GSERM 2020
Regression for Publishing

June 19, 2020 (first session)

Starting Points

- "Longitudinal" $\neq$ "Time Series"
- Terminology:
  - "Unit" / "Units" / "Units of observation" / "Panels" = Things we observe repeatedly
  - "Observations" = Each (one) measurement of a unit
  - "Time points" = When each observation on a unit is made
  - $i \in \{1....N\}$ indexes units
  - $t \in \{1...T\}$ or $\{1...T_i\}$ indexes observations / time points
  - If $T_i = T \; \forall i$ then we have "balanced" panels / units
  - NT = Total number of observations (if balanced)
- Averages:
  - Y_{it} indicates a variable that varies over both units and time,
  - $\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}$ = the over-time mean of Y,
  - $\bar{Y}_t = \frac{1}{N} \sum_{i=1}^{N} Y_{it}$ = the across-unit mean of Y, and
  - $\bar{Y} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} Y_{it}$ = the grand mean of Y.

More Terminology

- $N >> T \rightarrow$ "panel" data
  - (American) National Election Study panel studies ($N = 2000$, $T = 3$)
  - (U.S.) Panel Study of Income Dynamics ($N =$ large, $T \approx 12$)

- $T >> N$ or $T \approx N \rightarrow$ "time-series cross-sectional" ("TSCS") data

- $N = 1 \rightarrow$ "time series" data

Panel/TSCS Data Structure

id	t	Y	X_1	...
1	1	250	3.4	...
1	2	290	3.3	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...
2	1	160	4.7	...
2	2	150	4.9	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...

Within- and Between-Unit Variation

Define:

$$\bar{Y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} Y_{it}$$

Then:

$$Y_{it} = \bar{Y}_i + (Y_{it} - \bar{Y}_i).$$

- The *total* variation in Y_{it} can be decomposed into
- The *between-unit* variation in the $\bar{Y}_i$s, and
- The *within-unit* variation around $\bar{Y}_i$ (that is, $Y_{it} - \bar{Y}_i$).

Daily COVID-19 New Cases, 31/12/2019 - 18/6/2020

```
> summary(COVIDs)
      location              date           new_cases
 Australia:  171   2020-05-31:  208   Min.   : -2461.0
 Austria  :  171   2020-06-01:  208   1st Qu.:      0.0
 Belarus  :  171   2020-06-02:  208   Median :      4.0
 Belgium  :  171   2020-06-03:  208   Mean   :    690.2
 Brazil   :  171   2020-06-04:  208   3rd Qu.:     65.5
 Canada   :  171   2020-06-05:  208   Max.   :175841.0
 (Other)  :23295   (Other)   :23073   NA's   :218

> # Overall:
>
> describe(COVIDs$new_cases)
   vars     n   mean     sd median trimmed  mad   min    max  range  skew kurtosis    se
X1    1 24103 690.17 6009.5      4   45.25 5.93 -2461 175841 178302 15.77   283.66 38.71

> # Between:
>
> CountryMeans <- ddply(COVIDs,.(location),summarise,
+                       newcases = mean(new_cases,na.rm=TRUE))
> with(CountryMeans, describe(newcases))
   vars   n   mean      sd median trimmed   mad  min      max    range  skew kurtosis     se
X1    1 207 499.65 3520.26  17.76    68.4 25.94 0.04 48640.78 48640.74 12.57   166.67 244.68

> # Within:
>
> COVIDs <- ddply(COVIDs, .(location), mutate,
+                CaseMean = mean(new_cases,na.rm=TRUE))
> COVIDs$within <- with(COVIDs, new_cases-CaseMean)
> with(COVIDs, describe(within))
   vars     n mean      sd median trimmed   mad       min      max  range skew kurtosis    se
X1    1 24103    0 4261.94  -3.49  -22.86 33.35 -48640.78 127200.2 175841 3.74   188.88 27.45
```

Regression!

Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

assumes:

- All the usual OLS assumptions, plus
- $\beta_{0i} = \beta_0 \; \forall \; i$
- $\beta_{1i} = \beta_1 \; \forall \; i$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$$

(same)

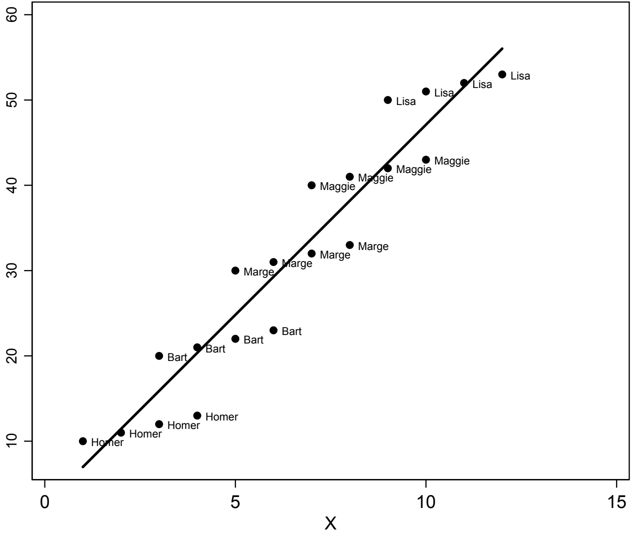
Variable Intercepts

$$Y_{it} = \beta_{0i} + \beta_1 X_{it} + u_{it} \text{ (by unit)}$$

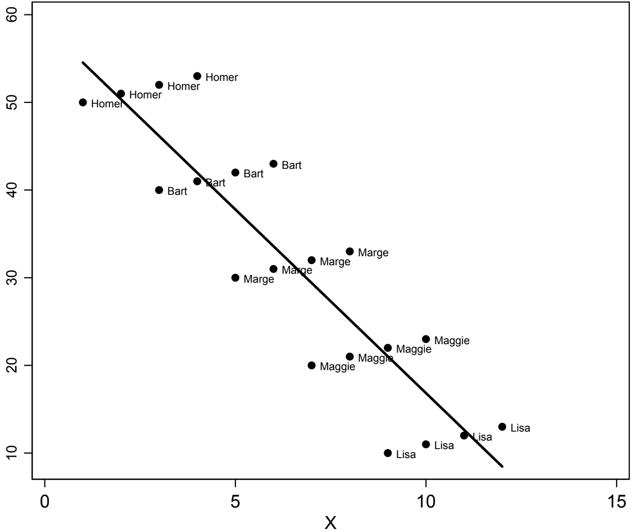
$$Y_{it} = \beta_{0t} + \beta_1 X_{it} + u_{it} \text{ (by time point)}$$

$$Y_{it} = \beta_{0it} + \beta_1 X_{it} + u_{it} \text{ (both)}$$

Varying Intercepts



Varying Intercepts



Varying Slopes (+ Intercepts)

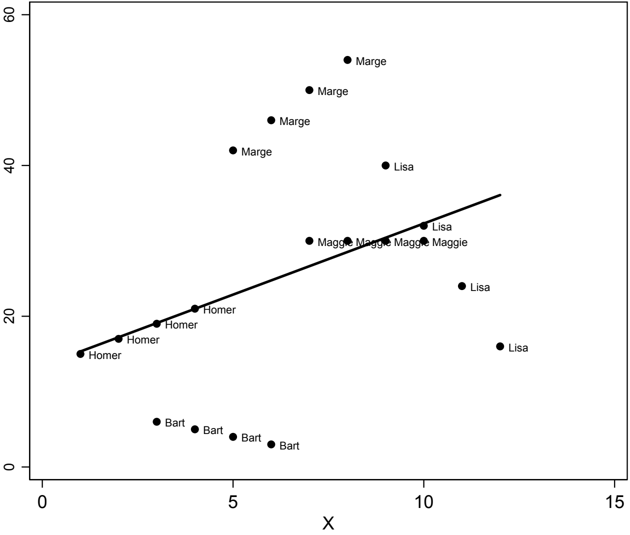
$$Y_{it} = \beta_0 + \beta_{1i} X_{it} + u_{it} \ \ \text{(slopes by unit)...}$$

