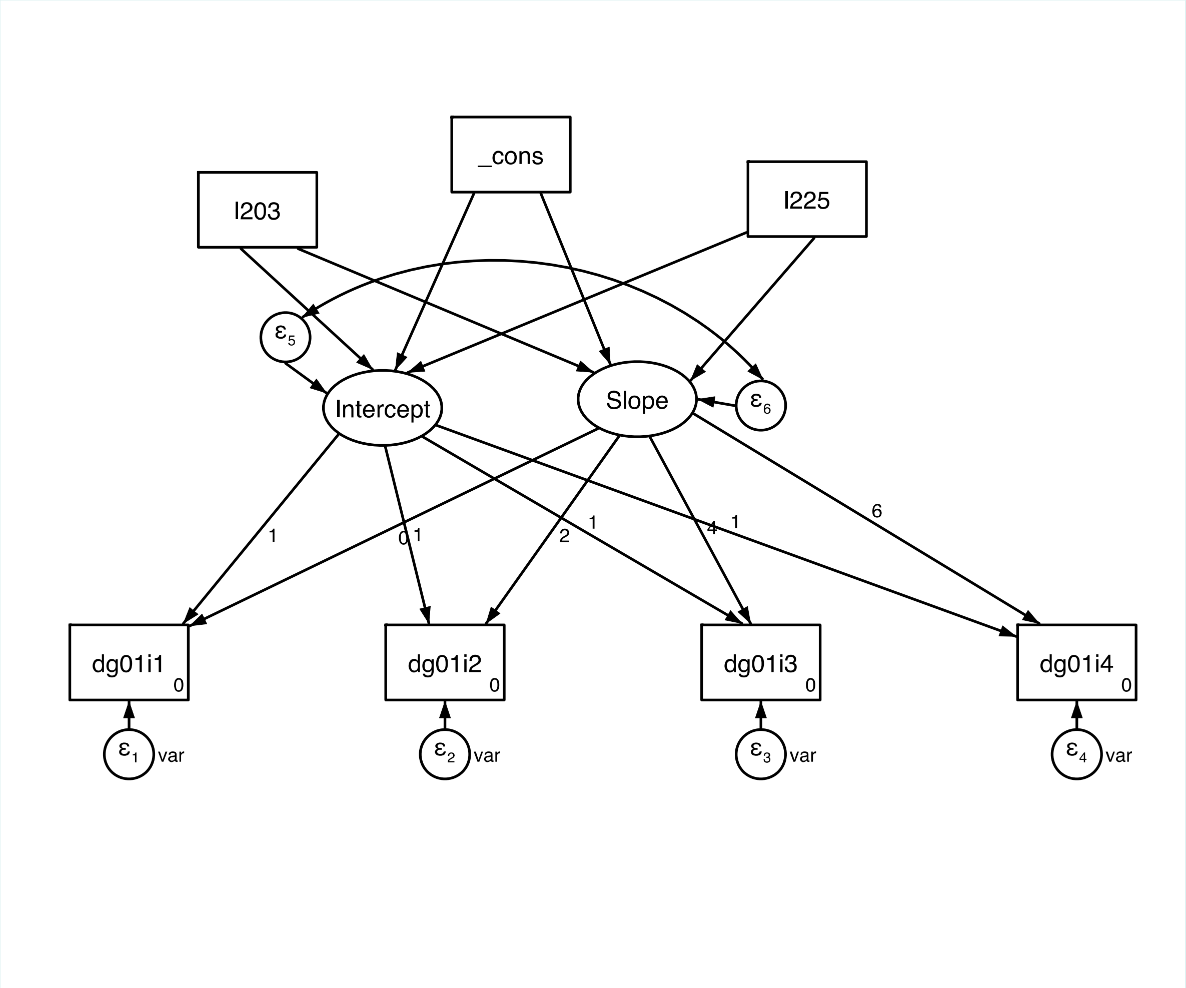
- Recall the Level 2 Growth Model:

$$(a) \quad \beta_{0i} = \beta_{00} + \beta_{01} X_{1i} + \beta_{02} X_{2i} + \zeta_{0i}$$

$$(b) \quad \beta_{1i} = \beta_{10} + \beta_{11} X_{1i} + \beta_{12} X_{2i} + \zeta_{1i}$$

- Since, in the SEM framework, the β_{0i} and β_{1i} are latent variables, we need only to model them now as Latent Endogenous Variables and have the specific Level 2 X variables as exogenous predictors.

