# Random Effects and Hybrid Models for Multi-Wave Panel Data

PS 2701-2019

Week 3

Professor Steven Finkel

# The Random Effects Model

- Alternative to FE/FD: "Random Effects" (RE). Gets around some of these problems, but has its own set of possibly problematic assumptions as well. (remember: **TANSTAAFL!)**
- Go back to original model of heterogeneity:

(1) $Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \ldots \beta_k X_{ikt} + U_i + \varepsilon_{it}$

- Look at composite error term: $(U_i + \varepsilon_{it})$
  - RE says: let's estimate the model by treating **both** components of the error term as arising from (independent) random processes. This is the way $\varepsilon_{it}$ is usually modeled. What is new here is treating $U_i$ the same way. Instead of being a "fixed" quantity at the level that is produced by our sample observations, we treat $U_i$ as a second normally distributed variable so that each unit's $U_i$ is the result of a random draw from this second distribution.
  - Some units higher on Y generally because of a randomly drawn high $U_i$ , some units lower on Y generally because of a randomly drawn low $U_i$
  - This is why the RE model is sometimes called the "**Random Intercept**" Model – the draw from the second distribution, added to the grand mean of Y, gives each unit its observed intercept.

# Assumptions of RE

1. The two components of the composite error, $U_i$ and $\varepsilon_{it}$ are independent, i.e. $E(U\varepsilon)=0$
2. The variances of both $U_i$ ($\sigma^2_u$) and $\varepsilon_{it}$ ($\sigma^2_\varepsilon$) are constant for all X (no heteroskedasticity)
3. The idiosyncratic residuals $\varepsilon_{it}$ at one point in time are not related to their value at another point in time (no autocorrelation in $\varepsilon_{it}$).

So far, so good, these three are relatively unproblematic.

4. Both $U_i$ and $\varepsilon_{it}$ are unrelated to the $X_{ik}$, i.e. $E(XU)=E(X\varepsilon)=0$

Yes, you read that right!! In order to identify and estimate the β in the RE model with *two* separate error terms, we need to treat them as *unrelated to the observed independent variables*. This is the usual OLS assumption for ε, and now we have to extend it to U as well. Otherwise we can't estimate the separate effects of X and the composite error.

- If assumption (4) is true, then RE is **definitely** the best estimator available. It is the most efficient, since (as we will see) we are only adding one additional parameter to the estimation --- the variance of $U_i$ ($\sigma_u$) --- instead of the *N-1* new parameters in FE/FD. But if this assumption is **not** true, then random effects gives wrong answers, as the RE estimator will be biased (technically, "inconsistent": biased even when $N \rightarrow \infty$)
- In fact the potential violation of this assumption is the reason that many panel analyses are done in the first place!!!!
- Thus, many people dismiss RE out of hand, and believe FE is the way to go. However, let us see what the RE estimator entails, how it works, what its advantages may be, and then how we might make use of it, if not in its "pure" form, then in some modified form to add to our toolkit of longitudinal methods.

# Estimation of RE Models

- Problem in estimating (1)? The composite error term ($U_i + \varepsilon_{it}$) has a complicated structure that is no longer independent over time within cases because of the presence of $U_i$
- The variance of the composite error, given assumptions (1)-(3) above, is: $\sigma^2_u + \sigma^2_e$. The covariance of the composite error is: $\sigma^2_u$ for all time periods.
- So the variance-covariance matrix of the error term, given RE assumptions, is (for a 4 wave panel as an example):

$$(2) \quad \begin{bmatrix} \sigma_u^2+\sigma_\varepsilon^2 & & & \\ \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & & \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 \end{bmatrix}$$

- So diagonal of this matrix has the variance of both $U_i + \varepsilon_{it}$, and the off-diagonals have variance of $U_i$ only. This is called an "**exchangeable**" error structure, since the given occasions of observation are all exchangeable in the sense that the covariance of time 1 and time 2 errors is the same as the covariance between time 1 and time 3, time 1 and time 4, time 2 and time 4, etc. The only thing producing covariation between the errors at different occasions is the constant unit effect $U_i$.
- Note: The error structure in (2) is not necessarily *correct*, it is simply the one implied by the standard RE assumptions. If there is heteroskedasticity or autocorrelation in the $\varepsilon_{it}$, then this error structure will need to be revised.
- Compare this to the error term structure of pooled OLS:

$$(3) \qquad \begin{bmatrix} \sigma_\varepsilon^2 & & & \\ 0 & \sigma_\varepsilon^2 & & \\ 0 & 0 & \sigma_\varepsilon^2 & \\ 0 & 0 & 0 & \sigma_\varepsilon^2 \end{bmatrix}$$

- How do we estimate a model such as equation (1) that has an error structure like matrix (2) as opposed to the "OLS-ready" error matrix (3)?

