GSERM - St. Gallen 2025
Analyzing Panel Data

June 17, 2025

Two-Way Variation

Two-way variation:

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

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

Note that we can write:

$$\alpha_i = \sum_{t=1}^{T_i}(\gamma V_i)$$

and

$$\eta_t = \sum_{i=1}^{N_t}(\delta W_t).$$

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-Way” and “Two-Way” Effects

“Two-way” effects:

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

“One-way” effects:

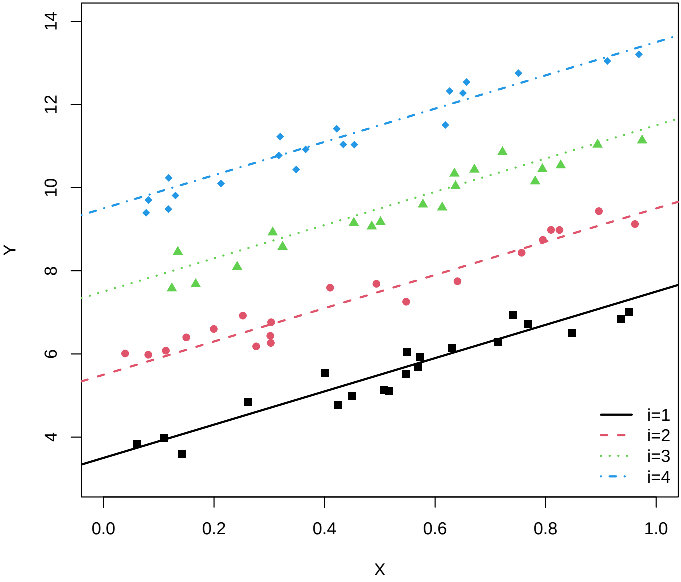
- Assuming $\alpha_i = 0$:

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

- Assuming $\eta_t = 0$:

$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it} \quad \text{(“unit effects”)}$$

Intuition: One-Way Unit Effects



(One-Way) "Fixed" Effects

"Brute force" model fits:

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

<u>In other words</u>:

- Specify $\mathbf{X}_{it}$,
- Fit a model that includes both $\mathbf{X}_{it}\boldsymbol{\beta}$ and N (or $N-1$) indicator variables (with parameters α_i), one for each unit i
  - If the model includes an intercept (β_0), then one unit is omitted and becomes the "reference" category
  - If the model omits an intercept, then all N $\hat{\alpha}$s can be estimated; $\hat{\alpha}_i$ is then the expected value of Y for unit i when all values of $\mathbf{X}_{it}$ are zero

“Fixed” Effects (continued)

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:

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

<u>But!</u>

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

“Fixed” Effects = “Within” Effects

This means that:

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

gives:

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

$\rightarrow$ **A “fixed effects” model is actually a “within-effects” model.**

- “Fixed effects” models <u>only</u> use the *within-unit* variation in Y and $\mathbf{X}$
- $\hat{\beta}$s for variables that do not vary within units <u>cannot be estimated</u>

The one-way FE model implies that $\alpha_i \neq \alpha_j$ for at least some $i \neq j$. We can test this assumption using a standard F-test of the hypothesis:

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

versus

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

Specifically, the test statistic

$$F = \frac{(\text{SSE}_{OLS} - \text{SSE}_{FE}) / N - 1}{\text{SSE}_{FE} / [NT - (N-1)]}$$

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

Example Redux: WDI, 1960-2024

The World Development Indicators

- Cross-national country-level time series data
- $N = 215$ countries, $T = 65$ years (1960-2025) + 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.

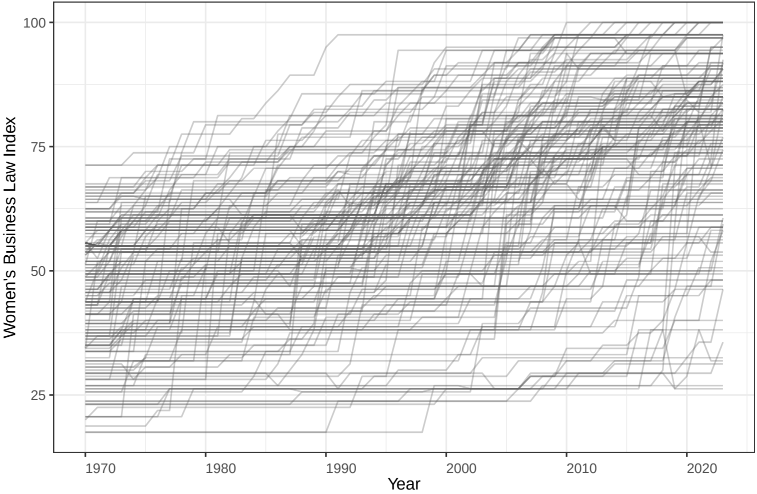
```
> describe(wdi,fast=TRUE,ranges=FALSE,check=TRUE)
                        vars     n            mean      sd  skew kurtosis          se
ISO3                       1 13975             NaN      NA    NA       NA          NA
Year                       2 13975         1992.00 1.9e+01  0.00    -1.20        0.16
Region                     3 13975             NaN      NA    NA       NA          NA
country                    4 13975             NaN      NA    NA       NA          NA
iso3c                      5 13975             NaN      NA    NA       NA          NA
LandArea                   6 12068       596744.44 1.6e+06  5.45    34.69    14960.16
ArablePercent              7 11662           13.40 1.4e+01  1.47     1.97        0.13
Population                 8 13730     25346115.55 1.1e+08  9.73   105.72   899764.29
PopGrowth                  9 13513            1.75 1.8e+00  0.73    21.38        0.02
RuralPopulation           10 13696           48.11 2.6e+01 -0.11    -1.00        0.22
UrbanPopulation           11 13696           51.89 2.6e+01  0.11    -1.00        0.22
BirthRatePer1K            12 13730           27.67 1.3e+01  0.24    -1.24        0.11
FertilityRate             13 13728            3.82 2.0e+00  0.43    -1.18        0.02
PrimarySchoolAge          14 11120            6.13 6.1e-01 -0.04     0.11        0.01
LifeExpectancy            15 13726           65.17 1.1e+01 -0.76     0.18        0.10
AgeDepRatioOld            16 13730           10.76 7.2e+00  1.79     5.00        0.06
CO2Emissions              17 10962            5.05 1.1e+01  8.70   102.13        0.11
GDP                       18 11040 240200887836.82 1.1e+12 11.48   160.11 10755237244.82
GDPPerCapita              19 11040        12313.79 1.9e+04  3.11    14.14      183.36
GDPPerCapGrowth           20 10970            1.89 6.5e+00  1.68    41.63        0.06
Inflation                 21  8882           22.90 3.2e+02 54.09  3625.88        3.40
TotalTrade                22  8777           77.93 5.3e+01  3.03    18.39        0.57
Exports                   23  8777           36.19 2.9e+01  2.99    16.80        0.30
Imports                   24  8786           41.75 2.7e+01  2.54    13.97        0.29
FDIIn                     25  8969            5.53 4.5e+01 15.13   541.91        0.48
AgriEmployment            26  6134           28.11 2.4e+01  0.66    -0.67        0.30
NetAidReceived            27  9043    506951242.00 1.0e+09  8.32   157.34 10484966.48
MobileCellSubscriptions   28 10212           36.32 5.2e+01  1.29     1.14        0.51
NaturalResourceRents      29  9211            6.85 1.1e+01  2.60     8.04        0.12
MilitaryExpenditures      30  7733            2.71 3.2e+00  9.43   239.85        0.04
GovtExpenditures          31  8421           16.34 8.3e+00  3.63    32.81        0.09
PublicEdExpend            32  5177            4.35 2.1e+00  6.89   166.21        0.03
PublicHealthExpend        33  4346            3.29 2.4e+00  1.33     3.08        0.04
HIVDeaths                 34  4656         6473.06 1.9e+04  5.78    45.97      277.31
WomenBusLawIndex          35 10152           59.85 1.9e+01  0.02    -0.58        0.19
PaidParentalLeave         36 10152            0.11 3.1e-01  2.50     4.27        0.00
PostColdWar               37 13975            0.54 5.0e-01 -0.15    -1.98        0.00
```

