

Hunting the Elusive Within-person and Between-person Effects in Random Coefficients Growth Models

Patrick J. Curran
University of North Carolina at Chapel Hill

Introduction

- ▶ Going to try to summarize work presented in three existing manuscripts
 - ▶ All work done in various collaborations with Dan Bauer, Taehun Lee, and Bud MacCallum
 - ▶ All models to be discussed are quite basic
 - ▶ Ultimately, issue is not statistical but conceptual
 - ▶ Raises more questions than answers
-

Why Do Longitudinal Research?

- ▶ Establish temporal precedence
 - ▶ Reduce alternative models
 - ▶ Increase statistical power
 - ▶ Increase psychometric rigor via invariance
 - ▶ Study inter-individual differences in intra-individual change
 - ▶ Less commonly articulated: explicit disaggregation of within-person & between-person effects
 - ▶ This last one might ultimately be one of most important
-

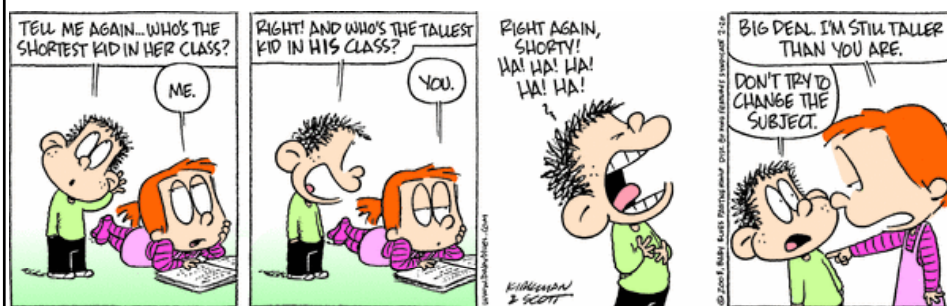
Within- and Between-Person Effects

- ▶ Sometimes disaggregation is explicit point of study
 - ▶ e.g., motivating theoretical question
 - ▶ Sometimes disaggregation is more implicit
 - ▶ e.g., time-invariant vs. time-varying covariates
 - ▶ Sometimes disaggregation is simply ignored
 - ▶ e.g., much of my own work
 - ▶ Lack of attention paid to disaggregation of effects less an error and more a lost opportunity
 - ▶ does not take full advantage of available data
 - ▶ But why is more attention not paid to this?
-

Lack of Attention to Disaggregation

- ▶ It is challenging to think about within-person and between-person effects from theoretical perspective
- ▶ Confusing with different types of between-person effects
 - ▶ between-person effect of a level-2 time-invariant covariate
 - ▶ between-person effect of a level-1 time-varying covariate
- ▶ Not widely known how to disaggregate in practice
 - ▶ some discussion in MLM -- absolutely none in SEM
- ▶ Existing methods impose rather strict assumptions
 - ▶ further analytical developments still needed
- ▶ It is helpful to orient to issue by better known disaggregation of within- and between-group effects

Within- vs. Between-Group Differences



Within- vs. Between-Group Differences

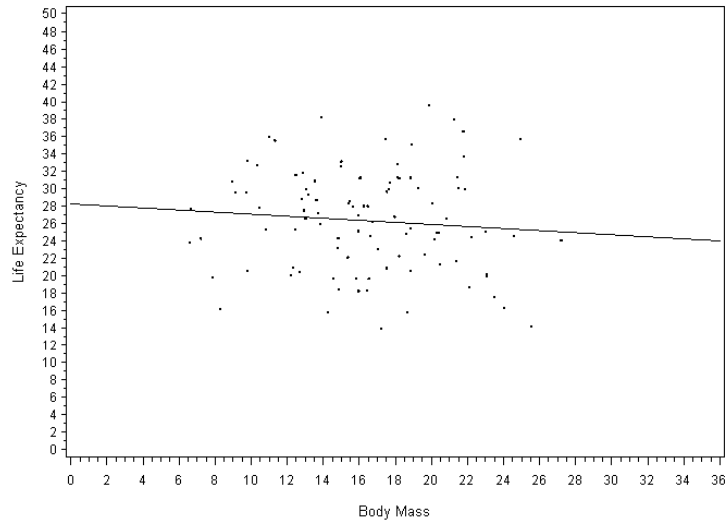
- ▶ Errors of inference have long been known
- ▶ Ecological fallacy: between-group relations do not inform about within-group relations
 - ▶ Durkheim's study of suicide & Catholicism vs. Protestantism
 - ▶ Robinson's study of illiteracy and immigrant status
- ▶ Simpson's Paradox: between-group relations can be opposite in direction from within-group relations
 - ▶ Berkeley sex bias case
 - ▶ Derek Jeter & David Justice batting averages

	1995		1996		1997		Combined	
Derek Jeter	12/48	.250	183/582	.314	190/654	.291	385/1284	.300
David Justice	104/411	.253	45/140	.321	163/495	.329	312/1046	.298

