GSERM - St. Gallen 2022
Analyzing Panel Data

June 7, 2022

Two-Way Variation

Two-way variation:

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

where V_i don't vary over time (within a unit), and W_i don't vary across units (for a given time point).

Note that we can write:

$$\alpha_i = \sum(\gamma V_i)$$

and

$$\eta_t = \sum(\delta W_t).$$

So:

$$
\begin{aligned}
Y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta} + \gamma V_i + \delta W_t + u_{it} \\
&= \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + \eta_t + u_{it}
\end{aligned}
$$

One- and Two-Way "Unit Effects"

"Two-way" unit effects:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + \eta_t + u_{it}$$
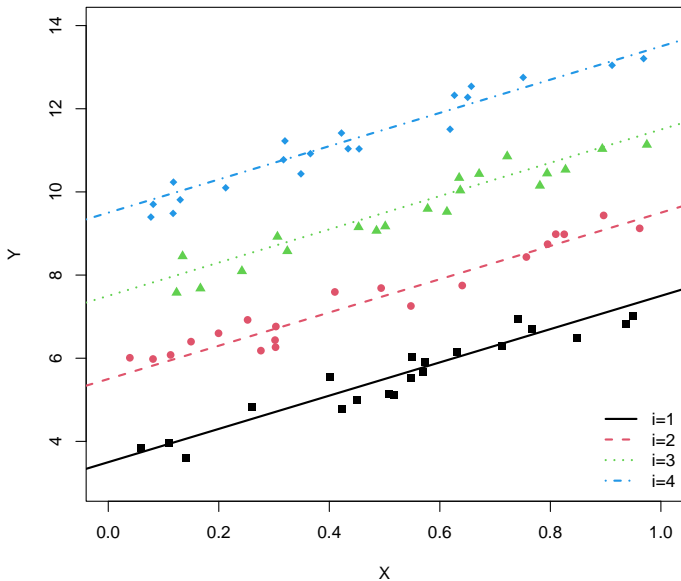
"One-way" effects:

- Assuming $\alpha_i = 0$ (w.l.o.g):

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

- Assuming $\eta_t = 0$ (w.l.o.g):

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

"Brute force" model fits:

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

Alternatively, decompose:

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

and

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

Yields:

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

But!

$$
\text{corr}(\bar{\mathbf{X}}_i \boldsymbol{\beta}_B, \alpha_i) = 1.0
$$

Means that:

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

gives:

$$
Y_{it}^* = \mathbf{X}_{it}^* \beta_{FE} + u_{it}.
$$

$\rightarrow$ **A "Fixed Effects" Model is actually a "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 $\sim F_{N-1, NT-(N-1)}$.

Running Example Data: WDI, 1960-2021

The World Development Indicators

- Cross-national country-level time series data
- $N = 215$ countries, $T = 72$ years (1960-2021) + missingness
- Variables:
    - Geography: land area, arable land
    - Population indicators
    - Demographics: Birth rates, life expectancy, etc.
    - Economics: GDP, inflation, trade, FDI, etc.
    - Governments: expenditures, policies, etc.

- Full descriptions are listed in the Github repo here

```
> describe(wdi,fast=TRUE,ranges=FALSE,check=TRUE)
                        vars     n            mean                 sd               se
ISO3                       1 13330             NaN                 NA               NA
Year                       2 13330         1990.50              17.90             0.16
Region                     3 13330             NaN                 NA               NA
country                    4 13330             NaN                 NA               NA
LandArea                   5 12906       613525.38         1766486.19         15549.43
ArablePercent              6 10935           13.51              13.49             0.13
Population                 7 13073     24638671.91       103129756.53        901978.87
PopGrowth                  8 12856            1.79               1.68             0.01
RuralPopulation            9 13045           48.61              25.74             0.23
UrbanPopulation           10 13045           51.39              25.74             0.23
BirthRatePer1K            11 12112           28.32              13.10             0.12
FertilityRate             12 11847            3.97               2.01             0.02
PrimarySchoolAge          13 10696            6.14               0.62             0.01
LifeExpectancy            14 11829           64.37              11.46             0.11
AgeDepRatioOld            15 11731           10.34               6.36             0.06
CO2Emissions              16  5535            4.28               5.41             0.07
GDP                       17  9585 242308268086.15 1101606170189.79 11252014966.83
GDPPerCapita              18  9582        11685.74           18675.05           190.78
GDPPerCapGrowth           19  9598            1.89               6.21             0.06
Inflation                 20  8275           23.88             332.39             3.65
TotalTrade                21  8363           78.18              54.14             0.59
Exports                   22  8363           36.41              28.85             0.32
Imports                   23  8372           41.78              27.87             0.30
FDIIn                     24  8195            5.57              45.42             0.50
AgriEmployment            25  5394           29.67              24.35             0.33
NetAidReceived            26  8633     453209476.19       867754118.55       9339331.98
MobileCellSubscriptions   27  9849           33.70              50.29             0.51
NaturalResourceRents      28  8745            6.61              10.87             0.12
MilitaryExpenditures      29  7333            2.74               3.23             0.04
GovtExpenditures          30  8012           16.20               8.19             0.09
HIVDeaths                 31  4041         7221.50           19857.71           312.38
WomenBusLawIndex          32  9776           59.23              18.46             0.19
PaidParentalLeave         33  9776            0.10               0.30             0.00
ColdWar                   34 13330            0.48               0.50             0.00
```
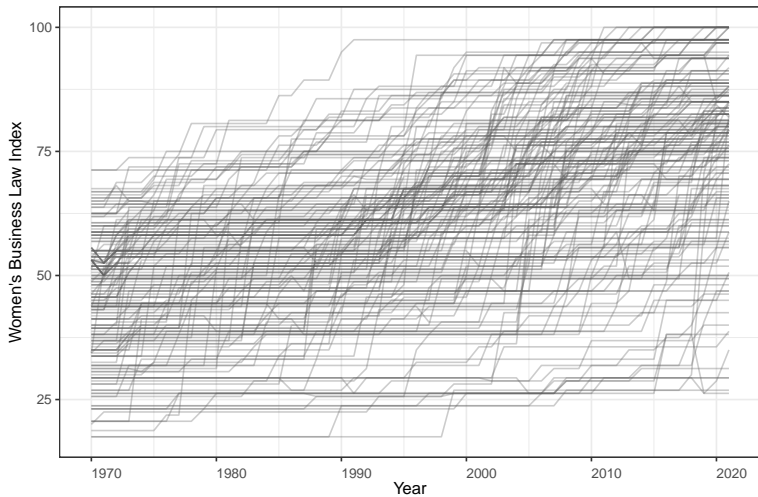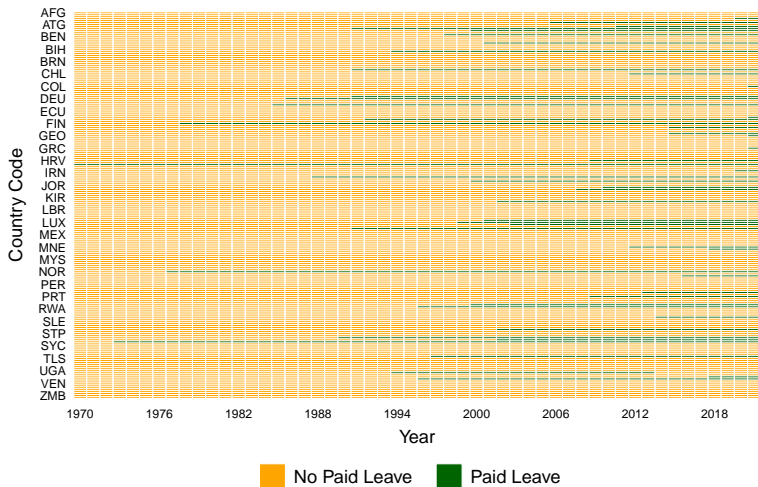
Categorical Variable Visualization

