Unit 3:
Multilevel Panel and Growth Models
III. Multilevel Analysis of Repeated Cross-Sectional Data (with Different Level 1 Units)

PS2701-2019

Longitudinal Analysis

Week 11

Professor Steven Finkel

Types of Multilevel Repeated Cross-Sectional Data

- Panel Data: Same Level 2 Cross-Sectional Units Observed at Multiple Level 1 Time Points
- Repeated Cross-Sectional Data
  - Two Level Structure: Different Level 1 Cross-Sectional Units at Different Level 2 Time Points (Example: ANES surveys at different election years)
  - Three Level Structure: Different Level 1 Cross-Sectional Units Nested within Different Level 2 Time Points Nested within Different Level 3 Units

    (Example: LAPOP/World Values surveys of different individuals at different points in time for multiple countries
- Straightforward application of multilevel framework covered thus far

Two-Level Structure

$$(1) \qquad Y_{it} = \beta_{0t} + \beta_{1t} X_{it} + \varepsilon_{it}$$

- At Level 1, the value of Y for a unit *i* at a given point in time *t* is a function of the overall intercept for the sample at time *t*, plus a time-specific regression coefficient multiplied by the value of an X for that unit at that time, plus an idiosyncratic error
- There is a time-specific intercept affecting all *i* and a time-specific slope associated with the IVs
- Goal is to explain why the intercept and slope may change over time, due to factors that change over time that are common to all units

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

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

- Level 2: The intercept and slope depend on X_1, which varies over time but is constant within units at a given point in time, plus random Level 2 error terms
- Example: Affective Party Polarization among individuals ($\mathbf{Y_{it}}$) is dependent on Media Exposure ($\mathbf{X_{it}}$) at Level 1; the overall level of party polarization depends on a population value(β_{00}) plus whether it is a presidential election year or not ($\mathbf{X_{1t}}$); the effect of media exposure on party polarization depends on a population value (β_{10}) plus whether it is a presidential election year or not (X1); there is random variation in the intercept and slope over time
- Could have a series of time dummies to make it a "fixed effects" time model instead

Two-Level Structure (Cross-Sectional Version)

- Exactly the same framework would apply to a fully cross-sectional model, with different Level 1 units nested within different Level 2 units at *one* point in time

$$(3) \qquad Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}$$

- At Level 1, the value of Y for a unit *i* in a given higher-level unit *j* (say, "country") is a function of the overall intercept for the country *j*, plus a country-specific regression coefficient multiplied by the value of an X for that unit (in that country), plus an idiosyncratic error
- There is a *j* country-specific intercept affecting all *i* and a *j* country-specific slope associated with the IV
- Goal is to explain why the intercept and slope may vary across *j* country units due to factors that vary over countries but are common to all units within a country. MUCH CP work is done this way!

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

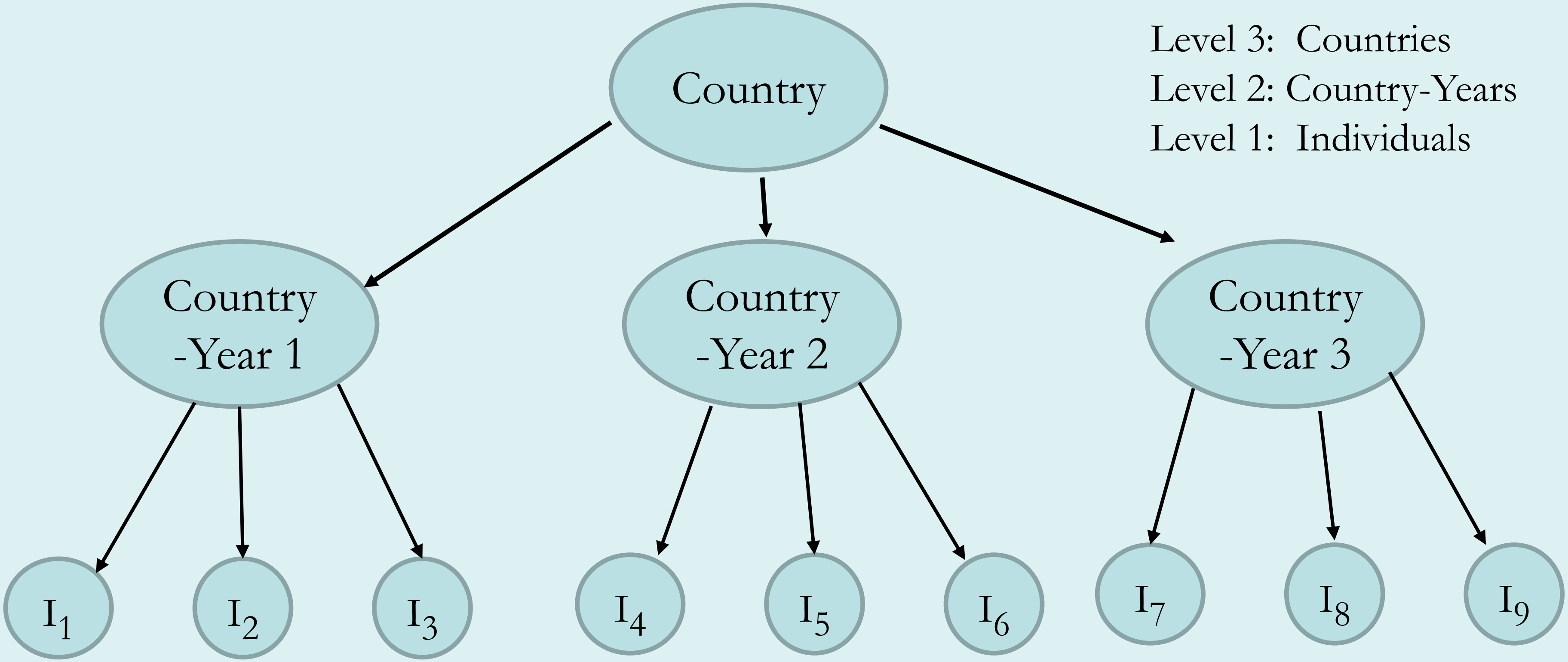
$$(b) \quad \beta_{1j} = \beta_{10} + \beta_{11} X_{1j} + \zeta_{1j}$$

