16. Clustered and Panel Data

Spring 2021

Matthew Blackwell

Gov 2002 (Harvard)

Where are we? Where are we going?

- Focus up until now on iid data, but often doesn't hold.

Where are we? Where are we going?

- Focus up until now on iid data, but often doesn't hold.
- **Panel** and **clustered** data are two common non-iid data.

Where are we? Where are we going?

- Focus up until now on iid data, but often doesn't hold.
- **Panel** and **clustered** data are two common non-iid data.
- Panel data also holds hope for removing unmeasured heterogeneity.

1/ Panel Data

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
  - they are richer or developed earlier
  - provide benefits more efficiently

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
    - they are richer or developed earlier
    - provide benefits more efficiently
    - posses some cultural trait correlated with better health outcomes

Motivation

Is Democracy Good for the Poor?

Michael Ross University of California, Los Angeles

- Relationship between democracy and infant mortality?
- Compare levels of democracy with levels of infant mortality, but...
- Democratic countries are different from non-democracies in ways that we can't measure?
    - they are richer or developed earlier
    - provide benefits more efficiently
    - posses some cultural trait correlated with better health outcomes
- If have data on countries over time, can we make any progress in spite of these problems?

Ross data

```r
ross <- foreign::read.dta("../assets/ross-democracy.dta")
head(ross[,c("cty_name", "year", "democracy", "infmort_unicef")])
```

```
##      cty_name year democracy infmort_unicef
## 1 Afghanistan 1965         0            230
## 2 Afghanistan 1966         0             NA
## 3 Afghanistan 1967         0             NA
## 4 Afghanistan 1968         0             NA
## 5 Afghanistan 1969         0             NA
## 6 Afghanistan 1970         0            215
```

Notation for panel data

- Units, $i = 1, \dots, n$

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Time is a typical application, but applies to other groupings:

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Time is a typical application, but applies to other groupings:
    - counties within states

Notation for panel data

- Units, $i = 1, \ldots, n$
- Time, $t = 1, \ldots, T$
- Time is a typical application, but applies to other groupings:
    - counties within states
    - states within countries

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Time is a typical application, but applies to other groupings:
  - counties within states
  - states within countries
  - people within coutries, etc.

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Time is a typical application, but applies to other groupings:
    - counties within states
    - states within countries
    - people within coutries, etc.
- **Panel data**: large n, relatively short T

Notation for panel data

- Units, $i = 1, \dots, n$
- Time, $t = 1, \dots, T$
- Time is a typical application, but applies to other groupings:
    - counties within states
    - states within countries
    - people within coutries, etc.
- **Panel data**: large n, relatively short T
- **Time series, cross-sectional (TSCS) data**: smaller n, large T (a political science term, mostly)

Model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)

Model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)
- c_i is an **unobserved** time-constant unit effect ("fixed effect")

Model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)
- c_i is an **unobserved** time-constant unit effect ("fixed effect")
    - Confusingly, we'll allow them to be random variables.

Model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)
- c_i is an **unobserved** time-constant unit effect ("fixed effect")
    - Confusingly, we'll allow them to be random variables.
- u_{it} are the unobserved time-varying "idiosyncratic" errors

Model

$$Y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)
- c_i is an **unobserved** time-constant unit effect ("fixed effect")
    - Confusingly, we'll allow them to be random variables.
- u_{it} are the unobserved time-varying "idiosyncratic" errors
- $v_{it} = c_i + u_{it}$ is the combined unobserved error: $Y_{it} = \mathbf{X}'_{it}\boldsymbol{\beta} + v_{it}$

Model

$$Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}$$

- $\mathbf{X}_{it}$ is a vector of covariates (possibly time-varying)
- c_i is an **unobserved** time-constant unit effect ("fixed effect")
    - Confusingly, we'll allow them to be random variables.
- u_{it} are the unobserved time-varying "idiosyncratic" errors
- $v_{it} = c_i + u_{it}$ is the combined unobserved error: $Y_{it} = \mathbf{X}_{it}'\boldsymbol{\beta} + v_{it}$
- Assume that if we could measure c_i, we would have the correct CEF:

$$\mathbb{E}[u_{it} \mid \mathbf{X}_{it}, c_i] = 0 \quad \Longrightarrow \quad \mathbb{E}[Y_{it} \mid \mathbf{X}_{it}, c_i] = \mathbf{X}_{it}'\boldsymbol{\beta} + c_i$$

Pooled OLS

- **Pooled OLS**: pool all observations into one regression

Pooled OLS

- **Pooled OLS**: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.

Pooled OLS

- **Pooled OLS**: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:

Pooled OLS

- **Pooled OLS**: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:
    1. Variance is probably wrong if there is dependence over time

Pooled OLS

- **Pooled OLS**: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:
  1. Variance is probably wrong if there is dependence over time
  2. Errors might be correlated with the covariates

Pooled OLS

- **Pooled OLS**: pool all observations into one regression
- Treats all unit-periods (each it) as an iid unit.
- Has two problems:
    1. Variance is probably wrong if there is dependence over time
    2. Errors might be correlated with the covariates
- Both problems arise out of ignoring the **unmeasured heterogeneity** inherent in c_i

Pooled OLS with Ross data

```r
pooled.mod <- lm(log(kidmort_unicef) ~ democracy + log(GDPcur),
                 data = ross)
summary(pooled.mod)
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.7640     0.3449    28.3   <2e-16 ***
## democracy    -0.9552     0.0698   -13.7   <2e-16 ***
## log(GDPcur)  -0.2283     0.0155   -14.8   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.795 on 646 degrees of freedom
##   (5773 observations deleted due to missingness)
## Multiple R-squared:  0.504,  Adjusted R-squared:  0.503
## F-statistic:  329 on 2 and 646 DF,  p-value: <2e-16
```

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it}u_{it}] = 0$.

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it} u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it}u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it}v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it} u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it} v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
    - Just run pooled OLS (but worry about SEs).

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it} u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it} v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
    - Just run pooled OLS (but worry about SEs).
- But c_i often correlated with $\mathbf{X}_{it}$ so that $\mathbb{E}[\mathbf{X}_{it} c_i] \neq 0$.