Full Multilevel SEM Growth Model

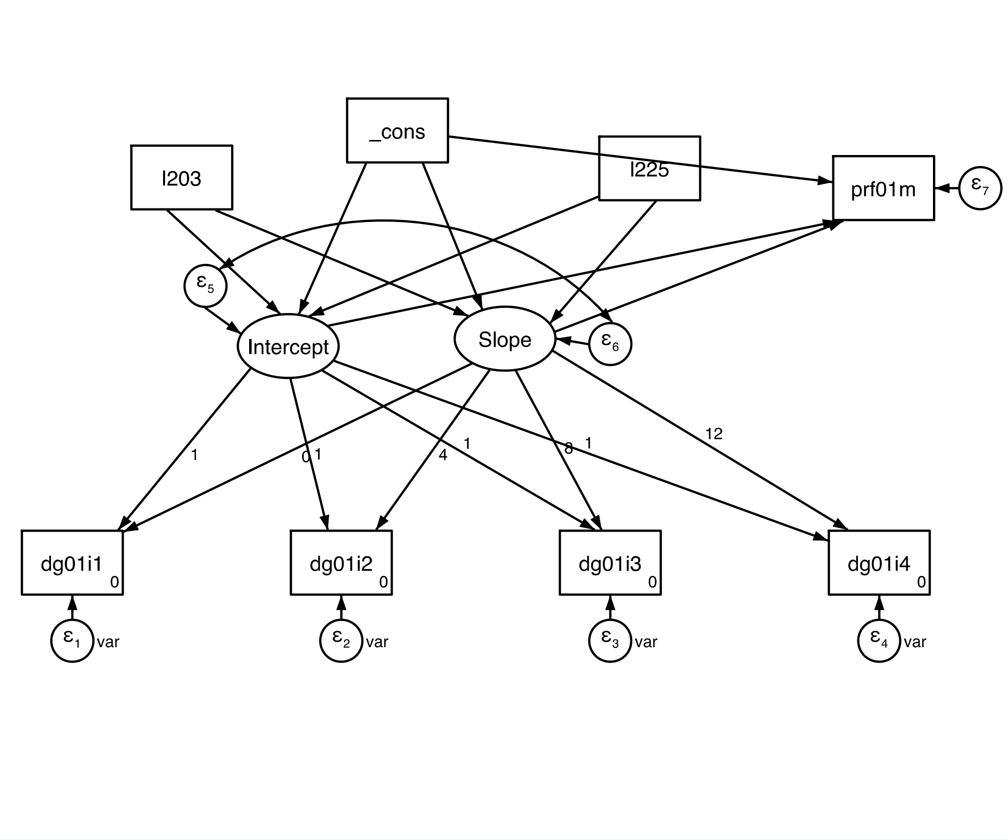


Important Note: Estimating the effect of _cons on INTERCEPT and SLOPE provides the population average fixed effects for these parameters!!

Extensions

- Multiple indicators for latent exogeneous varaiables
- Effects of growth curve parameters(INTERECEPT and/or SLOPE) on other variables, as in mediation models
- Effects of one growth curve on another
- Inclusion of time-varying covariates
- Non-linear growth with an additional latent variables representing, for example, the effects of a slope for "TIME-SQUARED"
- Heteroskedastic error variances
- Autocorrelated disturbances

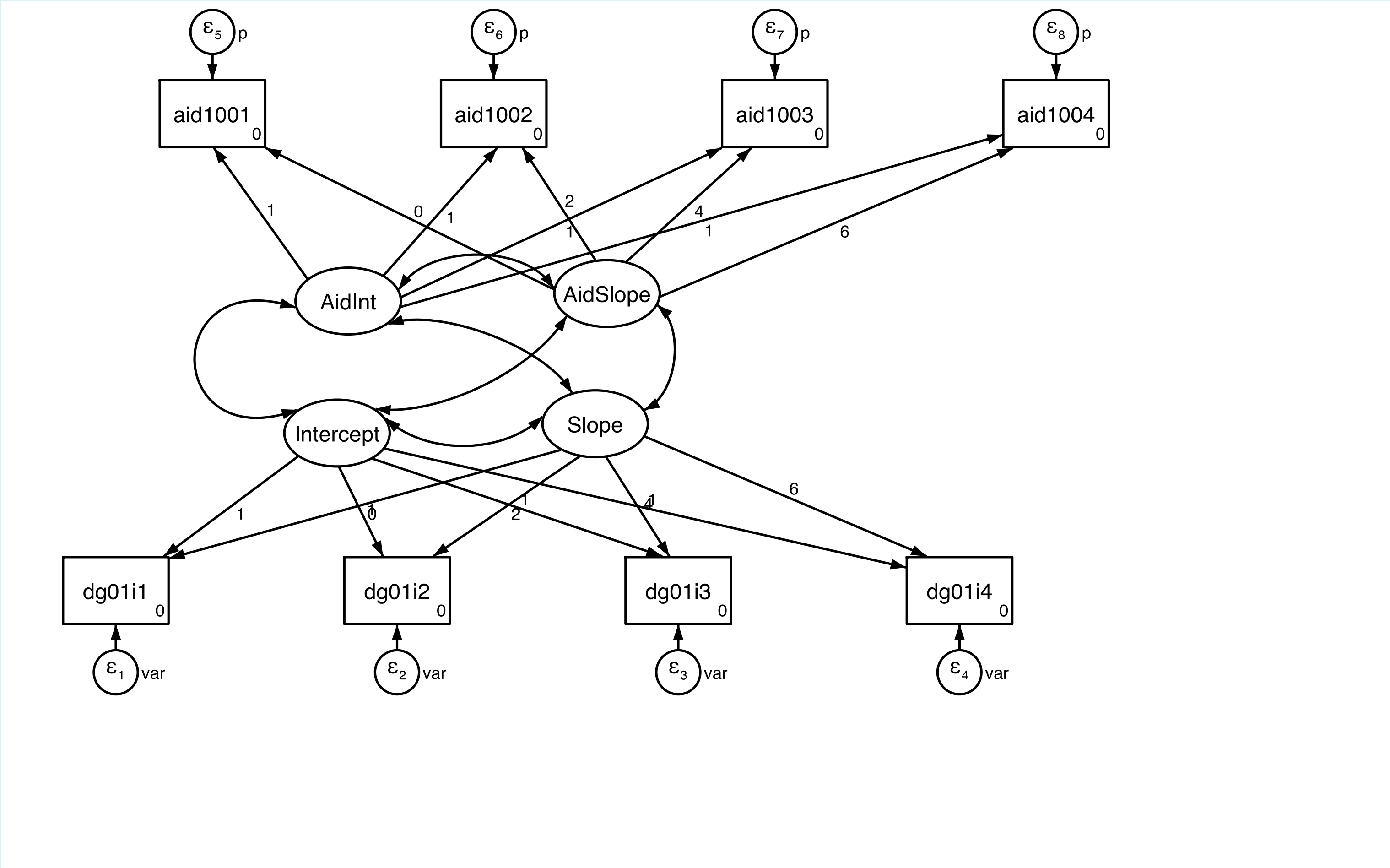
Example: Growth Parameters as Independent Variables



Hypothesis: Countries with Stronger Democratic Trajectories Have Higher Levels of GDP Growth During the Time Period

Note: Model is also an example of an **SEM mediation model**: Prior Democracy and Ethnic Fractionalization "cause" the Level 2 growth parameters, which then "cause" Level 2 GDP growth. See Selig, James, and K. Preacher. 2009. "Mediation Models for Longitudinal Data in Developmental Research, *Research in Human Development* 6(2-3), 144-69.

