3. Panel data

LPO 8852: Regression II

Sean P. Corcoran

Last update: September 25, 2025

Matching and weighting recap

Matching and weighting approaches seek to construct a comparison group where the conditional independence assumption is satisfied:

$$Y(0), Y(1) \perp\!\!\!\perp D|X$$

That is, conditional on X (or a one-number summary like the propensity score), potential outcomes are independent of treatment status D. If this holds, we can use mean outcomes of the matched or weighted comparison group as a stand-in for the treated group counterfactual.

$$\underbrace{E[Y(0)|D=1,X]}_{\text{unobserved}} = \underbrace{E[Y(0)|D=0,X]}_{\text{matched comparison group}}$$

Last update: September 25, 2025

Matching and weighting recap

Challenges:

- The conditional independence assumption (selection on observables) is strong! In most settings we have to be concerned about selection on *unobservables*.
- Constructing weights or matched samples is somewhat of an art, and results may be sensitive to model specification.
- We are typically comparing outcomes at one point in time (e.g., post "treatment").
- Matching/weighting designed for binary treatments (or at least a small number of categorical treatments). What if the "treatment" is continuous (e.g., class size)?

Panel data: definitions and Stata commands

Panel data

Panel, longitudinal, or *"cross-sectional time series"* data consist of observations on cross-sectional units (e.g., students, schools, hospitals, neighborhoods, counties, states) at multiple points in time.

- N cross-sectional (panel) units and T time periods ($T \geq 2$)
- A *balanced panel* has exactly $N \times T$ observations (T time observations for all N panel units)
- An *unbalanced panel* has T_i observations for panel unit i, where T_i is not the same for all i

Differs from a *pooled cross-section*, although panel data methods can often be used with this type of data.

Panel data - long

Panel data in *long* format, N students in $T = 4$ years:

studentID	year	readscore	mathscore	incomecat	...
1	1999	75	82	3	
1	2000	78	84	4	
1	2001	80	90	4	
1	2002	78	91	3	
2	1999	91	92	2	
2	2000	94	92	2	
2	2001	80	85	2	
2	2002	87	83	2	
3	1999	62	50	5	
3	2000	70	47	5	
3	2001	75	55	4	
3	2002	73	60	5	

Panel data - wide

Panel data in *wide* format, N students in $T = 4$ years:

studentID	read99	math99	inc99	read00	math00	inc00	read01	...
1	75	82	3	78	84	4	80	
2	91	92	2	94	92	2	80	
3	62	50	5	70	47	5	75	
4	...	...	...					

Panel data - reshape long

Moving between *long* and *wide* format in Stata with `reshape`, beginning with *wide* data:

- `i()` contains the time invariant variables (e.g., ID, gender)
- `j()` specifies the time variable to be created (e.g., year)
- The list of time varying variables are "stubs" that end in the *j* suffix

`reshape` **`long`** *`stubnames`*`, i(`*`varlist`*`) j(`*`varname`*`)`

- If `j()` consists of *string* rather than *numeric* values, use the `string` option
- Example time-varying variable names: *score98, score99, score00* (Stata will have problems with 00 as a `j()` value if `string` option is not used).

Panel data - reshape wide

Moving between *long* and *wide* format in Stata with `reshape`, beginning with *long* data:

- `i()` contains the time invariant variables (e.g., ID, gender)
- `j()` specifies the time variable (e.g., year)
- The list of time varying variables are "stubs" that *will* end in the *j* suffix, once converted to wide

```
reshape wide stubnames, i(varlist) j(varname)
```

- After reshaping, Stata allows you to revert back easily without losing information. E.g., after the above command just type `reshape long`
- Most panel regression commands expect the data to be in *long* format.

In-class example 1

Illustration of reshape commands using *Census_states_1970_2000* data:

- Cross sectional unit: state
- Time variable: year (decennial Census years)
- Time-varying variables: median household income, unemployment rate

Stata panel commands

Stata has many useful `xt` commands for working with panel data. Typically these require that you first declare the data to be a panel using `xtset`:

- `xtset` *panelvar timevar*
- The *panelvar* must be numeric. If it is not, you can use encode: **`encode`** *panelvar*, `gen(`*panelvar2*`)`
- It is possible to tell Stata in the `xtset` options what units of time the data represent—e.g., years, quarters, minutes (may be useful for some purposes–I don't usually do this)
- `xtset` alone will report back the current panel settings

Stata panel commands

Other useful Stata panel data commands for description:

- `xtdescribe`—to see patterns of participation/data availability
- `xtsum`—for descriptive statistics that show between- and within-unit variation
- `xttab`—for one-way tabulations with separate counts within and between units
- `xttrans`—for transition probabilities (movement between categories of a categorical variable)
- `xtline` and `xtline, overlay`—for separate line graphs by panel unit (see in-class example)

Stata panel commands

Other useful Stata commands when working with panels:

- `duplicates report` *`varnames`*—to affirm that there is one only observation per unit per time period (e.g., *`id year`*)
- `isid` *`varnames`*—same: affirms that combinations of *varnames* uniquely identify the observations

xtsum

Decomposition of variation in `xtsum`:

$$s_w^2 = \frac{1}{NT-1}\sum_i \sum_t (X_{it} - \bar{X}_i)^2$$

$$s_b^2 = \frac{1}{N-1}\sum_i (\bar{X}_i - \bar{X})^2$$

$$s_o^2 = \frac{1}{NT-1}\sum_i \sum_t (X_{it} - \bar{X})^2$$

Note $\bar{X}$ is the *grand mean* of X. Can also write:

$$s_w^2 = \frac{1}{NT-1}\sum_i \sum_t (X_{it} - \bar{X}_i + \bar{X})^2$$

because adding a constant ($\bar{X}$) will not affect s_w^2. (This formulation is used below).

xtdescribe and xtsum

```
. xtdescribe

classroom:  1, 2, ..., 3                                   n =          3
studentid:  1, 2, ..., 3                                   T =          3
           Delta(studentid) = 1 unit
           Span(studentid)  = 3 periods
           (classroom*studentid uniquely identifies each observation)

Distribution of T_i:   min      5%     25%      50%      75%     95%     max
                         3       3       3        3        3       3       3

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
        3    100.00  100.00 |  111
 ---------------------------+---------
        3    100.00         |  XXX

. xtsum y

Variable         |      Mean   Std. dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
y        overall |  48.33333   15.05822         14         64 |     N =       9
         between |              12.2202         35         59 |     n =       3
         within  |             10.71214   27.33333   64.33333 |     T =       3
```


xtsum

`xtsum` shows the min and max of:

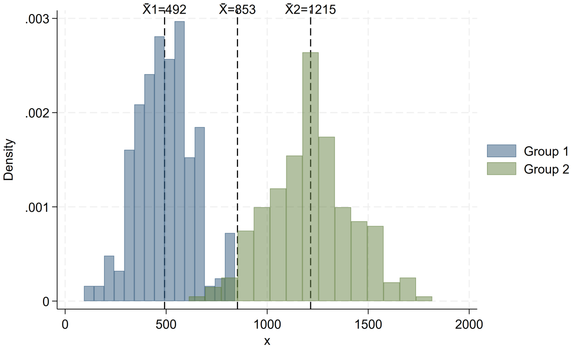
- X_{it}: overall
- $\bar{X}_i$: between
- $(X_{it} - \bar{X}_i + \bar{X})$: within

Note: on `xtsum`, see also `https://www.stata.com/support/faqs/statistics/decomposed-variances-in-xtsum/`

- Between and within variation do not exactly sum to overall
- One reason: they are bias corrected (multiplied by $\sqrt{n/n-1}$)
- With unbalanced panels, s_b is calculated using mean of panel means, not $\bar{X}$ (may not be the same)

xtsum

Simulated data, two groups (N=250 each): within vs. between variation



In-class example 2

Illustration of `xt` commands using *State_school_finance_panel* data:

- Cross sectional unit: state
- Time variable: year (annual 1990-2010)
- Time-varying variables: various school finance measures

Panel data and unobserved heterogeneity

Panel data - advantages

Why use panel data?