- Several ways, but the simplest is to estimate via a procedure known as **Generalized Least Squares (GLS),** which involves ***weighting*** the equation by a factor that will transform the problematic error term (2) into a variant of the unproblematic error term (3), so that OLS can be used on the *weighted* or *transformed* model.
  - This is what used to be the most common correction for heteroskedasticity: weight the data by the inverse of X or the square of X (because the unequal variance was an increasing or decreasing function of X), and this weighted equation would then have an error variance that satisfied OLS assumptions. So WLS (Weighted Least Squares) is a type of GLS estimation.
  - Another example of GLS is in time-series analysis, where one might weight the data by $\varrho$, the autocorrelation parameter for the $\varepsilon_{it}$, and then use OLS on the weighted data to estimate structural effects when the error term is autocorrelated.
- Generally: GLS proceeds by weighting the data by the *inverse* of the error variance-covariance matrix to ensure that the weighted equation has a normal structure with common variance on the diagonals and zero covariances on the off-diagonals. Then OLS is used on the weighted equation.

# GLS Random Effects Estimation

- Needed: estimates of the two variance terms $\sigma^2_u + \sigma^2_e$. If we could obtain those estimates, we can weight or transform equation (1) in the following way and then use OLS to estimate the effects:

$$(4) \quad Y_{it} - \theta\bar{Y}_i = (\alpha - \theta\alpha) + \beta_1(X_{1it} - \theta\bar{X}_{1i}) + \beta_2(X_{2it} - \theta\bar{X}_{2i}) + \beta_k(X_{ikt} - \theta\bar{X}_{ik}) + (U_{it} - \theta\bar{U}_i + \varepsilon_{it} - \theta\bar{\varepsilon}_i)$$

with:

$$(5) \quad \theta = 1 - \left( \frac{\sigma_\varepsilon}{\sqrt{T\sigma_u^2 + \sigma_e^2}} \right)$$

- If the observed Y and X in the model are transformed/weighted by equation 5's θ ("Theta") then the resulting error term in (4) will be OLS-ready, i.e. with constant variance on the diagonals and zero covariance on the off-diagonals.
- If we knew $\sigma^2_u + \sigma^2_e$. we could just plug them in to (5). But we don't know the population values of these two error variances. We need to *estimate* them from our data, which is why this application is called **"Feasible Generalized Least Squares" (FGLS)** and not the "real" thing. It is a bit more imprecise, but this is taken care of in various adjustments of degrees of freedom and thus the standard errors of the resulting coefficients.

- How do we get estimates of $\sigma^2_u + \sigma^2_e$? More tricks, this time from the "Within" and "Between" Regressions that we used in the FE transformation earlier.
- From the **Within** (FE) Regression we get an estimate of $\sigma^2_e$.
  - Why? FE eliminates $U_i$ altogether, and the error term is pure ε
- From the **Between** Regression (the means of $Y_i$ over time against the means of all the $X_i$ over time), we get an error term whose variance is: $\sigma_u^2 + \frac{\sigma_\varepsilon^2}{T}$ or the variance of U plus the *time-averaged* variance of ε.
- So from this information we can calculate an estimate of each error component, calculate θ or "THETA," transform equation (4) by this estimate and re-estimate the model with OLS
- Computational Formula for θ:

$$(6) \quad \hat{\theta} = 1 - \sqrt{\frac{\sigma_W^2}{T\sigma_B^2}}$$

# Notes on RE

(1) We can examine $\theta$ more closely to get a better idea of what RE is doing. As $\theta$ (theta) gets closer to 1, it means that more and more of the composite error variance is made up of $U_i$ unit-level or "between" variance. So what happens then? Then the weighted RE equation (4) reduces to the FE equation !! This is as it should be (right?), because if all of the error variation is from U, let's difference out U completely as the FE model does.

- As theta gets closer to 0, it means that more and more of the composite error variance is made up of random idisosyncratic variance $\varepsilon$, with no unit variance at all. So the RE equation reduces to POOLED OLS in this instance!! This is also as it should be (right?), because we only should take into account unit effects when they exist!!
- So we can look at the RE estimator as a ***weighted average* of FE and pooled OLS**, with the weight ($\theta$) depending on how much of the estimated composite error variance is from the units. This is a "middle ground," then, between the full unit-level differencing model of FE and the assumption of no unit effects in pooled OLS. If there is a lot of unit-level variation, then RE is closer to FE. If there is not so much unit-level variation, then RE is closer to pooled OLS.
- This seems reasonable, **IF** the RE assumption of zero correlation between X and U is tenable – a big "if"!!

