ECON60052: Cross Section Econometrics

Lecture 2: Heteroscedasticity and Clustering

Nicky L. Grant

Semester 2

Last lecture...

We talked about the question of causality

We discussed in detail the potential outcome framework

Matching (exact matching) was introduced as a method to compare outcomes for a treated person with a similar person or group of persons in a comparison group

The idea was that conditional on observables or 'matching away observable differences between treatment and control groups', we can analyse the treatment of a policy

Why Worry About Heteroskedasticity?

OLS is unbiased and consistent when zero conditional mean assumptions holds, even with heterokedasticity

The standard errors of the estimates *are* biased if we have heteroskedasticity

If the standard errors are biased, we can not form t statistics or F statistics or LM statistics using standard s.e formulas derived assuming homoskedasticity

Instead we must derive estimators of s.e($\hat{\beta}$)robust to heteroskedasticity ...

Heteroskedasticity

Recall the assumption of homoskedasticity implies that the variance of the unobserved error, u, is constant for all possible values of all the explanatory variables:

$$\text{Var}(u|x_1, \ldots, x_k) = \sigma^2$$

If this is not true, that is if the variance of u is different for different values of the x's, then the errors are **heteroskedastic**

This means our estimate of the variance of $\widehat{\beta}$ will be biased and we cannot use the sum of squared residuals for constructing F statistics

Example (Estimating Returns to Education)

Ability is unobservable in the simple wage/education model

It is likely that the variance in ability differs by educational attainment

In practice, often difficult to assess if there is heteroskedasticity

$\Rightarrow$ We need to test for it and/or correct it...

$\mathrm{Var}(\widehat{\beta})$ with Heteroskedasticity

For the simple regression $y = \beta_0 + \beta_1 x + u$

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum(x_i - \overline{x})u_i}{\sum(x_i - \overline{x})^2}$$

so

$$\mathrm{Var}(\widehat{\beta}_1) = \frac{\sum(x_i - \overline{x})^2 \sigma_i^2}{\mathrm{SST}_x^2} \quad \text{where } \mathrm{SST}_x = \Sigma(x_i - \overline{x})^2$$

When $\sigma_i^2 = \sigma^2$, the formula reduces to usual σ^2/SST_x.
A valid estimator when $\sigma_i^2 \neq \sigma^2$ is

$$\widehat{\mathrm{Var}}(\widehat{\beta}_1) = \frac{\sum(x_i - \overline{x})^2 \widehat{u}_i^2}{\mathrm{SST}_x^2}$$

where $\widehat{u}_i$ are OLS residuals
For the multiple regression case, the formula generalises to:

$$\widehat{\mathrm{Var}}(\widehat{\beta}_j) = \frac{\sum \widehat{r}_{ij}^2 \widehat{u}_i^2}{\mathrm{SSR}_j^2}$$

where $\widehat{r}_{ij}$ is the i-th residual from regressing x_j on all the other independent variables (the part that is uncorrelated with the other covariates), and SSR_j is the sum of squared residuals from this regression

Robust Standard Errors

Now that we have a consistent estimate of the variance, the square root can be used as a standard error for inference

Typically call these **robust standard errors**

Important to remember that these robust standard errors only have asymptotic justification $\Rightarrow$ with small sample sizes t statistics formed with robust standard errors will not have a distribution close to the t, and inferences will not be correct

In Stata, robust standard errors are easily obtained using the **robust** option of **regress**

Example (a Estimating Returns to Education)

```
. regress lwage educ

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  1,   524) =  119.58
       Model |  27.5606288     1  27.5606288            Prob > F      =  0.0000
    Residual |  120.769123   524  .230475425            R-squared     =  0.1858
-------------+------------------------------           Adj R-squared =  0.1843
       Total |  148.329751   525   .28253286            Root MSE      =  .48008

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0827444   .0075667    10.94   0.000     .0678796    .0976091
       _cons |   .5837727   .0973358     6.00   0.000     .3925563    .7749891
------------------------------------------------------------------------------
```

