PLSC 504: Fall 2024
Panel Data for Non-Continuous Responses + Causal Inference

October 16, 2024

Binary Y + Unit Effects ("Fixed")

One-way unit effects:

$$Y_{it} = f(\mathbf{X}_{it}\beta + \alpha_i + u_{it})$$

So, think about logit first:

$$\Pr(Y_{it} = 1) = \frac{\exp(\mathbf{X}_{it}\beta + \alpha_i)}{1 + \exp(\mathbf{X}_{it}\beta + \alpha_i)} \equiv \Lambda(\mathbf{X}_{it}\beta + \alpha_i)$$

Incidental Parameters:

- Nonlinearity $\rightarrow$ inconsistency in both $\hat{\alpha}$s *and* $\hat{\boldsymbol{\beta}}$.
- Anderson's *unconditional* estimator:

$$L^U = \prod_{i=1}^{N}\prod_{t=1}^{T} \Lambda(\mathbf{X}_{it} + \alpha_i)^{Y_{it}}[1 - \Lambda(\mathbf{X}_{it} + \alpha_i)]^{1-Y_{it}}$$

- Chamberlain's *conditional* estimator:

$$L^C = \prod_{i=1}^{N} \Pr\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ...Y_{iT} = y_{iT} \mid \sum_{t=1}^{T} Y_{it}\right)$$

Fixed-Effects (continued)

Intuition: Suppose we have $T = 2$. That means that:

- $\Pr(Y_{i1} = 0 \text{ and } Y_{i2} = 0 \mid \sum_T Y_{it} = 0) = 1.0$
- $\Pr(Y_{i1} = 1 \text{ and } Y_{i2} = 1 \mid \sum_T Y_{it} = 2) = 1.0$

and:

$$\Pr\left(Y_{i1} = 0 \text{ and } Y_{i2} = 1 \mid \sum_T Y_{it} = 1\right) = \frac{\Pr(0,1)}{\Pr(0,1) + \Pr(1,0)}$$

with a similar statement for $\Pr(Y_{i1} = 1 \text{ and } Y_{i2} = 0 \mid \sum_T Y_{it} = 1)$.

<u>The Point</u>:

$\sum_{t=1}^{T} Y_{it}$ **is a sufficient statistic for** α_i, so conditioning on it $\equiv$ "fixed effects."

Notes On Fixed-Effects

Things to bear in mind:

- Fixed effects = no estimates for $\boldsymbol{\beta}_b$
- Interpretation: per logit, but $\mid \hat{\alpha}_i$.
- Everything above is for **logit**...
  - For FE probit, there is no conditional model
  - Unconditional / "brute force" FE probit is biased (see here and here)
- BTSCS in international relations: Green et al. (2001) v. B&K (2001) ("Dirty Pool" debate)

Model is:

$$\begin{aligned} Y_{it}^* &= \mathbf{X}_{it}\beta + u_{it} \\ Y_{it} &= 0 \text{ if } Y_{it}^* \leq 0 \text{ ;} \\ &= 1 \text{ if } Y_{it}^* > 0 \end{aligned}$$

with:

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

with $\eta_{it} \sim$ i.i.d. N(0,1), and $\alpha_i \sim N(0, \sigma_\alpha^2)$. This implies:

$$\text{Var}(u_{it}) = 1 + \sigma_\alpha^2$$

and so:

$$\text{Corr}(u_{it}, u_{is},\ t \neq s) \equiv \rho = \frac{\sigma_\alpha^2}{1 + \sigma_\alpha^2}$$

which means that we can write $\sigma_\alpha^2 = \left(\frac{\rho}{1-\rho}\right)$.

Random Effects Variants

Probit:

$$\begin{aligned} L_i &= \text{Prob}(Y_{i1} = y_{i1},\ Y_{i2} = y_{i2},\ ...Y_{iT} = y_{iT}) \\ &= \int_{-\infty}^{X_{i1}\beta} \int_{-\infty}^{X_{i2}\beta} ... \int_{-\infty}^{X_{iT}\beta} \phi(u_{i1}, u_{i2}...u_{iT}) du_{iT}...du_{i2}\ du_{i1} \end{aligned}$$

Logit:

$$\begin{aligned} L_i &= \text{Prob}(Y_{i1} = y_{i1},\ Y_{i2} = y_{i2},\ ...Y_{iT} = y_{iT}) \\ &= \int_{-\infty}^{X_{i1}\beta} \int_{-\infty}^{X_{i2}\beta} ... \int_{-\infty}^{X_{iT}\beta} \lambda(u_{i1}, u_{i2}...u_{iT}) du_{iT}...du_{i2}\ du_{i1} \end{aligned}$$

Solution?

$$\phi(u_{i1}, u_{i2}, ...u_{iT}) = \int_{-\infty}^{\infty} \phi(u_{i1}, u_{i2}, ...u_{iT} \mid \alpha_i)\phi(\alpha_i) d\alpha_i$$

- $\hat{\rho} =$ proportion of the variance due to the α_is.
- Implementation: Gauss-Hermite quadrature or MCMC.
- Best with N large and T small.
- Critically requires $\text{Cov}(\mathbf{X}, \alpha) = 0$ (see notes re: Chamberlain's CRE Estimator).

