Panel Data

Po-Chun Huang

National Chengchi University

May 13, 2021

Today

- Panel Data (Wooldridge Ch13 and Ch14)
  - Why use panel data?
  - First Difference Estimation
  - Fixed Effects Estimation
  - STATA

Panel Data

- Recall that panel data are repeated observations of a given unit of observation over time
- Examples:
  - Data on 22 cities/counties in Taiwan, each state is observed in 3 years, for a total of 66 observations
  - Data on 1,000 individuals in four different years, for 4,000 observations total
  - Data on 10,000 firms in ten different years, for 100,000 observations total

Panel Data

- We can use use panel data to test whether various laws—like increasing beer taxes and raising the minimum legal drinking age—affect traffic fatalities

$$vfrall_{it} = \beta_0 + \beta_1 beertax_{it} + \beta_2 mlda_{it} + ... + u_{it}$$

- Another example using state-level panel would be an examination of the effect of a change in the minimum wage on employment

$$emp_{it} = \beta_0 + \beta_1 minwage_{it} + ... + u_{it}$$

- Other panel studies use individuals, employers, cities, or industries, as the unit of observations

Why Panel Data are Useful?

- With panel data we can control for factors that
  - vary across entities (states) but do not vary over time
  - could cause omitted variable bias if they are omitted
  - are unobserved or unmeasured—and therefore cannot be included in the regression using multiple regression
- Here's the key idea
  - An individual person, state, or any entity can **serve as its own "control"** if the unobserved characteristics associated with each observation do not change over time

Notations for Panel Data

- A double subscript distinguishes entities (states) and time periods (years)
- $i =$ entity (state), $n =$ number of entities, so $i = 1, ..., n$
- $t =$ time period (year), $T =$ number of time periods so $t = 1, ..., T$

Some Panel Jargon

- Another term for panel data is longitudinal data
- A "balanced panel" has no missing observations
- An "unbalanced panel" has some entities (states) that are not observed for some time periods (years)

Example: Ruhm's study of "Alcohol Policies and Highway Vehicle Fatalities"

- Christopher Ruhm (1996) examined the impact of several policies that were intended to reduce alcohol-related traffic deaths
  - increasing the tax on beer
  - raising the minimum legal drinking age (MLDA)
  - automatically suspending the driver's' license of a drunk driver
  - automatically suspending the license of a driver who refuses to take a blood alcohol test
  - permitting people who are injured by a drunk driver to sue whoever served alcohol to the driver

Example: Ruhm's study of "Alcohol Policies and Highway Vehicle Fatalities"

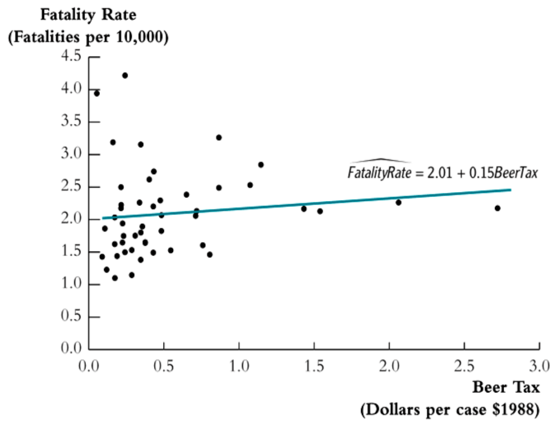
- Unit of observation: a US state in a given year
- 48 U.S. states, so $n =$ number of entities $= 48$
- 7 years (1982, ... , 1988), so $T =$ number of time periods $= 7$
- Balanced panel, so total number of observations $= 7 \cdot 48 = 336$

Example: Ruhm's study of "Alcohol Policies and Highway Vehicle Fatalities"

- Variables
  - Traffic fatality rate (number of traffic deaths in that state in that year, per 10,000 residents)
  - Tax on a case of beer
  - Legal driving age
  - Drunk driving laws, etc.