Bivariate Growth Curve Model



Extensions: Curran et al (2014), “Latent Curve Model with Structured Residuals” (LCM-SR)

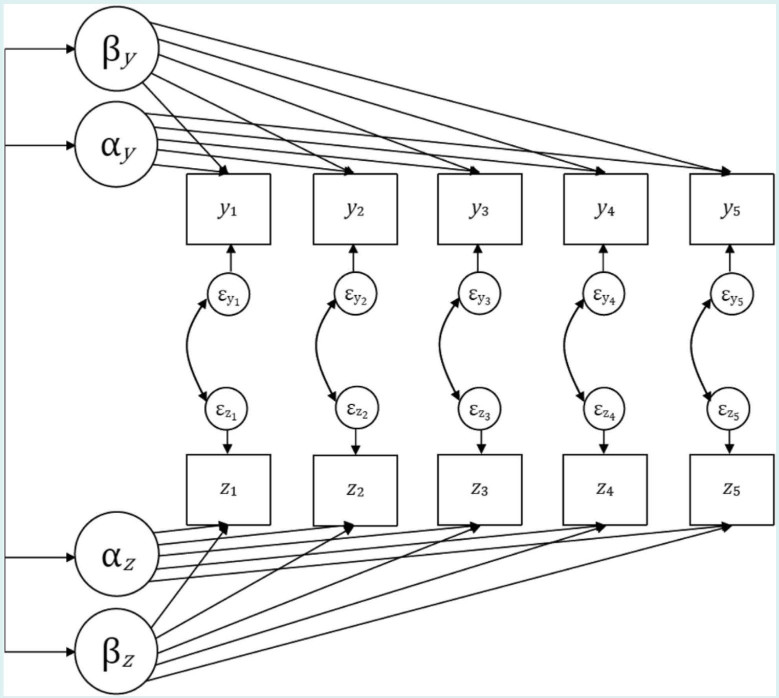


Figure 4. Bivariate unconditional linear latent curve model for five repeated measures.

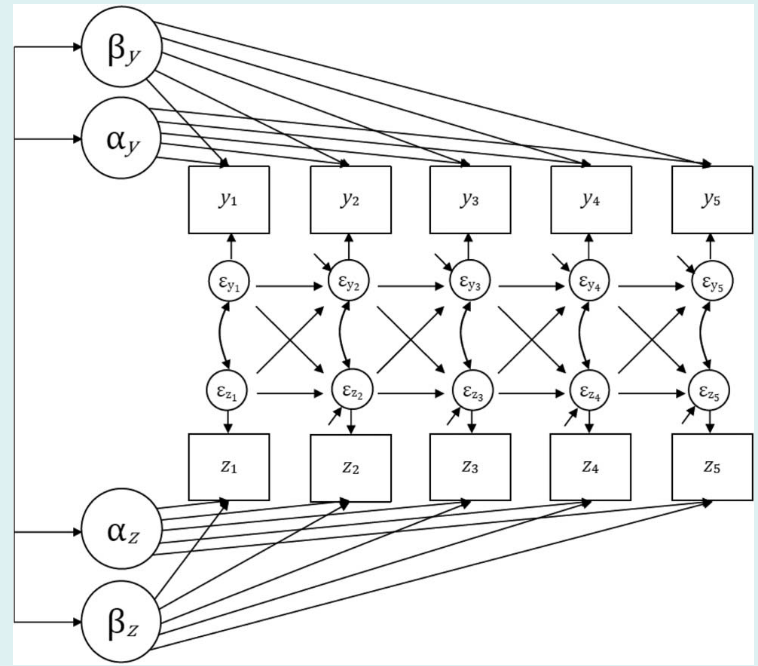
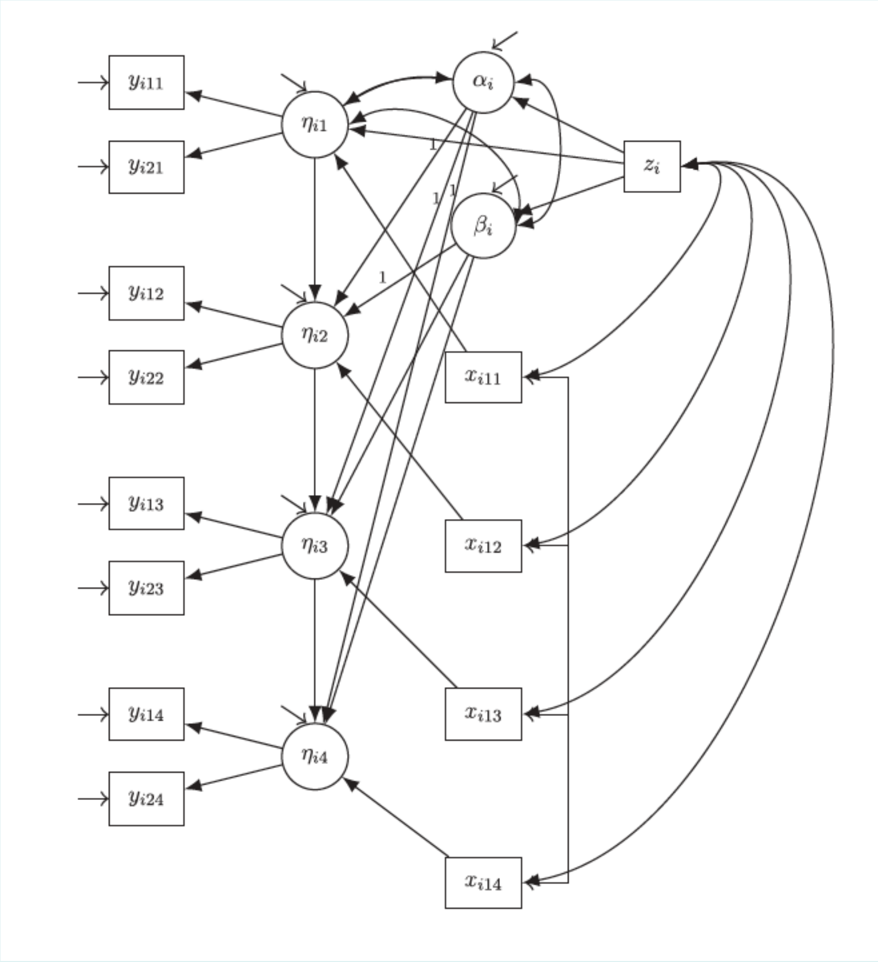


Figure 7. Bivariate unconditional linear latent curve model with structured residuals for five repeated measures.

Model on left can be estimated in Stata; Model on right in Mplus (and R?)

