GSERM - St. Gallen 2024
Analyzing Panel Data

June 11, 2024

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:

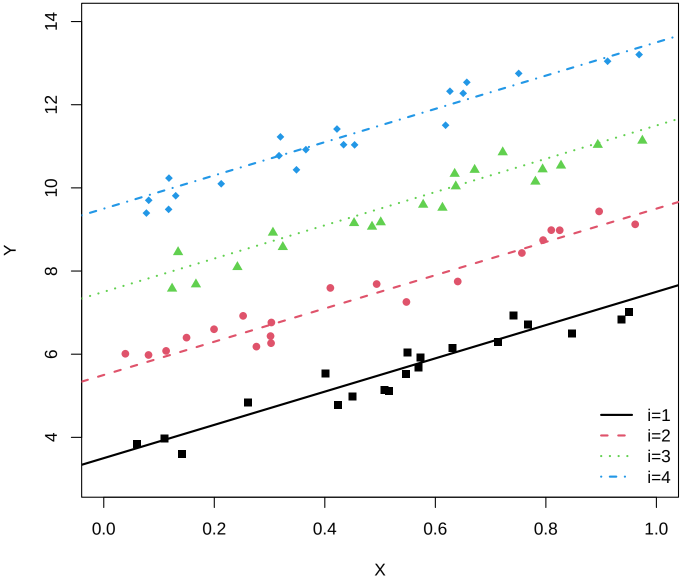
- Assuming $\alpha_i = 0$ (w.l.o.g):

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

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

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

Intuition: One-Way Unit Effects



(One-Way) "Fixed" Effects

"Brute force" model fits:

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

In other words:

- 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 cannot be estimated

"Fixed" Effects: Test(s)

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

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

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

Example Redux: WDI, 1960-2023

The World Development Indicators

- Cross-national country-level time series data
- $N = 215$ countries, $T = 64$ years (1960-2023) + 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.

Data Summary

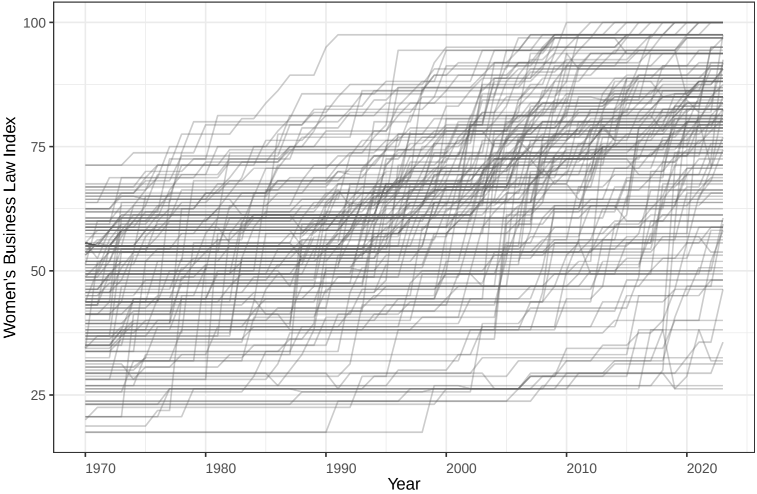
```
> describe(wdi,fast=TRUE,ranges=FALSE,check=TRUE)
                        vars     n            mean      sd  skew kurtosis            se
ISO3                       1 13760             NaN      NA    NA       NA            NA
Year                       2 13760         1991.50 1.8e+01  0.00    -1.20          0.16
Region                     3 13760             NaN      NA    NA       NA            NA
country                    4 13760             NaN      NA    NA       NA            NA
iso3c                      5 13760             NaN      NA    NA       NA            NA
LandArea                   6 11941       605302.93 1.6e+06  5.41    34.35      15006.31
ArablePercent              7 11542           13.35 1.4e+01  1.49     2.05          0.13
Population                 8 13515     25109276.98 1.0e+08  9.73   105.53     901048.65
PopGrowth                  9 13298            1.75 1.8e+00  0.83    22.79          0.02
RuralPopulation           10 13482           48.28 2.6e+01 -0.12    -1.00          0.22
UrbanPopulation           11 13482           51.72 2.6e+01  0.12    -1.00          0.22
BirthRatePer1K            12 12937           28.02 1.3e+01  0.21    -1.25          0.12
FertilityRate             13 12779            3.91 2.0e+00  0.38    -1.23          0.02
PrimarySchoolAge          14 10896            6.14 6.1e-01 -0.04     0.11          0.01
LifeExpectancy            15 12766           64.63 1.1e+01 -0.73    -0.03          0.10
AgeDepRatioOld            16 13515           10.70 7.0e+00  1.74     4.57          0.06
CO2Emissions              17  5920            4.24 5.5e+00  2.75    11.36          0.07
GDP                       18 10099 250284546944.35 1.1e+12 11.01   146.93 11352953710.99
Inflation                 19  8547           23.50 3.3e+02 53.07  3489.78          3.54
TotalTrade                20  8622           78.38 5.4e+01  2.99    17.71          0.58
Exports                   21  8622           36.51 2.9e+01  2.95    16.16          0.31
Imports                   22  8631           41.88 2.8e+01  2.54    13.54          0.30
FDIIn                     23  8484            5.49 4.5e+01 15.71   572.23          0.49
AgriEmployment            24  5951           28.79 2.4e+01  0.64    -0.74          0.31
NetAidReceived            25  9043    506951242.00 1.0e+09  8.32   157.34   10484966.48
MobileCellSubscriptions   26 10212           36.32 5.2e+01  1.29     1.14          0.51
NaturalResourceRents      27  9211            6.85 1.1e+01  2.60     8.04          0.12
MilitaryExpenditures      28  7555            2.72 3.2e+00  9.45   240.84          0.04
GovtExpenditures          29  8280           16.33 8.2e+00  3.82    34.97          0.09
PublicEdExpend            30  4925            4.35 1.9e+00  2.89    39.41          0.03
PublicHealthExpend        31  3930            3.27 2.3e+00  1.32     2.95          0.04
HIVDeaths                 32  4370         8137.20 2.5e+04  5.93    45.32        377.09
WomenBusLawIndex          33 10152           59.85 1.9e+01  0.02    -0.58          0.19
PaidParentalLeave         34 10152            0.11 3.1e-01  2.50     4.27          0.00
PostColdWar               35 13760            0.53 5.0e-01 -0.13    -1.98          0.00
GDPPerCapita              36 10099        12093.81 1.9e+04  3.19    14.84        190.34
```