(2) Why a "weighted" average between FE and OLS? Another way to look at RE: We need to correct FE to the extent that the information from our sample regarding the $U_i$ is unreliable, or the extent to which the unit effect is based on chance deviations and not "real" ones. (We said that this was one of the problems with FE last session). So we can say that RE "*shrinks*" the estimates of $U_i$ back to where they "should be," or rather back to the overall grand mean, to the extent that there is not a lot of reliable unit-level variation in the model, that is, unit-level variation based on lots of time periods of observations (see the denominator of (5) or (6)).

If there is lots of unit-level variation with lots of time points, then we more or less take FE "at its word," (so to speak), but if there is very little overall unit-level variation and few time points, we adjust FE and say "some of the observed variation is likely to have come about simply by chance, and is not to be believed." We then "*shrink*" the FE estimates back towards pooled OLS in that case.

(3) Another interesting aspect of the RE model in comparison with FE. Look again at the formula for θ in either (5) or (6) and the *T* in the denominator. This means that as we have more and more waves of observation, θ goes to 1.

This means that **FE and RE converge as T gets large.** So the RE/FE distinction is really important in panel analysis, but not as much in pooled time-series-cross-sectional analysis with large *T*.

- The implication of this is that the true error term "reveals" itself to us as T→∞, and the unobserved $U_i$ term becomes observable!!!
- Look at the "Between Model" again from the FE slides from last session:

$$\bar{Y}_i = \alpha + \beta_1 \bar{X}_{1i} + \beta_2 \bar{X}_{2i} + \beta_3 \bar{X}_{3i} + \ldots \beta_j \bar{X}_{ji} + U_i + \bar{\varepsilon}_i$$

- The error term here is composed of $U_i + \bar{\varepsilon}_i$. As *T* gets bigger and bigger, the $\bar{\varepsilon}$ converges to 0, meaning that all that is left is $U_i$. So with large *T*, no real difference between FE and RE. We can use either FE or RE (with a θ of 1 – see equation 4) or pooled OLS with the inclusion of $\hat{U}_i$ from the Between Model above -- all will yield the same results!!

# RE versus FE: The Hausman Test

- So, should you use FE or RE? Clearly, the choice is based on whether the RE assumptions (1) through (4) above are satisfied. If they are, the RE is more efficient, as stated earlier. If they are not, then RE will be biased and *inefficient*, and this is usually taken to mean that FE will be preferred. (This is becoming more controversial, though, as we will see below in the "compromise" section.)
- A test developed by Hausman exists to assist in this choice. The intuition of the Hausman test: If the assumptions of RE hold, then FE and RE are two different ways to get the correct estimates, but RE is more efficient. If the assumptions do not hold, then RE will be "inconsistent" (biased even as sample size gets larger and larger), and only FE will be "consistent". So we should see similar estimates between RE and FE if the RE assumptions hold, and different ones if they don't.
- So the Hausman test is:

$$\frac{\beta_{FE} - \beta_{GLS-RE}}{\operatorname{var}(\beta_{FE}) - \operatorname{var}(\beta_{GLS-RE})}$$

  with the test statistic being distributed as $\chi^2$ with degrees of freedom equal to the number of time-varying independent variables, under the null hypothesis that:

  $\beta_{FE} = \beta_{GLS-RE}$.
- Rejecting $H_0$ is usually taken to mean that the RE assumptions – in particular, no correlation between X and U – do not hold, and that therefore FE is preferred. If not, we would have seen similar estimates from RE and FE, given sampling error.

# RE versus FE: Is RE appropriate in "non-sampling" situations?

- A more substantive argument about RE: it is only appropriate when the unit effects were drawn randomly from some population, as would be the case in large-scale survey research, for example. In these cases, we are not that interested in the actual value of $U_i$, they are just representative of a larger population or universe of $U_i$ and we happened to get case 1, 2, 3, etc. So if you can think of sample as a collection of units where you are more interested in overall population, then the random effects set-up in theory might be appropriate.
- By contrast, country or aggregate level studies are not as well suited for RE, following this logic. Germany is Germany is Germany and India is India is India – the unit effects we observe *are* fixed for these units and the inferences we wish to make are **conditioned on these specific units**. Also, when you have a relatively small N, it is harder to justify the **normal** assumption for the unit effect distribution, so FE is more tenable, as it makes no assumptions whatsoever about the shape of the $U_i$ distribution.
- This is one way to look at the issue. Another, following Allison (1994): FE and RE are simply alternative ways of dealing with the "nuisance" posed by the presence of the $U_i$. We can deal with it by estimating one parameter ($\sigma^2_u$) under the assumption of a normal distribution and zero correlation between X and $U_i$ (RE), or we can deal with it by estimating *N-1* parameters (in the LSDV version of FE) under no assumptions about the distribution and no assumptions about the correlation between X and $U_i$. We choose the one that holds in the given sample that we are analyzing and hope that the "nuisance" posed by the $U_i$ is taken care of in the most efficient manner possible that does not produce bias in the estimates in the process.