Extensions: Bianconcini and Bollen (2018), “Latent Variable Autoregressive Trajectory Model (LV-ALT)



Features:

- Multiple Indicators of a Latent Variable Over Time
- Autoregressive Effects of the Latent Variable
- Random Intercept
- Random Slope for Time
- Time Varying and Time-Invariant Covariates
- Many models are nested within this model by imposing particular constraints

V. Multilevel SEM

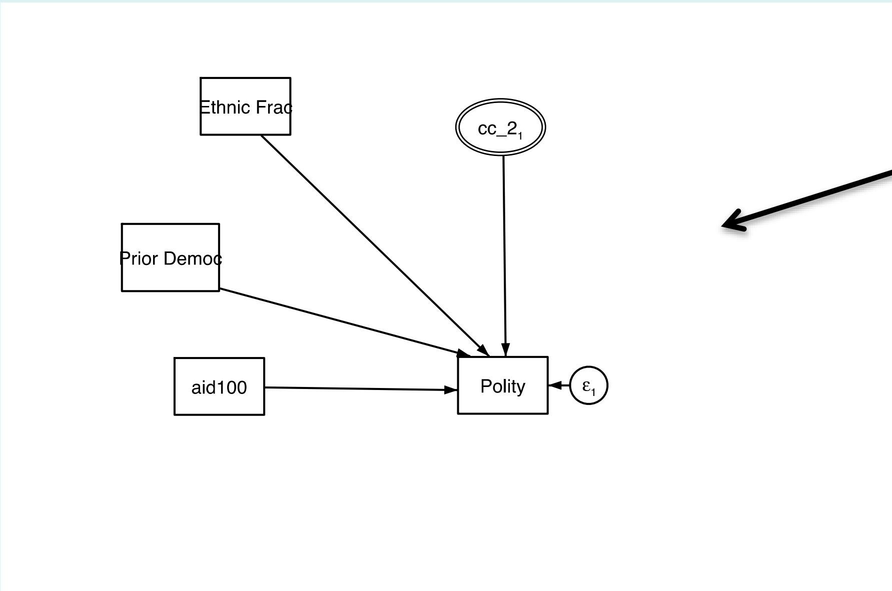
Multilevel Structural Equation Modeling (MSEM) with "Long" Data, Multilevel Random Effects/ Random Coefficients

- More general procedure (**GSEM** in Stata, MPLUS "Two Level" Module) allows estimation of multilevel models with random effects with long-form data
- Advantages: Can use for all kinds of random coefficient models, not only "time" which was made possible because we knew the exact parameter constraints to impose with wide data
- Allows easier "pooled" SEM estimation of data with larger T
- ****Allows estimation of RE-Hybrid model "between effects" via inclusion of cluster-level means of an independent variable *OR* via a latent variable random effect (which corrects for cluster mean unreliability)
- Straightforward incorporation of measurement error/multiple indicators as in "wide" SEM
- Straightforward extensions to mediation analysis, non-continuous DVs

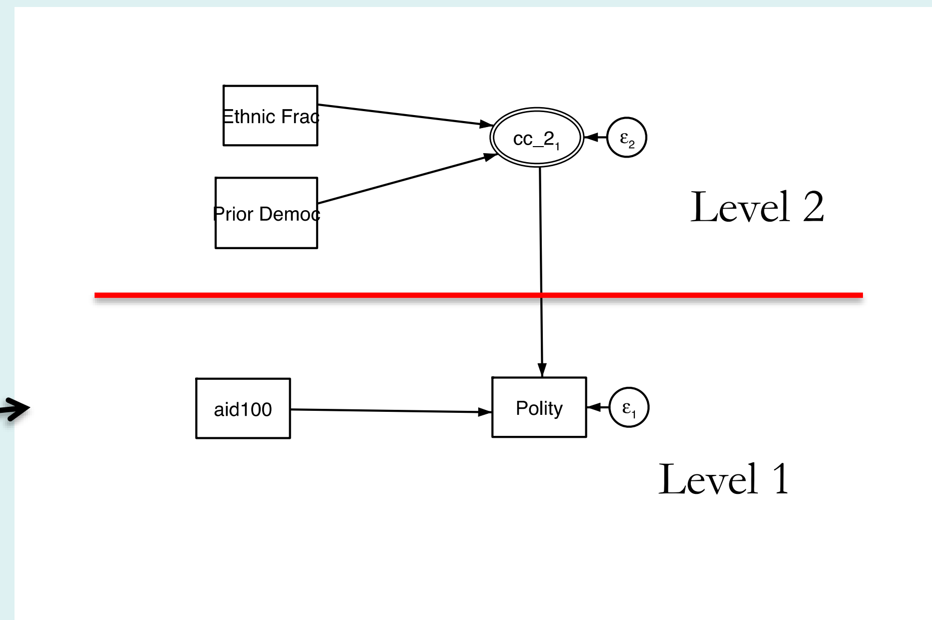
Disadvantages:

- More difficult to deal with heteroskedasticity and autocorrelation, relaxation of other equation-by-equation constraints (as in all long-form data analyses)
- Missing values more difficult to handle in Stata GSEM
- More difficult to assess model fit and test significance of nested models due to inapplicability of chi-square statistic as a summary measure (though still can compare nested models)
- Models sometimes have difficulty converging, especially in Stata compared to Mplus, the dominant multilevel/SEM stand-alone package
- Convergence and estimation problems compounded as the number of random effects in a model increases

