SEM IV:
Fixed Effects, Random Effects and Dynamic Structural Equation Models

PS2701-2019

Longitudinal Analysis

Week 8

Professor Steven Finkel

Unobserved Heterogeneity in the SEM Framework

- What is missing from the SEM models we've considered so far? Unobserved heterogeneity!! That is, no "U" term representing stable omitted unit-level variables
- But you **can** incorporate unobserved heterogeneity within the SEM framework. If you have enough waves of observation, you can add the "U" term as an **unobserved latent variable** with no indicators!!!!
- This is essentially what U_i is – an unobserved latent variable standing in for all stable factors that have not been included in the observed data. You then model this unobserved variable as influencing all the X and Y over time to capture the possible spuriousness induced by the unobservables. See the next slides, Finkel, chapter 5, or, more recently:

  - Teachman, Jay, Greg Duncan, W. Jean Yeung, and Dan Levy. 2001. "Covariance Structure Models for Fixed and Random Effects." *Sociological Methods and Research 30: 271-288, or*
  - Dormann, Christian. 2001. "Modeling Unmeasured Third Variables in Longitudinal Studies." *Structural Equation Modeling 8(4): 575-598.*
  - Allison, Paul. 2009. *Fixed Effects Regression Models*. Sage Publications. (Chapter 6).
  - Bollen, Kenneth and Jennie Brand. 2010. "A General Panel Model with Fixed and Random Effects: A Structural Equations Approach", *Social Forces* 89(1):1-34.

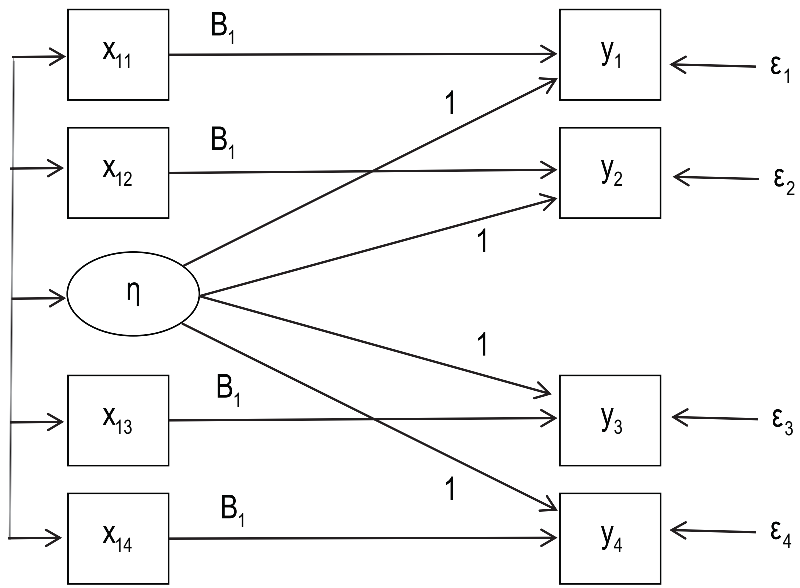
- Advantages of the SEM approach
  - Explicit wave by wave specification as opposed to pooling, possible changes in parameter values in different waves (and importantly, this can include U as well!)
  - No loss of data through differencing or use of lags
  - Explicit specification of all aspects of the model (as opposed to transforming, differencing, eliminating this term or that term!)
  - Ability to assess overall model fit and test nested models with difference in chi-square and GOF measures
  - Ability to incorporate measurement error via Wiley-Wiley or multiple indicator models
- Disadvantages:
  - difficulty for some models to converge quickly or at all
  - Cumbersome to specify and model as number of panel waves increases

An SEM Version of the FE Model (Bollen and Brand 2012)