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it}u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
  - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it}v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
  - Just run pooled OLS (but worry about SEs).
- But c_i often correlated with $\mathbf{X}_{it}$ so that $\mathbb{E}[\mathbf{X}_{it}c_i] \neq 0$.
  - Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it}u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it}v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
    - Just run pooled OLS (but worry about SEs).
- But c_i often correlated with $\mathbf{X}_{it}$ so that $\mathbb{E}[\mathbf{X}_{it}c_i] \neq 0$.
    - Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.
    - Ignore the heterogeneity $\rightsquigarrow$ correlation between the combined error and the independent variables.

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it} u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it} v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
    - Just run pooled OLS (but worry about SEs).
- But c_i often correlated with $\mathbf{X}_{it}$ so that $\mathbb{E}[\mathbf{X}_{it} c_i] \neq 0$.
    - Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.
    - Ignore the heterogeneity $\rightsquigarrow$ correlation between the combined error and the independent variables.
    - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it} v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = \mathbb{E}[\mathbf{X}_{it} c_i] \neq 0$

Unmeasured heterogeneity

- Since u_{it} is the CEF error, $\mathbf{X}_{it}$ are uncorrelated with it: $\mathbb{E}[\mathbf{X}_{it}u_{it}] = 0$.
- If unit-effect c_i is uncorrelated with $\mathbf{X}_{it}$, no problem for consistency!
  - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it}v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = 0$.
  - Just run pooled OLS (but worry about SEs).
- But c_i often correlated with $\mathbf{X}_{it}$ so that $\mathbb{E}[\mathbf{X}_{it}c_i] \neq 0$.
  - Example: democratic institutions correlated with unmeasured aspects of health outcomes, like quality of health system or a lack of ethnic conflict.
  - Ignore the heterogeneity $\rightsquigarrow$ correlation between the combined error and the independent variables.
  - $\rightsquigarrow \mathbb{E}[\mathbf{X}_{it}v_{it}] = \mathbb{E}[\mathbf{X}_{it}(c_i + u_{it})] = \mathbb{E}[\mathbf{X}_{it}c_i] \neq 0$
- Pooled OLS will be inconsistent for the CEF parameters, $\boldsymbol{\beta}$.

Strict exogeneity

- Panel data allows us to estimate $\boldsymbol{\beta}$ even in this setting

Strict exogeneity

- Panel data allows us to estimate $\boldsymbol{\beta}$ even in this setting
- Two approaches that leverage repeated observations:

Strict exogeneity

- Panel data allows us to estimate $\boldsymbol{\beta}$ even in this setting
- Two approaches that leverage repeated observations:
  - **Differencing** look at changes over time.

Strict exogeneity

- Panel data allows us to estimate $\boldsymbol{\beta}$ even in this setting
- Two approaches that leverage repeated observations:
  - **Differencing** look at changes over time.
  - **Fixed effects** look at relationships within units.

2/ First Differencing Methods

First differencing

- One approach: compare **changes over time**

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.
- Two time periods:

$$Y_{i1} = \mathbf{X}_{i1}'\boldsymbol{\beta} + c_i + u_{i1}$$
$$Y_{i2} = \mathbf{X}_{i2}'\boldsymbol{\beta} + c_i + u_{i2}$$

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.
- Two time periods:

$$Y_{i1} = \mathbf{X}_{i1}'\boldsymbol{\beta} + c_i + u_{i1}$$
$$Y_{i2} = \mathbf{X}_{i2}'\boldsymbol{\beta} + c_i + u_{i2}$$

- Look at the change in Y over time:

$$\Delta Y_i = Y_{i2} - Y_{i1}$$

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.
- Two time periods:

$$Y_{i1} = \mathbf{X}_{i1}'\boldsymbol{\beta} + c_i + u_{i1}$$
$$Y_{i2} = \mathbf{X}_{i2}'\boldsymbol{\beta} + c_i + u_{i2}$$

- Look at the change in Y over time:

$$\begin{aligned}\Delta Y_i &= Y_{i2} - Y_{i1} \\ &= (\mathbf{X}_{i2}'\boldsymbol{\beta} + c_i + u_{i2}) - (\mathbf{X}_{i1}'\boldsymbol{\beta} + c_i + u_{i1})\end{aligned}$$

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.
- Two time periods:

$$Y_{i1} = \mathbf{X}'_{i1}\boldsymbol{\beta} + c_i + u_{i1}$$
$$Y_{i2} = \mathbf{X}'_{i2}\boldsymbol{\beta} + c_i + u_{i2}$$

- Look at the change in Y over time:

$$\begin{aligned}\Delta Y_i &= Y_{i2} - Y_{i1} \\ &= (\mathbf{X}'_{i2}\boldsymbol{\beta} + c_i + u_{i2}) - (\mathbf{X}'_{i1}\boldsymbol{\beta} + c_i + u_{i1}) \\ &= (\mathbf{X}'_{i2} - \mathbf{X}'_{i1})\boldsymbol{\beta} + (c_i - c_i) + (u_{i2} - u_{i1})\end{aligned}$$

First differencing

- One approach: compare **changes over time**
- Intuitively, time-constant heterogeneity can't affect changes over time.
- Two time periods:

$$\begin{aligned} Y_{i1} &= \mathbf{X}'_{i1}\boldsymbol{\beta} + c_i + u_{i1} \\ Y_{i2} &= \mathbf{X}'_{i2}\boldsymbol{\beta} + c_i + u_{i2} \end{aligned}$$

- Look at the change in Y over time:

$$\begin{aligned} \Delta Y_i &= Y_{i2} - Y_{i1} \\ &= (\mathbf{X}'_{i2}\boldsymbol{\beta} + c_i + u_{i2}) - (\mathbf{X}'_{i1}\boldsymbol{\beta} + c_i + u_{i1}) \\ &= (\mathbf{X}'_{i2} - \mathbf{X}'_{i1})\boldsymbol{\beta} + (c_i - c_i) + (u_{i2} - u_{i1}) \\ &= \Delta\mathbf{X}'_i\boldsymbol{\beta} + \Delta u_i \end{aligned}$$