# “Hybrid” Models

- Recently, there has been more attention given to possible compromise models that fuse some of the attractive features of both RE and FE. Two ways of looking at what we want from the ideal model is:

- (1) Keep the FE set up, while trying to say something about the effects of time-invariant Xs; and

- (2) Keep the RE set up, while at the same time allowing possible correlation between the X and the $U_i$.

- One method exists that stems from (1), that is, from FE outward, so to speak, while two others stem from (2), that is, from RE outward.

# Plümper-Troeger's (2007) "Fixed Effects Variance Decomposition"

- Three step estimation:
  a. Use FE model to get estimates of TVs, and an estimate of the $U_i$. In this step we get the $\beta$ for the TV variables, controlling for the potential correlation with $U_i$ as before, and we can generate an estimated unit effect as well. (We did this manually in the DO file and saw that STATA generates it also internally).
  b. We predict the estimated $U_i$ from step (a) from all TIV variables in a "between"-type regression. This generates the effects of the TIV variables on the FE unit estimate, which we knew contains not only the "true" unit effect but the effects of all other TIVs. This step purges the FE unit effect of everything we can explain from the TIVs, and the residual from this step's equation leaves us with the best guess of the "pure" $U_i$ that we can come up with.
  c. Now throw all of this concoction into an OLS equation: the TV, the TIV, and the step (b) residual, which has our best estimate of $U_i$. This is our best construction of the actual panel data regression of our original heterogeneity equation (1) with the manually-arrived at unit effect as an "observed" variable. We use regular old OLS on this and get correct estimates of standard errors, effects of the TIV on the mean of Y, and effects of the TV on Y, controlling for their possible correlation with the $U_i$.
- See Thomas Plümper's web site to download a STATA ADO file to implement the procedures here

# Hausman-Taylor's (1981) Instrumental Variables Method

- Another possibility, somewhat popular in econometrics but with virtually no applications in political science, is a method from Hausman (the same one from the test)-Taylor.

- Idea: if some of the Xs are related to the error term (in this case $U_i$), then one way to estimate causal effects is through an instrumental variables (IV) approach, if appropriate exogenous instruments can be found for the problematic Xs.

- Categorize the independent variables in the following way:
  - TV1 = Time-varying variables that we assume are related to $U_i$
  - TV2= Time-varying variables that we assume are not related to $U_i$
  - TIV1= Time-invariant variables that we assume are related to $U_i$
  - TIV2= Time-invariant variables that we assume are not related to $U_i$

- Note that these assumptions are only that: assumptions!!

- How do we get coefficients for the TV variables? Regardless of whether they are or are not related to the $U_i$, we can use the de-meaning FE procedure estimation and get consistent results
- How to get coefficients for the TIV? If they are unrelated to $U_i$, then *any* procedure aside from FE or FD will work, including RE, OLS, whatever
- The real problem is getting estimates for the TIV that *are* related (by assumption) to the $U_i$
- We need to find an instrument for such variables to use in an IV or TSLS-type estimation. Such an instrument must:
  - covary with the problematic TIV
  - be unrelated to the $U_i$ (and $\varepsilon$ of course too)
- How about the mean (average) of TV2, the time-varying factors that are uncorrelated with $U_i$?
  - Since the values of these variables are by assumption uncorrelated with $U_i$, the average value is also uncorrelated with $U_i$.
  - So we use as instruments for TIV1: TIV2 and Mean(TV2)!!!
- To program this in Stata **(XTHT)**: TIV1 is "endogenous", TV1 is "endogenous" and all other variables are treated as exogenous. One needs at least one endogenous TV and one endogenous TIV, and at least one exogenous TV and one exogenous TIV to identify the model
- Great idea, but completely dependent on assumptions regarding exogeneity and finding appropriate instruments!!!