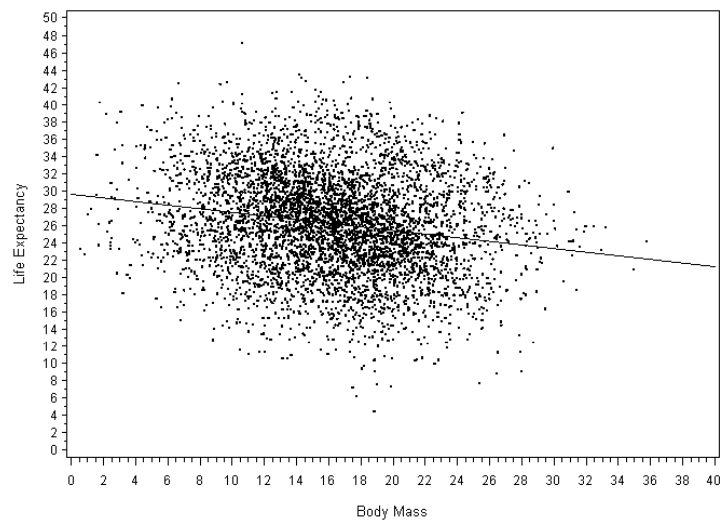
Simulated Data to Highlight Effects

- ▶ Data simulated with known population structure
 - ▶ 100 groups with 50 individuals within each group
 - ▶ single continuous outcome measure
 - ▶ single continuous level-I predictor
- ▶ Hypothetical example motivating data:
 - ▶ predictor (x) is **body mass** of individual animal
 - ▶ outcome (y) is **life expectancy** of individual animal
 - ▶ group is **species** of animal
 - ▶ Question: is body mass related to life expectancy?

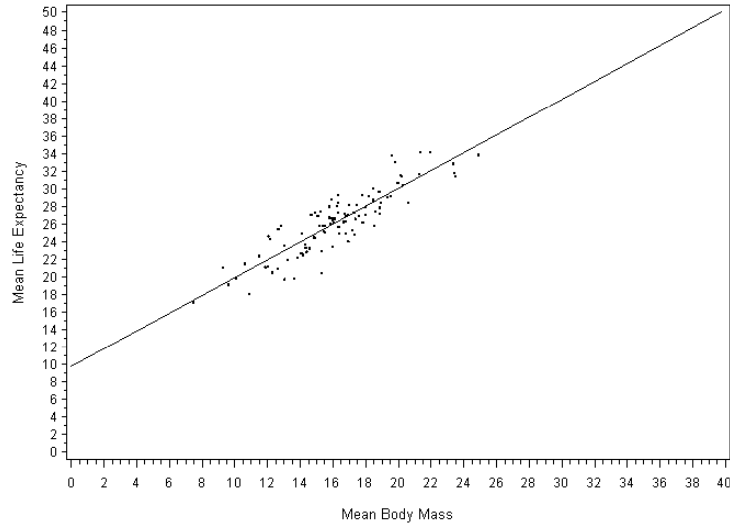
What if selected one observation from each group so sample is independent?



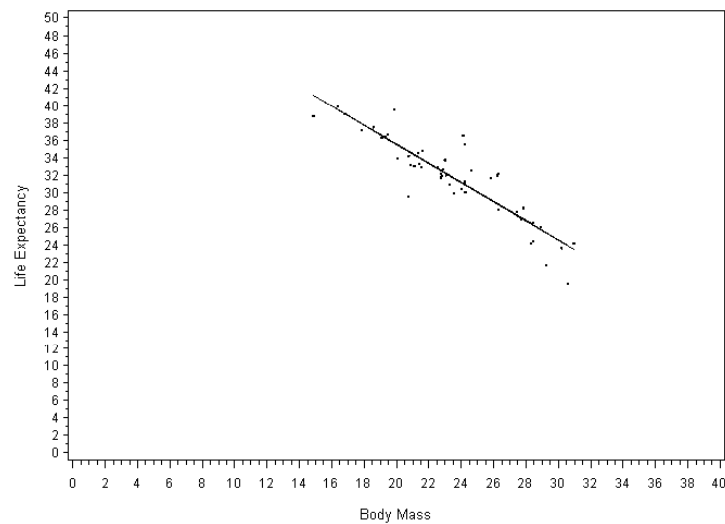
What if selected 5000 individuals in 50 groups and plotted individual scores?



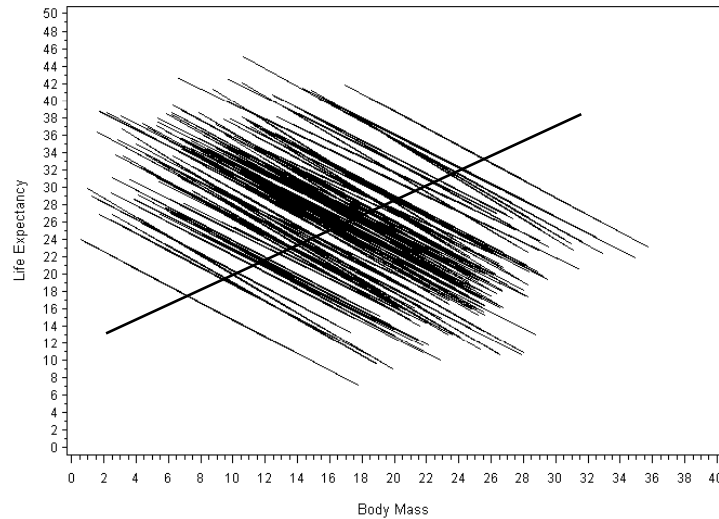
What if selected multiple individuals nested in multiple groups & plotted group means?



What if selected 50 members all within just a single group?



What if selected 50 individuals from 100 groups & plotted group regressions?