FIGURE 8.1 **The Traffic Fatality Rate and the Tax on Beer**

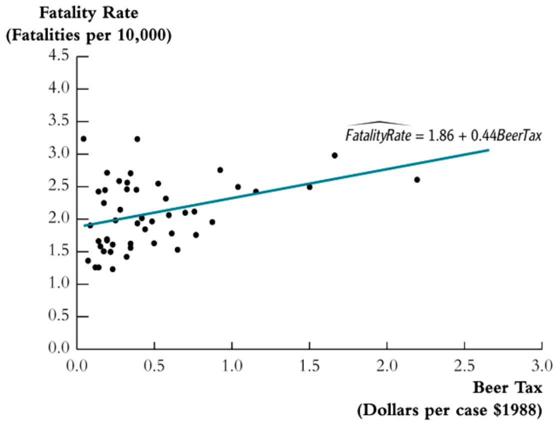
Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(a) 1982 data

FIGURE 8.1 The Traffic Fatality Rate and the Tax on Beer

Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(b) 1988 data

In both cases: higher alcohol taxes, more traffic deaths?

- Why might there be higher more traffic deaths in states that have higher alcohol taxes?
- Lots of other factors determine traffic fatality rates and correlate to beer taxes
  - "culture" around drinking and driving
  - density of cars on the road

Cultural attitudes towards drinking and driving

- Arguably are a determinant of traffic deaths AND
- Are potentially are correlated with the beer tax
- So beer taxes could be picking up cultural differences (omitted variable bias)
- Then the two conditions for omitted variable bias are satisfied
    - Specifically, ”high taxes” could reflect ”cultural attitudes towards drinking”

Traffic density

- Suppose
  - High traffic density means more traffic deaths
  - Western states with lower traffic density have lower alcohol taxes
- Then the two conditions for omitted variable bias are satisfied
  - Specifically, "high taxes" could reflect "high traffic density"
- Panel data lets us eliminate omitted variable bias when the omitted variables are constant over time within a given state

Panel Data Methods

Wooldridge Ch 13.3-13.5

- Start with a simple two-period model
  - For example, 1982 and 1988
- Then write the following model:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D2_t + a_i + u_{it}$$

- This could be a model of the effect of the beer tax (X) on the vehicle fatality rate (Y)
- Two things are new in this equation
- First, $D2_t$ is a dummy variable

$= 0$ if the observation comes from the first period $(t = 1)$

$= 1$ if it comes from the second period $(t = 2)$

Panel Data Methods

Wooldridge Ch 13.3-13.5

- Second, a_i denotes an unobserved effect or a fixed effect — it is meant to capture anything about observation i that is fixed over time and is otherwise unobserved
- a_i denotes an unobserved effect or a fixed effect
- You could think of a_i as representing a whole set of dummy variables, one for each cross-sectional observation in the data
  - In fact, one way to handle panel data is to include a dummy for each cross-sectional observation—for example, you could think of including one dummy variable for Alabama, one for Arkansas, and so on
  - We will see this later—it is called the "fixed-effects" estimator

The Problem

- If you simply pool your two years of panel data and estimate by OLS, you end up estimating this:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D2_t + v_{it}$$

- Compare this with the model we are considering:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D2_t + a_i + u_{it}$$

- and you can see that

$$v_{it} = a_i + u_{it}$$

- So if a_i is correlated with X_{it} you have a problem
- Why? Because then v_{it} will be correlated with $X_i t$ and OLS will be biased and inconsistent

Estimation

- Three ways of estimating fixed effects models:

1. First Difference Estimator
2. Demeaning (Fixed Effects Estimator)
3. Dummy Variable Estimator

First Difference (FD) Estimator

- Again, the model is:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D2_t + a_i + u_{it}$$

- Write down separate a equation for year 2 and for year 1:

$$Y_{i2} = (\beta_0 + \delta) + \beta_1 X_{i2} + a_i + u_{i2}$$

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

- Now subtract the second equation from the first:

$$Y_{i2} - Y_{i1} = \delta + \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$$

- You can rewrite this using deltas:

$$\Delta Y_i = \delta + \beta_1 \Delta X_i + \Delta u_i$$

- This is a "first-differenced" equation, and is easy to estimate

Note three things

- First, β_1 in the FD equation has exactly the same interpretation as β_1 in the first equation

