Longitudinal Data Analysis
Instructor: Natasha Sarkisian

Panel Data Analysis: Fixed Effects Models

Panel data are longitudinal data where multiple cases (N cases) are observed over T time periods (two or more), resulting in N x T observations. The name "panel data" is based on survey panels that typically have many respondents that are observed for only a few time periods. Sometimes people reserve the term "panel data" for those cases where N is relatively large and T is relatively small (say, 5 or less). Also, if N is very large relative to T, people say that the dataset is "cross-sectionally dominated." In contrast, if you have very long sequences of data over time, but very few cases, people refer to such data as "time series cross section data." Still, similar techniques can be used to analyze these data, although researchers should pay closer attention to time processes.

There are two kinds of information in such data, regardless of the type: the cross-sectional information reflected in the differences between subjects, and the time-series or within-subject information reflected in the changes within subjects over time. Panel data regression techniques allow us to take advantage of these different types of information.

You should differentiate panel data from the data where multiple independent cross-sections (e.g., different samples at each time point) are combined into a single dataset. There are a few reasons why we might want to combine cross-sections over time. We may be interested in changes over time and such data may be the only ones available. Or we may be simply interested in increasing sample size available for analysis. In any case, using independently pooled cross-sections raises only minor statistical complications. Typically, to reflect the fact that the population distribution may differ over time, we use year dummy variables in the models. This model assumes that the effect of each independent variable remains constant over time. This may or may not be true, and we can test that. If you suspect that effects of some variables change over time, these variables can be interacted with year dummies.

We will focus on data generated by longitudinal study designs – multiple observations on the same unit. Although it may seem that such data are too much hassle, they offer very important advantages, as we can:

1. Assess changes over time
2. Test causal arguments
3. Increase sample size
4. Control for some types of omitted variables even without observing them, by observing changes in the dependent variable over time. This controls for omitted variables that differ among cases but are constant over time. It is also possible to use panel data to control for omitted variables that vary over time but are constant across cases.

For our examples, we'll use a crossnational dataset.

```
. use "C:\Documents and Settings\SARKISIN\My Documents\crossnatfull.dta", clear
```

To be able to use panel data techniques, we want to make sure that the data are in long form (multiple observations per unit – each year is on the separate line). That is, if they are in the wide form to start with (if you have only one observation per case and separate variables for each time period), you should use reshape command to convert the data to long form.

Stata provides a number of tools for analyzing panel data. The commands all begin with the prefix xt. To use these commands, we first need to tell Stata that our dataset is a panel dataset. We need to have a variable that identifies the units (for example, a country or person id) and a time variable. To set a dataset as a panel, we need to sort the data by the panel variable and then by the date variable within the panel variable, and then use xtset command:

```
. sort v1 v2

. xtset v1 v2, yearly
       panel variable:  v1, 10 to 1300
        time variable:  v2, 1815 to 1973, but with gaps
```

To describe the data setup:

```
. xtdes

      v1:  10, 20, ..., 1300                                 n =        166
      v2:  1815, 1816, ..., 1973                             T =        148
           Delta(v2) = 1; (1973-1815)+1 = 159
           (v1*v2 uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       4      14        28        98     148     148

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+-----------------------------------------------------
       18     10.84   10.84 |
......................................................................................
...............
> ............................................11111111111111
       18     10.84   21.69 |
11111111111111111111111111111111111111111111111111111111111111111111111111111111111111
1111111111111..
> ...111111111111111111111......1111111111111111111111111111
        7      4.22   25.90 |
......................................................................................
...............
> ..............................................111111111111
        6      3.61   29.52 |
1111111111111111111111111111111111111111111111........................................
...............
> ..........................................................
        6      3.61   33.13 |
1111111111111111111111111111111111111111111111111111..................................
...............
> ..........................................................
        5      3.01   36.14 |
......................................................................................
...............
```

```
> .......................................................111
        5      3.01   39.16 |
......................................................................................
...............
> ................................11111111111111111111111111
        5      3.01   42.17 |
......................................................................................
...............
> ...111111111111111111111......1111111111111111111111111111
        4      2.41   44.58 |
......................................................................................
...............
> ....................................................111111
       92     55.42  100.00 | (other patterns)
 ---------------------------+---------------------------------------------------------
----------------------------------------------
> ----------------------------------------------------------
      166    100.00         |
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXX..
> ...XXXXXXXXXXXXXXXXXXXXX......XXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

Next, we want to describe some individual variables. Compare the output of regular summarize or tabulate with their xt versions:

```
. sum v41

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         v41 |      6280    156.2275     388.234          0       4500

. xtsum v41

Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
v41      overall |  156.2275    388.234          0       4500 |     N =    6280
         between |              197.6826         0   1301.811 |     n =     136
         within  |              279.3011  -885.5833  3354.417 | T-bar = 46.1765

. tab v114
    type of |
     regime |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |      7,877       90.91       90.91
          2 |        569        6.57       97.47
          3 |        110        1.27       98.74
          4 |        109        1.26      100.00
------------+-----------------------------------
      Total |      8,665      100.00