GSEM Random Effects Model

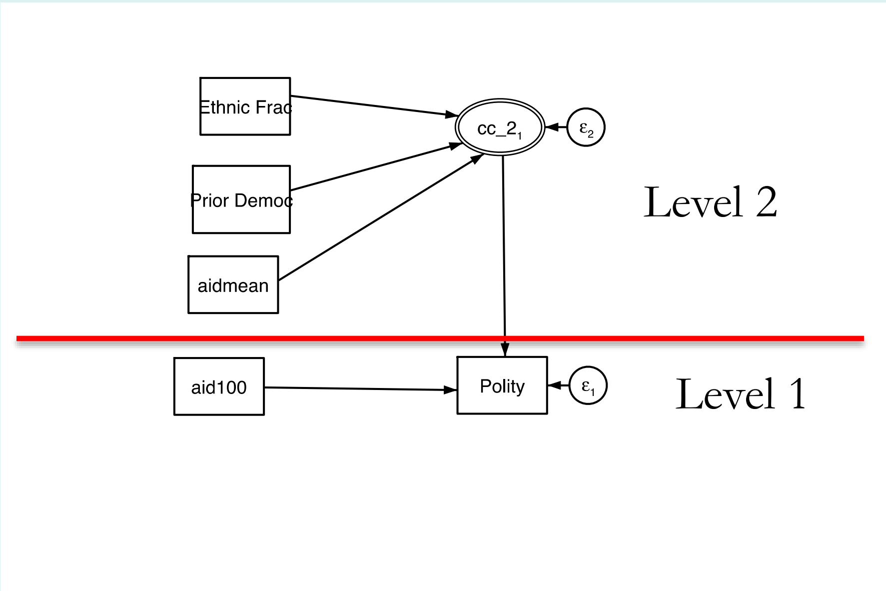


Linear Model with Random Effect

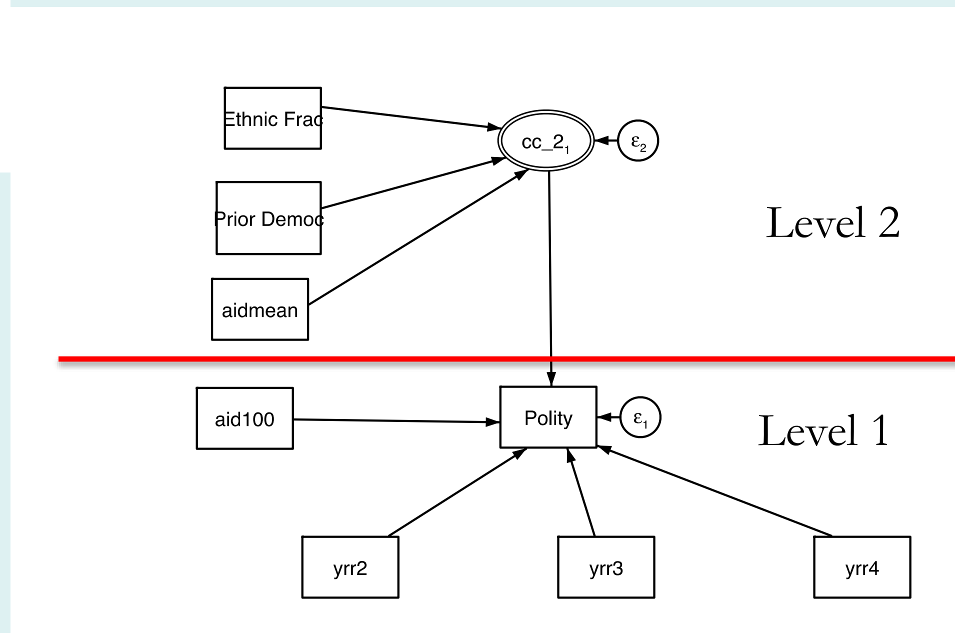


Multilevel Model Version

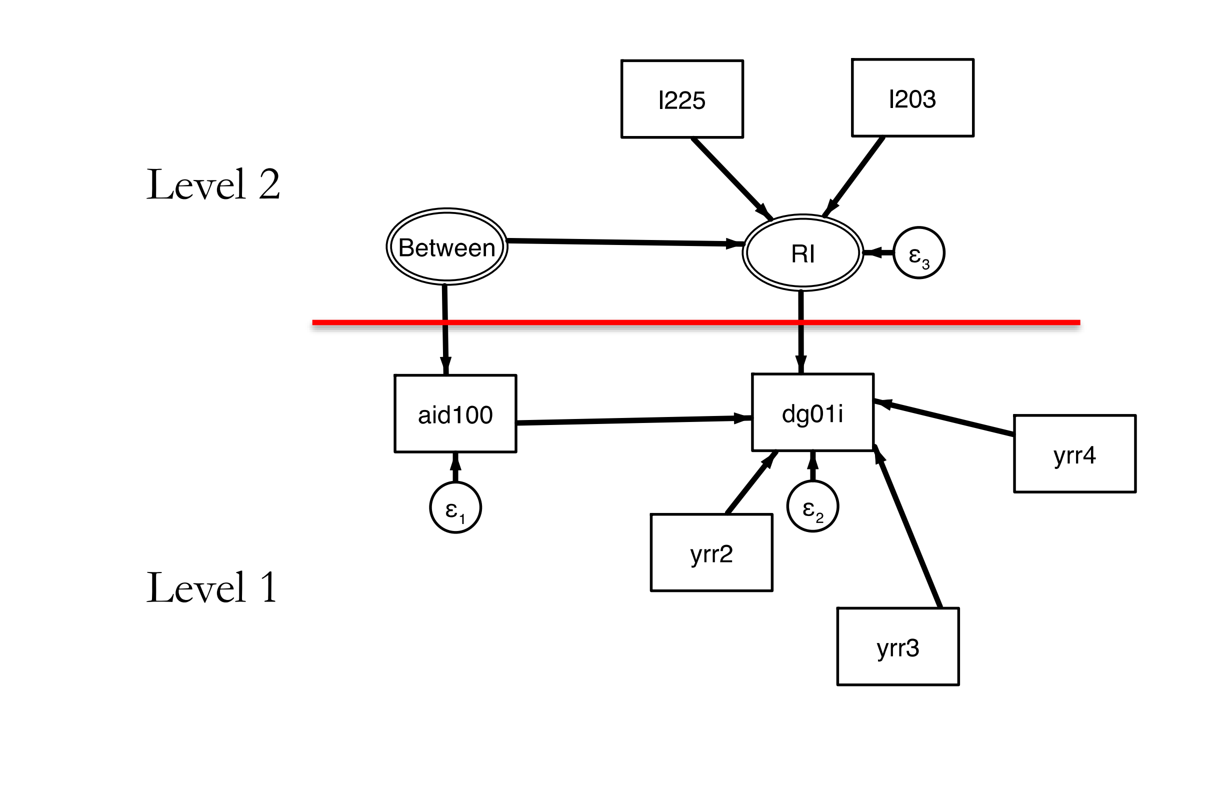
Multilevel Random Effects Hybrid Model



With time dummies

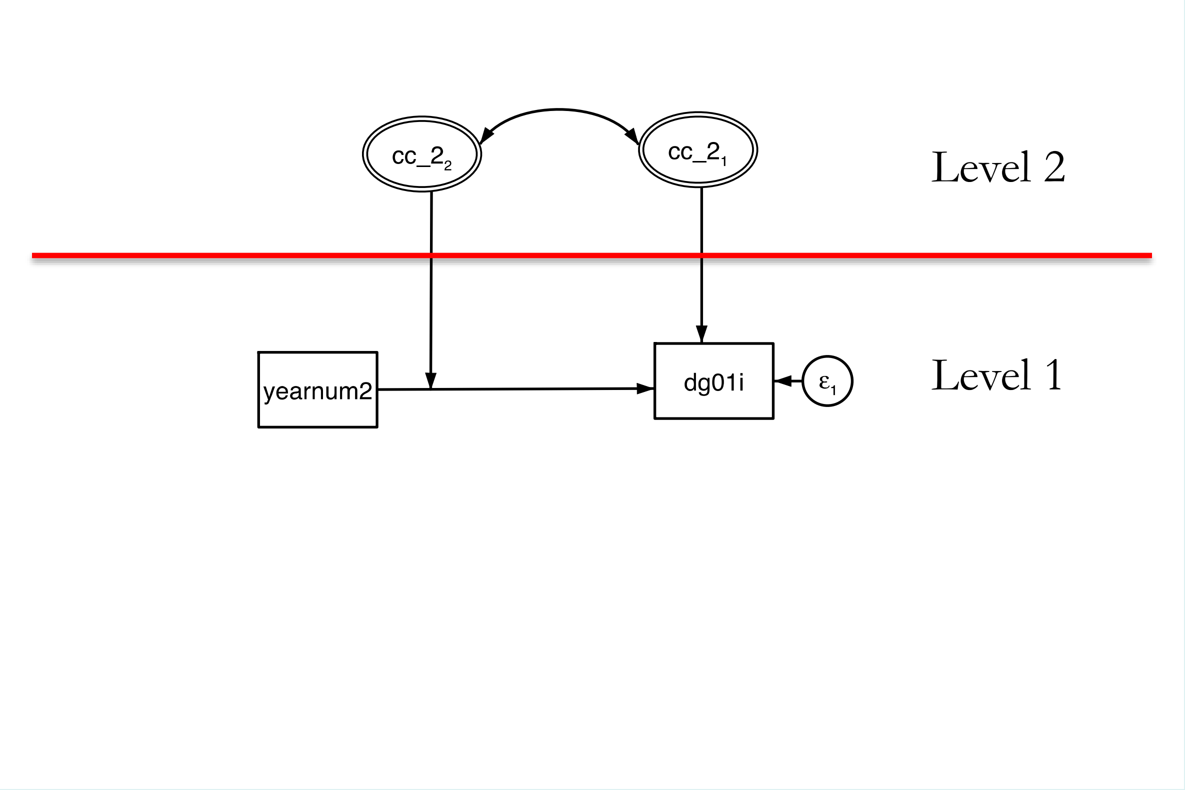


