

ECON207 Session 9

Generalized Least Squares / Intro to Panel Data Models

Anthony Tay

This Version: 29 Oct 2024

## Session 9 GLS / Panel Data Models

- Generalized Least Squares
- Applications
  - Heteroskedasticity
  - Clustered Standard Errors
  - Panel Data (Random Effects Models)
- Panel Data Models (Fixed Effects Model)

## Generalized Least Squares Theory

You know that in the regression model:

$$y = X\beta + \epsilon, \quad E(\epsilon \mid X) = 0, \quad E(\epsilon\epsilon^T \mid X) = \sigma^2 I_n$$

where  $y$  is  $n \times 1$  and  $X$  is a  $n \times (k + 1)$  matrix of regressors

- OLS estimator is  $\hat{\beta}^{ols} = (X^T X)^{-1} X^T y$ 
  - linear estimator
  - Unbiased and consistent
  - Unbiasedness and consistency requires  $y = X\beta + \epsilon, \quad E(\epsilon \mid X) = 0$

## Generalized Least Squares Theory

- Variance-covariance matrix of  $\hat{\beta}^{ols}$  is

$$Var(\hat{\beta}^{ols} \mid X) = \sigma^2 (X^T X)^{-1}$$

- Unbiased estimator for  $\sigma^2$  is

$$\widehat{\sigma^2} = \frac{\hat{\epsilon}_{ols}^T \hat{\epsilon}_{ols}}{n - k - 1} \quad \text{where} \quad \hat{\epsilon}_{ols} = y - X\hat{\beta}^{ols}$$

- Estimate of  $\hat{\beta}^{ols}$  is

$$\widehat{Var}(\hat{\beta}^{ols} \mid X) = \widehat{\sigma^2} (X^T X)^{-1}$$

- All assumptions required

## Generalized Least Squares Theory

- OLS is also *best* linear unbiased estimator

- if  $\tilde{\beta} = Ay$ ,  $A \neq (X^T X)^{-1} X^T$ , then for all non-zero  $(k+1) \times 1$  vector  $c$ ,

$$\text{Var}(c^T \tilde{\beta} | X) > \text{Var}(c^T \hat{\beta}^{ols} | X)$$

- Equivalent description of “best”

$$\text{Var}(\tilde{\beta} | X) - \text{Var}(\hat{\beta}^{ols} | X) \text{ is positive-definite}$$

- “best” requires all assumptions, in particular,  $E(\epsilon \epsilon^T | X) = \sigma^2 I_n$

## Generalized Least Squares Theory

Suppose now that the noise terms  $\epsilon$  are heteroskedastic or correlated (or both)

$$y = X\beta + \epsilon, \quad E(\epsilon | X) = 0, \quad E(\epsilon \epsilon^T | X) = \sigma^2 \Omega$$

where  $\Omega$  is an  $n \times n$  positive-definite matrix not equal to  $I_n$

- Assume that  $\sigma^2$  is unknown
- Assume (for the moment) that  $\Omega$  is **known**

OLS estimator continues to be unbiased / consistent

$$\hat{\beta}^{ols} = (X^T X)^{-1} X^T y = (X^T X)^{-1} X^T (X\beta + \epsilon) = \beta + (X^T X)^{-1} X^T \epsilon$$

$$E(\hat{\beta}^{ols} | X) = \beta + (X^T X)^{-1} X^T E(\epsilon | X) = \beta$$

## Generalized Least Squares Theory

Variance-covariance matrix of  $\hat{\beta}^{ols}$  becomes

$$\text{Var}(\hat{\beta}^{ols} | X) = \sigma^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

Proof:

$$\hat{\beta}^{ols} = \beta + (X^T X)^{-1} X^T \epsilon$$

$$\begin{aligned} \text{Var}(\hat{\beta}^{ols} | X) &= \text{Var}((X^T X)^{-1} X^T \epsilon | X) \\ &= (X^T X)^{-1} X^T \text{Var}(\epsilon | X) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1} \end{aligned}$$

## Generalized Least Squares Theory

But OLS estimator is no longer Best Linear Unbiased,

i.e., we can find another linear estimator that is more “efficient”

- Because  $\Omega$  is positive definite, we can find non-singular  $n \times n$  matrix  $P$  such that

$$P \Omega P^T = I_n$$

Since  $\Omega$  is known,  $P$  is known

- Pre-multiply regression equation by  $P$

$$Py = PX\beta + P\epsilon \quad \text{or} \quad y^* = X^*\beta + \epsilon^*$$

## Generalized Least Squares Theory

- The noise term in the modified equation satisfies

$$E(\epsilon^* | X) = E(P\epsilon | X) = PE(\epsilon | X) = 0$$

$$E(\epsilon^* \epsilon^{*\top} | X) = E(P\epsilon \epsilon^\top P^\top | X) = PE(\epsilon \epsilon^\top | X)P^\top = \sigma^2 P\Omega P^\top = \sigma^2 I_n$$

- Since  $y^* = X^*\beta + \epsilon^*$  satisfies all necessary conditions for BLU Estimators

- $\hat{\beta} = (X^{*\top} X^*)^{-1} X^{*\top} y^* = (X^\top P^\top P X)^{-1} X^\top P^\top P y$

- We refer to this estimator as  $\hat{\beta}^{gls}$

- $Var(\hat{\beta}^{gls} | X) = \sigma^2 (X^{*\top} X^*)^{-1} = \sigma^2 (X^\top P^\top P X)^{-1}$