$$Y_{it} = \beta_{0i} + \beta_{1i} X_{it} + u_{it} \ \ \text{(both by unit)}$$

$$Y_{it} = \beta_{0t} + \beta_{1t} X_{it} + u_{it} \ \ \text{(both by time point)}$$

$$Y_{it} = \beta_{0it} + \beta_{1it} X_{it} + u_{it} \ \ \text{(both by both)}$$

Varying Slopes + Intercepts



The Error

$$u_{it} \sim \text{i.i.d.} N(0, \sigma^2) \; \forall \; i, t$$

$$\begin{aligned}
\text{Var}(u_{it}) &= \text{Var}(u_{jt}) \; \forall \; i \neq j \text{ (i.e., no cross-unit heteroscedasticity)} \\
\text{Var}(u_{it}) &= \text{Var}(u_{is}) \; \forall \; t \neq s \text{ (i.e., no temporal heteroscedasticity)} \\
\text{Cov}(u_{it}, u_{js}) &= 0 \; \forall \; i \neq j, \; \forall \; t \neq s \text{ (i.e., no auto- or spatial correlation)}
\end{aligned}$$

“Unit Effects”

One- and Two-Way Unit Effects

Two-way variation:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \gamma V_i + \delta W_t + u_{it}$$

$\longrightarrow$ two-way effects:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + \eta_t + u_{it}$$

One-way effects:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \eta_t + u_{it} \quad \text{(time)}$$

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it} \quad \text{(units)}$$

"Brute force" model:

$$\begin{aligned} Y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it} \\ &= \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_1 I(i=1)_i + \alpha_2 I(i=2)_i + ... + u_{it} \end{aligned}$$

Alternatively:

$$\bar{X}_i = \frac{\sum_{N_i} X_{it}}{N_i}$$

and

$$\tilde{X}_{it} = X_{it} - \bar{X}_i.$$

Yields:

$$Y_{it} = \bar{\mathbf{X}}_i\boldsymbol{\beta}_B + (\mathbf{X}_{it} - \bar{X}_i)\boldsymbol{\beta}_W + \alpha_i + u_{it}$$

“Fixed” Effects

Means that:

$$
\begin{aligned}
Y^*_{it} &= Y_{it} - \bar{Y}_i \\
\mathbf{X}^*_{it} &= \mathbf{X}_{it} - \bar{\mathbf{X}}_i
\end{aligned}
$$

$$Y^*_{it} = \boldsymbol{\beta}_{FE}\mathbf{X}^*_{it} + u_{it}.$$

$\equiv$ “Within-Effects” Model.

Standard F-test for

$$H_0 : \alpha_i = \alpha_j \forall i \neq j$$

versus

$$H_A : \alpha_i \neq \alpha_j \text{ for some } i \neq j$$

is $\sim F_{N-1, NT-(N-1)}$.

An Example: Refugee Flows in Africa, 1992-2001

Data:

- 50 African countries $\rightarrow$ $(50 \times 49 =)$ 2450 directed dyads
- Ten years
- i indexes directed dyads, t indexes years

Model:

$$
\begin{aligned}
ln(\text{Refugees})_{A \rightarrow Bt} \quad = \quad & \beta_0 + \beta_1 \text{Population Difference}_{ABt} + \beta_2 \text{Distance}_{AB} + \\
& \beta_3 \text{POLITY Difference}_{ABt} + \beta_4 \text{War Difference}_{ABt} + u_{ABt}
\end{aligned}
$$