Example (Estimating Returns to Education, cont'd)

```
. regress lwage educ, robust

Linear regression                               Number of obs =     526
                                                F(  1,   524) =  114.32
                                                Prob > F      =  0.0000
                                                R-squared     =  0.1858
                                                Root MSE      =  .48008

------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0827444   .0077389    10.69   0.000     .0675413    .0979475
       _cons |   .5837727   .0982339     5.94   0.000     .3907921    .7767533
------------------------------------------------------------------------------
```

Testing for Heteroskedasticity

We want to test:

$$H_0 : \mathsf{Var}(u|x_1, x_2, \ldots, x_k) = \sigma^2$$

which is equivalent to

$$H_0 : \mathsf{E}(u^2|x_1, x_2, \ldots, x_k) = \mathsf{E}(u^2) = \sigma^2$$

If we assume the relationship between u^2 and x_j is linear, we can estimate:

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v$$

and test

$$H_0 : \delta_1 = \cdots = \delta_k = 0$$

$$H_A : \delta_1 \neq 0 \text{ or} \ldots \delta_k \neq 0$$

Example: Heteroskedasticity with Two Category Variable

For the simple regression with a single dummy variable d_i

$$y_i = \beta_0 + \beta_1 d_i + u_i \quad i = 1, \ldots, n$$

we know that $\widehat{\beta}_1 = \overline{y}_1 - \overline{y}_2$, with $n_1 \equiv \sum d_i$, let $p \equiv n_1/n$ and $n_1 + n_2 \equiv n$

$$\widehat{\mathsf{Var}}(\widehat{\beta}_1)_c = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{\sum_{i\in 1}(y_i - \overline{y}_1)^2 + \sum_{i\in 2}(y_i - \overline{y}_2)^2}{n-2}\right)$$

$$\widehat{\mathsf{Var}}(\widehat{\beta}_1)_r = \frac{\sum_{i\in 1}(y_i - \overline{y}_1)^2}{n_1^2} + \frac{\sum_{i\in 2}(y_i - \overline{y}_2)^2}{n_2^2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

'c' denotes conventional, and sums the squared-deviations. 'r' denotes robust, and basically adds together consistent (but biased) estimates for the 2 sub-samples.

When $s_1 = s_2 = s$, they coincide, obviously.

Less well-known is that when $n_1 = n_2 = n/2$, they also coincide

$\Rightarrow$ when the data are balanced, the robust standard error won't differ much from the traditional one, even when there is heteroskedasticity, $s_1 \neq s_2$

Testing heteroskedasticity with two categories

To interpret the test from above, apply the same algebra to:

$$\widehat{u}_i^2 = \delta_0 + \delta_1 d_i + v$$

$$\widehat{\delta}_0 = \frac{\sum_{i\in 2} \widehat{u}_i^2}{n_2} = \frac{\sum_{i\in 2}(y_i - \overline{y}_2)^2}{n_2} \approx s_2^2$$

$$\widehat{\delta}_1 = \frac{\sum_{i\in 1} \widehat{u}_i^2}{n_1} - \frac{\sum_{i\in 2} \widehat{u}_i^2}{n_2} \approx s_1^2 - s_2^2$$

$\Rrightarrow$ Testing $H_0 : \delta_1 = 0$ is equivalent to testing $\sigma_1^2 = \sigma_2^2$

$\hookrightarrow$ An example with dummy variables follows. *manual* is a dummy for having a manual occupation, *lrhrpay* is log real hourly pay