## Generalized Least Squares Theory

Since  $P\Omega P^\top = I_n$  and  $P$  is non-singular, we have

$$\Omega = (P^{-1})(P^\top)^{-1} \quad \text{and} \quad \Omega^{-1} = P^\top P$$

Therefore we can write the GLS estimator and its variance-covariance matrix as

$$\hat{\beta}^{gls} = (X^\top P^\top P X)^{-1} X^\top P^\top P y = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y$$

$$Var(\hat{\beta}^{gls} | X) = \sigma^2 (X^\top P^\top P X)^{-1} = \sigma^2 (X^\top \Omega^{-1} X)^{-1}$$

An unbiased estimator for  $\sigma^2$  is

$$\hat{\sigma}_*^2 = \frac{\hat{\epsilon}^{*\top} \hat{\epsilon}^*}{n - k - 1} \quad \text{where} \quad \hat{\epsilon}^* = y^* - X^* \hat{\beta}^{gls}$$

## Generalized Least Squares Theory

**Remark 1** GLS is equivalent to

$$\hat{\beta}^{gls} = \arg \min_{\hat{\beta}} (y - X\hat{\beta})^\top \Omega^{-1} (y - X\hat{\beta})$$

since

$$\hat{\beta}^{gls} = \arg \min_{\hat{\beta}} (y^* - X^* \hat{\beta})^\top (y^* - X^* \hat{\beta}) = \arg \min_{\hat{\beta}} (Py - PX\hat{\beta})^\top (Py - PX\hat{\beta})$$

$$= \arg \min_{\hat{\beta}} (P(y - X\hat{\beta}))^\top P(y - X\hat{\beta}) = \arg \min_{\hat{\beta}} (y - X\hat{\beta})^\top P^\top P (y - X\hat{\beta})$$

$$= \arg \min_{\hat{\beta}} (y - X\hat{\beta})^\top \Omega^{-1} (y - X\hat{\beta})$$

## Generalized Least Squares Theory

**Remark 2**

- Since  $\hat{\beta}_{gls}$  is OLS on  $y^* = X^*\beta + \epsilon^*$  which satisfies the “Gauss-Markov” assumptions, we know that  $\hat{\beta}_{gls}$  is BLUE

- Therefore  $\hat{\beta}_{gls}$  is more efficient than  $\hat{\beta}_{ols}$ , since  $\hat{\beta}_{ols}$  is OLS on  $y = X\beta + \epsilon$  which does not satisfy Gauss-Markov assumptions

The above argument is sufficient to show that

$$Var(\hat{\beta}^{ols} | X) - Var(\hat{\beta}^{gls} | X) \quad \text{is positive semi-definite}$$

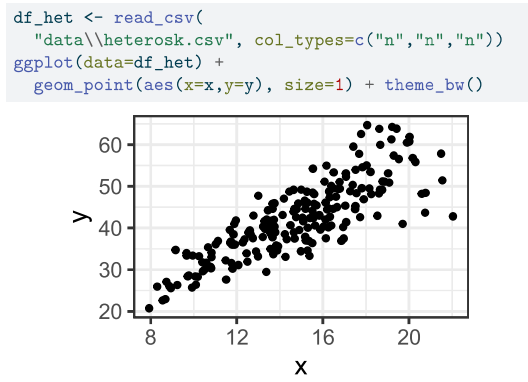
This fact can also be shown directly (exercise)

# Generalized Least Squares (Heteroskedasticity Example)

**Example:** A simple heteroskedasticity example

Suppose for all  $i, j = 1, \dots, n$

- $Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$
- $E(\epsilon_i \mid X_{11}, \dots, X_{n1}) = 0$
- $E(\epsilon_i^2 \mid X_{11}, \dots, X_{n1}) = \sigma_i^2 = \sigma^2 X_{i1}^2$
- $E(\epsilon_i \epsilon_j \mid X_{11}, \dots, X_{n1}) = 0$  for all  $i \neq j$
- $\sum_{i=1}^n (X_{i1} - \bar{X}_1)^2 > 0$



# Generalized Least Squares (Heteroskedasticity Example)

Writing this example in matrix form

$$y = X\beta + \epsilon, \quad E(\epsilon \mid X) = 0, \quad E(\epsilon\epsilon^T \mid X) = \sigma^2\Omega \quad \text{where} \quad \Omega = \begin{bmatrix} X_{11}^2 & 0 & 0 & \dots & 0 \\ 0 & X_{21}^2 & 0 & \dots & 0 \\ 0 & 0 & X_{31}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_{n1}^2 \end{bmatrix}$$

What is  $P$  such that  $P^T P = \Omega^{-1}$ ? We have

$$\Omega^{-1} = \begin{bmatrix} X_{11}^{-2} & 0 & 0 & \dots & 0 \\ 0 & X_{21}^{-2} & 0 & \dots & 0 \\ 0 & 0 & X_{31}^{-2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_{n1}^{-2} \end{bmatrix} \implies P = \begin{bmatrix} X_{11}^{-1} & 0 & 0 & \dots & 0 \\ 0 & X_{21}^{-1} & 0 & \dots & 0 \\ 0 & 0 & X_{31}^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_{n1}^{-1} \end{bmatrix}$$

