Panel data analysis
Day 1. (cont.)

David Pupovac
davidpupovac@gmail.com

So far…

- … we were focused only on non-spherical errors. We have emphasized that most common problems are:
  - panel heteroskedasticity **(it is about *N* (and *T*))**
  - contemporaneous correlation of the errors **(it is about *T*)**
  - serially correlation of errors **(it is about *T*)**
- In this respect we spoke about:
  - Clustered sandwich estimator which corrects for the dependence of errors within groups
  - Feasible generalized least squares estimator
  - **(…and today we will talk about)** Panel corrected standard errors estimator (PCSE) which correct for contemporaneously correlated and panel-heteroscedastitic errors, but can be extended to correct for autocorrelation as well

What are these methods good for?

- These methods alone are rarely ever sufficient to correct for all the issues inherent to typical TSCS/CSTS data

- Clustered sandwich error estimator and panel corrected standard errors estimator are **corrections of standard errors**. In this respect, they can be combined with other modeling approaches (some of which we will discuss today)

- Feasible generalized least squares transform the data and the method can not be combined with other modeling approaches

- These methods treat issues arising from TSCS/CSTS data structure as a problem which needs to be eliminated (rather than modeled) – thus, to certain extent these methods ignore the panel structure of the data instead of exploiting it.

PCSE vs. FGLS

- Panel-corrected standard error was conceived as an alternative to feasible generalized least squares (FGLS)
- The full FGLS variance–covariance estimates are typically unacceptably optimistic (anticonservative) when used with the type of TSCD data analyzed by most social scientist: 10–20 panels with 10–40 periods per panel.
- Beck and Katz (1995) show that the OLS or Prais–Winsten estimates with PCSEs coverage to probabilities that are closer to nominal.
- The correction works only because we what repeated information on the contemporaneous errors. **The method works only in TSCS context.**

Panel Corrected Standard Errors

- Panel corrected standard errors correctly measure the sampling variability of the OLS estimates even with <u>panel heteroskedastic</u> and <u>contemporaneously correlated errors</u>.

- **In normal setting any serial correlation must be removed before the panel corrected standard errors are calculated**

  Stata allows you to avoid this (the disturbances may be assumed to be autocorrelated within panel, and the autocorrelation parameter may be constant across panels or different for each panel) – more on this later

- Note that this is a “sandwich estimator” in the style of the Huber-White robust estimator. The only difference is that Ω (e.i. variance-covariance matrix of disturbances) in $Var\_cov(\hat{\beta}) = (X^T X)^{-1} X^T \Omega\, X\, (X^T X)^{-1}$ is estimated differently.

Ω matrix so far:

No assumption violation

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Heteroskedasticity

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^2 \end{bmatrix}$$

Autocorrelation

$$\begin{bmatrix} \sigma^2 & \sigma_{1,2} & \dots & \sigma_{1,m} \\ \sigma_{2,1} & \sigma^2 & \dots & \sigma_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1} & \sigma_{m,2} & \dots & \sigma^2 \end{bmatrix}$$

Panel heteroskedasticity

$$\begin{bmatrix} \sigma_1^2 I & 0 & 0 & 0 \\ 0 & \sigma_2^2 I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m^2 I \end{bmatrix}$$

Cross-sectional correlation

$$\begin{bmatrix} \sigma_1^2 I & \sigma_{1,2} I & \dots & \sigma_{1,m} I \\ \sigma_{2,1} I & \sigma_2^2 I & \dots & \sigma_{2,m} I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m,1} I & \sigma_{m,2} I & \dots & \sigma_m^2 I \end{bmatrix}$$

$$\boldsymbol{Var\_cor}(\hat{\boldsymbol{\epsilon}}) = \boldsymbol{\Omega} = \boldsymbol{\epsilon\epsilon^T}$$

….which, if no violation = $\boldsymbol{\sigma^2 I}$

$$\boldsymbol{Var\_cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X^T X})^{-1}\boldsymbol{X^T \Omega\, X}\,(\boldsymbol{X^T X})^{-1}$$

….which, if no violation = $\boldsymbol{\sigma^2}(\boldsymbol{X^T X})^{-1}$

PCSE - estimation

- TSCS models with contemporaneously correlated and panel-heteroscedastitic errors have the covariance matrix of the errors, Ω as an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances along the block diagonal Σ

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,N} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} & \cdots & \sigma_{2,N} \\ & & & \cdot & \\ & & & \cdot & \\ \sigma_{1,N} & \sigma_{2,N} & \sigma_{3,N} & \cdots & \sigma_N^2 \end{pmatrix} \quad \Omega = \begin{pmatrix} \Sigma & 0 & 0 & \ldots & 0 \\ 0 & \Sigma & 0 & \ldots & 0 \\ 0 & 0 & \Sigma & \ldots & 0 \\ & & & \cdot & \\ & & & \cdot & \\ 0 & 0 & 0 & \ldots & \Sigma \end{pmatrix}$$

- $\sum$ is the panel-by-panel covariance matrix. We thus have T replications of the error that can be used to estimate this block; for large T, this estimate is quite good.