Example: WDI "Plus"

Data from the WDI plus POLITY and the UCDP:

- `ISO3` - The country's International Standards Organization (ISO) three-letter identification code.
- `Year` - The year that row of data applies to.
- `CivilWar` - Civil conflict indicator: 1 if there was a civil conflict in that country in that year; 0 otherwise. From UCDP.
- `OnsetCount` - The sum of new conflict episodes in that country / year. From UCDP.
- `LandArea` - Land area (sq. km).
- `PopMillions` - Popluation (in millions).
- `PopGrowth` - Population Growth (percent).
- `UrbanPopulation` - Urban Population (percent of total).
- `GDPPerCapita` - GDP per capita (constant 2010 $US).
- `GDPPerCapGrowth` - GDP Per Capita Growth (percent annual).
- `PostColdWar` - 1 if `Year` $>$ 1989, 0 otherwise.
- `POLITY` - The POLITY score of democracy/autocracy. Scaled so that 0 = most autocratic, 10 = most democratic.

$N = 216$, $\bar{T} = 61$, NT varies (due to missingness).

```
> describe(DF,skew=FALSE)

                vars     n      mean         sd    median    min        max      range       se
ISO3*              1 13822    108.49      62.35    108.00   1.00      216.0     215.00     0.53
Year*              2 13822     32.50      18.47     32.00   1.00       64.0      63.00     0.16
country*           3 13760    108.00      62.07    108.00   1.00      215.0     214.00     0.53
CivilWar           4  9052      0.13       0.34      0.00   0.00        1.0       1.00     0.00
OnsetCount         5  9394      0.05       0.24      0.00   0.00        4.0       4.00     0.00
LandArea           6 11941 605302.93 1639812.91 107160.00   2.03 16389950.0 16389947.97 15006.31
PopMillions        7 13730     25.30     105.45      4.31   0.00     1428.6    1428.63     0.90
UrbanPopulation    8 13696     51.89      25.74     51.17   2.08      100.0      97.92     0.22
GDPPerCapita       9 11021  12146.44   18971.36   3927.71 122.52   228667.9  228545.42   180.71
GDPPerCapGrowth   10 10942      1.91       6.68      2.08 -64.43      150.4     214.86     0.06
PostColdWar       11 13760      0.53       0.50      1.00   0.00        1.0       1.00     0.00
POLITY            12  8279      5.55       3.71      6.50   0.00       10.0      10.00     0.04
POLITYSquared     13  8279     44.57      40.24     42.25   0.00      100.0     100.00     0.44


> # Make panel data:
>
> DF<-pdata.frame(DF,index=c("ISO3","Year"))

> # Variation in civil wars:
>
> summary(DF$CivilWar)
total sum of squares: 1051
      id     time
0.438820 0.009196

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
      0       0       0       0       0       1    4770
```

Pooled Logit

```
> Logit<-glm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+               GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared,data=DF,family="binomial")


> summary(Logit)

Call:
glm(formula = CivilWar ~ log(LandArea) + log(PopMillions) + UrbanPopulation +
    log(GDPPerCapita) + GDPPerCapGrowth + PostColdWar + POLITY +
    POLITYSquared, family = "binomial", data = DF)

Coefficients:
                  Estimate Std. Error z value     Pr(>|z|)
(Intercept)        0.26556    0.50133    0.53         0.60
log(LandArea)     -0.04554    0.03019   -1.51         0.13
log(PopMillions)   0.66233    0.03531   18.76      < 2e-16 ***
UrbanPopulation    0.01898    0.00317    5.99 0.0000000021 ***
log(GDPPerCapita) -0.60750    0.05810  -10.46      < 2e-16 ***
GDPPerCapGrowth   -0.03317    0.00593   -5.59 0.0000000226 ***
PostColdWar       -0.45766    0.08227   -5.56 0.0000000266 ***
POLITY             0.65350    0.05857   11.16      < 2e-16 ***
POLITYSquared     -0.06219    0.00553  -11.25      < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6284.3  on 7443  degrees of freedom
Residual deviance: 5056.3  on 7435  degrees of freedom
  (6378 observations deleted due to missingness)
AIC: 5074

Number of Fisher Scoring iterations: 6
```

```
> FELogit<-bife(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+                  GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared|ISO3,data=DF,model="logit")


> summary(FELogit)
binomial - logit link

CivilWar ~ log(LandArea) + log(PopMillions) + UrbanPopulation +
    log(GDPPerCapita) + GDPPerCapGrowth + PostColdWar + POLITY +
    POLITYSquared | ISO3

Estimates:
                   Estimate Std. error z value Pr(> |z|)
log(LandArea)     -13.43912    8.10323   -1.66     0.097 .
log(PopMillions)    0.61643    0.28027    2.20     0.028 *
UrbanPopulation     0.01777    0.01188    1.50     0.135
log(GDPPerCapita)  -0.40990    0.16248   -2.52     0.012 *
GDPPerCapGrowth    -0.04877    0.00782   -6.23   4.5e-10 ***
PostColdWar        -0.18605    0.17221   -1.08     0.280
POLITY              0.69588    0.09072    7.67   1.7e-14 ***
POLITYSquared      -0.07239    0.00862   -8.40   < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

residual deviance= 2997,
null deviance= 4769,
n= 4252, N= 87

( 6378 observation(s) deleted due to missingness )
( 3192 observation(s) deleted due to perfect classification )

Number of Fisher Scoring Iterations: 6

Average individual fixed effect= 167.7
```

```
> RELogit<-pglm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+                 GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared|ISO3,data=DF,family=binomial,
+               effect="individual",model="random")

> summary(RELogit)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 1 iterations
Return code 3: Last step could not find a value above the current.
Boundary of parameter space?
Consider switching to a more robust optimisation method temporarily.
Log-Likelihood: -1766
10  free parameters
Estimates:
                  Estimate Std. error t value Pr(> t)
(Intercept)        0.26556        NaN     NaN     NaN
log(LandArea)     -0.04554    0.09007   -0.51   0.613
log(PopMillions)   0.66233    0.09330    7.10 1.3e-12 ***
UrbanPopulation    0.01898    0.00771    2.46   0.014 *
log(GDPPerCapita) -0.60750    0.14278   -4.25 2.1e-05 ***
GDPPerCapGrowth   -0.03317    0.00730   -4.54 5.5e-06 ***
PostColdWar       -0.45766    0.11626   -3.94 8.3e-05 ***
POLITY             0.65350    0.08104    8.06 7.4e-16 ***
POLITYSquared     -0.06219    0.00732   -8.50 < 2e-16 ***
sigma              2.53151    0.17860   14.17 < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
```

Nice Table...