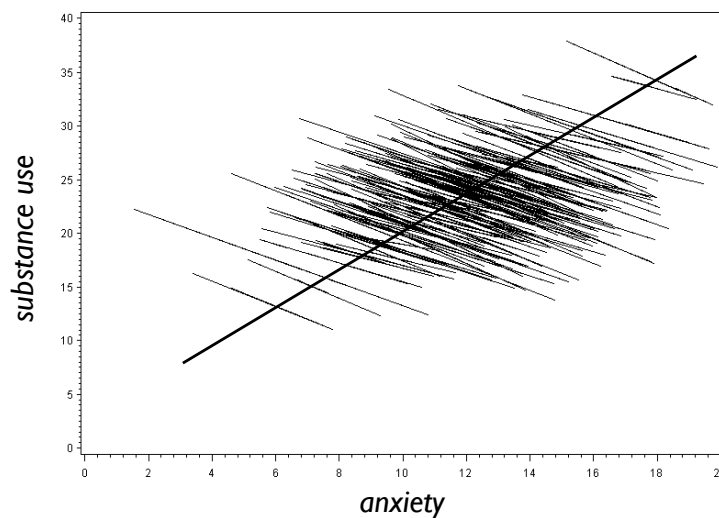
Extending to Repeated Measures

- ▶ Just as multiple individuals are nested within group, repeated measures are nested within individual
- ▶ Disaggregation of within- & between person effects precisely same as within- & between-group effects
 - ▶ two effects captured with time-varying covariates (TVCs)
- ▶ Disaggregation of effects more challenging because temporal ordering of level-I observations matter
- ▶ We would rarely ignore disaggregation in grouped data, yet limited attention paid in longitudinal data

Simulated Empirical Data

- ▶ Hypothetical example: weekly variations in anxiety and subsequent alcohol use
 - ▶ between-person effect: on average, do more anxious people drink more alcohol?
 - ▶ within-person effect: on average, do people drink less on days they are elevated on anxiety because they don't go out
- ▶ Nine repeated measures taken on 500 individuals
- ▶ Between-person effect is positive (equal to 1.5)
- ▶ Within-person effect is negative (equal to -1.0)
- ▶ Significant within- and between-person random effects

Simulated Anxiety and Substance Use



Multilevel Growth Model

$$y_{ti} = \beta_{0i} + \beta_{1i}z_{ti} + e_{ti}$$

$$\beta_{0i} = \gamma_{00} + u_{0i}$$

$$\beta_{1i} = \gamma_{10}$$

$$y_{ti} = (\gamma_{00} + \gamma_{10}z_{ti}) + (u_{0i} + e_{ti})$$

Disaggregating Effects in the MLM

$$\dot{z}_{ti} = z_{ti} - \bar{z}_i$$

$$y_{ti} = \beta_{0i} + \beta_{1i}\dot{z}_{ti} + e_{ti}$$

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\bar{z}_i + u_{0i}$$

$$\beta_{1i} = \gamma_{10}$$

$$y_{ti} = (\gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{10}\dot{z}_{ti}) + (u_{0i} + e_{ti})$$

MLM Results

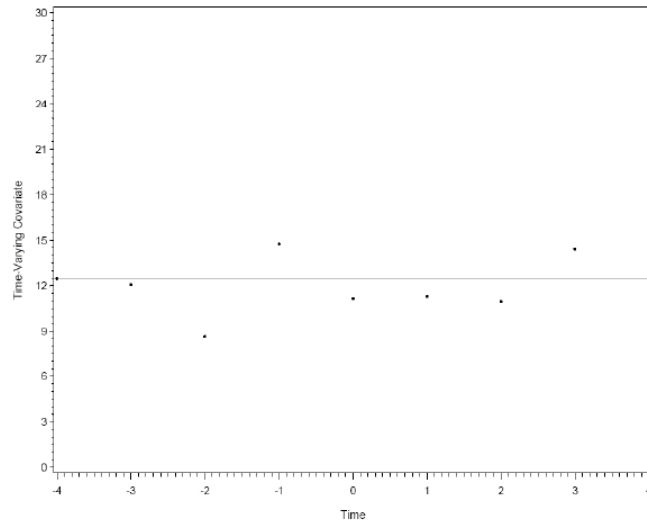
- ▶ **Population values:**
 - ▶ within-person = -1.0
 - ▶ between-person = 1.50

 - ▶ **Person mean-centered TVC at level-1 and person mean at level-2**
 - ▶ within-person = $-.99$
 - ▶ between-person = 1.51
-

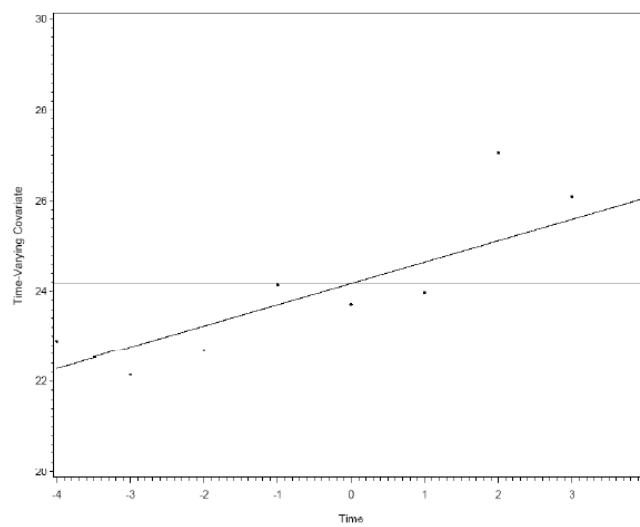
Assumptions of Multilevel TVC Model

- ▶ **TVC is not changing systematically over time**
 - ▶ allows us to deviate each time-specific measure of the TVC from the person-specific mean
 - ▶ **Person-specific mean estimated with perfect reliability**
 - ▶ allows us to take just person-specific mean without also needing to take person-specific variance
 - ▶ **How are these manifested?**
-