Figure 2. Classic Fixed Effects Model in Path Diagram

$$y_{it} = \mathbf{B}_{yx}\mathbf{x}_{it} + \eta_i + \varepsilon_{it}$$

$$E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$$



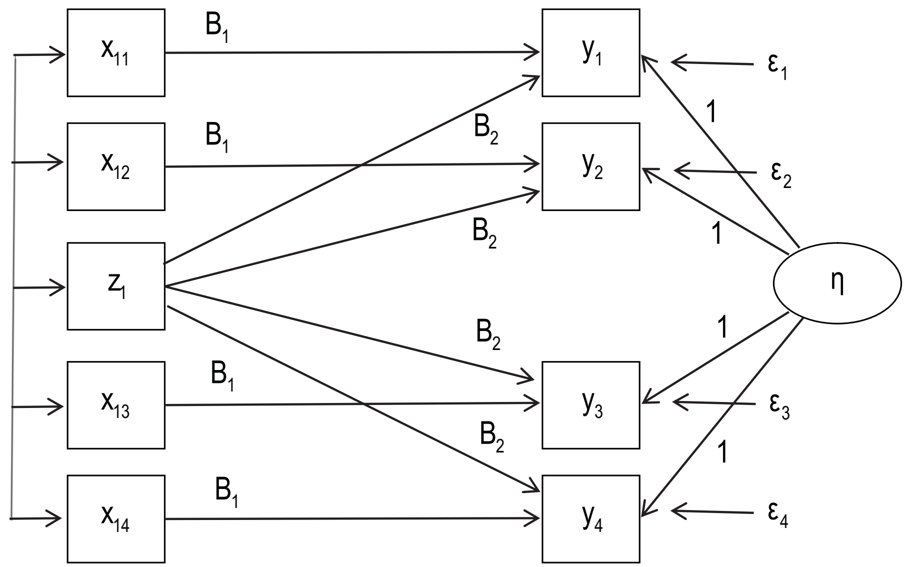
In this model, η is equivalent to U_i, the set of stable unobservables, and the model includes causal effects from η to the Y outcomes at each wave, and free covariances with all X independent variables. This is the three-wave SEM version of the basic unobserved heterogeneity model, allowing for correlation between η(or U_i) and the Xs

An SEM Version of the RE Model (Bollen and Brand 2010)

Figure 1. Classic Random Effects Model in Path Diagram

$$y_{it} = \mathbf{B}_{yx}\mathbf{x}_{it} + \mathbf{B}_{yz}\mathbf{z}_{i} + \eta_{i} + \varepsilon_{it}$$

$$E(\varepsilon_{it}^{2}) = \sigma_{\varepsilon}^{2}$$



- This model is equivalent to the FE model, with the additional constraint that the covariances between η(or U_i) and the Xs are 0. This is easily programmed in Stata SEM.
- This also shows that, from the SEM point of view, the RE model is "nested" within the FE model, and the improvement in fit of the FE model over RE can be assessed via a chi-square difference test
- The model includes time-invariant Z variable as an additional covariate – one of the major advantages of RE. But the Z cannot covary with η, nor can the Xs (the major disadvantage of RE).
- **IMPORTANT NOTE**: We **cannot** add the cluster mean (X-bar) as an additional variable to arrive at the RE-Hybrid model in the traditional SEM set-up. Why? The set of X_i and their respective cluster means are *"ipsative"*, or perfectly dependent (remember that that, within clusters, we lose one df in calculating the cluster mean). So estimation of both "within" and "between" effects is not (yet) possible in the same SEM model – we'll consider this in more detail later

- Bollen and Brand (2012) discuss *many* more possibilities for including time-varying causal effects of time-varying Xs, time-varying causal effects of time-invariant Zs, etc. within the SEM framework

Table 1: General Panel Model and Special Cases

General Panel Model

$y_{it} = \mathbf{B}_{yxt}\,\mathbf{x}_{it} + \mathbf{B}_{yzt}\,\mathbf{z}_i + \lambda_t \eta_i + \varepsilon_{it}$

(1) Time-invariant λ

$\lambda_t = 1$ for all t

(2) Time-invariant λ & $\mathbf{B}_{yz}$

$\lambda_t = 1$, $\mathbf{B}_{yzt} = \mathbf{B}_{yz}$ for all t

(3) Time-invariant λ, $\mathbf{B}_{yz}$, & $\mathbf{B}_{yx}$

$\lambda_t = 1$, $\mathbf{B}_{yxt} = \mathbf{B}_{yx}$, $\mathbf{B}_{yzt} = \mathbf{B}_{yz}$ for all t

(4) Time-invariant λ, $\mathbf{B}_{yz}$, $\mathbf{B}_{yx}$, & COV($\mathbf{x}_{it}$, η_i) = 0

$\lambda_t = 1$, $\mathbf{B}_{yxt} = \mathbf{B}_{yx}$, $\mathbf{B}_{yzt} = \mathbf{B}_{yz}$, COV($\mathbf{x}_{it}$, η_i) = 0 for all t

(5) (a) Fail to reject (4)

time-invariant λ, $\mathbf{B}_{yz}$, $\mathbf{B}_{yx}$, COV($\mathbf{x}_{it}$, η_i) = 0, & σ_ε