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

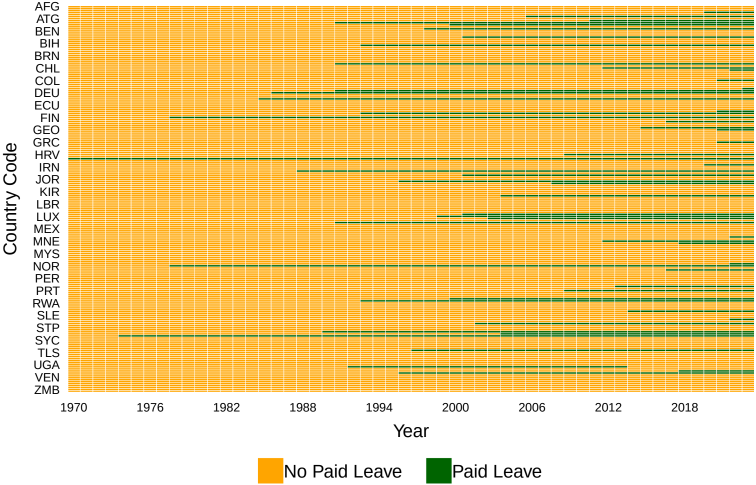
The basis for a 2021 World Bank report...

- 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 $\approx$ 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."

Visualization (using `panelView`)



Categorical Variable Visualization

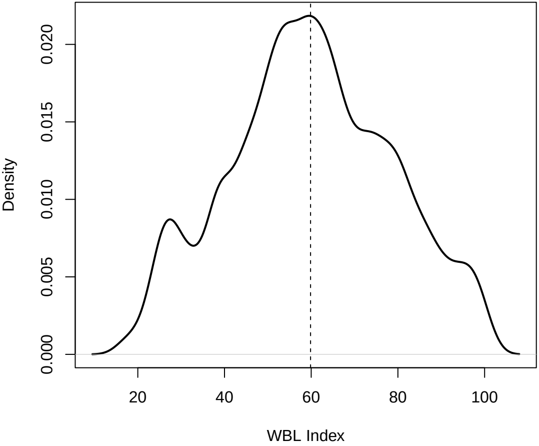


plm (panel linear models) is the workhorse R package for fitting linear models to panel data. To do so, you must first declare the data as panel data and indicate the variables containing the unit (i) and period (t) indices:

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

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

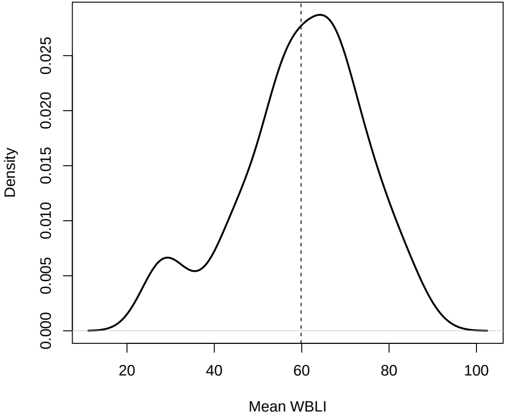
WBLI: Total Variation

```
> describe(WBLI,na.rm=TRUE) # all variation
   vars     n mean sd median trimmed mad min max range skew kurtosis   se
X1    1 10152   60 19     59      60  19  18 100    82 0.02    -0.58 0.19
```



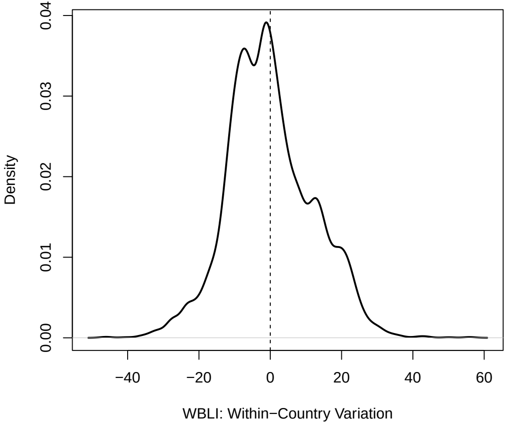
WBLI: "Between" Variation

```
> 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   60 14     61      61  13  24  90    66 -0.5    -0.12 1.1
```