Data: Refugee Flows in Africa, 1992-2001

```
> summary(Refugees)
   dirdyadID           year        ln_ref_flow         pop_diff
 Min.   :404411   Min.   :1992   Min.   :-0.6931   Min.   :-0.117949
 1st Qu.:451461   1st Qu.:1994   1st Qu.:-0.6931   1st Qu.:-0.008848
 Median :510520   Median :1996   Median :-0.6931   Median : 0.000000
 Mean   :512160   Mean   :1996   Mean   :-0.6011   Mean   : 0.000000
 3rd Qu.:565553   3rd Qu.:1999   3rd Qu.:-0.6931   3rd Qu.: 0.008848
 Max.   :651625   Max.   :2001   Max.   :14.1343   Max.   : 0.117949
    distance        regimedif          wardiff    pop_between
 Min.   :0.000   Min.   :-1.00   Min.   :-4   Min.   :-0.109517
 1st Qu.:1.299   1st Qu.:-0.25   1st Qu.: 0   1st Qu.:-0.008833
 Median :2.169   Median : 0.00   Median : 0   Median : 0.000000
 Mean   :2.200   Mean   : 0.00   Mean   : 0   Mean   : 0.000000
 3rd Qu.:3.066   3rd Qu.: 0.25   3rd Qu.: 0   3rd Qu.: 0.008833
 Max.   :5.652   Max.   : 1.00   Max.   : 4   Max.   : 0.109517
   pop_within           regime_between   regime_within     war_between
 Min.   :-0.0088492   Min.   :-0.955   Min.   :-1.180   Min.   :-2.3
 1st Qu.:-0.0004707   1st Qu.:-0.225   1st Qu.:-0.085   1st Qu.:-0.4
 Median : 0.0000000   Median : 0.000   Median : 0.000   Median : 0.0
 Mean   : 0.0000000   Mean   : 0.000   Mean   : 0.000   Mean   : 0.0
 3rd Qu.: 0.0004707   3rd Qu.: 0.225   3rd Qu.: 0.085   3rd Qu.: 0.4
 Max.   : 0.0088492   Max.   : 0.955   Max.   : 1.180   Max.   : 2.3
   war_within
 Min.   :-2.5
 1st Qu.:-0.3
 Median : 0.0
 Mean   : 0.0
 3rd Qu.: 0.3
 Max.   : 2.5
```

An Example: Refugee Flows in Africa, 1992-2001

Pooled OLS:

```
> RefOLS<-lm(ln_ref_flow~pop_diff+distance+regimedif+wardiff, data=Refugees)
> summary(RefOLS)

Residuals:
    Min      1Q  Median      3Q     Max
-0.6114 -0.2109 -0.0857  0.0335 14.3756

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.3224073  0.0119195 -27.049   <2e-16 ***
pop_diff    -0.1732934  0.2166658  -0.800    0.424
distance    -0.1266528  0.0047016 -26.938   <2e-16 ***
regimedif   -0.0002476  0.0157962  -0.016    0.987
wardiff      0.0743220  0.0068169  10.903   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.9097 on 23613 degrees of freedom
Multiple R-squared: 0.03467,Adjusted R-squared: 0.03451
F-statistic:   212 on 4 and 23613 DF,  p-value: < 2.2e-16
```

An Example: Refugee Flows in Africa, 1992-2001

"Fixed" effects:

```
> library(plm)
> RefFE<-plm(ln_ref_flow~pop_diff+distance+regimedif+wardiff,
  data=Refugees, effect="individual", model="within")
> summary(RefFE)
Oneway (individual) effect Within Model

Unbalanced Panel: n=2450, T=1-10, N=23618

Residuals :
     Min.   1st Qu.    Median   3rd Qu.      Max.
-9.03e+00 -5.74e-03 -9.18e-06  5.72e-03  1.14e+01

Coefficients :
           Estimate Std. Error t-value Pr(>|t|)
pop_diff  6.8642028  2.5516636  2.6901 0.007149 **
regimedif 0.0050497  0.0223160  0.2263 0.820984
wardiff   0.0104144  0.0073673  1.4136 0.157493
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    8149.6
Residual Sum of Squares: 8146
R-Squared      :  0.00043949
      Adj. R-Squared :  0.00039385
F-statistic: 3.102 on 3 and 21165 DF, p-value: 0.025509
```

