Longitudinal Data Analysis
Instructor: Natasha Sarkisian

Panel Data Analysis: Random Effects Models

In fixed effects models, each dummy variable removes one degree of freedom from the model; thus, fixed effects models work well when you have a substantial number of time periods. To avoid losing the degrees of freedom and to utilize both the information on change over time for a given unit and the information on differences across units, we can estimate random effects models. The model still decomposes the residuals: $Y_{it}= \alpha + X_{it}\beta + u_i + e_{it}$ where u_i represents the effect of unit i and e_{it} is the residual effect for time point t for that unit. But in a random effects model, unit residuals u_i do not have specific values – u_i is a normally distributed random variable (hence the name – random effects).

The nature of the coefficients β also changes as we go from a fixed effects to a random effects model – in a random effects model, we are not only predicting change over time but also explaining the differences among the units. Thus, the data on cross-sectional variation are utilized in estimating independent variables' effects. Because the predictors are used to explain not only change over time but also differences among units, the random unit residual variable u is assumed to be uncorrelated with Xβ: `corr(u_i, Xb)  = 0.` We can now use time-invariant variables in our model.

```
. xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, re cluster(hhid)
Random-effects GLS regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0229                         Obs per group: min =         1
       between = 0.0309                                        avg =       4.9
       overall = 0.0254                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(10)      =    529.71
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
                               (Std. Err. adjusted for 4635 clusters in hhid)
------------------------------------------------------------------------------
             |               Robust
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |   -.017518   .0013378   -13.09   0.000       -.02014   -.0148959
 rpoorhealth |  -.1027325   .0722552    -1.42   0.155     -.2443501    .0388852
    rmarried |  -.3439424   .0982397    -3.50   0.000     -.5364887   -.1513962
   rtotalpar |  -.2764635   .0419983    -6.58   0.000     -.3587785   -.1941484
     rsiblog |  -.3816662   .0643893    -5.93   0.000     -.5078669   -.2554656
    hchildlg |  -.0431438   .0651145    -0.66   0.508     -.1707658    .0844782
      female |   .4784234   .0581174     8.23   0.000      .3645154    .5923314
         age |   -.040811   .0118594    -3.44   0.001     -.0640551    -.017567
    minority |  -.1316851   .0900886    -1.46   0.144     -.3082556    .0448853
     raedyrs |   .0647266   .0110043     5.88   0.000      .0431586    .0862946
       _cons |   4.572378   .7227877     6.33   0.000       3.15574    5.989015
-------------+----------------------------------------------------------------
     sigma_u |  1.6329416
     sigma_e |  3.5375847
         rho |  .17564702   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Note that less variance is attributed to person level in this model than in the fixed effects model, but a significance test for unit-level variance is not included. But we can easily obtain it:

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        rallparhelptw[hhidpn,t] = Xb + u[hhidpn] + e[hhidpn,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                rallpar~w |   16.45761       4.056797
                        e |   12.51451       3.537585
                        u |   2.666498       1.632942

        Test:   Var(u) = 0
                             chi2(1) =  4211.99
                         Prob > chi2 =     0.0000
```

Thus, we reject the null hypothesis that person-specific residuals are all zero – there is a significant amount of variance across persons above and beyond that explained by our predictors.