Full ML-SEM RE-Hybrid Model with Latent “Between” Random Effect



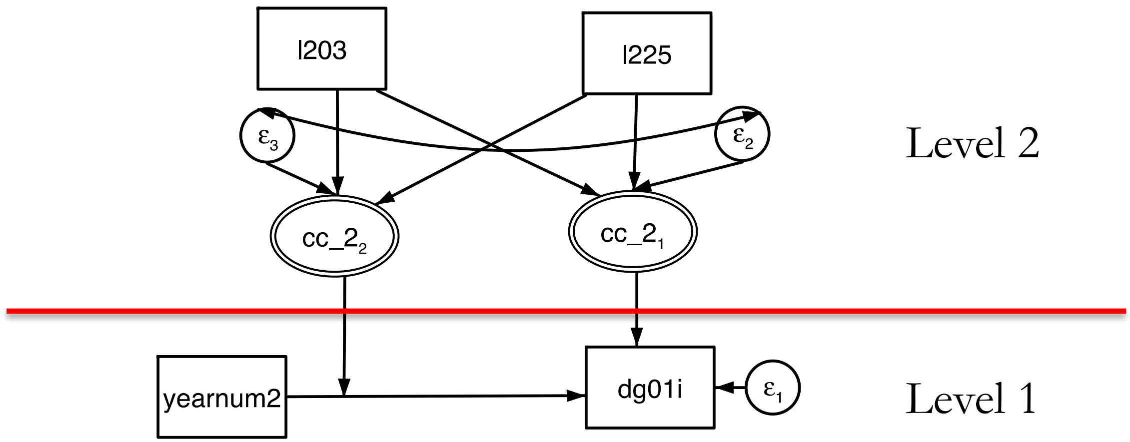
NOTE: This version uses the “raw” (not mean-deviated) Level 1 variable due to convergence issues. The ”within” effect here is identical; see STATA do file for post-estimation commands to obtain the corresponding “between” effect