An Example: Refugee Flows in Africa, 1992-2001

Models of Refugees in Africa

Variable	OLS	Fixed Effects
Constant	-0.32	-
	(0.01)	
Population Difference	-0.17	6.86
	(0.22)	(2.55)
Distance	-0.13	(dropped)
	(0.005)	
POLITY Difference	-0.0002	0.005
	(0.016)	(0.022)
War Difference	0.074	0.010
	(0.007)	(0.007)
$\hat{\rho}$	-	0.61

Note: $NT = 23618$ ($N = 2450$, $\bar{T} = 9.6$)

Issues (?) with "Fixed" Effects

Pros:

- Specification Bias
- Intuitive
- Widely Used/Understood

Cons:

- Can't Estimate $\boldsymbol{\beta}_B$
- Slowly-Changing **X**s
- (In)Efficiency / Inconsistency ("Incidental Parameters")

"Between" Effects

From:

$$Y_{it} = \bar{\mathbf{X}}_i \boldsymbol{\beta}_B + (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\beta_W + \alpha_i + u_{it}.$$

"Between" effects:

$$\bar{Y}_i = \bar{\mathbf{X}}_i \boldsymbol{\beta}_B + u_{it}$$

- Essentially cross-sectional
- Based on N observations

Refugee Flows in Africa, 1992-2001

"Between" effects:

```
> RefBE<-plm(ln_ref_flow~pop_diff+distance+regimedif+wardiff, data=Refugees,
  effect="individual", model="between")
> summary(RefBE)
Oneway (individual) effect Between Model

Unbalanced Panel: n=2450, T=1-10, N=23618

Residuals :
   Min. 1st Qu.  Median 3rd Qu.    Max.
-0.5850 -0.2200 -0.0840  0.0534  9.6500

Coefficients :
             Estimate Std. Error  t-value  Pr(>|t|)
(Intercept) -0.299703   0.029741 -10.0771 < 2.2e-16 ***
pop_diff    -0.246861   0.525232  -0.4700    0.6384
distance    -0.134874   0.011755 -11.4742 < 2.2e-16 ***
regimedif    0.010709   0.045117   0.2374    0.8124
wardiff      0.124185   0.022004   5.6439 1.855e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    1383.9
Residual Sum of Squares: 1296.7
R-Squared      :  0.063042
      Adj. R-Squared :  0.062913
F-statistic: 41.1269 on 4 and 2445 DF, p-value: < 2.22e-16
```

Refugee Example Redux

Variable	OLS	Fixed ("Within") Effects	Between Effects
Constant	-0.32	-	-0.30
	(0.01)		(0.03)
Population Difference	-0.17	6.86	-0.25
	(0.22)	(2.55)	(0.53)
Distance	-0.13	(dropped)	-0.13
	(0.005)		(0.01)
POLITY Difference	-0.0002	0.005	0.01
	(0.016)	(0.022)	(0.05)
War Difference	0.074	0.010	0.12
	(0.007)	(0.007)	(0.02)
$\hat{\rho}$	-	0.61	-

Note: $NT = 23618$ ($N = 2450$, $\bar{T} = 9.6$).

“Random” Effects

Model:

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

with:

$$u_{it} = \alpha_i + \lambda_t + \eta_{it}$$

and

$$\begin{aligned}
E(\alpha_i) = E(\lambda_t) = E(\eta_{it}) &= 0, \\
E(\alpha_i\lambda_t) = E(\alpha_i\eta_{it}) = E(\lambda_t\eta_{it}) &= 0, \\
E(\alpha_i\alpha_j) &= \sigma^2_\alpha \text{ if } i = j,\ 0 \text{ otherwise}, \\
E(\lambda_t\lambda_s) &= \sigma^2_\lambda \text{ if } t = s,\ 0 \text{ otherwise}, \\
E(\eta_{it}\eta_{js}) &= \sigma^2_\eta \text{ if } i = j,\ t = s,\ 0 \text{ otherwise}, \\
E(\alpha_i\mathbf{X}_{it}) = E(\lambda_t\mathbf{X}_{it}) = E(\eta_{it}\mathbf{X}_{it}) &= 0.
\end{aligned}$$

