Lecture 5: Panel data analysis (II)

Prof. dr. Wolter Hassink
Utrecht University School of Economics
w.h.j.hassink@uu.nl

Contents:

- Least Squares Dummy Variable estimator
- Within estimator (or fixed effect estimator)
- Fixed effects versus first differences
- Pooled OLS
- Random-effects estimator
- Random effects or fixed effects: Hausman test
- Line of reasoning with panel data: final example

Material:

Wooldridge:
Chapter 14: 14.1, 14.2, 14.3

MOTIVATION

- Example: panel information of 900 international businesses (firms). Years: 2019, 2020, 2021.
- Variable of interest: Profits of these international firms. Research question: can we explain the volume of profits by the country of origin (= location of headquarter)? E.g., firm has its headquarter in UK, The Netherlands, Germany or France (4 options).
- For one year (cross section)? Answer: **YES** (reference group: Germany)

$$profits_i = \beta_0 + \beta_1 UK_i + \beta_2 France_i + \beta_3 Netherlands_i + u_i$$
$$i = 1,...,900$$

- For three years? Answer: **IT DEPENDS**. Week 4:
- Regression equation is specified as:

$$profits_{it} = a_i + \beta_1 UK_i + \beta_2 France_i + \beta_3 Netherlands_i + u_{it}$$
$$i = 1,...,900; t = 2019, 2020, 2021$$

$UK, France, Netherlands$ and $Germany$ are 0-1 indicators:
$UK_i + France_i + Netherlands_i + Germany_i = 1$

- Can we estimate the regression parameters by the first difference estimator? Answer: **NO**, because $\Delta UK_i = 0$, $\Delta France_i = 0$ and $\Delta Netherlands_i = 0$

$$\Delta profits_{it} = \beta_1 \Delta UK_i + \beta_2 \Delta France_i + \beta_3 \Delta Netherlands_i + \Delta u_{it}$$
$$i = 1,...,900; t = 2019, 2020, 2021$$

- Can we estimate the parameters by the Pooled OLS estimator? Answer: **YES**

$$profits_{it} = \beta_0 + \beta_1 UK_i + \beta_2 France_i + \beta_3 Netherlands_i + v_{it}$$
$$v_{it} = a_i + u_{it} \qquad i = 1,...,900; t = 2019, 2020, 2021$$

THIS WEEK: THREE QUESTIONS

QUESTION 1: Can we apply alternative estimators, next to first difference estimator? This week: **LSDV estimator** and **fixed effects estimator**

QUESTION 2: Can we have a panel data estimator that includes country of headquarter? This week: **random effects estimator**

QUESTION 3: Is there any road map for the preferred estimator? Five options: 1. first difference estimator, 2. pooled OLS (both in week 4), 3. LSDV estimator, 4. fixed effects estimator, 5. random effects estimator (this week). These week: several testing procedures.

Least Squares Dummy Variable estimator

Remember the table of week 4

Table A: Estimation methods under different assumptions of strict exogeneity and on the correlation between the individual effect and RHS-variables:

	$E(a_i \mid x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}) \neq 0$ Correlation between a_i and all of the explanatory variables is allowed to be nonzero	$E(a_i \mid x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}) = 0$ A zero correlation between a_i and all of the explanatory variables is assumed.
All x_{it} strictly exogenous	1. First differences **2. LSDV procedure** **3. Within estimation**	**4. Random effects**
Some x_{it} not strictly exogenous	Instrumental Variables (IV)	5. Pooled OLS (no lagged dependent variables) 6. Instrumental variables (IV) (lagged dep. vars. Included)

$$y_{it} = a_i + \beta_1 x_{1it} + ... + \beta_k x_{kit} + u_{it} \quad i = 1,...,N; t = 1,...,T$$

Previous week:

- First-difference estimator (here only one explanatory variable)
  $\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it}$ $\quad\quad i = 1,...,N; t = 2,...,T$
- Pooled OLS
  $y_{it} = \beta_0 + \beta_1 x_{1it} + ... + \beta_k x_{kit} + v_{it}$ for which $v_{it} = a_i + u_{it}$

This week we will consider the following three estimators:

- LSDV estimator (least square dummy variable estimator)
- Within estimator (also referred to as the fixed-effects estimator)
- Random effects estimator

Estimation method 2: Least Squares Dummy Variable estimator (LSDV-method)

Aim: to introduce the LSDV-method.

The LSDV method is straightforward. In the regression equation we include a full set of dummy variables a_i indicating the cross-sectional unit. The OLS estimator is used to estimate the parameters of the equation, which thus includes all the parameters of the dummy variables

$$y_{it} = \underbrace{\beta_1 x_{1it} + \ldots + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T \qquad (1)$$

Each individual of the panel data set (e.g. firm, person, country) gets a separate dummy variable. Consequently:

- The correlation between the explanatory variable a_i and the other explanatory variables $x_{i1},\ldots x_{iT}$: $E(a_i \mid x_{1it},\ldots,x_{1iT},\ldots,x_{ki1},\ldots x_{kiT}) \neq 0$.
- Strict exogeneity (in simple words: which means that equation (1) has no lagged dependent variables (y_{it-1}), there is no feedback mechanism)

In the case of a country data set with a limited set of individual firms, this is not a problem (not too many cross-sectional units). But in a panel data set this may be hard (too many dummy variables) because of computational issues. For instance, a panel data set of about 1.5 million households (which is not uncommon nowadays), requires the inclusion of as many dummy variables in the regression equation.

- An application of the LSDV estimator in Stata will be given below.

Some useful Stata commands (I)

LSDV

- id-firm is the name of the variable indicating the cross sectional unit (e.g the firm indicator)
- important command to create dummies (it creates df1, df2,…, dfN for each of the units):
  - quietly tab id_firm, gen(df)
- An OLS regression of y on x and the set of dummies df* (you need to include the asterisk)
  - reg y x df*
- testparm df*
- reg y x df*, cluster(id_firm)

Example: leveragedata.dta

LSDV:

```
. reg leverage i.year cashflow_assets net_income_MV capex_MV volatility gdp_growth
inflation df*
note: df4453 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =   38018
-------------+------------------------------           F(4481, 33536) =   32.54
       Model |  2304.23662  4481  .514223749           Prob > F      =  0.0000
    Residual |  529.969924 33536  .015803015           R-squared     =  0.8130
-------------+------------------------------           Adj R-squared =  0.7880
       Total |  2834.20655 38017  .074551031           Root MSE      =  .12571

------------------------------------------------------------------------------
       leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           year |
          1991  |  -.0071953   .0128123    -0.56   0.574    -.0323079    .0179173
          1992  |   .0063964    .012046     0.53   0.595    -.0172141     .030007
          1993  |  -.0500511   .0120103    -4.17   0.000    -.0735917   -.0265104
          1994  |  -.0187898   .0118235    -1.59   0.112    -.0419643    .0043847
          1995  |  -.0041872   .0115466    -0.36   0.717     -.026819    .0184446
          1996  |  -.0273106   .0114664    -2.38   0.017    -.0497853    -.004836
          1997  |  -.0383464    .011436    -3.35   0.001    -.0607614   -.0159314
          1998  |  -.0394148   .0115158    -3.42   0.001    -.0619861   -.0168435
          1999  |  -.0286589   .0115514    -2.48   0.013       -.0513   -.0060179
          2000  |  -.0048752   .0112476    -0.43   0.665    -.0269209    .0171704
          2001  |  -.0107868   .0111608    -0.97   0.334    -.0326624    .0110889
          2002  |   .0005191   .0112197     0.05   0.963    -.0214719    .0225102
          2003  |    -.02291   .0112636    -2.03   0.042     -.044987    -.000833
          2004  |  -.0283002   .0112057    -2.53   0.012    -.0502636   -.0063367
          2005  |  -.0476214   .0111731    -4.26   0.000    -.0695211   -.0257217
          2006  |  -.0417042   .0111741    -3.73   0.000    -.0636059   -.0198025
          2007  |  -.0310703   .0111437    -2.79   0.005    -.0529123   -.0092282
          2008  |   .0297393   .0110511     2.69   0.007     .0080788    .0513997
          2009  |  -.0450109   .0120865    -3.72   0.000    -.0687009    -.021321
          2010  |   .0212174   .0112624     1.88   0.060    -.0008573     .043292
          2011  |   .0361837   .0111027     3.26   0.001      .014422    .0579454
          2012  |    .007615   .0112974     0.67   0.500    -.0145283    .0297583
          2013  |  -.0082831   .0115645    -0.72   0.474    -.0309498    .0143837
                |
cashflow_assets |  -.3282559   .0119771   -27.41   0.000    -.3517314   -.3047804
  net_income_MV |  -.0717367   .0022064   -32.51   0.000    -.0760613   -.0674121
       capex_MV |    .281408   .0048738    57.74   0.000     .2718552    .2909609
     volatility |   .0016269   .0001371    11.86   0.000     .0013581    .0018957
     gdp_growth |  -.0095725   .0004607   -20.78   0.000    -.0104755   -.0086695
      inflation |  -.0027681   .0009939    -2.78   0.005    -.0047163     -.00082
            df1 |  -.0250849   .0537778    -0.47   0.641    -.1304912    .0803215
            df2 |   .4366369   .0557651     7.83   0.000     .3273353    .5459384
            df3 |   .4939437   .0636731     7.76   0.000     .3691422    .6187452
             . . . . .
         df4452 |   .3410562   .0440643     7.74   0.000     .2546886    .4274237
         df4453 |          0  (omitted)
          _cons |  -.0390907   .0361524    -1.08   0.280    -.1099506    .0317692
-------------+------------------------------------------------------------
```