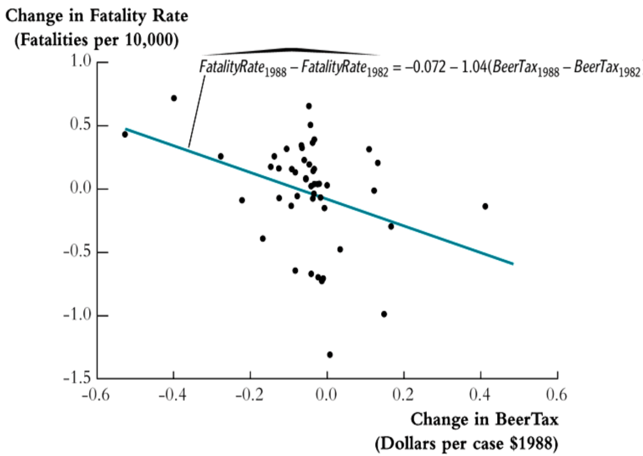
$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta D2_t + a_i + u_{it}$$

$$\Delta Y_i = \delta + \beta_1 \Delta X_i + \Delta u_i$$

- This seems a little odd, but review the derivation, and you will see that it makes sense
- By taking first differences, you eliminate the unobserved fixed effects that might be correlated with the error term

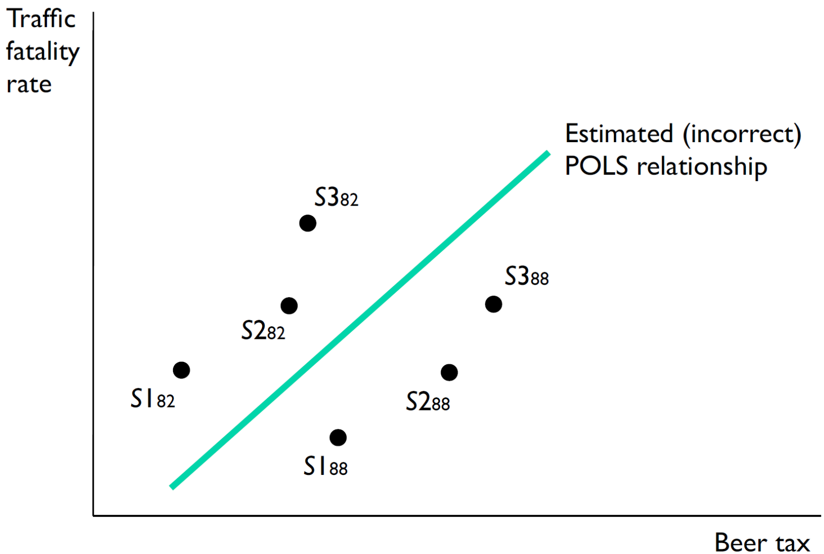
Scatterplot of $\Delta beertax_i$ and $\Delta vfrall_i$

FIGURE 8.2 Changes in Fatality Rates and Beer Taxes, 1982–1988

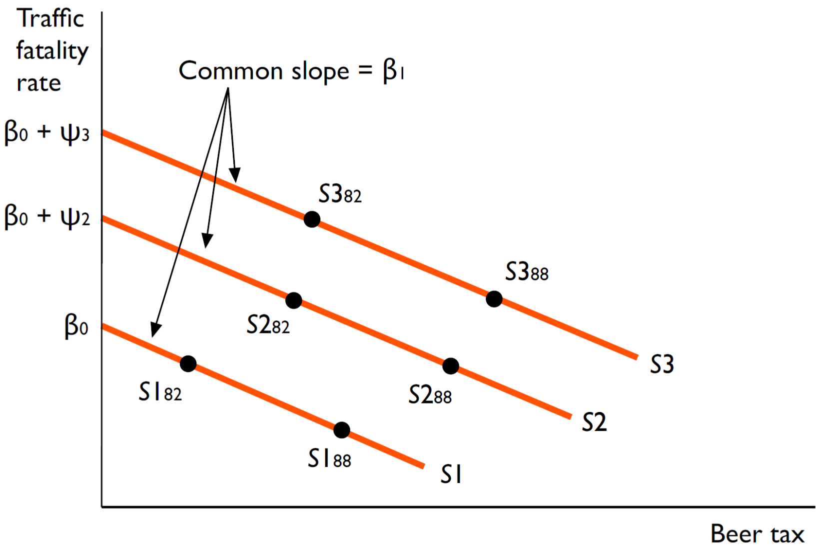


This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.

