Warwick Economics Summer School
Topics in Microeconometrics
Panel Data Methods

Michele Aquaro

University of Warwick

This version: July 22, 2016

Reading material

- Textbook: *Introductory Econometrics: A Modern Approach*, 5e, by J.M. Wooldridge, Chapters 13 and 14.
- Topics:
    1. What is a Panel Data Set?
    2. Formats for Storing Panel Data
    3. Analysis with Two-Periods of Panel Data
    4. Differencing with More than Two Years of Data
    5. Fixed Effects Estimation
    6. Random Effects Estimation
    7. Choosing Among POLS, FD, FE, and RE
- When the symbol  appears on the top-right of these slides, it means that the material presented is slightly more advanced.

Section 1

What is a Panel Data Set?

What is a Panel Data Set?

- With a panel data set, the same units (individuals, schools, cities, and so on) are sampled in two or more time periods. For each unit i we have multiple periods (e.g. years) of data.
- By contrast, with pooled cross sections we have different units sampled in each period. If there is some overlap, we ignore it (and usually would not know there is overlap).
- Main benefit of panel data: with multiple years of data we can control for unobserved characteristics that do not change (or change slowly) over time. Very useful for policy analysis.

What is a Panel Data Set?

• Notation here assumes a balanced panel. A *balanced panel* is one where we observe the same time periods for each unit. Easier to achieve for larger units (such as schools and cities). At a disaggregated level – such as individuals and families – following the same units over time can be challenging. (Attrition can be a serious problem.) Econometric methods extend to unbalanced panels, and software takes care of algebraic details.

• The notation we use is the following. For each cross-sectional unit i at time t the response variable is y_{it}. An explanatory variable is x_{it}. With more than one explanatory variable we have x_{it1}, x_{it2}, ..., x_{itk}.

What is a Panel Data Set?

- We will start with the case of two periods, so $t = 1$, 2 and (hopefully) many cross section observations, $i = 1, 2, ..., n$.
- For each i and t, we observe $(x_{it1}, x_{it2}, ..., x_{itk}, y_{it})$.
- Unobserved factors are put into two categories.

1. A component that does not change over time, a_i. Called an *unobserved effect* or *unobserved heterogeneity*. It varies by individual but not by time. At the individual level, can think of a_i as "ability" – something innate and not subject to change. Generally, a_i contains unobserved "attributes."
2. There are also unobservables that change across time, u_{it}. These are sometimes called "shocks"; we will call them *idiosyncratic errors*. They are specific to unit i but vary over time, and they affect the outcome, y_{it}.

Section 2

Formats for Storing Panel Data

Formats for Storing Panel Data

- A very common way to store panel data is to stack the time periods for each i on top of each other. In particular, the time periods for each unit should be adjacent, and stored in chronological order (from earliest period to the most recent). This is sometimes called the "long" storage format.

Formats for Storing Panel Data

```
. use gpa3

. describe id term trmgpa sat season

               storage  display     value
variable name   type    format      label      variable label
-------------------------------------------------------------------------------------
id              float   %9.0g                  student identifier
term            int     %4.0f                  fall = 1, spring = 2
trmgpa          float   %9.0g                  term GPA
sat             int     %4.0f                  SAT score
season          byte    %8.0g                  =1 if in season

. list id term trmgpa sat season in 1/10

     +--------------------------------------+
     |  id   term   trmgpa    sat   season |
     |--------------------------------------|
  1. |  22      1      1.5    920        0 |
  2. |  22      2     2.25    920        1 |
  3. |  35      1      2.2    780        0 |
  4. |  35      2      1.6    780        1 |
  5. |  36      1      1.6    810        0 |
     |--------------------------------------|
  6. |  36      2     1.29    810        1 |
  7. | 156      1        2   1080        1 |
  8. | 156      2     2.73   1080        0 |
  9. | 246      1      2.8    960        1 |
 10. | 246      2      2.6    960        0 |
     +--------------------------------------+
```

Formats for Storing Panel Data

• While not absolutely necessary for some procedures, it is best to tell Stata that you have a panel data set. In particular, what are i and t? In `GPA3.DTA`, $i = id$ and $t = term$.

```
. xtset id term
       panel variable:  id (strongly balanced)
        time variable:  term, 1 to 2
                delta:  1 unit

. tab term

  fall = 1, |
 spring = 2 |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        366       50.00       50.00
          2 |        366       50.00      100.00
------------+-----------------------------------
      Total |        732      100.00
```

- The same data structure is convenient for more than two years.

```
. use mathpnl

. xtset distid year
       panel variable:  distid (strongly balanced)
        time variable:  year, 1992 to 1998
                delta:  1 unit

. des distid year math4 expp found lunch

              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------
distid          float  %9.0g                  district identifier
year            int    %9.0g                  1992-1998
math4           float  %9.0g                  % satisfactory, 4th grade math
expp            int    %9.0g                  expenditure per pupil
found           int    %9.0g                  foundation grant, $:  1995-98
lunch           float  %9.0g                  % eligible for free lunch
```

```
. list distid year math4 expp found lunch in 1/21

     +---------------------------------------------+
     | distid   year   math4    expp   found   lunch |
     |---------------------------------------------|
  1. |   1010   1992    28.8    4227       .    36.3 |
  2. |   1010   1993    32.3    4809       .    39.2 |
  3. |   1010   1994    39.1    5214       .    38.6 |
  4. |   1010   1995      68    6019    5245   37.41 |
  5. |   1010   1996    68.4    6155    5398   40.79 |
     |---------------------------------------------|
  6. |   1010   1997      49    6134    5553   43.84 |
  7. |   1010   1998      75    6476    5707   42.61 |
  8. |   2010   1992      30    3445       .    36.4 |
  9. |   2010   1993    28.6    3446       .    38.6 |
 10. |   2010   1994    37.5    3583       .    52.6 |
     |---------------------------------------------|
 11. |   2010   1995      40    8525    5581   51.52 |
 12. |   2010   1996    83.3    8632    5734    47.3 |
 13. |   2010   1997      90    8854    5889   52.24 |
 14. |   2010   1998      67    9424    6043    56.6 |
 15. |   2020   1992      20    6379       .    59.3 |
     |---------------------------------------------|
 16. |   2020   1993    37.5    7600       .    59.2 |
 17. |   2020   1994     100    7412       .      50 |
 18. |   2020   1995    57.1    8041    8588   51.22 |
 19. |   2020   1996      50    9102    8741    62.5 |
 20. |   2020   1997      40    9942    8896   48.81 |
     |---------------------------------------------|
 21. |   2020   1998      50   10539    9050   56.41 |
     +---------------------------------------------+
```

Section 3

Analysis with Two-Periods of Panel Data

• We have time periods $t = 1$ and $t = 2$ for each unit i (balanced panel). These periods do not have to be, say, adjacent years. They could be periods far apart in time. Or, they could be close together.
• First consider the case with a single explanatory variable, x_{it}.
• The equation is

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it},\ t = 1, 2.$$

• We observe (x_{it}, y_{it}) for each of the two time periods. The variable $d2_t$ is a constructed time dummy for the second time period: $d2_t = 1$ if $t = 2$ and $d2_t = 0$ if $t = 1$.
• The variable a_i is the unobserved unit effect (or heterogeneity). u_{it} is the unobserved idiosyncratic error.
• We are interested in estimating β_1, the partial effect of x on y. Note that the model assumes this effect is constant over time.