. xttab  v114
                  Overall             Between            Within
     v114 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |    7877     90.91        158     99.37          91.17
        2 |     569      6.57         58     36.48          12.65
        3 |     110      1.27         27     16.98           5.08
        4 |     109      1.26          6      3.77          14.65
----------+-----------------------------------------------------
    Total |    8665    100.00        249    156.60          61.70
                                (n = 159)
```

There is also a special command to investigate transitions:

```
. xttrans v114

   type of |       type of regime
    regime |         1          2          3          4 |     Total
-----------+--------------------------------------------+----------
         1 |     98.28       1.16       0.53       0.03 |    100.00
         2 |     16.15      83.12       0.55       0.18 |    100.00
         3 |     20.00      18.10      61.90       0.00 |    100.00
         4 |      6.42       0.92       0.00      92.66 |    100.00
-----------+--------------------------------------------+----------
     Total |     90.87       6.62       1.28       1.22 |    100.00

. xttrans v114, freq

   type of |       type of regime
    regime |         1          2          3          4 |     Total
-----------+--------------------------------------------+----------
         1 |     7,610         90         41          2 |     7,743
           |     98.28       1.16       0.53       0.03 |    100.00
-----------+--------------------------------------------+----------
         2 |        88        453          3          1 |       545
           |     16.15      83.12       0.55       0.18 |    100.00
-----------+--------------------------------------------+----------
         3 |        21         19         65          0 |       105
           |     20.00      18.10      61.90       0.00 |    100.00
-----------+--------------------------------------------+----------
         4 |         7          1          0        101 |       109
           |      6.42       0.92       0.00      92.66 |    100.00
-----------+--------------------------------------------+----------
     Total |     7,726        563        109        104 |     8,502
           |     90.87       6.62       1.28       1.22 |    100.00
```

For our examples, we'll focus on predicting the size of the military:

```
. des v41 v76 v5 v7 v19
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
v41             int    %8.0g                  size of military
v76             int    %8.0g                  pri + sec enr per capita
v5              int    %8.0g                  population density
v7              int    %8.0g                  urb 100,000+ per capita
v19             long   %12.0g                 revenue per capita

. sum v41 v76 v5 v7 v19
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         v41 |      6280    156.2275     388.234          0       4500
         v76 |      6021    1092.985    648.7898          0       3444
          v5 |      8816    1196.446    1501.329          5      13020
          v7 |      7811    97.85418    110.9803          0        671
         v19 |      6334     6089.75    14940.87          7     301233
```

```
. xtsum v41 v76 v5 v7 v19
Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
v41      overall |  156.2275    388.234          0       4500 |     N =    6280
         between |             197.6826          0   1301.811 |     n =     136
         within  |             279.3011  -885.5833   3354.417 | T-bar = 46.1765
                 |                                            |
v76      overall |  1092.985   648.7898          0       3444 |     N =    6021
         between |             672.4557   92.55556       3222 |     n =     146
         within  |             397.2686  -118.4437   2827.602 | T-bar = 41.2397
                 |                                            |
v5       overall |  1196.446   1501.329          5      13020 |     N =    8816
         between |             1772.606   14.51064      12580 |     n =     158
         within  |             599.8466  -1616.034   7056.602 | T-bar = 55.7975
                 |                                            |
v7       overall |  97.85418   110.9803          0        671 |     N =    7811
         between |             100.4004          0   662.6667 |     n =     152
         within  |             71.53064   -204.003   511.7564 | T-bar = 51.3882
                 |                                            |
v19      overall |   6089.75   14940.87          7     301233 |     N =    6334
         between |             17454.64         35   157729.2 |     n =     153
         within  |             11396.27  -95325.42   149593.6 | T-bar = 41.3987
```

While it is possible to use ordinary multiple regression techniques on panel data, they are usually not appropriate because of non-independence of observations, heteroskedasticity (both across time and across units) and autocorrelation. For comparison purposes, however, let’s use OLS to estimate our model:

```
. reg v41 v76 v5 v7 v19
      Source |       SS       df       MS              Number of obs =    4945
-------------+------------------------------           F(  4,  4940) =  142.19
       Model |  77569027.8     4    19392257           Prob > F      =  0.0000
    Residual |   673742407  4940  136385.103           R-squared     =  0.1032
-------------+------------------------------           Adj R-squared =  0.1025
       Total |   751311435  4944  151964.287           Root MSE      =   369.3

