Day 2A
Inference for Clustered Data: Part 1
With a Panel Data Example

A. Colin Cameron
Univ. of Calif. - Davis

... for
Center of Labor Economics
Norwegian School of Economics
Advanced Microeconometrics

Aug 28 - Sep 1, 2017

Introduction: Why Does Clustering Matter?

- By clustered data we mean data with the property that
  - observations in the same cluster or group are correlated
  - observations in the different clusters or group are uncorrelated
  - the clusters are known (unlike in cluster analysis).
- Failure to control for clustering in regression usually
  - underestimates standard errors
  - overstates t statistics and p-values and provides too narrow confidence intervals.
- These slides focus on panel data example
  - individuals uncorrelated but observations over time correlated for a givein individual.
- The subsequent set of slides focus on cross-section example
  - individuals in groups such as a village uncorrelated but individuals in different villages.
- Focus is on OLS but most analysis generalizes to nonlinear models.

Outline

2. Motivation: A simple example

- Suppose we have univariate data $y_i \sim (\mu, \sigma^2)$.
- We estimate μ by $\bar{y}$ and

$$\mathrm{Var}[\bar{y}] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} y_i\right] = \frac{1}{N^2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}(y_i, y_j)\right].$$

- **Given independence** over i
  - this simplifies to $\mathrm{Var}[\bar{y}] = \frac{1}{N}\sigma^2$
  - since $\mathrm{Cov}(y_i, y_j) = 0$ so we have just $\frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{Var}(y_i)\right] = \frac{1}{N^2} \times N\sigma^2$.
- Now **suppose observations are positively correlated**
  - then there are lots of $\mathrm{Cov}(y_i, y_j)$ terms to add in, increasing $\mathrm{Var}[\bar{y}]$.

Intuition (continued)

- Assume **equicorrelation** with $\text{Cov}(y_i, y_j) = \rho\sigma^2$ for $i \neq j$

$$Var\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \cdots & \cdots & \sigma_{1N} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{N1} & \cdots & \cdots & \sigma_{NN} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \cdots & 1 \end{bmatrix}$$

- Then $\text{Var}[\bar{y}] = \frac{1}{N^2}\sigma^2[N + N(N-1)\rho] = \frac{1}{N}\sigma^2\{1 + (N-1)\rho\}$.
- The variance is $1 + (N-1)\rho$ times **larger**!
  - if $N = 81$ and $\rho = 0.1$ then $\text{Var}[\bar{y}]$ is 9 times larger than $\frac{1}{N}\sigma^2$!.
- Reason: **An extra observation is not providing a new independent piece of information.**

Example 1: Difference-in-Differences State-Year Panel

- Example: How do wages respond to a policy indicator variable $\mathbf{d}_{ts}$ that varies by state
  - e.g. $\mathbf{d}_{ts} = 1$ if minimum wage law in effect
- OLS estimate model for state s at time

$$y_{ts} = \alpha + \mathbf{x}'_{ts}\boldsymbol{\beta} + \gamma \times \mathbf{d}_{ts} + u_{ts}.$$

- The regressor $\mathbf{d}_{ts}$ will be highly correlated within cluster.
- Problems if additionally errors u_{ts} are correlated within cluster i.e. model systematically overpredicts (or underpredicts) wage in all years for state s
  - $\text{Cor}[u_{ts}, u_{t's'}] \neq 0$ if $s = s'$
- Natural model for error is a time series model
  - e.g. AR(1): $\text{Var}[u_{ts}] = \sigma^2$ and $\text{Cor}[u_{ts}, u_{t+k,s}] = \rho^k\sigma^2$.
  - We want to do inference without specifying such a model.
- Same problems if individuals in states over time

$$y_{its} = \alpha + \mathbf{x}'_{its}\boldsymbol{\beta} + \gamma \times \mathbf{d}_{ts} + u_{its}.$$

Example 2: Individuals in Cluster ("Moulton-type")

- Example: How do job injury rates effect wages? Hersch (1998).
  - CPS individual data on male wages $N = 5960$.
  - But there is no individual data on job injury rate.
  - Instead aggregated data on industry injury rates for 211 industries
- OLS estimate model for individual i in industry g

$$y_{ig} = \alpha + \mathbf{x}'_{ig}\boldsymbol{\beta} + \gamma \times rind_g + u_{ig}.$$

- The regressor $rind_g$ is perfectly correlated within cluster.
- Problems if additionally errors u_{ig} are correlated within cluster i.e. model systematically overpredicts (or underpredicts) wage for all individuals in industry g
  - $\text{Cor}[u_{ig}, u_{jh}] \neq 0$ if $g = h$
- Natural model for error is within-cluster equicorrelation
  - $\text{Var}[u_{ig}] = \sigma^2$ and $\text{Cor}[u_{ig}, u_{jg}] = \rho\sigma^2$.
  - We want to do inference without specifying such a model.

- Moulton (1986, 1990) and Bertrand, Duflo & Mullainathan (2004) showed
  - the practical importance of controlling for clustering
  - clustering can arise in a wider range of settings than obvious.
- To control for clustering
  - originally use a restrictive one-way random effects model
  - now use cluster-robust standard errors
  - White (1984), Liang and Zeger (1986), Arellano (1987), Rogers (1993)
  - Wooldridge (2003, 2006) and Cameron and Miller (2011, 2015) provide surveys.

3. Panel Data Summary: Wages Example

- **This handout considers a panel data example**
  - the next handout considers a cross-section clustering example
- PSID wage data 1976-82 on 595 individuals. Balanced.
- Source: Baltagi and Khanti-Akom (1990).
  - Corrected version of Cornwell and Rupert (1998).
- **Note: For investigating and estimating Moulton-type data**
  - **the xt panel commands can also be used**
  - though only a subset are relevant.
- We present Stata `xt` commands in detail.
- We consider OLS, FGLS, random effects and fixed effects estimators.

Reading in Panel Data

- Data organization may be
  - long form: each observation is an individual-time (i, t) pair
  - wide form: each observation is data on i for all time periods
  - wide form: each observation is data on t for all individuals
- xt commands require data in long form
  - use reshape long command to convert from wide to long form
  - see Cameron and Trivedi (2010) chapter 8.11.
- Data here are already in long form.

```
.  * Read in data set
.  use mus08psidextract.dta, clear
(PSID wage data 1976-82 from Baltagi and Khanti-Akom
(1990))
```

Summarize Data using Non-panel Commands

```
. * Describe dataset
. describe

Contains data from mus08psidextract.dta
  obs:         4,165                          PSID wage data 1976-82 from Baltagi and Khanti-Akom (1990)
 vars:            15                          16 Aug 2007 16:29
 size:       283,220 (97.5% of memory free)   (_dta has notes)
-----------------------------------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------------------------------------
exp             float  %9.0g                  years of full-time work experience
wks             float  %9.0g                  weeks worked
occ             float  %9.0g                  occupation; occ==1 if in a blue-collar occupation
ind             float  %9.0g                  industry; ind==1 if working in a manufacturing industry
south           float  %9.0g                  residence; south==1 if in the South area
smsa            float  %9.0g                  smsa==1 if in the Standard metropolitan statistical area
ms              float  %9.0g                  marital status
fem             float  %9.0g                  female or male
union           float  %9.0g                  if wage set be a union contract
ed              float  %9.0g                  years of education
blk             float  %9.0g                  black
lwage           float  %9.0g                  log wage
id              float  %9.0g
t               float  %9.0g
exp2            float  %9.0g
-----------------------------------------------------------------------------------------------------------
```