So far we estimated our model using GLS (generalized least squares) estimation method; we could also estimate the same model using maximum likelihood estimation option, although cluster option is not available with this method:

```
. xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, re  mle

Fitting constant-only model:
Iteration 0:   log likelihood = -84739.359
Iteration 1:   log likelihood = -84735.952
Iteration 2:   log likelihood = -84735.947

Fitting full model:
Iteration 0:   log likelihood = -84417.691
Iteration 1:   log likelihood = -84386.623
Iteration 2:   log likelihood = -84386.583

Random-effects ML regression                    Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       4.9
                                                               max =         9

                                                LR chi2(10)        =    698.73
Log likelihood  = -84386.583                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0177108   .0011737   -15.09   0.000    -.0200112   -.0154104
 rpoorhealth |  -.0888093   .0643735    -1.38   0.168     -.214979    .0373604
    rmarried |  -.3523333   .0784346    -4.49   0.000    -.5060623   -.1986043
   rtotalpar |  -.3073022   .0323089    -9.51   0.000    -.3706264    -.243978
     rsiblog |  -.3762714   .0551995    -6.82   0.000    -.4844604   -.2680823
    hchildlg |  -.0384941   .0582924    -0.66   0.509     -.152745    .0757568
      female |        .47   .0671802     7.00   0.000     .3383292    .6016708
```

```
         age |  -.0423231   .0107905    -3.92   0.000     -.063472   -.0211741
    minority |  -.1365561   .0806732    -1.69   0.091    -.2946727    .0215606
     raedyrs |   .0658393   .0115215     5.71   0.000     .0432574    .0884211
       _cons |   4.670711   .6493207     7.19   0.000     3.398066    5.943356
-------------+----------------------------------------------------------------
    /sigma_u |   1.882485   .0301588                      1.824293    1.942533
    /sigma_e |   3.524548   .0157879                       3.49374    3.555628
         rho |   .2219534   .0060177                      .2103391    .2339254
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)= 2611.48 Prob>=chibar2 = 0.000
```

The same model can be fit using xtmixed command – we will later use this command for mixed model, and the random effects model is a basic case of such a model:

```
. xtmixed rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs || hhidpn:
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log restricted-likelihood = -84415.598
Iteration 1:   log restricted-likelihood = -84415.597
Computing standard errors:
Mixed-effects REML regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
                                                Obs per group: min =         1
                                                               avg =       4.9
                                                               max =         9
                                                Wald chi2(10)      =    706.62
Log restricted-likelihood = -84415.597          Prob > chi2        =    0.0000
------------------------------------------------------------------------------
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0177123   .0011738   -15.09   0.000    -.0200128   -.0154117
 rpoorhealth |  -.0886995   .0643665    -1.38   0.168    -.2148554    .0374565
    rmarried |  -.3524069   .0784612    -4.49   0.000    -.5061881   -.1986257
   rtotalpar |   -.307533    .032103    -9.58   0.000    -.3704536   -.2446123
     rsiblog |  -.3762313   .0552292    -6.81   0.000    -.4844784   -.2679841
    hchildlg |  -.0384531   .0583245    -0.66   0.510     -.152767    .0758607
      female |    .469936   .0672173     6.99   0.000     .3381925    .6016796
         age |  -.0423344   .0107962    -3.92   0.000    -.0634946   -.0211742
    minority |  -.1365957   .0807244    -1.69   0.091    -.2948126    .0216212
     raedyrs |   .0658478   .0115284     5.71   0.000     .0432526    .0884431
       _cons |   4.671444   .6496436     7.19   0.000     3.398166    5.944722
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
hhidpn: Identity             |
                   sd(_cons) |   1.884641   .0301846        1.8264    1.944741
-----------------------------+------------------------------------------------
                sd(Residual) |   3.524762   .0157898       3.49395    3.555845
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  2616.90 Prob >= chibar2 = 0.0000
```

As mentioned above, random effects coefficients have a dual nature: They simultaneously explain change over time and the cross-sectional differences among units. The implicit assumption is that both types of effects are the same. That is, when we say that a one unit increase in X is associated with a b units increase in Y, a one unit increase might mean two things:

1. We observe two different individuals with a one unit difference in X between them.
2. We observe one person, and its X value increases by one unit.

In a random effects model, we are assuming that both of those produce the same effect on Y. That is, for instance, we assume that if one person works one hour more per week than another, and if a given person increases her or his work hours by one hour per week, the effect on hours of help to parents would be the same.

We test this assumption using the Hausman test. The Hausman test checks a more efficient model against a less efficient but consistent model to make sure that the more efficient model also gives consistent results. The null hypothesis is that the coefficients estimated by the efficient random effects estimator are the same as the ones estimated by the consistent fixed effects estimator. If they are, then it is safe to use a random effects model. If the two sets of coefficients are significantly different, then the random effects model is problematic. It is best to use hausman test with sigmamore option; it avoids problems with the matrix [V_b-V_B] not being positive definite.

```
. qui xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog
hchildlg female age minority raedyrs, fe

. est store fixed

. qui xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog
hchildlg female age minority raedyrs, re

. est store random

. hausman fixed random, sigmamore
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
rworkhours80 |   -.0193467     -.017518       -.0018287        .0009452
 rpoorhealth |    .0792176    -.1027325        .1819501        .0499086
    rmarried |   -.6578103    -.3439424       -.3138679        .1128988
   rtotalpar |     -.52481    -.2764635       -.2483466        .0223144
     rsiblog |   -.5767981    -.3816662       -.1951319        .1790009
    hchildlg |    .3859163    -.0431438        .4290601        .1652614
------------------------------------------------------------------------------
                           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(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =      263.59
                Prob>chi2 =      0.0000
```