- Can help us answer questions not possible with a cross-section or time-series approach
- Can generate *measures* not possible with cross-sectional or time series data (e.g., growth, work spells, turnover)
  - If 50% of women are working in year t, does this reflect 50% of women working at any given point, or 50% of women who work all the time?
- Allows us to address selection bias due to unobserved heterogeneity that is fixed over time ("fixed effects")

Selection bias revisited

Lecture 1: interpretation of regression coefficients as causal is often complicated by selection bias. Example:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

with $E(u_i|X_i) \neq 0$ because we believe potential outcomes are not independent of X. We can attempt to mitigate selection bias through the inclusion of additional covariates or matching/weighting, but this only solves the problem if adjusting for these observables eliminates OVB.

In practice we are often more concerned about selection on *unobservables*.

Unobserved heterogeneity

Suppose there are **unobserved, fixed differences** across units (C_i) that affect the outcome and are (potentially) correlated with the explanatory variable of interest (X_i):

$$Y_i = \beta_0 + \beta_1 X_i + C_i + u_i$$

C_i could represent the effects of ability, health, motivation, intelligence, county of birth, parental resources, managerial quality, organizational culture, state/local policies or regulations, etc.

There are several analytic strategies that exploit panel data to remove the bias associated with C_i.

First difference model

First difference model

Suppose we have two time periods (T=2) for each cross-sectional unit i, and assume the linear model above applies in both periods:

$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + C_i + u_{i2}$$
$$Y_{i1} = \beta_0 + \beta_1 X_{i1} + C_i + u_{i1}$$

Now subtract period 1 from period 2 for the "first difference":

$$\Delta Y_i = \beta_1 \Delta X_i + \Delta u_i$$
$$Y_i^* = \beta_1 X_i^* + u_i^*$$

Because C_i is time-invariant, it <u>differences out</u> of the model. Notice the constant β_0 also differences out.

First difference model

β_1 can be identified from the first difference model using OLS as long as the usual OLS assumptions apply to it:

- The new error term $u_i^* = \Delta u_i$ (change in the error over time) is uncorrelated with the new explanatory variable, $X_i^* = \Delta X_i$ (change in the observed X over time).
- This requires that we have no cross-period correlations between u and X: this is called **strict exogeneity**.
- The X_i must vary over time for at least some i, else they difference out too.

In-class example 3

Example using panel of Texas elementary schools:

- `use` *Texas_elementary_panel_2004_2007.dta*
- `xtset` *campus year*
- `xtdescribe`
- `rename` *ca311tar avgpassing*
- `egen` *avgclass* `= rowmean(`*cpctg01a-cpctgmea*`)`
- `reg` *avgpassing avgclass* `if` *year*==2007 (cross-sectional regression for 2007)

Note: *avgclass* is the mean class size across grades in a school, and *avgpassing* is the school average passing rate across all grades and subjects.

In-class example 3

Having declared the dataset as a panel, Stata recognizes the **`d.`** prefix as a "difference operator":

- `reg` *d.avgpassing d.avgclass* `if` *year*`==2007, noconstant`
- This is the first difference regression, using 2007 only (and its lag in the calculation of *d.avgpassing* and *d.avgclass*)
- `d.` can be used after `xtset` or `tsset` (time series set)
- Note suppression of the constant. In theory the constant term differences out (see earlier slide). In practice there is no harm in estimating with a constant, which would capture a year-on-year trend.

In-class example 3

A few things to note in example 3:

- Change in coefficient on class size: does it make sense?
- Change in sample size (re: unbalanced panel due to missing values)

A few things to think about:

- Is strict exogeneity likely to hold in this circumstance?
- Where is the identifying variation coming from?
- How much variation is there in the *change* in passing rates (ΔY) and class size (ΔX)?
- Do outliers dominate the variation in changes?

```
gen davgpassing = d.avgpassing
/* create variable containing FD that can be described */
```

In-class example 3

The first difference model is easily generalizable to multiple years ($T > 2$).

- Each year of data is differenced with the prior year
- 1st period is sacrificed
- Must continue to think about OLS assumptions, e.g. strict exogeneity

```
reg d.avgpassing d.avgclass, noconstant
table year if e(sample)
* note 1st year of data is not used
```


Fixed effects or “within” model

One-way fixed effects

Alternatively, in the *(one-way) fixed effects* model, we could treat the C_i as parameters to be estimated (the "fixed effects"):

$$Y_{it} = \beta_0 + \beta_1 X_{it} + C_i + u_{it}$$

Effectively we are allowing for a *unique intercept* for every cross-sectional unit i. This is feasible to estimate since each i is observed multiple times.

Note on notation: in practice we cannot estimate β_0 and intercepts for all the N cross-sectional units (re: collinearity). Often you will see fixed effects models (more precisely) written without the β_0 (next slide).

One-way fixed effects: notation

You may see the fixed effects model written something like this:

$$Y_{it} = \beta_1 X_{it} + \sum_{j=1}^{n} \mathbf{1}(i = j)\gamma_i + u_{it}$$

Where $\mathbf{1}(i = j)$ is an indicator function that equals 1 when $i = j$. The n γ_is are the fixed effect coefficients to be estimated. More often, it is written more concisely as:

$$Y_{it} = \beta_1 X_{it} + \gamma_i + u_{it}$$

where it is understood that the γ_i represent the n fixed effect coefficients to be estimated.

“Least squares dummy variables” estimation

Now we are estimating β_0, β_1, and $(N-1)$ fixed effect coefficients. This can be done by including $N-1$ dummy variables in the regression, one for each cross sectional unit (omitting one). This is sometimes called the “least squares dummy variable” (LSDV) model:

- `reg` *`avgpassing avgclass`* `i.`*`campus`*
- For this example limit to *year*>=2006 and *houston*==1 so that the number of schools is manageable.

Note omission of first cross-sectional unit effect with `i.campus`. Interpretation of the $N-1$ fixed effects are relative to this (arbitrary) omitted unit. You can control which unit effect is omitted if desired. See in-class example.

“Least squares dummy variables” estimation

There are a number of reasons why you might not want to do it this way:

- Could be time-consuming and harder on memory with very large datasets (re: you are creating dummy variables for each unique i)
- Soaks up degrees of freedom; may result in the number of regressors exceeding the number of observations
- Often we are not interested in the estimates of the fixed effects themselves, so there is no need to see/report them.
- Exception: recent “school effects” and “teacher effects” studies work explicitly with fixed effects estimates ($\hat{C}_i$)—but more on this later

Within transformation

Suppose that panel data are available with multiple observations per i and the model is:

$$Y_{it} = \beta_1 X_{it} + C_i + u_{it} \quad t = 1, ..., T \quad \forall i$$

Within each panel unit i, take the average over t on both sides and subtract the unit average from each it observation:

$$\bar{Y}_i = \beta_1 \bar{X}_i + C_i + \bar{u}_i$$

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i)$$

This is called "de-meaning" or the "within" transformation (sometimes denoted $\ddot{Y}_i$) or "absorbing" the fixed effect. Notice that the intercepts C_i "difference out." This can only happen if the C_i are truly *time invariant*.