WDI's Women, Business and the Law Index (WBLI)

The basis for a 2021 World Bank report...

- Examines "the laws and regulations that affect women's economic opportunity in 190 economies."
- An index comprising eight indicators "structured around women's interactions with the law as they move through their careers: *Mobility*, *Workplace*, *Pay*, *Marriage*, *Parenthood*, *Entrepreneurship*, *Assets*, and *Pension*."
- The WBL Index:
  - Theoretically ranges from 0 - 100
  - In practice: Lowest values $\approx$ 20
  - Higher values correspond to higher levels of women's empowerment and greater opportunities and support for women, particularly in business

- "Better performance in the areas measured by the Women, Business and the Law index is associated with a more narrow gender gap in development outcomes, higher female labor force participation, lower vulnerable employment, and greater representation of women in national parliaments."

```
> WDI<-pdata.frame(wdi)
> WBLI<-WDI$WomenBusLawIndex
> class(WBLI)
[1] "pseries" "numeric"

> describe(WBLI,na.rm=TRUE) # all variation

    vars    n mean   sd median trimmed  mad  min max range skew kurtosis   se
X1     1 9776 59.2 18.5   58.8    59.2 18.5 17.5 100  82.5 0.04    -0.55 0.19
```
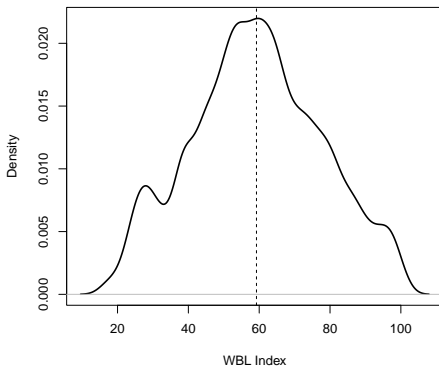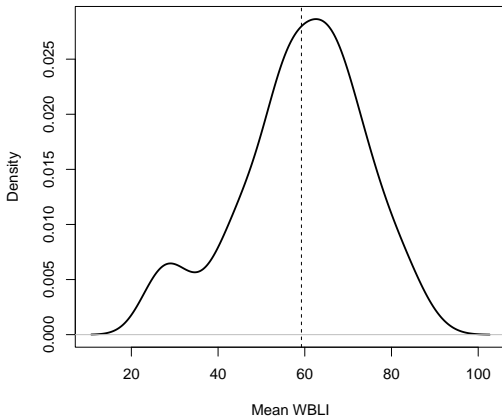
```
> describe(plm::between(WBLI,effect="individual",na.rm=TRUE)) # "between" variation

    vars   n mean   sd median trimmed  mad  min max range  skew kurtosis   se
X1     1 188 59.2 14.4   60.5    60.1 13.3 23.4  90  66.6 -0.47    -0.13 1.05
```
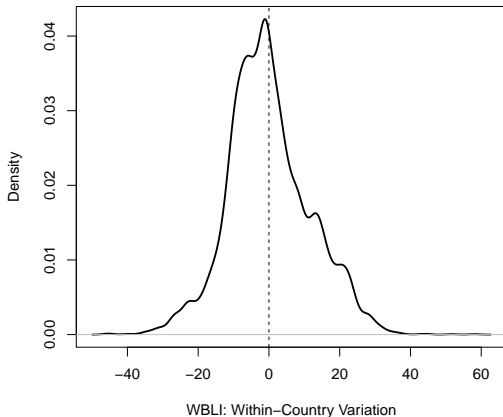
```
> describe(Within(WBLI,na.rm=TRUE)) # "within" variation

   vars    n mean   sd median trimmed  mad   min max range skew kurtosis   se
X1    1 9776    0 11.5  -0.93   -0.33 10.4 -45.3  58   103 0.26     0.45 0.12
```


WBLI: Within−Country Variation

A Regression Model

Regression model:

$$\text{WBLI}_{it} = \beta_0 + \beta_1\text{Population Growth}_{it} + \beta_2\text{Urban Population}_{it} + \beta_3\text{Fertility Rate}_{it} + \beta_4\ln(\text{GDP Per Capita})_{it} + \beta_5\text{Natural Resource Rents}_{it} + \beta_6\text{Cold War}_t + u_{it}$$