Models of Civil War

	Logit	FE Logit	RE Logit
Intercept	0.27		0.27
	(0.50)		
ln(Land Area)	-0.05	-13.44	-0.05
	(0.03)	(8.10)	(0.09)
ln(Population)	0.66^*	0.62^*	0.66^*
	(0.04)	(0.28)	(0.09)
Urban Population	0.02^*	0.02	0.02^*
	(0.00)	(0.01)	(0.01)
ln(GDP Per Capita)	-0.61^*	-0.41^*	-0.61^*
	(0.06)	(0.16)	(0.14)
GDP Growth	-0.03^*	-0.05^*	-0.03^*
	(0.01)	(0.01)	(0.01)
Post-Cold War	-0.46^*	-0.19	-0.46^*
	(0.08)	(0.17)	(0.12)
POLITY	0.65^*	0.70^*	0.65^*
	(0.06)	(0.09)	(0.08)
POLITY Squared	-0.06^*	-0.07^*	-0.06^*
	(0.01)	(0.01)	(0.01)
Estimated Sigma			2.53^*
			(0.18)
AIC	5074.27		3552.10
BIC	5136.51		
Log Likelihood	-2528.14	-1498.49	-1766.05
Deviance	5056.27	2996.98	
Num. obs.	7444	4252	

$^*p < 0.05$

Wrap-Up: Some Useful Packages

- `pglm`
  - Workhorse package for panel (FE, RE, BE) GLMs
  - Binary + ordered logit/probit, Poisson / negative binomial
  - Discussed + used extensively in Croissant and Millo (2018) *Panel Data Econometrics with R*
  - The one thing it won't (apparently) do is fixed-effects, binary-response models...
- `fixest`
  - Fast / efficient fitting of FE models
  - Fits linear models, logit, Poisson, and negative binomial
  - Includes easy coefficient plots & tables; simple multi-threading; built-in "robust" S.E.s
- `alpaca`
  - Fast / efficient fitting of GLMs with high-dimensional fixed effects
  - *Includes bias correction for incidental parameters after binary-response models*
  - Also includes useful panel data simulation routines + average partial effects

GEEs

Quick GLM review

Linear-normal model is:

$$Y_i = \mu_i + u_i$$

with:

$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}.$$

Generalize:

$$g(\mu_i) = \mathbf{X}_i \boldsymbol{\beta}$$

and:

$$Y_i \sim \text{i.i.d.}\, F[\mu_i, \mathbf{V}_i].$$

"Score" equations:

$$\mathbf{U}(\beta) = \sum_{i=1}^{N} \mathbf{D}_i' \mathbf{V}_i^{-1} [Y_i - \mu_i] = \mathbf{0}.$$

with:

- $\mathbf{D}_i = \frac{\partial \mu_i}{\partial \beta}$,
- $\mathbf{V}_i = \frac{h(\mu_i)}{\phi}$, and
- $(Y_i - \mu_i) \approx$ a "residual."
- Known as "quasi-likelihood" (e.g. Wedderburn 1974 *Biometrika*).

Now suppose:

$$Y_{it} = \mu_{it} + u_{it}$$

where

- $i \in \{1, ...N\}$ are i.i.d. "units,"
- $t \in \{1, ...T\}$, $T > 1$ are "time points,"
- we want $g(\mu_{it}) = \mathbf{X}_{it}\boldsymbol{\beta}$.

Key issue: Accounting for (conditional) dependence in Y over time.

Full joint distributions over T are hard. But...

Define:

$$\underset{T \times T}{\mathbf{R}_i(\boldsymbol{\alpha})} = \begin{pmatrix} 1.0 & \alpha_{12} & \cdots & \alpha_{1,T} \\ \alpha_{21} & 1.0 & \cdots & \alpha_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{T,1} & \cdots & \alpha_{T,T-1} & 1.0 \end{pmatrix},$$

$\rightarrow$ "working correlation" matrix.

- Completely defined by $\boldsymbol{\alpha}$,
- Structure specified by the analyst.

Liang and Zeger (1986): We can decompose the variance of Y_{it} as:

$$\mathbf{V}_i = \mathrm{diag}(\mathbf{V}_i^{\frac{1}{2}})\,\mathbf{R}_i(\boldsymbol{\alpha})\,\mathrm{diag}(\mathbf{V}_i^{\frac{1}{2}})$$

With a standard GLM assumption about the mean and variance, this is:

$$\mathbf{V}_i = \frac{(\mathbf{A}_i^{\frac{1}{2}})\,\mathbf{R}_i(\boldsymbol{\alpha})\,(\mathbf{A}_i^{\frac{1}{2}})}{\phi}$$

where

$$\underset{T\times T}{\mathbf{A}_i} = \begin{pmatrix} h(\mu_{i1}) & 0 & \cdots & 0 \\ 0 & h(\mu_{i2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & h(\mu_{iT}) \end{pmatrix}$$

What does that mean?

$\mathbf{V}_i = \text{Var}(Y_{it}|\mathbf{X}_{it}, \beta)$ has two parts:

- $\mathbf{A}_i =$ *unit-level* variation,
- $\mathbf{R}_i(\alpha) =$ within-unit *temporal* variation.