Example (Occupation and Pay)

```
. tab manual if wave==16, su(lrhrpay)

           | Summary of Log real usual gross pay
           |              per hour
manual occn |        Mean   Std. Dev.       Freq.
------------+------------------------------------
          0 |   2.3402384   .55775906        3048
          1 |   2.1171209   .42925469        1019
------------+------------------------------------
      Total |   2.2843356   .53722719        4067
```

```
. reg lrhrpay manual if wave==16, robust

Linear regression                                      Number of obs =    4067
                                                       F(  1,  4065) =  176.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0324
                                                       Root MSE      =  .52852

------------------------------------------------------------------------------
             |               Robust
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      manual |  -.2231176   .0168172   -13.27   0.000    -.2560884   -.1901467
       _cons |   2.340238   .0101036   231.62   0.000      2.32043    2.360047
------------------------------------------------------------------------------
```

```
reg lrhrpay manual if wave==16

      Source |       SS       df       MS              Number of obs =    4067
-------------+------------------------------           F(  1,  4065) =  136.10
       Model |  38.0174036     1  38.0174036           Prob > F      =  0.0000
    Residual |  1135.48325  4065  .279331674           R-squared     =  0.0324
-------------+------------------------------           Adj R-squared =  0.0322
       Total |  1173.50066  4066  .288613049           Root MSE      =  .52852

------------------------------------------------------------------------------
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      manual |  -.2231176   .0191251   -11.67   0.000    -.2606131    -.185622
       _cons |   2.340238   .0095731   244.46   0.000      2.32147    2.359007
```

```
. predict uhat if wave==16, res

. gen uhatsq=uhat*uhat if wave==16

. reg uhatsq manual if wave==16

      Source |       SS       df       MS              Number of obs =    4067
-------------+------------------------------           F(  1,  4065) =   44.33
       Model |  12.3008868     1  12.3008868            Prob > F      =  0.0000
    Residual |  1127.87929  4065   .27746108            R-squared     =  0.0108
-------------+------------------------------           Adj R-squared =  0.0105
       Total |  1140.18018  4066  .280418145            Root MSE      =  .52675

------------------------------------------------------------------------------
      uhatsq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      manual |  -.1269143   .0190609    -6.66   0.000    -.1642842   -.0895445
       _cons |   .3109931    .009541    32.60   0.000     .2922875    .3296987
------------------------------------------------------------------------------
```

Manual workers earn 0.223 log-points less, on average

There are far fewer manual workers in this data: $p = 1019/4067 = 0.251$

The histograms suggest that the variance of the two group's hourly wages might be different

The test of equal variances is rejected with a t-statistic of –6.66. ($s_1^2 - s_2^2 \approx -0.127$)

When testing null hypothesis of equal pay, the robust se is 0.0168 compared with a conventional se of 0.0191 $\Rrightarrow$ a 13% difference (not too much)

Wooldridge says . . .

Wooldridge (Chapter 8) discusses other tests for heteroskedasticity

He also discusses WLS and feasible GLS. Unless one knows that the heteroskedasticity has a particular form, it is safer to simply compute Robust SEs

$\Rightarrow$ using WLS etc is no longer used much

Computing Robust SEs is easy but have asymptotic justification

The clustering problem

The key assumption behind correct inference is that the data are independent of each other

- The error terms are not correlated with each other

But this is often not true in practice. Most important form of dependence comes when the data are in a group structure

Exam grades of children in the same class or school

Grades are correlated because of the same school (teacher) and family background

Earnings in the same region are correlated because of the same industrial structure

Workers in the same firms will be subject to common firm effects (earnings/tenure/promotion)

Classic panel data: observe individuals for T periods.

Pooling the data across individuals (as if 2 x-sections) and apply OLS ignores that the errors are correlated for each individual...

This correlation is known as **clustering** and its effect the **Moulton problem**

The problem is that ignoring the correlation within the cluster (usually) leads to *estimates of the variance* that are biased downwards.

A closely related problem is correlation over time. Occurs with Pooled Cross Sections and Panel Data.