In this case, we reject the null hypothesis – fixed effects and random effects coefficients are significantly different. Examining the coefficients, we might suspect that rpoorhealth or hchildlg are responsible.

To better understand the meaning of the Hausman test, let’s introduce the between effects model.

```
. xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, be
```

```
Between regression (regression on group means)  Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0008                         Obs per group: min =         1
       between = 0.0483                                        avg =       4.9
       overall = 0.0173                                        max =         9
                                                F(10,6232)         =     31.62
sd(u_i + avg(e_i.))=  2.539716                  Prob > F           =    0.0000
------------------------------------------------------------------------------
rallparhel~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0097168   .0019226    -5.05   0.000    -.0134858   -.0059478
 rpoorhealth |  -.3909062   .1052603    -3.71   0.000    -.5972526   -.1845598
    rmarried |  -.3108795   .0944656    -3.29   0.001    -.4960647   -.1256943
   rtotalpar |   .3335196   .0595554     5.60   0.000     .2167706    .4502686
     rsiblog |  -.3402857   .0571287    -5.96   0.000    -.4522776   -.2282937
    hchildlg |   -.139232   .0610987    -2.28   0.023    -.2590065   -.0194575
      female |    .683156   .0695158     9.83   0.000     .5468811    .8194309
         age |  -.0040194     .01099    -0.37   0.715    -.0255636    .0175247
    minority |  -.0596539    .079881    -0.75   0.455    -.2162482    .0969403
     raedyrs |   .0382127   .0116876     3.27   0.001     .0153009    .0611245
       _cons |    1.80804   .6821661     2.65   0.008     .4707594    3.145321
------------------------------------------------------------------------------
```