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.
- For consistency of OLS on the differences, we need $\mathbb{E}[\Delta \mathbf{X}_i \Delta u_i] = 0$.

$$\mathbb{E}[(\mathbf{X}_{i2} - \mathbf{X}_{i1})(u_{i2} - u_{i1})] = \mathbb{E}[\mathbf{X}_2 u_2] + \mathbb{E}[\mathbf{X}_1 u_1] - \mathbb{E}[\mathbf{X}_1 u_2] - \mathbb{E}[\mathbf{X}_2 u_1] = 0$$

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.
- For consistency of OLS on the differences, we need $\mathbb{E}[\Delta \mathbf{X}_i \Delta u_i] = 0$.

$$\mathbb{E}[(\mathbf{X}_{i2} - \mathbf{X}_{i1})(u_{i2} - u_{i1})] = \mathbb{E}[\mathbf{X}_2 u_2] + \mathbb{E}[\mathbf{X}_1 u_1] - \mathbb{E}[\mathbf{X}_1 u_2] - \mathbb{E}[\mathbf{X}_2 u_1] = 0$$

  - First two are 0 since we assume the CEF is correctly specified up to c_i

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.
- For consistency of OLS on the differences, we need $\mathbb{E}[\Delta \mathbf{X}_i \Delta u_i] = 0$.

$$\mathbb{E}[(\mathbf{X}_{i2} - \mathbf{X}_{i1})(u_{i2} - u_{i1})] = \mathbb{E}[\mathbf{X}_2 u_2] + \mathbb{E}[\mathbf{X}_1 u_1] - \mathbb{E}[\mathbf{X}_1 u_2] - \mathbb{E}[\mathbf{X}_2 u_1] = 0$$

  - First two are 0 since we assume the CEF is correctly specified up to c_i
  - $\mathbb{E}[\mathbf{X}_1 u_2]$ and $\mathbb{E}[\mathbf{X}_2 u_1]$ are additional assumptions: no **feedback between outcome and covariates**

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.
- For consistency of OLS on the differences, we need $\mathbb{E}[\Delta \mathbf{X}_i \Delta u_i] = 0$.

$$\mathbb{E}[(\mathbf{X}_{i2} - \mathbf{X}_{i1})(u_{i2} - u_{i1})] = \mathbb{E}[\mathbf{X}_2 u_2] + \mathbb{E}[\mathbf{X}_1 u_1] - \mathbb{E}[\mathbf{X}_1 u_2] - \mathbb{E}[\mathbf{X}_2 u_1] = 0$$

  - First two are 0 since we assume the CEF is correctly specified up to c_i
  - $\mathbb{E}[\mathbf{X}_1 u_2]$ and $\mathbb{E}[\mathbf{X}_2 u_1]$ are additional assumptions: no **feedback between outcome and covariates**
- Invertibility of $\mathbb{E}[\Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}']$ requires $\mathbf{X}_{it}$ to vary over time for someone

First differences model

$$\Delta Y_i = \Delta \mathbf{X}_i' \boldsymbol{\beta} + \Delta u_i$$

- Coefficient on the levels $\mathbf{X}_{it}$ = the coefficient on the changes $\Delta \mathbf{X}_i$
- Time-constant unobserved heterogeneity c_i drops out.
- For consistency of OLS on the differences, we need $\mathbb{E}[\Delta \mathbf{X}_i \Delta u_i] = 0$.

$$\mathbb{E}[(\mathbf{X}_{i2} - \mathbf{X}_{i1})(u_{i2} - u_{i1})] = \mathbb{E}[\mathbf{X}_2 u_2] + \mathbb{E}[\mathbf{X}_1 u_1] - \mathbb{E}[\mathbf{X}_1 u_2] - \mathbb{E}[\mathbf{X}_2 u_1] = 0$$

  - First two are 0 since we assume the CEF is correctly specified up to c_i
  - $\mathbb{E}[\mathbf{X}_1 u_2]$ and $\mathbb{E}[\mathbf{X}_2 u_1]$ are additional assumptions: no **feedback between outcome and covariates**
- Invertibility of $\mathbb{E}[\Delta \mathbf{X}_{it} \Delta \mathbf{X}_{it}']$ requires $\mathbf{X}_{it}$ to vary over time for someone
- Under these assumptions, pooled OLS on the differences is consistent.

3/ Fixed Effects Methods

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means
- First note that taking the average of the Y's over time for a given unit leaves us with a very similar model:

$$\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}\right]$$

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means
- First note that taking the average of the Y's over time for a given unit leaves us with a very similar model:

$$
\begin{aligned}
\overline{Y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it}'\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}c_i + \frac{1}{T}\sum_{t=1}^{T}u_{it}
\end{aligned}
$$

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means
- First note that taking the average of the Y's over time for a given unit leaves us with a very similar model:

$$
\begin{aligned}
\overline{Y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{X}'_{it}\boldsymbol{\beta} + c_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}'_{it}\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}c_i + \frac{1}{T}\sum_{t=1}^{T}u_{it} \\
&= \overline{\mathbf{X}}'_i\boldsymbol{\beta} + c_i + \overline{u}_i
\end{aligned}
$$

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means
- First note that taking the average of the Y's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}
\overline{Y}_i &= \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}\right] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it}'\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T}c_i + \frac{1}{T}\sum_{t=1}^{T}u_{it} \\
&= \overline{\mathbf{X}}_i'\boldsymbol{\beta} + c_i + \overline{u}_i
\end{aligned}$$

- Key fact: mean of the time-constant c_i is just c_i

Fixed effects models

- **Fixed effects model**: another way to remove unmeasured heterogeneity
- Focuses on **within-unit comparisons**: changes in Y_{it} and X_{it} relative to their within-group means
- First note that taking the average of the Y's over time for a given unit leaves us with a very similar model:

$$\begin{aligned}
\overline{Y}_i &= \frac{1}{T}\sum_{t=1}^{T}[\mathbf{X}_{it}'\boldsymbol{\beta} + c_i + u_{it}] \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_{it}'\right)\boldsymbol{\beta} + \frac{1}{T}\sum_{t=1}^{T} c_i + \frac{1}{T}\sum_{t=1}^{T} u_{it} \\
&= \overline{\mathbf{X}}_i'\boldsymbol{\beta} + c_i + \overline{u}_i
\end{aligned}$$