PCSE - estimation

- We need to estimate these diagonals Σ. As OLS is consistent we can use its residual. Let $e_{i,t}$ be the OLS residuals for unit i and time t. We can estimate typical element of Σ as:

$$\hat{\Sigma}_{i,j} = \frac{\sum_{t=1}^{T} e_{i,t} e_{j,t}}{T}$$

- To calculate PCSE organize the residuals from the fitted model according to the group. Letting E denote the $T \times N$ matrix of the OLS residuals, we can estimate Ω by:

$$\Omega = \frac{(E'E)}{T} \otimes I_t$$

- Panel corrected standard errors are then estimated by taking square root of the diagonal elements:

$$(X'X)^{-1} X' \left(\frac{E'E}{T} \otimes I_t \right) X (X'X)^{-1}$$

PCSE and Prais–Winsten

- Panel-corrected standard error (PCSE) estimates are estimated by either OLS or Prais–Winsten regression (FGLS).
- Prais–Winsten estimates are used when autocorrelation is specified

When to use PCSE vs. FGLS vs. clustered sandwich estimator?

- In small samples clustered sandwich estimator is not an improvement over the non-robust option. Clustered sandwich errors require $N > 10 - 15$. Consequently, method is appropriate for data which tend towards CSTS

- FGLS underestimates sampling variability if a large number of parameters need to be estimated. This is a case when T is large and where we need to account for cross-sectional correlation across a larger number of units (N). On the other hand, FGLS performs well in large samples. Consequently, the method is suitable for data which tend to CSTS

- Panel corrected standard errors are designed to account for cross-sectional correlation. Therefore, the method requires large T. Consequently, it is appropriate for TSCS data

Finally, remember that we can combine clustered sandwich estimator and PCSE with other methods, while we can not do this FGLS

Panel data analysis: fixed, within and between effects Day 2.

David Pupovac
davidpupovac@gmail.com

Pooling data

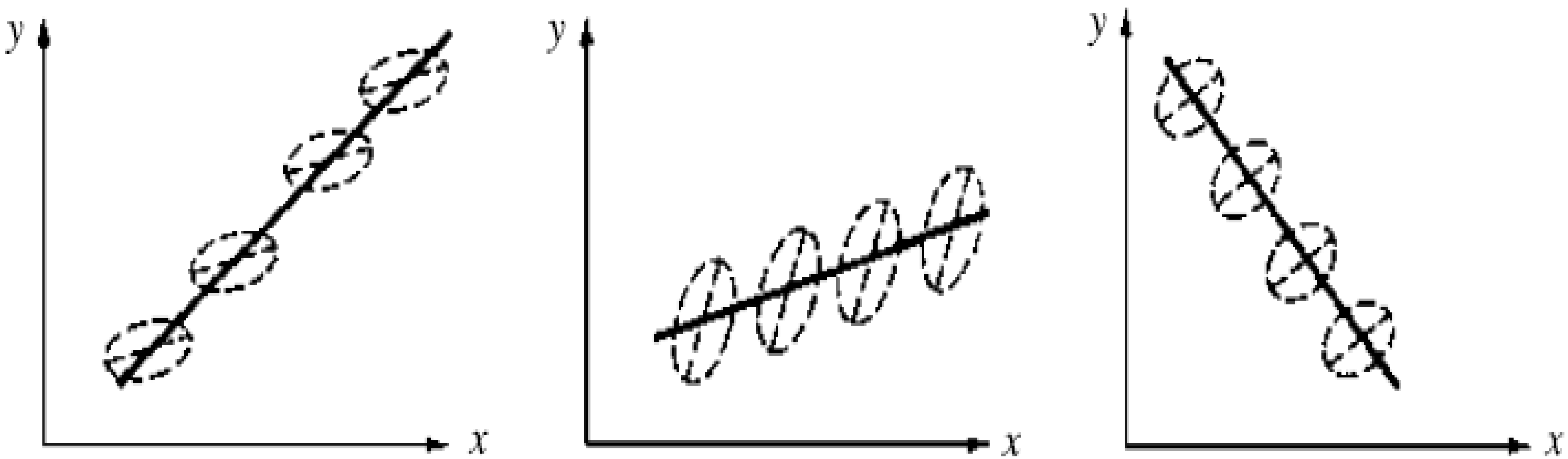
- As mentioned yesterday, even if we assume spherical errors, there are questions:

  - Are the intercepts same for each unit?
  - Are the slopes same for each unit?
  - Are the parameters constant over time?
  - And so on…

- These questions refer to assumption of the homogeneity. So, what is meant by heterogeneity?

What is heterogeneity?

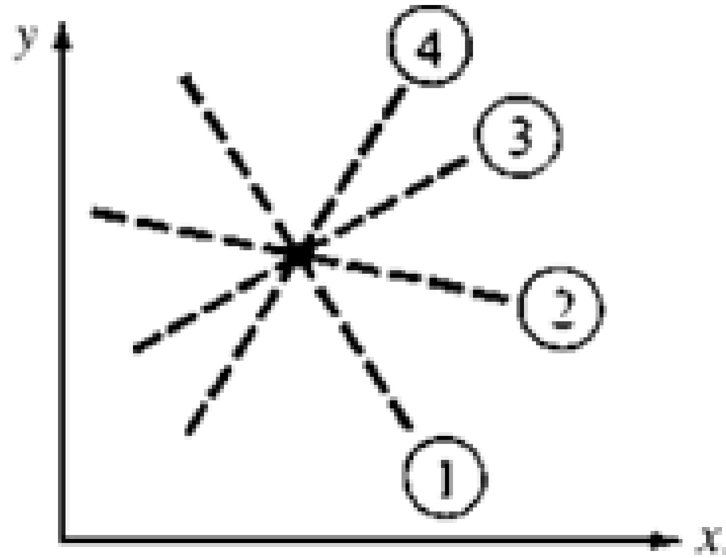
Example1: Heterogeneous intercepts, homogeneous slope



- In this case it is obvious that OLS applied to the pooled data in the presence of heterogeneity will lead to wrong estimate of slopes

What is heterogeneity?

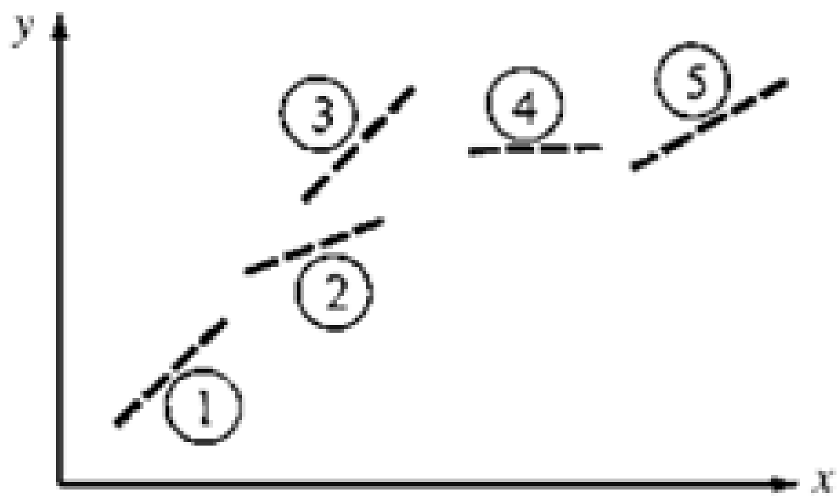
Example 2: Heterogeneous intercepts and slopes



- In this case assuming identical parameters for all cross-sectional units, would lead to nonsensical results because it would represent an average of coefficients that differ greatly across units.

What is heterogeneity?

Example 3: Different regression regimes across units



- In this case the pooled data implies a curvilinear relation between the variables.