Specifying $\mathbf{R}_i(\boldsymbol{\alpha})$

Independent: $$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0 & 0 & \cdots & 0 \\ 0 & 1.0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1.0 \end{pmatrix}$$

- Assumes no within-unit temporal correlation.
- Equivalent to GLM on pooled data.

Exchangeable: $$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0 & \alpha & \cdots & \alpha \\ \alpha & 1.0 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \cdots & \alpha & 1.0 \end{pmatrix}$$

- One free parameter in $\mathbf{R}_i(\boldsymbol{\alpha})$ ($\alpha_{ts} = \alpha \; \forall \; t \neq s$)
- Temporal correlation within units is constant across time points.
- Akin (in some respects) to a random-effects model...

Specifying $\mathbf{R}_i(\boldsymbol{\alpha})$

AR(p) (e.g., AR(1)): $$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0 & \alpha & \alpha^2 & \cdots & \alpha^{T-1} \\ \alpha & 1.0 & \alpha & \cdots & \alpha^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha^{T-1} & \cdots & \alpha^2 & \alpha & 1.0 \end{pmatrix}$$

- One free parameter in $\mathbf{R}_i(\boldsymbol{\alpha})$ ($\alpha_{ts} = \alpha^{|t-s|} \; \forall \; t \neq s$).
- Conditional within-unit correlation an exponential function of the lag.

Stationary(p): $$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0 & \alpha_1 & \cdots & \alpha_p & 0 & \cdots & 0 \\ \alpha_1 & 1.0 & \alpha_1 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \alpha_p & \cdots & \alpha_1 & 1.0 \end{pmatrix}$$

- AKA "banded," or "*p*-dependent."
- $p \leq T - 1$ free parameters in $\mathbf{R}_i(\boldsymbol{\alpha})$.
- Conditional within-unit correlation an exponential function of the lag, up to lag p, and zero thereafter.

Specifying $\mathbf{R}_i(\boldsymbol{\alpha})$

Unstructured: $$\mathbf{R}_i(\boldsymbol{\alpha}) = \begin{pmatrix} 1.0 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1,T-1} \\ \alpha_{12} & 1.0 & \alpha_{23} & \cdots & \alpha_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_{1,T-1} & \alpha_{2,T-1} & \cdots & \alpha_{T-1,T-1} & 1.0 \end{pmatrix}$$

- $\frac{T(T-1)}{2}$ free parameters in $\mathbf{R}_i(\boldsymbol{\alpha})$.
- Conditional within-unit correlation is completely data-dependent.

Score equations:

$$\boldsymbol{U}_{GEE}(\boldsymbol{\beta}_{GEE}) = \sum_{i=1}^{N} \mathbf{D}_i' \left[\frac{(\mathbf{A}_i^{\frac{1}{2}})\,\mathbf{R}_i(\boldsymbol{\alpha})\,(\mathbf{A}_i^{\frac{1}{2}})}{\phi} \right]^{-1} [Y_i - \mu_i] = \mathbf{0}$$

Two-step estimation:

- For fixed values of $\boldsymbol{\alpha}_s$ and ϕ_s at iteration s, use Newton scoring to estimate $\hat{\boldsymbol{\beta}}_s$,
- Use $\hat{\boldsymbol{\beta}}_s$ to calculate standardized residuals $(Y_i - \hat{\mu}_i)_s$, from which consistent estimates of $\boldsymbol{\alpha}_{s+1}$ and ϕ_{s+1} can be estimated.

Liang & Zeger (1986):

$$\hat{\boldsymbol{\beta}}_{GEE} \underset{N\rightarrow\infty}{\sim} \boldsymbol{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma}).$$

For $\hat{\boldsymbol{\Sigma}}$, two options:

$$\hat{\boldsymbol{\Sigma}}_{\text{Model}} = N\left(\sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)$$

$$\hat{\boldsymbol{\Sigma}}_{\text{Robust}} = N\left(\sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1} \left(\sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{S}}_i \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right) \left(\sum_{i=1}^{N} \hat{\boldsymbol{D}}_i' \hat{\boldsymbol{V}}_i^{-1} \hat{\boldsymbol{D}}_i\right)^{-1}$$

where $\hat{\boldsymbol{S}}_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$.

Inference (aka, magic!)

- $\hat{\Sigma}_{\text{Model}}$
  - Requires that $\mathbf{R}_i(\boldsymbol{\alpha})$ be "correct" for consistency.
  - Is slightly more efficient than $\hat{\boldsymbol{\Sigma}}_{\text{Robust}}$ if so.

- $\hat{\Sigma}_{\text{Robust}}$
  - Is consistent *even if* $\mathbf{R}_i(\boldsymbol{\alpha})$ *is misspecified*.
  - Is slightly less efficient than $\hat{\boldsymbol{\Sigma}}_{\text{Model}}$ if $\mathbf{R}_i(\boldsymbol{\alpha})$ is correct.

Moral: Use $\hat{\Sigma}_{\textbf{Robust}}$.

Summary

GEEs:

- Are a straightforward variation on GLMs, and so
- Can be applied to a range of data types (continuous, binary, count, proportions, etc.),
- Yield robustly consistent point estimates of βs,
- Account for within-unit correlation in an informed way, but also
- Yield consistent inferences even if that correlation is misspecified.

Practical Issues: Model Interpretation

- In general, GEEs = GLMs.
- GEEs are *marginal* models, so:
  - $\hat{\beta}$s have an interpretation as *average* / total effects.
  - Estimates / effect sizes generally be smaller than conditional (e.g. fixed/random) effects models.
  - E.g., for logit, $\hat{\boldsymbol{\beta}}_M \approx \frac{\hat{\boldsymbol{\beta}}_C}{\sqrt{1+0.35\sigma^2_\eta}}$, where $\sigma^2_\eta > 0$ is the variance of the unit effects.