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it}, \; t = 1, 2.$$

• The intercept in the first (base) period is β_0, and that for the second period is $\beta_0 + \delta_0$.

• For policy analysis, x_{it} is often a dummy variable.

• Occasionally x_{it} does not change over time for *any* unit. (Generally, we would expect x_{it} to be constant across time for *some* units.) For example, x_{it} could be gender, or years of schooling for people who have completed their schooling. We will be limited in what we can learn in that case.

• How should we estimate the slope β_1 (and β_0, δ_0 along with it)? One possibility is to just use a pooled OLS analysis. Effectively, define the *composite error* as

$$v_{it} = a_i + u_{it}, \, t = 1, 2$$

and write

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + v_{it}, \, t = 1, 2.$$

- Applying OLS we obtain the **pooled OLS estimator**. We simply regress y on $d2$ and x. (Stata need not even know we have a panel data set; it looks like a regression with one long cross section.)
- A few important issues arise with the POLS estimator:
  1. Even if we assume random sampling across i – which we do – we cannot reasonably assume the observations for i across $t = 1, 2$ are independent. In fact,

     $$\begin{aligned} v_{i1} &= a_i + u_{i1} \\ v_{i2} &= a_i + u_{i2} \end{aligned}$$

     must be correlated because of the presence of a_i.
  2. A more serious issue is that consistency of OLS (as n gets large, as usual) requires that x_{it} and v_{it} are uncorrelated. Because $v_{it} = a_i + u_{it}$, we need

     $$\begin{aligned} Cov(x_{it}, a_i) &= 0 \\ Cov(x_{it}, u_{it}) &= 0 \end{aligned}$$

• When $Cov(x_{it}, a_i) \neq 0$ it is often said that (pooled) OLS suffers from *heterogeneity bias.*
• If the explanatory variable changes over time – at least for some units in the population – heterogeneity bias can be solved by differencing away a_i.

Differencing the Two Years

• To remove the source of bias in POLS, a_i, write the time periods in reverse order for any unit i:

$$
\begin{aligned}
y_{i2} &= (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2} \\
y_{i1} &= \beta_0 + \beta_1 x_{i1} + a_i + u_{i1}
\end{aligned}
$$

Subtract time period one from time period two to get

$$y_{i2} - y_{i1} = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

• If we define $\Delta y_i = y_{i2} - y_{i1}$ (Δx_i and Δu_i are defined similarly), we can write the cross-sectional equation as

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

• Important: β_1 is the original coefficient we are interested in. We have obtained an estimating equation by taking changes or *differencing*.

• Notice that the intercept in the differenced equation,

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

is the *change* in the intercept over the two time periods. It is sometimes interesting to study this change.

• Differencing away the unobserved effect, a_i, is simple but can be very powerful for isolating causal effects.

• If $\Delta x_i = 0$ for all i, or even if Δx_i is the same nonzero constant, this strategy does not work. We need some variation in Δx_i across i.

- The OLS estimator applied to

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

is often called the *first-difference estimator* [With more than two time periods, other orders of differencing are possible, see Aquaro and Čížek (2013, 2014) and references therein]. We will refer to the *FD estimator*.

Example (CRIME2.DTA)

- Relationship between City Unemployment Rates and Crime Rates (Section 13.3).
- 46 cities over the two years (1982 and 1987).
- *crmrte* is the number of crimes per 1,000 people. *unem* is the unemployment rate, in percent.

$$crmrte_{it} = \beta_0 + \delta_0 d87_t + \beta_1 unem_{it} + a_i + u_{it}$$
$$(i = 1, \ldots, 46; t = 1982, 1987)$$