F-test:

```
. testparm df*

 ( 1)  df1 = 0
 ( 2)  df2 = 0
 ( 3)  df3 = 0
 ( 4)  df4 = 0
 ( 5)  df5 = 0
. . . .
 (4452)  df4452 = 0

       F(4452, 33536) =   22.42
            Prob > F =    0.0000
```

Within estimator (also referred to as fixed effect estimator)

Estimation method 3: within estimator

Aim: to introduce the within estimator (or fixed-effect estimator)

- It produces identical estimates for β as the LSDV method of above (see the example 1 below)
- It requires the following steps:

  Model: $y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it} \quad i = 1,...,N; t = 1,...,T$ (2)
- First, average equation (2) over t=1,…,T to get the following cross-sectional equation:
- $\bar{y}_i = \underbrace{\beta_1 \bar{x}_{1i} + ... + \beta_k \bar{x}_{ki}}_{k \text{ explanatory variables}} + a_i + \bar{u}_i \qquad i = 1,...,N$ (3)

  where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$; $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$; $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$

  Furthermore, we include a_i in (3), since it is equal to the average of a_i: $\bar{a}_i = \frac{1}{T}\sum_{t=1}^{T} a_i = \frac{1}{T} \cdot Ta_i = a_i$
- Subtracting equation (3) from equation (2) yields:

  $\ddot{y}_{it} = \underbrace{\beta_1 \ddot{x}_{1it} + ... + \beta_k \ddot{x}_{kit}}_{k \text{ explanatory variables}} + \ddot{u}_{it} \qquad i = 1,...,N; t = 1,...,T$ (4)

where

$\ddot{y}_{it} = y_{it} - \bar{y}_i$; $\ddot{x}_{it} = x_{it} - \bar{x}_i$; $\ddot{u}_{it} = u_{it} - \bar{u}_i$

- Time demeaning removes the individual effect a_i.
- The within estimator $\hat{\beta}_{within}$ can be obtained by applying OLS on equation (4).
- Estimate of the individual effect:
  - $\hat{a}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{1i} - ... - \hat{\beta}_k \bar{x}_{ki}$
- If the following assumptions about u_{it} hold:
  - $E(u_{it}) = 0$; $Var(u_{it}) = \sigma^2$ (expected value of zero; constant variance)
  - u_{it} is independent over time and across individuals (i.i.d.)

Then it is possible to show that:

- $\hat{u}_{it} = \ddot{y}_{it} - \hat{\beta}_1 \ddot{x}_{1it} - ... - \hat{\beta}_k \ddot{x}_{kit}$
- $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2}{N(T-1)-k}$
- Consistent estimates of $Var(\hat{\beta}_{within})$ (using $\hat{\sigma}^2$)

- The within R^2 can be obtained by applying OLS on equation (4).
- Note: $\hat{\sigma}^2$ has a slightly different denominator than what would be expected when running an OLS regression on equation (4). $N(T-1)-k$ instead of $NT-k$ (as used by OLS)
- Suppose that there is autocorrelation and heteroskedasticity in u_{it}. In that case, $Var(\hat{\beta}_{within})$ is incorrect, because $\hat{\sigma}^2$ is incorrect: robust Newey-West standard errors are needed.
  - Note that it is a different estimator for $Var(\hat{\beta}_{within})$ than the robust standard error (that only corrects for heteroskedasticity)
  - Stata: cluster option: xtreg y x, fe cluster(id)
  - The cluster option corrects for both autocorrelation and heteroskedasticity.

Within-estimation procedure: further remarks

Aim: to discuss issues about the within estimator.

$$y_{it} = \underbrace{\beta_1 x_{1it} + \ldots + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it}$$

$$\ddot{y}_{it} = \underbrace{\beta_1 \ddot{x}_{1it} + \ldots + \beta_k \ddot{x}_{kit}}_{k \text{ explanatory variables}} + \ddot{u}_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T \qquad (4)$$

where

$$\ddot{y}_{it} = y_{it} - \overline{y}_i;\ \ddot{x}_{it} = x_{it} - \overline{x}_i;\ \ddot{u}_{it} = u_{it} - \overline{u}_i$$

- The parameter vector β is identified ('can be estimated') due to time-variation in x_{it} for each individual: the variables differ across time at the level of the individual.
- Individual-specific variables that are constant over time (e.g. gender, year of birth) cannot be included in (4). Their parameters are not identified, as they are incorporated in the individual effect a_i.
- Estimator $\hat{\beta}_{within}$ and $\hat{a}_i$ are consistent if T and N are large.
- If T is small and N is large, then $\hat{\beta}_{within}$ is still consistent. $\hat{a}_i$ is inconsistent because of the small number of observations (T).
- Why is $\hat{\beta}_{within}$ a consistent estimator? Let's assume for simplicity we have a bivariate model: $y_{it} = x_{it}\beta + a_i + u_{it}$ or

  $$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \qquad (5)$$

  - Consistency requires that the error term is uncorrelated with the explanatory variable: In (5):
    $Corr(\ddot{x}_{it}, \ddot{u}_{it}) = Corr(x_{it} - \overline{x}_i, u_{it} - \overline{u}_i) = 0$
  - It means that u_{it} is uncorrelated with $\overline{x}_i$
  - It means that u_{it} is uncorrelated with $x_{i1},\ldots.,x_{iT}$
    - Uncorrelated with x in the past, present and future….

  Conclusion: strict exogeneity is needed (this excludes lagged dependent variables and feedback effects).

- Contemporaneous exogeneity is too weak to prove consistency of the fixed effects estimator $\hat{\beta}_{within}$ because it does not exclude correlation between u_{it} and $x_{i1},....,x_{iT}$.

Some useful Stata commands (II)

Fixed effects (do not forget the subcommand fe which says Stata should apply the FE estimator:

- xtreg y x, fe
- xtreg y x, fe cluster(id_firm)

Example 1 (continued): fixed effects

```
. xtreg leverage i.year i.ncountry cashflow_assets net_income_MV capex_MV volatility
gdp_growth inflation, fe
note: 2.ncountry omitted because of collinearity
note: 3.ncountry omitted because of collinearity
note: 4.ncountry omitted because of collinearity
note: 5.ncountry omitted because of collinearity
note: 6.ncountry omitted because of collinearity
note: 7.ncountry omitted because of collinearity
note: 8.ncountry omitted because of collinearity
note: 9.ncountry omitted because of collinearity
note: 10.ncountry omitted because of collinearity
note: 11.ncountry omitted because of collinearity
note: 12.ncountry omitted because of collinearity
note: 13.ncountry omitted because of collinearity
note: 14.ncountry omitted because of collinearity
note: 15.ncountry omitted because of collinearity
note: 16.ncountry omitted because of collinearity
note: 17.ncountry omitted because of collinearity
note: 18.ncountry omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =     38018
Group variable: id_firm                         Number of groups   =      4453

R-sq:  within  = 0.2635                         Obs per group: min =         1
       between = 0.1852                                        avg =       8.5
       overall = 0.2094                                        max =        24

                                                F(29,33536)        =    413.74
corr(u_i, Xb)  = 0.1416                         Prob > F           =    0.0000

------------------------------------------------------------------------------
    leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
       1991  |  -.0071953   .0128123    -0.56   0.574    -.0323079    .0179173
       1992  |   .0063964    .012046     0.53   0.595    -.0172141     .030007
       1993  |  -.0500511   .0120103    -4.17   0.000    -.0735917   -.0265104
       1994  |  -.0187898   .0118235    -1.59   0.112    -.0419643    .0043847
       1995  |  -.0041872   .0115466    -0.36   0.717     -.026819    .0184446
       1996  |  -.0273106   .0114664    -2.38   0.017    -.0497853   -.004836
       1997  |  -.0383464    .011436    -3.35   0.001    -.0607614   -.0159314
       1998  |  -.0394148   .0115158    -3.42   0.001    -.0619861   -.0168435
       1999  |  -.0286589   .0115514    -2.48   0.013       -.0513   -.0060179
       2000  |  -.0048752   .0112476    -0.43   0.665    -.0269209    .0171704
       2001  |  -.0107868   .0111608    -0.97   0.334    -.0326624    .0110889
       2002  |   .0005191   .0112197     0.05   0.963    -.0214719    .0225102
       2003  |    -.02291   .0112636    -2.03   0.042     -.044987    -.000833
       2004  |  -.0283002   .0112057    -2.53   0.012    -.0502636   -.0063367
       2005  |  -.0476214   .0111731    -4.26   0.000    -.0695211   -.0257217
       2006  |  -.0417042   .0111741    -3.73   0.000    -.0636059   -.0198025
       2007  |  -.0310703   .0111437    -2.79   0.005    -.0529123   -.0092282
       2008  |   .0297393   .0110511     2.69   0.007     .0080788    .0513997
       2009  |  -.0450109   .0120865    -3.72   0.000    -.0687009    -.021321
       2010  |   .0212174   .0112624     1.88   0.060    -.0008573     .043292
       2011  |   .0361837   .0111027     3.26   0.001      .014422    .0579454
       2012  |    .007615   .0112974     0.67   0.500    -.0145283    .0297583
       2013  |  -.0082831   .0115645    -0.72   0.474    -.0309498    .0143837
             |
    ncountry |
          2  |          0  (omitted)
          3  |          0  (omitted)
          4  |          0  (omitted)
          5  |          0  (omitted)
```

```
            6  |          0  (omitted)
            7  |          0  (omitted)
            8  |          0  (omitted)
            9  |          0  (omitted)
           10  |          0  (omitted)
           11  |          0  (omitted)
           12  |          0  (omitted)
           13  |          0  (omitted)
           14  |          0  (omitted)
           15  |          0  (omitted)
           16  |          0  (omitted)
           17  |          0  (omitted)
           18  |          0  (omitted)
               |
cashflow_assets |  -.3282559   .0119771   -27.41   0.000    -.3517314   -.3047804
  net_income_MV |  -.0717367   .0022064   -32.51   0.000    -.0760613   -.0674121
       capex_MV |    .281408   .0048738    57.74   0.000     .2718552    .2909609
     volatility |   .0016269   .0001371    11.86   0.000     .0013581    .0018957
     gdp_growth |  -.0095725   .0004607   -20.78   0.000    -.0104755   -.0086695
      inflation |  -.0027681   .0009939    -2.78   0.005    -.0047163     -.00082
          _cons |   .3082911   .0121223    25.43   0.000     .2845309    .3320513
----------------+----------------------------------------------------------------
        sigma_u |  .23427663
        sigma_e |  .12571004
            rho |  .77644166   (fraction of variance due to u_i)
---------------------------------------------------------------------------------
F test that all u_i=0:     F(4452, 33536) =    22.42         Prob > F = 0.0000

---------------------------------------------------------------------------------
```

Fixed effects versus first differences

Fixed effects versus first differences (I)

Similarities and differences

What do the estimators have in common?