This type of analysis is equivalent to taking the mean of each variable across time for each case and running a regression on the collapsed dataset of means. As this results in a loss of information, between effects are rarely used. The between effects estimator is mostly important because Stata's random-effects estimator is a weighted average of a fixed effects and a between effects coefficient. Thus, implicitly, the Hausman test assesses whether fixed effects and between effects produce the same coefficients. If they do, it is appropriate to combine them into a random effects model. Comparing these coefficients to the fixed effects coefficients in the Hausman output, we see some major differences for rpoorhealth and hchildlg but also rtotalpar. We could also estimate the two types of effects (over time and across units) separately in a single random effects model using the same kind of person-specific mean variables and mean-differenced variables that we created when examining fixed effects models (this is only done for time-varying variables):

```
. for var rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg: bysort hhidpn:
egen Xm=mean(X)  \ gen Xdiff=X-Xm

->  bysort hhidpn: egen rworkhours80m=mean(rworkhours80)
(36 missing values generated)

->  gen rworkhours80diff=rworkhours80-rworkhours80m
(8015 missing values generated)

->  bysort hhidpn: egen rpoorhealthm=mean(rpoorhealth)

->  gen rpoorhealthdiff=rpoorhealth-rpoorhealthm
(7535 missing values generated)

->  bysort hhidpn: egen rmarriedm=mean(rmarried)

->  gen rmarrieddiff=rmarried-rmarriedm
(7561 missing values generated)

->  bysort hhidpn: egen rtotalparm=mean(rtotalpar)

->  gen rtotalpardiff=rtotalpar-rtotalparm
```

```
(7846 missing values generated)

->  bysort hhidpn: egen rsiblogm=mean(rsiblog)
(6 missing values generated)

->  gen rsiblogdiff=rsiblog-rsiblogm
(81 missing values generated)

->  bysort hhidpn: egen hchildlgm=mean(hchildlg)
(2248 missing values generated)

->  gen hchildlgdiff=hchildlg-hchildlgm
(10457 missing values generated)

. xtreg  rallparhelptw  rworkhours80m rworkhours80diff rpoorhealthm rpoorhealthdiff
rmarriedm rmarrieddiff rtotalparm rtotalpardiff rsiblogm rsiblogdiff hchildlgm
hchildlgdiff female age  minority  raedyrs, re cluster(hhid)
Random-effects GLS regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0242                         Obs per group: min =         1
       between = 0.0409                                        avg =       4.9
       overall = 0.0332                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(16)      =    577.31
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
                                (Std. Err. adjusted for 4635 clusters in hhid)
------------------------------------------------------------------------------
             |               Robust
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhou~80m |  -.0115568   .0022162    -5.21   0.000    -.0159006   -.0072131
rworkhours~f |  -.0176429   .0016761   -10.53   0.000     -.020928   -.0143578
rpoorhealthm |  -.3904361   .1203335    -3.24   0.001    -.6262854   -.1545869
rpoorhealt~f |   .0658695   .0839301     0.78   0.433    -.0986304    .2303694
   rmarriedm |  -.2655983   .1099098    -2.42   0.016    -.4810175    -.050179
rmarrieddiff |   -.680859   .1555738    -4.38   0.000    -.9857781   -.3759399
  rtotalparm |   .1583439   .0615846     2.57   0.010     .0376404    .2790474
rtotalpard~f |  -.4481539   .0546544    -8.20   0.000    -.5552747   -.3410332
    rsiblogm |  -.3632242    .068526    -5.30   0.000    -.4975326   -.2289157
 rsiblogdiff |   -.683971   .1578554    -4.33   0.000     -.993362     -.37458
   hchildlgm |    -.09689   .0682514    -1.42   0.156    -.2306603    .0368802
hchildlgdiff |   .3307412   .1666087     1.99   0.047     .0041942    .6572882
      female |   .6542834   .0634002    10.32   0.000     .5300213    .7785454
         age |  -.0074142   .0122852    -0.60   0.546    -.0314927    .0166644
    minority |  -.0700329   .0907892    -0.77   0.440    -.2479765    .1079107
     raedyrs |   .0421826   .0112259     3.76   0.000     .0201802     .064185
       _cons |   2.440797   .7640383     3.19   0.001     .9433094    3.938285
-------------+----------------------------------------------------------------
     sigma_u |   1.627307
     sigma_e |  3.5375847
         rho |  .17464829   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Let's compare pairs of coefficients:

```
. test rworkhours80m=rworkhours80diff

 ( 1)  rworkhours80m - rworkhours80diff = 0

           chi2(  1) =    4.81
         Prob > chi2 =    0.0284

. test rpoorhealthm=rpoorhealthdiff
```

```
 ( 1)  rpoorhealthm - rpoorhealthdiff = 0

           chi2(  1) =   10.80
         Prob > chi2 =    0.0010

. test rmarriedm=rmarrieddiff

 ( 1)  rmarriedm - rmarrieddiff = 0

           chi2(  1) =    5.93
         Prob > chi2 =    0.0149

. test rtotalparm=rtotalpardiff

 ( 1)  rtotalparm - rtotalpardiff = 0

           chi2(  1) =   54.91
         Prob > chi2 =    0.0000

. test rsiblogm=rsiblogdiff

 ( 1)  rsiblogm - rsiblogdiff = 0

           chi2(  1) =    3.64
         Prob > chi2 =    0.0563

. test hchildlgm=hchildlgdiff

 ( 1)  hchildlgm - hchildlgdiff = 0

           chi2(  1) =    5.80
         Prob > chi2 =    0.0160
```

All differences except for effects of number of siblings are significant if we pick .05 alpha, but because of large sample size and because some of these have different numbers but similar substantive interpretation, I will use .01 alpha level. I will keep coefficients for number of children different for now because the story seems different. So we can constrain the model as follows:

```
. xtreg  rallparhelptw  rworkhours80 rpoorhealthm rpoorhealthdiff rmarried rtotalparm
rtotalpardiff rsiblog hchildlgm hchildlgdiff female age  minority  raedyrs, re
cluster(hhid)
Random-effects GLS regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0242                         Obs per group: min =         1
       between = 0.0401                                        avg =       4.9
       overall = 0.0327                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(13)      =    573.86
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
                                (Std. Err. adjusted for 4635 clusters in hhid)
------------------------------------------------------------------------------
             |               Robust
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0158588    .001346   -11.78   0.000    -.0184969   -.0132208
rpoorhealthm |  -.4760427    .115992    -4.10   0.000    -.7033829   -.2487026
rpoorhealt~f |   .0787295   .0839293     0.94   0.348    -.0857689     .243228
    rmarried |  -.4113396   .0988835    -4.16   0.000    -.6051476   -.2175315
  rtotalparm |   .1822215   .0605674     3.01   0.003     .0635115    .3009316
rtotalpard~f |  -.4767525   .0534871    -8.91   0.000    -.5815853   -.3719197
```

```
     rsiblog |   -.380791   .0641246    -5.94   0.000    -.5064729   -.2551091
   hchildlgm |  -.0930924   .0682771    -1.36   0.173     -.226913    .0407282
hchildlgdiff |   .2962874   .1650998     1.79   0.073    -.0273022    .6198771
      female |   .5895598   .0594126     9.92   0.000     .4731133    .7060063
         age |  -.0132152     .01195    -1.11   0.269    -.0366369    .0102064
    minority |  -.0819295   .0900856    -0.91   0.363    -.2584942    .0946351
     raedyrs |    .043404   .0112664     3.85   0.000     .0213222    .0654857
       _cons |   2.979462   .7336866     4.06   0.000     1.541463    4.417462
-------------+----------------------------------------------------------------
     sigma_u |  1.6326049
     sigma_e |  3.5375847
         rho |  .17558732   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. test hchildlgm=hchildlgdiff
 ( 1)  hchildlgm - hchildlgdiff = 0
           chi2(  1) =    4.88
         Prob > chi2 =    0.0271
```

Not much of a story left for number of children, so I will further constrain the model:

```
. xtreg  rallparhelptw  rworkhours80 rpoorhealthm rpoorhealthdiff rmarried rtotalparm
rtotalpardiff rsiblog hchildlg female age  minority  raedyrs, re cluster(hhid)

Random-effects GLS regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0239                         Obs per group: min =         1
       between = 0.0400                                        avg =       4.9
       overall = 0.0326                                        max =         9
Random effects u_i ~ Gaussian                   Wald chi2(12)      =    566.65
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
                                (Std. Err. adjusted for 4635 clusters in hhid)
------------------------------------------------------------------------------
             |               Robust
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0159143   .0013455   -11.83   0.000    -.0185514   -.0132772
rpoorhealthm |   -.477944    .115859    -4.13   0.000    -.7050234   -.2508646
rpoorhealt~f |   .0817012   .0839297     0.97   0.330    -.0827979    .2462004
    rmarried |  -.4003812   .0985202    -4.06   0.000    -.5934773    -.207285
  rtotalparm |   .1815216   .0605404     3.00   0.003     .0628647    .3001786
rtotalpard~f |  -.4785309   .0534485    -8.95   0.000    -.5832881   -.3737737
     rsiblog |  -.3861979   .0639649    -6.04   0.000    -.5115669    -.260829
    hchildlg |  -.0502547   .0641206    -0.78   0.433    -.1759288    .0754194
      female |   .5906606   .0594083     9.94   0.000     .4742225    .7070987
         age |  -.0138068   .0119468    -1.16   0.248    -.0372221    .0096084
    minority |  -.0829326   .0900776    -0.92   0.357    -.2594815    .0936163
     raedyrs |   .0445902   .0112573     3.96   0.000     .0225263    .0666541
       _cons |    2.95161   .7336181     4.02   0.000     1.513745    4.389475
-------------+----------------------------------------------------------------
     sigma_u |  1.6322975
     sigma_e |  3.5375847
         rho |  .17553282   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Thus, there are really two kinds of information in panel data:

1. The cross-sectional information reflected in the differences among units.
2. The time-series or within-unit information reflected in the changes within units.

For that reason, panel data is also called sometimes cross-sectional time-series data.

