Panel Data Regression Models

Simulated Data

Simulated Data with Fixed Effects

```
set more off
clear all
set obs 10
g id=_n

*Generate Fixed Effects
g fe=int(100*runiform())

*Generate x1 that correlated with FE and error term with FE component
forvalue i=1(1)12 {
g x1`i'=50+5*fe+rnormal(500,100)
g u`i'=4*fe+rnormal(0,10)
}

*Generate y
forvalue i=1(1)12 {
g y`i'=1+0.1*x1`i'+u`i'
}
```

```
*Transform data to be panel data
reshape long x1 u y, i(id) j(month)

(note: j = 1 2 3 4 5 6 7 8 9 10 11 12)

Data                                  wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                          10   ->     120
Number of variables                     38   ->       6
j variable (12 values)                       ->   month
xij variables:
                           x11 x12 ... x112   ->   x1
                              u1 u2 ... u12   ->   u
                              y1 y2 ... y12   ->   y
-----------------------------------------------------------------------------
```

```
. xtset id month
       panel variable:  id (strongly balanced)
        time variable:  month, 1 to 12
                delta:  1 unit

. reg y x1

      Source |       SS           df       MS      Number of obs   =       120
-------------+----------------------------------   F(1, 118)       =    248.54
       Model |  865133.963         1  865133.963   Prob > F        =    0.0000
    Residual |  410739.424       118  3480.84257   R-squared       =    0.6781
-------------+----------------------------------   Adj R-squared   =    0.6753
       Total |  1275873.39       119  10721.6251   Root MSE        =    58.999

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .5703078   .0361751    15.77   0.000     .4986712    .6419443
       _cons |  -182.4773   28.51509    -6.40   0.000    -238.9449   -126.0096
------------------------------------------------------------------------------
```

```
. xtreg y x1, fe

Fixed-effects (within) regression               Number of obs      =       120
Group variable: id                              Number of groups   =        10

R-sq:                                           Obs per group:
     within  = 0.4078                                         min =        12
     between = 0.9744                                         avg =      12.0
     overall = 0.6781                                         max =        12

                                                F(1,109)           =     75.06
corr(u_i, Xb)  = 0.7786                         Prob > F           =    0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .0896361    .010346     8.66   0.000     .0691306    .1101416
       _cons |   189.5934   8.063686    23.51   0.000     173.6114    205.5753
-------------+----------------------------------------------------------------
     sigma_u |  96.881565
     sigma_e |   10.31783
         rho |  .98878508   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(9, 109) = 416.58                    Prob > F = 0.0000

. predict uhat, u

. corr y fe x1 u
(obs=120)

             |        y       fe       x1        u
-------------+------------------------------------
           y |   1.0000
          fe |   0.9919   1.0000
          x1 |   0.8235   0.7797   1.0000
           u |   0.9957   0.9937   0.7674   1.0000
```

Example

1. Panel Least Squares

The model can be stated as:

$$GPP_{it} = \beta_0 + \beta_1 G_{it} + \beta_2 LOAN_{it} + \beta_3 AGRI_{it} + u_{it} \qquad (1)$$

where:

GPP_{it} = Gross of provincial product of province i at time t.

G_{it} = Government spending on province i at time t.

$LOAN_{it}$ = Loan outstanding of province i at time t.

$AGRI_{it}$ = Agricultural production of province i at time t.

u_{it} = Disturbance term of province i at time t.

To estimate the model using STATA, we begin by setting the data set to be panel data setting.

```
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 0 to 9
                delta:  1 year
```

The model without any problems

```
. xtgls gpp g loan agri

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        homoskedastic
Correlation:   no autocorrelation

Estimated covariances      =         1          Number of obs      =        80
Estimated autocorrelations =         0          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =    748.54
Log likelihood             = -779.4455          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   1.920918   .8032843     2.39   0.017     .3465095    3.495326
        loan |   .5097789    .046617    10.94   0.000     .4184112    .6011465
        agri |  -53.35561   14.93147    -3.57   0.000    -82.62075   -24.09048
       _cons |   11683.41   1378.756     8.47   0.000     8981.096    14385.72
------------------------------------------------------------------------------
```