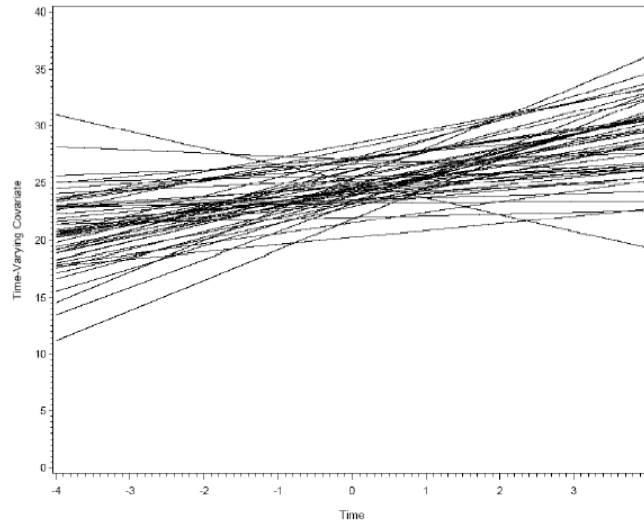
Assumes No Variability in TVC



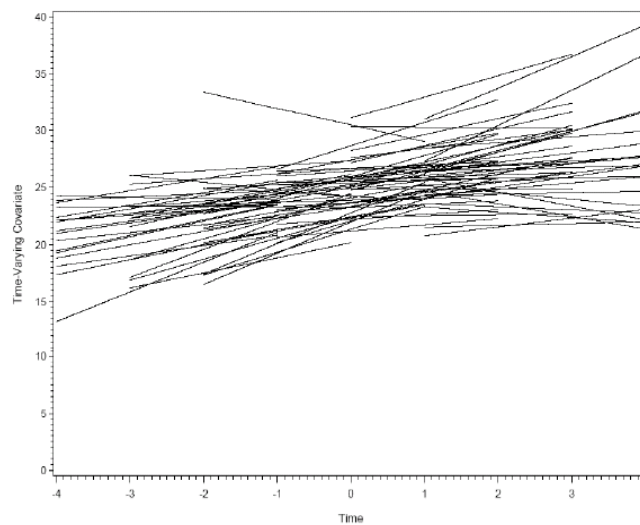
Assumes no Growth in TVC



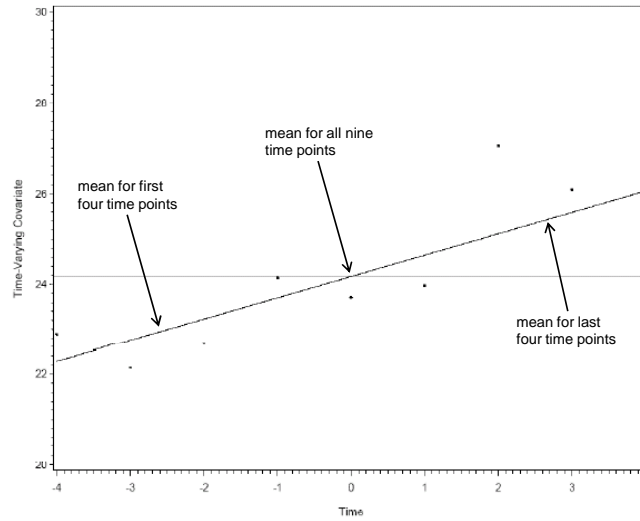
Simulated Data for Growth in TVC



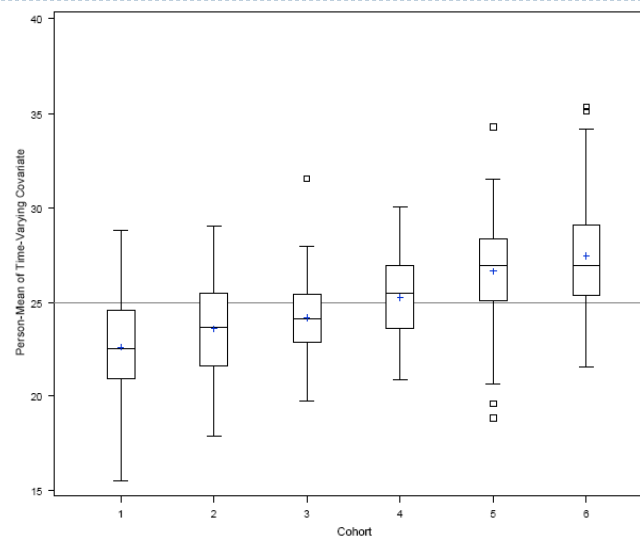
TVC Growth in Cohort-Sequential Design



Person-Means are Biased



Person-Specific Mean Depends on Cohort

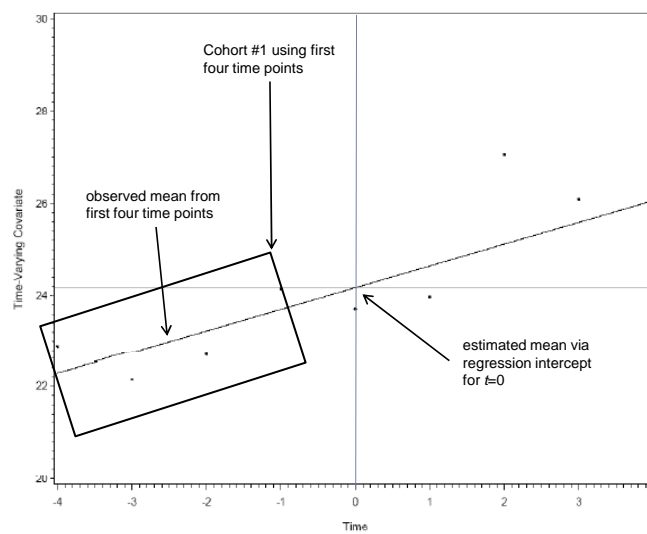


MLM Results for Growth in TVC

- ▶ Population values:
 - ▶ within-person = -1.0
 - ▶ between-person = 1.50

- ▶ Person mean-centered TVC at level-1 and person mean at level-2
 - ▶ within-person = $-.24$
 - ▶ between-person = $.71$

Detrending TVC via Individual Regressions



MLM Results from De-trended TVC

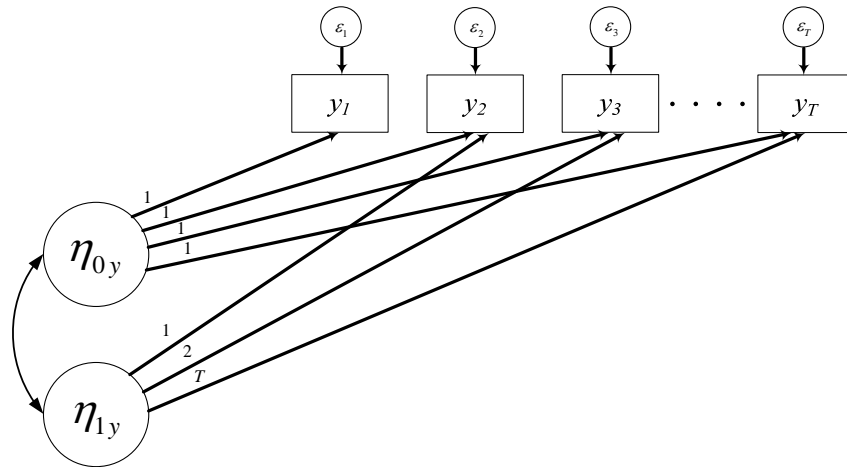
- ▶ Population values:
 - ▶ within-person = -1.0
 - ▶ between-person = 1.50

 - ▶ Trajectory-deviated TVC at level-1 and person-specific trajectory intercept at level-2
 - ▶ within-person = $-.95$ (was $-.24$ with mean centered TVC)
 - ▶ between-person = 1.25 (was $.71$ with person-mean)
-

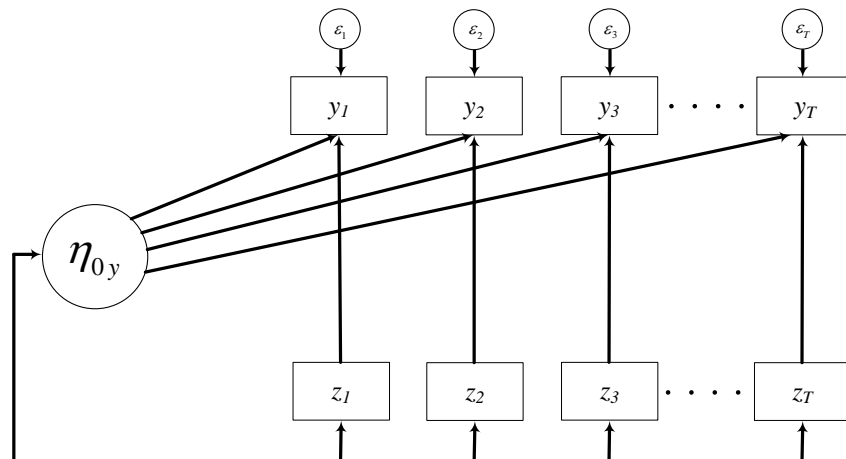
Summary of MLM

- ▶ Standard method works very well to disaggregate within- and between-person effects over time
 - ▶ But assumes perfect reliability and no trend in TVC
 - ▶ Using standard method in presence of trend is biased
 - ▶ Can deviate TVCs with respect to trajectory, but method is post hoc and has lots of sampling variability
 - ▶ No way to address unreliability in TVC over time
 - ▶ End up taking many steps to “fix” data prior to model
 - ▶ Can SEM do anything to help?
-