Clustering and the Moulton factor

Often we use cross-section or panel data to analyse the effect of a macro variable on an individual-level outcome

- Effect of school-type on exam grades
- Effect of industrial structure on firm's productivity
- Effect of a firm's profits on its workers' wages
- Effect of regional unemployment on individuals' wages

The data are 'matched in' to the aggregate indicator (e.g. regional unemployment to the individual's region)

This model can be effectively written as:

$$y_{ig} = \beta_0 + \beta_1 x_g + e_{ig} \quad g = 1, \ldots, G \quad i = 1, \ldots, N$$

x_g only varies at the group level

Data example

i	g	y_{ig}	1	x_g
1	1	y_{11}	1	x_1
2	1	y_{21}	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_1	1		1	x_1
1	2	y_{12}	1	x_2
2	2	y_{22}	1	x_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_2	2		1	x_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	G	y_{1G}	1	x_G
2	G	y_{2G}	1	x_G
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_G	G		1	x_G

The problem

Because x only varies by g, the error structure is

$$e_{ig} \equiv v_g + \eta_{ig}$$

$\Rightarrow$ This can increase the standard errors sharply.

Consider e_{ig} and e_{jg} for two individuals in the same group:

$$\mathsf{Cov}(e_{ig}, e_{jg}) = \mathsf{E}[(v_g + \eta_{ig})(v_g + \eta_{jg})] - \mathsf{E}[(v_g + \eta_{ig})]\mathsf{E}[(v_g + \eta_{jg})] = \mathsf{Var}(v_g) = \sigma_v^2$$

$$\mathsf{Var}(e_{ig}) = \mathsf{E}[(v_g + \eta_{ig})^2] - [\mathsf{E}[(v_g + \eta_{ig})]^2 = \sigma_v^2 + \sigma_\eta^2$$

where $\sigma_v^2 = \mathsf{Var}(v_g)$ and $\sigma_\eta^2 = \mathsf{Var}(\eta_{ig})$

So the intraclass correlation is:

$$\rho_e \equiv \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2};$$

$\Rightarrow$ it is the proportion of the variation in $v + \eta$ due to v

The Moulton factor

It can be shown that the Moulton factor is (c means conventional here) with regressors fixed at the group level and groups of equal size (n):

$$\frac{\text{Var}(\widehat{\beta}_1)}{\text{Var}_c(\widehat{\beta}_1)} = 1 + (n-1)\rho_e$$

Measures how much we overestimate precision because we are ignoring the intraclass correlation

Conventional standard errors $\text{Var}_c(\widehat{\beta}_1)$ become increasingly misleading when n and ρ_e increase

- When $\rho_e = 0$ there is no problem...
- If $\rho_e = 1$ (or $\eta_{ig} = 0$) all y_{ig}s are the same within a group
- Making a dataset larger by copying each obs n times generates no new information $\Rightarrow$ $\text{Var}_c(\widehat{\beta}_1)$ should be scaled up n times!
- Moulton factor increases with group size: increasing group size reduces the number of groups and data is independent across but not within groups

Moulton factor: example

The Moulton factor can be large

Say 4000 students (individuals) of 40 schools (groups) leads to an average $n = 100$

Interest lies in a school-level treatment indicator (all students in the school either treated or not)

Even with smallish ρ_e, say 0.1, we get

$$1 + \rho_e(n-1) = 1 + 0.1 \times (100 - 1) = \sqrt{10.9} = 3.3$$

↪ 'Standard' standard errors are only about one-third of what they should be!

Groups of different size and x vary

When x varies by g *and* i:

$$\frac{\text{Var}(\widehat{\beta}_1)}{\text{Var}_c(\widehat{\beta}_1)} = 1 + (n-1)\rho_e\rho_x$$

ρ_x is the within-cluster correlation coefficient for x (no variance components structure like before)

- it (sort of) decomposes x into ig and g components
- if x_{ig} is uncorrelated within groups, the problem disappears

The most general form takes into account variations in group size n_g:

$$\frac{\text{Var}(\widehat{\beta}_1)}{\text{Var}_c(\widehat{\beta}_1)} = 1 + \left[\frac{\text{Var}(n_g)}{\overline{n}} + \overline{n} - 1\right]\rho_e\rho_x$$

↪ Suppose i refers to workers, g to firms, then n_g means firm-size, which varies a lot...and $\overline{n}$ is average firm size...