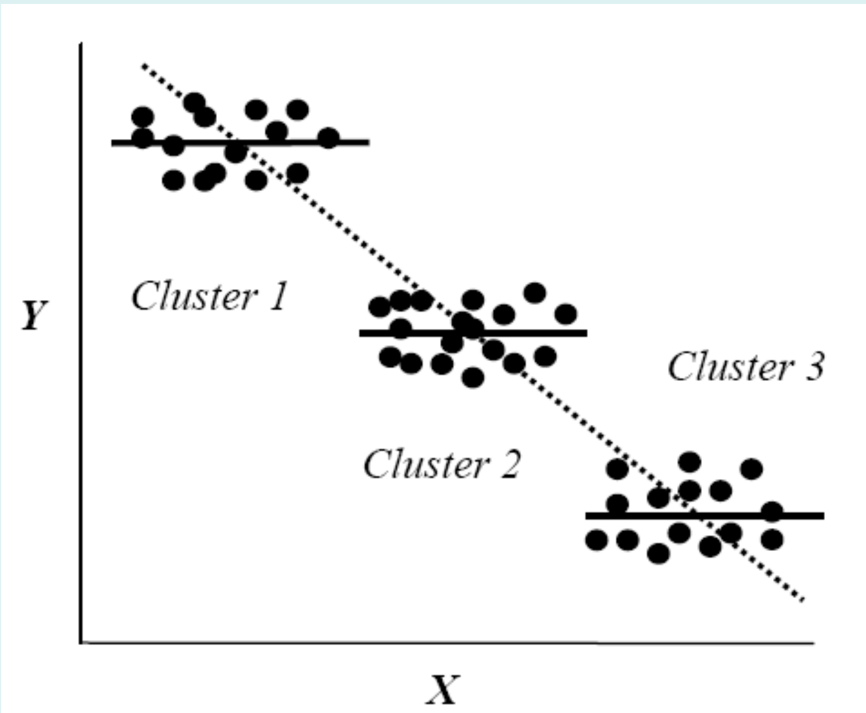
# Bell&Jones Random Effects-Hybrid Model

- The most important "compromise" method we will consider starts from RE. This procedure --- discussed by Bell&Jones and others from the multi-level modeling tradition, is extremely simple and might be labeled under the category: "what is all the fuss about?"
- The idea here starts with the notion that the possible covariation of time-varying Xs and the $U_i$ is what messes up RE. But this possible covariation is usually just the result of model misspecification --- something in the $U_i$ term is related to something in the X that we need to account for, and RE (at present) cannot account for it because of its assumption that E(XU)=0. But we can bring the covariation between X and $U_i$ into the model indirectly, by including the **mean of X** as an additional independent variable in (4):

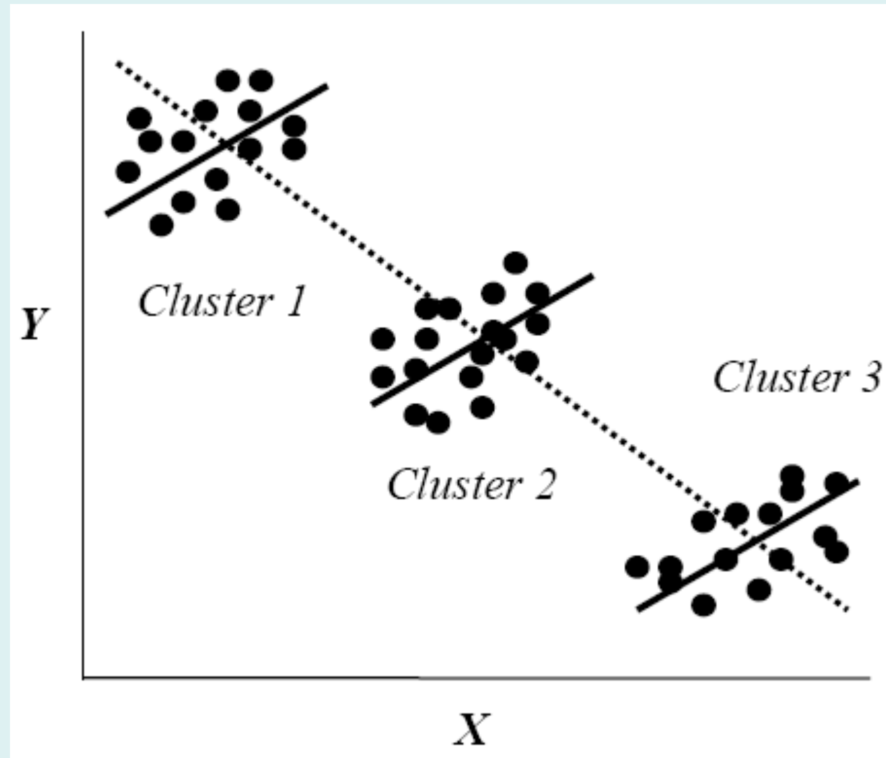
$$(7) \qquad Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 \bar{X}_{1i} + U_i + \varepsilon_{it}$$

- Whatever covariation between X and U that may exist is now accounted for; if units that are generally high (low) on X also have high (low) U terms, then the mean of X in the model will pick this up. The effect of "regular" X can now be estimated, controlling for this possible confounding problem. Sounds like FE!!