Regression model on pooled data

- The OLS model on pooled data can be represented with:

$$y_{i,t} = \alpha + x_{i,t}\beta + \varepsilon_{i,t}$$

- Under the assumption of no latent heterogeneity and if the remaining assumptions of the classical model are met (zero conditional mean of $\varepsilon_{i,t}$ independence, homoscedasticity across observations i, and strict exogeneity of $x_{i,t}$), then we can estimate equation with OLS

- <u>However, these assumptions are unlikely to be met in TSCS/CSTS datat.</u>

Dealing with heterogeneity

- **Discussions of TSCS/CSTS data analysis typically focus on the differences between <u>fixed effects</u> and <u>random effects</u>.**

TODAY WE WILL PREDOMINANTLY DEAL WITH <u>FIXED EFFECTS</u>

- Although we can perceive the TSCS/CSTS data analysis from the viewpoint of ether time or cross-sections, we will predominantly focus on the understanding of cross-sections as the basis of panel (this is also more logical for the discussion and understanding of time dependencies) – we will relax this a bit later

Fixed effects

- A fixed group effect model **examines group differences in intercepts**, assuming the **same slopes** and **constant variance** across entities or units. The fixed effects formulation implies that differences across groups can be captured in differences in the constant intercept terms

- Due to emphasis on the intercepts it is common to think about the fixed effect models in terms of unit specific dummies. <u>OLS regressions with unit specific (cross-sectional) dummies is fixed effect model.</u> This fixed effects approach takes α_i to be a group-specific constant terms in the regression model, resulting in following expression:

$$y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t}$$

Fixed effects and ignorance

- **The dummy variable coefficients reflect <u>ignorance</u> – they are inserted merely for the purpose of measuring shifts in the regression line arising from <u>unknown/unobserved variables</u>.**

- Some researchers feel that this type of ignorance should be treated in a fashion similar to the general ignorance represented by the error term, and have accordingly proposed the **random effects** (variance components or error components) model

Fixed effect and correlation with regressors

- The fixed effects model arises from the assumption that the omitted effects, in the general model, are correlated with the included variables.

- **Since a group (individual specific) effect is time invariant and considered a part of the intercept, changes in intercept are allowed to be correlated to other regressors.**

- **This is one of the main distinguishing properties of fixed effects in comparison to random effects model which assumes that the random error associated with each cross-section unit is uncorrelated with the other regressors**

- It should be noted that the term “fixed” as used here signifies that intercepts are nonstochastic

  – <u>NOTE: “Fixed effects” here has different meaning to what it has in multi-level, where it refers to unit effects, usually cross-sectional units. In TSCS language, multi-level fixed effects can be thought of as “random intercepts.”</u>

Characteristics of fixed effects model

- Fixed effects models are used to control for the effect of the different units - <u>essentially they are removing cross unit differences</u>. **This implies that the focus of fixed effects models is the variation over time.**

- Obviously this implies that the suitability of a fixed effects model depends on the number of units (N) we observe and the number of observations (T) of each unit. If T is small, we can't very effectively assess time-serial processes even if we want to; while if T is large, we can.

  – Recall that with small T, TSCS regressions mostly captures the effects of cross-sectional variation simply because there will tend to be less time-serial variation to begin with)

- Thus, if the effect of x on y is *entirely* driven by temporal variation in x, fixed effects estimation will have absolutely no effect on the estimated results.

Characteristics of fixed effects model

- As a consequence, if the effect is driven by both cross-sectional and time-serial variation, estimating using fixed effects will **attenuate** the coefficient for x, depending on the ratio of time-serial to cross-sectional variation. This attenuation of the estimated effect of x is one of the drawbacks of using fixed effects, and especially so as N increases and T decreases, e.g., classic panel data.

- Fixed effects are unbiased but inefficient in small samples (that it is efficient as the T is going to infinity).

Characteristics of fixed effects model

- Furthermore, the independent variable may exhibit extremely minimal variation within each unit – i.e they are slowly changing over time. **If the correlation between the slow moving covariate and the unit fixed effects is high enough, this can destabilize estimates of the effect of the independent variable** (see Plümper and Troeger, 2007, 2011).

- **Even more obvious drawback is that fixed effects will purge time-invariant variables from the model. Thus, any time-invariant x essentially mimic the individual specific constant term.**
  - (for example, if x is a political institutional variable that is the same over time in each country, fixed effects wipes out the explanatory power of the variable)

(This lack of identification is the price of the robustness of the specification to unmeasured correlation between the common effect and the exogenous variables)

Fixed effects model estimation: dummy variables and demeaning

- There are two approaches to estimating fixed effects:

  1. Including dummy variables for cross-sectional units – which we will call intercept (dummy) model
  2. Demeaning all of the variables – which we will call within effect model

- The approaches are identical

  – The second model is more convenient for calculation and it is typically used in statistical software.