GSEM Multilevel Growth Model: Level 2 Random Intercept and Random Time Slope

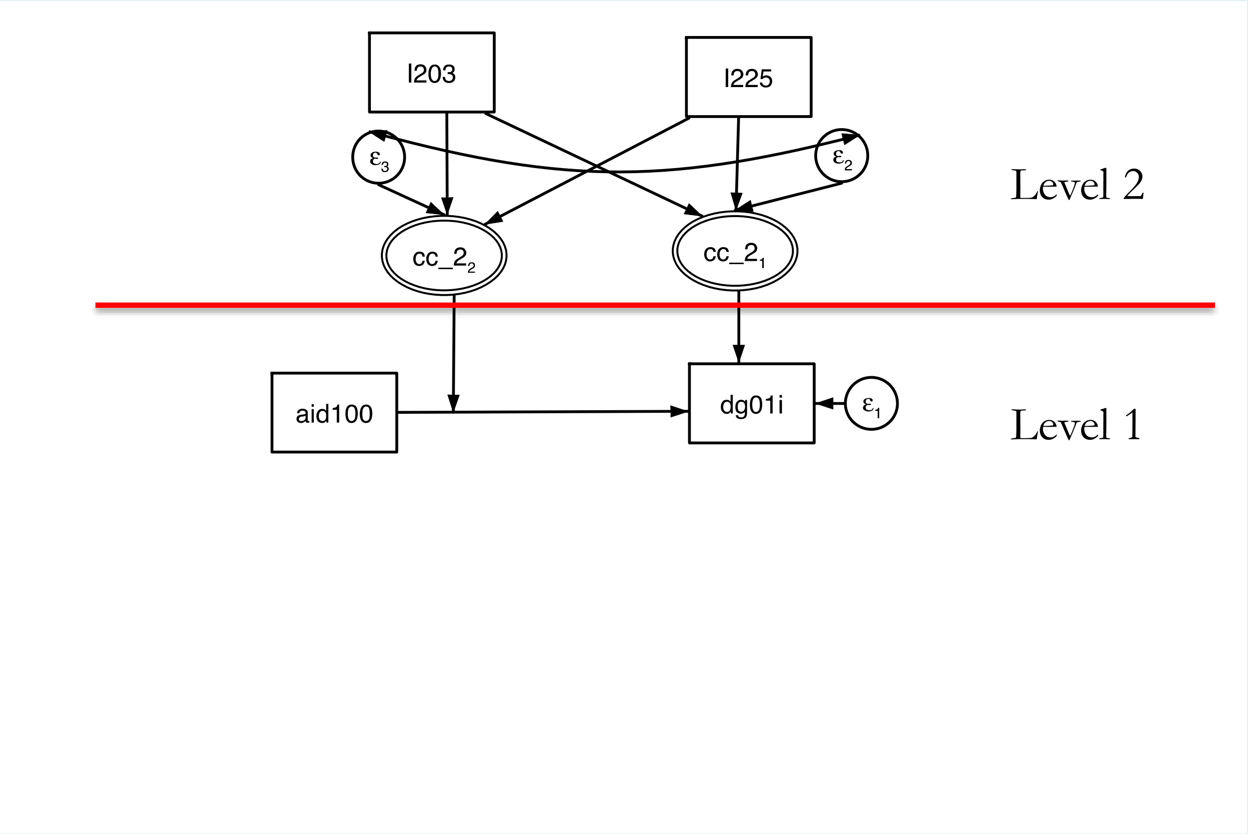


cc_2 is the Level 2 marker variable; double circles signifies a Level 2 random effect