# Generalized Least Squares (Heteroskedasticity Example)

Then

$$y^* = Py = \begin{bmatrix} X_{11}^{-1} & 0 & 0 & \dots & 0 \\ 0 & X_{21}^{-1} & 0 & \dots & 0 \\ 0 & 0 & X_{31}^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_{n1}^{-1} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1/X_{11} \\ Y_2/X_{21} \\ Y_3/X_{31} \\ \vdots \\ Y_n/X_{n1} \end{bmatrix}$$

$$X^* = PX = \begin{bmatrix} X_{11}^{-1} & 0 & 0 & \dots & 0 \\ 0 & X_{21}^{-1} & 0 & \dots & 0 \\ 0 & 0 & X_{31}^{-1} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & X_{n1}^{-1} \end{bmatrix} \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ 1 & X_{31} \\ \vdots & \vdots \\ 1 & X_{n1} \end{bmatrix} = \begin{bmatrix} 1/X_{11} & 1 \\ 1/X_{21} & 1 \\ 1/X_{31} & 1 \\ \vdots & \vdots \\ 1/X_{n1} & 1 \end{bmatrix}$$

That is, transformed regression is

$$\frac{Y_i}{X_{i1}} = \beta_0 \frac{1}{X_{i1}} + \beta_1 + \frac{\epsilon_i}{X_{i1}} \quad \text{i.e.,} \quad Y_i^* = \beta_1 + \beta_0 X_i^* \epsilon_i^*, \quad i = 1, \dots, n.$$

# Generalized Least Squares (Heteroskedasticity Example)

Equivalently,

$$\hat{\beta}^{gls} = \arg \min_{\hat{\beta}} (y - X\hat{\beta})^T \Omega^{-1} (y - X\hat{\beta}) = \arg \min_{\hat{\beta}} \sum_{i=1}^n w_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1})^2$$

where  $w_i = 1/X_{i1}^2$

This form of GLS is called ‘‘Weighted Least Squares (WLS)’’

- GLS for heteroskedastic, uncorrelated errors
- Impt: Actual form of the weights  $w_i$  depends on  $\Omega$

## Generalized Least Squares (Heteroskedasticity Example)

WLS using `lm()` (option weights refer to  $w_i$ )

```
df_het$wt <- 1/df_het$x^2
wls <- lm(y~x,data=df_het, weights=wt)
coef(summary(wls)) %>% round(3)
cat("\nR-squared: ", summary(wls)$r.squared,"\n")
```

|             | Estimate | Std. Error | t value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 4.826    | 1.407      | 3.431   | 0.001    |
| x           | 2.532    | 0.102      | 24.730  | 0.000    |

R-squared: 0.755433

- WLS and OLS estimates are different, but both consistent
- WLS s.e. smaller (not surprising, since WLS more efficient)
- But why is WLS  $R^2$  larger than OLS  $R^2$ ? (OLS maximizes  $R^2$ )

OLS with Heteroskedasticity-Robust s.e.

```
ols <- lm(y~x, data=df_het)
coeftest(ols, vcov=vcovHC, type="HC") %>%
  round(3)
cat("R-squared: ", summary(ols)$r.squared,"\n")
```

t test of coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )   |
|-------------|----------|------------|---------|------------|
| (Intercept) | 6.286    | 1.864      | 3.372   | 0.001 ***  |
| x           | 2.431    | 0.136      | 17.841  | <2e-16 *** |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.'

R-squared: 0.6826877

## Generalized Least Squares (Heteroskedasticity Example)

$R^2$  using `lm()` for WLS is a “Weighted R-squared”:

$$\text{weighted-}R^2 = \frac{\sum_{i=1}^n w_i (\hat{Y}_i^{wls} - \bar{Y}_{wls})^2}{\sum_{i=1}^n w_i (Y_i - \bar{Y}_{wls})^2}.$$

where

- $\bar{Y}_{wls}$  is the weighted mean of  $\{Y_i\}_{i=1}^n$ ,
- i.e., the WLS estimator of  $\beta_0$  from the regression  $Y_i = \beta_0 + \epsilon_i$  using same weights as in WLS estimator of the main equation

## Generalized Least Squares (Heteroskedasticity Example)

For “Unweighted”- $R^2$ , use

$$R^2_{wls} = 1 - \frac{\sum_{i=1}^n \hat{\epsilon}_{i,wls}^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

where  $\hat{\epsilon}_i^{wls} = Y_i - \hat{Y}_i^{wls} = Y_i - \hat{\beta}_0^{wls} - \hat{\beta}_1^{wls} X_i$

```
ehat <- df_het$y - coef(wls)[1] - coef(wls)[2]*df_het$x
ssr <- sum(ehat^2); sst <- sum((df_het$y - mean(df_het$y))^2)
R2 <- 1 - ssr/sst; cat("R-squared: ", R2,"\n")
```

R-squared: 0.6814966

- $R^2_{wls}$  is slightly lower than in the OLS regression
- Expected, since OLS minimizes SSR, equivalent to maximizing  $R^2$

## Generalized Least Squares (Heteroskedasticity Example)

For more general cases, weighted least squares is not so straightforward  
Heteroskedasticity specification may have parameters that need to be estimated, e.g.,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, \quad \text{Var}(\epsilon_i|X) = \alpha_0 + \alpha_1 X_{i1}^2 + \alpha_2 X_{i2}^2$$

- To do weighted least squares, have to estimate  $\alpha_0, \alpha_1, \alpha_2$ 
  - Maybe estimate regression by OLS, then estimate  $\alpha$ 's by regressing

$$\hat{\epsilon}_{i,ols}^2 = \alpha_0 + \alpha_1 X_{i1}^2 + \alpha_2 X_{i2}^2 + u_i$$

- Then use weights  $w_i = (\hat{\alpha}_0 + \hat{\alpha}_1 X_{i1}^2 + \hat{\alpha}_2 X_{i2}^2)^{-2}$
- Is form of heteroskedasticity correct? Other forms may be preferred/appropriate.