Descriptive Statistics:

	vars	n	mean	sd	min	max	range	se
WomenBusLawIndex	1	7566	59.99	18.79	17.50	100.00	82.50	0.22
PopGrowth	2	7566	1.70	1.46	-6.77	17.51	24.28	0.02
UrbanPopulation	3	7566	51.19	23.90	2.85	100.00	97.16	0.27
FertilityRate	4	7566	3.67	1.91	0.90	8.61	7.70	0.02
NaturalResourceRents	5	7566	6.82	10.45	0.00	87.51	87.51	0.12
ColdWar	6	7566	0.32	0.47	0.00	1.00	1.00	0.01
lnGDPPerCap	7	7566	8.28	1.44	5.32	11.63	6.31	0.02

```
> OLS<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+                log(GDPPerCapita)+NaturalResourceRents+ColdWar,
+        data=WDI,model="pooling")

> summary(OLS)
Pooling Model

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, model = "pooling")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -46.94   -8.02    1.05    9.30   49.73

Coefficients:
                      Estimate Std. Error t-value   Pr(>|t|)
(Intercept)           54.8395     1.7261   31.77    < 2e-16 ***
PopGrowth             -3.1926     0.1437  -22.21    < 2e-16 ***
UrbanPopulation       -0.0584     0.0109   -5.37 0.000000083 ***
FertilityRate         -1.7928     0.1652  -10.85    < 2e-16 ***
log(GDPPerCapita)      3.1544     0.1993   15.83    < 2e-16 ***
NaturalResourceRents  -0.3486     0.0162  -21.54    < 2e-16 ***
ColdWar              -11.3437     0.3716  -30.53    < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    2670000
Residual Sum of Squares: 1280000
R-Squared:      0.519
Adj. R-Squared: 0.519
F-statistic: 1361.01 on 6 and 7559 DF, p-value: <2e-16
```

```
> FE<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+          log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+          effect="individual",model="within")

> summary(FE)
Oneway (individual) effect Within Model

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, effect = "individual", model = "within")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-33.144  -4.832  -0.406   4.802  41.184

Coefficients:
                     Estimate Std. Error t-value Pr(>|t|)
PopGrowth              0.0735     0.1194    0.62   0.538
UrbanPopulation        0.2482     0.0205   12.09  <2e-16 ***
FertilityRate         -2.0657     0.1657  -12.47  <2e-16 ***
log(GDPPerCapita)      9.1607     0.3104   29.51  <2e-16 ***
NaturalResourceRents   0.0353     0.0182    1.94   0.052 .
ColdWar               -7.1920     0.2951  -24.37  <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    934000
Residual Sum of Squares: 434000
R-Squared:      0.535
Adj. R-Squared: 0.523
F-statistic: 1414.37 on 6 and 7374 DF, p-value: <2e-16
```

Table: Models of WBLI

	OLS	FE
Population Growth	−3.190***	0.073
	(0.144)	(0.119)
Urban Population	−0.058***	0.248***
	(0.011)	(0.021)
Fertility Rate	−1.790***	−2.070***
	(0.165)	(0.166)
ln(GDP Per Capita)	3.150***	9.160***
	(0.199)	(0.310)
Natural Resource Rents	−0.349***	0.035*
	(0.016)	(0.018)
Cold War	−11.300***	−7.190***
	(0.372)	(0.295)
Constant	54.800***	
	(1.730)	
Observations	7,566	7,566
R^2	0.519	0.535
Adjusted R^2	0.519	0.523
F Statistic	1,361.000*** (df = 6; 7559)	1,414.000*** (df = 6; 7374)

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The model is:

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

which is estimated via:

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

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

Comparison: Unit vs. Time Fixed Effects

Table: FE Models of WBLI (Units vs. Time)

	FE.Units	FE.Time
Population Growth	0.073	-3.630^{***}
	(0.119)	(0.135)
Urban Population	0.248^{***}	-0.054^{***}
	(0.021)	(0.010)
Fertility Rate	-2.070^{***}	-0.657^{***}
	(0.166)	(0.159)
ln(GDP Per Capita)	9.160^{***}	3.550^{***}
	(0.310)	(0.187)
Natural Resource Rents	0.035^{*}	-0.400^{***}
	(0.018)	(0.015)
Cold War	-7.190^{***}	
	(0.295)	
Observations	7,566	7,566
R^2	0.535	0.439
Adjusted R^2	0.523	0.435
F Statistic	$1,414.000^{***}$ (df = 6; 7374)	$1,175.000^{***}$ (df = 5; 7511)

*p$<$0.1; **p$<$0.05; ***p$<$0.01

The specification:

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

...suggests that we can use an F-test to examine the hypothesis:

$$H_0 : \alpha_i = 0 \; \forall \; i$$

(and a similar test for $\eta_t = 0$ in the time-centered case).

Arguably better are Lagrange multiplier-based tests:

- Breusch-Pagan (1980)
- King and Wu (1997)
- See (e.g.) Croissant and Millo (2018, §4.1) for details

```
> pFtest(FE,OLS)

F test for individual effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
F = 78, df1 = 185, df2 = 7374, p-value <2e-16
alternative hypothesis: significant effects


> plmtest(FE,effect=c("individual"),type=c("bp"))

Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
chisq = 44635, df = 1, p-value <2e-16
alternative hypothesis: significant effects


> plmtest(FE,effect=c("individual"),type=c("kw"))

Lagrange Multiplier Test - (King and Wu) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 211, p-value <2e-16
alternative hypothesis: significant effects
```

```
> pFtest(FE.Time,OLS)

F test for time effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
F = 25, df1 = 48, df2 = 7511, p-value <2e-16
alternative hypothesis: significant effects


> plmtest(FE.Time,effect=c("time"),type=c("bp"))

Lagrange Multiplier Test - time effects (Breusch-Pagan) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
chisq = 9920, df = 1, p-value <2e-16
alternative hypothesis: significant effects


> plmtest(FE.Time,effect=c("time"),type=c("kw"))

Lagrange Multiplier Test - time effects (King and Wu) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
normal = 100, p-value <2e-16
alternative hypothesis: significant effects
```

Fixed Effects: Interpretation

FE models are *subject-specific* models, that rely entirely on *within-unit* variability to estimate variable effects.

- This means that:

$$\hat{\beta}_k = \frac{\partial E(Y|\hat{\alpha})}{\partial X_k}$$

- That is, $\hat{\beta}_k$ is *the expected change in E(Y) associated with a one-unit increase in <u>observation i's</u> value of X_k*
- Key: *within-unit* changes in **X** are associated with *within-unit* expected changes in Y.
- In a linear model, the value of $\hat{\alpha}$ doesn't affect the value of that partial derivative...

Fixed Effects: Interpretation

Mummolo and Peterson (2018) note that:

> "...because the within-unit variation is always smaller (or at least, no larger) than the overall variation in the independent variable, researchers should use within-unit variation to motivate counterfactuals when discussing the substantive impact of a treatment" (2018, 829).

