Person-level Fixed Effects Model  
Vartanian: SW 541

We are examining all years of childhood available in the data. We will use a fixed effect model to determine the predicted income level during the child’s first 18 years of life.

```
. xtreg income povrate pabove age afdcinc kids if counting>1, fe

Fixed-effects (within) regression               Number of obs      =     60411
Group variable (i): pid                         Number of groups   =      5674
R-sq:  within  = 0.3583                         Obs per group: min =         2
between = 0.6764                                        avg =      10.6
overall = 0.5468                                        max =        18
F(5,54732)         =   6111.10
corr(u_i, Xb)  = 0.3202                         Prob > F           =    0.0000

------------------------------------------------------------------------------
income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
povrate |  -605.9632    13.5009   -44.88   0.000    -632.4251   -579.5013
pabove |  -78811.65   494.5185  -159.37   0.000    -79780.91   -77842.39
age |   733.2417   23.26633    31.52   0.000     687.6396    778.8439
afdcinc |   .0812813   .0339383     2.39   0.017      .014762    .1478006
kids |  -226.2886   89.22036    -2.54   0.011    -401.1611   -51.41603
_cons |   91155.66   635.0405   143.54   0.000     89910.97    92400.34
-------------+----------------------------------------------------------------
sigma_u |  21161.351
sigma_e |  20864.218
rho |   .50706993   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5673, 54732) =    10.42         Prob > F = 0.0000
```

If this analysis, we have 5,674 people, with 60,411 years of data. On average, each child has 10.6 years worth of data. Each child has at least 2 years worth of data.

This model factors out permanent factors in the child’s life over these childhood years. After factoring out these factors, we then look at income level in any year and subtract off the mean level of income over all of the childhood years. We then examine the neighborhood poverty rate in any year, and subtract off the mean level of neighborhood poverty for that child over the childhood years. This new variable, where the mean of the poverty rate is subtracted from the actual poverty rate for any year, is the independent variable. In this way, we can determine changes in the neighborhood poverty rate affect changes in income from their respective means. The same type of analysis is done with all of the other independent variables.

In essence, we are running this model with n-1 dummy variables, or in this case, 5774-1 dummy variables for the person.
We could also run a random effects model that does not require us to use these 5,773 dummy variables.

```
.xtreg income povrate pabove age afdcinc kids if counting>1, re
```

```
Number of obs = 60411
Number of groups = 5674
R-sq: within = 0.3567
between = 0.6839
overall = 0.5547
Random effects u_i ~ Gaussian
Wald chi2(5) = 41795.34
corr(u_i, X) = 0 (assumed)
```

```
------------------------------------------------------------------------------
income |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
povrate |  -772.3516   11.04467   -69.93   0.000    -793.9988   -750.7045
pabove |  -84226.16   454.6986  -185.24   0.000    -85117.36   -83334.97
age |    587.607   21.78837    26.97   0.000     544.9026    630.3114
afdcinc |  -.0130362   .0304859    -0.43   0.669    -.0727876    .0467151
kids |  -543.3005   76.45204    -7.11   0.000    -693.1437   -393.4572
_cons |   98629.63   594.4009   165.93   0.000     97464.63    99794.64
-------------+----------------------------------------------------------------
sigma_u |  17800.138
sigma_e |  20864.218
rho |  .42124641   (fraction of variance due to u_i)
------------------------------------------------------------------------------

From this, we can see that the models are not tremendously different from one another. To test differences between the models, use a Hausman test.

```
.hausman fixed random
```

```
---- Coefficients ----
|      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
|     fixed        random       Difference          S.E.
-------------+----------------------------------------------------------------
povrate |   -605.9632    -772.3516        166.3884        7.764624
pabove |   -78811.65    -84226.16         5414.51        194.4164
age |    733.2417      587.607        145.6347        8.160234
afdcinc |    .0812813   -.0130362        .0943175        .0149136
kids |   -226.2886    -543.3005        317.0119        45.99302
_cons |   98629.63   594.4009009    165.93   0.000     97464.63    99794.64
-------------+----------------------------------------------------------------
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