Practical Issues: Specifying $\mathbf{R}_i(\boldsymbol{\alpha})$

- Has been called "more art than science."
- Pointers:
  - Choose based on *substance* of the problem.
  - Remember that $\mathbf{R}_i(\boldsymbol{\alpha})$ is conditional on $\mathbf{X}$, $\hat{\boldsymbol{\beta}}$.
  - Consider unstructured when T is small and N large.
  - Try different ones, and compare.
- In general, it shouldn't matter terribly much...

GEEs: Software

Software	Command(s)/Package(s)
R	`gee` / `geepack` / `geeM` / `multgeeB` / `orth` / `repolr`
Stata	`xtgee` / `xtlogit` / `xtprobit` / `xtpois` / etc.
SAS	`genmod` (w/ `repeated`)

- Generally follow GLMs (specify "family" + "link")
- Certain combinations not possible/recommended
- Estimation: Fisher scoring, MLE, etc. (MCMC?)

From the `geepack` manual:

Warning

Use "unstructured" correlation structure only with great care. (It may cause R to crash).

Civil Wars (2013-17)... GEE: Independence

```
> GEE.ind<-geeglm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+
+                  log(GDPPerCapita)+GDPPerCapGrowth+POLITY+POLITYSquared,
+                  data=DF5,id=ISO3,family="binomial",corstr="independence")

> summary(GEE.ind)

Call:
geeglm(formula = CivilWar ~ log(LandArea) + log(PopMillions) +
    UrbanPopulation + log(GDPPerCapita) + GDPPerCapGrowth + POLITY +
    POLITYSquared, family = "binomial", data = DF5, id = ISO3,
    corstr = "independence")

 Coefficients:
                  Estimate Std.err  Wald Pr(>|W|)
(Intercept)        -3.9595  3.9226  1.02  0.31278
log(LandArea)       0.1139  0.2104  0.29  0.58840
log(PopMillions)    0.9252  0.2528 13.39  0.00025 ***
UrbanPopulation     0.0218  0.0172  1.61  0.20457
log(GDPPerCapita)  -0.5621  0.3208  3.07  0.07970 .
GDPPerCapGrowth    -0.0583  0.0331  3.10  0.07816 .
POLITY              0.9122  0.5752  2.51  0.11278
POLITYSquared      -0.0963  0.0494  3.81  0.05096 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = independence
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.681   0.822
Number of clusters:   159  Maximum cluster size: 5
```

GEE: Exchangeable

```
> GEE.exc<-geeglm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+
+                     log(GDPPerCapita)+GDPPerCapGrowth+POLITY+POLITYSquared,
+                   data=DF5,id=ISO3,family="binomial",corstr="exchangeable")

> summary(GEE.exc)

Call:
geeglm(formula = CivilWar ~ log(LandArea) + log(PopMillions) +
    UrbanPopulation + log(GDPPerCapita) + GDPPerCapGrowth + POLITY +
    POLITYSquared, family = "binomial", data = DF5, id = ISO3,
    corstr = "exchangeable")

 Coefficients:
                  Estimate Std.err  Wald Pr(>|W|)
(Intercept)        -1.9079  3.3163  0.33  0.56509
log(LandArea)       0.1264  0.2156  0.34  0.55762
log(PopMillions)    0.8859  0.2439 13.20  0.00028 ***
UrbanPopulation     0.0280  0.0178  2.46  0.11674
log(GDPPerCapita)  -0.7840  0.2868  7.47  0.00627 **
GDPPerCapGrowth    -0.0302  0.0353  0.73  0.39161
POLITY              0.4811  0.4439  1.17  0.27841
POLITYSquared      -0.0542  0.0383  2.01  0.15675
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.689    1.08
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.448   0.696
Number of clusters:   159  Maximum cluster size: 5
```

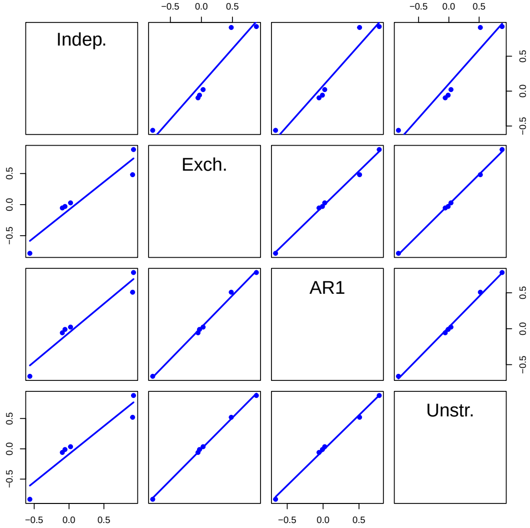
GEE: AR(1)

```
> GEE.ar1<-geeglm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+
+                   log(GDPPerCapita)+GDPPerCapGrowth+POLITY+POLITYSquared,
+                   data=DF5,id=ISO3,family="binomial",corstr="ar1")

> summary(GEE.ar1)

Call:
geeglm(formula = CivilWar ~ log(LandArea) + log(PopMillions) +
    UrbanPopulation + log(GDPPerCapita) + GDPPerCapGrowth + POLITY +
    POLITYSquared, family = "binomial", data = DF5, id = ISO3,
    corstr = "ar1")

 Coefficients:
                  Estimate  Std.err Wald Pr(>|W|)
(Intercept)       -3.53500  3.67327 0.93   0.3359
log(LandArea)      0.21779  0.23029 0.89   0.3443
log(PopMillions)   0.78204  0.26231 8.89   0.0029 **
UrbanPopulation    0.02267  0.01735 1.71   0.1912
log(GDPPerCapita) -0.66199  0.29178 5.15   0.0233 *
GDPPerCapGrowth   -0.00973  0.04749 0.04   0.8377
POLITY             0.50808  0.50065 1.03   0.3102
POLITYSquared     -0.05677  0.04463 1.62   0.2034
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = ar1
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.724    1.46
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.651    0.75
Number of clusters:   159  Maximum cluster size: 5
```