## Generalized Least Squares

GLS refers to any method that involves

- Linear transformation of regression equation so that the transformed errors meet Gauss-Markov Criterion
- OLS on transformed equation

In next section, we turn to models for Panel Data

- Random Effects Estimator (GLS)
- Fixed Effects Estimator

## Panel Data

### Introduction to Panel Data (Longitudinal Data)

## Panel Data

- Panel data comprising information on a **fixed** cross section of individual entities (people, firms, cities,...) observed repeatedly over several time periods

$$\{Y_{i,t}, X_{i,t,1}, X_{i,t,2}, \dots, X_{i,t,K}\}, i = 1, 2, \dots, N, t = 1, 2, \dots, T$$

- $i$  refers to the individual entities in the cross-section
- $t$  refers to period of observation
- $X_{i,t,K}$  is period  $t$  observation of variable  $K$  for individual  $i$

The same individuals are followed over time, e.g.,  $i = 1$  refers to the same individual for all time periods

## Panel Data

- National Longitudinal Surveys of Labor Market Experience (NLS)
  - *Data on attitudes/behaviors/events related to schooling, employment, marriage, fertility, health, ... Follows various cohorts ('older men', 'mature women', 'young men', ... annual/biennially since mid-1960s)*
- Panel Study of Income Dynamics (PSID)
  - *Data on employment, income, housing, travel, ... Follows 6000 families, 15000 individuals (and their descendants), still continuing*
- Singapore Life Panel
- German Social Economics Panel, British Household Panel Survey, ...

## Panel Data

One benefit of Panel Data is in dealing with certain omitted variables. Suppose

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \alpha_i + u_{i,t}$$

$$= \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \epsilon_{i,t}, \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

where  $\alpha_i$  is some unobserved variable or combination of variables, a.k.a. “time invariant individual effect”

- in returns to schooling application,  $\alpha_i$  might include ability

If observations are simply pooled, then

- if  $\alpha_i$  is correlated with included regressors, OLS estimator for  $\beta_1$  will be inconsistent

## Panel Data

The panel data structure allows us to remove the individual effects

For example, if we “time-demean” the sample, we get

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \alpha_i + u_{i,t}$$

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_{i,1} + \dots + \beta_K \bar{X}_{i,K} + \alpha_i + \bar{u}_i$$

$$Y_{i,t} - \bar{Y}_i = \beta_1 (X_{i,t,1} - \bar{X}_{i,1}) + \dots + \beta_K (X_{i,t,K} - \bar{X}_{i,K}) + u_{i,t} - \bar{u}_i$$

$$\dot{Y}_{i,t} = \beta_1 \dot{X}_{i,t,1} + \dots + \beta_K \dot{X}_{i,t,K} + \ddot{u}_{i,t}$$

The unobserved individual effect has been removed

OLS on the time-demeaned equation gives the “Fixed Effect Estimator”

## Panel Data

Assume Large N, Small T; Balanced Panel (observed regularly, no attrition)

- Pooled OLS
- Random Effect Estimator
  - Generalized Least Squares Estimator when  $\alpha_i$  **not** correlated with  $X_{i,t,k}$
- Fixed Effects Estimator
  - Consistent estimation when  $\alpha_i$  is correlated with  $X_{i,t,k}$
  - Least Squares Dummy Variable Estimator
  - First-Difference Estimator

## Pooled OLS

We begin with the simple case where  $\alpha_i$  is **not** correlated with regressors

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \alpha_i + u_{i,t}$$

$$= \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \epsilon_{i,t}, \quad i = 1, \dots, N; \quad t = 1, \dots, T$$

OLS on the “pooled” data set gives consistent estimator

However, observations are no longer uncorrelated across all  $i$  and  $t$

- Cannot use default standard errors

Pooled OLS

Setting up using matrix algebra

For any given individual  $i = 1, \dots, N$ , let

$$y_i = \begin{bmatrix} Y_{i,1} \\ Y_{i,2} \\ \vdots \\ Y_{i,T} \end{bmatrix}, \quad X_i = \begin{bmatrix} 1 & X_{i,1,1} & \dots & X_{i,1,K} \\ 1 & X_{i,2,1} & \dots & X_{i,2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{i,T,1} & \dots & X_{i,T,K} \end{bmatrix}, \quad \epsilon_i = \begin{bmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \vdots \\ \epsilon_{i,T} \end{bmatrix}$$

The regression for the  $i$ th individual is

$$y_i = X_i \beta + \epsilon_i$$

Pooled OLS

Stacking up the samples for each individual, we get the pooled Regression model

$$y = X\beta + \epsilon$$

where  $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$ ,  $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$ ,  $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}$

Pooled OLS estimator:

$$\hat{\beta}^{pols} = (X^T X)^{-1} X^T y$$

Pooled OLS

The pooled OLS estimator is consistent as long as  $E(\epsilon \mid X) = 0$

However, usual variance formula invalid because  $Var(\epsilon)$  does not have the form  $\sigma^2 I$

In  $\epsilon_{i,t} = \alpha_i + u_{i,t}$ , assume  $\alpha_i$  and  $u_{i,t}$  independent random variables with

- $\alpha_i$  zero-mean variance  $\sigma_\alpha^2$  (think of  $E(\alpha_i)$  as having been absorbed into  $\beta_0$ )
- $u_{i,t}$  zero-mean variance  $\sigma_u^2$ 
  - $Var(\epsilon_{i,t}) = \sigma_\alpha^2 + \sigma_u^2$
  - $Cov(\epsilon_{i,t}, \epsilon_{i,s}) = E((\alpha_i + u_{i,t})(\alpha_i + u_{i,s})) = \sigma_\alpha^2$  for  $s \neq t$
  - $Cov(\epsilon_{i,t}, \epsilon_{j,s}) = E((\alpha_i + u_{i,t})(\alpha_j + u_{j,s})) = 0$  for all  $i \neq j$ , all  $s, t$