------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0107072   .0107961     0.99   0.321    -.0104579    .0318723
          v5 |   .0038175   .0032207     1.19   0.236    -.0024966    .0101315
          v7 |   .2710273   .0597177     4.54   0.000     .1539542    .3881005
         v19 |   .0073414   .0004123    17.81   0.000      .006533    .0081497
       _cons |   44.82276   11.10465     4.04   0.000     23.05272     66.5928
```

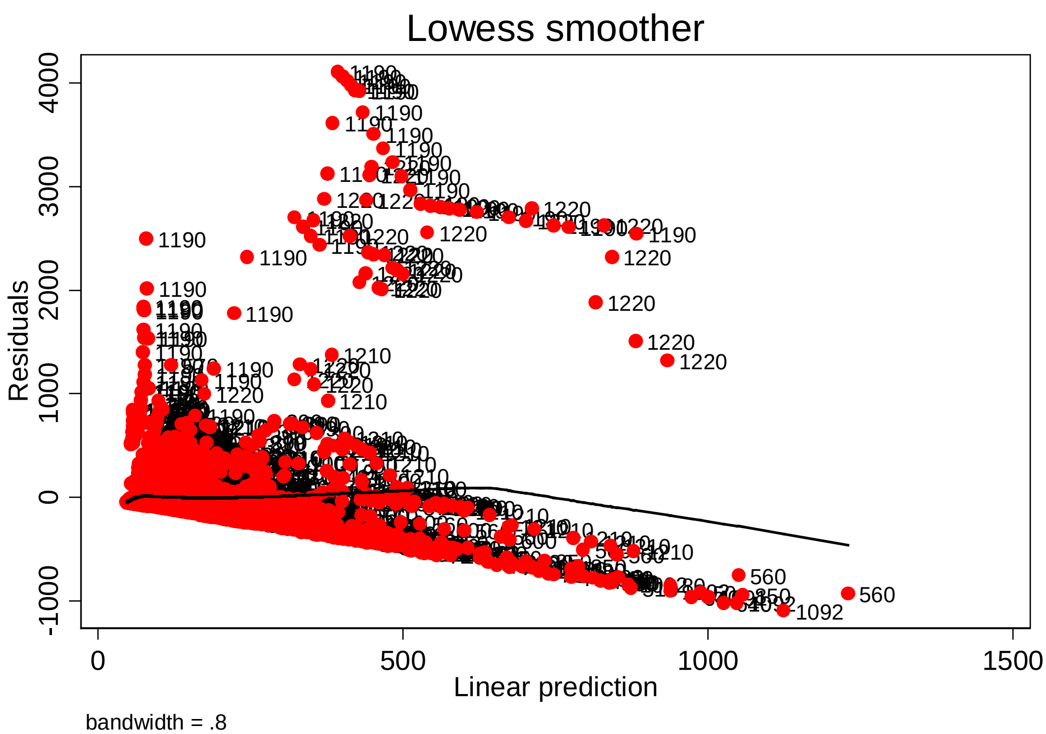
```
. reg v41 v76 v5 v7 v19, cluster(v1)
Linear regression                                      Number of obs =    4945
                                                       F(  4,   123) =    4.08
                                                       Prob > F      =  0.0039
                                                       R-squared     =  0.1032
Number of clusters (v1) = 124                          Root MSE      =   369.3
------------------------------------------------------------------------------
             |               Robust
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0107072   .0478825     0.22   0.823    -.0840734    .1054878
          v5 |   .0038175   .0243906     0.16   0.876    -.0444623    .0520972
          v7 |   .2710273   .2574062     1.05   0.294    -.2384924    .7805471
         v19 |   .0073414   .0051641     1.42   0.158    -.0028806    .0175634
       _cons |   44.82276   56.24492     0.80   0.427    -66.51061    156.1561
------------------------------------------------------------------------------
```

Here, the model we estimated remains unchanged, but the standard errors are adjusted to reflect the fact that, effectively, we have much fewer independent cases than the N=4945. Note that the resulting standard errors are much larger. But an even bigger problem of using OLS is that we are still assuming that error term variances are equal across countries. But even a simple residuals vs predicted values graph shows us that this is not true:

```
. predict resid, res
(4032 missing values generated)

. predict fitted, xb
(3685 missing values generated)

. lowess resid fitted, mlabel(v1) mcolor(red)
```



To avoid the problems of heteroscedasticity across units, we estimate a model that allows for each country to have its own intercept – a fixed effects model:

```
. xtreg  v41 v76 v5 v7 v19, fe
Fixed-effects (within) regression               Number of obs      =      4945
Group variable (i): v1                          Number of groups   =       124

R-sq:  within  = 0.1806                         Obs per group: min =         1
       between = 0.0051                                        avg =      39.9
       overall = 0.0600                                        max =       148

                                                F(4,4817)          =    265.43
corr(u_i, Xb)  = -0.2548                        Prob > F           =    0.0000
------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0445623    .011977     3.72   0.000      .021082    .0680426
```

```
         v5 |  -.0637468    .007078    -9.01   0.000    -.0776229   -.0498708
         v7 |   .5901123   .0757072     7.79   0.000     .4416916    .7385331
        v19 |   .0080274   .0003284    24.44   0.000     .0073836    .0086712
      _cons |   51.30225   11.77021     4.36   0.000     28.22727    74.37723
-------------+----------------------------------------------------------------
     sigma_u |  278.56116
     sigma_e |  242.31081
         rho |  .56926004   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(123, 4817) =    54.13           Prob > F = 0.0000
```

Although the individual country intercepts are not presented in the output, we can verify that we get the same model:

```
. xi: reg v41 v76 v5 v7 v19 i.v1
i.v1              _Iv1_10-1300        (naturally coded; _Iv1_10 omitted)

      Source |       SS       df       MS              Number of obs =    4945
-------------+------------------------------           F(127,  4817) =   62.83
       Model |   468483540   127  3688846.77           Prob > F      =  0.0000
    Residual |   282827895  4817  58714.5308           R-squared     =  0.6236
-------------+------------------------------           Adj R-squared =  0.6136
       Total |   751311435  4944  151964.287           Root MSE      =  242.31

------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0445623    .011977     3.72   0.000      .021082    .0680426
          v5 |  -.0637468    .007078    -9.01   0.000    -.0776229   -.0498708
          v7 |   .5901123   .0757072     7.79   0.000     .4416916    .7385331
         v19 |   .0080274   .0003284    24.44   0.000     .0073836    .0086712
     _Iv1_20 |  -63.98205   70.12371    -0.91   0.362    -201.4565    73.49244
     _Iv1_30 |  -191.1164   96.24602    -1.99   0.047    -379.8026   -2.430294
     _Iv1_40 |  -266.1745   65.57011    -4.06   0.000    -394.7219   -137.6272
     _Iv1_50 |  -564.9821    72.4991    -7.79   0.000    -707.1135   -422.8508
     _Iv1_51 |  (dropped)
     _Iv1_60 |   245.8339   67.40873     3.65   0.000     113.6821    377.9858
 [Output omitted]
   _Iv1_1300 |  -269.6325   96.63103    -2.79   0.005    -459.0734   -80.19153
       _cons |   75.37176   57.26024     1.32   0.188    -36.88446     187.628
------------------------------------------------------------------------------
```

These country-specific intercepts can also be viewed as part of the decomposed residuals: $Y_{it} = \alpha + Xit\beta + u_i + e_{it}$ where u_i is the effect of country i and e_{it} is the residual effect for time point t within that country. In a fixed effects model, each of country residuals u_i is assigned a specific value – it's a fixed intercept for each country. Because country intercepts are essentially separate independent variables in a fixed effects models, these intercepts are allowed to be correlated with the independent variables in the model –e.g., in our output we have

```
corr(u_i, Xb)  = -0.2548
```

What this means is that we do not use our independent variables to explain country-specific effects – they are just set aside and we focus on explaining change over time. One big advantage of doing this is that we eliminate all country-specific effects, including those that we could not explicitly model with the variables at hand (omitted variables), so we can focus explicitly on change over time. A disadvantage, however, is that the data on cross-sectional variation are available but not used in estimating independent variables' effects.

As a preliminary step to estimating a fixed effects model, it is usually helpful to estimate a fully unconditional model:

```
. xtreg v41, fe
Fixed-effects (within) regression               Number of obs      =      6280
Group variable (i): v1                          Number of groups   =       136

R-sq:  within  = 0.0000                         Obs per group: min =         1
       between = 0.0019                                        avg =      46.2
       overall = 0.0000                                        max =       148

                                                F(0,6144)          =      0.00
corr(u_i, Xb)  = -0.0000                        Prob > F           =         .
------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   156.2275   3.562973    43.85   0.000     149.2429    163.2122
-------------+----------------------------------------------------------------
     sigma_u |  197.68264
     sigma_e |  282.35297
         rho |  .32893849   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(135, 6144) =    42.42           Prob > F = 0.0000
```

Its most important function is to provide information about outcome variability at each of the two levels. Sigma_e will provide information about level-1 (across time) variability, and sigma_u will provide information on level-2 (across country) variability. So running this model allows us to decompose the variance in the dependent variable into variance components -- into within-group and between-group variance (although they are expressed as standard deviations – to get variances, we'd have to square them). This model does not explain anything, but it allows us to evaluate whether there is variation in group means (here, country means), and how much of it. That's why it is always a good idea to run this basic model when starting the analyses – it's the null model of our regression analysis. If we find that there is no significant variation across countries, then there is no need for a fixed effects model. That significance test is the F test below the model.

The proportion of variance due to group-level variation in means can be calculated as

$\rho = \text{sigma\_u}^2 / (\text{sigma\_u}^2 + \text{sigma\_e}^2)$

It can be interpreted as the proportion of variance explained by the unit effects. It can also be interpreted as the average correlation between two randomly chosen time points that are in the same unit; therefore, it is also known as intra-class correlation. Here, we get:

```
. di 197.68264^2 / (197.68264^2 + 282.35297^2)
.32893848
```

which is the rho number in the table. So 33% of the total variance in military size is due to country-specific effects.

Predict command after xtreg, fe allows us to get predicted values and residuals. It allows the following options:

```
      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
```

So to obtain two sets of residuals, level 1 (e) and level 2 (u), we run:

```
. qui xtreg v41 v76 v5 v7 v19, fe

. predict level1, e
(4032 missing values generated)

. predict level2, u
(4032 missing values generated)
```

We can use these residuals to conduct regression diagnostics. Moreover, we can apply all the OLS diagnostic tools to the model with many dummies we estimated earlier (for more details on OLS diagnostics, see SC704 class notes at http://www.sarkisian.net/sc704). Alternatively, Stata also has a convenient feature that makes it easier to search for the best model specification; xtdata command allows you to generate a dataset that will produce fixed effects models with regular regress commands. This will also enable you to use OLS diagnostics tools.