Within transformation

Under certain assumptions, an OLS regression of the de-meaned Y on the de-meaned X will yield unbiased and consistent estimates of β_1.

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i)$$

$$Y_{it}^* = \beta_1 X_{it}^* + u_{it}^*$$

This is also known as the fixed effects or "within" regression, and extends to more than one explanatory variable $(X_1, ..., X_k)$.

Explanatory variables X_j that are time *invariant* fall out of the model. (They all equal their within-group mean, so the within-transformation equals zero). Common examples: gender, race or ethnicity, birthplace...

Within transformation (FWL)

Another way to see the equivalency between LSDV and the within transformation is by applying the Frisch-Waugh-Lovell (FWL) theorem. FWL says the estimated coefficient on X_1 in a regression of Y on X_1 and X_2 will be the same as the coefficient resulting from the following:

1. Regress X_1 on X_2 and get the residuals
2. Regress Y on X_2 and get the residuals
3. Regress the residuals in (2) on the residuals in (1)

Intuition: the residuals in (1) represent the part of X_1 that is unexplained by X_2. The residuals in (2) represent the part of Y that is unexplained by X_2. Part (3) is the bivariate relationship between "what's leftover".

Within transformation (FWL)

Suppose we have the following one-way fixed effects model with one covariate X:

$$Y_{it} = \beta_1 X_{it} + \gamma_i + u_{it}$$

Regress X_{it} on a set of dummy variables (D_i) for the i units:

$$X_{it} = \sum_{i=1}^{n} D_i \alpha_i + v_{it}$$

The $\hat{\alpha}_i$ are just the unit means $\bar{X}_i$. So the residuals $\hat{v}_{it}$ are:

$$\hat{v}_{it} = X_{it} - \bar{X}_i$$

These are the "part of X that is unexplained by the fixed effects." They are within variation.

Within transformation (FWL)

Regress Y_{it} on a set of dummy variables (D_i) for the i units:

$$Y_{it} = \sum_{i=1}^{n} D_i \delta_i + w_{it}$$

The $\hat{\delta}_i$ are just the unit means $\bar{Y}_i$. So the residuals $\hat{w}_{it}$ are:

$$\hat{w}_{it} = Y_{it} - \bar{Y}_i$$

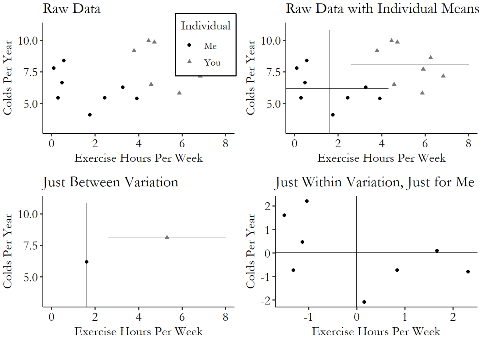
These are the "part of Y that is unexplained by the fixed effects." So, regressing $\hat{w}_{it}$ on $\hat{v}_{it}$ means you are regressing:

$$Y_{it} - \bar{Y}_i = \beta_1 (X_{it} - \bar{X}_i) + u_{it}$$

You will get the same estimate for β_1!

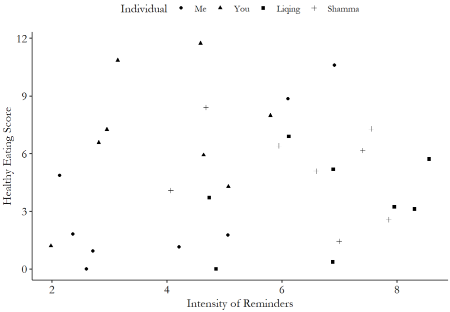
Within transformation: illustrated

From Huntington-Klein chapter 16: relationship between exercise and colds



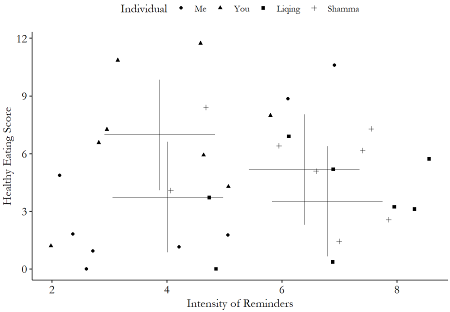
Within transformation: illustrated

Huntington-Klein Figure 16.4: intensity of reminders and healthy eating score, raw data ($r = 0.111$)



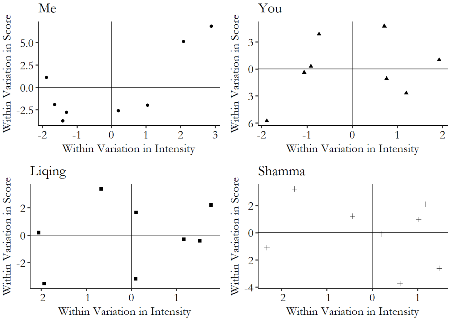
Within transformation: illustrated

Huntington-Klein Figure 16.5: four individuals differ in their mean y and x. If we use only the *between* variation, there is a clear negative relationship ($r = -0.440$)



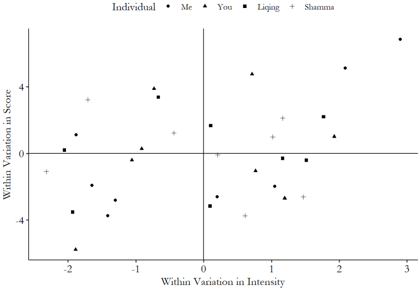
Within transformation: illustrated

Huntington-Klein Figure 16.6: remove between variation and focus on within. Data plotted again with person means centered at zero.



Within transformation: illustrated

Huntington-Klein Figure 16.7: overlay the "within" plots to see all of the within variation. The correlation is more positive ($r = 0.363$).



See also the animated gif *FEanimation* on Github

Fixed effects or “within” model in Stata

Within transformation: `xtreg`

The fixed effects or “within” regression model can be estimated using OLS using `xtreg`:

- `xtreg` *avgpassing avgclass*`, fe`
- Note `xtset` must have been declared, or specify the cross-sectional unit in the options, e.g. `i(`*campus*`)`
- While the fixed effects are not estimated directly, can “back out” a prediction: $\hat{C}_i = \bar{Y}_i - \bar{X}_i\hat{\beta}_1$
- `predict` *schlfe*`, u`

Note the estimated fixed effects from `xtreg` are not the same as the dummy coefficients from the LSDV model. See the do file *Simulated panel data* for an illustration.

Within transformation: areg

The `areg` command also absorbs fixed effects by de-meaning within group:

- `areg` *`avgpassing avgclass`*, `absorb(`*`campus`*`)`
- As with `xtreg` the fixed effect coefficients are not estimated/reported
- Here again can "back out" the fixed effects using `predict`. For example: `predict` *`schlfe`*`, d`
- The intercept is not specific to any omitted group, but rather the intercept that makes the prediction calculated at the means of the X equal to $\bar{Y}$

See also new `absorb()` option in Stata 19 regression commands.

Within transformation: areg

- `areg` is intended for applications where the number of cross-sectional units or categories is fixed (e.g., states) and does not grow as the sample size gets larger. Many panel applications are the latter, so use of `xtreg` is advised. See later notes on standard error calculations.
- The R^2, $RMSE$, standard errors, etc., in `areg` are the same as the LSDV using `reg`. The model F statistic is different, however. It only tests the joint significance of the X (excluding the fixed effects).

Within transformation: transforming the dataset

The command `xtdata, fe` can be used to transform your dataset using the within-transformation. However, this is rarely done (in my experience) since it transforms *all* of the variables in your dataset. `xtreg` will do the transformation on the fly without altering your dataset.