Pooled OLS

Then  $Var(\epsilon_i \mid X) = E(\epsilon_i \epsilon_i^T \mid X)$

$$= \begin{bmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_u^2 + \sigma_\alpha^2 \end{bmatrix} = \sigma_u^2 \underbrace{\left( I_T + \frac{\sigma_\alpha^2}{\sigma_u^2} i_T i_T^T \right)}_{\text{"}\Omega\text{"}}$$

$$Var(\epsilon \mid X) = E(\epsilon \epsilon^T \mid X) = \begin{bmatrix} E(\epsilon_1 \epsilon_1^T) & E(\epsilon_1 \epsilon_2^T) & \dots & E(\epsilon_1 \epsilon_N^T) \\ E(\epsilon_2 \epsilon_1^T) & E(\epsilon_2 \epsilon_2^T) & \dots & E(\epsilon_2 \epsilon_N^T) \\ \vdots & \vdots & \ddots & \vdots \\ E(\epsilon_N \epsilon_1^T) & E(\epsilon_N \epsilon_2^T) & \dots & E(\epsilon_N \epsilon_N^T) \end{bmatrix} = \sigma_u^2 \begin{bmatrix} \Omega & 0 & \dots & 0 \\ 0 & \Omega & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Omega \end{bmatrix}$$



# Pooled OLS

“Panel-Robust Estimate” of  $Var(\hat{\beta}^{pols})$ :

$$\widehat{Var}(\hat{\beta}^{pols}) = \left( \sum_{i=1}^N X_i^T X_i \right)^{-1} \left( \sum_{i=1}^N X_i^T \hat{\epsilon}_{pols} \hat{\epsilon}_{pols}^T X_i \right) \left( \sum_{i=1}^N X_i^T X_i \right)^{-1}$$

where  $\hat{\epsilon}_{pols} = y - X\hat{\beta}_{pols}$

This estimator of the variance in fact allows for  $Var(\epsilon_i | X) = \Omega_i$

# Random Effects Estimator

Pooled OLS estimator  $\hat{\beta}^{pols}$  is consistent but not efficient

- Panel-Robust variance estimator  $\widehat{Var}(\hat{\beta}^{pols})$  accounts for structure of  $Var(\epsilon)$
- But  $\hat{\beta}^{pols}$  does not make use of this structure

Generalized Least Squares

- Find  $P$  such that  $Var(P\epsilon | X) = E(P\epsilon\epsilon^T P^T | X) = \sigma_u^2 I$
- Transform regression  $Py = PX\beta + P\epsilon$  or  $y^* = X^*\beta + \epsilon^*$
- Apply OLS to transformed equation

GLS estimator  $\hat{\beta}^{gl} = (X^{*T} X^*)^{-1} X^{*T} y^*$

# Random Effects Estimator

What is the appropriate  $P$ ? Consider sample for just the  $i$ th individual:

$$y_i = X_i\beta + \epsilon_i, \quad Var(\epsilon_i | X) = \sigma_u^2 \left( I_T + \frac{\sigma_\alpha^2}{\sigma_u^2} i_T i_T^T \right) = \sigma_u^2 \Omega$$

Let

$$P = M_0 + \psi(I_T - M_0)$$

where  $M_0 = I_T - i_T(i_T^T i_T)^{-1} i_T^T$  and  $\psi = \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_\alpha^2}}$

Then  $E(P\epsilon_i\epsilon_i^T P^T) = \sigma_u^2 I_T$

# Random Effects Estimator

Proof:  $M_0$  and  $I - M_0$  are both symmetric, idempotent, and  $M_0(I - M_0) = (I - M_0)M_0 = 0$

Furthermore, for any non-zero scalar  $\psi$ , we have

$$(M_0 + \psi(I_T - M_0)) \left( M_0 - \frac{1}{\psi^2} (I - M_0) \right) (M_0 + \psi(I_T - M_0))^T = I_T \quad (\text{exercise!})$$

Finally, we have

$$\begin{aligned} \Omega &= I_T + \frac{\sigma_\alpha^2}{\sigma_u^2} i_T i_T^T = I_T + \frac{T\sigma_\alpha^2}{\sigma_u^2} i_T (i_T^T i_T)^{-1} i_T^T \\ &= I_T - i_T (i_T^T i_T)^{-1} i_T^T + i_T (i_T^T i_T)^{-1} i_T^T + \frac{T\sigma_\alpha^2}{\sigma_u^2} i_T (i_T^T i_T)^{-1} i_T^T \\ &= M_0 + \left( 1 + \frac{T\sigma_\alpha^2}{\sigma_u^2} \right) i_T (i_T^T i_T)^{-1} i_T^T \\ &= M_0 + \left( \frac{\sigma_u^2 + T\sigma_\alpha^2}{\sigma_u^2} \right) i_T (i_T^T i_T)^{-1} i_T^T = M_0 + \frac{1}{\psi^2} (I_T - M_0) \quad \text{where } \psi = \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}} \end{aligned}$$

## Random Effects Estimator

It follows that if  $Var(\epsilon_i) = \sigma_u^2 \Omega$  where

$$\Omega = I_T + \frac{\sigma_\alpha^2}{\sigma_u^2} i_T i_T^T = M_0 + \frac{1}{\psi^2} (I_T - M_0) \quad \text{where} \quad \psi = \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}}$$