```
. use "C:\Documents and Settings\SARKISIN\My Documents\Teaching Grad
Statistics\SC706\sc706\crossnatfull.dta", clear

. xtdata v2 v41 v76 v5 v7 v19, fe clear

. reg  v41 v5 v7 v19 v76
      Source |       SS       df       MS              Number of obs =    4945
-------------+------------------------------           F(  4,  4940) =  272.21
       Model |  62338344.4     4  15584586.1           Prob > F      =  0.0000
    Residual |   282827895  4940  57252.6103           R-squared     =  0.1806
-------------+------------------------------           Adj R-squared =  0.1799
       Total |   345166239  4944  69815.1778           Root MSE      =  239.28
------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0445623   .0118269     3.77   0.000     .0213763    .0677483
          v5 |  -.0637468   .0069893    -9.12   0.000     -.077449   -.0500447
          v7 |   .5901123   .0747588     7.89   0.000     .4435519    .7366727
         v19 |   .0080274   .0003243    24.75   0.000     .0073917    .0086632
       _cons |   51.30225   11.62275     4.41   0.000     28.51649    74.08801
------------------------------------------------------------------------------
```

Thus, you can search for model specification using this dataset, and then estimate the final model you select using your original dataset. (You cannot use the model from the modified dataset because significance tests will be incorrect as the degrees of freedom do not take into account the implicit presence of country dummies.)

It addition to the one-way fixed effects model that we just estimated, we could also consider estimating a two-way fixed-effects model. It is a good idea in most cases to include time into the model when estimating a fixed-effects model. Unfortunately, Stata does not automatically estimated two-way FE models – we have to introduce year dummies:

```
. xi: xtreg v41 v76 v5 v7 v19 i.v2, fe
i.v2              _Iv2_1815-1973      (naturally coded; _Iv2_1815 omitted)
Fixed-effects (within) regression               Number of obs      =      4945
Group variable (i): v1                          Number of groups   =       124
R-sq:  within  = 0.2297                         Obs per group: min =         1
       between = 0.0033                                        avg =      39.9
       overall = 0.0816                                        max =       148
                                                F(151,4670)        =      9.22
corr(u_i, Xb)  = -0.3136                        Prob > F           =    0.0000
------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