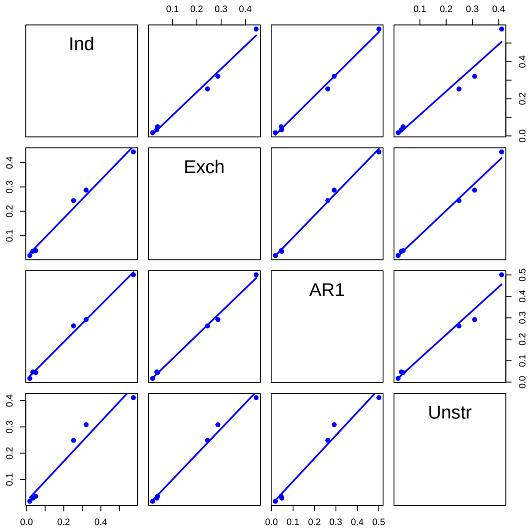
GEE: Unstructured (2013-2017)

```
> GEE.unstr<-geeglm(CivilWar~log(LandArea)+log(PopMillions)+UrbanPopulation+
+                    log(GDPPerCapita)+GDPPerCapGrowth+POLITY+POLITYSquared,
+                    data=DF5,id=ISO3,family="binomial",corstr="unstructured")
> summary(GEE.unstr)

Call:
geeglm(formula = CivilWar ~ log(LandArea) + log(PopMillions) +
    UrbanPopulation + log(GDPPerCapita) + GDPPerCapGrowth + POLITY +
    POLITYSquared, family = "binomial", data = DF5, id = ISO3,
    corstr = "unstructured")

 Coefficients:
                  Estimate Std.err  Wald Pr(>|W|)
(Intercept)        -2.1234  3.1797  0.45  0.50426
log(LandArea)       0.1467  0.1932  0.58  0.44766
log(PopMillions)    0.8796  0.2488 12.50  0.00041 ***
UrbanPopulation     0.0351  0.0171  4.20  0.04053 *
log(GDPPerCapita)  -0.8326  0.3084  7.29  0.00694 **
GDPPerCapGrowth    -0.0116  0.0294  0.15  0.69408
POLITY              0.5202  0.4112  1.60  0.20585
POLITYSquared      -0.0598  0.0362  2.73  0.09830 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.672    0.79
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.392   0.481
alpha.1:3    0.411   0.504
alpha.1:4    0.344   0.427
alpha.1:5    0.334   0.412
alpha.2:3    0.733   0.833
alpha.2:4    0.279   0.350
alpha.2:5    0.503   0.574
alpha.3:4    0.399   0.501
alpha.3:5    0.738   0.851
alpha.4:5    0.436   0.540
Number of clusters:   159  Maximum cluster size: 5
```

Comparing $\hat{\beta}$s



Comparing $\widehat{\text{s.e.s}}$



GEEs: Wrap-Up

GEEs are:

- Robust
- Flexible
- Extensible beyond panel/TSCS context

Causal Inference

Things We Can Always Do

Whether or not we have panel data, we can use:

- Regression!
- Matching...
- Instrumental variables...
- Regression discontinuity (RDD)...

Panel Data Approaches: Differences-In-Differences

Simple "DiD":

- Leverages two-group, two-period data ($T = 2$):

	Pre-Treatment ($T = 0$)	Post-Treatment ($T = 1$)
Treated ($W = 1$)	A	B
Untreated ($W = 0$)	C	D

- Process (simple version):
  - Calculate pre- vs. post-treatment difference for the treated group ($B - A$)
  - Calculate pre- vs. post-treatment difference for the untreated group ($D - C$)
  - Calculate the differences between the differences [$\widehat{\text{DiD}} = (B - A) - (D - C)$]

Differences-In-Differences (continued)

Simple DiD is the same as fitting the regression:

$$Y_{it} = \beta_0 + \beta_1 W_{it} + \beta_2 T_{it} + \beta_3 W_{it} T_{it} + u_{it}$$

- Validity depends on (a) all the usual assumptions required by OLS, plus (b) the *parallel trends assumption* – that there are no (non-treatment-related) time-varying differences between the two groups as we go from $T = 0$ to $T = 1$.
- Resources:
  - Our old friend Wikipedia
  - Pischke's slides on DiD
  - R: package `did`
  - Stata: `ieddtab` in the `ietoolkit`

Panel Data Approaches: Synthetic Controls

The "synthetic control method" (SCM):

- Addresses situations in which we have a *single* treated case (or small number of them)...
- Requires at least one (and ideally more) repeated measurements over time on the outcome of interest, and
- Also requires multiple (but not *too* many) non-treated cases
- Assumptions:
  - Possible control units are similar
  - Lack of spillover between treated and potential control units
  - Lack of exogenous shocks to potential control units

Synthetic Controls (continued)

SCM details:

- Intuition:
  - Create a counterfactual "control" unit that is as similar to the (pre-treatment) treated case as possible
  - Do so by weighting the observed predictors across "control" cases to minimize the difference (in a MSE sense)
  - Compare the pre-treatment trends in the synthetic control and treated cases
  - The weights are then used to create a post-treatment trend for the synthetic control
  - Inference is via placebo methods (varying the timing of the intervention)
- Advantages:
  - Works with (very) small N
  - Doesn't require parallel trends (a la DiD)
  - Abadie et al. claim that SCM controls for both observed and unobserved time-varying confounders
- A few references:
  - A nontechnical introduction in the *BMJ*
  - Method of the Month Blog
  - The Development Impact blog post on SCM

Panel Data Approaches: Unit Effects

Recall the two-way unit-effects model:

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

In this model:

- α_i captures *all* non-time-varying, unit-specific confounders
- η_t captures *all* non-unit-varying, period-specific confounders

(Some like to claim that) This makes unit-effects models a powerful tool for causal inference with panel data...