- Summary statistics combine variation over i and t.

```
. * Summarize dataset
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         exp |       4165    19.85378    10.96637          1         51
         wks |       4165    46.81152    5.129098          5         52
         occ |       4165    .5111645    .4999354          0          1
         ind |       4165    .3954382    .4890033          0          1
       south |       4165    .2902761    .4539442          0          1
-------------+---------------------------------------------------------
        smsa |       4165    .6537815     .475821          0          1
          ms |       4165    .8144058    .3888256          0          1
         fem |       4165     .112605    .3161473          0          1
       union |       4165    .3639856    .4812023          0          1
          ed |       4165    12.84538    2.787995          4         17
-------------+---------------------------------------------------------
         blk |       4165    .0722689    .2589637          0          1
       lwage |       4165    6.676346    .4615122    4.60517      8.537
          id |       4165         298    171.7821          1        595
           t |       4165           4     2.00024          1          7
        exp2 |       4165     514.405    496.9962          1       2601
```

- Since 4165 $(= 7 \times 595)$ observations for all variables the dataset is balanced and complete.

- Listing the first few observations is useful

```
. * Organization of data set
. list id t exp wks occ in 1/3, clean

       id   t   exp   wks   occ
  1.    1   1     3    32     0
  2.    1   2     4    43     0
  3.    1   3     5    40     0
```

- Data are in long form, sorted by `id` and then by `t`

Stata Commands for Panel Data Summary

- Commands `describe`, `summarize` and `tabulate` confound cross-section and time series variation.
- Instead use specialized panel commands after `xtset`:
  - `xtdescribe`: extent to which panel is unbalanced
  - `xtsum`: separate within (over time) and between (over individuals) variation
  - `xttab`: tabulations within and between for discrete data e.g. binary
  - `xttrans`: transition frequencies for discrete data
  - `xtline`: time series plot for each individual on one chart
  - `xtdata`: scatterplots for within and between variation.

Summarize Data using Panel Commands

- `xtset` command defines i and t.
  - Allows use of panel commands and some time series operators
  - For Moulton-type data only `xtset id`

```
.  * Declare individual identifier and time identifier
.  xtset id t
 panel variable:  id (strongly balanced)
 time variable:  t, 1 to 7
 delta:  1 unit
```

- `xtdescribe` command summarizes number of time periods each individual is observed.

```
. * Panel description of data set
. xtdescribe

      id:  1, 2, ..., 595                                    n =        595
       t:  1, 2, ..., 7                                      T =          7
           Delta(t) = 1 unit
           Span(t)  = 7 periods
           (id*t uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         7       7       7         7         7       7       7

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
      595    100.00  100.00 |  1111111
 ---------------------------+---------
      595    100.00         |  XXXXXXX
```

- Data are balanced with every individual i having 7 time periods of data.

- `xtsum` command splits overall variation into
  - between variation: variation in $\bar{x}_i = T_i^{-1} \sum_i x_{it}$ across individuals
  - within variation: variation in x_{it} around $\bar{x}_i$

```
. * Panel summary statistics: within and between variation
. xtsum lwage exp ed t
```

Variable		Mean	Std. Dev.	Min	Max	Observations
lwage	overall	6.676346	.4615122	4.60517	8.537	N = 4165
	between		.3942387	5.3364	7.813596	n = 595
	within		.2404023	4.781808	8.621092	T = 7
exp	overall	19.85378	10.96637	1	51	N = 4165
	between		10.79018	4	48	n = 595
	within		2.00024	16.85378	22.85378	T = 7
ed	overall	12.84538	2.787995	4	17	N = 4165
	between		2.790006	4	17	n = 595
	within		0	12.84538	12.84538	T = 7
t	overall	4	2.00024	1	7	N = 4165
	between		0	4	4	n = 595
	within		2.00024	1	7	T = 7

- For time-invariant variable `ed` the within variation is zero.
  For individual-invariant variable `t` the between variation is zero.
  For `lwage` the within variation $<$ between variation.

- xttab command provides more detail for discrete-valued variable.

```
. * Panel tabulation for a variable
. xttab south

                  Overall             Between            Within
    south |   Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |   2956     70.97       428     71.93          98.66
        1 |   1209     29.03       182     30.59          94.90
----------+-----------------------------------------------------
    Total |   4165    100.00       610    102.52          97.54
                               (n = 595)
```

- 29.03% on average were in the south.
  30.59% were ever in the south.
  94.9% of those ever in the south were always in the south.

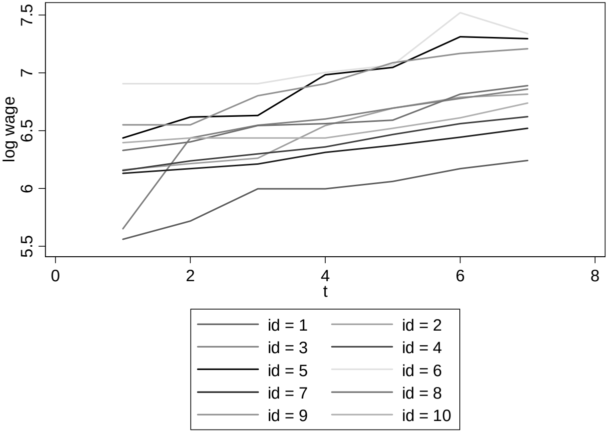
- xttrans provides transition probabilities for discrete-valued variable.

```
. * Transition probabilities for a variable
. xttrans south, freq

 residence;
   south==1 |    residence; south==1
  if in the |    if in the South area
 South area |         0          1 |     Total
------------+----------------------+----------
          0 |     2,527          8 |     2,535 
            |     99.68       0.32 |    100.00 
------------+----------------------+----------
          1 |         8      1,027 |     1,035 
            |      0.77      99.23 |    100.00 
------------+----------------------+----------
      Total |     2,535      1,035 |     3,570 
            |     71.01      28.99 |    100.00 
```

- For the 28.99% of the sample ever in the south, 99.23% remained in the south the next period.

```
.   * Time series plots of log wage for first 10 individuals
.   xtline lwage if id<=10, overlay
```



- Much autocorrelation in each person's wage.