- Key fact: mean of the time-constant c_i is just c_i
- This regression is sometimes called the “between regression”

Within transformation

- **Fixed effect** or **within transformation**:

$$(Y_{it} - \overline{Y}_i) = (\mathbf{X}'_{it} - \overline{\mathbf{X}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

Within transformation

- **Fixed effect** or **within transformation**:

$$(Y_{it} - \overline{Y}_i) = (\mathbf{X}'_{it} - \overline{\mathbf{X}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

  - Center every covariate and the outcome at its within-unit mean.

Within transformation

- **Fixed effect** or **within transformation**:

$$(Y_{it} - \overline{Y}_i) = (\mathbf{X}'_{it} - \overline{\mathbf{X}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

  - Center every covariate and the outcome at its within-unit mean.
  - c_i drops out because its within-unit mean is itself (time-constant).

Within transformation

- **Fixed effect** or **within transformation**:

$$(Y_{it} - \overline{Y}_i) = (\mathbf{X}'_{it} - \overline{\mathbf{X}}'_i)\boldsymbol{\beta} + (u_{it} - \overline{u}_i)$$

  - Center every covariate and the outcome at its within-unit mean.
  - c_i drops out because its within-unit mean is itself (time-constant).

- If we write $\ddot{Y}_{it} = Y_{it} - \overline{Y}_i$, then we can write this more compactly as:

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

Fixed effects with Ross data

```r
library(lfe)
fe.mod <- lfe::felm(log(kidmort_unicef) ~ democracy + log(GDPcur) | id, data = ross)
summary(fe.mod)
```

```
##
## Call:
##    lfe::felm(formula = log(kidmort_unicef) ~ democracy + log(GDPcur) |      id, data = ross)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7049 -0.1166  0.0063  0.1222  0.7575
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## democracy    -0.1432     0.0335   -4.28 0.000023 ***
## log(GDPcur)  -0.3752     0.0113  -33.12  < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.219 on 481 degrees of freedom
##   (5773 observations deleted due to missingness)
## Multiple R-squared(full model): 0.972   Adjusted R-squared: 0.962
## Multiple R-squared(proj model): 0.718   Adjusted R-squared: 0.621
## F-statistic(full model): 100 on 167 and 481 DF, p-value: <2e-16
## F-statistic(proj model):  613 on 2 and 481 DF, p-value: <2e-16
```

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.
    - Like with differencing, need u_{it} to be uncorrelated with all $\mathbf{X}_{is}$

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.
    - Like with differencing, need u_{it} to be uncorrelated with all $\mathbf{X}_{is}$
    - Why? $\ddot{u}_{it}$ and $\ddot{\mathbf{X}}_{it}$ are functions of errors/covariates in **all time periods**.

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.
    - Like with differencing, need u_{it} to be uncorrelated with all $\mathbf{X}_{is}$
    - Why? $\ddot{u}_{it}$ and $\ddot{\mathbf{X}}_{it}$ are functions of errors/covariates in **all time periods**.
- Key assumption is **strict exogeneity**:

$$\mathbb{E}[u_{it}|\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{iT}, c_i] = 0$$

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}_{it}'\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.
    - Like with differencing, need u_{it} to be uncorrelated with all $\mathbf{X}_{is}$
    - Why? $\ddot{u}_{it}$ and $\ddot{\mathbf{X}}_{it}$ are functions of errors/covariates in **all time periods**.
- Key assumption is **strict exogeneity**:

$$\mathbb{E}[u_{it}|\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{iT}, c_i] = 0$$

    - u_{it} uncorrelated with all covariates for unit i at any point in time.

Strict exogeneity

$$\ddot{Y}_{it} = \ddot{\mathbf{X}}'_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

- To use OLS on demeaned data, need $\mathbb{E}[\ddot{\mathbf{X}}_{it}\ddot{u}_{it}] = 0$.
- This is not implied by $\mathbb{E}[u_{it}|\mathbf{X}_{it}, c_i] = 0$.
    - Only implies u_{it} will be uncorrelated with $\mathbf{X}_{it}$.
    - Like with differencing, need u_{it} to be uncorrelated with all $\mathbf{X}_{is}$
    - Why? $\ddot{u}_{it}$ and $\ddot{\mathbf{X}}_{it}$ are functions of errors/covariates in **all time periods**.
- Key assumption is **strict exogeneity**:

$$\mathbb{E}[u_{it}|\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{iT}, c_i] = 0$$

    - u_{it} uncorrelated with all covariates for unit i at any point in time.
    - Rules out lagged dependent variables, since $Y_{i,t-1}$ is a function of $u_{i,t-1}$.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
    - $\rightsquigarrow X_{it} = \overline{X}_i$ and $\ddot{X}_{it} = 0$ for all periods t.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
    - $\rightsquigarrow X_{it} = \overline{X}_i$ and $\ddot{X}_{it} = 0$ for all periods t.
- If $\ddot{X}_{it} = 0$ for all i and t, violates invertibility.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
  - $\rightsquigarrow X_{it} = \overline{X}_i$ and $\ddot{X}_{it} = 0$ for all periods t.
- If $\ddot{X}_{it} = 0$ for all i and t, violates invertibility.
  - R/Stata and the like will drop it from the regression.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
  - $\rightsquigarrow X_{it} = \overline{X}_i$ and $\ddot{X}_{it} = 0$ for all periods t.
- If $\ddot{X}_{it} = 0$ for all i and t, violates invertibility.
  - R/Stata and the like will drop it from the regression.
  - Any time-constant variable gets "absorbed" by the fixed effect.