The model with Heteroskedasticity

```
. xtgls gpp g loan agri, igls panels(heteroskedastic)

Iteration 1: tolerance = .41982167
...
Iteration 22: tolerance = 6.777e-08

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic
Correlation:   no autocorrelation

Estimated covariances      =         8          Number of obs      =        80
Estimated autocorrelations =         0          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =   1140.96
Log likelihood             = -721.8528          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |  -.1485432   .2537412    -0.59   0.558    -.6458667    .3487803
        loan |   .6368605   .0246414    25.85   0.000     .5885644    .6851567
        agri |  -12.99571   2.658269    -4.89   0.000    -18.20582   -7.785598
       _cons |   6153.178   344.0983    17.88   0.000     5478.757    6827.598
------------------------------------------------------------------------------
```

Compute LR-Test for Heteroskedasticity

Firstly, keep estimated result of the model with Heteroskedaticity.

```
. estimates store hetero
```

Second, estimate model without Heteroskedaticity.

```
. xtgls gpp g loan agri

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        homoskedastic
Correlation:   no autocorrelation

Estimated covariances      =         1          Number of obs      =        80
Estimated autocorrelations =         0          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =    748.54
Log likelihood             = -779.4455          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   1.920918   .8032843     2.39   0.017     .3465095    3.495326
        loan |   .5097789    .046617    10.94   0.000     .4184112    .6011465
        agri |  -53.35561   14.93147    -3.57   0.000    -82.62075   -24.09048
       _cons |   11683.41   1378.756     8.47   0.000     8981.096    14385.72
------------------------------------------------------------------------------
```

Third, specify degree of freedom of the test.

```
. local df=e(N_g)-1
```

Finally, compute the LR-test = $2(\log L_{UR} - \log L_R)$.

```
. lrtest hetero, df(`df')
(log-likelihoods of null models cannot be compared)
```

```
Likelihood-ratio test                                 LR chi2(7)  =    115.19
(Assumption: . nested in hetero)                      Prob > chi2 =    0.0000
```

Wooldridge Test for Autocorrelation

```
. findit xtserial

. xtserial gpp g loan agri

Wooldridge test for autocorrelation in panel data
H0: no first order autocorrelation
    F(  1,       7) =    105.860
           Prob > F =      0.0000
```

Model with Heteroskedasticity and Cross-sectional Correlation

```
. xtgls gpp g loan agri, panels(correlated)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   no autocorrelation

Estimated covariances      =        36          Number of obs      =        80
Estimated autocorrelations =         0          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =   2262.05
Log likelihood             = -669.5133          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |    1.50661   .3003486     5.02   0.000     .9179374    2.095282
        loan |    .508701   .0178062    28.57   0.000     .4738015    .5436004
        agri |  -42.39837   3.959852   -10.71   0.000    -50.15954    -34.6372
       _cons |   10721.04    388.408    27.60   0.000     9959.776    11482.31
------------------------------------------------------------------------------
```

Model with Heteroskedasticity, Cross-sectional Correlation, and Autocorrelation

```
. xtgls gpp g loan agri, panels(correlated) corr(ar1)

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        heteroskedastic with cross-sectional correlation
Correlation:   common AR(1) coefficient for all panels  (0.8532)

Estimated covariances      =        36          Number of obs      =        80
Estimated autocorrelations =         1          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =    265.49
Log likelihood             = -610.0246          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |  -.2611345   .3368487    -0.78   0.438    -.9213458    .3990767
        loan |    .518529   .0439935    11.79   0.000     .4323034    .6047546
        agri |   .4994303   1.297904     0.38   0.700    -2.044414    3.043275
       _cons |   8643.839   311.1028    27.78   0.000     8034.088    9253.589
------------------------------------------------------------------------------
```

LR-Test for the Problems[1]

```
. estimates store hauto

. xtgls gpp g loan agri

Cross-sectional time-series FGLS regression

Coefficients:  generalized least squares
Panels:        homoskedastic
Correlation:   no autocorrelation

Estimated covariances      =         1          Number of obs      =        80
Estimated autocorrelations =         0          Number of groups   =         8
Estimated coefficients     =         4          Time periods       =        10
                                                Wald chi2(3)       =    748.54
Log likelihood             = -779.4455          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   1.920918   .8032843     2.39   0.017     .3465095    3.495326
        loan |   .5097789    .046617    10.94   0.000     .4184112    .6011465
        agri |  -53.35561   14.93147    -3.57   0.000    -82.62075   -24.09048
       _cons |   11683.41   1378.756     8.47   0.000     8981.096    14385.72
------------------------------------------------------------------------------

. local df2=e(N_g)-1

. lrtest hauto, df(`df')
(log-likelihoods of null models cannot be compared)

Likelihood-ratio test                                 LR chi2(7)  =    338.84
(Assumption: . nested in hauto)                       Prob > chi2 =    0.0000

. predict gpp_gls, xb
```

[1] In order to obtain an appropriated LR-test, the estimated likelihood-value should be computed from MLE process or MLE equivalent. In this case, GLS does not produce MLE equivalent estimated result. Thus, the LR-test computed from these estimated resulted might not be appropriated.