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 12   -1.2   -0.32  11 -46  56   102 0.26     0.35 0.12
```



Regression model:

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

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

Descriptive Statistics:

	vars	n	mean	sd	median	min	max	range	skew	kurtosis	se
WomenBusLawIndex	1	8127	60.6	18.98	60.6	17.50	100.0	82.5	-0.03	-0.67	0.21
PopGrowth	2	8127	1.6	1.54	1.6	-16.88	19.4	36.2	1.18	18.83	0.02
UrbanPopulation	3	8127	51.6	23.82	51.5	2.85	100.0	97.2	0.07	-1.02	0.26
FertilityRate	4	8127	3.6	1.90	3.1	0.77	8.6	7.8	0.52	-1.02	0.02
NaturalResourceRents	5	8127	7.0	10.76	2.5	0.00	88.6	88.6	2.52	7.59	0.12
PostColdWar	6	8127	0.7	0.46	1.0	0.00	1.0	1.0	-0.86	-1.26	0.01
lnGDPPerCap	7	8127	8.3	1.44	8.2	4.92	11.7	6.8	0.14	-0.87	0.02

Where is the Variation in our Data?

Variable	Dim	Mean	SD	Min	Max	Observations
WomenBusLawIndex	overall	60.645	18.978	17.5	100	N = 8127
	between		15.467	23.329	94.56	n = 187
	within		11.407	15.356	117.932	T = 43.46
PopGrowth	overall	1.652	1.543	-16.881	19.36	N = 8127
	between		1.268	-1.077	6.717	n = 187
	within		0.992	-17.672	16.634	T = 43.46
UrbanPopulation	overall	51.644	23.818	2.845	100	N = 8127
	between		22.962	7.649	100	n = 187
	within		6.881	13.353	80.27	T = 43.46
FertilityRate	overall	3.608	1.896	0.772	8.606	N = 8127
	between		1.673	1.26	7.585	n = 187
	within		0.915	0.968	7.376	T = 43.46
NaturalResourceRents	overall	7.031	10.761	0	88.592	N = 8127
	between		9.994	0	43.274	n = 187
	within		5.16	-24.193	61.138	T = 43.46
PostColdWar	overall	0.698	0.459	0	1	N = 8127
	between		0.162	0.6	1	n = 187
	within		0.438	-0.272	1.098	T = 43.46
lnGDPPerCap	overall	8.302	1.437	4.924	11.68	N = 8127
	between		1.388	5.803	11.199	n = 187
	within		0.352	6.462	10.248	T = 43.46

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 = 1-52, N = 8127

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
 -49.49   -8.52    1.10    9.27   44.23

Coefficients:
                     Estimate Std. Error t-value    Pr(>|t|)
(Intercept)           49.6205     1.8348   27.04     < 2e-16 ***
PopGrowth             -2.3344     0.1308  -17.84     < 2e-16 ***
UrbanPopulation       -0.0566     0.0105   -5.38 0.000000076 ***
FertilityRate         -2.5607     0.1590  -16.10     < 2e-16 ***
log(GDPPerCapita)      2.6396     0.1944   13.58     < 2e-16 ***
NaturalResourceRents  -0.3430     0.0155  -22.10     < 2e-16 ***
PostColdWar           10.8062     0.3707   29.15     < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    2930000
Residual Sum of Squares: 1470000
R-Squared:      0.499
Adj. R-Squared: 0.499
F-statistic: 1348.85 on 6 and 8120 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 = 1-52, N = 8127

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-33.544  -5.098  -0.433   4.842  53.156

Coefficients:
                     Estimate Std. Error t-value Pr(>|t|)
PopGrowth             -0.1060     0.0954   -1.11  0.26688
UrbanPopulation        0.2959     0.0198   14.93  < 2e-16 ***
FertilityRate         -2.0058     0.1617  -12.40  < 2e-16 ***
log(GDPPerCapita)      8.8327     0.2998   29.46  < 2e-16 ***
NaturalResourceRents   0.0665     0.0172    3.87  0.00011 ***
PostColdWar            7.0291     0.2942   23.89  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1060000
Residual Sum of Squares: 494000
R-Squared:      0.533
Adj. R-Squared: 0.521
F-statistic: 1507.19 on 6 and 7934 DF, p-value: <2e-16
```

A Nicer Table

Table: Models of WBLI

	OLS	FE
Population Growth	-2.300^{***}	-0.110
	(0.130)	(0.095)
Urban Population	-0.057^{***}	0.300^{***}
	(0.011)	(0.020)
Fertility Rate	-2.600^{***}	-2.000^{***}
	(0.160)	(0.160)
ln(GDP Per Capita)	2.600^{***}	8.800^{***}
	(0.190)	(0.300)
Natural Resource Rents	-0.340^{***}	0.066^{***}
	(0.016)	(0.017)
Post-Cold War	11.000^{***}	7.000^{***}
	(0.370)	(0.290)
Constant	50.000^{***}	
	(1.800)	
Observations	8,127	8,127
R^2	0.500	0.530
Adjusted R^2	0.500	0.520
F Statistic	$1,349.000^{***}$ (df = 6; 8120)	$1,507.000^{***}$ (df = 6; 7934)