Fixed effects and time-invariant covariates

- What if there is a covariate that doesn't vary over time?
  - $\rightsquigarrow X_{it} = \overline{X}_i$ and $\ddot{X}_{it} = 0$ for all periods t.
- If $\ddot{X}_{it} = 0$ for all i and t, violates invertibility.
  - R/Stata and the like will drop it from the regression.
  - Any time-constant variable gets "absorbed" by the fixed effect.
- Can include interactions between time-constant and time-varying variables, but lower order term of the time-constant variables get absorbed by fixed effects too.

Time-constant variables

- Pooled model with a time-constant variable, proportion Islamic:

```
library(lmtest)
p.mod <- felm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam, data = ross)
coeftest(p.mod)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.30608    0.35952   28.67  < 2e-16 ***
## democracy   -0.80234    0.07767  -10.33  < 2e-16 ***
## log(GDPcur) -0.25497    0.01607  -15.87  < 2e-16 ***
## islam        0.00343    0.00091    3.77  0.00018 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Time-constant variables

- FE model, where the islam variable drops out, along with the intercept:

```
fe.mod2 <- felm(log(kidmort_unicef) ~ democracy + log(GDPcur) + islam | id, data = ro
coeftest(fe.mod2)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## democracy    -0.1297     0.0359   -3.62  0.00033 ***
## log(GDPcur)  -0.3800     0.0118  -32.07  < 2e-16 ***
## islam            NaN         NA     NaN      NaN
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
    - OLS doesn't know you "used" the data to estimate the within-unit means.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
  - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \dots D_{in}$$

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
  - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \ldots D_{in}$$

  - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
    - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \ldots D_{in}$$

  - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.
  - Gives the **exact** same point estimates as within transformation.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
    - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \dots D_{in}$$

    - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.
    - Gives the **exact** same point estimates as within transformation.
- Comments:

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
  - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \ldots D_{in}$$

  - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.
  - Gives the **exact** same point estimates as within transformation.
- Comments:
  - Pros: easy to implement and gives correct SEs.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
    - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \ldots D_{in}$$

    - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.
    - Gives the **exact** same point estimates as within transformation.
- Comments:
    - Pros: easy to implement and gives correct SEs.
    - Con: computationally slow with large n.

Least squares dummy variable

- Naive OLS on demeaned data is ok for $\hat{\boldsymbol{\beta}}$ but the SEs are wrong.
  - OLS doesn't know you "used" the data to estimate the within-unit means.
- As an alternative, **dummy variable estimator** regressing:

$$Y_{it} \text{ on } \mathbf{X}_{it}, D_{i2}, D_{i3}, \ldots D_{in}$$

  - Here, D_{i2} is a binary variable which is 1 if $i = 2$ and 0 otherwise.
  - Gives the **exact** same point estimates as within transformation.
- Comments:
  - Pros: easy to implement and gives correct SEs.
  - Con: computationally slow with large n.
  - Usually better to use dedicated software like **lfe** package in R.

Example with Ross data

```r
library(lmtest)
lsdv.mod <- lm(log(kidmort_unicef) ~ democracy + log(GDPcur) + as.factor(id),
               data = ross)
coeftest(lsdv.mod)[1:6,]
```

```
##                     Estimate Std. Error t value  Pr(>|t|)
## (Intercept)           13.764     0.2660   51.75 1.01e-198
## democracy             -0.143     0.0335   -4.28  2.30e-05
## log(GDPcur)           -0.375     0.0113  -33.12 3.49e-126
## as.factor(id)AGO       0.300     0.1677    1.79  7.45e-02
## as.factor(id)ALB      -1.931     0.1901  -10.16  4.39e-22
## as.factor(id)ARE      -1.876     0.1702  -11.02  2.39e-25
```

```r
coeftest(fe.mod)[1:2,]
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## democracy     -0.143     0.0335   -4.28  2.30e-05
## log(GDPcur)   -0.375     0.0113  -33.12 3.49e-126
```

4/ Clustering

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.
- Zero conditional mean error holds here (random assignment)

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.
- Zero conditional mean error holds here (random assignment)
- Violation of **iid/random sampling**:

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.
- Zero conditional mean error holds here (random assignment)
- Violation of **iid/random sampling**:
    - errors of individuals within the same household are correlated.

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.
- Zero conditional mean error holds here (random assignment)
- Violation of **iid/random sampling**:
    - errors of individuals within the same household are correlated.
    - SEs are going to be wrong.

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
    - Randomly assign households to different treatment conditions.
    - But the measurement of turnout is at the individual level.
- Zero conditional mean error holds here (random assignment)
- Violation of **iid/random sampling**:
    - errors of individuals within the same household are correlated.
    - SEs are going to be wrong.
- Called **clustering** or **clustered dependence**

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:
    - voters in households

Clustered dependence: notation

- Clusters (groups): $g = 1, \dots, m$
- Units: $i = 1, \dots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:
    - voters in households
    - individuals in states

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:
    - voters in households
    - individuals in states
    - students in classes

Clustered dependence: notation

- Clusters (groups): $g = 1, \dots, m$
- Units: $i = 1, \dots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:
    - voters in households
    - individuals in states
    - students in classes
    - rulings in judges

Clustered dependence: notation

- Clusters (groups): $g = 1, \ldots, m$
- Units: $i = 1, \ldots, n_g$
- n_g is the number of units in cluster g
- $n = \sum_{g=1}^{m} n_g$ is the total number of units
- Units are (usually) belong to a single cluster:
    - voters in households
    - individuals in states
    - students in classes
    - rulings in judges
- Outcome varies at the unit-level, Y_{ig} and the main independent variable varies at the cluster level, X_g.