Fixed effects model in Stata

Compare `xtreg`, `areg` and first difference when $T = 2$

```
xtreg avgpassing avgclass3, fe
```

```
areg avgpassing avgclass3, absorb(campus)
```

```
reg d.avgpassing d.avgclass3, noconstant
```

Also compare to `regress` with `absorb`—new option in Stata:

```
reg avgpassing avgclass3, absorb(campus)
```

Fixed effects model in Stata

A few notes about `xtreg, fe`

- FE is more efficient (smaller standard errors) than first differencing if the error terms are serially uncorrelated and $T > 2$
- Assumes no correlation in u across units of panel i (some tests for this using user-written `xtscd, xttest3`)
- The estimates of the fixed effects themselves (C_i) are unbiased but inconsistent in large samples. (Why? As the number of panel units grows ($N \to \infty$) the number of parameters to estimate also grows).
- `xtreg` has not historically allowed `svy` specification (for complex sampling designs) but can use `pweights` and `cluster()` option. See also the `mixed` (or `xtmixed`) command for an alternative.

Fixed effects model in Stata

Stata actually fits the following model with `xtreg`:

$$(Y_{it} - \bar{Y}_i + \bar{Y}) = \beta_0 + \beta_1(X_{it} - \bar{X}_i + \bar{X}) + (u_{it} - \bar{u}_i + \bar{u})$$

Where the values with a bar but no subscript are the grand means. This includes an intercept which is the average of the fixed effects (C_i).

Fixed effects model in Stata

Other useful output from `xtreg`:

```
.
. xtreg avgpassing avgclass3, fe

Fixed-effects (within) regression               Number of obs      =       350
Group variable: campus                          Number of groups   =       180

R-sq:  within  = 0.0039                         Obs per group: min =         1
       between = 0.0087                                        avg =       1.9
       overall = 0.0032                                        max =         2

                                                F(1,169)           =      0.66
corr(u_i, Xb)  = -0.1079                        Prob > F           =    0.4179

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   avgclass3 |  -.2390704    .294385    -0.81   0.418     -.820216    .3420752
       _cons |   76.80355   6.032673    12.73   0.000     64.89445    88.71265
-------------+----------------------------------------------------------------
     sigma_u |  11.942612
     sigma_e |  7.1632112
         rho |   .7354221   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(179, 169) =     5.27             Prob > F = 0.0000

.     * fixed effects model
```


Fixed effects model in Stata

Other useful output from `xtreg`:

- F-test for joint significance of fixed effects (null hypothesis H_0 is that all fixed effects are zero). If rejected, fixed effects model is a reasonable assumption and regular OLS may provide inconsistent estimates. In practice, rarely rejected.
- R^2 *within*: variance "explained" by within-group deviations from group means
- R^2 *between*: variance in group means $\bar{Y}_i$ "explained" by the group mean X's: $\bar{X}_i$
- *sigma_u* estimate of the standard deviation in fixed effects (C_i)

Fixed effects model: assumptions

Fixed effects model: assumptions for inference

- **FE.1:** linear model $Y_{it} = \beta_1 X_{it1} + ... + \beta_k X_{itk} + C_i + u_{it}$
- **FE.2:** cross-sectional units are a random sample
- **FE.3:** X_{it} varies over time for some i, no perfect collinearity
- **FE.4:** $\forall t$, $E(u_{it}|X_i, C_i) = 0$ or the expected value of u given X in *all* time periods is zero (strict exogeneity)
- **FE.5:** $Var(u_{it}|X_i, C_i) = Var(u_{it}) = \sigma_u^2$ - homoskedasticity
- **FE.6:** for $t \neq s$ errors are uncorrelated: $Cov(u_{it}, u_{is}|X_i, C_i) = 0$. No serial correlation.

Under FE.1-FE.4, fixed effects model (and first difference model) is unbiased. Adding FE.5-FE.6, fixed effects model is BLUE. If FE.6 holds, fixed effects is more efficient than the first difference model. Can relax homoskedasticity assumption and calculate `robust` standard errors.

Fixed effects model: assumptions for inference

Note: the econometric theory described here is for "short" panels, with N large relative to T. If the opposite is true in your context, use FE model with caution (see Wooldridge chapter 14, Cameron & Trivedi).

Fixed effects model: practical considerations

Fixed effects considerations:

- Where is the identification coming from?
- How much variation is there within panel units? When small, one risks imprecise estimates
- For stats on within- and between- school variation can use `xtsum` (described earlier), `xttab` and `xttrans` for categorical variables
- Example:

```
xtsum avgpassing avgclass
```

Panel data models: on standard errors

With panel data it is hard to assume that errors are independent across observations. You have multiple observations from the same cross-sectional unit and there is likely to be something omitted from the model that is correlated across observations.

- Should almost always use clustered standard errors with FE (at the same level as FE): `vcecluster` *`varname`*
- Note `areg` by default uses traditional OLS standard errors, although this can be adjusted
- With `xtreg`, using `robust` standard errors is equivalent to clustering on the cross-sectional unit, since Stata takes the panel structure into account.

Two- and multi-way fixed effects models

Two-way fixed effects model

The two-way fixed effects model adds another dimension of fixed effects (often time periods). Notation is analagous to the one-way model shown earlier:

$$Y_{it} = \beta_1 X_{it} + \gamma_i + \theta_t + u_{it}$$

where it is understood that the γ_i represent the n unit fixed effect coefficients to be estimated and the θ_t represent the T time effect coefficients to be estimated.

Two-way fixed effects model

There is no explicit command for two-way models, rather can just include dummies (e.g., for each time period).

```
xtreg avgpassing avgclass i.year, fe
* the i.year syntax introduces (T-1) time effects
test _Iyear_2006 _Iyear_2007
* joint test that time effects = 0
```

Alternatively, could use `reghdfe` or new `regress` with `absorb`

```
reghdfe avgpassing avgclass, absorb(campus year)
reg avgpassing avgclass, absorb(campus year)
```

As with one-way fixed effects model, requires variation across units within time periods t.

Two-way fixed effects model

The generalized difference-in-differences model is a two-way fixed effects model (covered in Lecture 4):

$$Y_{it} = \beta_0 + \beta_1(treat_i \times post_t) + \theta_t + \gamma_i + \delta X_{it} + u_{it}$$

There are cross-sectional unit fixed effects (γ_i) which represent separate intercepts for each unit and time effects (θ_t) which represent common variation over time within group.

Estimation with multiple fixed effects

Above we used `xtreg` or `areg` with dummy variables for a second "fixed effect" (time). This works fine if your second fixed effect has a reasonable number of categories. What if all of your fixed effects have a large number of groups?

Example: regressing measure of between-country trade in a given sector on a measure of tariffs in place, with fixed effects for country, industry, and year. With 40 years, 160 countries, and 1,000 industries, there would be 1,200 parameters to estimate (and the output would be overwhelming). `areg` could absorb one of these (e.g., industry) but not the others.

In Stata:

```
reghdfe trade tariff, absorb(industry country year)
reg trade tariff, absorb(industry country year)
```

Estimation with multiple fixed effects

What is happening behind the scenes here? With one-way fixed effects, `areg` and `xtreg` do a simple within- transformation. By FWL:

- Residualize Y: find the part of Y that is unexplained by the fixed effects
- Residualize X: find the part of X that is unexplained by the fixed effects
- This amounts to de-meaning Y and X

With multiple fixed effects the idea is the same, but it turns out this is more complicated to carry out.