$\lambda_t = 1$, $\mathbf{B}_{yxt} = \mathbf{B}_{yx}$, $\mathbf{B}_{yzt} = \mathbf{B}_{yz}$, COV($\mathbf{x}_{it}$, η_i) = 0, $\sigma_{\varepsilon t} = \sigma_\varepsilon$ for all t

(5) (b) Reject (4)

time-invariant λ, $\mathbf{B}_{yz}$, $\mathbf{B}_{yx}$, & σ_ε

$\lambda_t = 1$, $\mathbf{B}_{yxt} = \mathbf{B}_{yx}$, $\mathbf{B}_{yzt} = \mathbf{B}_{yz}$, $\sigma_{\varepsilon t} = \sigma_\varepsilon$ for all t

(6) Classic Random Effects Model (REM)

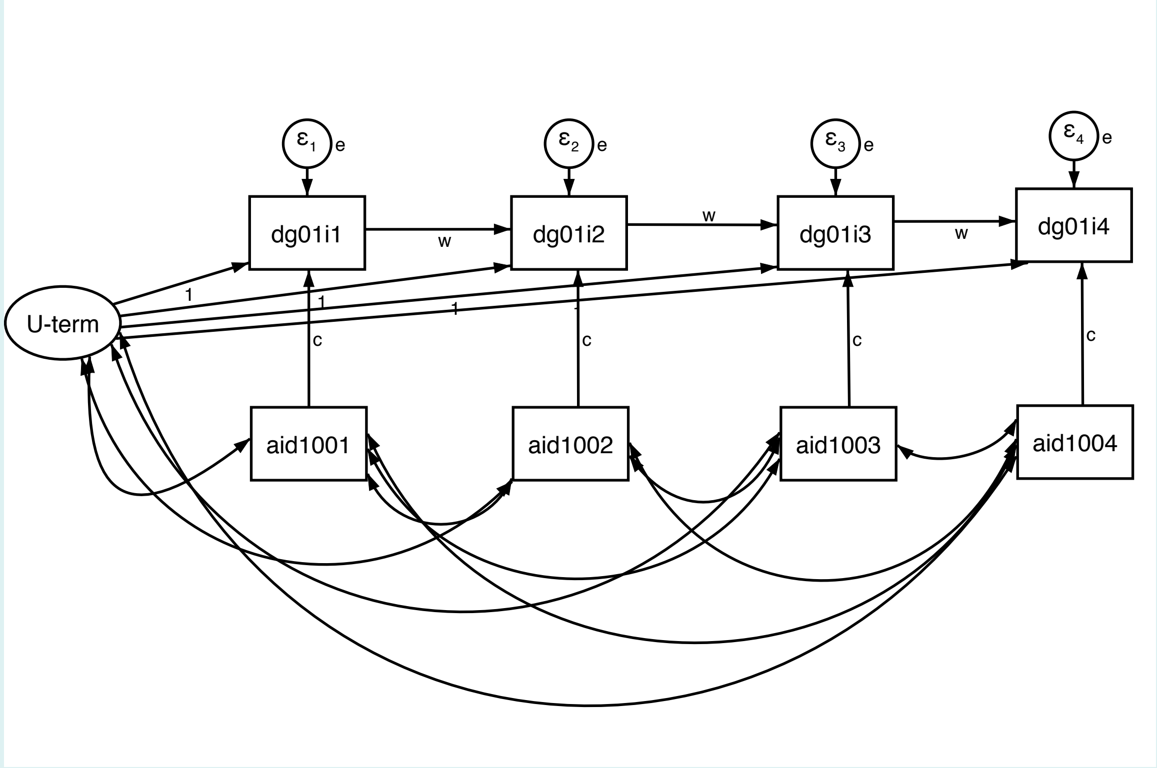
Equivalent to (5) (a)

(7) Classic Fixed Effects Model (FEM)