Significance:

- Predictors **X** in FE models typically have both cross-sectional and temporal variation
- FE models only consider *within-unit* variation in **X** and Y
- As a result, the degree of actual/observed (within-unit) variation in predictors is almost always less – and sometimes significantly less – than if cross-sectional variation were also considered

Interpretation Example: Urban Population

`UrbanPopulation` – All Variation:

```
> with(WDI, sd(UrbanPopulation,na.rm=TRUE)) # all variation
[1] 25.7
```

`UrbanPopulation` – "Within" Variation:

```
> WDI<-ddply(WDI, .(ISO3), mutate,
+                 UPMean = mean(UrbanPopulation,na.rm=TRUE))
> WDI$UPWithin<-with(WDI, UrbanPopulation-UPMean)
>
> with(WDI, sd(UPWithin,na.rm=TRUE)) # "within" variation
[1] 8.86
```

> *"While the overall variation in the independent variable may be large, the within-unit variation used to estimate β may be much smaller"* (M & P 2018, 830).

Pros and Cons of "Fixed" Effects

Pros:

- Mitigates (Some) Specification Bias
- Simple $+$ Intuitive
- Widely Used/Understood

Cons (see e.g. Collischon and Eberl 2020):

- Can't Estimate β_B
- Slowly-Changing **X**s
- (In)Efficiency / Inconsistency (Incidental Parameters)
- Cannot Control for (e.g.) Time-Varying Heterogeneity
- Sensitivity to Measurement Error

From the equation above:

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

...we can derive a "Between Effects" model:

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

This model:

- is essentially cross-sectional,
- is based on N observations,
- considers *only* between-unit (average) differences
- Interpretation:

  $\hat{\beta}_B$ *is the expected difference in Y between two units whose values on $\bar{X}$ differ by a value of 1.0.*

```
> BE<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+                 log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+           effect="individual",model="between")

> summary(BE)
Oneway (individual) effect Between Model

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, effect = "individual", model = "between")

Unbalanced Panel: n = 186, T = 1-50, N = 7566
Observations used in estimation: 186

Residuals:
   Min. 1st Qu. Median 3rd Qu.    Max.
-30.423  -6.319  0.332   8.067  22.183

Coefficients:
                    Estimate Std. Error t-value   Pr(>|t|)
(Intercept)          50.3106    10.1621    4.95 0.0000016996 ***
PopGrowth            -5.8648     0.9173   -6.39 0.0000000014 ***
UrbanPopulation      -0.0497     0.0549   -0.91        0.366
FertilityRate         0.0633     1.0766    0.06        0.953
log(GDPPerCapita)     3.4052     1.1308    3.01        0.003 **
NaturalResourceRents -0.3579     0.0888   -4.03 0.000823194 ***
ColdWar             -12.5095     4.8879   -2.56        0.011 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    44700
Residual Sum of Squares: 17600
R-Squared:      0.605
Adj. R-Squared: 0.592
F-statistic: 45.7424 on 6 and 179 DF, p-value: <2e-16
```

Table: Models of WBLI

	OLS	FE	BE
Population Growth	−3.190***	0.073	−5.870***
	(0.144)	(0.119)	(0.917)
Urban Population	−0.058***	0.248***	−0.050
	(0.011)	(0.021)	(0.055)
Fertility Rate	−1.790***	−2.070***	0.063
	(0.165)	(0.166)	(1.080)
ln(GDP Per Capita)	3.150***	9.160***	3.400***
	(0.199)	(0.310)	(1.130)
Natural Resource Rents	−0.349***	0.035*	−0.358***
	(0.016)	(0.018)	(0.089)
Cold War	−11.300***	−7.190***	−12.500**
	(0.372)	(0.295)	(4.890)
Constant	54.800***		50.300***
	(1.730)		(10.200)
Observations	7,566	7,566	186
R^2	0.519	0.535	0.605
Adjusted R^2	0.519	0.523	0.592
F Statistic	1,361.000*** (df = 6; 7559)	1,414.000*** (df = 6; 7374)	45.700*** (df = 6; 179)

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Model:

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

with:

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

and

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

If those assumptions are met, we can consider the "two-way variance components" model where:

$$
\begin{aligned}
Var(u_{it}) &= Var(Y_{it}|\mathbf{X}_{it}) \\
&= \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_\eta^2
\end{aligned}
$$

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:

$$
\sigma_u^2 = \sigma_\alpha^2 + \sigma_\eta^2.
$$

"Random" Effects: Estimation

The model above will violate the standard OLS assumptions of uncorrelated errors, because the (compound) "errors" u_{it} within each unit share a common component α_i.

Consider the within-i variance-covariance matrix of the errors $\mathbf{u}$:

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

Assuming conditional independence across units, we then have:

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

We can then show that:

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

where

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

is an unknown quantity to be estimated.

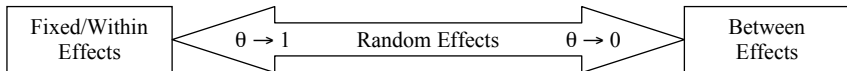Starting with an estimate of $\hat{\theta}$, calculate:

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

then 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 between the two processes until convergence.

"Random" Effects: An Alternative View

Fixed/Within Effects			Between Effects
	$\theta \to 1$ Random Effects $\theta \to 0$		

```
> RE<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+           log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+         effect="individual",model="random")

> summary(RE)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, effect = "individual", model = "random")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Effects:
               var std.dev share
idiosyncratic 58.91    7.68   0.4
individual    88.75    9.42   0.6
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.368   0.872   0.886   0.876   0.886   0.886

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -33.3    -5.2    -0.4     0.0     5.5    40.9

Coefficients:
                     Estimate Std. Error z-value  Pr(>|z|)
(Intercept)           2.56506    2.58528    0.99      0.32
PopGrowth            -0.07305    0.12134   -0.60      0.55
UrbanPopulation       0.12838    0.01920    6.69 0.000000000023 ***
FertilityRate        -2.37359    0.16536  -14.35   < 2e-16 ***
log(GDPPerCapita)     7.51273    0.29275   25.66   < 2e-16 ***
NaturalResourceRents  0.00581    0.01820    0.32      0.75
ColdWar              -8.48014    0.28934  -29.31   < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    982000
Residual Sum of Squares: 471000
R-Squared:      0.52
Adj. R-Squared: 0.52
```

A Nicer Table (Yet Again)

Table: Models of WBLI

	OLS	FE	BE	RE
Population Growth	−3.190***	0.073	−5.870***	−0.073
	(0.144)	(0.119)	(0.917)	(0.121)
Urban Population	−0.058***	0.248***	−0.050	0.128***
	(0.011)	(0.021)	(0.055)	(0.019)
Fertility Rate	−1.790***	−2.070***	0.063	−2.370***
	(0.165)	(0.166)	(1.080)	(0.165)
ln(GDP Per Capita)	3.150***	9.160***	3.400***	7.510***
	(0.199)	(0.310)	(1.130)	(0.293)
Natural Resource Rents	−0.349***	0.035*	−0.358***	0.006
	(0.016)	(0.018)	(0.089)	(0.018)
Cold War	−11.300***	−7.190***	−12.500**	−8.480***
	(0.372)	(0.295)	(4.890)	(0.289)
Constant	54.800***		50.300***	2.560
	(1.730)		(10.200)	(2.580)
Observations	7,566	7,566	186	7,566
R^2	0.519	0.535	0.605	0.520
Adjusted R^2	0.519	0.523	0.592	0.520
F Statistic	1,361.000*** (df = 6; 7559)	1,414.000*** (df = 6; 7374)	45.700*** (df = 6; 179)	7,910.000***

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Intuition:

- RE models require that $\text{Cov}(X_{it}, \alpha_i) = 0$.
- FE models do not.