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

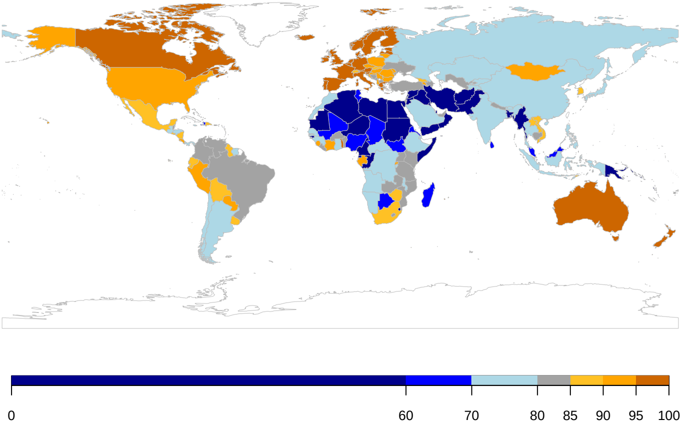
The basis for a 2021 World Bank report...

- Examines "the laws and regulations that affect women's economic opportunity in 190 economies" from 1970-2023.
- 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 ≈ 18
  - 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."

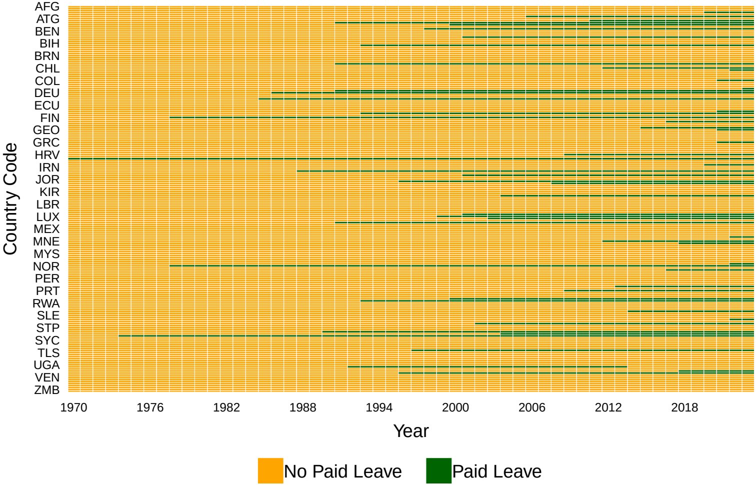
Trend Line Visualization (using `panelView`)



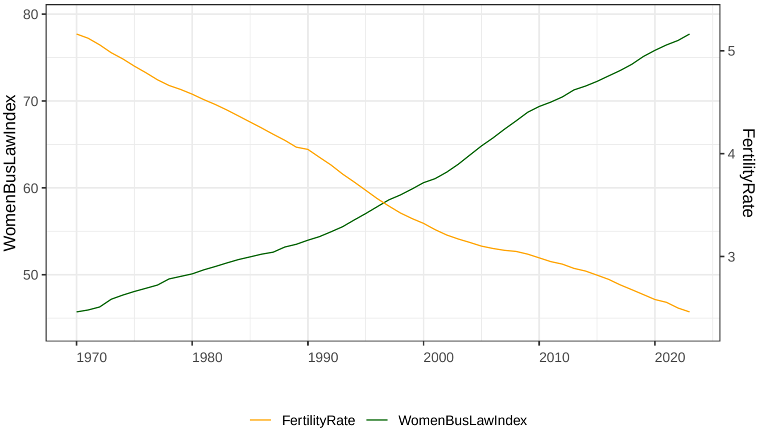
Map! WBLI (2023)



Categorical Variable Visualization (using `panelView`)



Bivariate Visualization (using `panelView`)



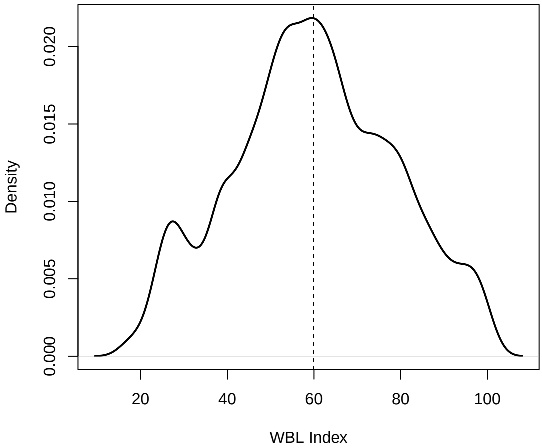
plm and Panel Data

plm (panel linear models) is the workhorse R package for fitting linear regression models to panel data. To do so, you must first declare the data as panel data and indicate the variables containing the unit (i) and period (t) indices:

```
> WDI<-pdata.frame(wdi,index=c("ISO3","Year"))
> class(WDI)
[1] "pdata.frame" "data.frame"

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

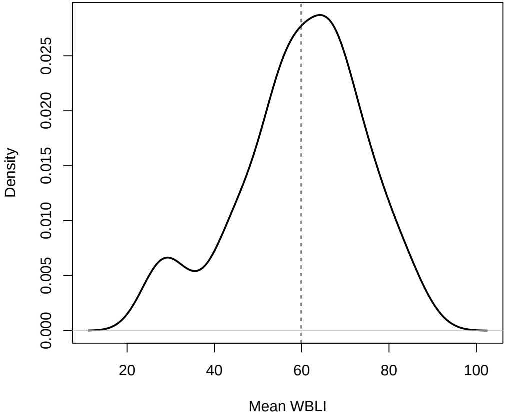
WBLI: Total Variation