Autocorrelations

- Because `xtset` set a time variable can use time series commands
  - `Lj.x` gives `x` lagged *j* periods.
- Can compute autocorrelations for a variable.

```
. * First-order autocorrelation in a variable
. sort id t

. correlate lwage L.lwage L2.lwage L3.lwage L4.lwage L5.lwage L6.lwage
(obs=595)
```

	lwage	L. lwage	L2. lwage	L3. lwage	L4. lwage	L5. lwage	L6. lwage
lwage							
--.	1.0000						
L1.	0.9238	1.0000					
L2.	0.9083	0.9271	1.0000				
L3.	0.8753	0.8843	0.9067	1.0000			
L4.	0.8471	0.8551	0.8833	0.8990	1.0000		
L5.	0.8261	0.8347	0.8721	0.8641	0.8667	1.0000	
L6.	0.8033	0.8163	0.8518	0.8465	0.8594	0.9418	1.0000

- High serial correlation: $\text{Cor}[y_t, y_{t-6}] = 0.80$ and not AR(1).
- Note that estimated autocorrelations without imposing stationarity.
- Weakness: Only 595 observations used as needed `L6.lwage`

Autocorrelations - better

- Command `pwcorr` uses all the available data

```
. * The previous uses few observations as needs L6.lwage
. * The following uses as much data as is available
. pwcorr lwage L.lwage L2.lwage L3.lwage L4.lwage L5.lwage L6.lwage
```

	lwage	L.lwage	L2.lwage	L3.lwage	L4.lwage	L5.lwage	L6.lwage
lwage	1.0000						
L.lwage	0.9189	1.0000					
L2.lwage	0.8858	0.9128	1.0000				
L3.lwage	0.8649	0.8748	0.9044	1.0000			
L4.lwage	0.8460	0.8565	0.8684	0.8988	1.0000		
L5.lwage	0.8220	0.8444	0.8602	0.8631	0.8944	1.0000	
L6.lwage	0.8033	0.8163	0.8518	0.8465	0.8594	0.9418	1.0000

Autocorrelations - nonstationary

- The preceding imposed stationarity: $\text{Corr}[y_{it}, y_{it-j}] = \text{Corr}[y_{it+k}, y_{it+k-j}]$.
- Following does not constrain e.g. correlation between years 1 and 2 to equal that between 2 and 3 (so no longer stationary)

```
. * First-order autocorrelation differs in different year pairs
. forvalues s = 2/7 {
  2.      quietly corr lwage L1.lwage if t == `s'
  3.      display "Autocorrelation at lag 1 in year `s' = " %6.3f r(rho)
  4.      }
Autocorrelation at lag 1 in year 2 =  0.942
Autocorrelation at lag 1 in year 3 =  0.867
Autocorrelation at lag 1 in year 4 =  0.899
Autocorrelation at lag 1 in year 5 =  0.907
Autocorrelation at lag 1 in year 6 =  0.927
Autocorrelation at lag 1 in year 7 =  0.924
```

4. Pooled OLS (a Population-Averaged Estimator)

- Pooled OLS is regular OLS of y_{it} on $\mathbf{x}_{it}$
  - Consistent if $\mathbf{x}_{it}$ is uncorrelated with the error u_{it}.

```
. * Pooled OLS with incorrect default standard errors
. regress lwage exp exp2 wks ed, noheader
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exp	.044675	.0023929	18.67	0.000	.0399838	.0493663
exp2	-.0007156	.0000528	-13.56	0.000	-.0008191	-.0006121
wks	.005827	.0011827	4.93	0.000	.0035084	.0081456
ed	.0760407	.0022266	34.15	0.000	.0716754	.080406
_cons	4.907961	.0673297	72.89	0.000	4.775959	5.039963

- Important: The default standard errors are too small
  - they erroneously assume errors are independent over t for given i.
  - this assumes more information content from data then is the case.

Cluster-Robust Standard Errors

- Should instead use cluster-robust standard errors