- For historical reasons, cities can have different crime rates. These historical unobserved factors are captured by a_i.

```
clear all
capture log close
set more off
set linesize 82
qui log using lg_wooldridge2013_chapter13_crime2.txt, text replace

/*
  Section 13.3
*/

use CRIME2.DTA
des crmrte unem year ccrmrte
sum crmrte unem year ccrmrte
list year crmrte ccrmrte in 1/10

* Suppose we use a single cross-section
reg crmrte unem if year == 87

* POLS
reg crmrte d87 unem

* FD
reg ccrmrte cunem

qui log close
```

```
.
. /*
>   Section 13.3
> */
.
. use CRIME2.DTA

. des crmrte unem year ccrmrte

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
crmrte          float  %9.0g                  crimes per 1000 people
unem            float  %9.0g                  unemployment rate
year            byte   %9.0g                  82 or 87
ccrmrte         float  %9.0g                  change in crmrte

. sum crmrte unem year ccrmrte

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      crmrte |        92    100.7908    29.84309   50.01925   179.4173
        unem |        92    7.971739    3.374218        2.4       20.3
        year |        92        84.5    2.513699         82         87
     ccrmrte |        46    6.163754    21.21634  -28.44835   65.21219

. list year crmrte ccrmrte in 1/10

     +----------------------------+
     | year     crmrte    ccrmrte |
     |----------------------------|
  1. |   82   74.65756          . |
  2. |   87   70.11729  -4.540268 |
  3. |   82   92.93487          . |
  4. |   87   89.97221  -2.962654 |
  5. |   82   83.61113          . |
     |----------------------------|
  6. |   87   77.19476  -6.416374 |
  7. |   82   88.94253          . |
  8. |   87   84.04099  -4.901543 |
  9. |   82   108.1728          . |
 10. |   87   103.5638  -4.608994 |
```

```
    +---------------------------+

.
. * Suppose we use a single cross-section
. reg crmrte unem if year == 87

      Source |       SS       df       MS              Number of obs =      46
-------------+------------------------------           F(  1,    44) =    1.48
       Model |  1775.90927     1  1775.90927           Prob > F      =  0.2297
    Residual |  52674.6465    44  1197.15106           R-squared     =  0.0326
-------------+------------------------------           Adj R-squared =  0.0106
       Total |  54450.5558    45  1210.01235           Root MSE      =    34.6

------------------------------------------------------------------------------
      crmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |  -4.161134   3.416456    -1.22   0.230    -11.04655     2.72428
       _cons |   128.3781   20.75663     6.18   0.000     86.54589    170.2104
------------------------------------------------------------------------------

.
. * POLS
. reg crmrte d87 unem

      Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  2,    89) =    0.55
       Model |  989.717223     2  494.858612           Prob > F      =  0.5788
    Residual |  80055.7995    89  899.503365           R-squared     =  0.0122
-------------+------------------------------           Adj R-squared = -0.0100
       Total |  81045.5167    91  890.610074           Root MSE      =  29.992

------------------------------------------------------------------------------
      crmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d87 |   7.940416   7.975325     1.00   0.322    -7.906385    23.78722
        unem |   .4265473   1.188279     0.36   0.720    -1.934538    2.787633
       _cons |   93.42025   12.73947     7.33   0.000     68.10719    118.7333
------------------------------------------------------------------------------

.
. * FD
. reg ccrmrte cunem
```

```
      Source |       SS       df       MS              Number of obs =      46
-------------+------------------------------           F(  1,    44) =    6.38
       Model |  2566.43744     1  2566.43744           Prob > F      =  0.0152
    Residual |  17689.5497    44  402.035219           R-squared     =  0.1267
-------------+------------------------------           Adj R-squared =  0.1069
       Total |  20255.9871    45  450.133047           Root MSE      =  20.051

------------------------------------------------------------------------------
     ccrmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       cunem |   2.217999   .8778658     2.53   0.015     .4487771    3.987222
       _cons |    15.4022   4.702117     3.28   0.002      5.92571     24.8787
------------------------------------------------------------------------------

.
. qui log close
```

Example (CRIME2.DTA)

- Cross-section model ($t = 1987$): an increase in unemployment rate was associated with a decrease in crime rate (not intuitive and coefficient not significant anyway).
- POLS: coefficient of *unem* is now positive but still insignificant.
- FD: The intercept means that, if the unemployment rate did not change, the crime rate would be predicted to increase by about 15 crimes per 1,000 people.
- FD: The coefficient on *cunem* is statistically significant and of the sign we might expect: a one percentage point increase in the unemployment rate was associated with about 2.2 more crimes per 1,000 people (0.2 percentage points) per year.

• The same differencing strategy works if x_{it} is a binary program indicator. The differenced equation is the same:

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

In particular, differencing a dummy variable is fine. We interpret the model as if we have estimated it in levels and controlled for a_i.
• In many program evaluation settings no units are treated at $t = 1$ and some are treated at $t = 2$, in which case $\Delta x_i = x_{i2} - x_{i1} = x_{i2}$, so putting in Δx_i is the same as putting in the treatment dummy for the second time period.

Example (JTAIN.DTA)

- Relationship between Job Training Grant and Scrap Rates (Section 13.4).
- Grants not randomly assigned.
- The years we use are 1987 and 1988. Because no firms received grants in 1987, $grant_{i,1987} = 0$ for all $i = 1, \ldots, n$.

$$\begin{aligned} \log(scrap)_{it} &= \beta_0 + \delta_0 d88_t + \beta_1 grant_{it} + a_i + u_{it} \\ \Delta \log(scrap)_i &= \delta_0 + \beta_1 \Delta grant_i + \Delta u_i \\ &= \delta_0 + \beta_1 grant_{i,1988} + \Delta u_i \end{aligned}$$

- a_i contains factors such as average employee ability, capital, managerial skill (roughly constant over a two-year period).

```
clear all
capture log close
set more off
set linesize 82
qui log using lg_wooldridge2013_chapter13_jtrain.txt, text replace

/*
  Section 13.4
*/

use JTRAIN.DTA
des fcode year scrap grant lscrap clscrap cgrant
sum fcode year scrap grant lscrap clscrap cgrant

* Let us focus on two periods only
drop if year > 1988

* FD
xtset fcode year
reg d.(scrap grant)

* FD with dependent variable in log
reg clscrap cgrant
reg d.(lscrap grant)

* POLS
reg lscrap d88 grant

qui log close
```

```
.
. /*
>   Section 13.4
> */
.
. use JTRAIN.DTA

. des fcode year scrap grant lscrap clscrap cgrant

              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------
fcode           float  %9.0g                  firm code number
year            int    %9.0g                  1987, 1988, or 1989
scrap           float  %9.0g                  scrap rate (per 100 items)
grant           byte   %9.0g                  = 1 if received grant
lscrap          float  %9.0g                  log(scrap)
clscrap         float  %9.0g                  lscrap - lscrap_1; year > 1987
cgrant          byte   %9.0g                  grant - grant_1

. sum fcode year scrap grant lscrap clscrap cgrant

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       fcode |        471    415708.9    4022.922     410032     419486
        year |        471        1988    .8173647       1987       1989
       scrap |        162    3.843642     6.00777        .01         30
       grant |        471    .1401274    .3474882          0          1
      lscrap |        162    .3936814    1.486471   -4.60517   3.401197
-------------+---------------------------------------------------------
     clscrap |        108   -.2211324    .5792481  -3.314186   2.397895
      cgrant |        471    .0636943    .4614712         -1          1

.
. * Let us focus on two periods only
. drop if year > 1988
(157 observations deleted)

.
. * FD
. xtset fcode year
       panel variable:  fcode (strongly balanced)
```

```
        time variable:  year, 1987 to 1988
                delta:  1 unit

. reg d.(scrap grant)

      Source |       SS       df       MS              Number of obs =      54
-------------+------------------------------           F(  1,    52) =    1.17
       Model |  6.73345587     1  6.73345587           Prob > F      =  0.2837
    Residual |  298.400031    52  5.73846213           R-squared     =  0.0221
-------------+------------------------------           Adj R-squared =  0.0033
       Total |  305.133487    53   5.7572356           Root MSE      =  2.3955

------------------------------------------------------------------------------
     D.scrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |
         D1. |  -.7394436   .6826276    -1.08   0.284    -2.109236    .6303488
             |
       _cons |  -.5637143   .4049149    -1.39   0.170    -1.376235    .2488069
------------------------------------------------------------------------------

.
. * FD with dependent variable in log
. reg clscrap cgrant

      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

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

. 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
------------------------------------------------------------------------------


.
. * POLS
. reg lscrap d88 grant

      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]
-------------+----------------------------------------------------------------
         d88 |  -.1889081   .3281441    -0.58   0.566    -.8395572     .461741
       grant |   .0566004     .43091     0.13   0.896    -.7978145    .9110152
       _cons |   .5974341   .2057802     2.90   0.005     .1894099    1.005458
------------------------------------------------------------------------------

.
. qui log close
```