This means that:

	Reality	
Model	$\text{Cov}(X_{it}, \alpha_i) = 0$	$\text{Cov}(X_{it}, \alpha_i) \neq 0$
Fixed Effects	Consistent, Inefficient	Consistent, Efficient
Random Effects	Consistent, Efficient	Inconsistent

Hausman test (FE vs. RE):

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

$$W \sim \chi_k^2$$

Null: The RE model is consistent ($\text{Cov}(X_{it}, \alpha_i) = 0$).

<u>Issues</u>:

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

Hausman test (FE vs. RE):

```
> phtest(FE, RE)

Hausman Test

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
chisq = 946, df = 6, p-value <2e-16

alternative hypothesis: one model is inconsistent
```

Practical "Fixed" vs. "Random" Effects

Factors to consider:

- "Panel" vs. "TSCS" Data

- Nature of the Data-Generating Processes

- Importance of Non-Time-Varying Covariate Effects

- Dimension/Location of the Variance in the Data (N vs. T)

Connections: Hierarchical Linear Models

Begin by considering a two-level "nested" data structure, with:

$$i \in \{1, 2, ...N\} \text{ indexing first-level units, and}$$
$$j \in \{1, 2, ...J\} \text{ indexing second-level groups.}$$

A general two-level HLM is an equation of the form:

$$Y_{ij} = \beta_{0j} + \mathbf{X}_{ij}\boldsymbol{\beta}_j + u_{ij} \tag{1}$$

where β_{0j} is a "constant" term, $\mathbf{X}_{ij}$ is a $NJ \times K$ matrix of K covariates, $\boldsymbol{\beta}_j$ is a $K \times 1$ vector of parameters, and $u_{ij} \sim$ i.i.d. $N(0, \sigma_u^2)$ is the usual random-disturbance assumption.

Each of the $K + 1$ "level-one" parameters is then allowed to vary across Q "level-two" variables $\mathbf{Z}_j$, so that:

$$\beta_{0j} = \gamma_{00} + \mathbf{Z}_j \gamma_0 + \varepsilon_{0j} \tag{2}$$

for the "intercept" and

$$\beta_{kj} = \gamma_{k0} + \mathbf{Z}_j \gamma_k + \varepsilon_{kj} \tag{3}$$

for the "slopes" of $\mathbf{X}$. The εs are typically assumed to be distributed multivariate Normal, with parameters for the variances and covariances selected by the analyst. Substitution of (2) and (3) into (1) yields:

$$Y_{ij} = \gamma_{00} + \mathbf{Z}_j \gamma_0 + \mathbf{X}_{ij} \gamma_{k0} + \mathbf{X}_{ij} \mathbf{Z}_j \gamma_k + \mathbf{X}_{ij} \varepsilon_{kj} + \varepsilon_{0j} + u_{ij} \tag{4}$$

The form is essentially a model with "saturated" interaction effects across the various levels, as well as "errors" which are multivariate Normal.

HLM Details

Model Assumptions

- Linearity / Additivity
- Normality of us
- Homoscedasticity
- Residual Independence:
  - $\mathrm{Cov}(\varepsilon_{\cdot j}, u_{ij}) = 0$
  - $\mathrm{Cov}(u_{ij}, u_{i\ell}) = 0$

Model Fitting

- MLE
- "Restricted" MLE ("RMLE")
- Choosing:
  - MLE is biased in small samples, especially for estimating variances
  - RMLE is not, but prevents use of LR tests when the models do not have identical fixed effects
  - In general: RMLE is better with small sample sizes, but MLE is fine in larger ones

Note that if we specify:

$$\beta_{0j} = \gamma_{00} + \varepsilon_{0j}$$

and

$$\beta_{kj} = \gamma_{k0}$$

we get:

$$Y_{ij} = \gamma_{00} + \mathbf{X}_{ij}\gamma_{k0} + \varepsilon_{0j} + u_{ij}$$

which is the RE model above (formally, a "one-level random-intercept" HLM).

In addition:

- HLMs can be expanded to 3- and 4- and higher-level models
- One can include cross-level interactions...

HLMs are widely used in education, psychology, ecology, etc. (less so in economics, political science). Also, there are many, many excellent books, websites, etc. that address HLMs

Random Effects Remix (using `lmer`)

```
> library(lme4)
> AltRE<-lmer(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+              log(GDPPerCapita)+NaturalResourceRents+ColdWar+(1|ISO3),
+              data=WDI)

> summary(AltRE)
Linear mixed model fit by REML ['lmerMod']
Formula: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
    log(GDPPerCapita) + NaturalResourceRents + ColdWar + (1 |      ISO3)
    Data: WDI

REML criterion at convergence: 53322

Scaled residuals:
    Min     1Q Median     3Q    Max
 -4.321 -0.632 -0.058  0.636  5.367

Random effects:
 Groups   Name        Variance Std.Dev.
 ISO3     (Intercept) 323      17.97
 Residual              59       7.68
Number of obs: 7566, groups: ISO3, 186

Fixed effects:
                      Estimate Std. Error t value
(Intercept)          -11.5141     2.9322   -3.93
PopGrowth              0.0323     0.1191    0.27
UrbanPopulation       0.2056     0.0199   10.31
FertilityRate        -2.1709     0.1643  -13.21
log(GDPPerCapita)     8.5722     0.3022   28.37
NaturalResourceRents  0.0269     0.0181    1.49
ColdWar              -7.6610     0.2906  -26.36

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.051
UrbanPopltn -0.183 -0.014
FertilityRt -0.399 -0.288  0.436
lg(GDPPrCp) -0.764 -0.059 -0.305  0.093
NtrlRsrcRnt -0.010 -0.113 -0.012 -0.053 -0.015
ColdWar     -0.109  0.013  0.219 -0.422  0.101  0.055
```