```
> describe(WBLI,na.rm=TRUE) # all variation
   vars     n  mean    sd median trimmed   mad  min max range skew kurtosis   se
X1    1 10152 59.85 18.74  59.38   59.86 19.46 17.5 100  82.5 0.02    -0.58 0.19
```



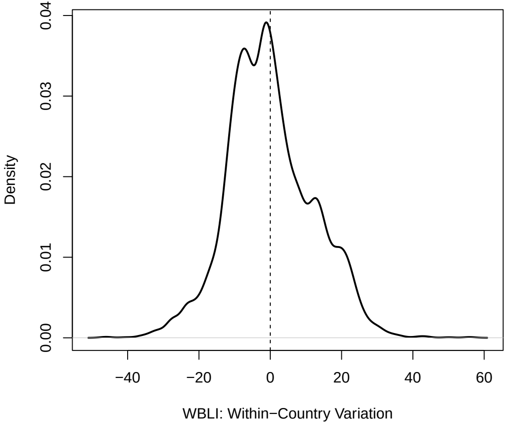
WBLI: "Between" Variation

```
> 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 59.85 14.48  61.04   60.73 13.08 23.67 89.94 66.27 -0.5    -0.12 1.06
```



WBLI: "Within" Variation

```
> describe(Within(WBLI,na.rm=TRUE)) # "within" variation
   vars     n mean    sd median trimmed   mad    min   max range skew kurtosis   se
X1    1 10152    0 11.94  -1.25   -0.32 11.14 -46.06 55.88 101.9 0.26     0.35 0.12
```



A Regression Model

Regression model:

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

$$\text{Expectations:}\quad \begin{array}{lll} \beta_1 < 0 & \beta_3 < 0 & \beta_5 < 0 \\ \beta_2 > 0 & \beta_4 > 0 & \beta_6 > 0 \end{array}$$

Descriptive Statistics:

	vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
WomenBusLawIndex	1	8318	60.44	19.03	60.62	17.50	100.00	82.50	-0.03	-0.69	0.21
PopGrowth	2	8318	1.67	1.67	1.64	-27.47	21.70	49.17	-0.06	30.06	0.02
UrbanPopulation	3	8318	52.09	24.06	51.93	2.85	100.00	97.16	0.06	-1.04	0.26
FertilityRate	4	8318	3.61	1.90	3.09	0.77	8.61	7.83	0.51	-1.04	0.02
NaturalResourceRents	5	8318	7.19	11.14	2.42	0.00	88.59	88.59	2.53	7.58	0.12
PostColdWar	6	8318	0.69	0.46	1.00	0.00	1.00	1.00	-0.83	-1.31	0.01
lnGDPPerCap	7	8318	8.33	1.46	8.24	4.93	11.68	6.75	0.12	-0.90	0.02

Where is the Variation in our Data?

Variable	Dimension	Mean	SD	Min	Max	Observations
Women Bus. Law Index	overall	60.435	19.028	17.5	100	$NT = 8318$
	between		15.363	23.329	89.555	$N = 187$
	within		11.346	15.147	117.722	$T = 44.481$
Pop. Growth	overall	1.673	1.673	-27.471	21.7	$NT = 8318$
	between		1.271	-1.077	6.328	$N = 187$
	within		1.145	-29.218	17.301	$T = 44.481$
Urban Population	overall	52.087	24.057	2.845	100	$NT = 8318$
	between		22.914	7.649	100	$N = 187$
	within		6.853	13.796	80.713	$T = 44.481$
Fertility Rate	overall	3.614	1.9	0.772	8.606	$NT = 8318$
	between		1.668	1.292	7.586	$N = 187$
	within		0.934	0.846	7.384	$T = 44.481$
Natural Resource Rents	overall	7.194	11.142	0	88.592	$NT = 8318$
	between		10.02	0	43.459	$N = 187$
	within		5.285	-24.03	61.301	$T = 44.481$
Post Cold War	overall	0.691	0.462	0	1	$NT = 8318$
	between		0.159	0.333	1	$N = 187$
	within		0.441	-0.278	1.358	$T = 44.481$
ln(GDP Per Cap)	overall	8.332	1.457	4.926	11.678	$NT = 8318$
	between		1.393	5.728	11.199	$N = 187$
	within		0.354	6.52	10.277	$T = 44.481$

Regression: Pooled OLS

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

> summary(OLS)
Pooling Model

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -78.61   -8.48    1.05    9.24   45.58

Coefficients:
                     Estimate Std. Error t-value      Pr(>|t|)
(Intercept)           52.5590     1.7972   29.25       < 2e-16 ***
PopGrowth             -1.7755     0.1165  -15.24       < 2e-16 ***
UrbanPopulation       -0.0659     0.0104   -6.33 0.00000000025 ***
FertilityRate         -2.9836     0.1540  -19.38       < 2e-16 ***
log(GDPPerCapita)      2.4691     0.1910   12.93       < 2e-16 ***
NaturalResourceRents  -0.3633     0.0150  -24.17       < 2e-16 ***
PostColdWar           10.2751     0.3652   28.14       < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    3010000
Residual Sum of Squares: 1500000
R-Squared:      0.5
Adj. R-Squared: 0.5
F-statistic: 1387.87 on 6 and 8311 DF, p-value: <2e-16
```

“Fixed” (Within) Effects

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

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

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-33.806  -5.109  -0.504   4.980  52.899