Example (`JTAIN.DTA`)

- FD level-level model: having a job training grant was associated with a scrap rate lower by about -0.739 percentage points (but estimate is not significant).
- FD log-level model: having a job training grant was associated with a scrap rate lower by about 31.7% (estimate significant at 10% significance level).
- POLS gives essentially a zero effect. Sign is actually positive. Since it differ so much from the first-difference estimates, it suggests that probably firms with lower-ability workers are more likely to receive a grant.

• In applications where x_{it} is a dummy variable with no assignment in the first period, the FD estimator has a simple interpretation. It is the same as applying OLS to

$$\Delta y_i = \delta_0 + \beta_1 x_{i2} + \Delta u_i$$

where x_{i2} is the second-period program participation (zero or one).
• Regression on a single dummy variable is easy to characterize: the estimate of β_1 is just the difference in means between the "treated" group and the control group:

$$\hat{\beta}_{FD} = \overline{\Delta y}_{treated} - \overline{\Delta y}_{control}.$$

• This has also been called a "difference-in-differences" estimator.

The Strict Exogeneity Assumption

- Recall that the estimating equation in changes or differences is

$$y_{i2} - y_{i1} = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1})$$

- We emphasized that this equation is free of a_i. But we still need the error in this equation to be uncorrelated with the explanatory variable for OLS to be consistent:

$$Cov(x_{i2} - x_{i1}, u_{i2} - u_{i1}) = Cov(\Delta x_i, \Delta u_i) = 0.$$

- When does this condition hold?

• Algebra gives

$$
\begin{aligned}
Cov(x_{i2} - x_{i1}, u_{i2} - u_{i1}) \quad &= \quad Cov(x_{i2}, u_{i2}) - Cov(x_{i1}, u_{i2}) \\
&\quad\; - Cov(x_{i2}, u_{i1}) + Cov(x_{i1}, u_{i2})
\end{aligned}
$$

• We need the explanatory variable in *both* time periods to be uncorrelated with the error in *both* time periods:

$$Cov(x_{is}, u_{it}) = 0,\ s, t = 1, 2.$$

- Note that it is not enough to just assume

$$Cov(x_{i1}, u_{i1}) = Cov(x_{i2}, u_{i2}) = 0$$

or

$$Cov(x_{it}, u_{it}) = 0,\ t = 1, 2,$$

which is known as *contemporaneous exogeneity*.

• The stronger assumption

$$Cov(x_{is}, u_{it}) = 0,\ s, t = 1, 2.$$

is known as *strict exogeneity*. For instance, it rules out feedback from u_{i1} to x_{i2}.

• Remember, if we difference we do not have to worry about $Cov(x_{it}, a_i) \neq 0$.

Multiple Explanatory Variables

• Having more than one explanatory variable causes no problems. A general two-period model is

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it},\ t = 1, 2,$$

Now the first-difference equation looks like

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_{i1} + \beta_2 \Delta x_{i2} + ... + \beta_k \Delta x_{ik} + \Delta u_i,$$

and we just estimate this by OLS.

• We want to interpret the β_j as in the original equation: What is the effect of changing x_j on y, holding fixed the other explanatory variables *and* the unobserved heterogeneity.
• All explanatory variables get differenced, whether they are dummy variables, squares, interactions, and so on. The FD equation is an estimating equation.

Section 4

Differencing with More than Two Years of Data

• First Differencing can be used with more than two years of panel data, but we must be careful to account for serial correlation (and, as usual, possibly heteroskedasticity) in the FD equation. This is because the FD equation is no longer just a single cross section.
• Generally, we should also include a full set of time dummies to account for time effects.
• With T time periods, where now $T \geq 2$, we can write

$$y_{it} = \delta_1 + \delta_2 d2_t + ... + \delta_T dT_t + \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it}$$

where now δ_1 denotes the intercept in the first year and δ_t, $t \geq 2$, is the difference between the intercept in period t and period 1.
• The model still contains an unobserved effect, a_i, and idiosyncratic error, u_{it}.

• When we difference we lose the first time period, as before, but we are left with a panel data set if we start with $T \geq 3$:

$$\Delta y_{it} = \delta_2 \Delta d2_t + ... + \delta_T \Delta dT_t + \beta_1 \Delta x_{it1} + \beta_2 \Delta x_{it2} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

for $t = 2, 3, ..., T$.
• Unless we are interested in the original δ_t, it is easier to include an overall intercept and not difference the time dummies. We lose a time dummy because we lose the first time period:

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + ... + \alpha_T dT_t + \beta_1 \Delta x_{it1} + \beta_2 \Delta x_{it2} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

• Now we just use POLS on the changes for $t = 2, ..., T$.
• Except for changing how we allow for different time intercepts, this is the same model as before. Estimates of the β_j are identical.

• As before, we have eliminated a_i – which we think is the main source of correlation between the x_{itj} and the composite error $v_{it} = a_i + u_{it}$. But we can see we need

$$Cov(\Delta x_{itj}, \Delta u_{it}) = 0,\ j = 1, ..., k$$

and this is a strict exogeneity assumption. Sufficient is for each j,

$$Cov(x_{isj}, u_{it}) = 0, \text{ all } s, t = 1, ..., T$$

• If we write the FD equation as

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + ... + \alpha_T dT_t + \beta_1 \Delta x_{it1} + \beta_2 \Delta x_{it2} + ... + \beta_k \Delta x_{itk} + e_{it}$$

with errors

$$e_{it} = \Delta u_{it},\ t = 2, 3, ...T,$$

then, in general, the e_{it} will be correlated across time (relevant only when $T > 2$).

• In fact, if the original errors $\{u_{it}\}$ are uncorrelated with constant variance σ_u^2, it can be shown that

$$Corr(e_{it}, e_{i,t+1}) = -1/2$$

• Generally, we should expect complicated serial correlation in $\{u_{it}\}$, which results in a complicated pattern for $\{e_{it}\}$. With a large cross-sectional sample size N, and assuming T is not larger than N, we can use clustering.