- Level 2: The intercept and slope depend on X_1, which varies across countries but is constant within units in the same country, plus random Level 2 error terms
- Example: Incumbent Voting among individuals in a given country ($\mathbf{Y_{it}}$) is dependent on their perceptions of the economy ($\mathbf{X_{it}}$) at Level 1; the overall level of incumbent voting depends on a population value(β_{00}) plus whether the country has "clarity of responsibility" in its institutional structure($\mathbf{X_{1j}}$); the effect of economic perceptions on incumbent voting depends on a population value (β_{10}) plus the "clarity of responsibility" in its institutional structure; there is random intercept and slope variation across countries. See classic article by Powell and Whitten (1993).
- Could have a series of country dummies to make it a "fixed effects" country model instead

Three-Level Framework

- What happens in the (fortunate) event that one has the *longitudinal* version of this data structure, such that there are different individual units at Level 1, which are nested within higher-level units observed at *multiple points* in time?
- Examples:
  - LAPOP/World Values Surveys – multiple waves of different individuals in the same countries at different points in time.
  - Aggregating individual ANES responses by US state, so different individuals in the 50(-ish) states at different points in time.
  - Election results for different parties in the same country at different points in time
- How can we (should we) exploit the *longitudinal* component of this data structure?

First: Get the Nesting Structure and Random Effects Right!