A between effects model uses only the cross-sectional information and asks: "What is the expected difference in Y between two individuals that differ by 1 in X?", while a fixed effects model uses only the time-series information and asks, "What is the expected change in a persons's value of Y if its value of X increases by 1?" A random effects model combines those two questions, but really, it may turn out that the answers to those two questions are the same or they may be different. If they are different, we could either use a fixed effects model, or we can separate the two types of effects within a random effects model, but we should be able to explain why the effects are different. Statistically, a fixed effects model is always a reasonable thing to do with panel data (it always gives consistent results) but it may not be the most efficient model to run. A random effects model will give you lower standard errors as it is a more efficient estimator.

To better understand these choices, see:
Bell, Andrew, Malcolm Fairbrother, and Andrew Bell. 2019. Fixed and random effects models: making an informed choice. *Quality and Quantity*, 53(2):1051–74.

<u>Autocorrelation</u>

Even though we took into account the fact that units have something in common (unit-specific residuals) and that observations are non-independent (by using cluster option), there can still be additional problems, especially with autocorrelation of residuals. We can test for and deal with autocorrelation the same way as in FE models, using xtserial and xtregar commands; the only difference is that we specify re rather than fe in xtregar.

```
. xtserial rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog
hchildlg female age minority  raedyrs

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,    4558) =     34.757
           Prob > F =      0.0000
```

Here, the hypothesis of no first order autocorrelation is rejected; therefore, we would want a model explicitly accounting for autoregressive error term. We can use xtregar:

```
. xtregar rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, re lbi

RE GLS regression with AR(1) disturbances       Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243

R-sq:  within  = 0.0231                         Obs per group: min =         1
       between = 0.0321                                        avg =       4.9
       overall = 0.0256                                        max =         9

                                                Wald chi2(11)      =    563.33
corr(u_i, Xb)      = 0 (assumed)                Prob > chi2        =    0.0000

------------------- theta --------------------
  min      5%       median        95%      max
0.0655   0.0995     0.2270     0.2647   0.2647

------------------------------------------------------------------------------
```

```
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0153351   .0012087   -12.69   0.000    -.0177041   -.0129662
 rpoorhealth |  -.0652898   .0637556    -1.02   0.306    -.1902484    .0596688
    rmarried |  -.3203856   .0794461    -4.03   0.000     -.476097   -.1646742
   rtotalpar |  -.2490934   .0334718    -7.44   0.000     -.314697   -.1834898
     rsiblog |  -.3716501   .0546967    -6.79   0.000    -.4788536   -.2644466
    hchildlg |  -.0463342   .0577381    -0.80   0.422    -.1594988    .0668305
      female |    .522334   .0660313     7.91   0.000      .392915     .651753
         age |  -.0382192   .0106209    -3.60   0.000    -.0590357   -.0174027
    minority |  -.1338564   .0792718    -1.69   0.091    -.2892263    .0215134
     raedyrs |   .0611931   .0113042     5.41   0.000     .0390372     .083349
       _cons |   4.320286   .6392352     6.76   0.000     3.067408    5.573164
-------------+----------------------------------------------------------------
      rho_ar |  .24444212   (estimated autocorrelation coefficient)
     sigma_u |  1.4158555
     sigma_e |  3.6044943
     rho_fov |  .13366962   (fraction of variance due to u_i)
------------------------------------------------------------------------------
modified Bhargava et al. Durbin-Watson = 1.5724772
Baltagi-Wu LBI = 2.0213364
```

<u>Diagnostics</u>

Same as after xtreg, fe, we can use predict command after xtreg, re to get predicted values and residuals:

```
      xb            xb, fitted values; the default
      stdp          standard error of the fitted values
      ue            u_i + e_it, the combined residual
      xbu           xb + u_i, prediction including effect
      u             u_i, the fixed- or random-error component
      e             e_it, the overall error component
```

Again, we can use these residuals to conduct regression diagnostics – examine normality, linearity, heteroskedasticity. Note that while in fixed effects models, we were not concerned about heteroskedasticity or non-normality for level 2 residuals, and expected to see some relationships between predictors and level 2 residuals, in random effects models, we have to ensure assumptions of multivariate normality, homoscedasticity, and linearity for both levels of residuals, and we should see no relationship at all between predictors and residuals on both levels.