Class exercise (solutions on Blackboard)

↪ Consider the following population model for worker i employed at firm g

↪ You want to analyse whether higher profits lead to higher earnings:

$$y_{ig} = \beta_0 + \beta_1 x_g + v_g + \eta_{ig} \quad g = 1, \ldots, G, \quad i = 1, \ldots, N$$

↪ There are G firms of sizes n_1, n_2, …, n_G; y_{ig} is the worker's earnings, x_g is the firm's profits

↪ η_{ig} is the worker-level error component ("unobserved productivity") and the error component v_g ("managerial know-how") is constant within firms

Class exercise ctd. (solutions on Blackboard)

1. Explain why Assumption MLR5 does not hold. Consider Var$(\widehat{\beta}_1)$, extracted from the traditional covariance matrix. Write down a well-known correction factor for dealing with this problem and explain all its elements.
2. What is this correction factor when $N = 1100$, $G = 100$, Var$(n_g) = 11$, $\sigma_v^2 = 4$ and $\sigma_\eta^2 = 7$?
3. How do your answers change if $v_g = 0$ for all firms? Explain in your own words the implications of this.

So what to do?

The model is

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + v_g + \eta_{ig} \quad g = 1, \ldots, G. \tag{1}$$

Because the key assumption is error independence *across* clusters, most solutions require the number of clusters to be 'large'

So these solutions aren't suitable for analysing variations in outcomes, such as pay, across 10 regions in a single cross-section...

Angrist and Pischke's solutions

Solution 1. Parametric

Fix conventional standard errors using AP's formula (8.2.5)

Use Stata's `loneway` command to estimate ρ_e and ρ_x

Solution 2. Use clustered standard errors

Use a cluster-robust estimator of the covariance matrix is, which is in line with standard robust version but also adjust for the fact that you have clusters

- ↪ In Stata, type **reg y x, cluster(g)**.
- ↪ Key point: these standard errors have asymptotic justification and are just valid if the *number of clusters* is large

Solution 3. Use group averages

Use weighted group data instead of microdata:

$$\bar{y}_g = \beta_0 + \beta_1 x_g + \bar{e}_g$$

by WLS, using n_g as weights. This simple solution:

- $\hookrightarrow$ does not work if x varies by i as well as g
- $\hookrightarrow$ ignores any other micro-level covariates
- $\hookrightarrow$ this can work if G is small

An example: the wage curve

The elasticity of real wages with respect to the unemployment rate is –0.10. There is a lot of supporting evidence on this...

Let's take a look at the BHPSWaves2-16

There are 11 regions (London, South East, South West, etc) and 15 years, so $G = 165$ clusters or region/years

There are: n_1 individuals in region 1/1992, n_2 individuals in region 2/1992, ..., n_G individuals in region 11/2006, totalling 65061 observations

Then we can take a look at AP's suggestions

Doing nothing, plain OLS

```
* conventional

. reg lrhrpay lurate if regsam==1

      Source |       SS       df       MS              Number of obs =   65061
-------------+------------------------------           F(  1, 65059) = 1816.78
       Model |  546.280371     1  546.280371            Prob > F      =  0.0000
    Residual |  19562.3267 65059  .300685942            R-squared     =  0.0272
-------------+------------------------------           Adj R-squared =  0.0272
       Total |  20108.6071 65060  .309077883           Root MSE      =   .54835

------------------------------------------------------------------------------
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lurate |  -.1671298   .0039211   -42.62   0.000     -.174815   -.1594445
       _cons |   -2.27937   .0055176  -413.11   0.000    -2.290185   -2.268556
------------------------------------------------------------------------------

scalar convSE = _se[lurate]

.  predict uhat, res
(1120 missing values generated)
```