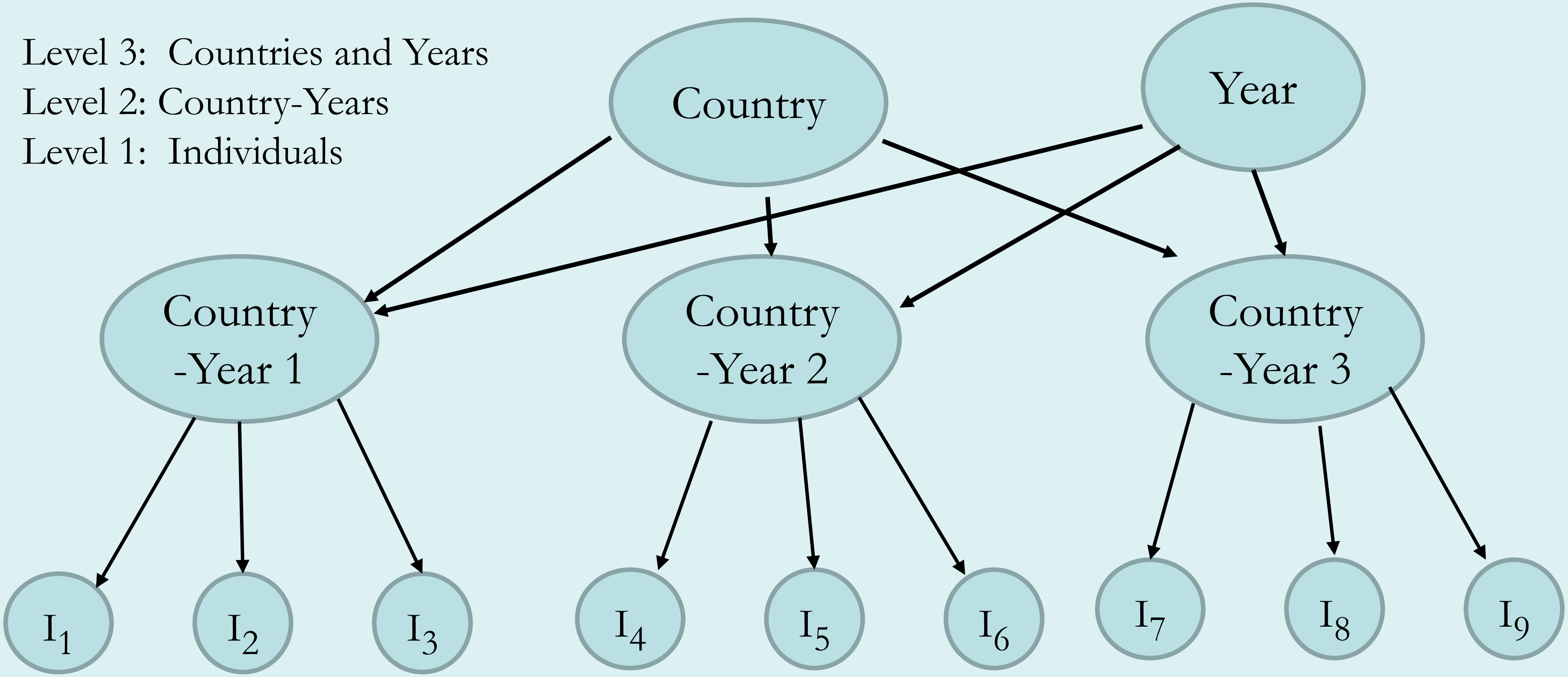


Level 3: Countries
Level 2: Country-Years
Level 1: Individuals

Note: Different individuals at Level 1 within each country-year, so **NOT PANEL DATA**

Actually: "Crossed-Effects" with Time at Level 3

Level 3: Countries and Years
Level 2: Country-Years
Level 1: Individuals



Country-Years Nested within BOTH Country and Year: Germany/Brazil/Turkey in 1990, 1995 and 2000 are Nested within "Germany"/"Brazil"/"Turkey"; and Germany in 1995, Brazil in 1995, and Turkey in 1995 are Nested within "1995", Germany/Brazil/Turkey in 2000 are nested within "2000", etc.. **With Small-ish T we ignore this and use time dummies as fixed effects**

Where Should We Include Random Effects?

- General Rule : Include random effects wherever clustering exists in the data (e.g., whenever observations at a lower level are not independent at a given higher level)

$$(4a) \quad Y_{itj} = \beta_{0tj} + \beta_1 X_{itj} + \varepsilon_{itj} \qquad \text{"Level 1"}$$

$$(4b) \quad \beta_{0tj} = \beta_{00j} + \varsigma_{tj} \qquad \text{"Level 2"}$$

$$(4c) \quad \beta_{00j} = \beta_{000} + \varsigma_{j} \qquad \text{"Level 3"}$$

$$(4d) \quad Y_{itj} = \beta_{000} + \beta_1 X_{itj} + \varsigma_{tj} + \varsigma_{j} + \varepsilon_{itj} \qquad \text{"Mixed"}$$

- So: Random Effects for Country-Year(ς_{tj}) and Country (ς_{j})
- Note: Slope of X assumed to be fixed across countries and time (this can be relaxed)
- Note: Time Dummies included but not shown

Extending the Model: Time-Varying Contextual Effects

- Model can be very useful for assessing the impact of contextual factors at Level 3 that vary over time (at Level 2)
  - For example, for country (Level 3) analyses: does individual protest behavior at Level 1 depend on the state of a country's economy, which varies at Level 2?
  - Does social trust among individuals at Level 1 depend on a country's number of immigrants, which varies at Level 2?
- Time-varying contextual factors at Level 2 can also be the country-year aggregation of Level 1 variables
  - For example, social trust among individuals at Level 1 may depend on trust in government at Level 1, **and** also depend on the average level of trust in government in a country at Level 2.
  - So *individuals* who distrust government may be more likely to distrust others, and *country-years* where there is a high level of distrust in government may also have high levels of social distrust, regardless of individual gov't distrust
- We are moving toward separating "within" and "between" effects!