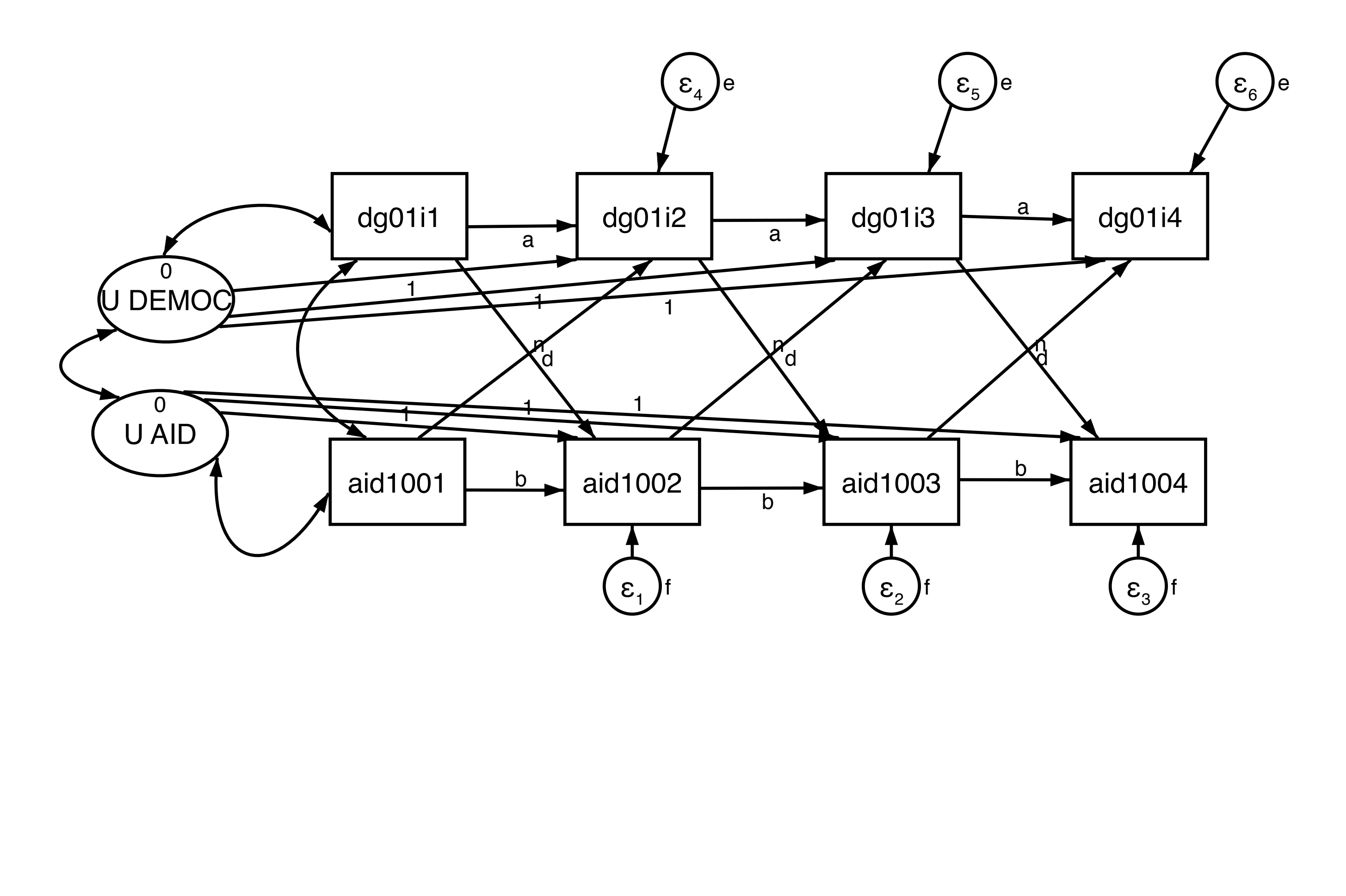
Equivalent to (5) (b) with $\mathbf{B}_{yz} = 0$ (no $\mathbf{z}_i$ in equation)

- **One important extension: allow the effects of the U-term to vary over time within the FE or Hybrid Model framework. This allows the unmeasured variable(s) to affect not only the level (intercept) of Y, but also the rate of change in Y, so it provides a partial control for the "selection-maturation" threat to causal inference discussed in Session 1. This is huge!!!**
- This is an active area of current research. Allison *et al.* (2017), e.g., argue via Monte-Carlo simulations that SEM is superior to Arellano-Bond in terms of efficiency, which makes sense considering the loss of information involved in the differencing and lag specification of A-B
- Allison argues (and I agree!) that **every econometric panel model can be estimated within the SEM framework**. This framework, with enough waves and enough information, can accommodate models with the lagged DV, measurement error, reciprocal effects, and thus be able to exploit all the other advantages of the SEM approach
- But programming can be tricky, models sometimes do not converge