Unconditional LCM



LCM with Time-specific TVCs



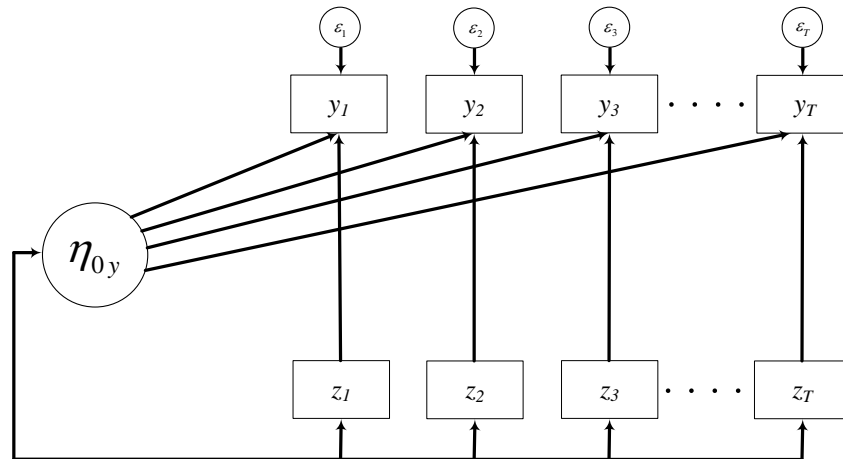
LCM with Time-specific TVCs

- ▶ Fit standard LCM to uncentered (i.e., raw-scale) TVC and resulting effect was equal to $-.99$
 - ▶ this is a pure estimate of the within-person effect
 - ▶ precisely matches MLM effect, even standard error
 - ▶ Leaves us with two weird things:
 1. Obtain within-person effect in MLM using centered TVC, but in SEM using uncentered TVC
 2. Obtain pure estimate of within-person effect in SEM, but complete omission of between-person effect
-

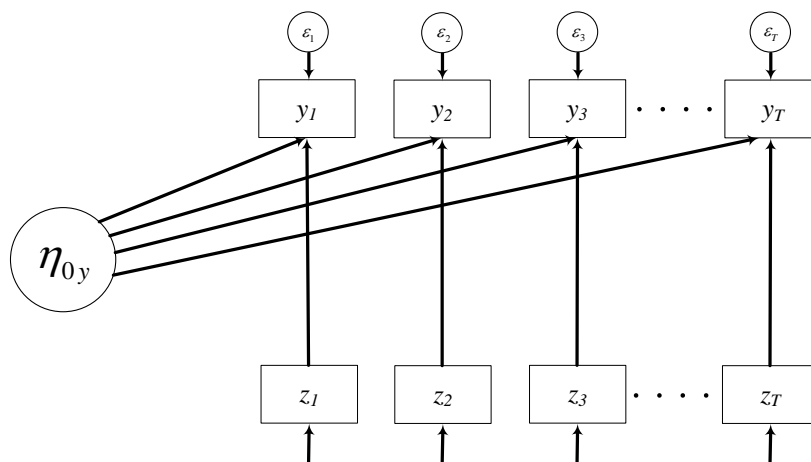
LCM with Time-specific TVCs

- ▶ In MLM, within effect obtained from centered TVC
 - ▶ In SEM, within effect obtained from uncentered TVC
 - ▶ can't even use centered TVC in SEM because ipsative and NPD
 - ▶ Why the difference?
 - ▶ MLM assumes TVCs and random intercept uncorrelated
 - ▶ SEM allows TVCs and random intercept to covary
 - ▶ Covariance between intercept & TVCs source of difference
 - ▶ if fix covariance to zero in SEM, get aggregate effect
-

Estimate of Within-person Effect



Estimate of Aggregate Effect



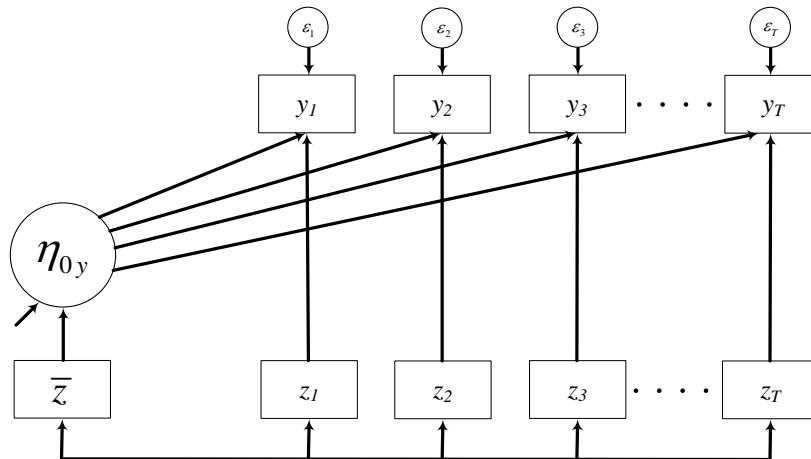
Can Derive Between-Effect in SEM

$$\hat{\delta} = \frac{T^{-1} \sum_{t=1}^T \text{cov}(z_{ti}, \zeta_{\alpha_i})}{T^{-2} \left\{ \sum_{t=1}^T \text{var}(z_{ti}) + \sum_{t \neq t'}^T \text{cov}(z_{ti}, z'_{t'i}) \right\}}$$