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

The model is:

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

which is estimated via "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

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

	OLS	FE.Units	FE.Time
Population Growth	-2.300^{***}	-0.110	-2.600^{***}
	(0.130)	(0.095)	(0.120)
Urban Population	-0.062^{***}	0.300^{***}	-0.063^{***}
	(0.011)	(0.020)	(0.010)
Fertility Rate	-2.500^{***}	-2.000^{***}	-1.600^{***}
	(0.160)	(0.160)	(0.150)
ln(GDP Per Capita)	2.800^{***}	8.800^{***}	3.200^{***}
	(0.200)	(0.300)	(0.180)
Natural Resource Rents	-0.340^{***}	0.066^{***}	-0.380^{***}
	(0.016)	(0.017)	(0.015)
Post-Cold War	11.000^{***}	7.000^{***}	
	(0.370)	(0.290)	
Constant	48.000^{***}		
	(1.800)		
Observations	8,127	8,127	8,127
R^2	0.500	0.530	0.410
Adjusted R^2	0.500	0.520	0.400
F Statistic	$1{,}356.000^{***}$ (df = 6; 8120)	$1{,}507.000^{***}$ (df = 6; 7934)	$1{,}103.000^{***}$ (df = 5; 8070)

*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 = 84, df1 = 186, df2 = 7934, 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 = 53439, 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 = 231, 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 = 8070, 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 = 8596, 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 = 93, 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 observation i's 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...

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 $\mathbf{X}$ 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	Dim	Mean	SD	Min	Max	Observations
UrbanPopulation	overall	51.644	23.818	2.845	100	N = 8127
	between		22.962	7.649	100	n = 187
	within		6.881	13.353	80.27	T = 43.46

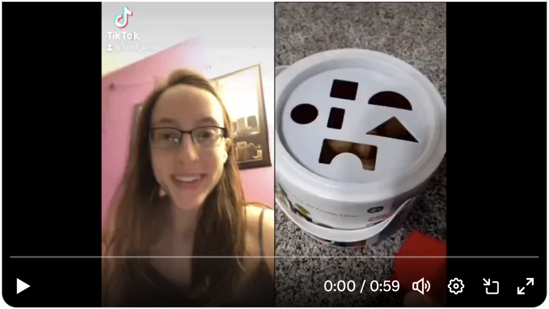
"While the overall variation in the independent variable may be large, the within-unit variation used to estimate β may be much smaller" (M & P 2018, p. 830).

Whither "Fixed" Effects"?

Andrew Charles Baker
@Andrew___Baker

Applied economists using fixed effects regressions

Alison Burke @TiredActor · Feb 19
I like to think a lot of stuff but mostly this x.com/billygrene/sta...



10:03 AM · Feb 20, 2024 · **27.9K** Views

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 = 1-52, N = 8127
Observations used in estimation: 187

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-30.595  -6.402   0.722   8.139  21.227

Coefficients:
                     Estimate Std. Error t-value   Pr(>|t|)
(Intercept)           50.0410    12.8414    3.90    0.00014 ***
PopGrowth             -5.7235     1.0403   -5.50 0.00000013 ***
UrbanPopulation       -0.0368     0.0553   -0.67    0.50643
FertilityRate         -0.6327     1.1314   -0.56    0.57673
log(GDPPerCapita)      2.3879     1.1571    2.06    0.04048 *
NaturalResourceRents  -0.3295     0.0914   -3.61    0.00040 ***
PostColdWar            9.6833     5.1642    1.88    0.06240 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    44500
Residual Sum of Squares: 18700
R-Squared:      0.58
Adj. R-Squared: 0.566
F-statistic: 41.4288 on 6 and 180 DF, p-value: <2e-16
```

A Nicer Table (Again)

Table: Models of WBLI

	OLS	FE	BE
Population Growth	-2.300^{***}	-0.110	-5.700^{***}
	(0.130)	(0.095)	(1.000)
Urban Population	-0.057^{***}	0.300^{***}	-0.037
	(0.011)	(0.020)	(0.055)
Fertility Rate	-2.600^{***}	-2.000^{***}	-0.630
	(0.160)	(0.160)	(1.100)
ln(GDP Per Capita)	2.600^{***}	8.800^{***}	2.400^{**}
	(0.190)	(0.300)	(1.200)
Natural Resource Rents	-0.340^{***}	0.066^{***}	-0.330^{***}
	(0.016)	(0.017)	(0.091)
Post-Cold War	11.000^{***}	7.000^{***}	9.700^{*}
	(0.370)	(0.290)	(5.200)
Constant	50.000^{***}		50.000^{***}
	(1.800)		(13.000)
Observations	8,127	8,127	187
R^2	0.500	0.530	0.580
Adjusted R^2	0.500	0.520	0.570
F Statistic	$1{,}349.000^{***}$ (df = 6; 8120)	$1{,}507.000^{***}$ (df = 6; 7934)	41.000^{***} (df = 6; 180)

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

"Random" Effects

Model:

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

with:

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

and

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

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:

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

"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 = 1-52, N = 8127

Effects:
                var std.dev share
idiosyncratic 62.28    7.89  0.39
individual    96.38    9.82  0.61
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.37    0.88    0.89    0.88    0.89    0.89

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -33.8    -5.4    -0.5     0.0     5.6    44.2

Coefficients:
                     Estimate Std. Error z-value Pr(>|z|)
(Intercept)           -6.1372     2.5605   -2.40    0.017 *
PopGrowth             -0.1989     0.0974   -2.04    0.041 *
UrbanPopulation        0.1807     0.0187    9.69   <2e-16 ***
FertilityRate         -2.3150     0.1614  -14.34   <2e-16 ***
log(GDPPerCapita)      7.2369     0.2843   25.46   <2e-16 ***
NaturalResourceRents   0.0362     0.0172    2.11    0.035 *
PostColdWar            8.2940     0.2891   28.69   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    1100000
Residual Sum of Squares: 532000
R-Squared:      0.516
Adj. R-Squared: 0.516
Chisq: 8439.21 on 6 DF, p-value: <2e-16
```