The SEM Dynamic Panel Model: Unidirectional Causality



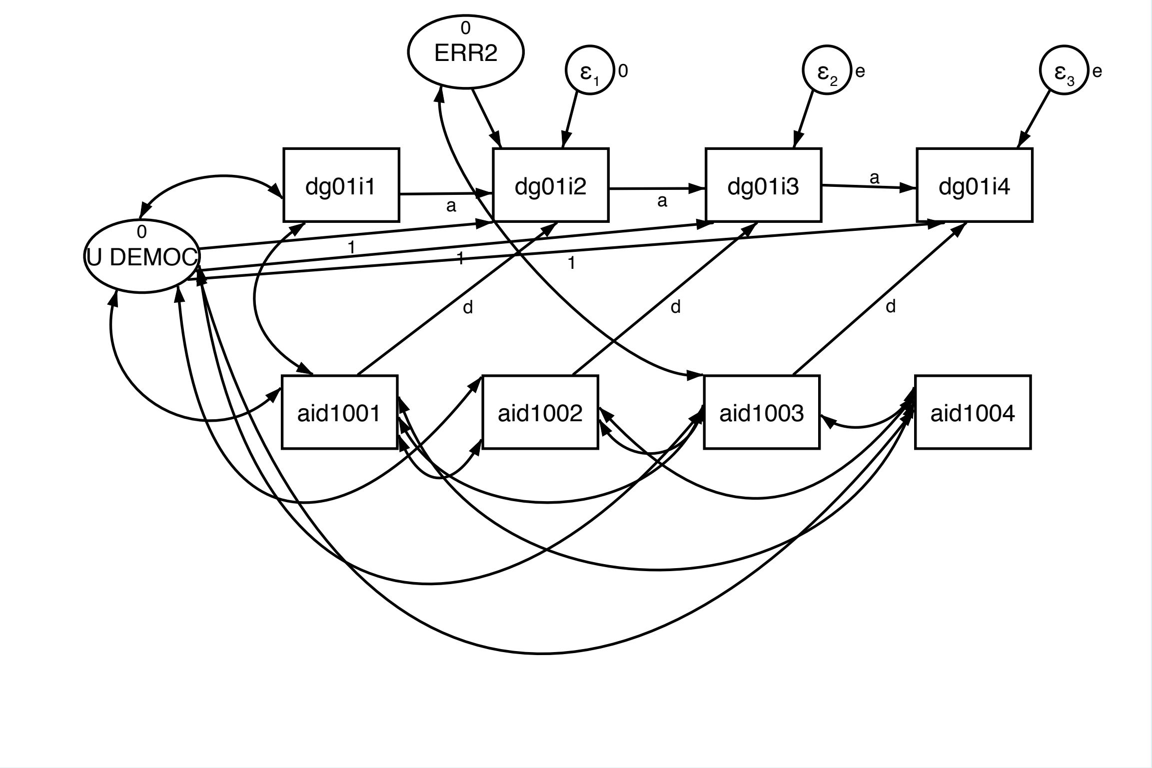
Dynamic Panel Model with Lagged Reciprocal Effects



Lagged Effects, Each variable with its own U-term

- Allison (2009; 2017) suggests alternative procedure, estimating the model “one dependent variable at a time”
  – Estimating both directions creates complex model with two U-terms that need to correlate, difficult to estimate
  – Separate models allows one variable to be quantitative, another to be categorical.
- If estimate “one by one”, must allow for possibility of reciprocal causation by modeling X (Y) as “sequentially exogenous”, following the logic of the econometric exposition earlier. This means that X_t (Y_t) must be allowed to correlate with all of the Y (X) error terms from previous time points
- This presents a (surmountable) programming challenge in Stata, since error terms cannot have effects on other variables
  – Trick: Create a latent variable to stands in for the “real” error term, and set the variance of the “real” error term to be zero