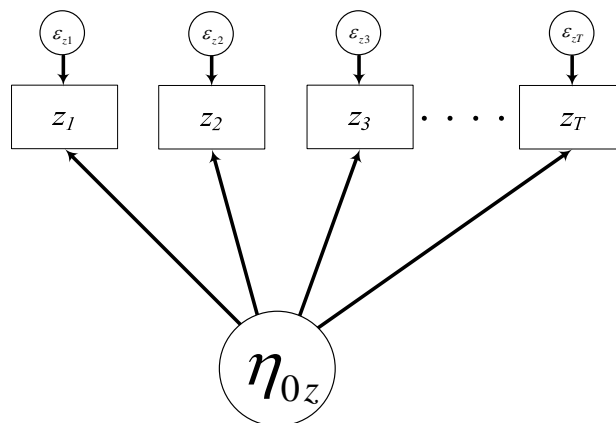
Between-Person Effect in SEM

- ▶ With standard TVC SEM, don't get any estimate of between effect
 - ▶ Can derive composite effect from covariance structure, but tedious, post hoc, and not explicit part of model
 - ▶ In MLM, just include person-specific mean as level-2 predictor -- simply use this here?
-

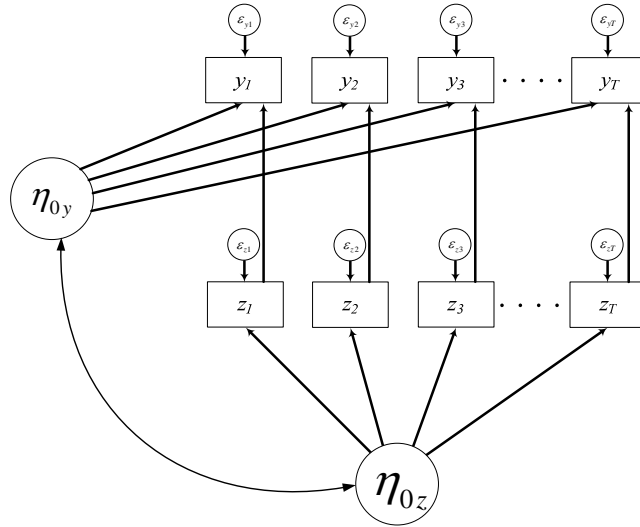
Can't Include Person Mean Because NPD



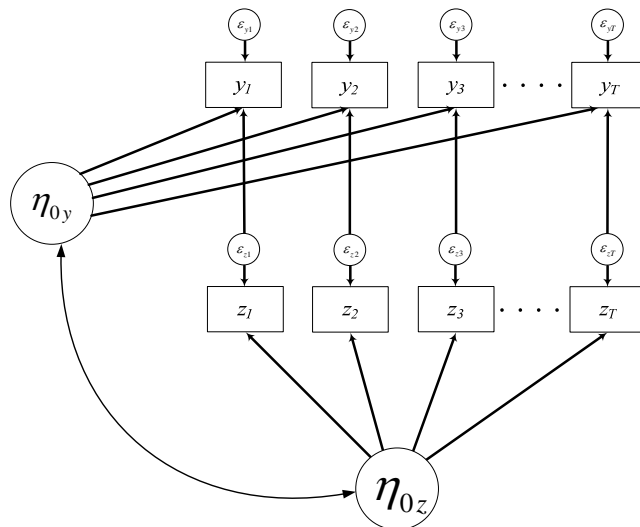
Estimating TVC Mean via the Model



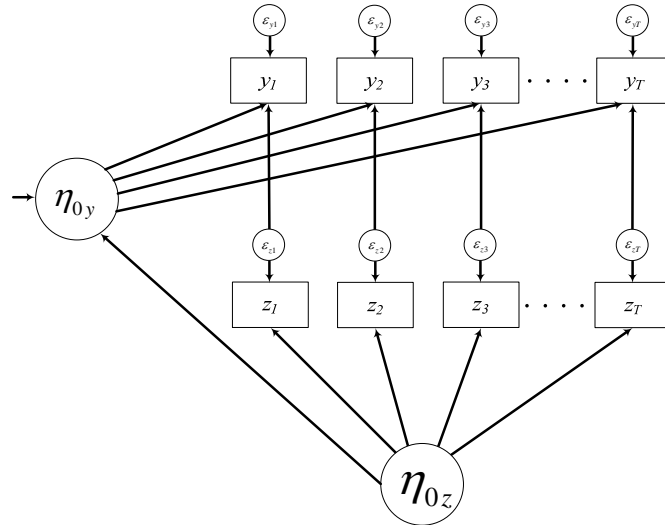
LCM with Latent TVC Intercept



LCM with Latent TVC Intercept



LCM with Latent TVC Intercept



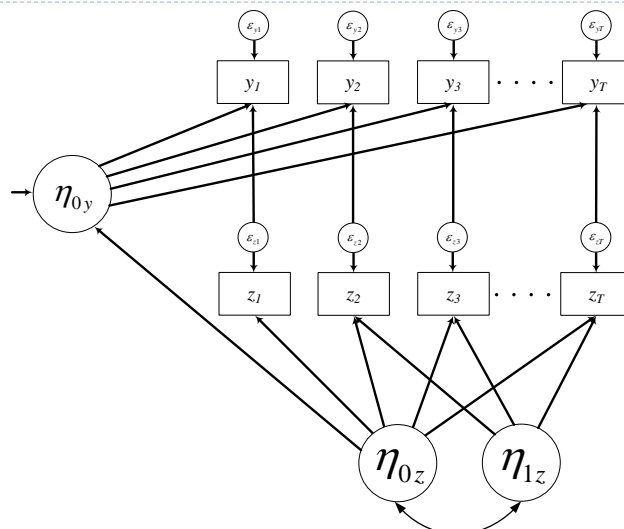
LCM with Latent TVC Intercept

- ▶ Fitted latent TVC SEM to data with no time trend in TVC
- ▶ Population values:
 - ▶ between: 1.5
 - ▶ within: -1.0
- ▶ MLM results (based on person-mean centered TVC):
 - ▶ between: 1.51
 - ▶ within: -.99
- ▶ SEM results
 - ▶ between: 1.78
 - ▶ within: -.99
- ▶ Why between effect higher? Because latent TVC factor accounting for within-person variability in TVC over time