Examples: X-bar picks up the X-U correlation: Clusters low on X(bar) are high on U, clusters high on X(bar) are low on U. Controlling for X-bar allows estimation of the "within" effect of X, just like FE did!




Null "within" cluster effect, negative "between" cluster effect




Positive "within" cluster effect, negative "between" cluster effect

- So we now have a $U_i$, $X_{it}$ and X-bar in (7), and the β for the time-varying values of X are thus estimated in RE while controlling for the possibly confounding correlation between X and $U_i$. So we can use RE for all the reasons we like RE:
  - only one lost df and thus more efficient estimates
  - the ability to model the intercepts and try to account for why some units are higher or lower than others
  - the ability to include other TIV in the model as predictors
- As we will see, the RE model is also the baseline model for the multilevel/hierarchical school or framework for longitudinal analysis. This simple little correction shows that we can use this framework and still not sacrifice some of the advantages of FE
- **HEY, MAYBE THERE IS A FREE LUNCH AFTER ALL?**
  - Not completely. Need to assume that (residualized) U is unrelated to all Z, so effects of Z may be overestimated (as Z takes any correlated effects it may have with U for itself). This is not a **huge** drawback but still a drawback. **TANSTAAFL!!**

- Can examine the "hybrid" model more closely:

$$(8) \qquad Y_{it} = \alpha + \beta_1 X_{1it} + \beta_2 \bar{X}_{1i} + U_i + \varepsilon_{it}$$

- Can write this model in terms of the mean-deviations in X too:

$$(9) \qquad Y_{it} = \alpha + \beta_1 (X_{1it} - \bar{X}_{1i}) + (\beta_1 + \beta_2) \bar{X}_{1i} + U_i + \varepsilon_{it}$$

$$(10) \qquad Y_{it} = \alpha + \beta_1 (X_{1it} - \bar{X}_{1i}) + \beta_3^* \bar{X}_{1i} + U_i + \varepsilon_{it}$$

- With $\beta_3^* = \beta_1 + \beta_2$ from equation (8). Equation (10) expresses the model in terms of the mean of X, and the deviation of $X_{it}$ from the mean of X.
- We can see from this expression that the Hybrid coefficient $\beta_1$ is going to give you the same value as the FE estimate, and coefficient $\beta_3$ is going to give you the BE estimate (the "between" estimate) of the causal effect of X on Y. In other words, the coefficient for $\beta_1$ will give you the effect of a **within-unit** change on $Y_{it}$ (i.e. changing a given case by one unit over time), and the coefficient for $\beta_3$ is going to give you the effect of a **between-unit** change on $Y_{it}$ (i.e. changing a given case into another case that is, on average, one unit higher).

- Will these two estimates generally be the same?
- Traditional RE models (see equation 1) assume that the answer is yes. That is, they assume that $\beta_1 = \beta^*_3$ so that the effect of X-bar on Y is zero. Thus, we can say that the traditional RE formulation says that the FE and BE estimates of the effect of X (or the "within" and "between" effects) are equal!!!
- In fact, we can conduct a statistical test of the equality of these coefficients in STATA using the "test" command after running an RE version of equation (10):

  test $\beta_1 = \beta^*_3$

- **THIS IS EXACTLY THE SAME AS THE HAUSMAN TEST FOR FE VERSUS RE DESCRIBED ABOVE!!!!!**

- So the traditional RE assumption is the same as saying that the *within-person (unit)* and *between-person(unit)* effects are the same. If this assumption is wrong, then traditional RE is wrong.

- When will this be wrong? When you can imagine that comparing <u>units</u> that differ from one another by one unit on X will lead to different effects on Y from the effect of changing a single unit at time 1 by one unit to time 2. (See earlier graphs on cluster bias).
- Examples: Typing Speed and Error Rates; Exercise and Mortality
- Social Science Examples:
  - Person X lives in the U.S. South, Person Y not. We expect Person X to have a *lower* salary than person Y, i.e. the **between** effect of "South" will be negative.
  - But Person Y now moves to the South. We may expect that changing from non-South to South for a given person may be associated with a *higher* salary (perhaps the motivation for moving). The "**within**" effect of South might therefore be positive.
- Or:
  - Gelman's (cross-sectional) work on US voting behavior: Persons living in rich states are more likely to vote Democratic than persons living in poor states (i.e. the "**between** effect" of income on the Republican vote is *negative*). But rich people in a given state are more likely to vote Republican (i.e "the **within** effect" of income on the Republican vote is *positive*).