Here is the underlying story: first Pooled OLS

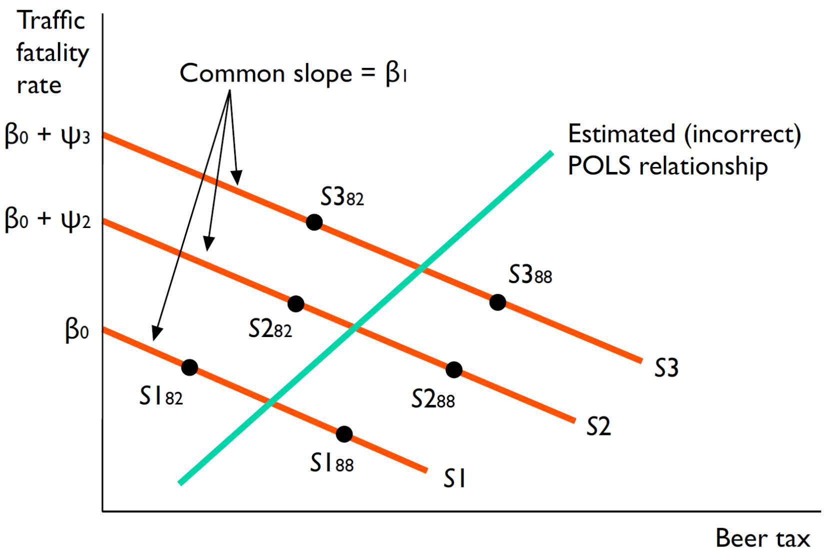


Now FD (or FE)



Traffic fatality rate
Common slope = β1
β0 + ψ3
β0 + ψ2
β0
S382
S388
S282
S288
S182
S188
S3
S2
S1
Beer tax

Finally, POLS and FD together



Demeaning (Fixed Effects Estimator)

- We have discussed the first–difference estimator. Next, I will introduce the "fixed effects" estimator
  - You can think of the FE estimator in either of two ways:
  - adding a dummy variable for each cross-sectional observation to the model—the least squares dummy variable (LSDV) approach
  - "time de-meaning" the data—the fixed-effects transformation approach
- LSDV gives us exactly the same estimates that we would obtain from the regression on time-demeaned data, and the standard errors are identical

Least squares dummy variable approach (LSDV)

- To estimate fixed effects using the LSDV approach, start with the following model:

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

- For example, this could be a model of how the traffic fatality rate (Y) is affected by the beer tax (X) in a state
- Remember that as denotes an unobserved effect or "fixed effect" that captures anything about state *i* that is fixed over time and is otherwise unobserved (aspects of a state's road that lead to traffic fatalities)

Least squares dummy variable approach (LSDV)

- A simple way to capture a fixed effect in a model is to add a dummy variable for each state (or individual, firm, etc.)
- Again, if we are estimating a model using panel of states

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

then we could add a dummy variable for each state to capture the state fixed effects we are estimating a model using panel of states

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \psi_2 S2_i + \psi_3 S3_i + \psi_4 S4_i + \ldots + \psi_4 8 S48_i + u_{it}$$

- In this case, we are assuming we have 48 states in the sample, so we can include only 47 dummies

The fixed-effects transformation

- Suppose you take the average of Y and X over time for each observation in the sample (state, individual, or whatever):

  where

$$\bar{Y}_i = \frac{\sum_{t=1}^{T} Y_{it}}{T}$$

$$\bar{X}_i = \frac{\sum_{t=1}^{T} X_{it}}{T}$$

- This is just the average of the dependent variable and independent variable for observation i