v76	.1150898	.0146829	7.84	0.000	.0863044	.1438753
v5	-.0575626	.0076677	-7.51	0.000	-.0725949	-.0425303
v7	.6014044	.0910177	6.61	0.000	.4229667	.779842
v19	.0105129	.0004057	25.91	0.000	.0097176	.0113082
_Iv2_1816	-4.611157	337.4329	-0.01	0.989	-666.139	656.9166
_Iv2_1817	-8.376042	337.433	-0.02	0.980	-669.904	653.152
_Iv2_1818	-13.15861	337.4332	-0.04	0.969	-674.6869	648.3697
_Iv2_1819	-13.14016	337.4332	-0.04	0.969	-674.6686	648.3883
_Iv2_1820	-15.4067	337.4335	-0.05	0.964	-676.9356	646.1222
_Iv2_1821	-19.04943	337.4337	-0.06	0.955	-680.5788	642.48
_Iv2_1822	-22.15934	337.434	-0.07	0.948	-683.6893	639.3707
_Iv2_1823	-25.60122	337.4344	-0.08	0.940	-687.1319	635.9295
_Iv2_1824	-27.20152	337.4348	-0.08	0.936	-688.733	634.33
_Iv2_1825	-30.20079	337.4352	-0.09	0.929	-691.7331	631.3315
_Iv2_1826	-33.83856	337.4357	-0.10	0.920	-695.3717	627.6946
_Iv2_1827	-35.70116	337.4362	-0.11	0.916	-697.2353	625.833
_Iv2_1828	-39.34672	337.4367	-0.12	0.907	-700.882	622.1886
_Iv2_1829	-42.06872	337.4373	-0.12	0.901	-703.6052	619.4678
_Iv2_1830	-44.2462	337.4375	-0.13	0.896	-705.7829	617.2905
_Iv2_1831	-45.12801	337.4377	-0.13	0.894	-706.6651	616.4091
_Iv2_1832	-42.19088	337.4374	-0.13	0.901	-703.7274	619.3457
_Iv2_1833	-68.38697	337.4659	-0.20	0.839	-729.9794	593.2054
_Iv2_1834	-72.5051	337.4749	-0.21	0.830	-734.1152	589.105
_Iv2_1835	-76.84239	337.4851	-0.23	0.820	-738.4724	584.7877
_Iv2_1836	-80.56155	337.496	-0.24	0.811	-742.213	581.0899
_Iv2_1837	-85.53126	337.5081	-0.25	0.800	-747.2065	576.144
_Iv2_1838	-87.45174	337.514	-0.26	0.796	-749.1385	574.235
_Iv2_1839	-89.98519	337.5199	-0.27	0.790	-751.6834	571.7131
_Iv2_1840	-91.83314	337.526	-0.27	0.786	-753.5434	569.8771
_Iv2_1841	-89.39411	337.5341	-0.26	0.791	-751.1204	572.3321
_Iv2_1842	-84.88844	337.5422	-0.25	0.801	-746.6304	576.8536
_Iv2_1843	-81.4473	337.5503	-0.24	0.809	-743.2052	580.3106
_Iv2_1844	-74.86112	337.5515	-0.22	0.824	-736.6214	586.8991
_Iv2_1845	-67.75075	337.5523	-0.20	0.841	-729.5126	594.0111
_Iv2_1846	-61.21163	337.5535	-0.18	0.856	-722.9758	600.5526
_Iv2_1847	-64.73685	337.5865	-0.19	0.848	-726.5657	597.092
_Iv2_1848	-58.4546	337.5786	-0.17	0.863	-720.268	603.3588
_Iv2_1849	-77.81048	294.6038	-0.26	0.792	-655.373	499.7521
_Iv2_1850	-74.84566	294.601	-0.25	0.799	-652.4028	502.7114
_Iv2_1851	-64.33836	294.6082	-0.22	0.827	-641.9095	513.2328
_Iv2_1852	-55.07556	294.6158	-0.19	0.852	-632.6616	522.5104
_Iv2_1853	-44.06957	294.6182	-0.15	0.881	-621.6604	533.5212
_Iv2_1854	-34.29104	294.6213	-0.12	0.907	-611.8878	543.3057
_Iv2_1855	-27.14247	294.6278	-0.09	0.927	-604.7521	550.4672
_Iv2_1856	-14.76211	294.633	-0.05	0.960	-592.3819	562.8577
_Iv2_1857	-3.908854	294.6396	-0.01	0.989	-581.5416	573.7239
_Iv2_1858	5.995196	294.647	0.02	0.984	-571.652	583.6424
_Iv2_1859	15.59062	294.6527	0.05	0.958	-562.0677	593.2489
_Iv2_1860	-119.469	268.0348	-0.45	0.656	-644.9437	406.0057
_Iv2_1861	-115.7792	268.0387	-0.43	0.666	-641.2617	409.7032
_Iv2_1862	-73.90396	262.6063	-0.28	0.778	-588.7362	440.9283
_Iv2_1863	-132.1926	268.1225	-0.49	0.622	-657.8392	393.454
_Iv2_1864	-129.4723	268.1259	-0.48	0.629	-655.1256	396.181
_Iv2_1865	79.99505	248.7892	0.32	0.748	-407.7492	567.7393
_Iv2_1866	78.2552	248.7831	0.31	0.753	-409.4771	565.9875
_Iv2_1867	75.95269	249.5615	0.30	0.761	-413.3057	565.211
_Iv2_1868	47.71354	248.2175	0.19	0.848	-438.91	534.337
_Iv2_1869	52.27472	248.2261	0.21	0.833	-434.3655	538.915
_Iv2_1870	50.76414	248.2284	0.20	0.838	-435.8806	537.4089
_Iv2_1871	63.57293	247.246	0.26	0.797	-421.146	548.2919
_Iv2_1872	58.33883	247.2524	0.24	0.813	-426.3926	543.0703
_Iv2_1873	57.08574	247.2581	0.23	0.817	-427.6568	541.8283
_Iv2_1874	52.42314	247.2684	0.21	0.832	-432.3397	537.186