Then

$$Var(P\epsilon_i) = P Var(\epsilon_i) P^T = \sigma_u^2 P \Omega P^T = \sigma_u^2 I_T$$

Apply the transformation to  $y_i = X_i \beta + \epsilon_i$  for every  $i$  and then pool the transformed regressions, which gives

$$\begin{bmatrix} Py_1 \\ Py_2 \\ \vdots \\ Py_N \end{bmatrix} = \begin{bmatrix} PX_1 \\ PX_2 \\ \vdots \\ PX_N \end{bmatrix} \beta + \begin{bmatrix} P\epsilon_1 \\ P\epsilon_2 \\ \vdots \\ P\epsilon_N \end{bmatrix} \quad \text{or} \quad y^* = X^* \beta + \epsilon^*, \quad Var(\epsilon^* | X) = \sigma_u^2 I_{NT}$$

## Random Effects Estimator

What is this transformation?

$$\begin{aligned} Py_i &= (M_0 + \psi(I_T - M_0))y_i \\ &= (I_T - i_T(i_T^T i_T)^{-1} i_T^T + \psi i_T(i_T^T i_T)^{-1} i_T^T) y_i \\ &= (I_T - (1 - \psi) i_T(i_T^T i_T)^{-1} i_T^T) y_i \\ &= y_i - \lambda i_T \bar{Y}_i \end{aligned}$$

$$\text{where } \lambda = 1 - \psi = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}}$$

## Random Effects Estimator

$$\text{i.e., } Py_i = y_i^* = \begin{bmatrix} Y_{i,1} - \lambda \bar{Y}_i \\ Y_{i,2} - \lambda \bar{Y}_i \\ \vdots \\ Y_{i,T} - \lambda \bar{Y}_i \end{bmatrix} \quad \text{where } \lambda = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_\alpha^2}}$$

Same for other regressors  $X_i^* = PX_i$

Note:

- To operationalize this theory, we have to estimate  $\sigma_u^2$  and  $\sigma_\alpha^2$
- We can obtain these from the Pooled OLS residuals (details omitted)
- Assumptions on  $\Omega$  more restrictive than in Pooled OLS with Panel-Robust S.E.

## Fixed Effects Estimator

We now consider the case where  $\alpha_i$  is correlated with regressors

$$\begin{aligned} Y_{i,t} &= \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \alpha_i + u_{i,t} \\ &= \beta_0 + \beta_1 X_{i,t,1} + \dots + \beta_K X_{i,t,K} + \epsilon_{i,t}, \quad i = 1, \dots, N; \quad t = 1, \dots, T \end{aligned}$$

We have to get deal with the  $\alpha_i$  in the  $\epsilon_{i,t}$

Fixed Effects Estimator

- subtract time-averages from each cross-sectional observation

## Fixed Effects Estimator

i.e., for every  $i$ , take  $\ddot{y}_i = M_0 y_i$ ,  $\ddot{X}_i = M_0 X_i$ ,  $\ddot{u}_i = M_0 u_i$

$$_{(NT \times 1)} \ddot{y} = \begin{bmatrix} \ddot{y}_1 \\ \ddot{y}_2 \\ \vdots \\ \ddot{y}_N \end{bmatrix} \quad \ddot{X}_{(NT \times K)} = \begin{bmatrix} \ddot{X}_1 \\ \ddot{X}_2 \\ \vdots \\ \ddot{X}_N \end{bmatrix} \quad \ddot{u}_{(NT \times 1)} = \begin{bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \\ \vdots \\ \ddot{u}_N \end{bmatrix}$$

where  $M_0 = I_T - i_T(i_T^T i_T)^{-1} i_T^T$

The transformed model is

$$\ddot{y} = \ddot{X}\beta + \ddot{u}$$

Estimate by OLS

$$\hat{\beta}^{fe} = (\ddot{X}^T \ddot{X})^{-1} \ddot{X}^T \ddot{y}$$

## Fixed Effects Estimator

Remarks:

- Also called “Within Estimator”
- Cannot include any time invariant variables such as race (will get removed along with unobserved individual effect)
- Identical to “Least Squares Dummy Variable” Models (LSDV)

$$Y_{i,t} = \theta_1 d_{1,i,t} + \theta_2 d_{2,i,t} + \dots + \theta_N d_{N,i,t} + \beta_1 X_{1,i,t} + \dots + \beta_K X_{K,i,t} + u_{i,t}$$

- proof omitted
- Intuition: individual dummies capture individual time effects
- Often  $N$  is very large, so direct OLS on the LSDV model may be infeasible

## Fixed Effects Estimator

Remarks (continued):

- Alternative to FE estimator: First Difference estimator

$$\begin{aligned} Y_{i,t} &= \beta_1 X_{1,i,t} + \dots + \beta_K X_{K,i,t} + \alpha_i + u_{i,t} \\ Y_{i,t-1} &= \beta_1 X_{1,i,t-1} + \dots + \beta_K X_{K,i,t-1} + \alpha_i + u_{i,t-1} \\ \Delta Y_{i,t} &= \beta_1 \Delta X_{1,i,t} + \dots + \beta_K \Delta X_{K,i,t} + \Delta u_{i,t} \end{aligned}$$

Estimate by OLS

- This approach also removes the unobserved individual effects
- If  $T = 2$ , then FD = FE
- If there are missing observations, FD can make many observations become unusable