• Stata has a few different ways of doing FD. First, after using `xtset`, we can construct the differences ourselves.

```
xtset id year
gen cy = D.y
gen cx1 = D.x1
...
gen cxk = D.xk
reg cy d3 d4 ... dT cx1 cx2 ... cxk
reg cy d3 d4 ... dT cx1 cx2 ... cxk, cluster(id)
```

• Question: Why do we need to use `xtset` before using the `D.` operator?

• Reminder: In using FD methods, it is important to remember that an equation such as

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + ... + \alpha_T dT_t + \beta_1 \Delta x_{it1} + \beta_2 \Delta x_{it2} + ... + \beta_k \Delta x_{itk} + \Delta u_{it}$$

is an *estimating* equation used to get rid of a_i. The estimates should be interpreted in the context of the original "levels" equation

$$y_{it} = \delta_1 + \delta_2 d2_t + ... + \delta_T dT_t + \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it}$$

• With the levels equation as the starting point, it is easy to choose explanatory variables for policy evaluation that allow, say, lagged effects. We can also add quadratics, interactions, and so on. These are constructed before applying differencing.

Example (CRIME4.DTA)

- County crime rates in North Carolina (Example 13.9).
- $n = 90$ and $t = 1981, \ldots, 1987$.
- As in standard criminometric studies, we use logs of variables to estimate elasticities.

```
clear all
capture log close
set more off
set linesize 82
qui log using lg_wooldridge2013_chapter13_crime4.txt, text replace

/*
  Example 13.9
*/

use CRIME4.DTA
des crmrte prbarr prbconv prbpris avgsen polpc year d82-d87
sum crmrte prbarr prbconv prbpris avgsen polpc year d82-d87
reg clcrmrte d83-d87 clprbarr clprbcon clprbpri clavgsen clpolpc

qui log close
```

```
.
. /*
>   Example 13.9
> */
.
. use CRIME4.DTA

. des crmrte prbarr prbconv prbpris avgsen polpc year d82-d87

              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------------------
crmrte          float  %9.0g                  crimes committed per person
prbarr          float  %9.0g                  'probability' of arrest
prbconv         float  %9.0g                  'probability' of conviction
prbpris         float  %9.0g                  'probability' of prison sentenc
avgsen          float  %9.0g                  avg. sentence, days
polpc           float  %9.0g                  police per capita
year            byte   %9.0g                  81 to 87
d82             byte   %9.0g                  =1 if year == 82
d83             byte   %9.0g                  =1 if year == 83
d84             byte   %9.0g                  =1 if year == 84
d85             byte   %9.0g                  =1 if year == 85
d86             byte   %9.0g                  =1 if year == 86
d87             byte   %9.0g                  =1 if year == 87

. sum crmrte prbarr prbconv prbpris avgsen polpc year d82-d87

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      crmrte |       630    .0315876    .0181209   .0018116    .163835
      prbarr |       630    .3073682    .1712047   .0588235       2.75
     prbconv |       630    .6886176    1.690345   .0683761         37
     prbpris |       630    .4255184    .0872452    .148936    .678571
      avgsen |       630     8.95454    2.658082       4.22      25.83
-------------+--------------------------------------------------------
       polpc |       630    .0019168    .0027349   .0004585   .0355781
        year |       630          84    2.001589         81         87
         d82 |       630    .1428571    .3502052          0          1
         d83 |       630    .1428571    .3502052          0          1
         d84 |       630    .1428571    .3502052          0          1
-------------+--------------------------------------------------------
```

```
         d85 |        630    .1428571    .3502052          0          1
         d86 |        630    .1428571    .3502052          0          1
         d87 |        630    .1428571    .3502052          0          1

. reg clcrmrte d83-d87 clprbarr clprbcon clprbpri clavgsen clpolpc

      Source |       SS       df       MS              Number of obs =     540
-------------+------------------------------           F( 10,   529) =   40.32
       Model |  9.6004283    10  .96004283             Prob > F      =  0.0000
    Residual | 12.5963755   529  .023811674            R-squared     =  0.4325
-------------+------------------------------           Adj R-squared =  0.4218
       Total | 22.1968038   539  .041181454            Root MSE      =  .15431

------------------------------------------------------------------------------
    clcrmrte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d83 |  -.0998658   .0238953    -4.18   0.000    -.1468071   -.0529246
         d84 |  -.0479374   .0235021    -2.04   0.042    -.0941063   -.0017686
         d85 |  -.0046111   .0234998    -0.20   0.845    -.0507756    .0415533
         d86 |   .0275143   .0241494     1.14   0.255    -.0199261    .0749548
         d87 |   .0408267   .0244153     1.67   0.095    -.0071361    .0887895
    clprbarr |  -.3274942   .0299801   -10.92   0.000    -.3863889   -.2685995
    clprbcon |  -.2381066   .0182341   -13.06   0.000    -.2739268   -.2022864
    clprbpri |  -.1650462    .025969    -6.36   0.000    -.2160613   -.1140312
    clavgsen |  -.0217607   .0220909    -0.99   0.325    -.0651574     .021636
     clpolpc |   .3984264    .026882    14.82   0.000     .3456177     .451235
       _cons |   .0077134   .0170579     0.45   0.651    -.0257961    .0412229
------------------------------------------------------------------------------

.
. qui log close
```

Example (CRIME4.DTA)

- The three probability variables (of arrest, conviction, and serving prison time) all have the expected sign, and they are all statistically significant.
- The coefficient on the police per capita is positive and significant (as in many study in this field). Interpreted causally, it says that 1% increase in police per capita *increases* crime rate by about 4%. What is going on here?

Section 5

Fixed Effects Estimation

• Differencing is one method of eliminating a_i (which itself is sometimes called a *fixed effect*).
• Alternatively, can use the "fixed effects" or "within" transformation: remove the within i time averages.
• In the simple model with only x_{it}:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

• Average this equation across t to get

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ is a "time average" for unit i. Similarly for $\bar{x}_i$ and $\bar{u}_i$.

• Subtract the time-averaged equation – sometimes called the "between equation" – from other time periods:

$$
\begin{aligned}
y_{it} - \bar{y}_i &= \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \\
\ddot{y}_{it} &= \beta_1 \ddot{x}_{it} + \ddot{u}_{it}
\end{aligned}
$$

for $t = 1, \ldots, T$ and $i = 1, \ldots, n$. As with the FD equation, this equation is free of a_i.