The fixed-effects transformation

- Now subtract this new equation from the original one

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

- You can rewrite the last equation as:

$$\ddot{Y}_i = \delta + \beta_1 \ddot{X}_i + \ddot{u}_i$$

- This is the "fixed-effects" estimator

Notice two things

- First, β_1 in the FE equation has exactly the same interpretation as the β_1 in the first equation (we saw something just like this with the FD estimator)

- Second, by using the FE estimator, you eliminate the unobserved fixed effects that might be correlated with the error term (again, just as with the FD estimator)

Random Effects (RE)

- FE assumes $Cov(x_{it}, a_i) \neq 0$: individual-specific effect is a random variable that is allowed to be correlated with the explanatory variables
- RE assumes $Cov(x_{it}, a_i) = 0$: individual-specific effect is a random variable that is assumed to be uncorrelated with the explanatory variables (of all past, current and future time periods)
- **Do not use it for observational data if you want to make a causal claim!**

Notes on Identification

- To consistently estimate β_1 using FE, FD or RE, the key assumption is strict exogeneity (holding a_i constant):

$$E(u_{it}|x_{i1}, x_{i2}, ..., x_{iT}, a_i) = 0, t = 1, ..., T \qquad (1)$$

- It means that u_{it} is uncorrelated to all periods of x once a_i is controlled for. Alternatively, only x_{it} affects the expected value of y_{it} once a_i is controlled for.
- This assumption will be violated if
  - there is feedback problem (x_{it} changes due to changes in u_{it-p}, $p > 0$)
  - lagged effects of x are not correctly specified

Estimation in STATA

1. First Difference Estimator:

   reg D.(Y X)

2. Dummy Variable Estimator:

   xi: reg Y X i.id

3. Fixed Effects Estimator:

   xtreg Y X, fe

Notes on FD and FE in STATA

- The key to doing FD and FE in Stata is to first "**tsset**" your data
- Doing FD requires you to generate first differences, which are simply changes in a given variable between two points in time (within a given unit of observation, like a state or individual)— this is much easier if you **tsset** your data (see the example above)
- Also, you may want to estimate a dynamic model (a model with lags) using either panel data or a time series—this is easy if you **tsset** your data

Note on heteroskedasticity and serial correlation in panel data

- To obtain standard errors that are robust to heteroskedasticity, specify the "robust" option in your regression command
- To obtain standard errors that are robust to serial correlation and heteroskedasticity, add the option "**cluster(panelvar)**" in your regression command, where "panelvar" is the variable identifying the cross-sectional unit of observation
- Remember that robust or cluster command does not change your estimates!

STATA

- Suppose we wish to evaluate the effect of a job training program on worker productivity of manufacturing firms
- We have data from 54 firms for two years (1987 and 1988)
- Consider the following model

$$scrap_{it} = \beta_0 + \delta y88_t + \beta_1 grant_{it} + a_i + u_{it}$$

$scrap_{it}$: the scrap rate of firm i during year t (the number of items, per 100, that must be scrapped due to defects). Let

$grant_{it}$: a binary indicator equal to one if firm i in year t received a job training grant.

a_i: the unobserved firm effect (the firm fixed effect)

STATA

- The unobserved effect contains such factors as average employee ability, capital, and managerial skill; these are roughly constant over a two-year period
- For example, administrators of the program might give priority to firms whose workers have lower skills (upward bias).
- Or, to make the job training program appear effective, administrators may give the grants to employers with more productive workers (downward bias).