Robust standard errors

```
* robust

. reg lrhrpay lurate if regsam==1, robust

Linear regression                               Number of obs =   65061
                                                F(  1, 65059) = 1798.97
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0272
                                                Root MSE      =  .54835

------------------------------------------------------------------------------
             |               Robust
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lurate |  -.1671298   .0039404   -42.41   0.000     -.174853   -.1594066
       _cons |   -2.27937   .0054741  -416.39   0.000    -2.290099   -2.268641
------------------------------------------------------------------------------
```

Cluster correct standard errors

```
* clustered AP's (8.2.6) and Solution 2

. reg lrhrpay lurate if regsam==1, cluster(regwave)

Linear regression                                   Number of obs =   65061
                                                    F(  1,   164) =   94.50
                                                    Prob > F      =  0.0000
                                                    R-squared     =  0.0272
                                                    Root MSE      =  .54835

                          (Std. Err. adjusted for 165 clusters in regwave)
------------------------------------------------------------------------------
             |               Robust
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lurate |  -.1671298   .0171922    -9.72   0.000    -.2010764   -.1331831
       _cons |   -2.27937   .0222698  -102.35   0.000    -2.323343   -2.235398
------------------------------------------------------------------------------
```

Adjusting for Moulton factor

```
* AP's Solution 1

. loneway uhat regwave

                One-way Analysis of Variance for uhat: Residuals

                                             Number of obs =     65061
                                                 R-squared =    0.0469

    Source                 SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between regwave        916.83715    164    5.5904704     19.46     0.0000
Within regwave          18645.49  64896    .28731339
------------------------------------------------------------------------------
Total                  19562.327  65060    .30068132

         Intraclass       Asy.
         correlation      S.E.       [95% Conf. Interval]
         -------------------------------------------------
  rho_e= 0.04477       0.00541       0.03417     0.05537
```

```
. scalar rho_e=r(rho)

. scalar rho_x=1

. scalar Var_ng=180.2965^2   //  can only be computed AFTER data are collapsed

. scalar nbar=65061/165

. di nbar
394.30909

* AP (8.2.5)
. scalar moulton = sqrt(1+(Var_ng/nbar+nbar-1)*rho_e*rho_x)

. di moulton
4.7221593

. di moulton*convSE
.01851583
```

```
. * AP's Solution 3

. * collapse data to 165 region/wave "clusters"
.
. collapse lrhrpay lurate regwaveN, by(wave region)

. su wave region lrhrpay lurate regwaveN

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wave |       165           9    4.333646          2         16
      region |       165    9.454545    5.712048          2         18
     lrhrpay |       165   -2.511969    .1517321  -2.753511  -2.042849
      lurate |       165    1.326503    .5225828   .2623643   2.370244
    regwaveN |       165    401.097     180.2965        155        964
                            =nbar      =se(n_g)
```

```
. reg lrhrpay lurate [pweight=regwaveN]  // WLS with n_g as weights
(sum of wgt is   6.6181e+04)

Linear regression                                      Number of obs =     165
                                                       F(  1,   163) =   93.96
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3743
                                                       Root MSE      =  .11939

------------------------------------------------------------------------------
             |               Robust
     lrhrpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lurate |  -.1672094   .0172501    -9.69   0.000    -.2012719   -.1331469
       _cons |  -2.279388   .0223493  -101.99   0.000    -2.323519   -2.235257
------------------------------------------------------------------------------
```

Conclusions on the Wage Curve

obs	est	st err	
65061	–0.1671	0.00392	OLS, conventional
65061	–0.1671	0.00394	OLS, robust
65061	–0.1671	0.01719	OLS, cluster(g)
65061	–0.1671	0.01852	multiply conv by Moulton=4.7222
165	–0.1672	0.01725	WLS

↪ It would appear that we should throw the individual-level data away...

↪ But, much more needs doing to the individual-level ests: adding covariates, regional and time dummies....

Readings

↪ Clustering: Angrist & Pischke, Section 8.2.1.

↪ Heteroskedasticity:

- Angrist & Pischke, Section 8.1. (dummy variable example)
- Wooldridge, Chapter 8