```
. * Pooled OLS with cluster-robust standard errors
. regress lwage exp exp2 wks ed, noheader vce(cluster id)
                                 (Std. Err. adjusted for 595 clusters in id)
```

lwage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
exp	.044675	.0054385	8.21	0.000	.0339941	.055356
exp2	-.0007156	.0001285	-5.57	0.000	-.0009679	-.0004633
wks	.005827	.0019284	3.02	0.003	.0020396	.0096144
ed	.0760407	.0052122	14.59	0.000	.0658042	.0862772
_cons	4.907961	.1399887	35.06	0.000	4.633028	5.182894

- Cluster-robust standard errors here are twice as large as default! Cluster-robust t-statistics are half as large as default!
- Typical result. Need to use cluster-robust se's if use pooled OLS.

OLS with Clustered Errors Theory

- Model for G clusters with N_g individuals per cluster:

$$\begin{aligned} y_{ig} &= \mathbf{x}_{ig}'\boldsymbol{\beta} + u_{ig}, \quad i = 1, ..., N_g,\ g = 1, ..., G, \\ \mathbf{y}_g &= \mathbf{X}_g\boldsymbol{\beta} + \mathbf{u}_g, \quad g = 1, ..., G, \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \end{aligned}$$

- OLS estimator

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \left(\sum_{g=1}^{G}\sum_{i=1}^{N_g} \mathbf{x}_{ig}\mathbf{x}_{ig}'\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{N_g} \mathbf{x}_{ig}y_{ig}\right) \\ &= \left(\sum_{g=1}^{G} \mathbf{X}_g'\mathbf{X}_g\right)^{-1}\left(\sum_{g=1}^{G} \mathbf{X}_g'\mathbf{y}_g\right) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \end{aligned}$$

- As usual

$$\begin{aligned}\widehat{\beta} &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{g=1}^{G}\mathbf{X}_g\mathbf{u}_g).\end{aligned}$$

- Assume independence over g and correlation within g

$$\mathrm{E}[u_{ig}u_{jg'}|\mathbf{x}_{ig},\mathbf{x}_{jg'}] = 0, \text{ unless } g = g'.$$

- Then $\widehat{\beta} \overset{a}{\sim} \mathcal{N}[\beta, \mathrm{V}[\widehat{\beta}]]$ with asymptotic variance

$$\begin{aligned}\mathrm{Avar}[\widehat{\beta}] &= (\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}(\textstyle\sum_{g=1}^{G}\mathrm{E}[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g])(\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1} \\ &\neq \sigma_u^2(\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}.\end{aligned}$$

Consequences - KEY RESULT FOR INSIGHT

- Suppose equicorrelation within cluster g

$$\text{Cor}[u_{ig}, u_{jg} | \mathbf{x}_{ig}, \mathbf{x}_{jg}] = \begin{cases} 1 & i = j \\ \rho_u & i \neq j \end{cases}$$

  - this arises in a random effects model with $u_{ig} = \alpha_g + \varepsilon_{ig}$, where α_g and ε_{ig} are i.i.d. errors.
  - an example is individual i in village g or student i in school g.

- The incorrect default OLS variance estimate should be inflated by

$$\tau_j \simeq 1 + \rho_{x_j} \rho_u (\bar{N}_g - 1),$$

  - (1) ρ_{x_j} is the within cluster correlation of x_j
  - (2) ρ_u is the within cluster error correlation
  - (3) $\bar{N}_g$ is the average cluster size.
  - Need both (1) and (2) and it also increases with (3)
  - Kloek (1981), Scott and Holt (1982).

- Moulton (1986, 1990) showed that the inflation can be large even if ρ_u is small
  - especially with a grouped regressor (same for all individuals in group) so that $\rho_x = 1$.
  - CPS data example: $N_g = 81$, $\rho_x = 1$ and $\rho_u = 0.1$ then $\tau_j \simeq 1 + \rho_{x_j}\rho_u(\bar{N}_g - 1) = 1 + 1 \times 0.1 \times 80 = 9$.
    - ★ true standard errors are three times the default!
- So should correct for clustering even in settings where not obviously a problem.
- Bertrand,Duflo and Mullainathan (2004) showed such problems also arise for difference-in-differences analysis with individual in state-year panel
  - and sluster on state, not state-year pair.

The Cluster-Robust Variance Matrix Estimate

- Recall for OLS with independent heteroskedastic errors

$$\mathrm{Avar}[\widehat{\beta}] = (\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}(\textstyle\sum_{i=1}^{N} \mathrm{E}[u_i^2\mathbf{x}_i\mathbf{x}_i'])(\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}$$

can be consistently estimated (White (1980)) as $N \rightarrow \infty$ by

$$\widehat{\mathrm{V}}[\widehat{\beta}] = (\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{i=1}^{N} \widehat{u}_i^2\mathbf{x}_i\mathbf{x}_i'])(\mathbf{X}'\mathbf{X})^{-1}.$$

- Similarly for OLS with independent clustered errors

$$\mathrm{Avar}[\widehat{\beta}] = (\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}(\textstyle\sum_{g=1}^{G} \mathrm{E}[\mathbf{X}_g'\mathbf{u}_g\mathbf{u}_g'\mathbf{X}_g])(\mathrm{E}[\mathbf{X}'\mathbf{X}])^{-1}$$

can be consistently estimated as $G \rightarrow \infty$ by the cluster-robust variance estimate (CRVE)

$$\widehat{\mathrm{V}}_{\mathrm{CR}}[\widehat{\beta}] = (\mathbf{X}'\mathbf{X})^{-1}(\textstyle\sum_{g=1}^{G} \mathbf{X}_g'\widetilde{\mathbf{u}}_g\widetilde{\mathbf{u}}_g'\mathbf{X}_g)(\mathbf{X}'\mathbf{X})^{-1}.$$

  - `regress` uses $\widetilde{\mathbf{u}}_g = c\widehat{\mathbf{u}}_g = c(\mathbf{y}_g - \mathbf{X}_g\widehat{\beta})$ where $c = \frac{G}{G-1}\frac{N-1}{N-K} \simeq \frac{G}{G-1}$.

- The CRVE was
  - proposed by White (1984) for balanced case
  - proposed by Liang and Zeger (1986) for grouped data
  - proposed by Arellano (1987) for FE estimator for short panels (group on individual)
  - Hansen (2007a) and Carter, Schnepel and Steigerwald (2013) also allow $N_g \longrightarrow \infty$.
  - popularized by incorporation in Stata as the cluster option (Rogers (1993)).
  - also allows for heteroskedasticity so is cluster- and heteroskedastic-robust.
- Stata with cluster identifier `id_clu`
  - `regress y x, vce(cluster id_clu)`
  - `xtreg y x, pa corr(ind) vce(robust)`
    - ★ after `xtset id_clu`
    - ★ from version 12.1 on Stata interprets `vce(robust)` as cluster-robust for all `xt` commands.

Some Limitations

- Theory requires number of clusters $G \longrightarrow \infty$
  - problem if e.g. the cluster is state and there are few states
  - this is considered in the subsequent of slides.
- The rank of $\widehat{V}_{CR}[\widehat{\beta}]$ is at most the minimum of k (the dimension of β) and $G - 1$
  - if $k > G - 1$ it is possible to test at most $G - 1$ restrictions
  - then output gives no overall F-test but individual t-tests are okay.

5. Feasible GLS with Cluster-Robust Inference

- Potential efficiency gains for feasible GLS compared to OLS.
- Specify a model for $\Omega_g = \mathrm{E}[\mathbf{u}_g\mathbf{u}_g'|\mathbf{X}_g]$, such as within-cluster equicorrelation.
- Given a consistent estimate $\widehat{\Omega}$ of Ω, the feasible GLS estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}} = \left(\sum_{g=1}^{G} \mathbf{X}_g'\widehat{\Omega}_g^{-1}\mathbf{X}_g\right)^{-1} \sum_{g=1}^{G} \mathbf{X}_g'\widehat{\Omega}_g^{-1}\mathbf{y}_g.$$

- Default $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}] = (\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X})^{-1}$ requires correct Ω.
- To guard against misspecified Ω_g uses cluster-robust

$$\widehat{\mathrm{V}}_{\mathrm{CR}}[\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}] = \left(\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{X}_g'\widehat{\Omega}_g^{-1}\widehat{\mathbf{u}}_g\widehat{\mathbf{u}}_g'\widehat{\Omega}_g^{-1}\mathbf{X}_g\right) \left(\mathbf{X}'\widehat{\Omega}^{-1}\mathbf{X}\right)^{-1}$$

  - where $\widehat{\mathbf{u}}_g = \mathbf{y}_g - \mathbf{X}_g\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$ and $\widehat{\Omega} = \mathrm{Diag}[\widehat{\Omega}_g]$
  - assumes $\mathbf{u}_g$ and $\mathbf{u}_h$ are uncorrelated, for $g \neq h$
  - and needs $G \to \infty$.

FGLS Example 1

- Example 2 - Time series correlation for panel data and DiD
  - AR(1) error $u_{it} = \rho u_{i,t-1} + \varepsilon_{it}$ and ε_{it} i.i.d.
  - Implies $\text{Corr}[u_{i,t}, u_{i,t-k}] = \rho^k$.
  - Default VE: `xtreg y x, pa corr(ar 1)`
  - Cluster-robust VE: `xtreg y x, pa corr(ar 1) vce(robust)`
- Stata allows a range of correlation structures
  - exchangeable; independent; AR(p); MA(q); Kiefer(1980)
- Puzzle - why is FGLS not used more?
  - Easily done in Stata with robust VCE if $G \to \infty$
  - Unless FE's present and N_g small (see later).

FGLS Example 2

- Example 1 - Random effects
  - $y_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta} + \alpha_g + \varepsilon_{ig}$ where α_g and ε_{ig} are i.i.d. errors.
  - Within-cluster errors are equicorrelated or exchangeable
  - Default VE: `xtreg y x, pa corr(exch)`
  - Cluster-robust VE: `xtreg y x, pa corr(exch) vce(robust)`
- Richer variation of random effects is hierarchical linear model or mixed model
  - $y_{ig} = \mathbf{x}'_{ig}\boldsymbol{\beta}_g + u_{ig}$
  - $\boldsymbol{\beta}_g = \mathbf{W}_g\gamma + \mathbf{v}_j$ where u_{ig} and $\mathbf{v}_g$ are errors
  - In Stata use `mixed`.
  - Special case is RE MLE
    - ★ `xtmixed y x || id_clu:  , covar(unstr) mle`

AR(2) Error - Default Standard Errors

- More efficient Feasible GLS assuming an error correlation model over time.
  - Here specify AR(2): $u_{it} = \rho_1 u_{i,t-1} + \rho_2 u_{i,t-2} + \varepsilon_{it}$