## Example

Data jtrain from wooldridge library

- Several firms followed over three years (87, 88, 89)
- will use
  - grant (whether job training grant was given to firm that year)
  - grant\_1 (whether training grant was given in previous year)
  - lscrap (logged scrap rates)
  - fcode firm id
  - year, d87, d88, d89 year and year dummies

## Example

```
library(plm); library(lmtest); library(sandwich); library(wooldridge)
dat <- jtrain

# Pooled OLS
pool_mdl <- plm(lscrap~d88+d89+grant+grant_1, data=dat, model="pooling",
               index=c("fcode", "year"))
summary(pool_mdl)$coefficients %>% round(4)
```

|             | Estimate | Std. Error | t-value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.5974   | 0.2031     | 2.9421  | 0.0038   |
| d88         | -0.2394  | 0.3109     | -0.7700 | 0.4424   |
| d89         | -0.4965  | 0.3379     | -1.4693 | 0.1437   |
| grant       | 0.2000   | 0.3383     | 0.5913  | 0.5552   |
| grant_1     | 0.0489   | 0.4361     | 0.1122  | 0.9108   |

## Example

```
# Pooled OLS with Panel-Robust SE
pool_Var <- vcovHC(pool_mdl, method="arellano", type="HC1", cluster="group")
coeftest(pool_mdl, vcov=pool_Var) %>% round(4)
```

t test of coefficients:

|                | Estimate | Std. Error | t value  | Pr(> t )           |
|----------------|----------|------------|----------|--------------------|
| (Intercept)    | 0.5974   | 0.2184     | 2.7357   | 0.0069 **          |
| d88            | -0.2394  | 0.1251     | -1.9135  | 0.0575 .           |
| d89            | -0.4965  | 0.2317     | -2.1427  | 0.0337 *           |
| grant          | 0.2000   | 0.3206     | 0.6239   | 0.5336             |
| grant_1        | 0.0489   | 0.4691     | 0.1043   | 0.9171             |
| ---            |          |            |          |                    |
| Signif. codes: | 0 '***'  | 0.001 '**' | 0.01 '*' | 0.05 '.' 0.1 ' ' 1 |

## Example

```
# Random Effects
re_mdl <- plm(lscrap~d88+d89+grant+grant_1, data=dat, model="random",
               index=c("fcode", "year"))
summary(re_mdl)$coefficients %>% round(4)
```

|             | Estimate | Std. Error | z-value | Pr(> z ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 0.5974   | 0.2033     | 2.9389  | 0.0033   |
| d88         | -0.0935  | 0.1090     | -0.8584 | 0.3907   |
| d89         | -0.2714  | 0.1315     | -2.0643 | 0.0390   |
| grant       | -0.2144  | 0.1476     | -1.4529 | 0.1463   |
| grant_1     | -0.3729  | 0.2051     | -1.8182 | 0.0690   |

- Results very different from Pooled OLS (which doesn't account for individual fixed effects in any way)

## Example

Can use Panel-Robust SE after RE estimation

```
# Random Effects with Panel-Robust SE
re_Var <- vcovHC(re_mdl, method="arellano", type="HC1", cluster="group")
coeftest(re_mdl, vcov=re_Var) %>% round(4)
```

t test of coefficients:

|                | Estimate | Std. Error | t value  | Pr(> t )           |
|----------------|----------|------------|----------|--------------------|
| (Intercept)    | 0.5974   | 0.2184     | 2.7357   | 0.0069 **          |
| d88            | -0.0935  | 0.0930     | -1.0061  | 0.3159             |
| d89            | -0.2714  | 0.1865     | -1.4548  | 0.1477             |
| grant          | -0.2144  | 0.1303     | -1.6463  | 0.1017             |
| grant_1        | -0.3729  | 0.2659     | -1.4021  | 0.1629             |
| ---            |          |            |          |                    |
| Signif. codes: | 0 '***'  | 0.001 '**' | 0.01 '*' | 0.05 '.' 0.1 ' ' 1 |

## Example

```
# Fixed Effects
fe_md1 <- plm(lscrap~d88+d89+grant+grant_1, data=dat, model="within",
             index=c("fcode", "year"))
summary(fe_md1)$coefficients %>% round(4)
```

|         | Estimate | Std. Error | t-value | Pr(> t ) |
|---------|----------|------------|---------|----------|
| d88     | -0.0802  | 0.1095     | -0.7327 | 0.4654   |
| d89     | -0.2472  | 0.1332     | -1.8556 | 0.0663   |
| grant   | -0.2523  | 0.1506     | -1.6751 | 0.0969   |
| grant_1 | -0.4216  | 0.2102     | -2.0057 | 0.0475   |

## Example

Can also use Panel-Robust SE after FE estimation

```
# Fixed Effects with Panel-Robust SE
fe_Var <- vcovHC(fe_md1, method="arellano", type="HC1", cluster="group")
coeftest(fe_md1, vcov=fe_Var) %>% round(4)
```

t test of coefficients:

|         | Estimate | Std. Error | t value | Pr(> t ) |
|---------|----------|------------|---------|----------|
| d88     | -0.0802  | 0.0969     | -0.8276 | 0.4098   |
| d89     | -0.2472  | 0.1949     | -1.2681 | 0.2076   |
| grant   | -0.2523  | 0.1421     | -1.7757 | 0.0787   |
| grant_1 | -0.4216  | 0.2798     | -1.5067 | 0.1349   |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Example

```
fixef(fe_md1) %>% round(4)
```