# Moral(s) of the Story

- Rejecting the null hypothesis, or rejecting RE assumptions in the Hausman test either means that FE is valid, *or* that there is some omitted effect of Xbar on Y, independent of the effect of ($X_i$-Xbar). As Skrondal and Rabe-Hesketh state in their book *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models* (Chapman and Hall 2004, p. 53):
  - *"Some economists believe that a significant Hausman test implies that the random intercept model must be abandoned in favor of a fixed effects model. However, this is misguided since $\beta_1$ can be estimated without bias as long as the cluster mean is included as a covariate in addition to $X_{it}$"*
- The model with both Xbar and ($X_i$-Xbar ) as independent variables allows RE estimation in the context of potential $XU_i$ correlation, *and* it allows interpretation of possible differences in the *within* and *between* estimates, which may have substantive implications in a given analysis.
- Staying within the RE framework has other advantages, e.g. allowing estimation of stable observables, and allowing estimation of the random coefficient models we will consider later. So the "hybrid model" could very well be the best model that we have considered so far.

# Multilevel Derivation of Random Effects Models

- it is increasingly common to view panel data as a kind of *multilevel* or *hierarchical* data structure, where the waves of observation are "nested" within individual units (people, neighborhoods, cities, countries, country-dyads, etc.)
- So a three-wave panel of individuals would have 3 observations at the lowest level of the data hierarchy -- "Level 1" -- for each "Level 2" individual at the next level of the data hierarchy
- This fits well into "long" data logic, whereby each wave of observation is its own row in the data matrix, with multiple rows of data per each individual unit.
- The idea of multilevel analysis is to build models that account for outcomes at *both* levels of the data hierarchy, and that means accounting for *within-unit* (Level 1) and *between-unit* (Level 2) processes simultaneously

# Multilevel Models are Ubiquitous in Social Sciences, Panels or No Panels!!

- In **education research**, children (Level 1) are nested within schools (Level 2), nested within districts (Level 3).
- **Survey research**: respondents (Level 1) are (hypothetically) nested within blocks (Level 2) which are nested within Primary Sampling Units (Level 3) which are nested within, for example, U.S. states (Level 4)
- **Trend analysis**: respondents (Level 1) are nested within Time (Level 2), since in this kind of design *different individuals* from the same aggregate unit (country, state) are interviewed at multiple points in time. This is the *reverse* of panel data, where time of measurement (Level 1) is nested within individuals (Level 2). Example: Pooled NES Election data, 1948-2012
- **Cross-national survey research**: respondents (Level 1) are nested within countries (Level 2), and perhaps Time (Level 3) as well. Example: World Values Surveys and Latin America/Afro/Asian Barometers.
- Whenever data structures are nested or hierarchical, you can use the multilevel framework we are discussing now (and later in Unit 4)!

# The Basic Multilevel Panel Model

- Multilevel models proceed by specifying the hypothesized causal processes at both (all) levels of the data hierarchy
- A simple example would be a model with only one time-varying independent variable $X_{it}$. Since the variable changes over time, it is a "Level 1" explanatory variable – it potentially accounts for time-specific levels/changes in an outcome or dependent variable. So we could write the model as:

$$(11)\quad \text{Level 1:}\quad Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it}$$

$$\text{Level 2:}\quad \beta_{0i} = \beta_{00} + U_i$$

- Here $X_{it}$ is the sole explanatory variable at Level 1, along with an intercept $\beta_{0i}$ that varies across individual units (note the *i* subscript) and a unit-time error term

$$(11) \quad \text{Level 1: } Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it}$$

$$\text{Level 2: } \beta_{0i} = \beta_{00} + U_i$$

- At Level 2 – the level of the individual or unit – we have the unit's mean or average value predicted by a grand mean that is common to all units ($\beta_{00}$, note no "*i*" subscript), *plus* a unit specific error term ($U_{i.}$).
- $U_i$ represents – as we have discussed so far – all of the unobserved factors that push a given individual to be, on average, higher or lower on Y than another unit. It is the Level 2 error term in the multilevel framework – the portion of the unit-level average that cannot be predicted from the other parts of the model
- It is also necessary to include the $U_i$ term for another reason. The observations at Level 1 – by virtue of being "nested" in Level 2 units – are not *independent observations*. The $U_i$ term accounts for the clustering of this data structure, that is, it represents what is common to all *i* observations over time for each particular *i*. This is another aspect of the issue of "unobservables" that need to be taken into account in panel models