- We will firstly focus on intercept model

Intercept (dummy) model

- With least square dummy approach you have various options:

1. The first approach includes all dummies expect one (the reference category)
2. The second approach includes all dummies and, in turn, suppresses the intercept
3. The third approach includes the intercept and all dummies, and then impose a restriction that the sum of parameters of all dummies is zero

Interpretation

1. The first approach reports differences from the reference category
2. In second approach estimates are actual intercepts of groups, making it easy to interpret substantively
3. Third approach computes how far parameter estimates are away from the average group effect

(Keep in mind that the R^2 of second approach is not correct. Because second approach suppresses the intercept, you will get incorrect F and R^2 statistics.)

Intercept model – cross-sections and time

- The least squares dummy variable approach can be easily extended to include a time-specific effect as well.

- This model is obtained from the preceding one by the inclusion of additional $T - 1$ dummy variables.

- In this context it is often assumed that the intercept varies across the N cross-sectional units and/or across the T time periods. In the general case $(N-1)+(T-1)$ dummies can be used for this, with computational short-cuts available to avoid having to run a regression with all these extra variables.

- If one cross-sectional or time-series variable is considered (e.g., country, firm), this is called a **one-way fixed effect** model. **Two-way fixed effect** models have two sets of dummy variables (e.g., state and year).

Two-way fixed effects

- In estimating two-way fixed effects model you may combine the dummy approaches to avoid perfect multicollinearity. There are five strategies when combining the approaches:
  - Drop one cross-section and one time-series dummy variables
  - Drop one cross-section dummy and suppress the intercept; or drop one time-series and suppress the intercept
  - Drop one cross-section dummy and impose a restriction on the time-series dummies; or drop one time-series dummy and impose a restriction on the cross-section dummies
  - Suppress the intercept and impose a restriction on either cross-section or time-series dummies
  - Include all dummy variables and impose two restrictions on the cross-section and time-series dummies
- Each strategy produces different dummy coefficients but returns exactly same parameter estimates of regressors. **The first strategy of dropping two dummies is generally recommended**

Within estimator

- **The within effect model does not use dummy variables, but uses deviations from group means. The model is defined as:**

$$y_{i,t} - \bar{y}_i = (x_{i,t} - \bar{x}_i)\beta + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

- **Thus, this model is the OLS without an intercept.**

- The within effect model has several disadvantages:
  - Since no dummy is used, the within effect model has a larger degree of freedom for error, resulting in a small MSE (mean square error) and incorrect standard errors of parameter estimates. Thus, you have to adjust the standard error using the formula
  - R^2 of the within effect model is not correct because an intercept is suppressed.

Within effect estimator

- Despite its limitations within estimator, with appropriate corrections, is typically used to estimate fixed effects in statistical software. Stata modifies equation:

$$y_{i,t} - \bar{y}_i = (x_{i,t} - \bar{x}_i)\beta + \nu_i + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

and estimates within estimator by running OLS on:

$$y_{i,t} - \bar{y}_i + \bar{\bar{y}} = \alpha + (x_{i,t} - \bar{x}_i + \bar{\bar{x}})\beta + (\varepsilon_{i,t} - \bar{\varepsilon}_i + \bar{\nu}) + \bar{\bar{\varepsilon}}$$

- where $\bar{y}_i = \sum_{t=1}^{Ti} y_{i,t} / T_i$ and similarly $\bar{\bar{y}} = \Sigma_i \Sigma_t y_{it} / (nT_i)$. The conventional covariance matrix of the estimators is adjusted for the extra n-1 estimated means, so results are the same as using OLS on $y_{i,t} = x_{i,t}\beta + \alpha_i + \varepsilon_{i,t}$ to estimate α_i directly.

First difference estimator

- In dealing with cross-sectional unit effects, some scholars just estimate first difference model. Assume equations:

$$y_{i,t} = \alpha_i + x_{i,t}\beta + \varepsilon_{i,t}$$
$$y_{i,t-1} = \alpha_i + x_{i,t-1}\beta + \varepsilon_{i,t-1}$$

- Differencing both equations gets us:

$$y_{i,t} - y_{i,t-1} = \alpha_i - \alpha_i + (x_{i,t} - x_{i,t-1})\beta + (\varepsilon_{i,t} - \varepsilon_{i,t-1})$$
$$\Delta y_{i,t} = \Delta x_{i,t}\beta + \Delta\varepsilon_{i,t}$$