Q: Are They The Same? [A: More Or Less]

Table: RE and HLM Models of WBLI

	RE	AltRE
Population Growth	−0.073	0.032
	(0.121)	(0.119)
Urban Population	0.128***	0.206***
	(0.019)	(0.020)
Fertility Rate	−2.370***	−2.170***
	(0.165)	(0.164)
ln(GDP Per Capita)	7.510***	8.570***
	(0.293)	(0.302)
Natural Resource Rents	0.006	0.027
	(0.018)	(0.018)
Cold War	−8.480***	−7.660***
	(0.289)	(0.291)
Constant	2.560	−11.500***
	(2.580)	(2.930)
Observations	7,566	7,566
R^2	0.520	
Adjusted R^2	0.520	
Log Likelihood		-26,661.000
Akaike Inf. Crit.		53,340.000
Bayesian Inf. Crit.		53,402.000
F Statistic	7,910.000***	

*p<0.1; **p<0.05; ***p<0.01

For more, see here.

HLM with Country-Level Random βs for `ColdWar`

```
> HLM1<-lmer(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+               log(GDPPerCapita)+NaturalResourceRents+ColdWar+(ColdWar|ISO3),
+               data=WDI,control=lmerControl(optimizer="bobyqa"))

> summary(HLM1)
Linear mixed model fit by REML ['lmerMod']
Formula: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
    log(GDPPerCapita) + NaturalResourceRents + ColdWar + (ColdWar |      ISO3)
   Data: WDI
Control: lmerControl(optimizer = "bobyqa")

REML criterion at convergence: 50165

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ISO3     (Intercept) 560.7    23.68
          ColdWar     131.6    11.47    -0.20
 Residual             35.4      5.95
Number of obs: 7566, groups: ISO3, 186

Fixed effects:
                        Estimate Std. Error t value
(Intercept)          -29.248209   3.374962   -8.67
PopGrowth             -0.137795   0.099032   -1.39
UrbanPopulation        0.285673   0.023264   12.28
FertilityRate         -4.130147   0.177531  -23.26
log(GDPPerCapita)     10.946306   0.339078   32.28
NaturalResourceRents  -0.000932   0.015744   -0.06
ColdWar               -2.647474   0.995578   -2.66

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.052
UrbanPopltn -0.132 -0.037
FertilityRt -0.479 -0.197  0.480
lg(GDPPrCp) -0.722 -0.056 -0.384  0.168
NtrlRsrcRnt  0.031 -0.079  0.058 -0.011 -0.095
ColdWar     -0.089 -0.012  0.056 -0.118 -0.003  0.011
```

```
> anova(AltRE,HLM1)

refitting model(s) with ML (instead of REML)
Data: WDI

Models:
AltRE: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
        log(GDPPerCapita) + NaturalResourceRents + ColdWar + (1 | ISO3)
HLM1: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
        log(GDPPerCapita) + NaturalResourceRents + ColdWar + (ColdWar | ISO3)

       npar   AIC   BIC logLik deviance Chisq Df Pr(>Chisq)
AltRE     9 53323 53386 -26653    53305
HLM1     11 50174 50250 -25076    50152  3154  2     <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1


> VarCorr(HLM1)

 Groups   Name        Std.Dev. Corr
 ISO3     (Intercept) 23.68
          ColdWar     11.47    -0.20
 Residual              5.95
```

```
> Bs<-data.frame(coef(HLM1)[1])

> head(Bs)
    ISO3..Intercept. ISO3.PopGrowth ISO3.UrbanPopulation ISO3.FertilityRate
AFG           -19.2         -0.138                0.286              -4.13
AGO           -13.5         -0.138                0.286              -4.13
ALB           -11.3         -0.138                0.286              -4.13
ARE          -105.3         -0.138                0.286              -4.13
ARG           -49.2         -0.138                0.286              -4.13
ARM           -24.6         -0.138                0.286              -4.13
    ISO3.log.GDPPerCapita. ISO3.NaturalResourceRents ISO3.ColdWar
AFG                   10.9                 -0.000932       -3.632
AGO                   10.9                 -0.000932      -14.877
ALB                   10.9                 -0.000932       -7.589
ARE                   10.9                 -0.000932       -0.729
ARG                   10.9                 -0.000932      -22.651
ARM                   10.9                 -0.000932       -3.106

> mean(Bs$ISO3..Intercept.)
[1] -29.2

> mean(Bs$ISO3.ColdWar)
[1] -2.65
```
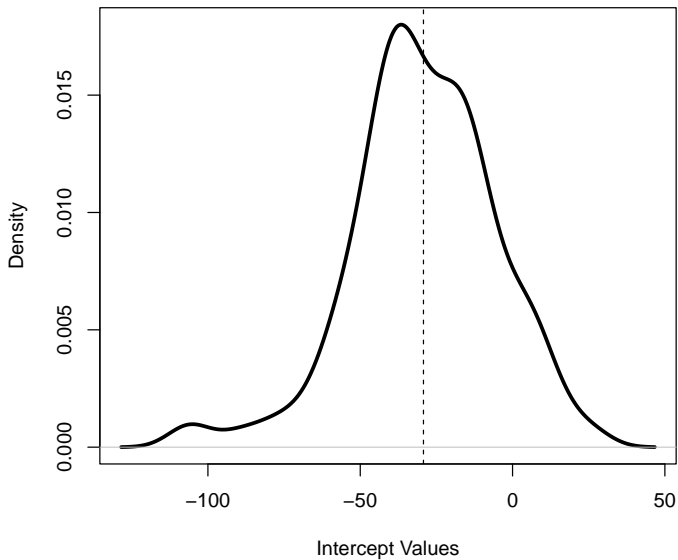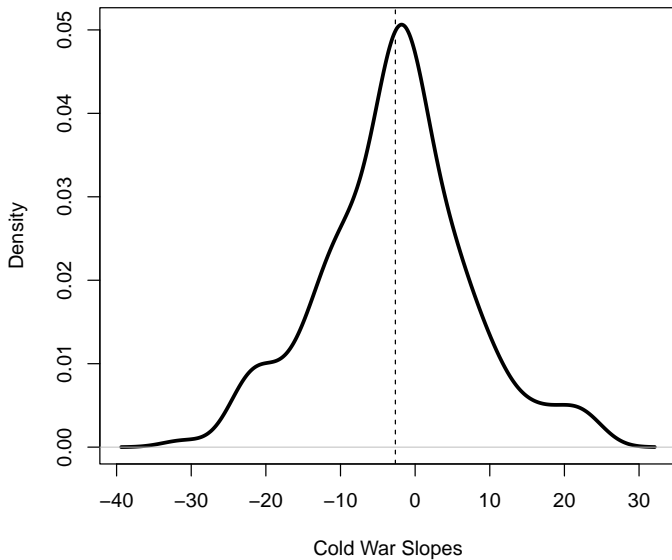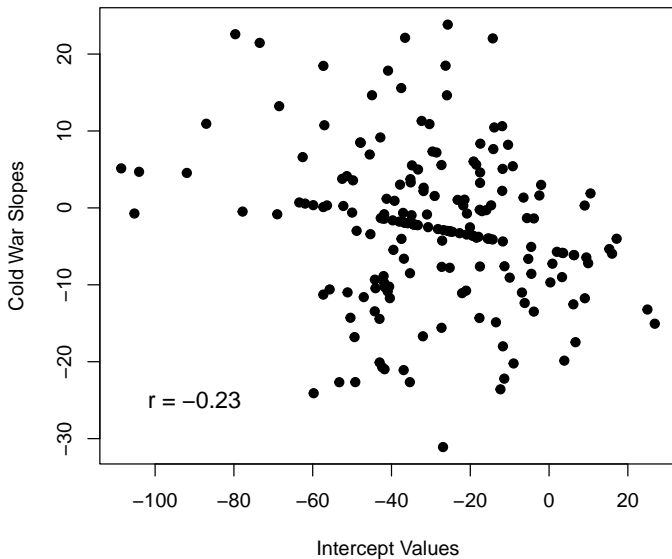
Random Intercepts (Plotted)