Clustered dependence: example model

$$
\begin{aligned}
Y_{ig} &= \beta_0 + X_g\beta_1 + v_{ig} \\
&= \beta_0 + X_g\beta_1 + c_g + u_{ig}
\end{aligned}
$$

- u_{ig} unit error component with $\mathbb{V}[u_{ig}|X_g] = \sigma_u^2$

Clustered dependence: example model

$$
\begin{aligned}
Y_{ig} &= \beta_0 + X_g\beta_1 + v_{ig} \\
&= \beta_0 + X_g\beta_1 + c_g + u_{ig}
\end{aligned}
$$

- u_{ig} unit error component with $\mathbb{V}[u_{ig}|X_g] = \sigma_u^2$
- c_g cluster error component with $\mathbb{V}[c_g|X_g] = \sigma_c^2$

Clustered dependence: example model

$$\begin{aligned} Y_{ig} &= \beta_0 + X_g\beta_1 + v_{ig} \\ &= \beta_0 + X_g\beta_1 + c_g + u_{ig} \end{aligned}$$

- u_{ig} unit error component with $\mathbb{V}[u_{ig}|X_g] = \sigma_u^2$
- c_g cluster error component with $\mathbb{V}[c_g|X_g] = \sigma_c^2$
- c_g and u_{ig} are assumed to be independent of each other.

Clustered dependence: example model

$$
\begin{aligned}
Y_{ig} &= \beta_0 + X_g\beta_1 + v_{ig} \\
&= \beta_0 + X_g\beta_1 + c_g + u_{ig}
\end{aligned}
$$

- u_{ig} unit error component with $\mathbb{V}[u_{ig}|X_g] = \sigma_u^2$
- c_g cluster error component with $\mathbb{V}[c_g|X_g] = \sigma_c^2$
- c_g and u_{ig} are assumed to be independent of each other.
  - $\rightsquigarrow \mathbb{V}[v_{ig}|X_g] = \sigma_c^2 + \sigma_u^2$

Clustered dependence: example model

$$\begin{aligned} Y_{ig} &= \beta_0 + X_g\beta_1 + v_{ig} \\ &= \beta_0 + X_g\beta_1 + c_g + u_{ig} \end{aligned}$$

- u_{ig} unit error component with $\mathbb{V}[u_{ig}|X_g] = \sigma_u^2$
- c_g cluster error component with $\mathbb{V}[c_g|X_g] = \sigma_c^2$
- c_g and u_{ig} are assumed to be independent of each other.
    - $\rightsquigarrow \mathbb{V}[v_{ig}|X_g] = \sigma_c^2 + \sigma_u^2$
- What if we ignore this structure and just use v_{ig} as the error?

Lack of independence

- Covariance between two units i and s in the same cluster:

$$\text{Cov}[v_{ig}, v_{sg}] = \sigma_c^2$$

Lack of independence

- Covariance between two units i and s in the same cluster:

$$\text{Cov}[v_{ig}, v_{sg}] = \sigma_c^2$$

- Correlation between units in the same group is called the **intra-class correlation coefficient**, or ρ_c:

$$\text{Cor}[v_{ig}, v_{sg}] = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_u^2} = \rho_c$$

Lack of independence

- Covariance between two units i and s in the same cluster:

$$\text{Cov}[v_{ig}, v_{sg}] = \sigma_c^2$$

- Correlation between units in the same group is called the **intra-class correlation coefficient**, or ρ_c:

$$\text{Cor}[v_{ig}, v_{sg}] = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_u^2} = \rho_c$$

- Zero covariance of two units i and s in different clusters g and k:

$$\text{Cov}[v_{ig}, v_{sk}] = 0$$

Example covariance matrix

- $\mathbf{v}' = \begin{bmatrix} v_{1,1} & v_{2,1} & v_{3,1} & v_{4,2} & v_{5,2} & v_{6,2} \end{bmatrix}$

Example covariance matrix

- $\mathbf{v}' = \begin{bmatrix} v_{1,1} & v_{2,1} & v_{3,1} & v_{4,2} & v_{5,2} & v_{6,2} \end{bmatrix}$
- Variance matrix under clustering:

Example covariance matrix

- $\mathbf{v}' = \begin{bmatrix} v_{1,1} & v_{2,1} & v_{3,1} & v_{4,2} & v_{5,2} & v_{6,2} \end{bmatrix}$
- Variance matrix under clustering:

$$\mathbb{V}[\mathbf{v}|\mathbf{X}] = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \sigma_c^2 \\ 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{bmatrix}$$

Example covariance matrix

- $\mathbf{v}' = \begin{bmatrix} v_{1,1} & v_{2,1} & v_{3,1} & v_{4,2} & v_{5,2} & v_{6,2} \end{bmatrix}$
- Variance matrix under clustering:

$$\mathbb{V}[\mathbf{v}|\mathbf{X}] = \begin{bmatrix} \sigma_c^2+\sigma_u^2 & \sigma_c^2 & \sigma_c^2 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2+\sigma_u^2 & \sigma_c^2 & 0 & 0 & 0 \\ \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma_u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_c^2+\sigma_u^2 & \sigma_c^2 & \sigma_c^2 \\ 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2+\sigma_u^2 & \sigma_c^2 \\ 0 & 0 & 0 & \sigma_c^2 & \sigma_c^2 & \sigma_c^2+\sigma_u^2 \end{bmatrix}$$

- Variance matrix under i.i.d.:

$$\mathbb{V}[\mathbf{v}|\mathbf{X}] = \begin{bmatrix} \sigma_u^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_u^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_u^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_u^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_u^2 \end{bmatrix}$$

Effects of clustering

$$Y_{ig} = \beta_0 + X_g\beta_1 + c_g + u_{ig}$$

- $\mathbb{V}^0[\hat{\beta}_1]$ = **conventional** OLS variance assuming i.i.d./homoskedasticity.

Effects of clustering

$$Y_{ig} = \beta_0 + X_g\beta_1 + c_g + u_{ig}$$

- $\mathbb{V}^0[\widehat{\beta}_1] =$ **conventional** OLS variance assuming i.i.d./homoskedasticity.
- Let $\mathbb{V}[\widehat{\beta}_1]$ be the true sampling variance under clustering.