2. Fixed Effect Model

The model can also be applied as fixed effect model:

$$GPP_{it} = \beta_{i0} + \beta_1 G_{it} + \beta_2 LOAN_{it} + \beta_3 AGRI_{it} + u_{it} \tag{2}$$

```
. xtreg gpp g loan agri, fe

Fixed-effects (within) regression               Number of obs      =        80
Group variable (i): id                          Number of groups   =         8

R-sq:  within  = 0.7585                         Obs per group: min =        10
       between = 0.8963                                        avg =      10.0
       overall = 0.8346                                        max =        10

                                                F(3,69)            =     72.25
corr(u_i, Xb)  = 0.8353                         Prob > F           =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   1.442314   .2248462     6.41   0.000      .993758     1.89087
        loan |   .1171806   .0358101     3.27   0.002     .0457415    .1886197
        agri |   11.94431   5.654815     2.11   0.038     .6632614    23.22536
       _cons |   12472.26   667.2543    18.69   0.000     11141.12     13803.4
-------------+----------------------------------------------------------------
     sigma_u |  10363.779
     sigma_e |  906.01137
         rho |  .99241554   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(7, 69) =   226.72             Prob > F = 0.0000

. estimates store fixed

. predict gpp_fe, xbu
```

To compute fixed effect, use the following command:

```
. predict ai, u

. table id, c(mean ai)

----------------------
       ID |   mean(ai)
----------+-----------
        1 |   21100.33
        2 |   2711.693
        3 |   5643.245
        4 |  -10611.55
        5 |   2260.344
        6 |  -6132.393
        7 |   -7620.03
        8 |  -7351.638
----------------------
```

3. Random Effect Model

The model can also be specified as random effect model:

$$GPP_{it} = \beta_0 + \beta_1 G_{it} + \beta_2 LOAN_{it} + \beta_3 AGRI_{it} + u_{it} \quad (3)$$

and $u_{it} = v_i + w_t + \varepsilon_{it}$

where: v_i = cross-section random effect

w_t = time random effect

ε_{it} = residual term

```
. xtreg gpp g loan agri, re

Random-effects GLS regression                   Number of obs      =        80
Group variable (i): id                          Number of groups   =         8

R-sq:  within  = 0.6104                         Obs per group: min =        10
       between = 0.9007                                        avg =      10.0
       overall = 0.8840                                        max =        10

Random effects u_i ~ Gaussian                   Wald chi2(3)       =    246.85
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |   .4391357   .4329641     1.01   0.310    -.4094584     1.28773
        loan |   .4686367    .045272    10.35   0.000     .3799053    .5573682
        agri |   4.008584   11.51057     0.35   0.728    -18.55171    26.56888
       _cons |   8736.887   1368.123     6.39   0.000     6055.414    11418.36
-------------+----------------------------------------------------------------
     sigma_u |  1009.6082
     sigma_e |  906.01137
         rho |  .55392236   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. estimates store random

. predict gpp_re, xbu
```

To compute random effect, use the following command:

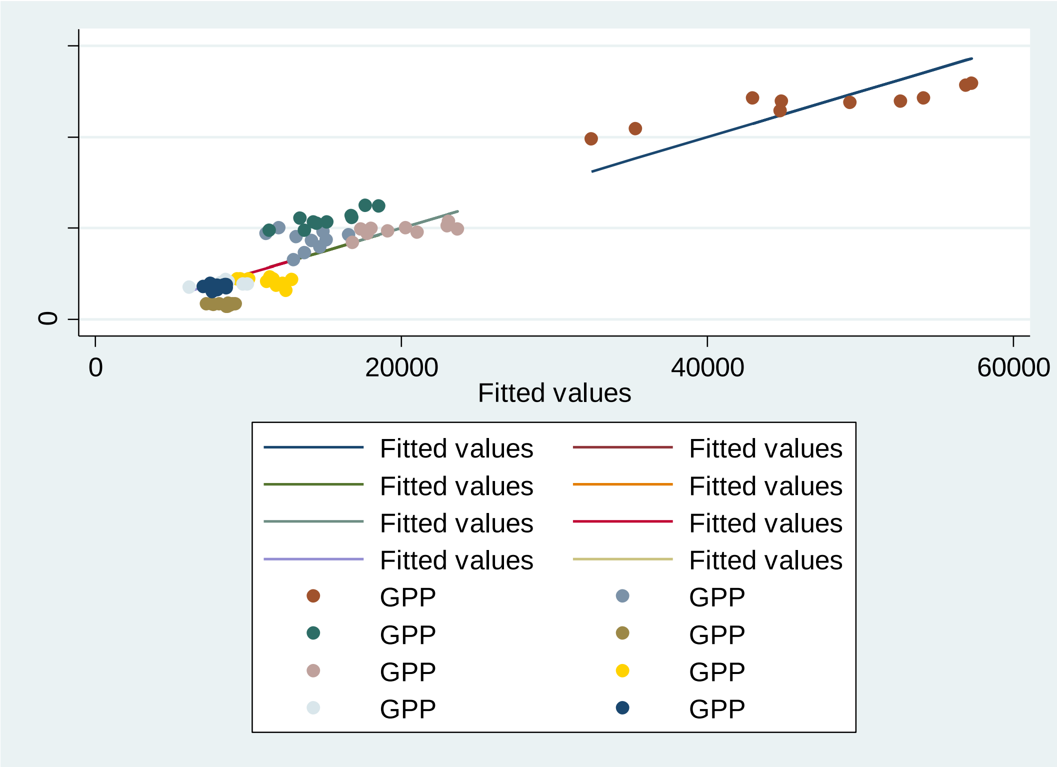
```
. predict ri, u

. table id, c(mean ri)

----------------------
       ID |   mean(ri)
----------+-----------
        1 |    5131.64
        2 |   4695.539
        3 |   6389.667
        4 |  -6270.812
        5 |   624.1134
        6 |  -3240.626
        7 |  -3939.543
        8 |  -3389.978
----------------------
```

To compute perform Hausman test, use the following command:

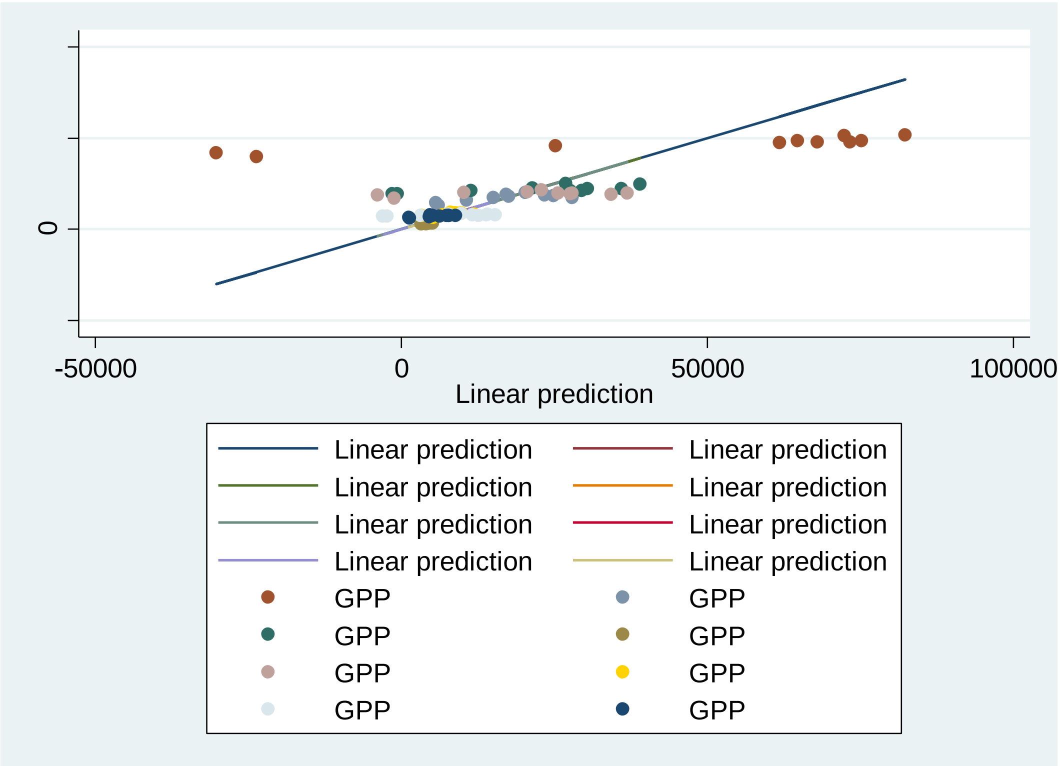
```
. hausman random fixed

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     random       fixed        Difference          S.E.
-------------+-------------------------------------------------------------
           g |    .4391357     1.442314       -1.003178        .3700028
        loan |    .4686367     .1171806        .3514562        .0276982
        agri |    4.008584     11.94431       -7.935725        10.02578
---------------------------------------------------------------------------
                           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(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      285.14
                Prob>chi2 =      0.0000

. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects:

        gpp[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                     gpp |   1.78e+08       13349.59
                       e |   820856.6       906.0114
                       u |    1019309       1009.608

        Test:   Var(u) = 0
                            chi2(1) =   156.18
                        Prob > chi2 =     0.0000

. xttest1

Tests for the error component model:

        gpp[id,t] = Xb + u[id] + v[id,t]
           v[id,t] = rho v[id,(t-1)] + e[id,t]

        Estimated results:
                                 Var     sd = sqrt(Var)
                ---------+-----------------------------
                     gpp |   1.78e+08       13349.59
                       e |   820856.6      906.01137
                       u |    1019309      1009.6082

        Tests:
           Random Effects, Two Sided:
           LM(Var(u)=0)       =  156.18   Pr>chi2(1) =  0.0000
           ALM(Var(u)=0)      =   93.54   Pr>chi2(1) =  0.0000

           Random Effects, One Sided:
           LM(Var(u)=0)       =   12.50   Pr>N(0,1)  =  0.0000
           ALM(Var(u)=0)      =    9.67   Pr>N(0,1)  =  0.0000

           Serial Correlation:
           LM(rho=0)          =   73.99   Pr>chi2(1) =  0.0000
           ALM(rho=0)         =   11.35   Pr>chi2(1) =  0.0008

           Joint Test:
           LM(Var(u)=0,rho=0) =  167.53   Pr>chi2(2) =  0.0000
```

```
. xtreg gpp g loan agri, be

Between regression (regression on group means)  Number of obs      =        80
Group variable: id                              Number of groups   =         8

R-sq:  within  = 0.6574                         Obs per group: min =        10
       between = 0.9968                                        avg =      10.0
       overall = 0.5026                                        max =        10

                                                F(3,4)             =    417.66
sd(u_i + avg(e_i.))=  1049.473                  Prob > F           =    0.0000

------------------------------------------------------------------------------
         gpp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           g |    25.9389   2.744986     9.45   0.001      18.3176     33.5602
        loan |  -.6168582   .1320047    -4.67   0.009     -.983362   -.2503544
        agri |  -75.30935   14.90695    -5.05   0.007    -116.6977   -33.92101
       _cons |   9538.695   1442.741     6.61   0.003     5533.004    13544.39
------------------------------------------------------------------------------

. predict gpp_be, xb

. twoway (line gpp_gls gpp_gls if id==1) (line gpp_gls gpp_gls if id==2) (line
gpp_gls gpp_gls if id==3) (line gpp_gls gpp_gls if id==4) (line gpp_gls gpp_gls
if id==5) (line gpp_gls gpp_gls if id==6) (line gpp_gls gpp_gls if id==7) (line
gpp_gls gpp_gls if id==8) (scatter gpp gpp_gls if id==1) (scatter gpp gpp_gls if
id==2) (scatter gpp gpp_gls if id==3) (scatter gpp gpp_gls if id==4) (scatter gpp
gpp_gls if id==5) (scatter gpp gpp_gls if id==6) (scatter gpp gpp_gls if id==7)
(scatter gpp gpp_gls if id==8)

. twoway (line gpp_be gpp_be if id==1) (line gpp_be gpp_be if id==2) (line gpp_be
gpp_be if id==3) (line gpp_be gpp_be if id==4) (line gpp_be gpp_be if id==5)
(line gpp_be gpp_be if id==6) (line gpp_be gpp_be if id==7) (line gpp_be gpp_be
if id==8) (scatter gpp gpp_be if id==1) (scatter gpp gpp_be if id==2) (scatter
gpp gpp_be if id==3) (scatter gpp gpp_be if id==4) (scatter gpp gpp_be if id==5)
(scatter gpp gpp_be if id==6) (scatter gpp gpp_be if id==7) (scatter gpp gpp_be
if id==8)

. twoway (line gpp_fe gpp_fe if id==1) (line gpp_fe gpp_fe if id==2) (line gpp_fe
gpp_fe if id==3) (line gpp_fe gpp_fe if id==4) (line gpp_fe gpp_fe if id==5)
(line gpp_fe gpp_fe if id==6) (line gpp_fe gpp_fe if id==7) (line gpp_fe gpp_fe
if id==8) (scatter gpp gpp_fe if id==1) (scatter gpp gpp_fe if id==2) (scatter
gpp gpp_fe if id==3) (scatter gpp gpp_fe if id==4) (scatter gpp gpp_fe if id==5)
(scatter gpp gpp_fe if id==6) (scatter gpp gpp_fe if id==7) (scatter gpp gpp_fe
if id==8)

. twoway (line gpp_re gpp_re if id==1) (line gpp_re gpp_re if id==2) (line gpp_re
gpp_re if id==3) (line gpp_re gpp_re if id==4) (line gpp_re gpp_re if id==5)
(line gpp_re gpp_re if id==6) (line gpp_re gpp_re if id==7) (line gpp_re gpp_re
if id==8) (scatter gpp gpp_re if id==1) (scatter gpp gpp_re if id==2) (scatter
gpp gpp_re if id==3) (scatter gpp gpp_re if id==4) (scatter gpp gpp_re if id==5)
(scatter gpp gpp_re if id==6) (scatter gpp gpp_re if id==7) (scatter gpp gpp_re
if id==8)
```