• We view this "time-demeaned" (or "within") equation as an estimating equation. As with FD, we interpret β_1 in the levels equation

$$
y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}
$$

• We can use pooled OLS on the deviations from time averages to estimate β_1. Called the "fixed effects" (FE) estimator or the "within estimator."

• Like FD, FE can be applied with $T \geq 2$. An interesting algebraic fact is that FD and FD are numerically identical when $T = 2$. For $T > 2$, they are different. Typically find larger differences for larger T.
• Like FD, FE allows arbitrary correlation between x_{it} and a_i, but it requires a strict exogeneity assumption with respect to $\{u_{it}\}$. As with FD, a sufficient condition is

$$Cov(x_{is}, u_{it}) = 0, \; s, t = 1, ..., T.$$

• FE has an advantage over FD when strict exogeneity fails: under reasonable assumptions with large T, FE tends to have less bias (Wooldridge, 2010, Chapter 10).

• Because a_i is removed using the FE transformation, it cannot be a source of serial correlation. But the u_{it} might have serial correlation (and heteroskedasticity), and so use “cluster robust” inference. (Using “robust” by itself is never a good idea with FE.)
• In Stata the command is `xtreg`. It computes proper standard errors and test statistics. But the default is that $\{u_{it}\}$ is homoskedastic and serially uncorrelated, so add a “cluster” option.

• As in other situations, it is easy to include a full set of year dummies and many explanatory variables.
• Computing pooled OLS, FD, and FE estimators can be informative and should be done in applications. If FD and FE are very different, it is a sign that strict exogeneity fails. If POLS is different from FE (say), it indicates explanatory variables correlated with a_i.
• Do not worry too much about goodness-of-fit with FE. The "within" R-squared is probably most informative (based on time-demeaned equation)

Example (JTRAIN.DTA)

- Relationship between job training and firm scrap rates (Example 14.1).
- We must allow for the possibility that the additional job training in 1988 made workers more productive in 1989 (i.e. include $grant_{i,t-1}$ as explanatory variable).

```
clear all
capture log close
set more off
set linesize 82
qui log using lg_wooldridge2013_chapter14_jtrain.txt, text replace

/*
  Example 14.1
*/

use JTRAIN.DTA
xtset fcode year
xtreg lscrap grant grant_1 d88 d89, fe
xtreg lscrap grant d88 d89, fe

qui log close
```

```
.
. /*
>   Example 14.1
> */
.
. use JTRAIN.DTA

. xtset fcode year
       panel variable:  fcode (strongly balanced)
        time variable:  year, 1987 to 1989
                delta:  1 unit

. xtreg lscrap grant grant_1 d88 d89, fe

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

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

                                                F(4,104)           =      6.54
corr(u_i, Xb)  = -0.0714                        Prob > F           =    0.0001

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.2523149    .150629    -1.68   0.097    -.5510178    .0463881
     grant_1 |  -.4215895      .2102    -2.01   0.047    -.8384239   -.0047551
         d88 |  -.0802157   .1094751    -0.73   0.465     -.297309    .1368776
         d89 |  -.2472028   .1332183    -1.86   0.066    -.5113797    .0169741
       _cons |   .5974341   .0677344     8.82   0.000     .4631142    .7317539
-------------+----------------------------------------------------------------
     sigma_u |   1.438982
     sigma_e |  .49774421
         rho |  .89313867   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(53, 104) =    24.66             Prob > F = 0.0000

. xtreg lscrap grant d88 d89, fe

Fixed-effects (within) regression               Number of obs      =       162
```

```
Group variable: fcode                          Number of groups   =        54

R-sq:  within  = 0.1701                        Obs per group: min =         3
       between = 0.0189                                       avg =       3.0
       overall = 0.0130                                       max =         3

                                               F(3,105)           =      7.18
corr(u_i, Xb)  = -0.0109                       Prob > F           =    0.0002

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       grant |  -.0822141   .1262632    -0.65   0.516    -.3325706    .1681424
         d88 |   -.140066    .106835    -1.31   0.193       -.3519    .0717681
         d89 |    -.42704   .0999338    -4.27   0.000    -.6251903   -.2288897
       _cons |   .5974341   .0687024     8.70   0.000     .4612098    .7336583
-------------+----------------------------------------------------------------
     sigma_u |  1.4283441
     sigma_e |  .50485774
         rho |  .88894293   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(53, 105) =    23.90             Prob > F = 0.0000

.
. qui log close
```

Example (JTRAIN.DTA)

- Coefficient on $grant_{it}$ significant at 10% level.
- Coefficient on $grant_{i,t-1}$ significant at 5% level: Obtaining a grant in 1988 was associated with a 42.2% lower firm scrap rate in 1989.
- We must allow for the possibility that the additional job training in 1988 made workers more productive in 1989 (i.e. include $grant_{i,t-1}$ as explanatory variable).
- The coefficient on $d89$ indicates that the scrap rate was substantially lower in 1989 than in the base year (1987), even in absence of job training (if we omit this dummies, this effect could be captured by $grant_{it}$ and $grant_{i,t-1}$.
- If we omit $grant_{i,t-1}$, the estimate on *grant* becomes much smaller and statistically insignificant.

Section 6

Random Effects Estimation

• Suppose we start with the same equation as before, written for a unit i:

$$y_{it} = \delta_t + \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it},\ t = 1, 2, ..., T$$

where δ_t represent different time intercepts. In some of what follows we drop them for simplicity.

• Unlike FD and FE, Random Effects (RE) estimation leaves a_i in the error term, and then accounts for the serial correlation over time in $v_{it} = a_i + u_{it}$ via a *generalized least squares* (GLS) procedure.

• RE estimation allows time-constant explanatory variables and this is often important in RE applications. (With FD and FE, we can think of all time-constant variables as being captured by a_i.)

• How come RE can include time-constant variables (such as gender)? Like pooled OLS, RE estimation maintains

$$Cov(x_{itj}, a_i) = 0,\ j = 1, ..., k$$

and also strict exogeneity with respect to $\{u_{it}\}$:

$$Cov(x_{isj}, u_{it}) = 0,\ s, t = 1, ..., T$$

• For consistency, RE maintains that the composite error term, $v_{it} = a_i + u_{it}$, is uncorrelated with the explanatory variables in all time periods.

• For policy analysis, RE is typically less *convincing* than FD or FE: we want participation to be correlated with the time-constant factors in a_i.