A Nicer Table (Yet Again)

Table: Models of WBLI

	OLS	FE	BE	RE
Population Growth	-2.300^{***}	-0.110	-5.700^{***}	-0.200^{**}
	(0.130)	(0.095)	(1.000)	(0.097)
Urban Population	-0.057^{***}	0.300^{***}	-0.037	0.180^{***}
	(0.011)	(0.020)	(0.055)	(0.019)
Fertility Rate	-2.600^{***}	-2.000^{***}	-0.630	-2.300^{***}
	(0.160)	(0.160)	(1.100)	(0.160)
ln(GDP Per Capita)	2.600^{***}	8.800^{***}	2.400^{**}	7.200^{***}
	(0.190)	(0.300)	(1.200)	(0.280)
Natural Resource Rents	-0.340^{***}	0.066^{***}	-0.330^{***}	0.036^{**}
	(0.016)	(0.017)	(0.091)	(0.017)
Post-Cold War	11.000^{***}	7.000^{***}	9.700^{*}	8.300^{***}
	(0.370)	(0.290)	(5.200)	(0.290)
Constant	50.000^{***}		50.000^{***}	-6.100^{**}
	(1.800)		(13.000)	(2.600)
Observations	8,127	8,127	187	8,127
R^2	0.500	0.530	0.580	0.520
Adjusted R^2	0.500	0.520	0.570	0.520
F Statistic	$1{,}349.000^{***}$ (df = 6; 8120)	$1{,}507.000^{***}$ (df = 6; 7934)	41.000^{***} (df = 6; 180)	$8{,}439.000^{***}$

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

"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**

The Hausman Test

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 ($\text{Cov}(X_{it}, \alpha_i) = 0$).

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 = 5151, 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

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 *u*s
- 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: 57676

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.253 -0.648 -0.067  0.620  6.622

Random effects:
 Groups   Name        Variance Std.Dev.
 ISO3     (Intercept) 344.2    18.6
 Residual              62.4     7.9
Number of obs: 8127, groups:  ISO3, 187

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          -18.9801     2.8893   -6.57
PopGrowth             -0.1338     0.0953   -1.40
UrbanPopulation        0.2554     0.0193   13.23
FertilityRate         -2.1097     0.1605  -13.14
log(GDPPerCapita)      8.2665     0.2925   28.26
NaturalResourceRents   0.0575     0.0171    3.37
PostColdWar            7.4860     0.2902   25.80

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.045
UrbanPopltn -0.180  0.004
FertilityRt -0.446 -0.272  0.435
lg(GDPPrCp) -0.751 -0.047 -0.282  0.102
NtrlRsrcRnt -0.012 -0.081 -0.021 -0.099  0.005
PostColdWar  0.018 -0.042 -0.213  0.429 -0.116 -0.061
```

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

Table: RE and HLM Models of WBLI

	RE	AltRE
Population Growth	-0.200^{**}	-0.130
	(0.097)	(0.095)
Urban Population	0.180^{***}	0.260^{***}
	(0.019)	(0.019)
Fertility Rate	-2.300^{***}	-2.100^{***}
	(0.160)	(0.160)
ln(GDP Per Capita)	7.200^{***}	8.300^{***}
	(0.280)	(0.290)
Natural Resource Rents	0.036^{**}	0.057^{***}
	(0.017)	(0.017)
Post-Cold War	8.300^{***}	7.500^{***}
	(0.290)	(0.290)
Constant	-6.100^{**}	-19.000^{***}
	(2.600)	(2.900)
Observations	8,127	8,127
R^2	0.520	
Adjusted R^2	0.520	
Log Likelihood		-28,838.000
Akaike Inf. Crit.		57,694.000
Bayesian Inf. Crit.		57,757.000
F Statistic	$8,439.000^{***}$	

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

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: 54302

Scaled residuals:
   Min     1Q Median     3Q    Max
-4.351 -0.535  0.008  0.536  8.029

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 ISO3     (Intercept) 616.2    24.82
          PostColdWar 140.6    11.86    -0.30
 Residual              37.7     6.14
Number of obs: 8127, groups:  ISO3, 187

Fixed effects:
                     Estimate Std. Error t value
(Intercept)          -30.7610     3.3941   -9.06
PopGrowth             -0.2562     0.0777   -3.30
UrbanPopulation        0.3304     0.0225   14.70
FertilityRate         -4.0847     0.1734  -23.56
log(GDPPerCapita)     10.5726     0.3258   32.45
NaturalResourceRents   0.0079     0.0149    0.53
PostColdWar            2.5235     1.0190    2.48

Correlation of Fixed Effects:
            (Intr) PpGrwt UrbnPp FrtltR l(GDPP NtrlRR
PopGrowth    0.049
UrbanPopltn -0.129 -0.016
FertilityRt -0.509 -0.194  0.481
lg(GDPPrCp) -0.698 -0.055 -0.368  0.183
NtrlRsrcRnt  0.014 -0.059  0.058 -0.041 -0.069
PostColdWar -0.218  0.001 -0.051  0.114  0.001 -0.010
```