```
. xtreg lwage $xlist, pa corr(ar 2) nolog

GEE population-averaged model                   Number of obs      =      4165
Group and time vars:                   id t     Number of groups   =       595
Link:                              identity     Obs per group: min =         7
Family:                            Gaussian                    avg =       7.0
Correlation:                          AR(2)                    max =         7
                                                Wald chi2(4)       =    837.36
Scale parameter:                   .1966639     Prob > chi2        =    0.0000
```

lwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
exp	.0718915	.0042152	17.06	0.000	.0636298	.0801531
exp2	-.0008966	.000093	-9.64	0.000	-.0010788	-.0007143
wks	.0002964	.0006389	0.46	0.643	-.0009558	.0015485
ed	.0905069	.0059771	15.14	0.000	.0787921	.1022217
_cons	4.526381	.0965424	46.88	0.000	4.337162	4.715601

AR(2) Error - Cluster Robust Standard Errors

- Repeat previous estimator but with cluster-robust standard errors.
  - Robust se's similar except for wks

```
. xtreg lwage exp exp2 wks ed, pa corr(ar 2) vce(robust) nolog

GEE population-averaged model                   Number of obs      =      4165
Group and time vars:                 id t       Number of groups   =       595
Link:                            identity       Obs per group: min =         7
Family:                          Gaussian                      avg =       7.0
Correlation:                        AR(2)                      max =         7
                                                Wald chi2(4)       =    873.28
Scale parameter:                 .1966639       Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semirobust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0718915    .003999    17.98   0.000     .0640535    .0797294
        exp2 |  -.0008966   .0000933    -9.61   0.000    -.0010794   -.0007137
         wks |   .0002964   .0010553     0.28   0.779     -.001772    .0023647
          ed |   .0905069   .0060161    15.04   0.000     .0787156    .1022982
       _cons |   4.526381   .1056897    42.83   0.000     4.319233    4.733529
------------------------------------------------------------------------------
```

  - Same as `xtgee lwage $xlist, pa corr(ar 2) vce(robust) nolog`

6. Random Effects Estimator

- Random effects estimator is FGLS estimator for the RE model

$$
\begin{aligned}
y_{it} &= \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\
\alpha_i &\sim \text{i.i.d.}[\alpha, \sigma^2_\alpha] \\
\varepsilon_{it} &\sim \text{i.i.d.}[0, \sigma^2_\varepsilon]
\end{aligned}
$$

- The RE model implies equicorrelated (or exchangeable) errors
  - $\text{Var}[\alpha_i + \varepsilon_{it}] = \text{Var}[\alpha_i] + \text{Var}[\varepsilon_{it}] = \sigma^2_\alpha + \sigma^2_\varepsilon$
  - For $s \neq t$, $\text{Cov}[\alpha_i + \varepsilon_{it}, \alpha_i + \varepsilon_{is}] = \text{Cov}[\alpha_i, \alpha_i] = \sigma^2_\alpha$
  - So $\text{Corr}[\alpha_i + \varepsilon_{it}, \alpha_i + \varepsilon_{is}] = \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2_\varepsilon)$ for all $s \neq t$.
- FGLS can be shown to equal OLS in the transformed model

$$
(y_{it} - \widehat{\theta}_i \bar{y}_i) = (\mathbf{x}_{it} - \widehat{\theta}_i \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + error,
$$

where $\widehat{\theta}_i$ is a consistent estimate of $\theta_i = 1 - \sqrt{\sigma^2_\varepsilon / (T_i \sigma^2_\alpha + \sigma^2_\varepsilon)}$.

- Random effects estimates with cluster-robust standard errors:

```
. * Random effects estimator with cluster-robust standard errors
. xtreg lwage exp exp2 wks ed, re vce(robust) theta

Random-effects GLS regression                   Number of obs      =      4165
Group variable: id                              Number of groups   =       595

R-sq:  within  = 0.6340                         Obs per group: min =         7
       between = 0.1716                                        avg =       7.0
       overall = 0.1830                                        max =         7

                                                Wald chi2(4)       =   1598.50
corr(u_i, X)     = 0 (assumed)                  Prob > chi2        =    0.0000
theta            = .82280511

                                  (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0888609   .0039992    22.22   0.000     .0810227    .0966992
        exp2 |  -.0007726   .0000896    -8.62   0.000    -.0009481   -.000597
         wks |   .0009658   .0009259     1.04   0.297    -.000849    .0027806
          ed |   .1117099   .0083954    13.31   0.000     .0952552    .1281647
       _cons |   3.829366   .1333931    28.71   0.000     3.567921    4.090812
-------------+----------------------------------------------------------------
     sigma_u |  .31951859
     sigma_e |  .15220316
         rho |  .81505521   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Option theta gives $\widehat{\theta} = 0.82 = 1 - \sqrt{0.152^2/(7 \times 0.319^2 + 0.152^2)}$.

7. Fixed Effects Estimator

- Mean-differencing eliminates α_i

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \\ \Rightarrow \quad \bar{y}_i &= \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i \\ \Rightarrow \quad (y_{it} - \bar{y}_i) &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i) \end{aligned}$$

- The within or fixed effects estimator is OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$
  - Efficiency loss as use only within variation
  - Coefficient of any time-invariant regressor is not identified $(x_{it} = \overline{x}_i)$
  - Use cluster-robust standard errors
  - Stata command `xtreg, fe`

- Within or FE estimates: Default SE's
  - Variable ed is not identified because time-invariant regressor in this dataset!

```
. xtreg lwage $xlist, fe
note: ed omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =      4165
Group variable: id                              Number of groups   =       595

R-sq:  within  = 0.6566                         Obs per group: min =         7
       between = 0.0276                                        avg =       7.0
       overall = 0.0476                                        max =         7

                                                F(3,3567)          =   2273.74
corr(u_i, Xb)  = -0.9107                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .1137879   .0024689    46.09   0.000     .1089473    .1186284
        exp2 |  -.0004244   .0000546    -7.77   0.000    -.0005315   -.0003173
         wks |   .0008359   .0005997     1.39   0.163    -.0003399    .0020116
          ed |          0  (omitted)
       _cons |   4.596396   .0389061   118.14   0.000     4.520116    4.672677
-------------+----------------------------------------------------------------
     sigma_u |  1.0362039
     sigma_e |  .15220316
         rho |  .97888036   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(594, 3567) =    56.52           Prob > F = 0.0000

.
```