- They allow for correlation between a_i and the explanatory variables $x_{i1},...x_{iT}$: $E(a_i \mid x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}) \neq 0$
- The assumption of strict exogeneity (which means that the regression equation contains no feedback mechanism; there is no lag of dependent variable)

Consequence:

- The parameter estimates of both estimators should be about the same (if the assumption of strict exogeneity is true).

How do they differ?

- Assumption about the error term *u*:
    - Fixed effects estimator: u_{it} is independent over time and across individuals
- First difference estimator (see week 4): $u_{it} = u_{it-1} + e_{it}$
    - e_{it} is independent over time and across individuals
    - Thus Δu_{it} is independent over time and across individuals
    - Thus it is assumed that u_{it} follows a random walk, i.e.:

      $u_{it} = u_{it-1} + e_{it}$

      $E(e_{it}) = 0$; $Var(e_{it}) = \sigma_e^2$ (expected value of zero; constant variance)

Consequence:

- Fixed-effect estimator gives smaller standard errors if the specification is correct and there is strict exogeneity. This estimator is more efficient.

- First-difference estimator is preferred if there is a unit root in the error terms: $u_{it} = u_{it-1} + e_{it}$

Issue: First differences or fixed effects? (II)

Aim: to discuss whether to apply first differences or fixed effects

- If the regression equation is correctly specified,
  $$y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it}$$
  the within estimation procedure and the 'first-difference estimation procedure should yield similar estimates for the parameter vector β.

- Question: which of the two estimation procedures is preferable? Answer: It depends on the time series behaviour of u_{it}. If it is a white noise error term, use the within estimation procedure $\hat{\beta}_{within}$. If it follows a random walk ($u_{it} = u_{it-1} + e_{it}$), use the first-difference procedure $\hat{\beta}_{fdif}$.

Issue: First differences or fixed effects? (III)

Motivation for the procedure

Assumption: strict exogeneity (no feedback, no lagged dependent variables).

For ease of exposition here we take one explanatory variable *x*

Fixed-effects estimator: u_{it} is identically and independently distributed.
First-difference estimator: Δu_{it} is identically and independently distributed. In other words, u_{it} has a unit root.

As will be shown and derived in exercise 14.1, the correlation between the error terms of a first-difference estimator is as follows.

Assumption: u_{it} is identically and independently distributed:

$$Cov(u_{it}, u_{is}) = 0 \quad \text{for } t \neq s \quad \text{(same individual } i\text{)}$$
$$Cov(u_{it}, u_{is}) = \sigma_u^2 \quad \text{for } t = s$$

Next, we estimate the model:

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it} \qquad\qquad i = 1,...,N; t = 2,...,T$$

The correlation between Δu_{it} and Δu_{it-1} is:

$$Corr(\Delta u_{it}, \Delta u_{it-1}) = Corr(u_{it} - u_{it-1}, u_{it-1} - u_{it-2}) = -0.5.$$

Hence, the intertemporal correlation of the FD-estimator is -0.5 if u_{it} is i.i.d.

Motivation:

Covariance for the same individual across time:

$$Cov[aX+bY,cW+dZ]=acCov[X,W]+$$
$$adCov[X,Z]+bcCov[Y,W]+bdCov[Y,Z]$$

$$Cov(u_{it}-u_{it-1},u_{it-1}-u_{it-2})=$$
$$=\underbrace{Cov(u_{it},u_{it-1})}_{=0}+\underbrace{Cov(u_{it},-u_{it-2})}_{=0}+\underbrace{Cov(-u_{it-1},u_{it-1})}_{\neq 0}+\underbrace{Cov(-u_{it-1},-u_{it-2})}_{=0}$$
$$=0+0-Var(u_{it-1})+0$$
$$=-\sigma_u^2$$

$$Var(u_{it}-u_{it-1})=\underbrace{Var(u_{it})}_{=\sigma_u^2}+\underbrace{Var(-u_{it-1})}_{=\sigma_u^2}+2\underbrace{Cov(u_{it},-u_{it-1})}_{=0}$$
$$=2\sigma_u^2+0=2\sigma_u^2$$

$$Corr(\Delta u_{it},\Delta u_{it-1})=\frac{Cov(\Delta u_{it},\Delta u_{it-1})}{\sqrt{Var(\Delta u_{it})}\cdot\sqrt{Var(\Delta u_{it-1})}}=\frac{-\sigma_u^2}{2\sigma_u^2}=-0.5$$

(the derivation will be discussed in further detail at the tutorial)

Fixed effects or first differences? (IV)

- To check for FE versus FD, follow the below procedure:
  - Step 1: Run a first-difference regression equation of Δy_{it} on Δx_{it}
  - Step 2: Predict the residuals (which gives $\Delta \hat{u}_{it}$) and run a Breusch-Godfrey test for autocorrelation of $\Delta \hat{u}_{it}$ on $\Delta \hat{u}_{it-1}$ and Δx_{it}.
  - Step 3A: If the estimated coefficient on $\Delta \hat{u}_{it-1}$ is about -0.5 (i.e. -0.5 is within the 95% confidence interval) then there is an indication that u_{it} is an independent error term (u_{it} is i.i.d.). Conclusion: prefer within-estimates (fixed effects).
  - Step 3B: If the estimated coefficient on $\Delta \hat{u}_{it-1}$ is not equal to -0.5 (i.e. -0.5 is outside the 95% confidence interval) then there is an indication that u_{it} is not an independent error term. Conclusion: prefer first-differences.
- If the two procedures yield dramatically different estimates for β, the two conclusions are possible, either:
  - For some RHS-variables, the assumption of strict exogeneity does not hold.
  - The regression model is incorrectly specified. Some important time-varying regressors are missing in the equation.
- It is useful to compare the results of both regression procedures.

Example 1 (continued): first differences, autocorrelation, and fixed effects

leveragedata.dta

- Estimate the first-difference equation:

```
. reg d.leverage i.year d.cashflow_assets d.net_income_MV d.capex_MV d.volatility
d.gdp_growth d.inflation

      Source |       SS       df       MS              Number of obs =   32665
-------------+------------------------------           F( 28, 32636) =  346.35
       Model |  99.3476528    28  3.54813046           Prob > F      =  0.0000
    Residual |  334.332667 32636  .010244291           R-squared     =  0.2291
-------------+------------------------------           Adj R-squared =  0.2284
       Total |   433.68032 32664  .013277012           Root MSE      =  .10121

------------------------------------------------------------------------------
  D.leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        year |
       1992  |   .0248175   .0102758     2.42   0.016     .0046764    .0449585
       1993  |  -.0551676   .0096015    -5.75   0.000    -.0739868   -.0363484
       1994  |   .0062496   .0096471     0.65   0.517    -.0126591    .0251583
       1995  |   .0139463   .0093621     1.49   0.136    -.0044038    .0322964
       1996  |  -.0269879   .0090235    -2.99   0.003    -.0446742   -.0093015
       1997  |  -.0234538   .0089761    -2.61   0.009    -.0410472   -.0058603
       1998  |  -.0061028   .0088369    -0.69   0.490    -.0234234    .0112178
       1999  |   .0087492   .0088319     0.99   0.322    -.0085615      .02606
       2000  |   .0150779   .0089385     1.69   0.092     -.002442    .0325978
       2001  |   .0046464   .0087413     0.53   0.595    -.0124869    .0217797
       2002  |   .0115113   .0087246     1.32   0.187    -.0055892    .0286118
       2003  |   -.031665   .0086835    -3.65   0.000     -.048685    -.014645
       2004  |  -.0240877    .008755    -2.75   0.006    -.0412479   -.0069276
       2005  |  -.0259071   .0086258    -3.00   0.003     -.042814   -.0090003
       2006  |  -.0125877   .0087022    -1.45   0.148    -.0296443     .004469
       2007  |   .0048807   .0086034     0.57   0.571    -.0119823    .0217438
       2008  |   .0802273   .0086554     9.27   0.000     .0632625    .0971922
       2009  |  -.0450363   .0089832    -5.01   0.000    -.0626437    -.027429
       2010  |   .0138347   .0092961     1.49   0.137     -.004386    .0320553
       2011  |   .0132293   .0086416     1.53   0.126    -.0037086    .0301672
       2012  |  -.0247699   .0085823    -2.89   0.004    -.0415916   -.0079482
       2013  |  -.0260382   .0086973    -2.99   0.003    -.0430852   -.0089913
             |
cashflow_assets |
         D1. |  -.1576294   .0086324   -18.26   0.000    -.1745493   -.1407096
             |
 net_income_MV |
         D1. |  -.0456648   .0015237   -29.97   0.000    -.0486512   -.0426784
             |
    capex_MV |
         D1. |   .1665012   .0034578    48.15   0.000     .1597238    .1732786
             |
  volatility |
         D1. |   .0005507   .0002147     2.57   0.010       .00013    .0009715
             |
  gdp_growth |
         D1. |  -.0043542   .0003905   -11.15   0.000    -.0051196   -.0035888
             |
   inflation |
         D1. |   .0029569   .0007924     3.73   0.000     .0014037    .0045101
             |
       _cons |   .0072149   .0083321     0.87   0.387    -.0091162     .023546
------------------------------------------------------------------------------
```