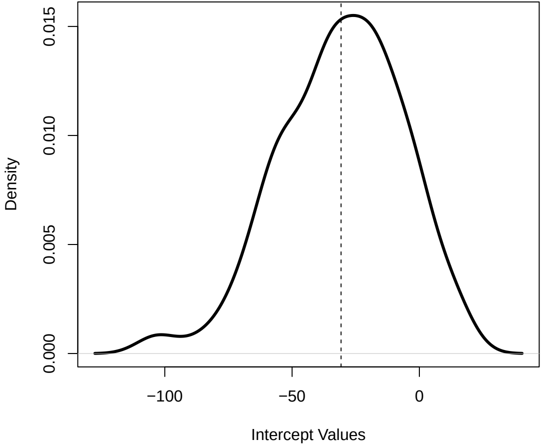
Testing

```
> 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 deviance Chisq Df Pr(>Chisq)
AltRE    9 57676 57740 -28829    57658
HLM1    11 54310 54387 -27144    54288  3370  2     <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

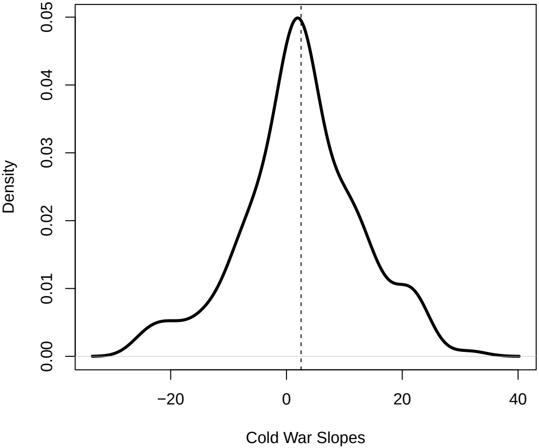
```
> VarCorr(HLM1)
 Groups   Name        Std.Dev. Corr
 ISO3     (Intercept) 24.82
          PostColdWar 11.86    -0.30
 Residual              6.14
```

```
> Bs<-data.frame(coef(HLM1)[1])
>
> head(Bs)
    ISO3..Intercept. ISO3.PopGrowth ISO3.UrbanPopulation ISO3.FertilityRate
AFG              -21          -0.26                 0.33               -4.1
AGO              -24          -0.26                 0.33               -4.1
ALB              -18          -0.26                 0.33               -4.1
ARE             -105          -0.26                 0.33               -4.1
ARG              -72          -0.26                 0.33               -4.1
ARM              -26          -0.26                 0.33               -4.1
    ISO3.log.GDPPerCapita. ISO3.NaturalResourceRents ISO3.PostColdWar
AFG                     11                    0.0079              3.5
AGO                     11                    0.0079             13.9
ALB                     11                    0.0079              6.4
ARE                     11                    0.0079              2.7
ARG                     11                    0.0079             22.8
ARM                     11                    0.0079              3.0