- Within or FE estimates: Cluster Robust se's
  - standard errors are 50% larger even after inclusion of fixed effect!

```
. * Within or FE estimator with cluster-robust standard errors
. xtreg lwage exp exp2 wks ed, fe vce(robust)

Fixed-effects (within) regression               Number of obs      =      4165
Group variable: id                              Number of groups   =       595

R-sq:  within  = 0.6566                         Obs per group: min =         7
       between = 0.0276                                        avg =       7.0
       overall = 0.0476                                        max =         7

                                                F(3,594)           =   1059.72
corr(u_i, Xb)  = -0.9107                        Prob > F           =    0.0000

                                   (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .1137879   .0040289    28.24   0.000     .1058753    .1217004
        exp2 |  -.0004244   .0000822    -5.16   0.000    -.0005858   -.0002629
         wks |   .0008359   .0008697     0.96   0.337    -.0008721    .0025439
          ed |  (dropped)
       _cons |   4.596396   .0600887    76.49   0.000     4.478384    4.714408
-------------+----------------------------------------------------------------
     sigma_u |  1.0362039
     sigma_e |  .15220316
         rho |  .97888036   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Least Squares Dummy Variable Model for FE

- Several ways to compute FE estimator aside from `xtreg, fe`.
- Least squares dummy variables:
  - $d_{ji,t}$ for $j = 1, ..., N$ are N dummies equal to 1 if $i = j$
  - Estimate directly using `regress` or use areg

$$y_{it} = \sum_{j=1}^{N} \alpha_i d_{ji,t} + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

- Implementation in Stata
  - ```
    * FE model fitted as LSDV using areg
    areg lwage exp exp2 wks ed, absorb(id) vce(cluster id)
    ```
  - ```
    * FE model fitted using LSDV using regress
    set matsize 800
    quietly xi:  regress lwage exp exp2 wks ed i.id, ///
    vce(cluster id)
    estimates table, keep(exp exp2 wks ed _cons) ///
    b se b(%12.7f)
    ```

- Within or FE estimates obtained using areg
  - cluster-robust se's are approximatelu $\sqrt{7/6}$ too large

```
. * FE model fitted as LSDV using areg with cluster-robust standard errors
. areg lwage exp exp2 wks ed, absorb(id) vce(cluster id)
note: ed omitted because of collinearity

Linear regression, absorbing indicators         Number of obs     =      4,165
                                                F(   3,     594)  =     908.44
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9068
                                                Adj R-squared     =     0.8912
                                                Root MSE          =     0.1522

                                 (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .1137879   .0043514    26.15   0.000     .1052418    .1223339
        exp2 |  -.0004244   .0000888    -4.78   0.000    -.0005988     -.00025
         wks |   .0008359   .0009393     0.89   0.374    -.0010089    .0026806
          ed |          0  (omitted)
       _cons |   4.596396   .0648993    70.82   0.000     4.468936    4.723856
-------------+----------------------------------------------------------------
          id |   absorbed                                     (595 categories)
```

- IMPORTANT: if the LSDV approach is used with `regress` or `areg` then
  - cluster-robust standard errors in Stata are overstated
  - due to wrong degrees of freedom correction
  - especially if there are few observations per cluster.
- Stata `regress` and `areg` uses $c = \frac{G}{G-1}\frac{N-1}{N-Kall}$
  - Here $Kall = K + G$ (the β's plus the dummies)
  - For $N_g = 2$ so $N = 2G$: $c = \frac{G}{G-1}\frac{2G-1}{2G-K-G} \simeq 1 \times 2 = 2$.
  - Where `xtreg` knows to use $c = G/(G-1) \simeq 1$
  - Lesson: with G small use `xtreg, fe`
  - Otherwise if use `areg` or `regress i.` get se's that are approximately $\sqrt{G/(G-1)}$ times too big.
  - In panel setting $G = T$.

Aside: Mundlak/Chamberlain Model

- Mundlak and Chamberlain suppose the fixed effects

$$\alpha_i = \bar{\mathbf{x}}_i' \boldsymbol{\pi} + error.$$

- So OLS regress y_{it} on intercept $\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_i$
- Yields same $\boldsymbol{\beta}$ estimate as the FE estimator.

```
* FE model fitted by add mean of x as a regressor
global xlist exp exp2 wks ed
sort id
foreach x of varlist $xlist {
   by id:  egen mean`x' = mean(`x')
   }
regress lwage exp exp2 wks ed mean*, vce(robust)
```

Aside: Between Estimator

- OLS of $\bar{y}_i$ on intercept and $\bar{\mathbf{x}}_i$
  - `xtreg, be` has no heteroskedastic robust option but can bootstrap.

```
. *******  BETWEEN ESTIMATOR
.
. * Between estimator with default standard errors
. xtreg lwage exp exp2 wks ed, be

Between regression (regression on group means)  Number of obs      =      4165
Group variable: id                              Number of groups   =       595

R-sq:  within  = 0.1357                         Obs per group: min =         7
       between = 0.3264                                        avg =       7.0
       overall = 0.2723                                        max =         7

                                                F(4,590)           =     71.48
sd(u_i + avg(e_i.))=   .324656                  Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |    .038153   .0056967     6.70   0.000     .0269647    .0493412
        exp2 |  -.0006313   .0001257    -5.02   0.000    -.0008781   -.0003844
         wks |   .0130903   .0040659     3.22   0.001     .0051048    .0210757
          ed |   .0737838   .0048985    15.06   0.000     .0641632    .0834044
       _cons |   4.683039   .2100989    22.29   0.000     4.270407    5.095672
------------------------------------------------------------------------------
```

8. Fixed versus Random Effects Estimators

- RE has advantages: estimates all parameters & may be more efficient.
  - But RE is inconsistent if fixed effects present.
- Use Hausman test to discriminate between FE and RE.
  - This tests difference between FE and RE estimates is statistically significantly different from zero.
- Do not use `hausman` command – it requires that RE estimator is fully efficient (see next slide).
- Instead do one of the following
  - 1. Do a panel bootstrap of the Hausman test.
  - 2. Do the Wooldridge (2002) robust version of Hausman test.
    - ★ Test $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ in the auxiliary OLS regression

      $$(y_{it} - \widehat{\theta}\bar{y}_i) = (1 - \widehat{\theta})\alpha + (\mathbf{x}_{it} - \widehat{\theta}\overline{\mathbf{x}}_i)'\boldsymbol{\beta} + (\mathbf{x}_{1it} - \overline{\mathbf{x}}_{1i})'\boldsymbol{\gamma} + v_{it},$$

      where $\mathbf{x}_{1i} \subset \mathbf{x}_i$ denotes time-varying regressors only.
    - ★ Use cluster-robust standard errors for this test.
    - ★ Stata add-on `xtoverid` after `xtreg,re` does this.

Hausman Test Theory

- Hausman test compares to estimators $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$
  - Test $H_0 : \text{plim}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) = 0$ against $H_a : \text{plim}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) \neq 0$.
  - e.g. OLS versus 2SLS with possible endogenous regressor
  - e.g. RE versus FE with possible fixed effect.
- Under H_0, as usual $(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) \overset{a}{\sim} N[\mathbf{0}, \mathsf{V}[\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}])$.
- So form χ^2 statistic: $H = (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})' \left[\mathsf{V}[\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}]\right]^{-1} (\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})$
  - reject H_0 if $H > \chi^2$ critical value.
- Problem: To implement we need estimate of $\mathsf{V}[\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}]$.
- Hausman (1978) assumed $\widehat{\boldsymbol{\theta}}$ is fully efficient under H_0
  - then $\text{Cov}[\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}] = \text{Var}[\widehat{\boldsymbol{\theta}}]$
  - implying $\mathsf{V}[\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}] = \mathsf{V}[\widehat{\boldsymbol{\theta}}] + \mathsf{V}[\widetilde{\boldsymbol{\theta}}] - 2 \times \mathsf{V}[\widehat{\boldsymbol{\theta}}] = \mathsf{V}[\widetilde{\boldsymbol{\theta}}] - \mathsf{V}[\widehat{\boldsymbol{\theta}}]$.
  - but we rarely have $\widehat{\boldsymbol{\theta}}$ fully efficient.

Hausman Test Wrong