Random Slopes for `ColdWar` (Plotted)

Scatterplot: Random Intercepts and Slopes

r = −0.23

Cold War Slopes

Intercept Values

Separating Within and Between Effects

Recall that we can decompose the variance of our regression model as:

$$Y_{it} = \bar{\mathbf{X}}_i \boldsymbol{\beta}_B + (\mathbf{X}_{it} - \bar{\mathbf{X}}_i)\boldsymbol{\beta}_W + u_{it} \qquad (5)$$

This raises the possibility of fitting the model in (5) directly...

- Simple to estimate (via OLS or other)
- Relatively easy interpretation
- Easy to test $\hat{\boldsymbol{\beta}}_B = \hat{\boldsymbol{\beta}}_W$

Example data: Separate effects for within- and between-country *Natural Resource Rents*...

Combining Within- and Between-Effects

Table: BE + WE Model of WBLI

	WEBE.OLS
Population Growth	-3.000^{***}
	(0.141)
Urban Population	-0.046^{***}
	(0.011)
Fertility Rate	-1.470^{***}
	(0.163)
ln(GDP Per Capita)	3.210^{***}
	(0.195)
Within-Country Nat. Resource Rents	0.090^{***}
	(0.030)
Between-Country Nat. Resource Rents	-0.515^{***}
	(0.019)
Cold War	-11.800^{***}
	(0.365)
Constant	53.500^{***}
	(1.700)
Observations	7,566
R^2	0.538
Adjusted R^2	0.537
Residual Std. Error	12.800 (df = 7558)
F Statistic	$1,256.000^{***}$ (df = 7; 7558)

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Our original decomposition considered "two-way" effects:

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

This implies that we can use (e.g.) an F-test to examine the hypothesis:

$$H_0 : \alpha_i = \eta_t = 0 \ \forall \ i, \ t$$

...that is, whether adding the (two-way) effects improves the model's fit.

We can also consider the partial hypotheses:

$$H_0 : \alpha_i = 0 \ \forall \ i$$

and

$$H_0 : \eta_t = 0 \ \forall \ t$$

separately.

Two-Way Effects: Good & Bad

<u>The Good</u>:

- Addresses both time-invariant, unit-level heterogeneity and cross-sectionally-invariant, time-point-specific heterogeneity
- May be "fixed" or "random"...
- Two-way FE is equivalent to differences-in-differences when $X \in \{0,1\}$ and $T = 2$ (more on that later)

<u>The (Potentially) Bad</u>:

- Ability to control for unobserved heterogeneity depends on (probably invalid) functional form assumptions (Imai and Kim 2020)
- FE *requires* predictors that vary both within and between both units and time points
- RE requires the (additional) assumption that $Cov(\mathbf{X}_{it}, \eta_t) = Cov(\alpha_i, \eta_t) = 0$
- Two-way effects models ask a *lot* of your data (effectively fits $N + T + k$ parameters using NT observations)

```
> TwoWayFE<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+                log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+                effect="twoway",model="within")

> summary(TwoWayFE)
Twoways effects Within Model

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, effect = "twoway", model = "within")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-31.756  -3.866   0.251   4.100  29.715

Coefficients:
                      Estimate Std. Error t-value Pr(>|t|)
PopGrowth            -0.26475    0.10023   -2.64   0.0083 **
UrbanPopulation      -0.00333    0.01798   -0.19   0.8529
FertilityRate         1.03517    0.15072    6.87   7e-12 ***
log(GDPPerCapita)     2.50015    0.28544    8.76   <2e-16 ***
NaturalResourceRents -0.00420    0.01588   -0.26   0.7913
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    306000
Residual Sum of Squares: 300000
R-Squared:      0.0179
Adj. R-Squared: -0.0141
F-statistic: 26.7123 on 5 and 7326 DF, p-value: <2e-16
```

```
> # Two-way effects:

> pFtest(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+        log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+        effect="twoway",model="within")

F test for twoways effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
F = 103, df1 = 233, df2 = 7326, p-value <2e-16
alternative hypothesis: significant effects

> plmtest(TwoWayFE,c("twoways"),type=("kw"))

Lagrange Multiplier Test - two-ways effects (King and Wu) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
normal = 187, p-value <2e-16
alternative hypothesis: significant effects


> # One-way effects in the two-way model:
>
> plmtest(TwoWayFE,c("individual"),type=("kw"))

Lagrange Multiplier Test - (King and Wu) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
normal = 211, p-value <2e-16
alternative hypothesis: significant effects

> plmtest(TwoWayFE,c("time"),type=("kw"))

Lagrange Multiplier Test - time effects (King and Wu) for unbalanced panels

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate + ...
normal = 100, p-value <2e-16
alternative hypothesis: significant effects
```

```
> TwoWayFE.BF<-lm(WomenBusLawIndex~PopGrowth+UrbanPopulation+
+                 FertilityRate+log(GDPPerCapita)+NaturalResourceRents+
+                 factor(ISO3)+factor(Year),data=WDI)

> summary(TwoWayFE.BF)

Call:
lm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    factor(ISO3) + factor(Year), data = WDI)

Residuals:
   Min    1Q Median     3Q    Max
-31.76 -3.87   0.25   4.10  29.72

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -17.26971    2.50039   -6.91 5.4e-12 ***
PopGrowth              -0.26475    0.10023   -2.64  0.0083 **
UrbanPopulation        -0.00333    0.01798   -0.19  0.8529
FertilityRate           1.03517    0.15072    6.87 7.0e-12 ***
log(GDPPerCapita)       2.50015    0.28544    8.76 < 2e-16 ***
NaturalResourceRents   -0.00420    0.01588   -0.26  0.7913
factor(ISO3)AGO        29.30071    2.01487   14.54 < 2e-16 ***
factor(ISO3)ALB        51.72646    2.02286   25.57 < 2e-16 ***
factor(ISO3)ARE        -5.95970    2.54027   -2.35  0.0190 *
.
.
.
factor(Year)1978        4.82791    0.88259    5.47 4.6e-08 ***
factor(Year)1979        5.23708    0.88928    5.89 4.1e-09 ***
 [ reached getOption("max.print") -- omitted 40 rows ]
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 6.4 on 7326 degrees of freedom
  (5764 observations deleted due to missingness)
Multiple R-squared:  0.888,	Adjusted R-squared:  0.884
F-statistic: 242 on 239 and 7326 DF,  p-value: <2e-16
```