_Iv2_1875	46.58909	247.2769	0.19	0.851	-438.1904	531.3685
_Iv2_1876	50.30211	247.2793	0.20	0.839	-434.482	535.0863
_Iv2_1877	53.44548	247.281	0.22	0.829	-431.3421	538.233
_Iv2_1878	58.14777	246.8737	0.24	0.814	-425.8411	542.1367
_Iv2_1879	60.72059	246.8735	0.25	0.806	-423.2679	544.7091
_Iv2_1880	60.96324	246.8777	0.25	0.805	-423.0336	544.9601
_Iv2_1881	54.47918	246.8807	0.22	0.825	-429.5235	538.4819
_Iv2_1882	120.5233	243.6352	0.49	0.621	-357.1167	598.1632
_Iv2_1883	116.9519	243.6477	0.48	0.631	-360.7126	594.6163
_Iv2_1884	117.4552	243.6576	0.48	0.630	-360.2288	595.1391
_Iv2_1885	118.2812	243.6641	0.49	0.627	-359.4154	595.9779
_Iv2_1886	134.1985	243.0422	0.55	0.581	-342.2789	610.6759
_Iv2_1887	133.8483	243.0501	0.55	0.582	-342.6446	610.3412
_Iv2_1888	133.5984	243.0589	0.55	0.583	-342.9119	610.1086
_Iv2_1889	135.4529	242.9935	0.56	0.577	-340.929	611.8348
_Iv2_1890	134.2302	243.0041	0.55	0.581	-342.1727	610.633
_Iv2_1891	135.1609	243.0125	0.56	0.578	-341.2583	611.5801
_Iv2_1892	133.7902	243.021	0.55	0.582	-342.6457	610.226
_Iv2_1893	134.3006	243.0302	0.55	0.581	-342.1533	610.7545
_Iv2_1894	132.3594	243.0394	0.54	0.586	-344.1125	608.8313
_Iv2_1895	133.6205	243.0482	0.55	0.583	-342.8688	610.1098
_Iv2_1896	135.5602	243.0535	0.56	0.577	-340.9395	612.0599
_Iv2_1897	139.8735	243.0602	0.58	0.565	-336.6393	616.3862
_Iv2_1898	148.6621	243.0655	0.61	0.541	-327.861	625.1853
_Iv2_1899	144.8076	243.0735	0.60	0.551	-331.7311	621.3464
_Iv2_1900	146.5521	243.0816	0.60	0.547	-330.0026	623.1069
_Iv2_1901	152.3022	243.0262	0.63	0.531	-324.1438	628.7482
_Iv2_1902	151.904	243.0339	0.63	0.532	-324.5572	628.3651
_Iv2_1903	157.344	242.967	0.65	0.517	-318.9861	633.6741
_Iv2_1904	157.613	242.9243	0.65	0.516	-318.6333	633.8593
_Iv2_1905	158.4004	242.9345	0.65	0.514	-317.8658	634.6666
_Iv2_1906	158.2652	242.9479	0.65	0.515	-318.0274	634.5579
_Iv2_1907	169.0869	242.8378	0.70	0.486	-306.9897	645.1636
_Iv2_1908	177.2497	242.8495	0.73	0.466	-298.8501	653.3495
_Iv2_1909	184.7033	242.8653	0.76	0.447	-291.4274	660.8339
_Iv2_1910	189.9074	242.8777	0.78	0.434	-286.2475	666.0623
_Iv2_1911	191.379	242.8929	0.79	0.431	-284.8059	667.5638
_Iv2_1912	191.4271	242.9081	0.79	0.431	-284.7874	667.6417
_Iv2_1913	197.4497	242.9075	0.81	0.416	-278.7635	673.663
_Iv2_1919	243.9048	242.9426	1.00	0.315	-232.3774	720.187
_Iv2_1920	179.8599	242.9599	0.74	0.459	-296.4561	656.176
_Iv2_1921	146.7884	242.9769	0.60	0.546	-329.561	623.1378
_Iv2_1922	133.2224	242.9451	0.55	0.583	-343.0646	609.5095
_Iv2_1923	120.2023	242.9519	0.49	0.621	-356.0982	596.5028
_Iv2_1924	123.8842	242.8121	0.51	0.610	-352.1421	599.9105
_Iv2_1925	116.5109	242.8215	0.48	0.631	-359.5339	592.5557
_Iv2_1926	111.7439	242.85	0.46	0.645	-364.3568	587.8446
_Iv2_1927	113.539	242.8615	0.47	0.640	-362.5842	589.6623
_Iv2_1928	108.6446	242.889	0.45	0.655	-367.5324	584.8216
_Iv2_1929	107.1459	242.9128	0.44	0.659	-369.0779	583.3697
_Iv2_1930	103.1109	242.978	0.42	0.671	-373.2407	579.4625
_Iv2_1931	101.4672	243.006	0.42	0.676	-374.9392	577.8736
_Iv2_1932	106.7174	242.9568	0.44	0.661	-369.5927	583.0274
_Iv2_1933	107.8447	242.9841	0.44	0.657	-368.5189	584.2084
_Iv2_1934	109.5116	243.0077	0.45	0.652	-366.8982	585.9213
_Iv2_1935	116.3512	243.0374	0.48	0.632	-360.1169	592.8193
_Iv2_1936	118.0992	243.0929	0.49	0.627	-358.4776	594.676
_Iv2_1937	120.2123	243.1124	0.49	0.621	-356.4029	596.8275
_Iv2_1938	127.829	243.1484	0.53	0.599	-348.8567	604.5147
_Iv2_1939	136.3224	243.1843	0.56	0.575	-340.4336	613.0783
_Iv2_1946	213.2124	243.4516	0.88	0.381	-264.0677	690.4925
_Iv2_1947	147.499	243.4782	0.61	0.545	-329.8332	624.8311
_Iv2_1948	128.6544	243.5303	0.53	0.597	-348.78	606.0888

```
   _Iv2_1949 |   135.9022   243.4572     0.56   0.577    -341.3888    613.1932
   _Iv2_1950 |   136.2776   243.5207     0.56   0.576    -341.1381    613.6932
   _Iv2_1951 |   163.5896   243.4656     0.67   0.502    -313.7179    640.8972
   _Iv2_1952 |   161.1812   243.5296     0.66   0.508    -316.2518    638.6142
   _Iv2_1953 |   156.4396   243.5878     0.64   0.521    -321.1075    633.9867
   _Iv2_1954 |   143.5888   243.6682     0.59   0.556     -334.116    621.2936
   _Iv2_1955 |   131.2964   243.6301     0.54   0.590    -346.3336    608.9264
   _Iv2_1956 |   115.5452   243.4662     0.47   0.635    -361.7635     592.854
   _Iv2_1957 |   105.6775   243.5525     0.43   0.664    -371.8004    583.1554
   _Iv2_1958 |   96.81203    243.639     0.40   0.691    -380.8353    574.4594
   _Iv2_1959 |   81.72898    243.706     0.34   0.737    -396.0499    559.5078
   _Iv2_1960 |     57.125   243.4462     0.23   0.814    -420.1445    534.3945
   _Iv2_1961 |   44.27294   243.4722     0.18   0.856    -433.0475    521.5934
   _Iv2_1962 |   34.69931   243.4746     0.14   0.887    -442.6258    512.0244
   _Iv2_1963 |   20.52712   243.5572     0.08   0.933    -456.9601    498.0143
   _Iv2_1964 |   5.671429   243.6451     0.02   0.981     -471.988    483.3308
   _Iv2_1965 |   -6.05434   243.7497    -0.02   0.980    -483.9189    471.8102
   _Iv2_1966 |  -15.69307   243.8276    -0.06   0.949    -493.7103    462.3241
   _Iv2_1967 |  -21.36904   243.8979    -0.09   0.930    -499.5241    456.7861
   _Iv2_1968 |  -30.24353   244.0054    -0.12   0.901    -508.6093    448.1222
   _Iv2_1969 |  -48.20492   244.1324    -0.20   0.843    -526.8197    430.4099
   _Iv2_1970 |  -72.38037   244.2656    -0.30   0.767    -551.2562    406.4955
   _Iv2_1971 |   -97.7743   244.4566    -0.40   0.689    -577.0245     381.476
   _Iv2_1972 |  -136.0872    244.659    -0.56   0.578    -615.7343    343.5599
   _Iv2_1973 |   -176.762   244.8916    -0.72   0.470    -656.8652    303.3412
       _cons |   -133.205   239.7758    -0.56   0.579    -603.2787    336.8687
-------------+----------------------------------------------------------------
     sigma_u |  295.44271
     sigma_e |  238.60108
         rho |  .60524399   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(123, 4670) =    55.28           Prob > F = 0.0000
```