. predict uhat, resid
(5353 missing values generated)

- Breusch-Godfrey after first-difference estimate:

```
. reg uhat l.uhat i.year d.cashflow_assets d.net_income_MV d.capex_MV d.volatility
d.gdp_growth d.inflation

      Source |       SS       df       MS              Number of obs =   28102
-------------+------------------------------           F( 28, 28073) =   10.02
       Model |  2.74238351    28  .097942268           Prob > F      =  0.0000
    Residual |  274.433137 28073  .009775697           R-squared     =  0.0099
-------------+------------------------------           Adj R-squared =  0.0089
       Total |  277.175521 28101  .009863547           Root MSE      =  .09887

------------------------------------------------------------------------------
           uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           uhat |
            L1. | -.0972879   .0059722   -16.29   0.000     -.1089937    -.085582
                |
           year |
           1993 |   .0118092   .0101402     1.16   0.244     -.008066    .0316845
           1994 |    .014257    .009538     1.49   0.135     -.004438    .0329519
           1995 |     .01849   .0092818     1.99   0.046     .0002973    .0366827
           1996 |   .0139628   .0091525     1.53   0.127    -.0039766    .0319021
           1997 |   .0135198    .008873     1.52   0.128    -.0038717    .0309113
           1998 |   .0115234   .0087537     1.32   0.188    -.0056341     .028681
           1999 |   .0165504   .0086718     1.91   0.056    -.0004467    .0335475
           2000 |   .0143471   .0088124     1.63   0.104    -.0029257    .0316198
           2001 |   .0150142   .0087248     1.72   0.085    -.0020869    .0321152
           2002 |   .0162071   .0086254     1.88   0.060    -.0006991    .0331133
           2003 |    .018279   .0085907     2.13   0.033     .0014408    .0351172
           2004 |   .0163362   .0085818     1.90   0.057    -.0004847     .033157
           2005 |   .0155979   .0085044     1.83   0.067    -.0010712     .032267
           2006 |    .016494   .0085168     1.94   0.053    -.0001993    .0331874
           2007 |   .0165834   .0084451     1.96   0.050     .0000306    .0331361
           2008 |   .0137903   .0085929     1.60   0.109    -.0030522    .0306327
           2009 |   .0178115   .0089722     1.99   0.047     .0002257    .0353974
           2010 |   .0149909   .0090753     1.65   0.099    -.0027972     .032779
           2011 |   .0157722   .0085215     1.85   0.064    -.0009304    .0324748
           2012 |   .0160332   .0084817     1.89   0.059    -.0005914    .0326579
           2013 |    .017986    .008481     2.12   0.034     .0013629    .0346091
                |
cashflow_assets |
            D1. |  -.0198735     .00956    -2.08   0.038    -.0386116   -.0011355
                |
  net_income_MV |
            D1. |   .0013649   .0016652     0.82   0.412     -.001899    .0046288
                |
       capex_MV |
            D1. |   .0024205   .0037537     0.64   0.519    -.0049369    .0097779
                |
     volatility |
            D1. |   .0005136    .000237     2.17   0.030     .0000492    .0009781
                |
     gdp_growth |
            D1. |   .0001906   .0004156     0.46   0.647    -.0006241    .0010053
                |
      inflation |
            D1. |   .0003954   .0008525     0.46   0.643    -.0012754    .0020662
                |
          _cons |  -.0162629   .0081466    -2.00   0.046    -.0322306   -.0002951
------------------------------------------------------------------------------
```

- Conclusion: -0.5 is not included in the 95% confidence interval of the parameter on l.uhat
- Implication: we prefer the first-difference estimator.
- We re-estimate the equation with clustered standard errors (which also corrects for heteroskedasticity).

```
. reg d.leverage i.year d.cashflow_assets d.net_income_MV d.capex_MV d.volatility
d.gdp_growth d.inflation, cluster(id_firm)

Linear regression                                      Number of obs =   32665
                                                       F( 28,  4033) =  151.44
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2291
                                                       Root MSE      =  .10121

                                (Std. Err. adjusted for 4034 clusters in id_firm)
---------------------------------------------------------------------------------
                |               Robust
     D.leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+----------------------------------------------------------------
           year |
          1992  |   .0248175    .009056     2.74   0.006     .0070627    .0425722
          1993  |  -.0551676   .0079563    -6.93   0.000    -.0707664   -.0395688
          1994  |   .0062496   .0080243     0.78   0.436    -.0094825    .0219816
          1995  |   .0139463    .007127     1.96   0.050    -.0000266    .0279192
          1996  |  -.0269879   .0072956    -3.70   0.000    -.0412912   -.0126845
          1997  |  -.0234538   .0073051    -3.21   0.001    -.0377758   -.0091317
          1998  |  -.0061028   .0073293    -0.83   0.405    -.0204722    .0082666
          1999  |   .0087492   .0071142     1.23   0.219    -.0051985    .0226969
          2000  |   .0150779   .0073423     2.05   0.040     .0006829    .0294729
          2001  |   .0046464   .0071457     0.65   0.516    -.0093631    .0186559
          2002  |   .0115113   .0070606     1.63   0.103    -.0023313    .0253539
          2003  |   -.031665   .0069352    -4.57   0.000    -.0452619   -.0180681
          2004  |  -.0240877   .0070915    -3.40   0.001     -.037991   -.0101845
          2005  |  -.0259071   .0069892    -3.71   0.000    -.0396098   -.0122045
          2006  |  -.0125877   .0070929    -1.77   0.076    -.0264936    .0013183
          2007  |   .0048807   .0068994     0.71   0.479    -.0086459    .0184073
          2008  |   .0802273    .007301    10.99   0.000     .0659133    .0945414
          2009  |  -.0450363   .0075672    -5.95   0.000    -.0598722   -.0302004
          2010  |   .0138347   .0078845     1.75   0.079    -.0016232    .0292925
          2011  |   .0132293   .0069604     1.90   0.057    -.0004168    .0268755
          2012  |  -.0247699    .006876    -3.60   0.000    -.0382506   -.0112892
          2013  |  -.0260382   .0070397    -3.70   0.000      -.03984   -.0122365
                |
cashflow_assets |
            D1. |  -.1576294   .0134112   -11.75   0.000    -.1839227   -.1313362
                |
  net_income_MV |
            D1. |  -.0456648   .0030126   -15.16   0.000    -.0515712   -.0397584
                |
       capex_MV |
            D1. |   .1665012   .0064726    25.72   0.000     .1538114     .179191
                |
     volatility |
            D1. |   .0005507    .000276     2.00   0.046     9.54e-06    .0010919
                |
     gdp_growth |
            D1. |  -.0043542   .0004465    -9.75   0.000    -.0052295   -.0034789
                |
      inflation |
            D1. |   .0029569   .0009011     3.28   0.001     .0011903    .0047236
                |
          _cons |   .0072149   .0065703     1.10   0.272    -.0056664    .0200963
---------------------------------------------------------------------------------
```

Pooled OLS (again!)

Estimators for a zero correlation between a_i and the explanatory variables

- Now we assume that there is no correlation between a_i and x_{it}.
- Consider the following methods:
    - Pooled OLS (method 5; previous week; next example)
    - Random effects (method 4; this week)