"Random" Effects

"Variance Components":

$$Var(Y_{it}|\mathbf{X}_{it}) = \sigma^2_\alpha + \sigma^2_\lambda + \sigma^2_\eta$$

If we assume $\lambda_t = 0$, then we get a model like:

$$Y_{it} = \mathbf{X}_{it}\beta + \alpha_i + \eta_{it}$$

with total error variance:

$$\sigma^2_u = \sigma^2_\alpha + \sigma^2_\eta.$$

"Random" Effects: Estimation

$$
\begin{aligned}
E(\mathbf{u}_i\mathbf{u}_i') \equiv \boldsymbol{\Sigma}_i &= \sigma_\eta^2 \mathbf{I}_T + \sigma_\alpha^2 \mathbf{i}\mathbf{i}' \\
&= \begin{pmatrix} \sigma_\eta^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\eta^2 + \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\eta^2 + \sigma_\alpha^2 \end{pmatrix}
\end{aligned}
$$

$$
\mathrm{Var}(\mathbf{u}) \equiv \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Sigma}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Sigma}_N \end{pmatrix}
$$

"Random" Effects: Estimation

Can estimate:

$$\boldsymbol{\Sigma}^{-1/2} = \frac{1}{\sigma_\eta}\left[\mathbf{I}_T - \left(\frac{\theta}{T}\mathbf{i}\mathbf{i}'\right)\right]$$

where

$$\theta = 1 - \sqrt{\frac{\sigma^2_\eta}{T\sigma^2_\alpha + \sigma^2_\eta}}.$$

With $\hat{\theta}$, calculate:

$$\begin{aligned} Y^*_{it} &= Y_{it} - \hat{\theta}\bar{Y}_i \\ X^*_{it} &= X_{it} - \hat{\theta}\bar{X}_i, \end{aligned}$$

estimate:

$$Y^*_{it} = (1-\hat{\theta})\alpha + X^*_{it}\beta_{RE} + [(1-\hat{\theta})\alpha_i + (\eta_{it} - \hat{\theta}\bar{\eta}_i)]$$

and iterate...

"Random" Effects: An Alternative View

Refugees Redux

```
> RefRE<-plm(ln_ref_flow~pop_diff+distance+regimedif+wardiff, data=Refugees,
   effect="individual", model="random")
> summary(RefRE)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Unbalanced Panel: n=2450, T=1-10, N=23618

Effects:
                 var std.dev share
idiosyncratic 0.3849  0.6204 0.466
individual    0.4416  0.6645 0.534
theta  :
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3176  0.7168  0.7168  0.7141  0.7168  0.7168

Coefficients :
              Estimate Std. Error  t-value  Pr(>|t|)
(Intercept) -0.3063941  0.0285299 -10.7394 < 2.2e-16 ***
pop_diff     0.0638665  0.4974613   0.1284  0.897845
distance    -0.1324536  0.0112685 -11.7544 < 2.2e-16 ***
regimedif    0.0005633  0.0198580   0.0284  0.977370
wardiff      0.0228523  0.0069775   3.2751  0.001058 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    9216.6
Residual Sum of Squares: 9158.9
R-Squared      :  0.0062699
Adj. R-Squared :  0.0062686
F-statistic: 37.177 on 4 and 23613 DF, p-value: < 2.22e-16
```

Refugees Redux, Remix

```
> library(lme4)

> AltRefRE<-lmer(ln_ref_flow~pop_diff+distance+regimedif+wardiff+(1|dirdyadID), data=Refugees)
> summary(AltRefRE)
Linear mixed model fit by REML
Formula: ln_ref_flow ~ pop_diff + distance + regimedif + wardiff + (1 |      dirdyadID)
   Data: Refugees
   AIC   BIC logLik deviance REMLdev
 50733 50790 -25360    50692   50719
Random effects:
 Groups    Name        Variance Std.Dev.
 dirdyadID (Intercept) 0.46653  0.68303
 Residual              0.38592  0.62123
Number of obs: 23618, groups: dirdyadID, 2450

Fixed effects:
              Estimate Std. Error t value
(Intercept) -0.3061471  0.0291477 -10.503
pop_diff     0.0758989  0.5075942   0.150
distance    -0.1325429  0.0115127 -11.513
regimedif    0.0007138  0.0199078   0.036
wardiff      0.0223476  0.0069779   3.203

Correlation of Fixed Effects:
          (Intr) pp_dff distnc regmdf
pop_diff   0.000
distance  -0.869  0.000
regimedif  0.000  0.036  0.000
wardiff    0.000 -0.004  0.000  0.109
```