```
. * Wrong Hausman test assuming RE estimator is fully efficient under null hypothesi
. hausman FE_def RE_def, sigmamore

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     FE_def       RE_def       Difference          S.E.
-------------+----------------------------------------------------------------
         exp |    .1137879     .0888609        .0249269         .0012778
        exp2 |   -.0004244    -.0007726        .0003482         .0000285
         wks |    .0008359     .0009658       -.0001299         .0001108
------------------------------------------------------------------------------
                           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(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =     1513.02
                Prob>chi2 =      0.0000
```

Hausman Test Correct

- Following is manual

```
. * Correct Robust Hausman test using method of Wooldridge (2002)
. global xlist exp exp2 wks

. foreach x of varlist $xlist {
  2.    by id: egen mean`x' = mean(`x')
  3.    }

. quietly regress lwage exp exp2 wks meanexp meanexp2 meanwks, vce(cluster id)

. test meanexp meanexp2 meanwks

 ( 1)  meanexp = 0
 ( 2)  meanexp2 = 0
 ( 3)  meanwks = 0

       F(  3,   594) =  630.59
            Prob > F =    0.0000
```

- Get exactly same result using simpler
  `quietly regress lwage exp exp2 wks ///`
  `meanexp meanexp2 meanwks, vce(cluster id)`
  `test meanexp meanexp2 meanwks`
- Can also use Stata add-on `xtoverid` after `xtreg,re`

9. First Difference Estimator

- First-differencing eliminates α_i

$$\begin{array}{rrcl} & y_{it} & = & \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \\ \Rightarrow & y_{i,t-1} & = & \alpha_i + \mathbf{x}_{i,t-1}'\boldsymbol{\beta} + \varepsilon_{i,t-1} \\ \Rightarrow & (y_{it} - y_{i,t-1}) & = & (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) \end{array}$$

- First differences estimator
  - OLS regression of Δy_{it} on $\Delta \mathbf{x}_{it}$, i.e. use first differences.
  - Coefficient of any time-invariant regressor is not identified $(x_{it} = x_{i,t-1})$.
- Not used much for basic FE model
  - FE estimator is fully efficient if ε_{it} is iid $(0, \sigma_\varepsilon^2)$
  - FD estimator is fully efficient if $\varepsilon_{it} = \varepsilon_{it-1} + v_{it}$ where v_{it} is iid $(0, \sigma_v^2)$
  - FE=FD if $T = 2$ as then $y_{i2} - \bar{y}_i = y_{i2} - \frac{y_{i1}+y_{i2}}{2} = (y_{i2} - y_{i1})/2$.

First Difference Estimator (continued)

- No direct Stata command.
- Can `regress D.(lwage $xlist), vce(cluster id)`
- More comparable to FE is regress with noconstant
  `regress D.(lwage $xlist), noconstant vce(cluster id)`
  - FE is also noconstant, but then adds back in $\bar{y}$.
- FD is used in models with fixed effects and lagged dependent variable
  - e.g. $y_{it} = \alpha_i + \rho y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}$
  - then within estimator is inconsistent if short panel
  - instead do IV on FD model with lagged $y'_{i,t}$s as instruments
  - this is Arellano-Bond estimator `xtabond` (also add-on `xtabond2`).

- First differences estimator

```
. regress d.lwage d.exp d.exp2 d.wks d.ed, noconstant vce(cluster id)
note: _delete omitted because of collinearity

Linear regression                               Number of obs =     3570
                                                F(  3,    594) = 1035.19
                                                Prob > F      =  0.0000
                                                R-squared     =  0.2209
                                                Root MSE      =  .18156

                                (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     D.lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |
         D1. |   .1170654   .0040974    28.57   0.000     .1090182    .1251126
             |
        exp2 |
         D1. |  -.0005321   .0000808    -6.58   0.000    -.0006908   -.0003734
             |
         wks |
         D1. |  -.0002683   .0011783    -0.23   0.820    -.0025824    .0020459
             |
          ed |
         D1. |  (omitted)
------------------------------------------------------------------------------
```

10.Estimator comparison

```
.  * Compare various estimators (with cluster-robust se's)
.  global xlist exp exp2 wks ed
.  quietly regress lwage $xlist, vce(cluster id)
.  estimates store OLS
.  quietly xtgee lwage exp exp2 wks ed, corr(ar 2)
     vce(robust)
.  estimates store PFGLS
.  quietly xtreg lwage $xlist, be
.  estimates store BE
.  quietly xtreg lwage $xlist, re vce(robust)
.  estimates store RE
.  quietly xtreg lwage $xlist, fe vce(robust)
.  estimates store FE
.  estimates table OLS PFGLS BE RE FE, b(%9.4f) se stats(N)
```

Variable	OLS	PFGLS	BE	RE	FE
exp	0.0447	0.0719	0.0382	0.0889	0.1138
	0.0054	0.0040	0.0057	0.0040	0.0040
exp2	-0.0007	-0.0009	-0.0006	-0.0008	-0.0004
	0.0001	0.0001	0.0001	0.0001	0.0001
wks	0.0058	0.0003	0.0131	0.0010	0.0008
	0.0019	0.0011	0.0041	0.0009	0.0009
ed	0.0760	0.0905	0.0738	0.1117	(omitted)
	0.0052	0.0060	0.0049	0.0084	
_cons	4.9080	4.5264	4.6830	3.8294	4.5964
	0.1400	0.1057	0.2101	0.1334	0.0601
N	4165	4165	4165	4165	4165

legend: b/se

- Coefficients vary considerably across OLS, RE, FE and RE estimators.
  - FE and RE similar as $\widehat{\theta} = 0.82 \simeq 1$.
- Not shown is that even for FE and RE cluster-robust changes se's.
- Coefficient of `ed` not identified for FE as time-invariant regressor!

Standard Errors Comparison

- Compares default to panel-robust standard errors for RE and FE.

Variable	RE_def	RE	FE_def	FE
exp	0.0889	0.0889	0.1138	0.1138
	0.0028	0.0040	0.0025	0.0040
exp2	-0.0008	-0.0008	-0.0004	-0.0004
	0.0001	0.0001	0.0001	0.0001
wks	0.0010	0.0010	0.0008	0.0008
	0.0007	0.0009	0.0006	0.0009
ed	0.1117	0.1117	(omitted)	(omitted)
	0.0061	0.0084		
_cons	3.8294	3.8294	4.5964	4.5964
	0.0936	0.1334	0.0389	0.0601
N	4165	4165	4165	4165

legend: b/se

11. Panel Bootstrap

- Do pairs bootstrap where resample $(y, \mathbf{x})$ over individuals i rather than observations (i, t)
- Do B iterations of this step. On the b^{th} iteration:
  - form a sample of G clusters $\{(\mathbf{y}_1^*, \mathbf{X}_1^*), ..., (\mathbf{y}_G^*, \mathbf{X}_G^*)\}$ by resampling with replacement G times from the original sample
  - obtain estimate $\widehat{\boldsymbol{\beta}}_b$, $b = 1, ..., B$.
- Then $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}] = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\boldsymbol{\beta}}_b - \overline{\widehat{\boldsymbol{\beta}}})(\widehat{\boldsymbol{\beta}}_b - \overline{\widehat{\boldsymbol{\beta}}})'$, $\quad \overline{\widehat{\boldsymbol{\beta}}} = \frac{1}{B}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_b$.
- In Stata use `cluster(id)` option of `bootstrap`
  - but need to replace `xtset id t` with simply `xtset id`
  - and if FE then also add `idcluster(newid)` option
- This pairs panel (or clustered) bootstrap
  - yields essentially the same results as usual cluster-robust standard errors
  - is a bootstrap without asymptotic refinement.
- NOTE: Stata 14 introduced new random number generator giving different bootstrap results.

Panel Bootstrap OLS Estimator