Let's first think about *one-way* unit- (fixed-)effects models, a la:

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

Causal Inference: One-Way (FE) Models

Imai and Kim (2019):

- The punch line first: "(t)he ability of unit fixed effects regression models to adjust for unobserved time-invariant confounders comes at the expense of dynamic causal relationships between treatment and outcome variables."
- Also dependent on functional form assumptions (specifically, linearity)

Intuition: For the model:

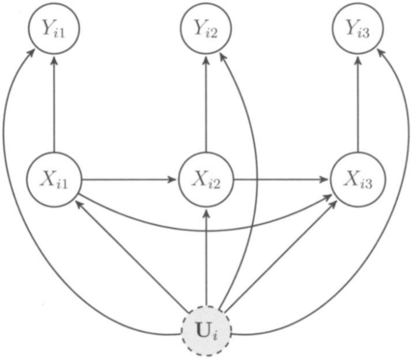
$$Y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$

where (for simplicity) X is a binary treatment for which we want to know a causal effect on Y:

- Identification is via $\text{Cov}[(\mathbf{X}_{it}, \alpha_i), u_{it}] = 0$
- In this framework, $\boldsymbol{\beta} = \tau$, the typical causal estimand (that is, the expected difference between $Y_{it}(0)$ and $Y_{it}(1)$)

A more flexible approach is to think of a FE model as a DAG...

Fixed-Effects DAG



Source: Imai and Kim (2019)

Key FE Takeaways

Summarizing Imai and Kim (2019):

- Three key identifying assumptions for FE models:
  - No unobserved time-varying confounders
  - Past treatments / values of **X** do not affect current values of Y[1]
  - Past outcomes Y do not affect current values of **X**.
- Alternatively, one can select on observables (a la Blackwell and Glynn 2018) and model dynamics (albeit at the cost of failing to control for unobserved time-constant confounders).

"...researchers must choose either to adjust for unobserved time-invariant confounders through unit fixed effects models or to model dynamic causal relationships between treatment and outcome under a selection-on-observables approach. No existing method can achieve both objectives without additional assumptions" (Imai and Kim 2019, 484).

[1] Can be relaxed via IV, but that requires independence of past and present values of Y.

Imai and Kim redux (2020):

- In the simple $T = 2$ case, DiD is equivalent to a two-way FE model:

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

- Imai & Kim: The same is not true for $T > 2$...
- More important: two-way FEs' ability to control for unmeasured confounders depends on the (linearity of the) functional form...
- Upshot: two-way FEs aren't a (nonparametric) cure-all...
- Related: When we control for both α_i and η_t, what – exactly – is the counterfactual?

Software Matters

In general:

- R
  - RDD has `rddtools`, `rdd`, etc.
  - DiD has `did`, `DRDID`, `bacondecomp`...
  - IV regression: `ivreg` (in `AER`), `tsls` (in `sem`), `ivmodel`, others...
  - Synthetic controls are in `Synth`, `MicroSynth`, `gsynth`, `tidysynth`...
  - See generally the CRAN Task View on *Causal Inference*.
- Stata also has a large suite of routines for attempting causal inference with observational data...
- And there's a pretty good NumPy/SciPy-dependent package for Python, called (creatively) Causalinference

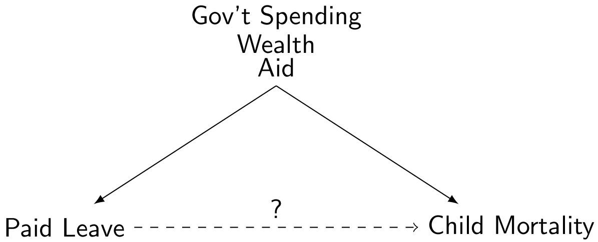
Back To The WDI

```
> 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             NaN               NA    NA       NA            NA
Region                     3 13760             NaN               NA    NA       NA            NA
country                    4 13760             NaN               NA    NA       NA            NA
iso3c                      5 13760             NaN               NA    NA       NA            NA
RuralPopulation            6 13696           48.11            25.74 -0.11    -1.00          0.22
UrbanPopulation            7 13696           51.89            25.74  0.11    -1.00          0.22
BirthRatePer1K             8 13150           27.86            13.10  0.23    -1.24          0.11
FertilityRate              9 12987            3.88             2.00  0.40    -1.21          0.02
PrimarySchoolAge          10 11119            6.13             0.61 -0.04     0.11          0.01
LifeExpectancy            11 12974           64.75            11.28 -0.73    -0.01          0.10
AgeDepRatioOld            12 13730           10.78             7.15  1.76     4.66          0.06
ChildMortality            13 11372           73.03            76.54  1.49     1.81          0.72
GDP                       14 11016 240364127688.72 1129708390544.55 11.46   159.31 10763521700.11
GDPPerCapita              15 11021        12146.44         18971.36  3.13    14.36        180.71
GDPPerCapGrowth           16 10942            1.91             6.68  2.58    61.16          0.06
TotalTrade                17  8843           78.67            53.83  2.99    17.70          0.57
FDIIn                     18  8861            5.33            44.04 16.01   599.63          0.47
NetAidReceived            19  9043    506951242.00     997064633.65  8.32   157.34   10484966.48
MobileCellSubscriptions   20 10212           36.32            51.76  1.29     1.14          0.51
NaturalResourceRents      21  9211            6.85            11.06  2.60     8.04          0.12
GovtExpenditures          22  8475           16.27             7.98  3.71    36.33          0.09
PublicHealthExpend        23  4098            3.31             2.38  1.34     3.10          0.04
WomenInLegislature        24  4892           18.06            11.86  0.70     0.06          0.17
PaidParentalLeave         25 10152            0.11             0.31  2.50     4.27          0.00
lnGDPPerCap               26 11021            8.38             1.51  0.10    -0.91          0.01
lnNetAidReceived          27  8876           18.81             1.97 -1.06     1.99          0.02
YearNumeric               28 13760         1991.50            18.47  0.00    -1.20          0.16
PostColdWar               29 13760            0.53             0.50 -0.13    -1.98          0.00
```