Estimation with multiple fixed effects

We want to find the part of Y and X that are orthogonal to all of the fixed effects. You could residualize one FE at a time, but the second residualization will "undo" part of the first one.

`reghdfe` and `regress` with `absorb` use an iterative procedure that repeats residualization until the residuals converge to zero. These commands use advanced algorithms to do this very quickly.

More on fixed effects designs

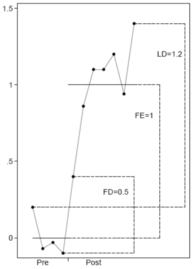
Comparison of models

It is important to be attentive to where the variation in each type of FE model is coming from:

- Fixed effects ("within") model: uses deviations from group (i) means, e.g., mean "pre" vs. mean "post"
- First differences model: uses variation in successive time periods, e.g., just prior to and just after a "treatment" (a change in x)
- **Long differences** is like first differences, but there is a long time span between observations. Here outcomes may be compared well before and well after a "treatment".

To evaluate these in your situation, need some idea of the speed in which X affects Y

Comparison of models



Source: Nichols (2007). Figure shows one panel (i)'s pre to post change following a treatment ($t > 4$). Notice the within-panel unit change differs depending on whether FE, FD, or LD is used.

Fixed effects models in other applications

Fixed effects models are not exclusively used with panel data in which cross-sectional units i are observed in multiple time periods. They are also used with grouped or clustered data. For example:

- Family fixed effects, where the family is the cross-sectional unit and siblings are the group members (akin to the time dimension)
- School fixed effects with student-level data, where each school has its own intercept
- Individual fixed effects within person-level data.

Fixed effects models in other applications

Fixed effects models are not exclusively used with panel data in which cross-sectional units i are observed in multiple time periods. They are also used with grouped or clustered data. For example:

- Family fixed effects, where the family is the cross-sectional unit and siblings are the group members (akin to the time dimension)
- School fixed effects with student-level data, where each school has its own intercept
- Individual fixed effects within person-level data.

Fixed effects models in other applications

The researcher needs to provide a convincing rationale for why the unobserved confounding variable should be considered fixed across multiple observations (e.g., siblings, years)

- Why did a mother's employment status change between siblings?
- Why did only 1 of 2 siblings participate in Head Start?
- Why did a student switch from a traditional school to a charter school?
- Why did an elementary school receive a new principal?

Fixed effects models: advantages and disadvantages

Advantages:

- Unobserved C_i can be correlated with the explanatory variables
- Slopes estimated using *within-group* (i) variation in X, Y

Disadvantages:

- Cannot estimate slope coefficients for time-invariant X
- Fixed effects "remove" a lot of the variation in Y
- The "within" model is less efficient (higher standard errors)
- There may be more measurement error (and attenuation bias, see Lecture 7) when relying on within-group *changes* vs. levels
- Group intercepts use up a lot of degrees of freedom

Jack et al., (2023)

Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts

- Data: district level test scores (passing rates) for 2016-2019 and 2021, plus district demographic data and other county-level controls.
- Instruction mode from the COVID-19 School Data Hub (public): schooling mode classified as in person, remote, or hybrid.

The authors are interested in the effects of schooling mode (X) on passing rates (Y). What would be problematic about estimating a cross-sectional regression with these data for 2021? What might the advantages of a panel be?

Jack et al., (2023)

TABLE 2—PAIRWISE CORRELATIONS BETWEEN IN-PERSON LEARNING ON DISTRICT DEMOGRAPHIC AND PANDEMIC VARIABLES

	Correlation (no fixed effects)		Correlation (state fixed effects)		Correlation (commute zone fixed effects)	
Previous pass rate	0.440	(0.066)	0.611	(0.062)	0.598	(0.053)
Share Black	−0.465	(0.039)	−0.752	(0.043)	−0.757	(0.041)
Share Hispanic	−0.442	(0.067)	−0.328	(0.063)	−0.296	(0.061)
Share FRPL	−0.160	(0.048)	−0.255	(0.048)	−0.365	(0.046)
Share ELL	−1.290	(0.121)	−0.879	(0.104)	−0.764	(0.099)
Avg. case rate	0.803	(0.199)	0.367	(0.107)	0.115	(0.051)
Repub. vote share	0.010	(0.000)	0.010	(0.000)	0.009	(0.001)

Notes: This table shows the pairwise correlations of the share of days in person during the 2020–2021 school year with district demographic and pandemic characteristics. We present the correlations of the sample overall, without fixed effects included ("no fixed effects"), with state-year fixed effects ("state fixed effects"), and with commuting zone fixed effects ("commute zone fixed effects"). The share in person measures the share of time during the 2020–2021 school year that the district offered full-time in-person instruction (rather than hybrid or virtual instruction). "Previous pass rate" represents the average pass rate on state

Jack et al., (2023)

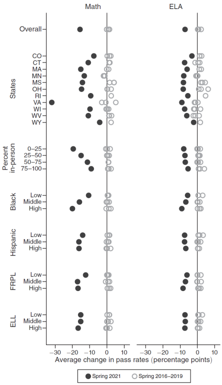


FIGURE 1. AVERAGE CHANGE IN PASS RATES ON STATE STANDARDIZED ASSESSMENTS IN SPRING 2021

Jack et al., (2023)

They use a fixed effects regression:

$$pass_{ict} = \alpha + \beta_1(\%inperson_{it}) + \beta_2(\%hybrid_{it}) + \gamma_{ct} + \delta_t + \nu_i + \Pi X_{ict} + u_{ict}$$

- i is school district, t is year, and c is county
- $\%inperson$ and $\%hybrid$ are the percentage of time the district spent in person and in hybrid modes
- δ_t are the time fixed effects (year)
- ν_i are the school district fixed effects
- The γ_{ct} are effects for county specific trends—this is a twist on the time FE that allow you to capture separate local time trends

Jack et al., (2023)

TABLE 3—SCHOOLING MODE AND CHANGES IN PASS RATES

	Math			ELA		
	(1) Pass rate	(2) Pass rate	(3) Pass rate	(4) Pass rate	(5) Pass rate	(6) Pass rate
Panel A. Main specifications						
% in-person	0.140	0.134	0.128	0.0813	0.0828	0.0872
	(0.0137)	(0.0147)	(0.0156)	(0.0102)	(0.0105)	(0.0105)
% hybrid	0.0776	0.0722	0.0743	0.0608	0.0537	0.0637
	(0.0143)	(0.0148)	(0.0161)	(0.0116)	(0.00949)	(0.00994)
Observations	11,041	11,041	11,041	11,064	11,064	11,064
Commute zone × year	No	Yes	No	No	Yes	No
County × year	No	No	Yes	No	No	Yes
Panel B. Demographic interactions						
% in-person × 2021	0.0960	0.156	0.0872	0.0686	0.0784	0.0729
	(0.0174)	(0.0196)	(0.0388)	(0.0138)	(0.0123)	(0.0276)
% hybrid × 2021	0.0379	0.0907	0.0280	0.0381	0.0409	0.0360
	(0.0169)	(0.0205)	(0.0388)	(0.0129)	(0.0123)	(0.0279)
% Black × % in-person × 2021	0.0943			0.0193		
	(0.0398)			(0.0240)		
% Black × % hybrid × 2021	0.0855			0.0508		
	(0.0472)			(0.0279)		
% Hispanic × % in-person × 2021		−0.135			−0.0178	
		(0.0680)			(0.0482)	
% Hispanic × % hybrid × 2021		−0.0664			0.0247	
		(0.0734)			(0.0421)	
% FRPL × % in-person × 2021			0.0810			0.000259
			(0.0582)			(0.0371)
% FRPL × % hybrid × 2021			0.0689			0.0153
			(0.0605)			(0.0380)
Observations	11,041	11,041	9,620	11,064	11,064	9,643
Commute zone × year	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the relationship between district in-person share, hybrid share, and pass rates in math and ELA on state standardized assessments for students in grades 3–8. Virtual share is the reference group. In panel A, we present our results for state-year fixed effects in columns 1 and 4 for math and ELA, respectively, for commuting zone–year fixed effects in columns 2 and 5, and for county-year fixed effects in columns 3 and 6. All regressions are weighted by district enrollment and include district fixed effects, year fixed effects, state-year fixed effects, demographic controls (race/ethnicity shares, share of students eligible for FRPL, and share of ELLs), county-level