Effects of clustering

$$Y_{ig} = \beta_0 + X_g\beta_1 + c_g + u_{ig}$$

- $\mathbb{V}^0[\widehat{\beta}_1] =$ **conventional** OLS variance assuming i.i.d./homoskedasticity.
- Let $\mathbb{V}[\widehat{\beta}_1]$ be the true sampling variance under clustering.
- When clusters are balanced, $n^* = n_g$, comparison of clustered to conventional:

$$\mathbb{V}[\widehat{\beta}_1] \approx \mathbb{V}^0[\widehat{\beta}_1]\left(1 + (n^* - 1)\rho_c\right)$$

Effects of clustering

$$Y_{ig} = \beta_0 + X_g\beta_1 + c_g + u_{ig}$$

- $\mathbb{V}^0[\widehat{\beta}_1]$ = **conventional** OLS variance assuming i.i.d./homoskedasticity.
- Let $\mathbb{V}[\widehat{\beta}_1]$ be the true sampling variance under clustering.
- When clusters are balanced, $n^* = n_g$, comparison of clustered to conventional:

$$\mathbb{V}[\widehat{\beta}_1] \approx \mathbb{V}^0[\widehat{\beta}_1]\,(1 + (n^* - 1)\rho_c)$$

- True variance will be higher than conventional when within-cluster correlation is positive, $\rho_c > 0$.

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
  - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
  - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
  - $\mathbb{E}[v_{ig} v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
    - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
    - $\mathbb{E}[v_{ig}v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.
    - Correlated errors allowed within groups, uncorrelated across. Allows heteroskedasticity.

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
    - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
    - $\mathbb{E}[v_{ig}v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.
    - Correlated errors allowed within groups, uncorrelated across. Allows heteroskedasticity.
- Pooled OLS under clustered dependence:

$$\mathbf{Y}_g = \mathbb{X}_g\boldsymbol{\beta} + \mathbf{v}_g$$

Linear model with clustering

$$Y_{ig} = \mathbf{X}_{ig}'\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
  - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
  - $\mathbb{E}[v_{ig}v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.
  - Correlated errors allowed within groups, uncorrelated across. Allows heteroskedasticity.
- Pooled OLS under clustered dependence:

$$\mathbf{Y}_g = \mathbb{X}_g\boldsymbol{\beta} + \mathbf{v}_g$$

  - $\mathbf{Y}_g$ is the $n_g \times 1$ vector of responses for cluster g

Linear model with clustering

$$Y_{ig} = \mathbf{X}'_{ig}\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
  - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
  - $\mathbb{E}[v_{ig}v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.
  - Correlated errors allowed within groups, uncorrelated across. Allows heteroskedasticity.
- Pooled OLS under clustered dependence:

$$\mathbf{Y}_g = \mathbb{X}_g\boldsymbol{\beta} + \mathbf{v}_g$$

  - $\mathbf{Y}_g$ is the $n_g \times 1$ vector of responses for cluster g
  - $\mathbb{X}_g$ is the $n_g \times k$ matrix of data for the gth cluster.

Linear model with clustering

$$Y_{ig} = \mathbf{X}'_{ig}\boldsymbol{\beta} + v_{ig}$$

- Assumptions:
  - $\mathbb{E}[v_{ig} \mid \mathbf{X}_{ig}] = 0$ so we have the correct CEF.
  - $\mathbb{E}[v_{ig}v_{jg'} \mid \mathbf{X}_{ig}, \mathbf{X}_{jg'}] = 0$ unless $g = g'$.
  - Correlated errors allowed within groups, uncorrelated across. Allows heteroskedasticity.
- Pooled OLS under clustered dependence:

$$\mathbf{Y}_g = \mathbb{X}_g\boldsymbol{\beta} + \mathbf{v}_g$$

  - $\mathbf{Y}_g$ is the $n_g \times 1$ vector of responses for cluster g
  - $\mathbb{X}_g$ is the $n_g \times k$ matrix of data for the gth cluster.
- We can write the OLS estimator as:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{g=1}^{m} \mathbb{X}'_g\mathbb{X}_g\right)\left(\sum_{g=1}^{m} \mathbb{X}'_g\mathbf{Y}_g\right)$$

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
    - Key intuition: we're sampling clusters, not individual units.

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
    - Key intuition: we're sampling clusters, not individual units.
- CLT implies $\sqrt{m}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will be asymp. normal with mean 0 and variance:

$$\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1} \mathbb{E}[\mathbb{X}_g'\mathbf{v}_g\mathbf{v}_g'\mathbb{X}_g] \left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}$$

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
    - Key intuition: we're sampling clusters, not individual units.
- CLT implies $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will be asymp. normal with mean 0 and variance:

$$\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1} \mathbb{E}[\mathbb{X}_g'\mathbf{v}_g\mathbf{v}_g'\mathbb{X}_g] \left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}$$

- Similar to the iid case, replace population quantities with sample versions:

$$\widehat{\mathbb{V}}_{CL}[\hat{\boldsymbol{\beta}}] = (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{g=1}^{m} \mathbb{X}_g'\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\mathbb{X}_g\right) (\mathbb{X}'\mathbb{X})^{-1}$$

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
    - Key intuition: we're sampling clusters, not individual units.
- CLT implies $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will be asymp. normal with mean 0 and variance:

$$\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}\mathbb{E}[\mathbb{X}_g'\mathbf{v}_g\mathbf{v}_g'\mathbb{X}_g]\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}$$

- Similar to the iid case, replace population quantities with sample versions:

$$\widehat{\mathbb{V}}_{CL}[\hat{\boldsymbol{\beta}}] = \left(\mathbb{X}'\mathbb{X}\right)^{-1}\left(\sum_{g=1}^{m}\mathbb{X}_g'\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\mathbb{X}_g\right)\left(\mathbb{X}'\mathbb{X}\right)^{-1}$$

    - Noting: $\mathbb{X}'\mathbb{X}/m = m^{-1}\sum_{g=1}^{m}\mathbb{X}_g'\mathbb{X}_g$

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
    - Key intuition: we're sampling clusters, not individual units.
- CLT implies $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will be asymp. normal with mean 0 and variance:

$$\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}\mathbb{E}[\mathbb{X}_g'\mathbf{v}_g\mathbf{v}_g'\mathbb{X}_g]\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}$$