LCM with TVC that Changes with Time

- ▶ In the MLM the person-mean centered approach assumes TVC unrelated to passage of time
- ▶ We addressed this by estimating an OLS estimate of each trajectory and deviated the TVC relative to trajectory
 - ▶ high sampling variability in individual OLS regressions
 - ▶ doesn't account for unreliability of TVC over time
 - ▶ let's be honest: pretty ugly data management solution
- ▶ But we can obtain person-mean of TVC in SEM via parameterization of model
- ▶ Can we also account for time trend via parameterization?

LCM with Latent TVC Intercept & Slope



LCM with Latent TVC Intercept & Slope

- ▶ Fitted latent TVC SEM to data with time trend in TVC
 - ▶ Population values:
 - ▶ between: 1.5
 - ▶ within: -1.0
 - ▶ MLM results (based on de-trended TVC):
 - ▶ between = 1.25
 - ▶ within = -.95
 - ▶ SEM results (with random intercept & slope for TVC)
 - ▶ between: 1.61
 - ▶ within: -.95
-

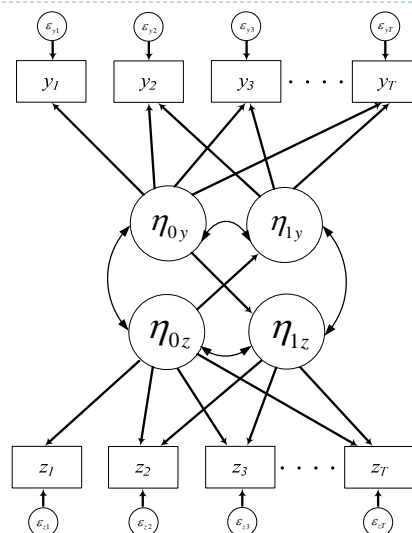
Summary Thus Far

- ▶ Standard methods work well in MLM when assumptions related to TVC are met
 - ▶ can modify standard methods in presence of growth
 - ▶ can't modify standard methods in presence of unreliability
 - ▶ Can't use standard MLM methods in SEM
 - ▶ person-mean centered TVCs are ipsative
 - ▶ person-mean is collinear with TVCs
 - ▶ In SEM, obtain pure estimate of within-effect based on uncentered TVC & within-effect based on latent mean
 - ▶ Can expand to include growth factor for trend, and includes information about variability in TVC
 - ▶ But many unresolved issues....
-

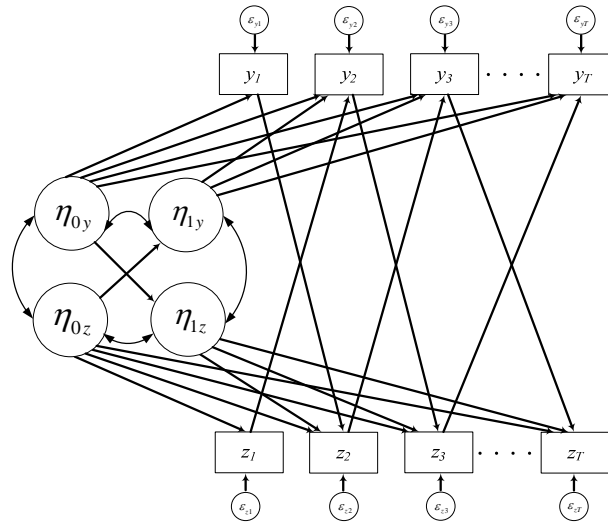
Conditional Between-Person Effect

- ▶ Usually think about “the” between-person effect
- ▶ But with growth in TVC, the between-person effect is now conditional on time
- ▶ With no growth in TVC, intercept is constant and between-person effect is constant
- ▶ With growth in TVC, intercept defined where time = 0 and between-person is not constant
- ▶ We must think much more carefully about what this implies substantively

Implications for Multivariate LCM



Implications for Multivariate ALT



Other Unresolved Issues

- ▶ What if TVC is binary?
- ▶ What if more than one TVC? What if they interact?
- ▶ How know if not over-parameterizing model?
- ▶ How handle bi-directional effects between TVC and DV?
- ▶ What is role of stability parameter in ALT model?
- ▶ How know if not just absorbing misspecification of growth model in time-specific relations?
- ▶ How best estimate possible interaction between within-person and between-person effect in MLM or SEM?
- ▶ How best estimate random effects for within-person effect within the SEM?

Walk Away Point

- ▶ Core issue is less statistical and more theoretical
- ▶ What is common denominator to all of these problems?

How do we maximize the correspondence between the substantive model and the statistical model?

- ▶ If assess two or more constructs at two or more time points, must at least consider the disaggregation of within- and between-person effects
 - ▶ The statistical models are simple
 - ▶ We must refine statistical models to best test theory, but we must refine theories to better explicate relations
-