• However, with good time-constant controls, RE may be *convincing*. This is because more is taken out of a_i as we add time-constant variables.

• When it is consistent, RE is typically more efficient – sometimes much more efficient – than FD or FE.

• The "standard" RE assumptions also include that

$$
\begin{aligned}
Cov(a_i, u_{it}) &= 0 \quad \text{(not especially controversial)} \\
Var(u_{it}) &= \sigma_u^2 \text{ for all } t \quad \text{(constant variance over time)} \\
Cov(u_{is}, u_{it}) &= 0,\ t \neq s \quad \text{(no serial correlation)}
\end{aligned}
$$

• These are assumed to hold for random draws i from the population.
• The first is "uncontroversial" because we are just separating the time-constant and time-varying unobservables.
• There are good reasons why the second and third of these fail, and empirically they often do. Fortunately, the RE estimator does not rely on any of these assumptions for *consistency*. But if they fail, we must use cluster-robust inference.

• The GLS procedure uses properties of the composite error, $v_{it} = a_i + u_{it}$ Under the "standard" RE assumptions,

$$\begin{aligned} Var(v_{it}) &= \sigma_a^2 + \sigma_u^2 \\ Cov(v_{it}, v_{is}) &= Var(a_i) = \sigma_a^2 \end{aligned}$$

where the first follows from $Cov(a_i, u_{it}) = 0$ and the last follows from $Cov(u_{it}, u_{is}) = 0$, $t \neq s$.

• Using these two equations, the serial correlation in $\{v_{it}\}$ is easy to characterize:

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \equiv \rho$$

• ρ is the share of the total variance, $\sigma_a^2 + \sigma_u^2$, due to the variance in a_i, the unobserved effect, σ_a^2.
• Note that the correlation is the same no matter how far apart t and s are. Can be unrealistic.
• In estimation, RE acts as if there is a single correlation parameter, ρ.

• A very useful characterization of RE is in terms of a "partially" time demeaned equation. Define a parameter, θ, that is between zero and one as

$$\theta = 1 - \left[\frac{1}{1 + T(\sigma_a^2/\sigma_u^2)} \right]^{1/2}.$$

• The variances σ_a^2 and σ_u^2 can be estimated after POLS (or FE) estimation, which then allows us to estimate θ by $\hat{\theta}$. Then RE estimate can be obtained from the pooled OLS regression

$$y_{it} - \hat{\theta}\bar{y}_i \text{ on } \mathbf{x}_{it} - \hat{\theta}\bar{\mathbf{x}}_i, t = 1, ..., T; i = 1, ..., N.$$

• Call $y_{it} - \hat{\theta}\bar{y}_i$ a "partially-time-demeaned" variable: only a fraction of the mean is removed. Same with the k elements of $\mathbf{x}_{it} - \hat{\theta}\bar{\mathbf{x}}_i$.
• It is easy to see that

$$
\begin{aligned}
\hat{\theta} &\approx 0 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{POLS} \\
\hat{\theta} &\approx 1 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{FE}
\end{aligned}
$$

• When is θ close to one? (i) σ_a^2/σ_u^2 is "large" or (ii) T is large. With large T, FE and RE can be similar.
• If $\mathbf{x}_{it}$ includes time-constant variables $\mathbf{z}_i$, then $(1-\hat{\theta})\mathbf{z}_i$ appears as a regressor in the RE estimation. If $\hat{\theta} = 1$ (FE) are they eliminated.

• No need to do the transformation by hand. Stata does it.

```
xtset id year
xtreg y x1 x2 ...  xK, re
xtreg y x1 x2 ...  xK, re cluster(id)
```

The first `xtreg` command produces the usual, nonrobust inference, the third makes the inference fully robust.

Example (WAGEPAN.DTA)