- Similar to the iid case, replace population quantities with sample versions:

$$\widehat{\mathbb{V}}_{CL}[\hat{\boldsymbol{\beta}}] = (\mathbb{X}'\mathbb{X})^{-1}\left(\sum_{g=1}^{m}\mathbb{X}_g'\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\mathbb{X}_g\right)(\mathbb{X}'\mathbb{X})^{-1}$$

    - Noting: $\mathbb{X}'\mathbb{X}/m = m^{-1}\sum_{g=1}^{m}\mathbb{X}_g'\mathbb{X}_g$
    - Tricky to account for all the $1/m$ terms, but this does $\rightarrow 0$ as $m \rightarrow \infty$

Cluster-robust variance estimator

- Independence is across clusters so the CLT holds as m gets big.
  - Key intuition: we're sampling clusters, not individual units.
- CLT implies $\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ will be asymp. normal with mean 0 and variance:

$$\left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1} \mathbb{E}[\mathbb{X}_g'\mathbf{v}_g\mathbf{v}_g'\mathbb{X}_g] \left(\mathbb{E}[\mathbb{X}_g'\mathbb{X}_g]\right)^{-1}$$

- Similar to the iid case, replace population quantities with sample versions:

$$\widehat{\mathbb{V}}_{CL}[\hat{\boldsymbol{\beta}}] = (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{g=1}^{m} \mathbb{X}_g'\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\mathbb{X}_g\right) (\mathbb{X}'\mathbb{X})^{-1}$$

  - Noting: $\mathbb{X}'\mathbb{X}/m = m^{-1}\sum_{g=1}^{m}\mathbb{X}_g'\mathbb{X}_g$
  - Tricky to account for all the $1/m$ terms, but this does $\to 0$ as $m \to \infty$
- With small-sample adjustment (reported by most software):

$$\widehat{\mathbb{V}}_{CL1}[\hat{\boldsymbol{\beta}}] = \frac{m}{m-1}\frac{n-1}{n-k} (\mathbb{X}'\mathbb{X})^{-1} \left(\sum_{g=1}^{m} \mathbb{X}_g'\hat{\mathbf{v}}_g\hat{\mathbf{v}}_g'\mathbb{X}_g\right) (\mathbb{X}'\mathbb{X})^{-1}$$

Example: Gerber, Green, Larimer

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______

Social pressure model

```
load("../assets/gerber_green_larimer.RData")
library(lmtest)
social$voted <- 1 * (social$voted == "Yes")
social$treatment <- factor(
  social$treatment,
  levels = c("Control", "Hawthorne", "Civic Duty", "Neighbors", "Self")
)
mod1 <- lm(voted ~ treatment, data = social)
coeftest(mod1)
```

```
##
## t test of coefficients:
##
##                         Estimate Std. Error t value
## (Intercept)              0.29664    0.00106  279.53
## treatmentHawthorne       0.02574    0.00260    9.90
## treatmentCivic Duty      0.01790    0.00260    6.88
## treatmentNeighbors       0.08131    0.00260   31.26
## treatmentSelf            0.04851    0.00260   18.66
##                         Pr(>|t|)
## (Intercept)              < 2e-16 ***
## treatmentHawthorne       < 2e-16 ***
## treatmentCivic Duty      5.8e-12 ***
## treatmentNeighbors       < 2e-16 ***
## treatmentSelf            < 2e-16 ***
## ---
```

Social pressure model, CRSEs

```r
library(sandwich)
coeftest(mod1, vcov = sandwich::vcovCL(mod1, cluster = social$hh_id))
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value
## (Intercept)           0.29664    0.00131  226.52
## treatmentHawthorne    0.02574    0.00326    7.90
## treatmentCivic Duty   0.01790    0.00324    5.53
## treatmentNeighbors    0.08131    0.00337   24.13
## treatmentSelf         0.04851    0.00330   14.70
##                      Pr(>|t|)
## (Intercept)           < 2e-16 ***
## treatmentHawthorne    2.8e-15 ***
## treatmentCivic Duty   3.2e-08 ***
## treatmentNeighbors    < 2e-16 ***
## treatmentSelf         < 2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\beta}$, cannot fix bias

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
    - Relies on independence between clusters

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
    - Relies on independence between clusters
    - Allows for arbitrary dependence within clusters

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
    - Relies on independence between clusters
    - Allows for arbitrary dependence within clusters
    - CRSEs usually $>$ conventional SEs—use when you suspect clustering

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
  - Relies on independence between clusters
  - Allows for arbitrary dependence within clusters
  - CRSEs usually $>$ conventional SEs—use when you suspect clustering
- When X_{ig} not constant within cluster, but just correlated $\rightsquigarrow$ more complicated.

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
    - Relies on independence between clusters
    - Allows for arbitrary dependence within clusters
    - CRSEs usually $>$ conventional SEs—use when you suspect clustering
- When X_{ig} not constant within cluster, but just correlated $\rightsquigarrow$ more complicated.
    - See Abadie, Athey, Imbens, and Wooldridge (2021).

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\boldsymbol{\beta}}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
  - Relies on independence between clusters
  - Allows for arbitrary dependence within clusters
  - CRSEs usually $>$ conventional SEs—use when you suspect clustering
- When X_{ig} not constant within cluster, but just correlated $\rightsquigarrow$ more complicated.
  - See Abadie, Athey, Imbens, and Wooldridge (2021).
- Consistency of the CRSE are in the number of groups, not the number of individuals

Cluster-robust standard errors

- CRSE do not change our estimates $\hat{\beta}$, cannot fix bias
- Valid under **clustered dependence** when main variable is constant within cluster
    - Relies on independence between clusters
    - Allows for arbitrary dependence within clusters
    - CRSEs usually $>$ conventional SEs—use when you suspect clustering
- When X_{ig} not constant within cluster, but just correlated $\rightsquigarrow$ more complicated.
    - See Abadie, Athey, Imbens, and Wooldridge (2021).
- Consistency of the CRSE are in the number of groups, not the number of individuals
    - CRSEs can be incorrect with a small ($<$ 50 maybe) number of clusters