Variable	OLS	Fixed Effects	Between Effects	Random Effects
Constant	-0.32	-	-0.30	-0.31
	(0.01)		(0.03)	(0.03)
Population Difference	-0.17	6.86	-0.25	0.09
	(0.22)	(2.55)	(0.53)	(0.52)
Distance	-0.13	(dropped)	-0.13	-0.13
	(0.005)		(0.01)	(0.01)
POLITY Difference	-0.0002	0.005	0.01	0.0005
	(0.016)	(0.022)	(0.05)	(0.0199)
War Difference	0.074	0.010	0.12	0.023
	(0.007)	(0.007)	(0.02)	(0.007)
$\hat{\rho}$	-	0.61	-	0.56

Note: $NT = 23618$ ($N = 2450$, $\bar{T} = 9.6$).

"Random" Effects: Testing

Hausman test (FE vs. RE):

$$\hat{W} = (\hat{\beta}_{\mathbf{FE}} - \hat{\beta}_{\mathbf{RE}})'(\hat{\mathbf{V}}_{\mathbf{FE}} - \hat{\mathbf{V}}_{\mathbf{RE}})^{-1}(\hat{\beta}_{\mathbf{FE}} - \hat{\beta}_{\mathbf{RE}})$$

$$W \sim \chi^2_k$$

Issues:

- Asymptotic
- No guarantee $(\hat{\mathbf{V}}_{\mathbf{FE}} - \hat{\mathbf{V}}_{\mathbf{RE}})^{-1}$ is positive definite
- A general specification test...

Example:

```
> phtest(RefFE, AltRefRE)

Hausman Test

data:  ln_ref_flow ~ pop_diff + distance + regimedif + wardiff
chisq = 34.712, df = 3, p-value = 0.0000001401
alternative hypothesis: one model is inconsistent
```

Practical “Fixed” vs. “Random” Effects

- “Panel” vs. “TSCS” Data
- Data-Generating Process
- Covariate Effects

Separating Within and Between Effects

$$Y_{it} = \bar{\mathbf{X}}_i \boldsymbol{\beta}_B + (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\boldsymbol{\beta}_W + u_{it}$$

- Simple...
- Easy interpretation
- Easy to test $\hat{\boldsymbol{\beta}}_B = \hat{\boldsymbol{\beta}}_W$

Again With The Refugees...

Variable	Estimate
Constant	-0.32
	(0.01)
Distance	-0.13
	(0.004)
Between (Mean) Population Difference	-0.22
	(0.22)
Within Population Difference	6.86
	(3.74)
Between (Mean) POLITY Difference	0.01
	(0.02)
Within POLITY Difference	0.005
	(0.032)
Between (Mean) War Difference	0.12
	(0.01)
Within War Difference	0.01
	(0.01)

Note: $NT = 23618$ ($N = 2450$, $\bar{T} = 9.6$).

Unit Effects Models: Software

R :

- the `lme4` package; command is `lmer`
- the `plm` package; `plm` command
- the `nlme` package; command `lme`

Stata : `xtreg`

- the `re` (the default) = random effects
- the `fe` = fixed (within) effects
- the `be` = between-effects

Other Panel / TSCS Models

- Dynamic models
  - GLS-ARMA models
  - Lagged-response / distributed lag models
  - "Panel Time-Series" (panel unit roots, etc.)
- Conditional / "unit-specific" models: HLMs, etc.
- Marginal / "population averaged" models: GEEs
- Models for binary / ordinal / event count responses in panel data