|         |         |         |         |         |         |        |         |         |         |
|---------|---------|---------|---------|---------|---------|--------|---------|---------|---------|
| 410523  | 410538  | 410563  | 410565  | 410566  | 410567  | 410577 | 410592  | 410593  | 410596  |
| -2.8258 | 1.0794  | 1.8915  | 1.6178  | 1.7956  | -0.5462 | 0.5973 | 3.3008  | 0.1091  | 1.9360  |
| 410606  | 410626  | 410629  | 410653  | 410665  | 410685  | 418011 | 418021  | 418035  | 418045  |
| -0.7028 | -0.1176 | -0.2795 | 0.1821  | -2.9002 | -1.6470 | 1.6986 | 0.3338  | 1.8337  | 0.8023  |
| 418051  | 418054  | 418065  | 418076  | 418083  | 418091  | 418097 | 418107  | 418118  | 418125  |
| -0.3945 | 0.5385  | 0.5564  | -0.2259 | 0.9310  | 0.7076  | 0.6739 | -0.1636 | -0.6648 | -0.2674 |
| 418140  | 418163  | 418168  | 418177  | 418237  | 419198  | 419201 | 419242  | 419268  | 419272  |
| 1.7944  | 2.3388  | -2.6982 | 3.2422  | -1.1806 | 1.9810  | 0.8023 | 1.8211  | 0.3174  | 3.2038  |
| 419289  | 419297  | 419307  | 419339  | 419343  | 419357  | 419378 | 419381  | 419388  | 419409  |
| 0.4258  | -1.2542 | 0.1002  | -0.5640 | 1.5580  | 0.1932  | 0.7627 | -0.4722 | 2.2829  | 1.0000  |
| 419432  | 419459  | 419482  | 419483  |         |         |        |         |         |         |
| 1.7180  | 0.6233  | 1.1006  | 3.3144  |         |         |        |         |         |         |

## Example

LSDV gives the same result as FE (only part of output shown...)

```
# LSDV
lsdv_md1 <- lm(lscrap~d88+d89+grant+grant_1+factor(fcode), data=dat)
summary(lsdv_md1)$coefficients %>% round(4)
```

|                     | Estimate | Std. Error | t value | Pr(> t ) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | -2.8258  | 0.2962     | -9.5407 | 0.0000   |
| d88                 | -0.0802  | 0.1095     | -0.7327 | 0.4654   |
| d89                 | -0.2472  | 0.1332     | -1.8556 | 0.0663   |
| grant               | -0.2523  | 0.1506     | -1.6751 | 0.0969   |
| grant_1             | -0.4216  | 0.2102     | -2.0057 | 0.0475   |
| factor(fcode)410538 | 3.9053   | 0.4064     | 9.6092  | 0.0000   |
| factor(fcode)410563 | 4.7173   | 0.4064     | 11.6074 | 0.0000   |
| factor(fcode)410565 | 4.4437   | 0.4064     | 10.9340 | 0.0000   |
| factor(fcode)410566 | 4.6214   | 0.4064     | 11.3715 | 0.0000   |
| factor(fcode)410567 | 2.2796   | 0.4064     | 5.6091  | 0.0000   |
| factor(fcode)410577 | 3.4231   | 0.4064     | 8.4230  | 0.0000   |
| factor(fcode)410592 | 6.1266   | 0.4064     | 15.0751 | 0.0000   |
| factor(fcode)410593 | 2.9350   | 0.4064     | 7.2217  | 0.0000   |
| factor(fcode)410596 | 4.7618   | 0.4064     | 11.7169 | 0.0000   |

## Example

```
# First difference
fd_md1 <- plm(lscrap~d88+d89+grant+grant_1, data=dat, model="fd",
              index=c("fcode", "year"))
summary(fd_md1)$coefficients %>% round(4)
```

|             | Estimate | Std. Error | t-value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | -0.1387  | 0.0752     | -1.8450 | 0.0679   |
| d88         | 0.0481   | 0.0627     | 0.7669  | 0.4449   |
| grant       | -0.2228  | 0.1307     | -1.7040 | 0.0914   |
| grant_1     | -0.3512  | 0.2351     | -1.4941 | 0.1382   |

## Example

```
# First difference with Panel Robust SE
fd_Var <- vcovHC(fd_md1, method="arellano", type="HC1", cluster="group")
coeftest(fd_md1, vcov=fd_Var) %>% round(4)
```

t test of coefficients:

|                | Estimate | Std. Error | t value  | Pr(> t )           |
|----------------|----------|------------|----------|--------------------|
| (Intercept)    | -0.1387  | 0.0770     | -1.8009  | 0.0746 .           |
| d88            | 0.0481   | 0.0459     | 1.0483   | 0.2969             |
| grant          | -0.2228  | 0.1063     | -2.0957  | 0.0385 *           |
| grant_1        | -0.3512  | 0.2188     | -1.6052  | 0.1115             |
| ---            |          |            |          |                    |
| Signif. codes: | 0 '***'  | 0.001 '**' | 0.01 '*' | 0.05 '.' 0.1 ' ' 1 |

## Roadmap

- (Previous) Session 1: Statistics Review
- (Previous) Session 2: Simple Linear Regression
- (Previous) Session 3: Estimator Standard Errors; Multiple Linear Regression
- (Previous) Session 4: Matrix Algebra
- (Previous) Session 5: OLS using Matrix Algebra
- (Previous) Session 6: Hypothesis Testing
- (Previous) Session 7: Prediction
- (Previous) Session 8: Instrumental Variable Regression
- *This Session 9: Panel Data Regressions*
- **Next Session 10: MLE / Limited Dependent Variable Models**
- Session 11-12: Introduction to Time Series / Time Series Regressions