```
. * OLS panel bootstrap using xtreg, pa and vce(boot)
. xtreg lwage exp exp2 wks ed, pa corr(ind) vce(boot, reps(400) ///
>   seed(10101) nodots)

GEE population-averaged model                   Number of obs      =      4,165
Group variable:                         id      Number of groups   =        595
Link:                             identity      Obs per group:
Family:                           Gaussian                     min =          7
Correlation:                   independent                     avg =        7.0
                                                               max =          7
                                                Wald chi2(4)       =     306.46
Scale parameter:                  .1525603      Prob > chi2        =     0.0000

Pearson chi2(4165):                 635.41      Deviance           =     635.41
Dispersion (Pearson):             .1525603      Dispersion         =   .1525603

                               (Replications based on 595 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |    .044675    .004895     9.13   0.000      .035081    .0542691
        exp2 |  -.0007156   .0001166    -6.14   0.000    -.0009442   -.0004871
         wks |    .005827   .0018287     3.19   0.001     .0022429    .0094111
          ed |   .0760407   .0052064    14.61   0.000     .0658364     .086245
       _cons |   4.907961   .1362406    36.02   0.000     4.640934    5.174987
------------------------------------------------------------------------------
```

- Note: the following gives exactly the same
  - `xtreg lwage exp exp2 wks ed, /// pa corr(ind) vce(boot, reps(400) seed(10101))`

Panel Bootstrap RE Estimator

```
. * RE panel bootstrap using vce(boot)
. xtreg lwage exp exp2 wks ed, re vce(boot, reps(400) seed(10101) nodots)

Random-effects GLS regression                   Number of obs     =      4,165
Group variable: id                              Number of groups  =        595

R-sq:                                           Obs per group:
     within  = 0.6340                                         min =          7
     between = 0.1716                                         avg =        7.0
     overall = 0.1830                                         max =          7

                                                Wald chi2(4)      =     683.66
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                   (Replications based on 595 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0888609   .0044227    20.09   0.000     .0801926    .0975293
        exp2 |  -.0007726   .0000935    -8.26   0.000    -.0009559   -.0005892
         wks |   .0009658   .0009027     1.07   0.285    -.0008035    .0027351
          ed |   .1117099    .009053    12.34   0.000     .0939664    .1294535
       _cons |   3.829366    .156578    24.46   0.000     3.522479    4.136254
-------------+----------------------------------------------------------------
     sigma_u |  .31951859
     sigma_e |  .15220316
         rho |  .81505521   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Same as `xtreg lwage exp exp2 wks ed, /// re vce(boot, reps(400) seed(10101) nodots)`

Panel Bootstrap FE Estimator - add idcluster()

```
. * FE panel bootstrap using bootstrap: with cluster(id) and idcluster
. * Need to add idcluster for bootstrap of FE
. bootstrap _b, reps(400) seed(10101) cluster(id) idcluster(newid) ///
>     nodots: xtreg lwage exp exp2 wks ed, fe

Fixed-effects (within) regression               Number of obs     =      4,165
Group variable: id                              Number of groups  =        595

R-sq:                                           Obs per group:
     within  = 0.6566                                         min =          7
     between = 0.0276                                         avg =        7.0
     overall = 0.0476                                         max =          7

                                                Wald chi2(3)      =    3475.91
corr(u_i, Xb)  = -0.9107                        Prob > chi2       =     0.0000

                                 (Replications based on 595 clusters in id)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .1137879   .0042744    26.62   0.000     .1054102    .1221655
        exp2 |  -.0004244   .0000855    -4.96   0.000     -.000592   -.0002568
         wks |   .0008359   .0008405     0.99   0.320    -.0008115    .0024833
          ed |          0  (omitted)
       _cons |   4.596396   .0733031    62.70   0.000     4.452725    4.740068
-------------+----------------------------------------------------------------
     sigma_u |  1.0362039
     sigma_e |  .15220316
         rho |  .97888036   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- Same as `xtreg lwage exp exp2 wks ed, /// fe vce(boot, reps(400) seed(10101) nodots)`

Panel Jackknife

- An alternative re-sampling scheme is a leave-one-cluster-out jackknife.
- Let $\widehat{\boldsymbol{\beta}}_{-i}$ denote the estimator of $\boldsymbol{\beta}$ when the i^{th} cluster (here i^{th} if N individuals) is deleted

$$\widehat{\mathrm{V}}_{\text{jack;boot}}[\widehat{\boldsymbol{\beta}}] = \frac{N-1}{N} \sum_{i=1}^{N} (\widehat{\boldsymbol{\beta}}_{-i} - \overline{\widehat{\boldsymbol{\beta}}})(\widehat{\boldsymbol{\beta}}_{-i} - \overline{\widehat{\boldsymbol{\beta}}})',$$

where $\overline{\widehat{\boldsymbol{\beta}}} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_{-i}$.

- For Stata `xt` commands this is option `vce(jackknife)`

12. Clustered Cross-section Data

- **xt** commands are not just for panel data
  - can apply several of them to clustered cross-section data.
- Consider data on individual i in village j with **clustering on village**.
- A **cluster-specific model** (here village-specific) specifies

$$y_{ji} = \alpha_i + \mathbf{x}_{ji}'\boldsymbol{\beta} + \varepsilon_{ji}.$$

- Here clustering is on village (not individual) and the repeated measures are over individuals (not time).
- Assuming **equicorrelated errors** can be more reasonable here than with panel data (where correlation dampens over time).
  - So perhaps less need for `vce(robust)` after `xtreg`.
- This is done in the subsequent set of slides.

13. Summary of Stata Panel Commands

- Linear panel estimators for short panels with exogenous regressors

Panel summary	`xtset; xtdescribe; xtsum; xtdata;` `xtline; xttab; xttran`
Pooled OLS	`regress`
Feasible GLS	`xtreg, pa` `xtgee, family(gaussian)`
Random effects	`xtreg, re; xtregar, re`
Fixed effects	`xtreg, fe; xtregar, fe`
Random slopes	`mixed; quadchk`
First differences	`regress` (with differenced data)