- Including a Country-Year Variable at Level 2

(5a) $Y_{itj} = \beta_{0tj} + \beta_1 X_{1itj} + \varepsilon_{itj}$ "Level 1"

(5b) $\beta_{0tj} = \beta_{00j} + \beta_2 X_{2tj} + \varsigma_{tj}$ "Level 2"

(5c) $\beta_{00j} = \beta_{000} + \varsigma_j$ "Level 3"

(5d) $Y_{itj} = \beta_{000} + \beta_1 X_{1itj} + \beta_2 X_{2tj} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$

- $\mathbf{X_{2tj}}$ is the country-year level 2 factor, with fixed effect β_2
- Y at Level 1 depends on X_{1itj}, a Level 1 independent variable, as well as X_{2jt}, a Level 2 contextual variables that affects all individuals at Level 1 at that time in that country
- Protest is a function of an individual's level of education plus the state of the country's economy at a given time
- Can add Level 3 time-invariant variables in 5(c) as well

- But: we know, following the same logic as with the hybrid model earlier in Unit 1, that β_2, the coefficient associated with the X_{2jt} Level 2 variable, is a mixture of its "within" Level 2 effect and its "between" effect at Level 3.
- That is, we need to separate the variable's country-year ("within") effect from its country "between" effect; otherwise we could mistake the effect of changing X_{2jt} by one unit from one time to another for the effect of a country being one unit higher than another country on X_{2jt} at all points in time
- This is one of the main advantages of the 3 level set-up: we can treat Level 2 as over-time observations of the same Level 3 units, and distinguish "within-unit" change from "between unit" levels on important contextual variables. **We have a Level 2/3 panel!!!**
- Technically, this overcomes potential endogeneity bias due to the relationship between X_{2jt} and ς_j, the Level 3 random effect

(6a) $Y_{itj} = \beta_{0tj} + \beta_1 X_{1itj} + \varepsilon_{itj}$ "Level 1"

(6b) $\beta_{0tj} = \beta_{00j} + \beta_2 X_{2tj} + \varsigma_{tj}$ "Level 2"

(6c) $\beta_{00j} = \beta_{000} + \beta_3 \bar{X}_{2j} + \varsigma_j$ "Level 3"

(6d) $$Y_{itj} = \beta_{000} + \beta_1 X_{1itj} + \beta_2 X_{2tj} + \beta_3 \bar{X}_{2j} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$$

- Or, expressing X_{2tj} in mean deviation form:

(6e) $$Y_{itj} = \beta_{000} + \beta_1 X_{1itj} + \beta_2 (X_{2tj} - \bar{X}_{2j}) + \beta_4 \bar{X}_{2j} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$$

- where $\beta_4 = \beta_2 + \beta_3$ from (6d)

- We have a nice **Level 2/Level 3** hybrid random effects model!!!

- So in our example, β_2 is the effect on individuals' protest propensity at a given time from a country *changing* by one unit on economic performance, relative to its average economic performance, and β_4 is the effect on individual's protest propensity at a given time from a country being one unit higher than another country at all times
- As with Level 1/Level 2 hybrid models, we give the "within" effect a stronger causal interpretation than the "between effect", as the within effect controls for confounding due to stable unobservables at the higher level (as the inclusion of $\bar{X}_{2j}$ picks up this correlation)
- The "between" effect of $\bar{X}_{2j}$ is possibly biased since (by assumption) $\bar{X}_{2j}$ is unrelated to ς_j, the Level 3 random effect (so we may overestimate its effect if there *is* correlation)

When a Level 2 variable is an aggregate of a Level 1 variable:

(7a) $$Y_{itj} = \beta_{0tj} + \beta_1 X_{1itj} + \varepsilon_{itj}$$ "Level 1"

(7b) $$\beta_{0tj} = \beta_{00j} + \beta_5 \bar{X}_{1tj} + \varsigma_{tj}$$ "Level 2"

(7c) $$\beta_{00j} = \beta_{000} + \beta_6 \bar{X}_{1j} + \varsigma_j$$ "Level 3"

(7d) $$Y_{itj} = \beta_{000} + \beta_1 X_{1itj} + \beta_5 \bar{X}_{1tj} + \beta_6 \bar{X}_{1j} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$$

or

(7e) $$Y_{itj} = \beta_{000} + \beta_1 (X_{1itj}) + \beta_5 (\bar{X}_{1tj} - \bar{X}_{1j}) + (\beta_5 + \beta_6) \bar{X}_{1j} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$$

or

(7f) $$Y_{itj} = \beta_{000} + \beta_1 (X_{1itj} - \bar{X}_{1tj}) + (\beta_1 + \beta_5)(\bar{X}_{1tj} - \bar{X}_{1j}) + (\beta_1 + \beta_5 + \beta_6) \bar{X}_{1j} + \varsigma_{tj} + \varsigma_j + \varepsilon_{itj}$$

We obtain Level 1 “within” and Level 2 “between” effects, **and** Level 2 “within” and Level 3 “between” effects! Pretty cool!!

Extension: Level 2 Growth Model

- Final variation: add TIME to the Level 2 equation (instead of time dummies), and you have a Level 2 Growth Model, where you predict the trajectory of a Level 3 unit over time with stable Level 3 characteristics
- All of this exploits the repeated measures of the Level 3 units, even though we do not have a true panel with the same units being observed at Level 1. We *do* have a Level 2 panel of observations nested within Level 3
- See Fairbrother (2014), "Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets", *Political Science Research Methods*