Coefficients:
                     Estimate Std. Error t-value Pr(>|t|)
PopGrowth             -0.1890     0.0807   -2.34    0.019 *
UrbanPopulation        0.3101     0.0196   15.83  < 2e-16 ***
FertilityRate         -1.7632     0.1542  -11.43  < 2e-16 ***
log(GDPPerCapita)      8.7789     0.2935   29.91  < 2e-16 ***
NaturalResourceRents   0.0718     0.0166    4.32 0.000016 ***
PostColdWar            6.9444     0.2882   24.09  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1070000
Residual Sum of Squares: 505000
R-Squared:      0.528
Adj. R-Squared: 0.517
F-statistic: 1514.57 on 6 and 8125 DF, p-value: <2e-16
```

A Nicer Table

	OLS	FE
Population Growth	-1.776^{***}	-0.189^{**}
	(0.117)	(0.081)
Urban Population	-0.066^{***}	0.310^{***}
	(0.010)	(0.020)
Fertility Rate	-2.984^{***}	-1.763^{***}
	(0.154)	(0.154)
ln(GDP Per Capita)	2.469^{***}	8.779^{***}
	(0.191)	(0.294)
Natural Resource Rents	-0.363^{***}	0.072^{***}
	(0.015)	(0.017)
Post-Cold War	10.280^{***}	6.944^{***}
	(0.365)	(0.288)
Constant	52.560^{***}	
	(1.797)	
Observations	8,318	8,318
R^2	0.500	0.528
Adjusted R^2	0.500	0.517
F Statistic	$1{,}388.000^{***}$ (df = 6; 8311)	$1{,}515.000^{***}$ (df = 6; 8125)

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

One-Way Period Fixed Effects

The model is:

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

which is estimated via "demeaning" data at each time point:

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

then fitting the OLS model:

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

Comparison: Unit vs. Time Fixed Effects

	OLS	FE.Units	FE.Time
Population Growth	-1.776^{***}	-0.189^{**}	-1.939^{***}
	(0.117)	(0.081)	(0.110)
Urban Population	-0.066^{***}	0.310^{***}	-0.067^{***}
	(0.010)	(0.020)	(0.010)
Fertility Rate	-2.984^{***}	-1.763^{***}	-2.064^{***}
	(0.154)	(0.154)	(0.149)
ln(GDP Per Capita)	2.469^{***}	8.779^{***}	2.871^{***}
	(0.191)	(0.294)	(0.181)
Natural Resource Rents	-0.363^{***}	0.072^{***}	-0.398^{***}
	(0.015)	(0.017)	(0.014)
Post-Cold War	10.280^{***}	6.944^{***}	
	(0.365)	(0.288)	
Constant	52.560^{***}		
	(1.797)		
Observations	8,318	8,318	8,318
R^2	0.500	0.528	0.402
Adjusted R^2	0.500	0.517	0.398
F Statistic	$1{,}388.000^{***}$ (df = 6; 8311)	$1{,}515.000^{***}$ (df = 6; 8125)	$1{,}112.000^{***}$ (df = 5; 8261)

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

Fixed Effects: Testing

As we noted previously, 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:

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

(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

FE (by Unit) Model Tests

```
> pFtest(FE,OLS)

 F test for individual effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
F = 86, df1 = 186, df2 = 8125, p-value <2e-16
alternative hypothesis: significant effects

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

 Lagrange Multiplier Test - (Breusch-Pagan)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
chisq = 56650, df = 1, p-value <2e-16
alternative hypothesis: significant effects

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

 Lagrange Multiplier Test - (King and Wu)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 238, p-value <2e-16
alternative hypothesis: significant effects
```

FE (by Period) Model Tests

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

 F test for time effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
F = 22, df1 = 50, df2 = 8261, p-value <2e-16
alternative hypothesis: significant effects

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

 Lagrange Multiplier Test - time effects (Breusch-Pagan)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
chisq = 8402, 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)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 92, 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 $\mathbf{X}$ 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...

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 counter-factuals when discussing the substantive impact of a treatment" (2018, 829).*

Significance:

- Predictors $\mathbf{X}$ in FE models typically have both cross-sectional and temporal variation
- FE models only consider *within-unit* variation in $\mathbf{X}$ 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

Q: How much is a "one standard deviation change in Urban Population"?

Recall its variation:

Variable	Dimension	Mean	SD	Min	Max	Observations
Urban Population	overall	52.087	24.057	2.845	100	$NT = 8318$
	between		22.914	7.649	100	$N = 187$
	within		6.853	13.796	80.713	$T = 44.481$

"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, p. 830).

Whither "Fixed" Effects?

Khoa Vu
@KhoaVuUmn

Real footage of an applied economist adding fixed effects in his regressions

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 $\boldsymbol{\beta}_B$
- Slowly-Changing **X**s
- (In)Efficiency / Inconsistency (Incidental Parameters)
- Cannot Control for (e.g.) Time-Varying Heterogeneity
- Sensitivity to Measurement Error

"Between" Effects

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.*

“Between” Effects

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

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

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318
Observations used in estimation: 187

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-29.640  -6.886   0.809   8.117  21.124

Coefficients:
                     Estimate Std. Error t-value Pr(>|t|)
(Intercept)           57.6089    12.8651    4.48 0.000013 ***
PopGrowth             -4.9920     1.1283   -4.42 0.000017 ***
UrbanPopulation       -0.0508     0.0550   -0.92  0.35657
FertilityRate         -1.5067     1.1726   -1.28  0.20048
log(GDPPerCapita)      1.9512     1.1458    1.70  0.09032 .
NaturalResourceRents  -0.3131     0.0920   -3.41  0.00082 ***
PostColdWar            7.7174     5.3256    1.45  0.14904
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    43900
Residual Sum of Squares: 18400
R-Squared:      0.58
Adj. R-Squared: 0.566
F-statistic: 41.4088 on 6 and 180 DF, p-value: <2e-16
```

A Nicer Table (Again)

	OLS	FE	BE
Population Growth	-1.776^{***}	-0.189^{**}	-4.992^{***}
	(0.117)	(0.081)	(1.128)
Urban Population	-0.066^{***}	0.310^{***}	-0.051
	(0.010)	(0.020)	(0.055)
Fertility Rate	-2.984^{***}	-1.763^{***}	-1.507
	(0.154)	(0.154)	(1.173)
ln(GDP Per Capita)	2.469^{***}	8.779^{***}	1.951^{*}
	(0.191)	(0.294)	(1.146)
Natural Resource Rents	-0.363^{***}	0.072^{***}	-0.313^{***}
	(0.015)	(0.017)	(0.092)
Post-Cold War	10.280^{***}	6.944^{***}	7.717
	(0.365)	(0.288)	(5.326)
Constant	52.560^{***}		57.610^{***}
	(1.797)		(12.870)
Observations	8,318	8,318	187
R^2	0.500	0.528	0.580
Adjusted R^2	0.500	0.517	0.566
F Statistic	$1,388.000^{***}$ (df = 6; 8311)	$1,515.000^{***}$ (df = 6; 8125)	41.410^{***} (df = 6; 180)

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

Model:

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

with:

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

and

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

“Random” Effects

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

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

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:

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

a.k.a. the “one-way error components model.”

"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 <u>within-i</u> 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:

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

“Random” Effects: Estimation

Under our assumptions, we can show that:

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

where

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

is an unknown quantity to be estimated.

Estimation: 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 (a la FGLS).