Bonus slides: random effects models

Random effects

An alternative conceptualization of C_i is as a *random* effect, uncorrelated with X_{it}.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \underbrace{C_i + e_{it}}_{v_{it}}$$

Think of v_{it} as a *composite* error consisting of a between-group component (C_i) common to all observations within the group, and a within-group component (e_{it}). It is assumed C_i and e_{it} are independent of one another and:

$$C_i \sim N(0, \sigma_c^2)$$
$$e_{it} \sim N(0, \sigma_e^2)$$

This is the **random effects** or **random intercepts** model.

Random effects

If C_i is uncorrelated with X_{it}, then the composite error term v_{it} is uncorrelated with X_{it}. (We already assume e_{it} is uncorrelated with X_{it}). This means the OLS estimator for β_1 will be unbiased and consistent.

Note: estimation of this model does *not* involve estimating the C_i's as parameters as was done in the LSDV model. These are now part of the error term.

Random effects

The composite error term v_{it} is not, however, i.i.d.:

$$Corr(v_{it}, v_{is}) = \rho = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_e^2} \text{ for } s \neq t$$

The common error for observations in group i (C_i) results in correlation between the composite error in period t (v_{it}) and in period s (v_{is}).

This means OLS is consistent but not efficient, and that traditional standard error formulas assuming i.i.d. errors are incorrect. The ratio above (ρ) is called the **intra-class correlation**.

Estimation using GLS (details later): `xtreg, re`.

Success for All example

- Success for All is a whole-school literacy intervention.
- Borman et al. (2005) conducted a randomized evaluation of SFA in 2001-02 and 2002-03 (21 treatment schools and 20 control).
- This was a *cluster-randomized* design with randomization at the school level.
- The data used by Murnane & Willett (*ch7_sfa.dta*) include grade 1 only. The outcome of interest is *wattack*, the student's score on a "Word-Attack" test.

Next slide: an "unconditional" model with no X_{it} estimates variance components σ_c^2 and σ_e^2 and the intra-class correlation ρ. This provides a sense of the degree of between- vs. within-group variance in the outcome.

Random effects with xtreg

```
. xtreg wattack, re i(schid)

Random-effects GLS regression                   Number of obs     =      2,334
Group variable: schid                           Number of groups  =         41

R-sq:                                           Obs per group:
     within  = 0.0000                                         min =         10
     between = 0.0000                                         avg =       56.9
     overall = 0.0000                                         max =        134

                                                Wald chi2(0)      =          .
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =          .

------------------------------------------------------------------------------
     wattack |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   477.5356   1.447118   329.99   0.000     474.6994    480.3719
-------------+----------------------------------------------------------------
     sigma_u |  8.8705267
     sigma_e |  17.725757
         rho |  .20027618   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

This example: Success for All impact evaluation (from Murnane & Willett). $\sigma_c^2 = 8.87^2 = \mathbf{78.7}$ and $\sigma_e^2 = 17.73^2 = \mathbf{314.35}$. $\rho = 0.200$.

loneway

`loneway` (one-way ANOVA) is another handy command for estimating variance components and ICC. (Note the difference in σ_c and ρ from `xtreg, re`. With unbalanced panels, these will differ slightly).

```
. loneway wattack schid

           One-way Analysis of Variance for wattack: word attack posttest

                                          Number of obs =        2,334
                                              R-squared =       0.2185

    Source                SS         df      MS            F     Prob > F
-------------------------------------------------------------------------
Between schid         201450.43      40   5036.2607     16.03     0.0000
Within schid          720466.21   2,293   314.20244
-------------------------------------------------------------------------
Total                 921916.63   2,333   395.16358

         Intraclass       Asy.
         correlation      S.E.       [95% Conf. Interval]
         ------------------------------------------------
            0.20993     0.04402       0.12366     0.29621

         Estimated SD of schid effect               9.137203
         Estimated SD within schid                  17.72576
         Est. reliability of a schid mean            0.93761
              (evaluated at n=56.56)
```


Random effects with `xtreg`

```
. xtreg wattack sfa ppvt, re i(schid)

Random-effects GLS regression                   Number of obs     =      2,334
Group variable: schid                           Number of groups  =         41

R-sq:                                           Obs per group:
     within  = 0.1101                                         min =         10
     between = 0.3960                                         avg =       56.9
     overall = 0.1820                                         max =        134

                                                Wald chi2(2)      =     308.21
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     wattack |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sfa |   3.440921   2.297268     1.50   0.134    -1.061642    7.943485
        ppvt |   .4851754   .0278075    17.45   0.000     .4306737    .5396771
       _cons |   432.0475   2.972263   145.36   0.000      426.222     437.873
-------------+----------------------------------------------------------------
     sigma_u |  6.9082397
     sigma_e |  16.725172
         rho |  .14574141   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

This regression: includes the treatment indicator (*sfa*) and one covariate (*ppvt*). Note changes in σ_c and σ_e, ρ. The residual variability is reduced with the inclusion of X's.

Random effects

Class size and passing rates in TX class size example:

```
. xtreg avgpassing avgclass, re i(campus)

Random-effects GLS regression                   Number of obs     =     16,062
Group variable: campus                          Number of groups  =      4,326

R-sq:                                           Obs per group:
     within  = 0.0018                                         min =          1
     between = 0.0098                                         avg =        3.7
     overall = 0.0060                                         max =          4

                                                Wald chi2(1)      =       2.74
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0978

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0442893   .0267548    -1.66   0.098    -.0967277    .0081491
       _cons |   76.21828   .5503649   138.49   0.000     75.13959    77.29698
-------------+----------------------------------------------------------------
     sigma_u |  12.391941
     sigma_e |  6.4870883
         rho |  .78490199   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```


Random effects

Compare to fixed effects: very different slope coefficient estimate.

```
. xtreg avgpassing avgclass, fe i(campus)

Fixed-effects (within) regression               Number of obs     =     16,062
Group variable: campus                          Number of groups  =      4,326

R-sq:                                           Obs per group:
     within  = 0.0018                                         min =          1
     between = 0.0098                                         avg =        3.7
     overall = 0.0060                                         max =          4

                                                F(1,11735)        =      21.30
corr(u_i, Xb)  = -0.1189                        Prob > F          =     0.0000

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.1339024   .0290105    -4.62   0.000    -.1907678   -.0770371
       _cons |   78.09211   .5590819   139.68   0.000     76.99621      79.188
-------------+----------------------------------------------------------------
     sigma_u |  12.997022
     sigma_e |  6.4870883
         rho |  .80056238   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(4325, 11735) = 13.83               Prob > F = 0.0000
```

Random vs. fixed effects

- The RE model is biased and inconsistent if the FE assumptions are more appropriate (correlation between X_{it} and C_i).
- If the RE assumptions hold (no correlation between X_{it} and C_i), both RE and FE are *consistent*. They should give "similar" answers in large samples, but the FE model will be *inefficient* (larger standard errors).
- A sufficiently large difference in point estimates suggests the FE assumption is probably correct and RE is inconsistent.
- The **Hausman test** is a formal test of this.

Hausman test

First use `estimates store` to save your `fe` and `re` estimates. Name them FE and RE, for example.