> mean(Bs$ISO3..Intercept.)
[1] -31

> mean(Bs$ISO3.PostColdWar)
[1] 2.5
```

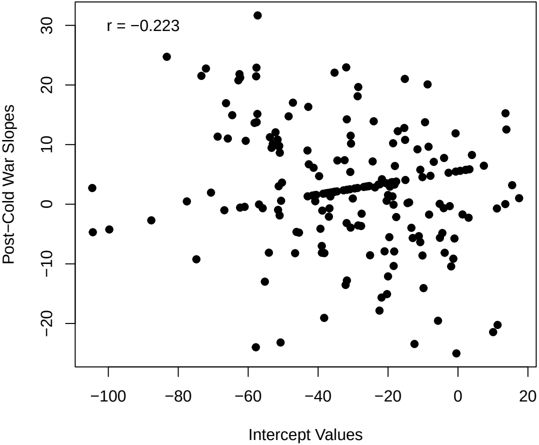
Random Intercepts (Plotted)



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 lower WBLI on average, *but*
- When natural resource rents increase *within* a country, that will be associated with higher expected WBLI...

Combining Within- and Between-Effects

Table: BE + WE Model of WBLI

	WEBE.OLS
Population Growth	-2.100^{***}
	(0.130)
Urban Population	-0.042^{***}
	(0.010)
Fertility Rate	-2.300^{***}
	(0.160)
ln(GDP Per Capita)	2.600^{***}
	(0.190)
Within-Country Nat. Resource Rents	0.097^{***}
	(0.028)
Between-Country Nat. Resource Rents	-0.510^{***}
	(0.018)
Post-Cold War	11.000^{***}
	(0.360)
Constant	49.000^{***}
	(1.800)
Observations	8,127
R^2	0.520
Adjusted R^2	0.520
Residual Std. Error	13.000 (df = 8119)
F Statistic	$1{,}254.000^{***}$ (df = 7; 8119)
	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Interpretation

Two important things to remember:

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

Variable	Dim	Mean	SD	Min	Max	Observations
NaturalResourceRents	overall	7.031	10.761	0	88.592	N = 8127
	between		9.994	0	43.274	n = 187
	within		5.16	-24.193	61.138	T = 43.46

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 $5.1 \times 0.098 = 0.5$
- A one-s.d. difference in `NaturalResourceRents` between two countries yields an expected change in Y of $9.6 \times -0.51 = -4.9$

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

Hypothesis:

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   8120 1465748
2   8119 1406260  1     59487 343 <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 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

	Fixed Effects	Mundlak
Population Growth	−0.110	−0.110
	(0.095)	(0.150)
Urban Population	0.300***	0.300***
	(0.020)	(0.031)
Fertility Rate	−2.000***	−2.000***
	(0.160)	(0.260)
ln(GDP Per Capita)	8.800***	8.800***
	(0.300)	(0.470)
Natural Resource Rents	0.066***	0.066**
	(0.017)	(0.027)
Post-Cold War	7.000***	7.000***
	(0.290)	(0.460)
Between-Country Population Growth		−6.000***
		(0.260)
Between-Country Urban Population		−0.320***
		(0.033)
Between-Country Fertility Rate		2.100***
		(0.330)
Between-Country ln(GDP Per Capita)		−5.800***
		(0.520)
Between-Country Nat. Resource Rents		−0.450***
		(0.032)
Between-Country Post-Cold War		3.000**
		(1.200)
Constant		42.000***
		(2.400)
Observations	8,127	8,127
R^2	0.530	0.570
Adjusted R^2	0.520	0.570
F Statistic	1,507.000*** (df = 6; 7934)	901.000*** (df = 12; 8114)

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

“Mundlak Device”: Testing

```
> # Testing:

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

Linear hypothesis test

Hypothesis:
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   8115
2   8114  1    26 0.00000033 ***
---
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 = 1-52, N = 8127

Residuals:
   Min. 1st Qu.  Median 3rd Qu.    Max.
-31.873  -4.069   0.247   4.142  43.134

Coefficients:
                     Estimate Std. Error t-value       Pr(>|t|)
PopGrowth             -0.2597     0.0798   -3.25         0.0011 **
UrbanPopulation        0.0240     0.0173    1.39         0.1655
FertilityRate          1.2307     0.1471    8.37        < 2e-16 ***
log(GDPPerCapita)      1.8678     0.2762    6.76 0.000000000014 ***
NaturalResourceRents   0.0289     0.0149    1.94         0.0523 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    344000
Residual Sum of Squares: 338000
R-Squared:      0.0165
Adj. R-Squared: -0.0137
F-statistic: 26.4366 on 5 and 7884 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 = 111, df1 = 236, df2 = 7884, 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 = 191, 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 = 231, 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 = 93, 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)

Residuals:
   Min     1Q Median     3Q    Max
-31.87  -4.07   0.25   4.14  43.13

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
PopGrowth             -0.2597     0.0798   -3.25  0.00114 **
UrbanPopulation        0.0240     0.0173    1.39  0.16547
FertilityRate          1.2307     0.1471    8.37  < 2e-16 ***
log(GDPPerCapita)      1.8678     0.2762    6.76  1.4e-11 ***
NaturalResourceRents   0.0289     0.0149    1.94  0.05225 .
factor(ISO3)AFG      -15.2358     2.4485   -6.22  5.1e-10 ***
factor(ISO3)AGO       14.0840     2.6724    5.27  1.4e-07 ***
factor(ISO3)ALB       36.8307     2.4462   15.06  < 2e-16 ***
.
.
.
factor(Year)1977       4.1587     0.8909    4.67  3.1e-06 ***
factor(Year)1978       5.0831     0.8914    5.70  1.2e-08 ***
 [ reached getOption("max.print") -- omitted 43 rows ]
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.5 on 7884 degrees of freedom
  (5633 observations deleted due to missingness)
Multiple R-squared:  0.99, Adjusted R-squared:  0.989
F-statistic: 3.11e+03 on 243 and 7884 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 = 1-52, N = 8127

Effects:
                 var std.dev share
idiosyncratic 42.930   6.552  0.31
individual    96.826   9.840  0.69
time           0.390   0.624  0.00
theta:
      Min. 1st Qu. Median Mean 3rd Qu. Max.
id    0.45    0.90   0.91 0.90    0.91 0.91
time  0.28    0.34   0.38 0.36    0.39 0.39
total 0.28    0.34   0.38 0.36    0.39 0.39

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -64.1    -6.9     2.2     0.3    11.0    33.1