Example: Two-Way Random Effects

```
> TwoWayRE<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+
+                log(GDPPerCapita)+NaturalResourceRents+ColdWar,data=WDI,
+                effect="twoway",model="random")

> summary(TwoWayRE)

Twoways effects Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    ColdWar, data = WDI, effect = "twoway", model = "random")

Unbalanced Panel: n = 186, T = 1-50, N = 7566

Effects:
                  var std.dev share
idiosyncratic  40.993   6.403  0.31
individual     89.185   9.444  0.68
time            0.612   0.782  0.00
theta:
        Min. 1st Qu. Median  Mean 3rd Qu.  Max.
id     0.439   0.893  0.905 0.896   0.905 0.905
time   0.372   0.423  0.469 0.449   0.477 0.479
total  0.330   0.422  0.468 0.448   0.476 0.478

Residuals:
   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
  -61.3    -6.4    2.4    0.4    10.6   32.2

Coefficients:
                      Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)           21.29333    0.35391    60.17   < 2e-16 ***
PopGrowth             -0.23814    0.01549   -15.37   < 2e-16 ***
UrbanPopulation        0.06588    0.00254    25.97   < 2e-16 ***
FertilityRate         -0.58746    0.02216   -26.51   < 2e-16 ***
log(GDPPerCapita)      5.02484    0.03939   127.55   < 2e-16 ***
NaturalResourceRents  -0.01087    0.00239    -4.54 0.0000056 ***
ColdWar              -12.66453    0.05391  -234.91   < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Total Sum of Squares:    2670000
Residual Sum of Squares: 1740000
R-Squared:      0.351
Adj. R-Squared: 0.351
Chisq: 161011 on 6 DF, p-value: <2e-16
```

A Prettier Table

Table: Models of WBLI

	OLS	FE	BE	RE	TwoWayFE	TwoWayRE
Population Growth	-3.190^{***}	0.073	-5.870^{***}	-0.073	-0.265^{***}	-0.238^{***}
	(0.144)	(0.119)	(0.917)	(0.121)	(0.100)	(0.015)
Urban Population	-0.058^{***}	0.248^{***}	-0.050	0.128^{***}	-0.003	0.066^{***}
	(0.011)	(0.021)	(0.055)	(0.019)	(0.018)	(0.003)
Fertility Rate	-1.790^{***}	-2.070^{***}	0.063	-2.370^{***}	1.030^{***}	-0.587^{***}
	(0.165)	(0.166)	(1.080)	(0.165)	(0.151)	(0.022)
ln(GDP Per Capita)	3.150^{***}	9.160^{***}	3.400^{***}	7.510^{***}	2.500^{***}	5.030^{***}
	(0.199)	(0.310)	(1.130)	(0.293)	(0.285)	(0.039)
Natural Resource Rents	-0.349^{***}	0.035^{*}	-0.358^{***}	0.006	-0.004	-0.011^{***}
	(0.016)	(0.018)	(0.089)	(0.018)	(0.016)	(0.002)
Cold War	-11.300^{***}	-7.190^{***}	-12.500^{**}	-8.480^{***}		-12.700^{***}
	(0.372)	(0.295)	(4.890)	(0.289)		(0.054)
Constant	54.800^{***}		50.300^{***}	2.560		21.300^{***}
	(1.730)		(10.200)	(2.580)		(0.354)
Observations	7,566	7,566	186	7,566	7,566	7,566
R^2	0.519	0.535	0.605	0.520	0.018	0.351
Adjusted R^2	0.519	0.523	0.592	0.520	-0.014	0.351

*p<0.1; **p<0.05; ***p<0.01

Other Variations: FEIS

"Fixed Effects Individual Slope" models

- Cite: Brueder, Josef, and Volker Ludwig. 2015. "Fixed-Effects Panel Regression." In *The Sage Handbook of Regression Analysis and Causal Inference*, Eds. Henning Best and Christof Wolf. Los Angeles: Sage, pp. 327-357.

- FE + unit-level slopes for (some / all) predictor variables

- Equivalent to including $N-1$ interactions between a predictor **X** and each of the α_is

- Also can test for homogeneity of estimated slopes (Hausman-like test)

- See the feisr R package, and its accompanying vignette, or xtfeis in Stata

Unit Effects Models: Software

R :

- the `plm` package; `plm` command
  - · Fits one- and two-way FE, BE, RE models
  - · Also fits first difference (FD) and instrumental variable (IV) models
- the `fixest` package; fast/scalable FE estimation for OLS and GLMs
- the `lme4` package; command is `lmer`
- the `nlme` package; command `lme`
- the `Paneldata` package

Stata : `xtreg`

- option `re` (the default) = random effects
- option `fe` = fixed (within) effects
- option `be` = between-effects
- Stata package `fect` = two-way models