Dynamic Panel Model with Lagged Effects under Sequential Exogeneity



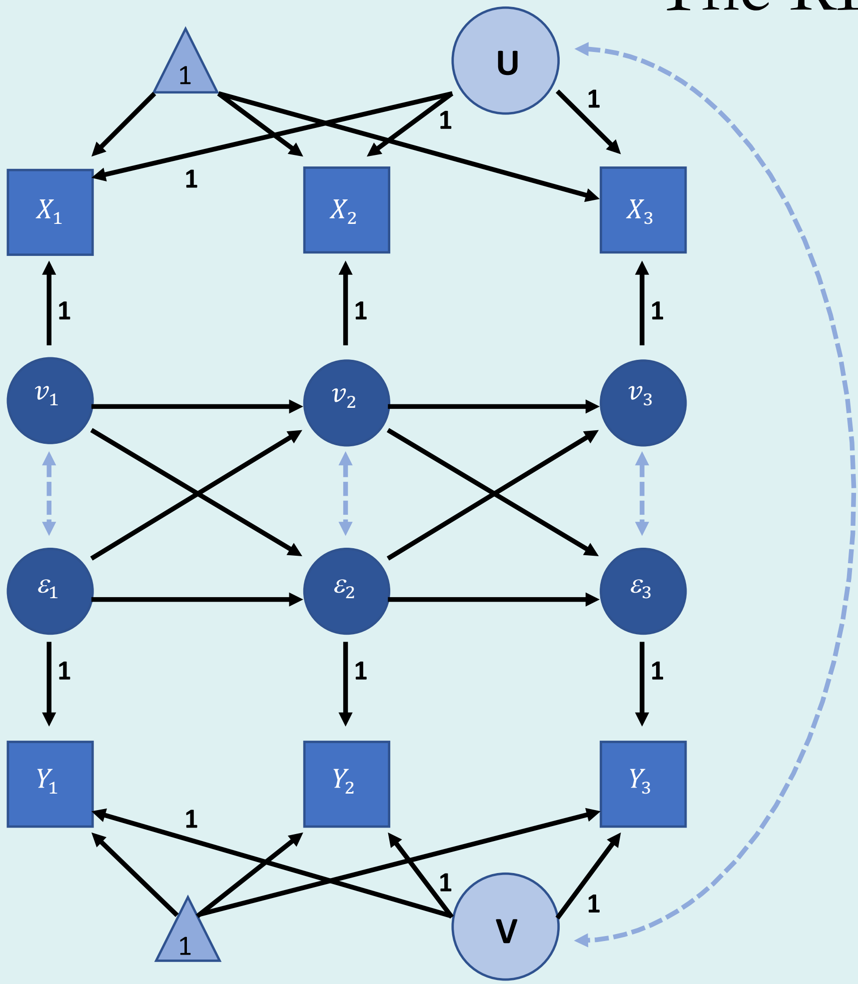
Note: Given sequential exogeneity, the error terms for the "dg01" variables are correlated with future values of "aid100". The only correlation that has any consequences for the analysis in four wave data is the one between dg01i2 (ERR2) and AID1003. See Allison, *Fixed Effects Regression Models*, pp. 93-96, and Allison *et al.* (2017).

The Random Effects-Cross-Lagged Panel Model

- Hamaker *et al.* (2015) criticize both traditional cross-lagged panel models *and* the dynamic panel model
  - Traditional CL models do not incorporate unit-level heterogeneity on either the "dependent" or "independent" variable sides of a given equation. So for predicting the dependent variable we do not control for stable, unit-level factors that influence all outcomes over time and may be confounded with the independent variables in the model.
  - Including the "U" term for the dependent variable and correlating this term with the independent variables at all times (as in the diagram on the previous slide) allows the estimation of the "within" effect, but does not provide any information on the "between" effect of the independent variable
  - As noted earlier, it is impossible to include the cluster mean as an additional independent variable to obtain a "hybrid model-type" between-effect estimate in the traditional SEM set-up
- Importantly, these criticisms also apply to the **autoregressive** effects present in the cross-lagged and dynamic panel models. The estimated effects are an "uninterpretable mixture" of within and between effects

- Solution: separate the "within" processes as operating at the level of the time-specific *residuals* of each observation
- That is, we specify the time-specific residuals as the "mean-deviated" forms of each variable over time, with the unit-level heterogeneity term representing the cluster "averages" of the independent and dependent variables that vary **between** units but not **within** units over time
- We then model the effects of deviations from unit-level means at one point in time on the deviations from its own unit-level mean at subsequent points in time (to capture the "within" level autoregressive effects) **and** on the deviation from the unit-level mean on the "opposite" variable (to capture the "within" level cross-lagged effects)
- The relationship between the unit-level means of the two variables captures the "between level" correlation
- Note this unit-level mean is a *latent* variable on both sides of the relationship and thus estimates the "average" value purged of measurement error – a real bonus for this model!

The RE-CLPM Model



Note: Difficult in Stata to estimate this model (as of 2019)
R code in Flourney, "A Better Cross-Lagged Model"
https://jflournoy.github.io/2017/10/20/riclpm-lavaan-demo/