It looks like on average, there is an increase in military size over time, so we could consider modeling it as a linear trend:

```
. xi: xtreg v41 v76 v5 v7 v19 v2, fe
Fixed-effects (within) regression              Number of obs      =      4945
Group variable (i): v1                         Number of groups   =       124
R-sq:  within  = 0.1837                        Obs per group: min =         1
       between = 0.0375                                       avg =      39.9
       overall = 0.0347                                       max =       148
                                               F(5,4816)          =    216.83
corr(u_i, Xb)  = -0.3282                       Prob > F           =    0.0000
------------------------------------------------------------------------------
         v41 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         v76 |   .0219487   .0130567     1.68   0.093    -.0036484    .0475457
          v5 |  -.0754687   .0075709    -9.97   0.000    -.0903111   -.0606263
          v7 |   .3923511   .0884183     4.44   0.000     .2190108    .5656913
         v19 |   .0075463   .0003463    21.79   0.000     .0068674    .0082252
          v2 |   1.185237   .2751054     4.31   0.000     .6459043    1.724569
       _cons |  -2168.702   515.4194    -4.21   0.000    -3179.159   -1158.245
-------------+----------------------------------------------------------------
     sigma_u |  293.26756
     sigma_e |  241.87032
         rho |  .59516697   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(123, 4816) =    53.09           Prob > F = 0.0000
```

To test whether it is appropriate to assume a linear trend, we test this model against the previous one in terms of its fit:

```
. estat ic
------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |   4945   -34593.44   -34091.44      6     68194.88    68233.92
------------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

. xi: qui xtreg v41 v76 v5 v7 v19 i.v2, fe
. estat ic
------------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+----------------------------------------------------------------
           . |   4945   -34593.44   -33948.03    152     68200.06    69188.99
------------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note
```

BIC difference: 69188.99-68233.92=955.07

The model with smaller BIC has better fit, and the strength of evidence in its favor is evaluated as follows:

BC Difference	Evidence
0-2	Weak
2-6	Positive
6-10	Strong
>10	Very strong

So in this case, the model with linear trend is clearly a better choice. We can also examine the trend graphically in the FE exploratory dataset we created. We could also follow with some linearity diagnostics, e.g. using mrunning, boxtid, etc.

```
.lowess v41 v2, mcolor(yellow)
```