"Random" Effects: An Alternative View

Random Effects

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318

Effects:
                var std.dev share
idiosyncratic 62.20    7.89  0.39
individual    96.27    9.81  0.61
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.727   0.877   0.889   0.881   0.889   0.889

Residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-34.0055  -5.3346  -0.5703  -0.0314   5.5855  44.0815

Coefficients:
                     Estimate Std. Error z-value Pr(>|z|)
(Intercept)           -7.3135     2.4992   -2.93   0.0034 **
PopGrowth             -0.2321     0.0824   -2.82   0.0048 **
UrbanPopulation        0.1928     0.0184   10.46   <2e-16 ***
FertilityRate         -2.1064     0.1539  -13.68   <2e-16 ***
log(GDPPerCapita)      7.2254     0.2785   25.95   <2e-16 ***
NaturalResourceRents   0.0450     0.0167    2.70   0.0069 **
PostColdWar            8.1810     0.2830   28.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1110000
Residual Sum of Squares: 543000
R-Squared:      0.511
Adj. R-Squared: 0.511
Chisq: 8493.57 on 6 DF, p-value: <2e-16
```

A Nicer Table (Yet Again)

	OLS	FE	BE	RE
Population Growth	-1.776^{***}	-0.189^{**}	-4.992^{***}	-0.232^{***}
	(0.117)	(0.081)	(1.128)	(0.082)
Urban Population	-0.066^{***}	0.310^{***}	-0.051	0.193^{***}
	(0.010)	(0.020)	(0.055)	(0.018)
Fertility Rate	-2.984^{***}	-1.763^{***}	-1.507	-2.106^{***}
	(0.154)	(0.154)	(1.173)	(0.154)
ln(GDP Per Capita)	2.469^{***}	8.779^{***}	1.951^{*}	7.225^{***}
	(0.191)	(0.294)	(1.146)	(0.278)
Natural Resource Rents	-0.363^{***}	0.072^{***}	-0.313^{***}	0.045^{***}
	(0.015)	(0.017)	(0.092)	(0.017)
Post-Cold War	10.280^{***}	6.944^{***}	7.717	8.181^{***}
	(0.365)	(0.288)	(5.326)	(0.283)
Constant	52.560^{***}		57.610^{***}	-7.314^{***}
	(1.797)		(12.870)	(2.499)
Observations	8,318	8,318	187	8,318
R^2	0.500	0.528	0.580	0.511
Adjusted R^2	0.500	0.517	0.566	0.511
F Statistic	$1{,}388.000^{***}$ (df = 6; 8311)	$1{,}515.000^{***}$ (df = 6; 8125)	41.410^{***} (df = 6; 180)	$8{,}494.000^{***}$

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

"Random" Effects: Testing

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):

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

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

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

Issues / Concerns:

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

Hausman Test Results

Hausman test (FE vs. RE):

```
> phtest(FE, RE)  # ugh...

 Hausman Test

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
chisq = 5341, df = 6, p-value <2e-16

alternative hypothesis: one model is inconsistent
```

Things to think about:

- Are we in asymptopia?
- Do we *believe* our model specification?...

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

HLM Starting Points

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

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

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.

Model Assumptions

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

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`)

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

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

REML criterion at convergence: 59006

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.289 -0.649 -0.074  0.636  6.599

Random effects:
 Groups   Name        Variance Std.Dev.
 ISO3     (Intercept) 341.1    18.47
 Residual              62.3     7.89
Number of obs: 8318, groups:  ISO3, 187

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          -19.9879     2.8298   -7.06
PopGrowth             -0.2012     0.0806   -2.49
UrbanPopulation        0.2688     0.0191   14.09
FertilityRate         -1.8775     0.1531  -12.27
log(GDPPerCapita)      8.2262     0.2865   28.71
NaturalResourceRents   0.0639     0.0165    3.87
PostColdWar            7.3931     0.2842   26.01

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.051
UrbanPopltn -0.179  0.007
FertilityRt -0.441 -0.238  0.434
lg(GDPPrCp) -0.748 -0.057 -0.283  0.096
NtrlRsrcRnt  0.005 -0.096 -0.023 -0.110 -0.012
PostColdWar  0.027 -0.025 -0.218  0.427 -0.121 -0.055
```

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

	RE	AltRE
Population Growth	-0.232^{***}	-0.201^{**}
	(0.082)	(0.081)
Urban Population	0.193^{***}	0.269^{***}
	(0.018)	(0.019)
Fertility Rate	-2.106^{***}	-1.878^{***}
	(0.154)	(0.153)
ln(GDP Per Capita)	7.225^{***}	8.226^{***}
	(0.278)	(0.286)
Natural Resource Rents	0.045^{***}	0.064^{***}
	(0.017)	(0.017)
Post-Cold War	8.181^{***}	7.393^{***}
	(0.283)	(0.284)
Constant	-7.314^{***}	-19.990^{***}
	(2.499)	(2.830)
Observations	8,318	8,318
R^2	0.511	
Adjusted R^2	0.511	
Log Likelihood		-29,503.000
Akaike Inf. Crit.		59,023.000
Bayesian Inf. Crit.		59,087.000
F Statistic	$8{,}494.000^{***}$	

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

For more, see here.

Why HLMs?

Some reasons:

- Flexibility / verisimilitude
- Maximizing information (via pooling/shrinkage)
- Modeling / conceptual consistency

See especially Gelman and Hill (2007, Chapter 11).

Example: What if we think that the end of the Cold War had (slightly) different effects on WBLI in each country?...

HLM with Country-Level Random βs for `ColdWar`

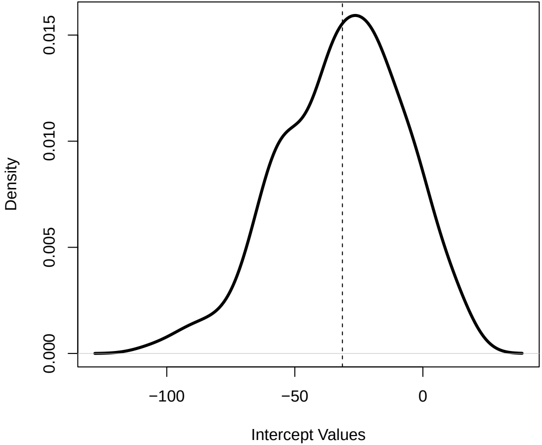
```
> HLM1<-lmer(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+
PostColdWar+(PostColdWar|ISO3),data=WDI,control=lmerControl(optimizer="bobyqa"))

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

REML criterion at convergence: 55471

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.346 -0.537  0.008  0.534  7.987

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ISO3     (Intercept) 605.2    24.60
          PostColdWar 141.2    11.88    -0.29
 Residual              37.3     6.11
Number of obs: 8318, groups:  ISO3, 187

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          -31.4227     3.3116   -9.49
PopGrowth             -0.3185     0.0647   -4.93
UrbanPopulation        0.3400     0.0222   15.33
FertilityRate         -3.8570     0.1652  -23.35
log(GDPPerCapita)     10.5059     0.3173   33.11
NaturalResourceRents   0.0114     0.0143    0.80
PostColdWar            2.5196     1.0030    2.51

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.049
UrbanPopltn -0.133 -0.012
FertilityRt -0.501 -0.168  0.488
lg(GDPPrCp) -0.693 -0.057 -0.368  0.170
NtrlRsrcRnt  0.021 -0.078  0.060 -0.043 -0.077
PostColdWar -0.210  0.004 -0.051  0.112 -0.001 -0.009
```

```
> anova(AltRE,HLM1)

refitting model(s) with ML (instead of REML)

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

      npar   AIC   BIC logLik -2*log(L) Chisq Df Pr(>Chisq)
AltRE    9 59006 59069 -29494     58988
HLM1    11 55479 55556 -27728     55457  3531  2     <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

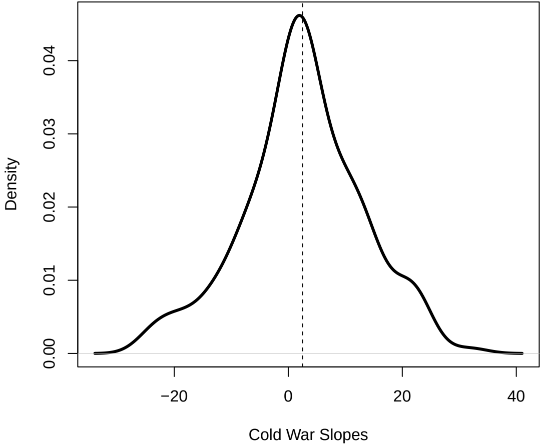
```
> VarCorr(HLM1)
 Groups   Name        Std.Dev. Corr
 ISO3     (Intercept) 24.60
          PostColdWar 11.88    -0.29
 Residual              6.11
```