Long-Form Multilevel Growth Model: Level 2 Explanatory Predictors

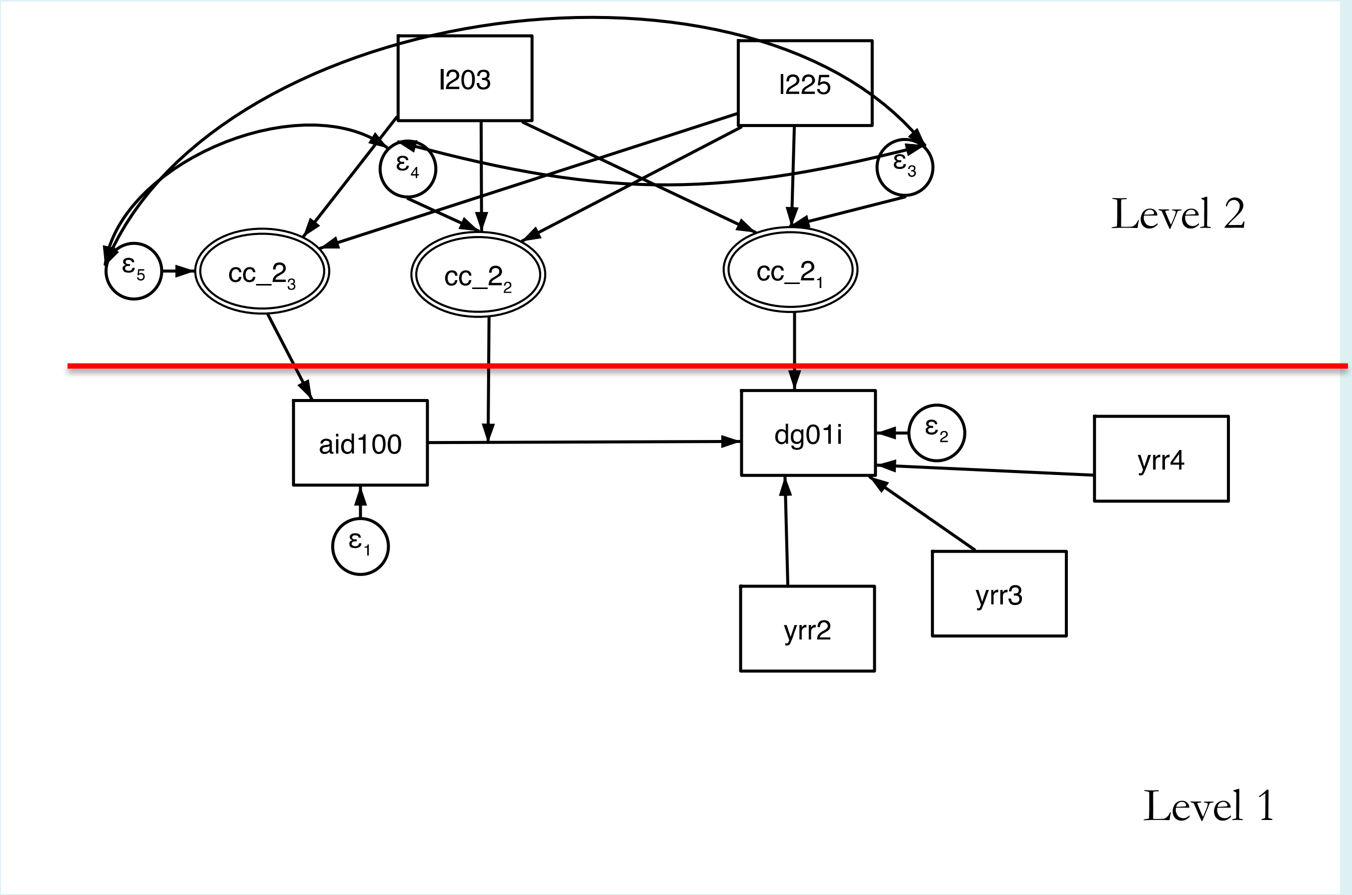


Multilevel Random Coefficient Model with Level 2 Explanatory Predictors

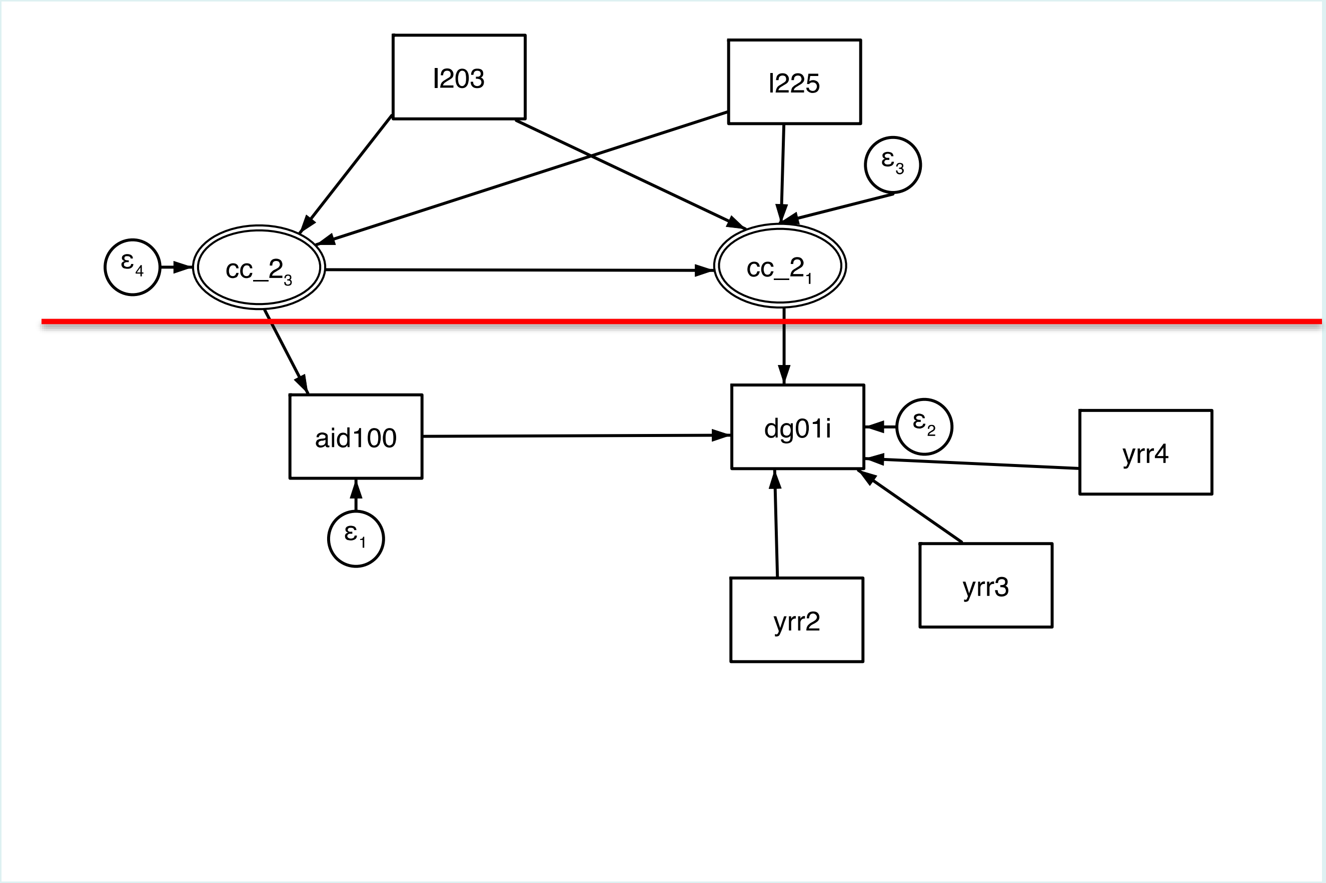


Exact same structure for this model as the growth model, but IV here is another time-varying covariate, Aid100

Full ML-SEM Multilevel Random Coefficient Model with Latent "Between" Random Effect



Alternative Model: Between and Within "Causal" Effects, No Random Slope



Extensions: Measurement Error and "Slopes as Predictors"

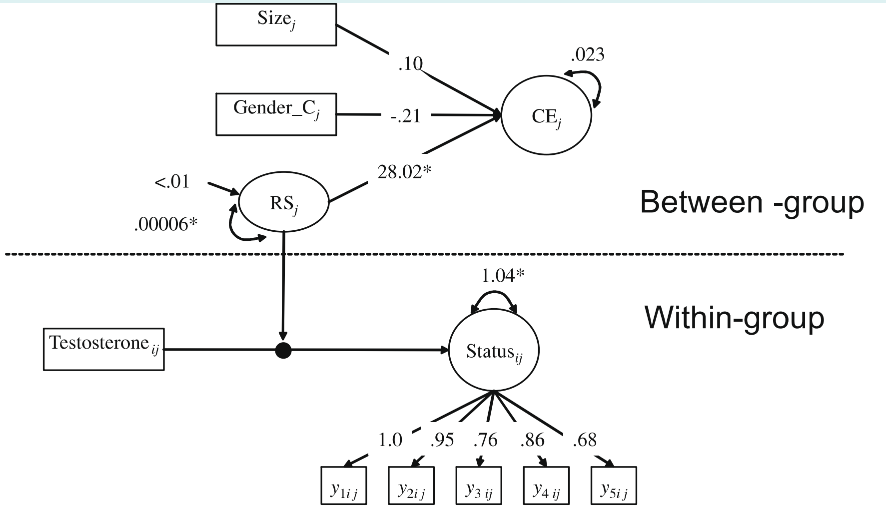


Fig. 2. Results from a multilevel structural equation model, showing unstandardized effects; $y_{1ij} - y_{5ij}$ = the five status items for each individual i in a group j; RS = random slope of the within-group T–status effect; gender_C = gender composition; CE = collective efficacy and size = group size.

Testosterone–status mismatch lowers collective efficacy in groups: Evidence from a slope-as-predictor multilevel structural equation model

Michael J. Zyphur [a,1], Jayanth Narayanan [b,1,*], Gerald Koh [c], David Koh [c]

[a] *Department of Management and Marketing, Level 10, 198 Berkeley Street, The University of Melbourne, Victoria 3010, Australia*
[b] *Department of Management & Organization, NUS Business School, National University of Singapore, 1 Business Link, Singapore 117592, Singapore*
[c] *Department of Epidemiology and Public Health, Yong Loo Lin School of Medicine, National University of Singapore, National University Health System, Block MD3, 16, Medical Drive, Singapore 117597, Singapore*

Organizational Behavior and Human Decision Processes 110 (2009) 70–79