\[ \text{chi2}(4) = (b-B)'[(V_b-V_B)^{-1}](b-B) \]
\[ = 903.54 \]
\[ \text{Prob>chi2} = 0.0000 \]
```

.end of do-file
These types of models, fixed, between, and random effects, are used with cross-sectional time-series data. We have a number of observations from a person over a period of time, or we have information about different members of families over a period of time. What we’re doing with the fixed effect model is examining changes within the person (or family). If we had run a between effects model, this would indicate the effects of the variables between persons (or families). Remember that this will be different than an OLS regression model because we have a number of observations from a single person or family. A between model examines the effect of the variable between persons or families. A random effects model assumes that the two effects will be the same. If this is the case, then running a random effects model is more efficient because it uses far fewer degrees of freedom. However, we may gain information from using a fixed effects model that we do not get by examining differences between subjects (persons or families). That is, we are able to control for unobserved, permanent factors within the person or family by using a fixed effect model. For example, if a person has some condition that is not controlled in the model because there is no information about this condition, the fixed effect model is able to ‘difference it out’, or because we are examining the same individual over many years, and if that condition stays the same over all those years, the effects of that condition is controlled. Again, these unobserved, permanent effects get ‘factored out’ of the analysis, or controlled. If we had used an observed variable such as level of education (say for the head of household) that did not change over time, this would also get factored out of the model. We couldn’t use such a variable because there is no variation in the variable over time. We need variation in variables over time in

1 What is a cross section time series data set? Panel data, also called longitudinal data or cross-sectional time series data, are data where multiple cases (people, firms, countries etc) were observed at two or more time periods. An example is the National Longitudinal Survey of Youth, where a nationally representative sample of young people were each surveyed repeatedly over multiple years.

There are two kinds of information in cross-sectional time-series data: 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 you to take advantage of these different types of information.

While it is possible to use ordinary multiple regression techniques on panel data, they may not be optimal. The estimates of coefficients derived from regression may be subject to omitted variable bias - a problem that arises when there is some unknown variable or variables that cannot be controlled for that affect the dependent variable. With panel data, it is possible to 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 between 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 between cases. Source for this footnote: http://dss.princeton.edu/online_help/analysis/panel.htm
order to use it within our analysis. Otherwise, the variable is controlled, even if it is not explicitly in the model, but we cannot determine the effects of such variables.

From this analysis, we find that the coefficients from the random effects model are biased. We have failed to control for crucial, unobserved, permanent variables that a fixed effect model allows us to control.

### Between regression (regression on group means)

- **Number of obs**: 60411
- **Number of groups**: 5674
- **R-sq**: within = 0.3352, between = 0.6898, overall = 0.5522
- **F(5, 5668)** = 2520.28
- **sd(u_i + avg(e_i.))** = 19238

### Coefficients

|        | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|---------------------|
| povrate | -981.1416 | 20.03724  | -48.97 | 0.000 | -1020.422, -941.8609 |
| pabove  | -106099.2 | 1241.098  | -85.54 | 0.000 | -108532.2, -103666.2 |
| age     | -237.208  | 98.62704  | -2.41  | 0.016 | -430.5548, -43.86132 |
| afdcinc | 0.200111  | 0.070019  | 2.86   | 0.004 | 0.062847, 0.337351   |
| kids    | -285.4378 | 157.8479  | -1.81  | 0.071 | -594.88, 24.00444   |
| _cons   | 121018.4  | 1359.504  | 89.02  | 0.000 | 118353.2, 123683.5   |

---

### .reg income povrate pabove age afdcinc kids if counting>1

- **Number of obs**: 60411
- **F(5, 60405)** = 15191.47
- **R-squared**: 0.5570
- **Adj R-squared**: 0.5570
- **Root MSE**: 28640

|        | Coef.  | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|---------------------|
| povrate | -990.4916 | 8.362997  | -118.44 | 0.000 | -1006.883, -974.1001 |
| pabove  | -98815.32 | 435.3202  | -226.99 | 0.000 | -99668.55, -97962.09 |
| age     | 334.4932  | 24.92703  | 13.42  | 0.000 | 285.6361, 383.3503   |
| afdcinc | 0.0709422 | 0.0276484 | 2.57   | 0.010 | 0.0167512, 0.1251332 |
| kids    | -586.8704 | 63.40039  | -9.26  | 0.000 | -711.1354, -462.6055 |
| _cons   | 112758.7  | 444.3516  | 253.76 | 0.000 | 111887.8, 113629.7   |