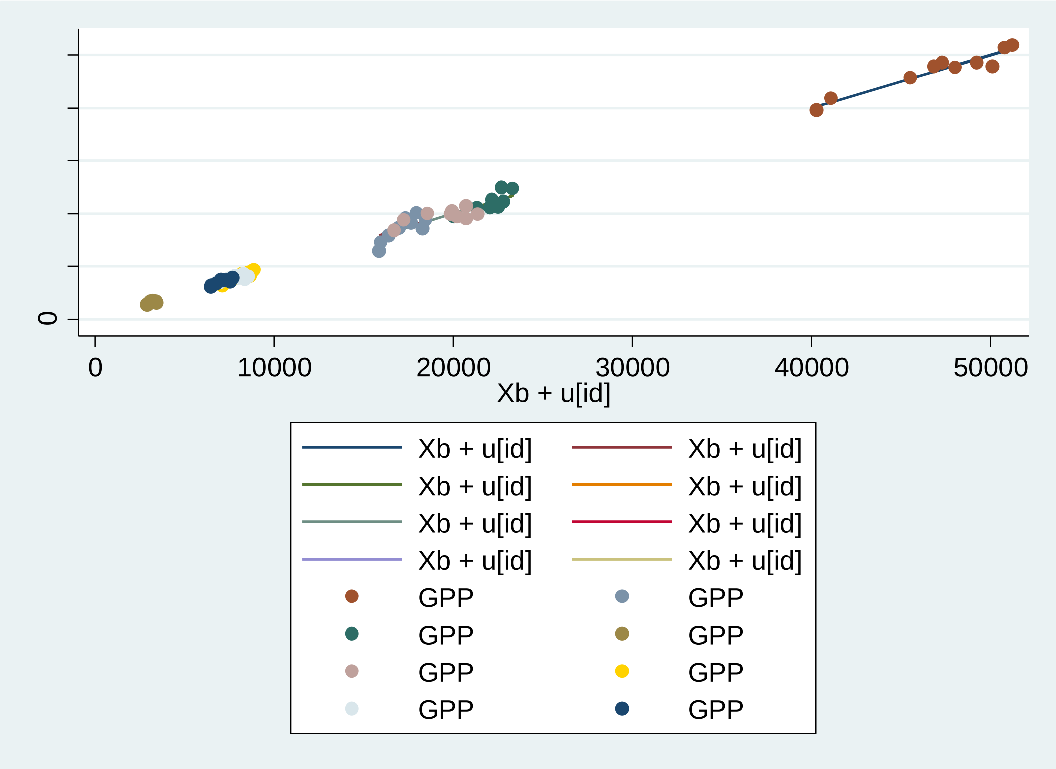
Pooled GLS



Between



Fixed Effect



Random Effect