```
xtreg avgpassing avglcass, fe i(campus)
estimates store FE
xtreg avgpassing avgclass, re i(campus)
estimates store RE
hausman FE RE
```

Hausman test

Null hypothesis: RE assumptions hold, both estimators consistent but RE is efficient. Alternative: RE assumptions do *not* hold and the RE estimator is inconsistent. In the TX example we can reject H_0:

```
. hausman FE RE

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      FE           RE          Difference          S.E.
-------------+----------------------------------------------------------------
    avgclass |   -.1339024    -.0442893       -.0896131        .0112156
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        63.84
                Prob>chi2 =       0.0000
```


Review of GLS

In a linear regression with known heteroskedasticity, we can transform the original data and apply OLS to the transformed data. E.g.:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

and assume $\text{Var}(u_i) = k_i \sigma_u^2$. The GLS transformation divides the data by $\sqrt{k_i}$. Observations with greater variance get *less* weight. The transformed model satisfies homoskedasticity. Note:

$$\begin{aligned} Var\left(\frac{u_i}{\sqrt{k_i}}\right) &= \frac{1}{k_i} Var(u_i) \\ &= \frac{1}{k_i} k_i \sigma_u^2 \\ &= \sigma_u^2 \end{aligned}$$

GLS estimation of random effects models

The random effects model with one covariate is:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \underbrace{C_i + e_{it}}_{v_{it}}$$

GLS estimation again involves a transformation. Let:

$$\theta = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_c^2}}$$

(and note the term under the square root looks like but is different from the ICC). T is the number of observations per group, assuming a balanced panel.

GLS estimation of random effects models

The transformations of Y_{it} and X_{it} are:

$$Y_{it} - \theta \bar{Y}_i$$
$$X_{it} - \theta \bar{X}_i$$

and OLS is estimated on the transformed model:

$$Y_{it} - \theta \bar{Y}_i = \beta_0(1-\theta) + \beta_1(X_{it} - \theta \bar{X}_i) + (v_{it} - \theta \bar{v}_i)$$

The transformed Y_{it} and X_{it} are *quasi-demeaned*. If $\theta = 1$, we have the demeaned (within) model.

GLS estimation of random effects models

θ is not known so it must first be estimated with consistent estimators for σ_e^2 and σ_c^2. Then, $\hat{\theta}$ is used in OLS estimation ("feasible GLS").

$$\hat{\theta} = 1 - \sqrt{\frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2 + T\hat{\sigma}_c^2}}$$

Consistent estimators for σ_c^2 and σ_e^2 can be obtained using pooled OLS or fixed effects residuals.

GLS estimation of random effects models

One method for estimating σ_c^2 and σ_e^2: note that

$$v_{it} = C_i + e_{it}$$

$$v_{it}v_{is} = (C_i + e_{it})(C_i + e_{is})$$

$$E(v_{it}v_{is}) = \underbrace{E(C_i^2)}_{\sigma_c^2} + \underbrace{E(C_i e_{is})}_{0} + \underbrace{E(C_i e_{it})}_{0} + \underbrace{E(e_{it}e_{is})}_{0}$$

Get the composite residuals $\hat{v}_{it}$ using pooled OLS. The square of the RMSE in this regression estimates σ_v^2. The within-group (i) covariance in $\hat{v}_{it}$ (the sample analog of $E(v_{it}v_{is})$ above) provides a consistent estimate of σ_c^2. Then, $\hat{\sigma}_e^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$.

GLS estimation of random effects models

$$Y_{it} - \theta \bar{Y}_i = \beta_0(1-\theta) + \beta_1(X_{it} - \theta \bar{X}_i) + (v_{it} - \theta \bar{v}_i)$$

$$\theta = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_c^2}}$$

Notice the transformation subtracts a *fraction* of the within-group mean, where the fraction depends on σ_e^2, σ_c^2, and T.

- When $\theta = 0$, the model reduces to pooled OLS
- When $\theta = 1$, the model reduces to fixed effects (within)
- So, the value of θ is indicative of which model RE is closer to

θ gets closer to 1 as between-group variation σ_c^2 grows relative to within-group variation σ_e^2, and as the number of time periods T grows.

GLS estimation of random effects models

Can request $\hat{\theta}$ in `xtreg, re`:

```
. xtreg avgpassing avgclass, re i(campus) theta

Random-effects GLS regression                   Number of obs     =     16,062
Group variable: campus                          Number of groups  =      4,326

R-sq:                                           Obs per group:
     within  = 0.0018                                         min =          1
     between = 0.0098                                         avg =        3.7
     overall = 0.0060                                         max =          4

                                                Wald chi2(1)      =       2.74
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0978

------------------- theta --------------------
  min      5%       median        95%      max
0.5362   0.6529     0.7468     0.7468   0.7468

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0442893   .0267548    -1.66   0.098    -.0967277    .0081491
       _cons |   76.21828   .5503649   138.49   0.000     75.13959    77.29698
-------------+----------------------------------------------------------------
     sigma_u |  12.391941
     sigma_e |  6.4870883
         rho |  .78490199   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

This uses the original unbalanced panel, so $\hat{\theta}$ varies with group size.

GLS estimation of random effects models

Can request $\hat{\theta}$ in `xtreg, re`:

```
. xtreg avgpassing avgclass, re i(campus) theta

Random-effects GLS regression                   Number of obs      =     14,796
Group variable: campus                          Number of groups   =      3,699

R-sq:                                           Obs per group:
     within  = 0.0020                                         min =          4
     between = 0.0138                                         avg =        4.0
     overall = 0.0061                                         max =          4

                                                Wald chi2(1)       =       2.97
corr(u_i, X)    = 0 (assumed)                   Prob > chi2        =     0.0848
theta           = .73287384

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0484254   .0280999    -1.72   0.085    -.1035003    .0066494
       _cons |   76.51251   .5742248   133.24   0.000     75.38705    77.63797
-------------+----------------------------------------------------------------
     sigma_u |  11.706021
     sigma_e |  6.4897977
         rho |  .76490175   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

This uses the <u>balanced</u> panel, so $\hat{\theta}$ is constant.

GLS estimation of random effects models

It is useful to consider the error term in the quasi-demeaned model:

$$v_{it} - \theta\bar{v}_i = (1-\theta)C_i + (e_{it} - \theta\bar{e}_i)$$

Suppose the RE assumption that C_i is uncorrelated with x_{it} does *not* hold. As $\theta \to 1$, the C_i component of the error term diminishes in importance, the RE estimator tends toward the FE estimator, and any bias associated with RE tends to zero.

MLE estimation of random effects models

Random effects models can also be estimated using **maximum likelihood** in which case all parameters of the model (β's, σ's) are estimated jointly:

```
. xtreg avgpassing avgclass, mle i(campus)

Fitting constant-only model:
Iteration 0:   log likelihood = -53584.523
Iteration 1:   log likelihood = -53584.523

Fitting full model:
Iteration 0:   log likelihood = -53674.187
Iteration 1:   log likelihood = -53583.763
Iteration 2:   log likelihood = -53582.969
Iteration 3:   log likelihood = -53582.969

Random-effects ML regression                    Number of obs     =     14,796
Group variable: campus                          Number of groups  =      3,699