Note that for both fixed effects and between effects, there are straightforward transformations of variables that can be made to obtain the same coefficients without xtreg (i.e., mean-differencing or collapsing dataset to person-mean level). For random effects, such transformation does not exist, but xtdata command in Stata (with re option) does offer an approximation that can be used to conduct faster searches for model specification for a random effects model if you have a lot of predictors and are trying to select the best model. The random effects models estimated in the exploratory dataset generated by xtdata command will not be identical to those estimated in the full dataset--they will be a very close approximation.

```
. xtreg rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, re
Random-effects GLS regression                   Number of obs      =     30541
Group variable: hhidpn                          Number of groups   =      6243
R-sq:  within  = 0.0229                         Obs per group: min =         1
```

```
       between = 0.0309                                         avg =       4.9
       overall = 0.0254                                         max =         9
Random effects u_i ~ Gaussian                    Wald chi2(10)      =    714.51
corr(u_i, X)       = 0 (assumed)                 Prob > chi2        =    0.0000
------------------------------------------------------------------------------
rallparhel~w |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |   -.017518   .0011596   -15.11   0.000    -.0197907   -.0152452
 rpoorhealth |  -.1027325   .0636668    -1.61   0.107    -.2275171    .0220522
    rmarried |  -.3439424   .0757802    -4.54   0.000    -.4924689    -.195416
   rtotalpar |  -.2764635   .0318816    -8.67   0.000    -.3389502   -.2139767
     rsiblog |  -.3816662   .0523548    -7.29   0.000    -.4842798   -.2790526
    hchildlg |  -.0431438   .0552126    -0.78   0.435    -.1513586     .065071
      female |   .4784234   .0632272     7.57   0.000     .3545003    .6023465
         age |   -.040811   .0101534    -4.02   0.000    -.0607114   -.0209107
    minority |  -.1316851   .0759759    -1.73   0.083    -.2805951    .0172248
     raedyrs |   .0647266   .0108469     5.97   0.000     .0434671    .0859861
       _cons |   4.572378   .6117239     7.47   0.000     3.373421    5.771334
-------------+----------------------------------------------------------------
     sigma_u |  1.6329416
     sigma_e |  3.5375847
         rho |  .17564702   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Xtdata command requires that we specify the ratio of sigma_u to sigma_e as standard deviations rather than variances; so we calculate it:

```
. di 1.6329416/3.5375847
.46159788

. xtdata rallparhelptw  rworkhours80  rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority  raedyrs, re ratio(.46159788) clear

------------------- theta --------------------
  min      5%       median        95%      max
0.0921   0.0921     0.3042     0.4146   0.4146

. reg  rallparhelptw rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority raedyrs, cluster(hhidpn)
Linear regression                                      Number of obs =   30541
                                                       F( 10,  6242) =   51.37
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0228
                                                       Root MSE      =  3.5801
                              (Std. Err. adjusted for 6243 clusters in hhidpn)
------------------------------------------------------------------------------
             |               Robust
rallparhel~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
rworkhours80 |  -.0158976   .0012812   -12.41   0.000    -.0184092   -.0133861
 rpoorhealth |  -.0677607   .0706261    -0.96   0.337    -.2062121    .0706907
    rmarried |  -.3536576   .0956405    -3.70   0.000    -.5411458   -.1661693
   rtotalpar |  -.3049411    .037739    -8.08   0.000    -.3789226   -.2309597
     rsiblog |  -.3732796   .0583542    -6.40   0.000    -.4876739   -.2588852
    hchildlg |  -.0502717   .0575627    -0.87   0.383    -.1631144     .062571
      female |   .5318233   .0672479     7.91   0.000     .3999942    .6636524
         age |  -.0000753    .004729    -0.02   0.987    -.0093458    .0091952
    minority |  -.0763812   .0816494    -0.94   0.350    -.2364422    .0836797
     raedyrs |    .065813     .01006     6.54   0.000     .0460919    .0855342
       _cons |   1.529079   .1619631     9.44   0.000     1.211575    1.846582
------------------------------------------------------------------------------
```

After converting the data, you may form linear transformations your predictors, but all nonlinear transformations must be done before conversion. You can, however, use some OLS-based diagnostic tools, e.g., examine linearity:

```
. mrunning  rallparhelptw rworkhours80 rpoorhealth rmarried rtotalpar rsiblog hchildlg
female age minority redyrs

30541 observations, R-sq = 0.0333
```