$$(11) \quad \text{Level 1: } Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it}$$

$$\text{Level 2: } \beta_{0i} = \beta_{00} + U_{i}$$

- Another interesting feature of multilevel models: You can view the "Level 2" model as "explaining" or "predicting" the size of a Level 1 parameter from Level 2 characteristics. In this case, the Level 1 parameter we are predicting is the unit-specific intercept $\beta_{0i}$, and we are explaining the size of the parameter with $U_i$ and a grand mean that is common to all units. So this model has the unit-specific intercept at Level 1 as a **dependent variable** at Level 2.
- This suggests that a more fully specified multilevel model would include other independent variables at Level 2 as well. We can add time-invariant $Z_i$ variables to the Level 2 model, so that the unit-specific intercept is predicted by the grand mean, a series of unit-specific characteristics (e.g. educational, income, etc. in the case of individuals), along with the unobservable unit-specific $U_i$ "error" component.

$$(11) \quad \text{Level 1: } Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it}$$

$$\text{Level 2: } \beta_{0i} = \beta_{00} + \beta_{01} Z_i + U_i$$

- Another possibility would be to include the Zs, as well as the unit-level average of $X_i$ at Level 2; this would mean that you believe that units generally high (low) on X will be generally high (low) on $Y_{it}$; or in other words, that the unit-specific intercept $\beta_{0i}$ will be higher (lower) than the grand mean to the extent that X-bar is high (low)

$$(12) \quad \text{Level 1: } Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + \varepsilon_{it}$$

$$\text{Level 2: } \beta_{0i} = \beta_{00} + \beta_{01} Z_i + \beta_{02} \bar{X}_i + U_i$$

- This formulation makes clear that, for a given independent variable, one may expect a different effect at Level 1 ($\beta_1$) than at Level 2 ($\beta_{02}$). Or, that you expect a different *within* unit (Level 1) effect from the *between* unit (Level 2) effect. This is exactly what our RE-Hybrid model assumes; we've just re-expressed this in multilevel language and notation.

- One can also combine the two separate equations into a single "mixed" equation, by substituting the Level 2 equations into the Level 1 expression for a given parameter

$$(13) \quad \text{Mixed:} \quad Y_{it} = \beta_{00} + \beta_{1i} X_{it} + \beta_{01} Z_i + \beta_{02} \bar{X}_i + U_i + \varepsilon_{it}$$

- If we assume that the $U_i$ and the $\varepsilon_{it}$ are both normally distributed random variables, we arrive **exactly** at the Random Effects-Hybrid model of (8). We can easily conduct the algebraic manipulations necessary to convert this into model (10) with *mean-deviated X* and *mean X* as the two independent variables in the model.
- So both the "classic" RE model and the RE-Hybrid model are examples of multilevel panel models – one omits X-bar (*mean X*) from the Level 2 formulation, while the RE-Hybrid model includes it.
- We'll examine maximum likelihood estimation methods (and Stata syntax) for multilevel "mixed" panel models in Unit 3

# Final observations (for now) on Multilevel Panel Models

- One can extend the multilevel logic by treating *other parameters* in the Level 1 equation as Level 2 dependent variables, not only the intercept. You can imagine the *slope* of a given $X_{it}$ effect – $\beta_{1i}$ – also varying across Level 2 units, and perhaps being predicted by a series of Level 2 independent variables along with unit-specific error component. We will discuss these kinds of "random coefficient models" later in the class.
- Another attractive feature of the multilevel framework is that one can also extend levels. For example, Ames/Renno/Baker (2005) analyze Brazilian election data at three levels: time (Level 1) nested within individuals (Level 2) nested within neighborhoods (Level 3). Different models could and should be specified for each of the levels, and then combined into a "mixed" formulation for estimation purposes.

# Final observations (for now) on Multilevel Panel Models

- While RE-Hybrid does indeed look like close to a "free lunch", we'll see later that there is one real problem with it: *mean*-*X* may not be a very reliable measure of the *true* unit or cluster-level mean, especially in short panels. Recent work advocates treating *mean*-X as a "latent" variable that is potentially contaminated with "measurement error". This idea builds on the Structural Equation Modeling framework that we'll be talking about in the next few sessions.
- It turns out that virtually **every panel model can be expressed within the multilevel framework**, and this is one of the insights most responsible for the integration and synthesis of panel analysis that is taking place across fields and traditions. We'll pick up on this theme as we progress.