Example 1: Pooled OLS and autocorrelation

Pooled OLS (with clustered standard errors):

```
. reg leverage i.year i.ncountry cashflow_assets net_income_MV capex_MV volatility
gdp_growth inflation
      Source |       SS       df       MS              Number of obs =   38018
-------------+------------------------------           F( 46, 37971) =  331.47
       Model |  812.032429    46  17.6528789           Prob > F      =  0.0000
    Residual |  2024.17412 37971  .053255751           R-squared     =  0.2865
-------------+------------------------------           Adj R-squared =  0.2856
       Total |  2834.20655 38017  .074551031           Root MSE      =  .23077
--------------------------------------------------------------------------------
       leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
           year |
          1991  |   .0029504   .0232836     0.13   0.899    -.0426861    .0485868
          1992  |   .0000561   .0217867     0.00   0.998    -.0426463    .0427586
          1993  |   -.036609   .0216375    -1.69   0.091    -.0790191    .0058011
          1994  |   .0008144   .0213076     0.04   0.970     -.040949    .0425779
          1995  |   .0095802   .0206564     0.46   0.643    -.0309068    .0500672
          1996  |  -.0102466   .0204811    -0.50   0.617    -.0503902     .029897
          1997  |  -.0179961   .0204068    -0.88   0.378     -.057994    .0220018
          1998  |   -.017542   .0205455    -0.85   0.393    -.0578118    .0227278
          1999  |  -.0105355   .0206114    -0.51   0.609    -.0509343    .0298633
          2000  |   .0044244   .0200754     0.22   0.826     -.034924    .0437728
          2001  |   .0000975   .0199214     0.00   0.996     -.038949    .0391441
          2002  |   .0149194    .019984     0.75   0.455    -.0242498    .0540887
          2003  |  -.0042183    .020029    -0.21   0.833    -.0434757    .0350391
          2004  |  -.0159808    .019941    -0.80   0.423    -.0550657    .0231042
          2005  |  -.0366836   .0198668    -1.85   0.065    -.0756231    .0022558
          2006  |  -.0384295   .0198872    -1.93   0.053    -.0774089    .0005499
          2007  |  -.0290793   .0198382    -1.47   0.143    -.0679627     .009804
          2008  |   .0201234   .0196752     1.02   0.306    -.0184406    .0586874
          2009  |  -.0335673   .0212667    -1.58   0.114    -.0752506     .008116
          2010  |   .0173953   .0199924     0.87   0.384    -.0217904    .0565811
          2011  |   .0278749   .0197154     1.41   0.157    -.0107678    .0665176
          2012  |   .0049275   .0200135     0.25   0.806    -.0342996    .0441545
          2013  |   -.006493   .0204539    -0.32   0.751    -.0465832    .0335972
       ncountry |
             2  |  -.0445892   .0081852    -5.45   0.000    -.0606325   -.0285459
             3  |   .0603322   .0212166     2.84   0.004      .018747    .1019173
             4  |   -.048074   .0287972    -1.67   0.095    -.1045173    .0083692
             5  |  -.0455876   .0081123    -5.62   0.000    -.0614879   -.0296872
             6  |  -.0490725   .0066374    -7.39   0.000     -.062082   -.0360631
             7  |  -.0895072   .0066424   -13.48   0.000    -.1025264    -.076488
             8  |   .0499154    .008389     5.95   0.000     .0334726    .0663581
             9  |  -.0513545   .0100213    -5.12   0.000    -.0709965   -.0317125
            10  |    .057862   .0073695     7.85   0.000     .0434175    .0723064
            11  |  -.0569571    .021772    -2.62   0.009    -.0996307   -.0142834
            12  |  -.0598471   .0133481    -4.48   0.000    -.0860097   -.0336845
            13  |  -.0964434   .0302854    -3.18   0.001    -.1558035   -.0370832
            14  |  -.0768503   .0076046   -10.11   0.000    -.0917555   -.0619452
            15  |   .1175568   .0102683    11.45   0.000     .0974307    .1376829
            16  |   .0016017   .0240471     0.07   0.947    -.0455312    .0487347
            17  |   .0071014   .0145235     0.49   0.625    -.0213651     .035568
            18  |   .0154778   .0079679     1.94   0.052    -.0001395    .0310952
                |
cashflow_assets |  -.7094728   .0148948   -47.63   0.000     -.738667   -.6802787
  net_income_MV |  -.0686996   .0033831   -20.31   0.000    -.0753304   -.0620687
       capex_MV |   .5044541   .0064815    77.83   0.000     .4917501     .517158
     volatility |  -.0010765   .0001137    -9.47   0.000    -.0012994   -.0008536
     gdp_growth |  -.0067869   .0007704    -8.81   0.000    -.0082969    -.005277
      inflation |  -.0046364   .0017127    -2.71   0.007    -.0079934   -.0012794
          _cons |   .4091819   .0212779    19.23   0.000     .3674766    .4508872

  --------------------------------------------------------------------------------.
```

. predict uhat, resid

. reg uhat l.uhat i.year i.ncountry cashflow_assets net_income_MV capex_MV volatility gdp_growth inflation

Source	SS	df	MS
Model	1236.9336	46	26.8898608
Residual	446.475075	32618	.013687997
Total	1683.40867	32664	.051537126

Number of obs =	32665
F(46, 32618) =	1964.48
Prob > F =	0.0000
R-squared =	0.7348
Adj R-squared =	0.7344
Root MSE =	.117

uhat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
uhat						
L1.	.8681897	.0028916	300.25	0.000	.8625221	.8738573
year						
1992	.0054429	.0118535	0.46	0.646	-.0177903	.0286761
1993	-.0104477	.011141	-0.94	0.348	-.0322845	.0113891
1994	-.0307748	.011023	-2.79	0.005	-.0523803	-.0091693
1995	-.0178197	.0109172	-1.63	0.103	-.0392179	.0035786
1996	-.0173844	.0105893	-1.64	0.101	-.0381399	.0033711
1997	-.018507	.0105939	-1.75	0.081	-.0392715	.0022575
1998	-.0214503	.0105839	-2.03	0.043	-.0421951	-.0007055
1999	-.0178809	.010631	-1.68	0.093	-.038718	.0029562
2000	-.0067058	.0104137	-0.64	0.520	-.0271169	.0137054
2001	-.0081796	.0102159	-0.80	0.423	-.0282031	.0118439
2002	-.0142949	.0102462	-1.40	0.163	-.0343778	.005788
2003	-.0093085	.0102263	-0.91	0.363	-.0293523	.0107353
2004	-.0121306	.0102212	-1.19	0.235	-.0321646	.0079033
2005	-.0203213	.0101848	-2.00	0.046	-.0402839	-.0003586
2006	-.025473	.0102126	-2.49	0.013	-.0454901	-.0054559
2007	-.0238148	.0101682	-2.34	0.019	-.0437449	-.0038847
2008	-.0077731	.0099828	-0.78	0.436	-.0273397	.0117935
2009	-.0107624	.0107814	-1.00	0.318	-.0318944	.0103697
2010	-.0164746	.0102141	-1.61	0.107	-.0364947	.0035454
2011	-.0148785	.010036	-1.48	0.138	-.0345494	.0047924
2012	-.0182539	.0101256	-1.80	0.071	-.0381004	.0015927
2013	-.023887	.0103999	-2.30	0.022	-.0442711	-.0035028
ncountry						
2	-.0128124	.0044356	-2.89	0.004	-.0215064	-.0041184
3	-.0025386	.0162013	-0.16	0.875	-.0342937	.0292165
4	-.0207784	.0167837	-1.24	0.216	-.0536751	.0121182
5	-.0164043	.0043752	-3.75	0.000	-.0249799	-.0078287
6	-.0113643	.0036	-3.16	0.002	-.0184204	-.0043082
7	-.0112784	.0036052	-3.13	0.002	-.0183448	-.004212
8	.009326	.0046566	2.00	0.045	.0001989	.0184531
9	-.0173592	.0054397	-3.19	0.001	-.0280212	-.0066972
10	-.002379	.0039989	-0.59	0.552	-.0102169	.005459
11	.0116422	.0130029	0.90	0.371	-.013844	.0371283
12	-.0147346	.0074554	-1.98	0.048	-.0293475	-.0001217
13	-.0207819	.0178103	-1.17	0.243	-.0556907	.0141268
14	-.0130139	.0041147	-3.16	0.002	-.0210788	-.0049489
15	.0091127	.0055611	1.64	0.101	-.0017872	.0200127
16	-.0132585	.0137125	-0.97	0.334	-.0401356	.0136186
17	.0027798	.0081756	0.34	0.734	-.0132446	.0188043
18	-.0080106	.004312	-1.86	0.063	-.0164622	.000441
cashflow_assets	.101437	.008489	11.95	0.000	.0847983	.1180758
net_income_MV	.0134957	.0019354	6.97	0.000	.0097023	.0172891
capex_MV	-.0964637	.0037497	-25.73	0.000	-.1038133	-.0891141
volatility	.0002336	.0000651	3.59	0.000	.000106	.0003611
gdp_growth	.0014666	.0004292	3.42	0.001	.0006253	.0023079
inflation	-.0001985	.0009722	-0.20	0.838	-.0021041	.001707
_cons	.0233256	.0109373	2.13	0.033	.001888	.0447631

Conclusion: there is autocorrelation
So: compute clustered standard errors:

```
. reg leverage i.year i.ncountry cashflow_assets net_income_MV capex_MV volatility
gdp_growth inflation, cluster(id_firm)
Linear regression                                      Number of obs =   38018
                                                       F( 46,  4452) =  125.91
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2865
                                                       Root MSE      =  .23077
                               (Std. Err. adjusted for 4453 clusters in id_firm)
```