```
> Bs<-data.frame(coef(HLM1)[1])
>
> head(Bs)
    ISO3..Intercept. ISO3.PopGrowth ISO3.UrbanPopulation ISO3.FertilityRate
AFG           -21.06        -0.3185                 0.34             -3.857
AGO           -25.58        -0.3185                 0.34             -3.857
ALB           -18.36        -0.3185                 0.34             -3.857
ARE          -105.35        -0.3185                 0.34             -3.857
ARG           -72.93        -0.3185                 0.34             -3.857
ARM           -26.76        -0.3185                 0.34             -3.857
    ISO3.log.GDPPerCapita. ISO3.NaturalResourceRents ISO3.PostColdWar
AFG                  10.51                   0.01144            3.618
AGO                  10.51                   0.01144           13.978
ALB                  10.51                   0.01144            6.373
ARE                  10.51                   0.01144            2.628
ARG                  10.51                   0.01144           22.861
ARM                  10.51                   0.01144            3.014

> mean(Bs$ISO3..Intercept.)
[1] -31.42

> mean(Bs$ISO3.PostColdWar)
[1] 2.52
```

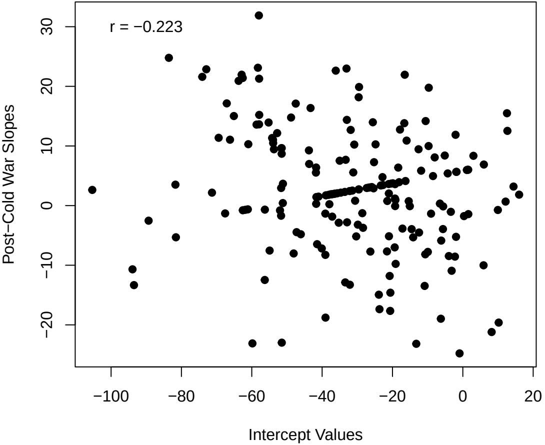
Random Intercepts (Plotted)



Random Slopes for `Post-ColdWar` (Plotted)



Scatterplot: Random Intercepts and Slopes



Separating Within and Between Effects

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

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

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

- Theory: Countries with large natural resource endowments will have <u>lower</u> WBLI on average, *but*
- When natural resource rents increase *within* a country, that will be associated with <u>higher</u> expected WBLI...

Combining Within- and Between-Effects

	WEBE.OLS
Population Growth	-1.638^{***}
	(0.114)
Urban Population	-0.047^{***}
	(0.010)
Fertility Rate	-2.633^{***}
	(0.151)
ln(GDP Per Capita)	2.460^{***}
	(0.186)
Within-Country Nat. Resource Rents	0.109^{***}
	(0.027)
Between-Country Nat. Resource Rents	-0.540^{***}
	(0.017)
Post-Cold War	10.930^{***}
	(0.358)
Constant	50.960^{***}
	(1.755)
Observations	8,318
R^2	0.524
Adjusted R^2	0.524
Residual Std. Error	13.130 (df = 8310)
F Statistic	$1,309.000^{***}$ (df = 7; 8310)

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

Interpretation

Two important things to remember:

1. Recall the variation in *Natural Resource Rents*:

Variable	Dimension	Mean	SD	Min	Max	Observations
Natural Resource Rents	overall	7.194	11.142	0	88.592	$NT = 8318$
	between		10.02	0	43.459	$N = 187$
	within		5.285	-24.03	61.301	$T = 44.481$

It is important to keep this in mind when discussing the relative sizes of effects:

- A one-s.d. increase in `NaturalResourceRents` within a country yields an expected change in Y of $5.3 \times 0.11 = 0.58$
- A one-s.d. difference in `NaturalResourceRents` between two countries yields an expected change in Y of $10.0 \times -0.55 = -5.5$

Interpretation (continued)

2. We can formally test whether $\hat{\beta}_B = \hat{\beta}_W$ for the *Natural Resource Rents* variable:

```
> library(car)
> linearHypothesis(WEBE.OLS,c("NRR.Within=NRR.Between"))

Linear hypothesis test:
NRR.Within - NRR.Between = 0

Model 1: restricted model
Model 2: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
    log(GDPPerCapita) + NRR.Within + NRR.Between + PostColdWar

  Res.Df     RSS Df Sum of Sq   F Pr(>F)
1   8311 1504209
2   8310 1431987  1     72222 419 <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Extension: The "Mundlak Device"

Mundlak (1978)[1] notes that in:

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

the only portion of α_i that can be correlated with u_{it} must also be correlated with *only* the means of $\mathbf{X}$ for each individual. Thus, a model that includes / "controls for" the unit-level means of $\mathbf{X}_{it}$:

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

...will ensure that the remaining variation in $\mathbf{X}_{it}$ is uncorrelated with u_{it}.

Moreover, a test for $\hat{\gamma} = 0$ is similar (and, as it turns out, superior) to the Hausman test for FE vs. RE.

[1] Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46:69-85.