STATA

OLS

```
. reg lscrap grant d88

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(2, 105)       =      0.18
       Model |  .810536068         2  .405268034   Prob > F        =    0.8378
    Residual |  240.098947       105  2.28665664   R-squared       =    0.0034
-------------+----------------------------------   Adj R-squared   =   -0.0156
       Total |  240.909484       107   2.2514905   Root MSE        =    1.5122

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |   .0566004     .43091     0.13   0.896    -.7978145    .9110152
         d88 |  -.1889081   .3281441    -0.58   0.566    -.8395572     .461741
       _cons |   .5974341   .2057802     2.90   0.005     .1894099    1.005458
------------------------------------------------------------------------------
```

STATA

tsset

```
tsset fcode year
     panel variable:  fcode (strongly balanced)
      time variable:  year, 1987 to 1988
              delta:  1 unit
```

STATA

FD

```
. reg D.(lscrap grant)

      Source |       SS           df       MS      Number of obs   =        54
-------------+----------------------------------   F(1, 52)        =      3.74
       Model |  1.23795567         1  1.23795567   Prob > F        =    0.0585
    Residual |  17.1971851        52  .330715099   R-squared       =    0.0672
-------------+----------------------------------   Adj R-squared   =    0.0492
       Total |  18.4351408        53  .347832845   Root MSE        =    .57508

------------------------------------------------------------------------------
    D.lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |
         D1. |  -.3170579   .1638751    -1.93   0.058    -.6458974    .0117816
             |
       _cons |  -.0574357    .097206    -0.59   0.557    -.2524938    .1376224
------------------------------------------------------------------------------
```

STATA

FE

```
. xtreg lscrap grant d88, fe

Fixed-effects (within) regression               Number of obs      =       108
Group variable: fcode                           Number of groups   =        54

R-sq:                                           Obs per group:
     within  = 0.1392                                          min =         2
     between = 0.0049                                          avg =       2.0
     overall = 0.0006                                          max =         2

                                                F(2,52)            =      4.20
corr(u_i, Xb)  = -0.0674                        Prob > F           =    0.0203

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.3170579   .1638751    -1.93   0.058    -.6458975    .0117816
         d88 |  -.0574357    .097206    -0.59   0.557    -.2524938    .1376224
       _cons |   .5974341   .0553369    10.80   0.000     .4863924    .7084757
-------------+----------------------------------------------------------------
     sigma_u |  1.4833025
     sigma_e |   .4066418
         rho |  .93009745   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(53, 52) = 26.42                     Prob > F = 0.0000
```

STATA
LSDV

```
. reg lscrap grant d88 i.fcode

      Source |       SS           df       MS      Number of obs   =       108
-------------+----------------------------------   F(55, 52)       =     25.54
       Model |  232.310891        55  4.22383438   Prob > F        =    0.0000
    Residual |  8.59859282        52  .165357554   R-squared       =    0.9643
-------------+----------------------------------   Adj R-squared   =    0.9266
       Total |  240.909484       107   2.2514905   Root MSE        =    .40664

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.3170579   .1638751    -1.93   0.058    -.6458975    .0117816
         d88 |  -.0574357    .097206    -0.59   0.557    -.2524938    .1376224
             |
       fcode |
     410538  |    3.89394   .4066418     9.58   0.000     3.077954    4.709926
     410563  |   4.773406   .4066418    11.74   0.000      3.95742    5.589393
     410565  |    5.62578   .4066418    13.83   0.000     4.809794    6.441767
     410566  |   4.682246   .4066418    11.51   0.000     3.866259    5.498232
```

STATA

Cluster SE

```
. reg lscrap grant d88 i.fcode, cl(fcode)

Linear regression                               Number of obs     =        108
                                                F(1, 53)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.9643
                                                Root MSE          =     .40664

                                 (Std. Err. adjusted for 54 clusters in fcode)
------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.3170579   .2435688    -1.30   0.199    -.8055951    .1714792
         d88 |  -.0574357   .1301601    -0.44   0.661    -.3185038    .2036324
             |
       fcode |
     410538  |    3.89394   2.86e-14  1.4e+14   0.000      3.89394     3.89394
     410563  |   4.773406   2.86e-14  1.7e+14   0.000     4.773406    4.773406
     410565  |    5.62578   2.86e-14  2.0e+14   0.000      5.62578     5.62578
     410566  |   4.682246   2.86e-14  1.6e+14   0.000     4.682246    4.682246
     410567  |   2.557998   2.86e-14  8.9e+13   0.000     2.557998    2.557998
     410577  |   3.389506   2.86e-14  1.2e+14   0.000     3.389506    3.389506
     410592  |   6.194197   2.86e-14  2.2e+14   0.000     6.194197    6.194197
```