leverage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
year						
1991	.0029504	.0138055	0.21	0.831	-.0241154	.0300161
1992	.0000561	.0150799	0.00	0.997	-.029508	.0296203
1993	-.036609	.0159728	-2.29	0.022	-.0679236	-.0052944
1994	.0008144	.0161321	0.05	0.960	-.0308125	.0324414
1995	.0095802	.0166859	0.57	0.566	-.0231326	.0422929
1996	-.0102466	.0167745	-0.61	0.541	-.043133	.0226397
1997	-.0179961	.0167746	-1.07	0.283	-.0508826	.0148904
1998	-.017542	.0171381	-1.02	0.306	-.0511413	.0160573
1999	-.0105355	.0171958	-0.61	0.540	-.0442477	.0231767
2000	.0044244	.0166613	0.27	0.791	-.0282401	.0370889
2001	.0000975	.0168222	0.01	0.995	-.0328823	.0330774
2002	.0149194	.0170239	0.88	0.381	-.0184558	.0482947
2003	-.0042183	.0171922	-0.25	0.806	-.0379236	.0294869
2004	-.0159808	.016918	-0.94	0.345	-.0491486	.017187
2005	-.0366836	.0168973	-2.17	0.030	-.0698107	-.0035565
2006	-.0384295	.016741	-2.30	0.022	-.0712502	-.0056088
2007	-.0290793	.0167073	-1.74	0.082	-.0618339	.0036753
2008	.0201234	.0168652	1.19	0.233	-.0129407	.0531875
2009	-.0335673	.0197538	-1.70	0.089	-.0722946	.00516
2010	.0173953	.0171219	1.02	0.310	-.0161721	.0509628
2011	.0278749	.016915	1.65	0.099	-.0052869	.0610368
2012	.0049275	.0174995	0.28	0.778	-.0293802	.0392351
2013	-.006493	.0179112	-0.36	0.717	-.0416078	.0286219
ncountry						
2	-.0445892	.0237515	-1.88	0.061	-.0911539	.0019755
3	.0603322	.0456017	1.32	0.186	-.0290698	.1497342
4	-.048074	.0451443	-1.06	0.287	-.1365792	.0404312
5	-.0455876	.0225627	-2.02	0.043	-.0898217	-.0013534
6	-.0490725	.0203556	-2.41	0.016	-.0889795	-.0091655
7	-.0895072	.0205905	-4.35	0.000	-.1298748	-.0491396
8	.0499154	.02224	2.24	0.025	.0063138	.0935169
9	-.0513545	.0314302	-1.63	0.102	-.1129733	.0102642
10	.057862	.0231333	2.50	0.012	.0125092	.1032147
11	-.0569571	.0534926	-1.06	0.287	-.1618292	.0479151
12	-.0598471	.0417232	-1.43	0.152	-.1416454	.0219511
13	-.0964434	.0490531	-1.97	0.049	-.1926119	-.0002749
14	-.0768503	.0221101	-3.48	0.001	-.1201972	-.0335035
15	.1175568	.0308953	3.81	0.000	.0569867	.1781269
16	.0016017	.070134	0.02	0.982	-.1358958	.1390992
17	.0071014	.0323112	0.22	0.826	-.0562445	.0704474
18	.0154778	.0243164	0.64	0.524	-.0321944	.06315
cashflow_assets	-.7094728	.0364888	-19.44	0.000	-.7810091	-.6379366
net_income_MV	-.0686996	.0056015	-12.26	0.000	-.0796812	-.0577179
capex_MV	.5044541	.0132879	37.96	0.000	.4784033	.5305049
volatility	-.0010765	.0002885	-3.73	0.000	-.001642	-.000511
gdp_growth	-.0067869	.0009723	-6.98	0.000	-.0086931	-.0048808
inflation	-.0046364	.0019291	-2.40	0.016	-.0084184	-.0008544
_cons	.4091819	.0276264	14.81	0.000	.3550203	.4633434

Random-effects estimator

Estimation method 5: random-effects estimator

Aim: to introduce the random-effects estimator.

Consider the static model:

$$y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it} \qquad (6)$$

$$i = 1,...,N; t = 1,...,T$$

- The equation (6) does not allow correlation between a_i and all of the right hand side variables $x_1,...x_k$ (over all T periods)

$$E(a_i \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$$

- The exogenous time-varying regressors are assumed to be strictly exogenous (conditional on the unobserved effect):

$$E(u_{it} \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}, a_i) = 0$$

  - In simple words: there are neither lagged dependent variables nor is there any feedback mechanism
- The following is assumed about u_{it}:
  - $E(u_{it}) = 0$; $Var(u_{it}) = \sigma_u^2$ (expected value of zero; constant variance)
  - u_{it} is independent over time and across individuals
- The following is assumed about a_i:
  - $E(a_i) = 0$; $Var(a_i) = \sigma_a^2$ (expected value of zero; constant variance)
  - a_i is independent over time and across individuals
- Equation (6) can be rewritten as follows:

Model: $$y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + v_{it} \qquad (6)$$

$$i = 1,...,N; t = 1,...,T$$

where

$$v_{it} = a_i + u_{it} \qquad (7)$$

- The random effects estimator works as follows. We estimate equation (6), in which the error structure of equation (7) is taken into account.

More on the correlation structure between the error terms

Aim: to discuss the correlation structure of the error terms.

- Error term (see slide above):
  $v_{it} = a_i + u_{it}$
- Covariance across two individuals *i* and *j*:
  - $Cov(v_{it}, v_{jt}) = 0 \qquad i \neq j$
  - $Cov(v_{it}, v_{js}) = 0 \qquad i \neq j; t \neq s$

- Covariance for the same individual across time:

$$Cov[aX + bY, cW + dZ] = acCov[X, W] + adCov[X, Z] + bcCov[Y, W] + bdCov[Y, Z]$$

- So that

$$\begin{aligned} Cov(v_{it}, v_{is}) &= Cov(a_i + u_{it}, a_i + u_{is}) \\ &= \underbrace{Cov(a_i, a_i)}_{=\sigma_a^2} + \underbrace{Cov(a_i, u_{is})}_{=0} + \underbrace{Cov(u_{it}, a_i)}_{=0} + \underbrace{Cov(u_{it}, u_{is})}_{=0} \\ &= Var(a_i) + 0 + 0 + 0 \\ &= \sigma_a^2 \end{aligned}$$

- Variance for an individual:

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

$$\begin{aligned} Var(v_{it}) = Var(a_i + u_{it}) &= \underbrace{Var(a_i)}_{\sigma_a^2} + \underbrace{Var(u_{it})}_{=\sigma_u^2} + 2\underbrace{Cov(a_i, u_{it})}_{=0} \\ &= \sigma_a^2 + \sigma_u^2 \end{aligned}$$

- Thus: $Corr(v_{it}, v_{is}) = \dfrac{Cov(v_{it}, v_{is})}{\sqrt{Var(v_{it})}\sqrt{Var(v_{is})}} = \dfrac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}$

- Conclusion: random effects estimator $\hat{\beta}_{re}$ takes account of the autocorrelation structure!

Random effects or fixed effects: Hausman test

Random effects or fixed effects

Aim: to establish when to choose fixed effects or random effects

- Model: $y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it}$ (8)

  $E(a_i \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$

  $E(u_{it} \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}, a_i) = 0$

  x_{it} is strictly exogenous (in simple words: no lagged dependent variables; no feedback mechanism).

Estimation method: Random effects

- Equation (8) can be estimated with random effects, which requires the assumption $E(a_i \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$
- $\hat{\beta}_{re}$ is a consistent estimator of β

Estimation method: Fixed effects

- Equation (8) can be estimated with fixed effects, which does NOT assume that $E(a_i \mid \underbrace{x_{1it},...,x_{1iT},...,x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$
  - $\hat{\beta}_{within}$ is a consistent estimator of β
  - $\hat{\beta}_{within}$ is less efficient than $\hat{\beta}_{re}$: It means that $Var(\hat{\beta}_{re}) < Var(\hat{\beta}_{within})$
    - Consequently, random effects yields significant *t*-statistics more easily than fixed effects.

Random effects: further remarks (II)

- Both random effects and pooled OLS allow for the inclusion of time-invariant individual variables (e.g. gender in a wage equation):

$$y_{it} = a_i + \beta_1 x_{1it} + ... + \beta_k x_{kit} + \gamma_1 z_{1i} + ... + \gamma_l z_{li} + u_{it}$$
$$i = 1,...,N; t = 1,...,T$$

- Where z_i is a vector of individual-specific regressors.
- Remember: the effect of *z* on *y* cannot be estimated ('is not identified') in the fixed-effects specification.

Fixed effects of Random effects? Hausman test
Aim: to establish when to choose fixed effects or random effects

- Model: $y_{it} = \underbrace{\beta_1 x_{1it} + ... + \beta_k x_{kit}}_{k \text{ explanatory variables}} + a_i + u_{it}$ (8)

$$E(a_i \mid \underbrace{x_{1it},..., x_{1iT},..., x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$$

$$E(u_{it} \mid \underbrace{x_{1it},..., x_{1iT},..., x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}, a_i) = 0$$

Structure:

- Null hypothesis: $H_0 : E(a_i \mid \underbrace{x_{1it},..., x_{1iT},..., x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) = 0$
- Alternative hypothesis: $H_1 : E(a_i \mid \underbrace{x_{1it},..., x_{1iT},..., x_{ki1},...x_{kiT}}_{\text{all } k \text{ explanatory variables in all } T \text{ time periods}}) \neq 0$

- Under the null hypothesis (H_0), random effects is preferred (because of the zero correlation between the a_i and the explanatory variables).
  Equation (8) can be estimated by random effects and fixed effects. **If H_0 is true**, both estimators are unbiased (consistent)
  However, random effects yields smaller standard errors, so that it is preferred to fixed effects.
  If alternative hypothesis (H_1) is true: fixed effects is preferred, because random effects estimator is biased!