- Return to Union Membership Using Panel Data (Example 14.4).
- Working men from 1980 to 1987, so eight years. $n = 545$. Use $lwage$ as the dependent variable and standard panel data methods.
- Union status and marital status change over time. Education does not. Experience does, but if we know the experience in 1980 we know it in any other year (increases by one each year).
- Might be worried about strict exogeneity: Why are union status and marital status changing over time? Do shocks to wages contribute?

```
clear all
capture log close
set more off
set linesize 82
qui log using lg_wooldridge2013_chapter14_wagepan.txt, text replace

/*
  Example 14.4
*/

use wagepan.dta

* POLS
reg lwage educ black hisp exper expersq married union d81-d87

xtset nr year

* RE
xtreg lwage educ black hisp exper expersq married union d81-d87, re

* FE
xtreg lwage expersq married union d81-d87, fe

qui log close
```

```
.
. /*
>   Example 14.4
> */
.
. use wagepan.dta

.
. * POLS
. reg lwage educ black hisp exper expersq married union d81-d87

      Source |       SS       df       MS              Number of obs =    4360
-------------+------------------------------           F( 14,  4345) =   72.46
       Model |  234.048277    14  16.7177341           Prob > F      =  0.0000
    Residual |  1002.48136  4345  .230720682           R-squared     =  0.1893
-------------+------------------------------           Adj R-squared =  0.1867
       Total |  1236.52964  4359  .283672779           Root MSE      =  .48033

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0913498   .0052374    17.44   0.000     .0810819    .1016177
       black |  -.1392342   .0235796    -5.90   0.000    -.1854622   -.0930062
        hisp |   .0160195   .0207971     0.77   0.441    -.0247535    .0567925
       exper |   .0672345   .0136948     4.91   0.000     .0403856    .0940834
     expersq |  -.0024117     .00082    -2.94   0.003    -.0040192   -.0008042
     married |   .1082529   .0156894     6.90   0.000     .0774937    .1390122
       union |   .1824613   .0171568    10.63   0.000     .1488253    .2160973
         d81 |     .05832   .0303536     1.92   0.055    -.0011886    .1178286
         d82 |   .0627744   .0332141     1.89   0.059    -.0023421    .1278909
         d83 |   .0620117   .0366601     1.69   0.091    -.0098608    .1338843
         d84 |   .0904672   .0400907     2.26   0.024      .011869    .1690654
         d85 |   .1092463   .0433525     2.52   0.012     .0242533    .1942393
         d86 |   .1419596    .046423     3.06   0.002     .0509469    .2329723
         d87 |   .1738334    .049433     3.52   0.000     .0769194    .2707474
       _cons |   .0920558   .0782701     1.18   0.240    -.0613935    .2455051
------------------------------------------------------------------------------

.
. xtset nr year
       panel variable:  nr (strongly balanced)
        time variable:  year, 1980 to 1987
```

```
              delta:  1 unit

.
. * RE
. xtreg lwage educ black hisp exper expersq married union d81-d87, re

Random-effects GLS regression                   Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1799                         Obs per group: min =         8
       between = 0.1860                                        avg =       8.0
       overall = 0.1830                                        max =         8

                                                Wald chi2(14)      =    957.77
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0918763   .0106597     8.62   0.000     .0709836    .1127689
       black |  -.1393767   .0477228    -2.92   0.003    -.2329117   -.0458417
        hisp |   .0217317   .0426063     0.51   0.610    -.0617751    .1052385
       exper |   .1057545   .0153668     6.88   0.000     .0756361    .1358729
     expersq |  -.0047239   .0006895    -6.85   0.000    -.0060753   -.0033726
     married |    .063986   .0167742     3.81   0.000     .0311091    .0968629
       union |   .1061344   .0178539     5.94   0.000     .0711415    .1411273
         d81 |    .040462   .0246946     1.64   0.101    -.0079385    .0888626
         d82 |   .0309212   .0323416     0.96   0.339    -.0324672    .0943096
         d83 |   .0202806    .041582     0.49   0.626    -.0612186    .1017798
         d84 |   .0431187   .0513163     0.84   0.401    -.0574595    .1436969
         d85 |   .0578155   .0612323     0.94   0.345    -.0621977    .1778286
         d86 |   .0919476   .0712293     1.29   0.197    -.0476592    .2315544
         d87 |   .1349289   .0813135     1.66   0.097    -.0244427    .2943005
       _cons |   .0235864   .1506683     0.16   0.876     -.271718    .3188907
-------------+----------------------------------------------------------------
     sigma_u |  .32460315
     sigma_e |  .35099001
         rho |  .46100216   (fraction of variance due to u_i)
------------------------------------------------------------------------------

.
. * FE
. xtreg lwage expersq married union d81-d87, fe
```

```
Fixed-effects (within) regression               Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1806                         Obs per group: min =         8
       between = 0.0286                                        avg =       8.0
       overall = 0.0888                                        max =         8

                                                F(10,3805)         =     83.85
corr(u_i, Xb)  = -0.1222                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     expersq |  -.0051855   .0007044    -7.36   0.000    -.0065666   -.0038044
     married |   .0466804   .0183104     2.55   0.011     .0107811    .0825796
       union |   .0800019   .0193103     4.14   0.000     .0421423    .1178614
         d81 |   .1511912   .0219489     6.89   0.000     .1081584     .194224
         d82 |   .2529709   .0244185    10.36   0.000     .2050963    .3008454
         d83 |   .3544437   .0292419    12.12   0.000     .2971125    .4117749
         d84 |   .4901148   .0362266    13.53   0.000     .4190894    .5611402
         d85 |   .6174823   .0452435    13.65   0.000     .5287784    .7061861
         d86 |   .7654966   .0561277    13.64   0.000     .6554532    .8755399
         d87 |   .9250249   .0687731    13.45   0.000     .7901893    1.059861
       _cons |   1.426019   .0183415    77.75   0.000     1.390058    1.461979
-------------+----------------------------------------------------------------
     sigma_u |  .39176195
     sigma_e |  .35099001
         rho |  .55472817   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3805) =     9.16           Prob > F = 0.0000

.
. qui log close
```

Example (`WAGEPAN.DTA`)

- RE estimates of return to schooling, experience are similar to POLS. But union and marriage premiums are much smaller. Robust SEs somewhat larger than nonrobust ones. Valid to compare POLS robust SEs and RE robust SEs. The latter are smaller on union and married, so RE appears to be more precise.
- FE: Two variables are dropped: educ does not change over time for anyone in the sample. *exper* must also be dropped because it increases by one for every year in the sample. The starting points for *exper* are different across people but not distinguishable from the fixed effect. Union and marriage premiums drop even further. FE probably the most reliable estimates.

Section 7

Choosing Among POLS, FD, FE, and RE

Summary

- We have covered four estimators:
  1. POLS, which is on the levels.
  2. FD, which is POLS but on the differences (changes)
  3. FE, which is POLS on the time-demeaned variables
  4. RE, which is POLS on the partially time-demeaned variables
- POLS on the levels is usually deficient, unless we include things like lagged y (not allowed in the other methods). With good controls and lags of y, we might be able to make a convincing analysis. But many economists prefer unobserved effects models.

Summary: FD versus FE

• If FD and FE are different in important ways, the strict exogeneity assumption may be violated. Wooldridge (2010, Chapter 10) shows how to test whether shocks u_{it} at time t are correlated with future explanatory variables. (We need $T \geq 3$ time periods for this test.)

Summary: RE versus FE

• Time-constant variables drop out of FE estimation and the estimates may be imprecise.
• On the time-varying covariates, are FE and RE so different after all? Define the parameter

$$\theta = 1 - \left[\frac{1}{1 + T(\sigma_a^2/\sigma_u^2)} \right]^{1/2},$$

which is consistently estimated (for fixed T) by $\hat{\theta}$. The RE estimate can be obtained from the pooled OLS regression

$$y_{it} - \hat{\theta}\bar{y}_i \text{ on } \mathbf{x}_{it} - \hat{\theta}\bar{\mathbf{x}}_i, t = 1, ..., T; i = 1, ..., N.$$

• Call $y_{it} - \hat{\theta}\bar{y}_i$ a "quasi-time-demeaned" variable: only a fraction of the mean is removed.

$$\begin{aligned} \hat{\theta} &\approx 0 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{POLS} \\ \hat{\theta} &\approx 1 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{FE} \end{aligned}$$

θ increases to unity as (i) σ_a^2/σ_u^2 increases or (ii) T increases. With large T, FE and RE are often similar.

• If $\mathbf{x}_{it}$ includes time-constant variables $\mathbf{z}_i$, then $(1 - \hat{\theta})\mathbf{z}_i$ appears as a regressor.

• Recall the key RE assumption is $Cov(\mathbf{x}_{it}, a_i) = 0$. With lots of good time-constant controls ("observed heterogeneity," such as industry dummies) might be able to make this condition roughly true.
• But if RE and FE give estimates that differ in economically important ways, we want evidence that RE should be rejected in favor of FE. (RE usually gives smaller standard errors; sometimes much smaller. So we would use it if we can.)
• There is a way to directly compare the RE and FE estimators on explanatory variables that change across i and t. Called the *Hausman Test*, see Wooldridge (2010, Chapter 10).