- which obviously removes α_i

- The advantage of the first difference approach is that it removes the latent heterogeneity (α_i) from the model whether the fixed or random effects model is appropriate. Similarly to fixed effects model, first differences allows for assumption that the omitted effects are correlated with the included variables

Disadvantages of first difference estimator

- Under the assumption of spherical errors, i.e. homoscedasticity and no serial correlation in, the fixed effect estimator is more efficient than the first difference estimator. **However, if $\varepsilon_{i,t}$ follows a random walk, the first difference estimator is more efficient**

- Costs of first difference:
  - degrees of freedom;
  - you end up with T – 1 observations, which for large N may be too large.
  - why not differences between t and t-2 or other lags?
  - the differencing also removes any time-invariant variables from the model

- In general, the first difference estimator does not have much to recommend it, save for one very important application. Many studies involve two period “panels,” a before and after treatment. In these cases, the phenomenon of interest may be the change in the outcome variable—the “treatment effect.”

However, differencing has its role in accounting for dynamics
(we will get back to this)

Between Group Effect Model: Group Mean Regression

- One of the useful features of panel data analysis as we are doing here is the ability to analyze the between-groups variation (heterogeneity) to learn about the main regression relationships and the within-groups variation to learn about dynamic effects

- The between effect model uses aggregate information, group means of variables. In other words, the unit of analysis is not an individual observation, but groups or subjects. The number of observations jumps down to N from NT. This group mean regression produces different goodness-of-fits and parameter estimates from those of OLS with dummies and the within effect model.

$$\bar{y}_i = \alpha + \bar{x}_i \beta + \nu_i + \bar{\varepsilon}_i$$

Between Group Effect Model

- Between regression estimator is rarely used in practice, but between estimates are an ingredient in the random effects estimate. Implementation of random effect uses the OLS estimates for this ingredient

- Newly required is that we assume that random effect ν_i and $\bar{x}_i$ are uncorrelated. If random effect ν_i and $\bar{x}_i$ correlated, the estimator can not determine how much of the change in $\bar{y}_i$ associated with an increase in $\bar{x}_i$, to assign to β versus how much to attribute to the unknown correlation.

Fixed effect , between effects, random effects

- **There are trade-offs between the models:**

- If interested in the over-time effects of x on y (or worried about the contaminating effects of cross sectional unit effects), fixed effects is a preferred approach, as it controls for differences between cross-sectional units and thus isolates temporal change.

- If interested in the cross-sectional effects of x on y (or worried about the contaminating effects of temporal "unit" effects), between effects is preferred, as it controls for (fixed) differences across time and so isolates variation across units.

- If interested in both, random effects is preferred, as it controls for neither temporal nor cross-sectional variation and so models the effects of both (we will come back to this in the next class).

Diagnostics

- In general, whatever was assumed by OLS is still valid for TSCS/CSTS data analysis. However, have in mind that TSCS/CSTS data lead to additional complications. Therefore, we will assume:
  - No endogenity
  - No autocorrelation
  - No heteroscedasticity
  - No multicollinearity
  - Linear relation
  - No omitted variables
  - No irrelevant variables
  - No outliers
- One problem with TSCS data is that it is by no means certain that each time series will show the same relationship as the rest, so it is important to conduct “jack-knife” tests in which each unit, e.g., country, in turn is dropped from the dataset and the model re-run to see if the results are the same.

Fixed effects and non-spherical errors

- It might seem appropriate to compute the robust standard errors for the fixed effects estimator. However, in principle, that should be unnecessary. If the model is correct and completely specified, then the individual effects should be capturing the omitted heterogeneity, and what remains is a classical, homoscedastic, non-autocorrelated disturbance

- **Nevertheless, the fixed effects model can be combined with corrected standard errors**

- Thus, if we suspect that there is heteroskedasticity or within-panel serial correlation in the idiosyncratic error term $\varepsilon_{i,t}$, we could use robust ***clustered sandwich estimator.***

Fixed effects and non-spherical errors

- You can also use ***panel corrected standard errors*** (PCSE)

- **Since Beck's and Katz's (1995) article, use of panel corrected standard errors became sort of a standard in dealing with TSCS (i.e. time dominant) data**

- In its normal form, PCSE corrects only for heteroskedasticity and cross-panel correlation

- Naturally, depending of specification you are free to use PCSE based options (i.e. control for various combinations of autocorrelation, heteroskedasticity and cross-panel correlation of errors)