Some useful Stata commands (III)

Random effects:

- xtreg y x, re
- xtreg y x, re robust

Hausman test:

- xtreg y x, fe
- est store fixed
- xtreg y x, re
- hausman fixed, force

Example 1 (continued): Pooled OLS and Random effects

Pooled OLS (with clustered standard errors):

```
. reg leverage i.year i.ncountry cashflow_assets net_income_MV capex_MV volatility
gdp_growth inflation, cluster(id_firm)

Linear regression                                      Number of obs =   38018
                                                       F( 46,  4452) =  125.91
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2865
                                                       Root MSE      =  .23077

                                (Std. Err. adjusted for 4453 clusters in id_firm)
------------------------------------------------------------------------------
                |               Robust
       leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
                |
cashflow_assets |  -.7094728   .0364888   -19.44   0.000    -.7810091   -.6379366
  net_income_MV |  -.0686996   .0056015   -12.26   0.000    -.0796812   -.0577179
       capex_MV |   .5044541   .0132879    37.96   0.000     .4784033    .5305049
     volatility |  -.0010765   .0002885    -3.73   0.000     -.001642    -.000511
     gdp_growth |  -.0067869   .0009723    -6.98   0.000    -.0086931   -.0048808
      inflation |  -.0046364   .0019291    -2.40   0.016    -.0084184   -.0008544
          _cons |   .4091819   .0276264    14.81   0.000     .3550203    .4633434
------------------------------------------------------------------------------
```

```
Random effects:
. xtreg leverage i.year cashflow_assets net_income_MV capex_MV volatility gdp_growth
inflation, re

Random-effects GLS regression                   Number of obs      =     38018
Group variable: id_firm                         Number of groups   =      4453

R-sq:  within  = 0.2629                         Obs per group: min =         1
       between = 0.2026                                        avg =       8.5
       overall = 0.2212                                        max =        24

                                                Wald chi2(29)      =  12901.32
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       leverage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
cashflow_assets |  -.3440569   .0116573   -29.51   0.000    -.3669048    -.321209
  net_income_MV |  -.0733575   .0021795   -33.66   0.000    -.0776291   -.0690858
       capex_MV |   .2973662   .0047898    62.08   0.000     .2879783    .3067541
     volatility |    .001116   .0001244     8.97   0.000     .0008721    .0013598
     gdp_growth |  -.0100331    .000456   -22.00   0.000    -.0109269   -.0091392
      inflation |  -.0012062   .0009845    -1.23   0.221    -.0031359    .0007235
          _cons |   .3032823   .0125461    24.17   0.000     .2786923    .3278723
----------------+-------------------------------------------------------------
        sigma_u |  .21045753
        sigma_e |  .12571004
            rho |  .73703448   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The Hausman test

```
. xtreg leverage i.year cashflow_assets net_income_MV capex_MV volatility gdp_growth
inflation, fe

Fixed-effects (within) regression               Number of obs      =     38018
Group variable: id_firm                         Number of groups   =      4453

R-sq:  within  = 0.2635                         Obs per group: min =         1
       between = 0.1852                                        avg =       8.5
       overall = 0.2094                                        max =        24

                                                F(29,33536)        =    413.74
corr(u_i, Xb)  = 0.1416                         Prob > F           =    0.0000

------------------------------------------------------------------------------
       leverage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------

cashflow_assets |  -.3282559   .0119771   -27.41   0.000    -.3517314   -.3047804
  net_income_MV |  -.0717367   .0022064   -32.51   0.000    -.0760613   -.0674121
       capex_MV |    .281408   .0048738    57.74   0.000     .2718552    .2909609
     volatility |   .0016269   .0001371    11.86   0.000     .0013581    .0018957
     gdp_growth |  -.0095725   .0004607   -20.78   0.000    -.0104755   -.0086695
      inflation |  -.0027681   .0009939    -2.78   0.005    -.0047163     -.00082
          _cons |   .3082911   .0121223    25.43   0.000     .2845309    .3320513
----------------+-------------------------------------------------------------
        sigma_u |  .23427663
        sigma_e |  .12571004
            rho |  .77644166   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(4452, 33536) =    22.42         Prob > F = 0.0000
```

```
. est store fixed
```

```
. xtreg leverage i.year cashflow_assets net_income_MV capex_MV volatility gdp_growth
inflation, re

Random-effects GLS regression                   Number of obs      =     38018
Group variable: id_firm                         Number of groups   =      4453

R-sq:  within  = 0.2629                         Obs per group: min =         1
       between = 0.2026                                        avg =       8.5
       overall = 0.2212                                        max =        24

                                                Wald chi2(29)      =  12901.32
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       leverage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
cashflow_assets |  -.3440569   .0116573   -29.51   0.000    -.3669048    -.321209
  net_income_MV |  -.0733575   .0021795   -33.66   0.000    -.0776291   -.0690858
       capex_MV |   .2973662   .0047898    62.08   0.000     .2879783    .3067541
     volatility |    .001116   .0001244     8.97   0.000     .0008721    .0013598
     gdp_growth |  -.0100331    .000456   -22.00   0.000    -.0109269   -.0091392
      inflation |  -.0012062   .0009845    -1.23   0.221    -.0031359    .0007235
          _cons |   .3032823   .0125461    24.17   0.000     .2786923    .3278723
------------------------------------------------------------------------------
```

```
. hausman fixed, force

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
        year |
        1991 |   -.0071953    -.0071565       -.0000388               .
        1992 |    .0063964     .0056499        .0007466               .
        1993 |   -.0500511    -.0492874       -.0007636               .
        1994 |   -.0187898    -.0147991       -.0039907               .
        1995 |   -.0041872    -.0000976       -.0040896               .
        1996 |   -.0273106    -.0220795       -.0052312               .
        1997 |   -.0383464    -.0315306       -.0068158               .
        1998 |   -.0394148     -.031532       -.0078828               .
        1999 |   -.0286589    -.0208409        -.007818               .
        2000 |   -.0048752      .000976       -.0058512               .
        2001 |   -.0107868    -.0065929       -.0041939               .
        2002 |    .0005191     .0049294       -.0044103               .
        2003 |     -.02291    -.0180144       -.0048956               .
        2004 |   -.0283002    -.0230235       -.0052766               .
        2005 |   -.0476214    -.0428228       -.0047987               .
        2006 |   -.0417042    -.0365223       -.0051819               .
        2007 |   -.0310703    -.0267037       -.0043666               .
        2008 |    .0297393     .0302732       -.0005339               .
        2009 |   -.0450109    -.0409076       -.0041034               .
        2010 |    .0212174     .0249711       -.0037537               .
        2011 |    .0361837     .0377859       -.0016022               .
        2012 |     .007615     .0089219       -.0013069               .
        2013 |   -.0082831    -.0046373       -.0036457               .
cashflow_a~s |   -.3282559    -.3440569         .015801        .0027492
net_income~V |   -.0717367    -.0733575        .0016208        .0003438
    capex_MV |     .281408     .2973662       -.0159582        .0009009
  volatility |    .0016269      .001116         .000511        .0000577
  gdp_growth |   -.0095725    -.0100331        .0004606        .0000653
   inflation |   -.0027681    -.0012062       -.0015619        .0001363
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                 chi2(29) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      798.44
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

- Conclusion: Reject Ho (thus fixed-effects specification is preferred).
- However, we prefer the first-difference specification to the fixed-effects specification (see the conclusion above with the test on -0.5).
- It means that overall, we conclude that the first-difference specification is preferred.

Line of reasoning with panel data: final example

To wind up: Line of reasoning with panel data (5 steps)

- **Step 1a:** Apply pooled OLS

$$y_{it} = \underbrace{\beta_1 x_{1it} + \ldots + \beta_k x_{kit}}_{k \text{ explanatory variables}} + v_{it}$$

- **Step 1b:** Test for autocorrelation of error term with Breusch-Godfrey. Estimated parameter on lagged v_{it} is

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

- **Step 1c:** Re-estimate pooled OLS with clustered standard errors

Step 2a: Apply first-differences estimator (FD):

$$\Delta y_{it} = \beta_1 \Delta x_{1it} + \ldots + \beta_1 \Delta x_{kit} + \Delta u_{it}$$

- **Step 2b:** Compute autocorrelation: is the parameter on Δu_{it-1} of Breusch-Godfrey equal to -0.5?
- **Step 2c:** Re-estimate FD with clustered standard errors

- **Step 3a:** Apply LSDV/ fixed-effects estimator (FE):

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{1it} + \ldots + \beta_k \ddot{x}_{kit} + \ddot{u}_{it}$$

- **Step 3b:** Compute autocorrelation
- **Step 3c:** Compare the outcome of the FD estimator with the outcome of the FE estimator

- **Step 4a:** Apply random effects estimator (RE)
- **Step 4b:** Test for autocorrelation
- **Step 4c:** Compare the outcome of the Random effects estimator and Pooled OLS estimator

- **Step 5:** Test for Random effects versus fixed effects using the Hausman test.

Example 2 (for the entire procedure)

We apply the five steps on the dataset wagepan.dta

. xtset nr year

. xtsum lwage married educ union

```
Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
lwage    overall |  1.649147    .5326094  -3.579079    4.05186 |     N =    4360
         between |              .3907468   .3333435   3.174173 |     n =     545
         within  |              .3622636  -2.467201   3.204687 |     T =       8
                 |                                             |
married  overall |  .4389908    .4963208          0          1 |     N =    4360
         between |              .3766116          0          1 |     n =     545
         within  |              .3236137  -.4360092   1.313991 |     T =       8
                 |                                             |
educ     overall |  11.76697    1.746181          3         16 |     N =    4360
         between |              1.747585          3         16 |     n =     545
         within  |                     0   11.76697   11.76697 |     T =       8
                 |                                             |
union    overall |  .2440367    .4295639          0          1 |     N =    4360
         between |              .3294467          0          1 |     n =     545
         within  |              .2759787  -.6309633   1.119037 |     T =       8
```

Step 1a: Apply pooled OLS

. reg lwage married educ union

```
      Source |       SS       df       MS              Number of obs =    4360
-------------+------------------------------           F(  3,  4356) =  203.27
       Model |  151.850323     3  50.6167742           Prob > F      =  0.0000
    Residual |  1084.67932  4356  .249008108           R-squared     =  0.1228
-------------+------------------------------           Adj R-squared =  0.1222
       Total |  1236.52964  4359  .283672779           Root MSE      =  .49901
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2090758   .0152444    13.71   0.000      .179189    .2389626
        educ |   .0760449   .0043292    17.57   0.000     .0675574    .0845324
       union |   .1710456   .0176107     9.71   0.000     .1365197    .2055716
       _cons |   .6208054    .051991    11.94   0.000     .5188766    .7227343
------------------------------------------------------------------------------
```

Step 1b: Test for autocorrelation of error term with Breusch-Godfrey.