STATA

Use data from 1987, 1988, and 1989

```
. xtreg lscrap grant grant_1 d88 d89, fe cl(fcode)

Fixed-effects (within) regression               Number of obs      =       162
Group variable: fcode                           Number of groups   =        54

R-sq:                                           Obs per group:
     within  = 0.2010                                         min =         3
     between = 0.0079                                         avg =       3.0
     overall = 0.0068                                         max =         3

                                                F(4,53)            =      7.07
corr(u_i, Xb)  = -0.0714                        Prob > F           =    0.0001

                                  (Std. Err. adjusted for 54 clusters in fcode)
------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.2523149   .1434399    -1.76   0.084    -.5400188     .035389
     grant_1 |  -.4215895   .2824604    -1.49   0.141    -.9881333    .1449543
         d88 |  -.0802157   .0978408    -0.82   0.416    -.2764594    .1160281
         d89 |  -.2472028   .1967819    -1.26   0.215    -.6418973    .1474917
       _cons |   .5974341   .0638746     9.35   0.000     .4693177    .7255504
-------------+----------------------------------------------------------------
     sigma_u |   1.438982
     sigma_e |  .49774421
         rho |  .89313867   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

STATA

Use data from 1987, 1988, and 1989

```
. reg lscrap grant grant_1 d88 d89 i.fcode, cl(fcode)

Linear regression                               Number of obs     =        162
                                                F(3, 53)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.9276
                                                Root MSE          =     .49774

                                  (Std. Err. adjusted for 54 clusters in fcode)
------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.2523149   .1762394    -1.43   0.158    -.6058063    .1011766
     grant_1 |  -.4215895   .3470489    -1.21   0.230    -1.117682    .2745025
         d88 |  -.0802157   .1202135    -0.67   0.507    -.3213333     .160902
         d89 |  -.2472028   .2417789    -1.02   0.311    -.7321498    .2377442
             |
       fcode |
     410538  |   3.905259   1.83e-14  2.1e+14   0.000     3.905259    3.905259
     410563  |   4.717328   1.83e-14  2.6e+14   0.000     4.717328    4.717328
     410565  |   4.443668   1.83e-14  2.4e+14   0.000     4.443668    4.443668
     410566  |   4.621434   1.83e-14  2.5e+14   0.000     4.621434    4.621434
     410567  |   2.279588   1.83e-14  1.2e+14   0.000     2.279588    2.279588
     410577  |   3.423147   1.83e-14  1.9e+14   0.000     3.423147    3.423147
     410592  |    6.12662   1.83e-14  3.3e+14   0.000      6.12662     6.12662
     410593  |   2.934958   1.83e-14  1.6e+14   0.000     2.934958    2.934958
```

STATA

Use data from 1987, 1988, and 1989

```
. reg D.(lscrap grant grant_1 d88), cl(fcode)

Linear regression                               Number of obs     =        108
                                                F(3, 53)          =       1.98
                                                Prob > F          =     0.1284
                                                R-squared         =     0.0365
                                                Root MSE          =     .57672

                                  (Std. Err. adjusted for 54 clusters in fcode)
-------------------------------------------------------------------------------
             |               Robust
    D.lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
       grant |
         D1. |   -.222781   .1316461    -1.69   0.096    -.4868297    .0412676
             |
     grant_1 |
         D1. |  -.3512459   .2709732    -1.30   0.201    -.8947493    .1922575
             |
         d88 |
         D1. |   .0481041   .0568246     0.85   0.401    -.0658717    .1620798
             |
       _cons |  -.1387113   .0953842    -1.45   0.152    -.3300278    .0526053
-------------------------------------------------------------------------------
```

STATA

- FD, FE, LSDV are equivalent if there are only two periods of data
- FE and LSDV are equivalent regardless of time periods.
- FD and FE do not generate the same results in general.
- Most importantly, FD, FE, and LSDV are more convincing than OLS because they remove unobserved heterogeneity.