"Mundlak Device" (continued)

```
# Create unit-level means of *all* the (time-varying) predictors:

smol$PGBetween<-plm::Between(smol$PopGrowth,effect="individual")
smol$UPBetween<-plm::Between(smol$UrbanPopulation,effect="individual")
smol$FRBetween<-plm::Between(smol$FertilityRate,effect="individual")
smol$GDPBetween<-plm::Between(smol$lnGDPPerCap,effect="individual")
smol$NRRBetween<-plm::Between(smol$NaturalResourceRents,effect="individual")
smol$PCWBetween<-plm::Between(smol$PostColdWar,effect="individual")


# Regress Y on both using plain-old OLS:

MD<-plm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+lnGDPPerCap+
        NaturalResourceRents+PostColdWar+PGBetween+UPBetween+FRBetween+GDPBetween+
        NRRBetween+PCWBetween,data=smol,effect="individual",model="pooling")
```

FE and Mundlak Models

	FE2	Mundlak
Population Growth	-0.189^{**}	-0.189
	(0.081)	(0.127)
Urban Population	0.310^{***}	0.310^{***}
	(0.020)	(0.031)
Fertility Rate	-1.763^{***}	-1.763^{***}
	(0.154)	(0.243)
ln(GDP Per Capita)	8.779^{***}	8.779^{***}
	(0.294)	(0.463)
Natural Resource Rents	0.072^{***}	0.072^{***}
	(0.017)	(0.026)
Post-Cold War	6.944^{***}	6.944^{***}
	(0.288)	(0.455)
Between-Country Population Growth		-5.275^{***}
		(0.247)
Between-Country Urban Population		-0.340^{***}
		(0.033)
Between-Country Fertility Rate		1.189^{***}
		(0.326)
Between-Country ln(GDP Per Capita)		-6.179^{***}
		(0.510)
Between-Country Nat. Resource Rents		-0.466^{***}
		(0.032)
Between-Country Post-Cold War		0.712
		(1.211)
Constant		49.070^{***}
		(2.412)
Observations	8,318	8,318
R^2	0.528	0.572
Adjusted R^2	0.517	0.572
F Statistic	$1{,}515.000^{***}$ (df = 6; 8125)	926.800^{***} (df = 12; 8305)

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

“Mundlak Device”: Testing

```
> # Testing:

> linearHypothesis(MD,"PGBetween+UPBetween+FRBetween+
+                  GDPBetween+NRRBetween+PCWBetween")

Linear hypothesis test:
PGBetween  + UPBetween  + FRBetween  + GDPBetween  + NRRBetween  + PCWBetween = 0

Model 1: restricted model
Model 2: WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +
    lnGDPPerCap + NaturalResourceRents + PostColdWar + PGBetween +
    UPBetween + FRBetween + GDPBetween + NRRBetween + PCWBetween

  Res.Df Df Chisq      Pr(>Chisq)
1   8306
2   8305  1  49.5 0.0000000000019 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Two-Way Unit Effects

Our original decomposition considered "two-way" effects:

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

This "controls for" both non-time-varying unit-level differences and non-cross-sectionally-varying differences between periods. It also 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

The Good:

- 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)

The (Potentially) Bad:

- 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 $\text{Cov}(\mathbf{X}_{it}, \eta_t) = \text{Cov}(\alpha_i, \eta_t) = 0$
- Two-way effects models ask a *lot* of your data (effectively fits $N + T + k$ parameters using NT observations)

Example: Two-Way Fixed Effects

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

> summary(TwoWayFE)
Twoways effects Within Model

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-32.072  -4.062   0.233   4.131  43.464

Coefficients:
                     Estimate Std. Error t-value Pr(>|t|)
PopGrowth             -0.2801     0.0677   -4.14  3.5e-05 ***
UrbanPopulation        0.0314     0.0172    1.82    0.069 .
FertilityRate          1.3204     0.1411    9.36  < 2e-16 ***
log(GDPPerCapita)      2.0950     0.2710    7.73  1.2e-14 ***
NaturalResourceRents   0.0363     0.0145    2.51    0.012 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    358000
Residual Sum of Squares: 350000
R-Squared:      0.0209
Adj. R-Squared: -0.0084
F-statistic: 34.5476 on 5 and 8075 DF, p-value: <2e-16
```

Two-Way Effects: Testing

```
> # Two-way effects:

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

 F test for twoways effects

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
F = 113, df1 = 236, df2 = 8075, p-value <2e-16
alternative hypothesis: significant effects

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

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

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 192, 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)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 238, p-value <2e-16
alternative hypothesis: significant effects

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

 Lagrange Multiplier Test - time effects (King and Wu)

data:  WomenBusLawIndex ~ PopGrowth + UrbanPopulation + FertilityRate +  ...
normal = 92, p-value <2e-16
alternative hypothesis: significant effects
```

Two-Way Fixed Effects via `lm`

```
> TwoWayFE.BF<-lm(WomenBusLawIndex~PopGrowth+UrbanPopulation+FertilityRate+log(GDPPerCapita)+NaturalResourceRents+
+                  factor(ISO3)+factor(Year)-1,data=WDI)

> summary(TwoWayFE.BF)

Call:
lm(formula = WomenBusLawIndex ~ PopGrowth + UrbanPopulation +
    FertilityRate + log(GDPPerCapita) + NaturalResourceRents +
    factor(ISO3) + factor(Year) - 1, data = WDI)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
PopGrowth             -0.2801     0.0677   -4.14  3.5e-05 ***
UrbanPopulation        0.0314     0.0172    1.82  0.06889 .
FertilityRate          1.3204     0.1411    9.36  < 2e-16 ***
log(GDPPerCapita)      2.0950     0.2710    7.73  1.2e-14 ***
NaturalResourceRents   0.0363     0.0145    2.51  0.01224 *
factor(ISO3)AFG      -17.0383     2.4282   -7.02  2.5e-12 ***
factor(ISO3)AGO       11.5212     2.6218    4.39  1.1e-05 ***
factor(ISO3)ALB       34.8172     2.4144   14.42  < 2e-16 ***
factor(ISO3)ARE      -21.5963     3.2954   -6.55  6.0e-11 ***
.
.
.
factor(Year)1977       4.2474     0.8758    4.85  1.3e-06 ***
factor(Year)1978       5.1409     0.8764    5.87  4.6e-09 ***
 [ reached 'max' / getOption("max.print") -- omitted 43 rows ]
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.59 on 8075 degrees of freedom
  (5657 observations deleted due to missingness)
Multiple R-squared:  0.99, Adjusted R-squared:  0.989
F-statistic: 3.14e+03 on 243 and 8075 DF,  p-value: <2e-16
```

Two-Way Random Effects

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

Unbalanced Panel: n = 187, T = 8-52, N = 8318

Effects:
                 var std.dev share
idiosyncratic 43.365   6.585  0.31
individual    96.697   9.833  0.69
time           0.306   0.553  0.00
theta:
        Min. 1st Qu. Median   Mean 3rd Qu.   Max.
id    0.7696  0.8972 0.9075 0.9010  0.9075 0.9075
time  0.2497  0.2957 0.3342 0.3179  0.3404 0.3424
total 0.2490  0.2953 0.3336 0.3173  0.3393 0.3419

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-65.454  -7.400   1.976  -0.223  10.725  34.264