. predict uhat, resid

. reg uhat l.uhat married educ union

```
      Source |       SS       df       MS              Number of obs =    3815
-------------+------------------------------           F(  4,  3810) =  511.61
       Model |  315.938702     4  78.9846754           Prob > F      =  0.0000
    Residual |   588.20123  3810  .154383525           R-squared     =  0.3494
-------------+------------------------------           Adj R-squared =  0.3488
       Total |  904.139931  3814  .237058189           Root MSE      =  .39292
------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |   .5733121   .0126968    45.15   0.000     .5484188    .5982053
     married |  -.0441766   .0127575    -3.46   0.001    -.0691889   -.0191644
        educ |   .0033008   .0036449     0.91   0.365    -.0038453     .010447
       union |  -.0400586   .0148631    -2.70   0.007     -.069199   -.0109182
       _cons |   .0355483   .0437596     0.81   0.417    -.0502461    .1213428
------------------------------------------------------------------------------
```

Conclusion: there is autocorrelation

- **Step 1c:** Re-estimate pooled OLS with clustered standard errors

. reg lwage married educ union, cluster(nr)

```
Linear regression                                      Number of obs =    4360
                                                       F(  3,   544) =   60.91
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1228
                                                       Root MSE      =  .49901

                                  (Std. Err. adjusted for 545 clusters in nr)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2090758   .0243617     8.58   0.000     .1612213    .2569303
        educ |   .0760449   .0088779     8.57   0.000     .0586057     .093484
       union |   .1710456   .0282567     6.05   0.000     .1155399    .2265513
       _cons |   .6208054   .1060566     5.85   0.000     .4124747    .8291361
------------------------------------------------------------------------------
```

- **Step 2a:** Apply first-differences estimator (FD):

. reg d.lwage d.married d.educ d.union, nocons

```
      Source |       SS       df       MS              Number of obs =    3815
-------------+------------------------------           F(  2,  3813) =    8.92
       Model |  3.57932445     2  1.78966222           Prob > F      =  0.0001
    Residual |  765.033676  3813  .200638258           R-squared     =  0.0047
-------------+------------------------------           Adj R-squared =  0.0041
       Total |  768.613001  3815  .201471298           Root MSE      =  .44793

------------------------------------------------------------------------------
     D.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |
         D1. |   .0820425   .0226819     3.62   0.000     .0375727    .1265123
        educ |
         D1. |  (dropped)
       union |
         D1. |   .0430192   .0198737     2.16   0.030      .004055    .0819833
------------------------------------------------------------------------------
```

- **Step 2b:** Compute autocorrelation: is the parameter on Δu_{it-1} of Breusch-Godfrey equal to -0.5?

. predict uhat, resid
(545 missing values generated)

  - Breusch-Godfrey after first-difference estimate:

. reg uhat l.uhat d.married d.educ d.union

```
      Source |       SS       df       MS              Number of obs =    3270
-------------+------------------------------           F(  3,  3266) =  242.68
       Model |  104.390207     3  34.7967355           Prob > F      =  0.0000
    Residual |  468.291419  3266  .143383778           R-squared     =  0.1823
-------------+------------------------------           Adj R-squared =  0.1815
       Total |  572.681626  3269  .175185569           Root MSE      =  .37866

------------------------------------------------------------------------------
        uhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        uhat |
         L1. |  -.3950886   .0146761   -26.92   0.000    -.4238638   -.3663134
     married |
         D1. |  -.0280538   .0214986    -1.30   0.192    -.0702059    .0140983
        educ |
         D1. |  (dropped)
       union |
         D1. |   .0155607   .0185722     0.84   0.402    -.0208536     .051975
       _cons |   .0806243   .0067844    11.88   0.000     .0673222    .0939265
------------------------------------------------------------------------------
```

  - Conclusion: -0.5 is not within interval of parameter on l.uhat
  - Thus we prefer first-difference estimator above the fixed-effects estimator

- **Step 2c:** Re-estimate FD with clustered standard errors

. reg d.lwage d.married d.educ d.union, nocons cluster(nr)

```
Linear regression                                      Number of obs =    3815
                                                       F(  2,   544) =    6.91
                                                       Prob > F      =  0.0011
                                                       R-squared     =  0.0047
                                                       Root MSE      =  .44793

                                   (Std. Err. adjusted for 545 clusters in nr)
------------------------------------------------------------------------------
             |               Robust
     D.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |
         D1. |   .0820425   .0239245     3.43   0.001     .0350467    .1290383
        educ |
         D1. |  (dropped)
       union |
         D1. |   .0430192    .022077     1.95   0.052    -.0003474    .0863857
------------------------------------------------------------------------------
```

- **Step 3a:** Apply LSDV/fixed-effects estimator (FE):

LSDV:

. reg lwage married educ union dnum*

```
      Source |       SS       df       MS              Number of obs =    4360
-------------+------------------------------           F(546,  3813) =    8.90
       Model |   692.98604   546   1.2692052           Prob > F      =  0.0000
    Residual |  543.543602  3813  .142550118           R-squared     =  0.5604
-------------+------------------------------           Adj R-squared =  0.4975
       Total |  1236.52964  4359  .283672779           Root MSE      =  .37756

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2416845   .0176735    13.67   0.000     .2070341    .2763348
        educ |   .0328052   .0145304     2.26   0.024     .0043171    .0612934
       union |   .0700438    .020724     3.38   0.001     .0294127    .1106749
. . . .
. . . .
-------------+----------------------------------------------------------------
```

F-test:

. testparm dnum*

```
       Constraint 209 dropped
       Constraint 395 dropped

       F(543,  3813) =    6.99
            Prob > F =    0.0000
```

Fixed-effects, within regression

```
. xtreg lwage married educ union, fe i(nr)
```

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

R-sq:  within  = 0.0498                         Obs per group: min =         8
       between = 0.0573                                        avg =       8.0
       overall = 0.0538                                        max =         8

                                                F(2,3813)          =    100.00
corr(u_i, Xb)  = -0.0035                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2416845   .0176735    13.67   0.000     .2070341    .2763348
        educ |  (dropped)
       union |   .0700438    .020724     3.38   0.001     .0294127    .1106749
       _cons |   1.525957    .010825   140.97   0.000     1.504733     1.54718
-------------+----------------------------------------------------------------
     sigma_u |  .37939434
     sigma_e |   .3775581
         rho |  .50242582   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3813) =     6.98           Prob > F = 0.0000
```

- **Step 4a:** Apply random effects estimator

Random effects:

. xtreg lwage married educ union, re i(nr)

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

R-sq:  within  = 0.0493                         Obs per group: min =         8
       between = 0.1791                                        avg =       8.0
       overall = 0.1191                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(3)       =    317.44
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2328783   .0161908    14.38   0.000      .201145    .2646117
        educ |   .0758092   .0086334     8.78   0.000      .058888    .0927303
       union |   .0991208   .0189042     5.24   0.000     .0620693    .1361723
       _cons |   .6306824   .1029879     6.12   0.000     .4288299    .8325349
-------------+----------------------------------------------------------------
     sigma_u |  .32508679
     sigma_e |   .3775581
         rho |  .42573728   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- **Step 5:** Test for Random effects versus fixed effects using **Hausman test.**

```
. xtreg lwage married educ union, fe i(nr)

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

R-sq:  within  = 0.0498                         Obs per group: min =         8
       between = 0.0573                                        avg =       8.0
       overall = 0.0538                                        max =         8

                                                F(2,3813)          =    100.00
corr(u_i, Xb)  = -0.0035                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2416845   .0176735    13.67   0.000     .2070341    .2763348
        educ |  (dropped)
       union |   .0700438    .020724     3.38   0.001     .0294127    .1106749
       _cons |   1.525957    .010825   140.97   0.000     1.504733     1.54718
-------------+----------------------------------------------------------------
     sigma_u |  .37939434
     sigma_e |   .3775581
         rho |  .50242582   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3813) =     6.98           Prob > F = 0.0000

. est store fixed

. xtreg lwage married educ union, re i(nr)

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

R-sq:  within  = 0.0493                         Obs per group: min =         8
       between = 0.1791                                        avg =       8.0
       overall = 0.1191                                        max =         8

Random effects u_i ~ Gaussian                   Wald chi2(3)       =    317.44
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     married |   .2328783   .0161908    14.38   0.000      .201145    .2646117
        educ |   .0758092   .0086334     8.78   0.000      .058888    .0927303
       union |   .0991208   .0189042     5.24   0.000     .0620693    .1361723
       _cons |   .6306824   .1029879     6.12   0.000     .4288299    .8325349
-------------+----------------------------------------------------------------
     sigma_u |  .32508679
     sigma_e |   .3775581
         rho |  .42573728   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. hausman fixed, force

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
     married |    .2416845     .2328783        .0088061        .0070859
       union |    .0700438     .0991208        -.029077         .008492
------------------------------------------------------------------------------
                             b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       13.45
                Prob>chi2 =      0.0012
```

Conclusion: Reject Ho (Thus fixed-effect specification is preferred)