Coefficients:
                      Estimate Std. Error z-value Pr(>|z|)
(Intercept)            5.74650    0.33690    17.1   <2e-16 ***
PopGrowth             -0.22894    0.01209   -18.9   <2e-16 ***
UrbanPopulation        0.12664    0.00239    52.9   <2e-16 ***
FertilityRate         -0.84844    0.02091   -40.6   <2e-16 ***
log(GDPPerCapita)      5.18252    0.03708   139.8   <2e-16 ***
NaturalResourceRents   0.03114    0.00219    14.2   <2e-16 ***
PostColdWar           11.87196    0.04770   248.9   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2930000
Residual Sum of Squares: 2040000
R-Squared:      0.321
Adj. R-Squared: 0.321
Chisq: 236462 on 6 DF, p-value: <2e-16
```

A Prettier Table

Table: Models of WBLI

	OLS	FE	BE	RE	TwoWayFE	TwoWayRE
Population Growth	-2.300^{***}	-0.110	-5.700^{***}	-0.200^{**}	-0.260^{***}	-0.230^{***}
	(0.130)	(0.095)	(1.000)	(0.097)	(0.080)	(0.012)
Urban Population	-0.057^{***}	0.300^{***}	-0.037	0.180^{***}	0.024	0.130^{***}
	(0.011)	(0.020)	(0.055)	(0.019)	(0.017)	(0.002)
Fertility Rate	-2.600^{***}	-2.000^{***}	-0.630	-2.300^{***}	1.200^{***}	-0.850^{***}
	(0.160)	(0.160)	(1.100)	(0.160)	(0.150)	(0.021)
ln(GDP Per Capita)	2.600^{***}	8.800^{***}	2.400^{**}	7.200^{***}	1.900^{***}	5.200^{***}
	(0.190)	(0.300)	(1.200)	(0.280)	(0.280)	(0.037)
Natural Resource Rents	-0.340^{***}	0.066^{***}	-0.330^{***}	0.036^{**}	0.029^{*}	0.031^{***}
	(0.016)	(0.017)	(0.091)	(0.017)	(0.015)	(0.002)
Post-Cold War	11.000^{***}	7.000^{***}	9.700^{*}	8.300^{***}		12.000^{***}
	(0.370)	(0.290)	(5.200)	(0.290)		(0.048)
Constant	50.000^{***}		50.000^{***}	-6.100^{**}		5.700^{***}
	(1.800)		(13.000)	(2.600)		(0.340)
Observations	8,127	8,127	187	8,127	8,127	8,127
R^2	0.500	0.530	0.580	0.520	0.016	0.320
Adjusted R^2	0.500	0.520	0.570	0.520	-0.014	0.320

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

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)	49.621		-6.137
	(1.835)		(2.560)
PopGrowth	-2.334	-0.106	-0.199
	(0.131)	(0.095)	(0.097)
UrbanPopulation	-0.057	0.296	0.181
	(0.011)	(0.020)	(0.019)
FertilityRate	-2.561	-2.006	-2.315
	(0.159)	(0.162)	(0.161)
log(GDPPerCapita)	2.640	8.833	7.237
	(0.194)	(0.300)	(0.284)
NaturalResourceRents	-0.343	0.066	0.036
	(0.016)	(0.017)	(0.017)
PostColdWar	10.806	7.029	8.294
	(0.371)	(0.294)	(0.289)
Num.Obs.	8127	8127	8127
R2	0.499	0.533	0.516
R2 Adj.	0.499	0.521	0.516
AIC	65298.6	56459.4	57066.1
BIC	65354.6	56508.4	57122.1
RMSE	13.43	7.80	8.09

Using modelsummary (continued)

A better-looking table:

```
> modelsummary(models,output="MS-Table-24.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"))
```

Models of WDBI

	OLS	Within	Random
(Intercept)	49.62***		-6.14*
	(1.83)		(2.56)
Population Growth	-2.33***	-0.11	-0.20*
	(0.13)	(0.10)	(0.10)
Urban Population	-0.06***	0.30***	0.18***
	(0.01)	(0.02)	(0.02)
Fertility Rate	-2.56***	-2.01***	-2.32***
	(0.16)	(0.16)	(0.16)
ln(GDP Per Capita)	2.64***	8.83***	7.24***
	(0.19)	(0.30)	(0.28)
Natural Resource Rents	-0.34***	0.07***	0.04*
	(0.02)	(0.02)	(0.02)
Post-Cold War	10.81***	7.03***	8.29***
	(0.37)	(0.29)	(0.29)
Num.Obs.	8127	8127	8127
R2	0.499	0.533	0.516
R2 Adj.	0.499	0.521	0.516

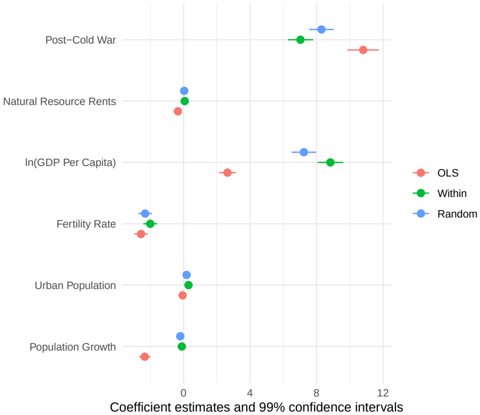
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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