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4

                                                LR chi2(1)        =       3.11
Log likelihood  = -53582.969                    Prob > chi2       =     0.0780

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0496391   .0281539    -1.76   0.078    -.1048197    .0055415
       _cons |   76.53576   .5755876   132.97   0.000     75.40763    77.66389
-------------+----------------------------------------------------------------
    /sigma_u |    11.8066   .1481004                      11.51987    12.10047
    /sigma_e |   6.492198   .0436102                      6.407283    6.578237
         rho |   .7678329   .0051631                      .7575916    .7778289
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 1.2e+04           Prob >= chibar2 = 0.000
```

Last update: September 25, 2025

Getting estimates of random effects C_i

As with `xtreg, fe`, one can obtain the $\hat{C}_i$ estimates of the group random effects. Unlike `fe`, these are not coefficient estimates but rather estimated from residuals. The random effects $\hat{C}_i$ can be calculated in two ways:

- Maximum likelihood (following `xtreg, mle`): `predict`
- Empirical Bayes / shrinkage approach: the Best Linear Unbiased Predictors (BLUPs)

Shrinkage approach: multiply $\hat{C}_i$ by a shrinkage factor $\hat{R}_i = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_c^2 + \frac{\hat{\sigma}_e^2}{T_i}}$

where T_i is the number of observations in group i. Examples on next 3 slides.

Last update: September 25, 2025

Getting estimates of random effects C_i: MLE

```
. xtreg avgpassing avgclass, re mle i(campus)

Fitting constant-only model:
Iteration 0:   log likelihood = -53584.523
Iteration 1:   log likelihood = -53584.523

Fitting full model:
Iteration 0:   log likelihood = -53674.187
Iteration 1:   log likelihood = -53583.763
Iteration 2:   log likelihood = -53582.969
Iteration 3:   log likelihood = -53582.969

Random-effects ML regression                    Number of obs     =     14,796
Group variable: campus                          Number of groups  =      3,699

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4

                                                LR chi2(1)        =       3.11
Log likelihood  = -53582.969                    Prob > chi2       =     0.0780

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0496391   .0281539    -1.76   0.078    -.1048197    .0055415
       _cons |   76.53576   .5755876   132.97   0.000     75.40763    77.66389
-------------+----------------------------------------------------------------
    /sigma_u |    11.8066   .1481004                      11.51987    12.10047
    /sigma_e |   6.492198   .0436102                      6.407283    6.578237
         rho |   .7678329   .0051631                      .7575916    .7778289
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 1.2e+04            Prob >= chibar2 = 0.000

. predict uhat1, u

. sum uhat1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       uhat1 |     14,796    8.39e-09    12.24512  -47.43509   23.42125
```


Getting estimates of random effects C_i: BLUP

```
. xtreg avgpassing avgclass, re mle i(campus)

Fitting constant-only model:
Iteration 0:   log likelihood = -53584.523
Iteration 1:   log likelihood = -53584.523

Fitting full model:
Iteration 0:   log likelihood = -53674.187
Iteration 1:   log likelihood = -53583.763
Iteration 2:   log likelihood = -53582.969
Iteration 3:   log likelihood = -53582.969

Random-effects ML regression                    Number of obs     =     14,796
Group variable: campus                          Number of groups  =      3,699

Random effects u_i ~ Gaussian                   Obs per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4

                                                LR chi2(1)        =       3.11
Log likelihood  = -53582.969                    Prob > chi2       =     0.0780

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0496391   .0281539    -1.76   0.078    -.1048197    .0055415
       _cons |   76.53576   .5755876   132.97   0.000     75.40763    77.66389
-------------+----------------------------------------------------------------
    /sigma_u |    11.8066   .1481004                      11.51987    12.10047
    /sigma_e |   6.492198   .0436102                      6.407283    6.578237
         rho |   .7678329   .0051631                      .7575916    .7778289
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 1.2e+04            Prob >= chibar2 = 0.000

. gen shrink = _b[/sigma_u]^2 / (_b[/sigma_u]^2 + (_b[/sigma_e]^2)/4)

. gen uhat1s = uhat1*shrink

. summ uhat1s shrink

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      uhat1s |     14,796    1.16e-08    11.38455  -44.10139   21.77522
      shrink |     14,796    .9297209           0   .9297209   .9297209
```

Getting estimates of C_i: BLUP using `xtmixed`

```
. xtmixed avgpassing avgclass || campus: , mle

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -53582.969
Iteration 1:   log likelihood = -53582.969

Computing standard errors:

Mixed-effects ML regression                     Number of obs     =     14,796
Group variable: campus                          Number of groups  =      3,699

                                                Obs per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4

                                                Wald chi2(1)      =       3.13
Log likelihood = -53582.969                     Prob > chi2       =     0.0770

------------------------------------------------------------------------------
  avgpassing |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    avgclass |  -.0496392   .0280727    -1.77   0.077    -.1046606    .0053823
       _cons |   76.53576   .5741313   133.31   0.000     75.41048    77.66103
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
campus: Identity             |
                   sd(_cons) |    11.8066   .1481006      11.51987    12.10047
-----------------------------+------------------------------------------------
                sd(Residual) |   6.492197   .0436102      6.407283    6.578236
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 11666.05      Prob >= chibar2 = 0.0000

. predict uhat2, reffects

. sum uhat2

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       uhat2 |     14,796   -6.21e-10    11.38455  -44.10139   21.77523
```


Getting estimates of random effects C_i

The shrinkage factor is smaller for groups with fewer observations (T_i). Their $\hat{C}_i$ is “shrunk” more toward the overall mean group effect of 0.

- RE estimates generally smaller than FE estimates in absolute value
- True for both MLE and EB estimates of the RE, but especially the EB
- The rank order of the $\hat{C}_i$ is usually preserved whether one assumes RE or FE

Random vs. fixed effects

When and where random effects are appropriate:

- As a rule, if the FE assumption holds the RE model is inappropriate. See the Texas class size example, where the Hausman test rejected RE.

- RE is appropriate with grouped or clustered data. See the Success for All example: assignment to treatment was random at the school level, so we need not be concerned about correlation between treatment and the error term. However, the errors are not i.i.d.

See Rabe-Hesketh and Skrondal MLM text for more guidance on RE vs. FE decision.

xttest0

The command `xttest0` (following `xtreg`) provides a formal test for the presence of random effects. H_0 in this case is that the variance across panel units is zero, and thus RE is unnecessary.

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        wattack[schid,t] = Xb + u[schid] + e[schid,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                 wattack |   395.1636       19.87872
                       e |   279.7314       16.72517
                       u |   47.72378        6.90824

        Test:   Var(u) = 0
                             chibar2(01) =   1266.18
                          Prob > chibar2 =    0.0000
```

Standard errors in panel models

Whether using a FE or RE model, the assumption that errors u_{it} are i.i.d. is not often satisfied in panel data. With repeat observations on the same cross-sectional unit, it is likely that errors are correlated across observations for the same i.

- If Y is over-predicted in one period for a given i, it is likely to be over-predicted in the next period.

Standard errors in panel models

The RE model explicitly models the correlation across observations within group. There is an increasing preference, however, for <u>not</u> doing this and adjusting standard errors for within-panel clustering.

- For "short" panels (large N small T) use cluster-robust standard errors
- The "cluster" is typically the cross-sectional unit, although when the regressor of interest is aggregated at a higher level (e.g., state), can cluster at that level. Theory requires large N and that higher levels nest the cross-sectional units.
- `vce(robust)` or `robust` in `xtreg` assumes data are clustered
- Cluster-robust standard errors from `areg` are different from those using `xtreg, fe`. It is recommended that you use `xtreg`.