Coefficients:
                     Estimate Std. Error z-value Pr(>|z|)
(Intercept)           2.50627    0.32839    7.63  2.3e-14 ***
PopGrowth            -0.25242    0.01023  -24.68  < 2e-16 ***
UrbanPopulation       0.14627    0.00236   61.86  < 2e-16 ***
FertilityRate        -0.83970    0.01989  -42.22  < 2e-16 ***
log(GDPPerCapita)     5.51216    0.03618  152.33  < 2e-16 ***
NaturalResourceRents  0.04082    0.00212   19.26  < 2e-16 ***
PostColdWar          11.24183    0.04494  250.16  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    3010000
Residual Sum of Squares: 2250000
R-Squared:      0.293
Adj. R-Squared: 0.292
Chisq: 267472 on 6 DF, p-value: <2e-16
```

A Prettier Table

	OLS	FE	BE	RE	TwoWayFE	TwoWayRE
Population Growth	-1.776^{***}	-0.189^{**}	-4.992^{***}	-0.232^{***}	-0.280^{***}	-0.252^{***}
	(0.117)	(0.081)	(1.128)	(0.082)	(0.068)	(0.010)
Urban Population	-0.066^{***}	0.310^{***}	-0.051	0.193^{***}	0.031^{*}	0.146^{***}
	(0.010)	(0.020)	(0.055)	(0.018)	(0.017)	(0.002)
Fertility Rate	-2.984^{***}	-1.763^{***}	-1.507	-2.106^{***}	1.320^{***}	-0.840^{***}
	(0.154)	(0.154)	(1.173)	(0.154)	(0.141)	(0.020)
ln(GDP Per Capita)	2.469^{***}	8.779^{***}	1.951^{*}	7.225^{***}	2.095^{***}	5.512^{***}
	(0.191)	(0.294)	(1.146)	(0.278)	(0.271)	(0.036)
Natural Resource Rents	-0.363^{***}	0.072^{***}	-0.313^{***}	0.045^{***}	0.036^{**}	0.041^{***}
	(0.015)	(0.017)	(0.092)	(0.017)	(0.015)	(0.002)
Post-Cold War	10.280^{***}	6.944^{***}	7.717	8.181^{***}		11.240^{***}
	(0.365)	(0.288)	(5.326)	(0.283)		(0.045)
Constant	52.560^{***}		57.610^{***}	-7.314^{***}		2.506^{***}
	(1.797)		(12.870)	(2.499)		(0.328)
Observations	8,318	8,318	187	8,318	8,318	8,318
R^2	0.500	0.528	0.580	0.511	0.021	0.293
Adjusted R^2	0.500	0.517	0.566	0.511	-0.008	0.292

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

Interpretation: `modelsummary` and `marginaleffects`

Two broadly useful packages by Vincent Arel-Bundock:

1. `modelsummary`
    - "`modelsummary` creates tables and plots to present descriptive statistics and to summarize statistical models in R."
    - In short, it... summarizes models (and data)

2. `marginaleffects`
    - An "R package to compute and plot predictions, slopes, marginal means, and comparisons (contrasts, risk ratios, odds ratios, etc.) for over 70 classes of statistical models in R."
    - Focus is on individual covariates / predictors
    - Especially handy for nonlinear models (GLMs, etc.)

Both are highly flexible and customizable, and **both play well with the models fit by** `plm`.

Using `modelsummary`

The most basic version of the code is:

```
> modelsummary(list(OLS,FE,RE))
```

...from which we get this (in the Viewer window):

	(1)	(2)	(3)
(Intercept)	52.559		-7.314
	(1.797)		(2.499)
PopGrowth	-1.776	-0.189	-0.232
	(0.117)	(0.081)	(0.082)
UrbanPopulation	-0.066	0.310	0.193
	(0.010)	(0.020)	(0.018)
FertilityRate	-2.984	-1.763	-2.106
	(0.154)	(0.154)	(0.154)
log(GDPPerCapita)	2.469	8.779	7.225
	(0.191)	(0.294)	(0.278)
NaturalResourceRents	-0.363	0.072	0.045
	(0.015)	(0.017)	(0.017)
PostColdWar	10.275	6.944	8.181
	(0.365)	(0.288)	(0.283)
Num.Obs.	8318	8318	8318
R2	0.500	0.528	0.511
R2 Adj.	0.500	0.517	0.511
AIC	66855.1	57781.0	58384.9
BIC	66911.3	57830.2	58441.1
RMSE	13.45	7.79	8.08

Using modelsummary (continued)

A better-looking table:

```
> modelsummary(models,output="MS-Table-25.tex",title="Models of WDBI",stars=TRUE,fmt=2,
+              gof_map=c("nobs","r.squared","adj.r.squared"),coef_rename=c("PopGrowth"="Population Growth",
+              "UrbanPopulation"="Urban Population","FertilityRate"="Fertility Rate","log(GDPPerCapita)"="ln(GDP Per Capita)",
+              "NaturalResourceRents"="Natural Resource Rents","PostColdWar"="Post-Cold War"))
```

Table 1: Models of WDBI

	OLS	Within	Random
(Intercept)	52.56***		−7.31**
	(1.80)		(2.50)
Population Growth	−1.78***	−0.19*	−0.23**
	(0.12)	(0.08)	(0.08)
Urban Population	−0.07***	0.31***	0.19***
	(0.01)	(0.02)	(0.02)
Fertility Rate	−2.98***	−1.76***	−2.11***
	(0.15)	(0.15)	(0.15)
ln(GDP Per Capita)	2.47***	8.78***	7.23***
	(0.19)	(0.29)	(0.28)
Natural Resource Rents	−0.36***	0.07***	0.05**
	(0.02)	(0.02)	(0.02)
Post-Cold War	10.28***	6.94***	8.18***
	(0.37)	(0.29)	(0.28)
Num.Obs.	8318	8318	8318
R2	0.500	0.528	0.511
R2 Adj.	0.500	0.517	0.511

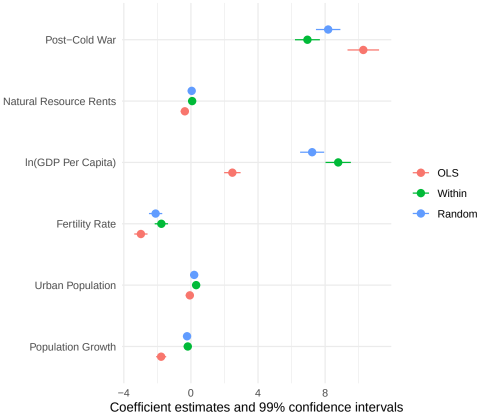
+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

modelsummary: Coefficient Plots

Example code:

```
> modelplot(models,conf_level=0.99,coef_omit="(Intercept)",coef_rename=c("PopGrowth"="Population Growth",
+                           "UrbanPopulation"="Urban Population","FertilityRate"="Fertility Rate",
+                           "log(GDPPerCapita)"="ln(GDP Per Capita)","NaturalResourceRents"="Natural Resource Rents",
+                           "PostColdWar"="Post-Cold War"))
```

...from which we get:



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 panelr package (various commands)
- 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