A New Question

Do paid parental leave policies decrease child mortality?

- Y = `ChildMortality` (N of deaths of children under 5 per 1000 live births) **(logged)**
- T = `PaidParentalLeave` (1 if provided, 0 if not)
- $\mathbf{X}$s:
  - `GDPPerCapita` (Wealth; in constant $US) (logged)
  - `NetAidReceived` (Net official development aid received; in constant $US) (logged)
  - `GovtExpenditures` (Government Expenditures, as a percent of GDP)



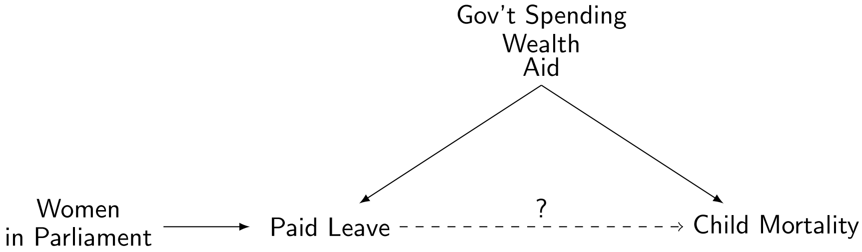
Preliminary Regressions

Table: Models of log(Child Mortality)

	Bivariate OLS	OLS	One-Way FE	Two-Way FE	FE w.Lagged Y
Paid Parental Leave	-1.830	-0.901^{***}	-0.059	-0.148^{***}	-0.207^{***}
	(0.034)	(0.037)	(0.043)	(0.024)	(0.027)
ln(GDP Per Capita)		-0.669^{***}	-1.130^{***}	-0.272^{***}	-0.561^{***}
		(0.009)	(0.017)	(0.013)	(0.012)
ln(Net Aid Received)		-0.089^{***}	-0.106^{***}	0.004	-0.011^{***}
		(0.006)	(0.006)	(0.004)	(0.004)
Government Expenditures		-0.001	0.002	0.0003	0.0002
		(0.001)	(0.001)	(0.001)	(0.001)
Lagged Child Mortality					0.009^{***}
					(0.0001)
Constant	3.770^{*}	10.900^{***}			
	(0.011)	(0.172)			
Observations	9,572	5,306	5,306	5,306	5,301
R^2	0.232	0.586	0.494	0.102	0.802
Adjusted R^2	0.231	0.586	0.479	0.066	0.796

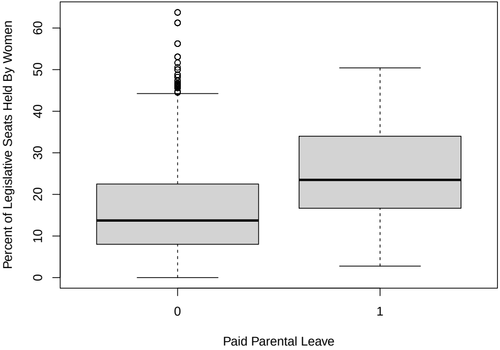
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Conceptually:



Assessing Cov(W, $\mathbf{Z}$):

```
> with(WDI,t.test(WomenInLegislature~PaidParentalLeave))

 Welch Two Sample t-test

data:  WomenInLegislature by PaidParentalLeave
t = -22, df = 1433, p-value <0.0000000000000002
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -9.60 -8.01
sample estimates:
mean in group 0 mean in group 1
           16.2            25.0
```



Instrumental Variables: Syntax

E.g., one-way fixed effects with IV:

```
FE.IV<-plm(lnCM~PaidParentalLeave+log(GDPPerCapita)+
              log(NetAidReceived)+GovtExpenditures |
              . - PaidParentalLeave+WomenInLegislature,
              data=WDI,effect="individual",model="within")
```

Instrumental Variable Results

Table: IV Models of log(Child Mortality)

	OLS	One-Way FE	FE w/IV	RE w/IV
Paid Parental Leave	−0.901	−0.059	452.000	−5.010***
	(0.037)	(0.043)	(5, 530.000)	(1.750)
ln(GDP Per Capita)	−0.669	−1.130***	−72.000	−0.514***
	(0.009)	(0.017)	(867.000)	(0.078)
ln(Net Aid Received)	−0.089	−0.106***	6.780	−0.049*
	(0.006)	(0.006)	(83.700)	(0.025)
Government Expenditures	−0.001	0.002	−0.090	−0.002
	(0.001)	(0.001)	(1.080)	(0.003)
Constant	10.900*			9.040***
	(0.172)			(0.866)
Observations	5,306	5,306	2,764	2,764
R^2	0.586	0.494	0.0001	0.271
Adjusted R^2	0.586	0.479	-0.055	0.270

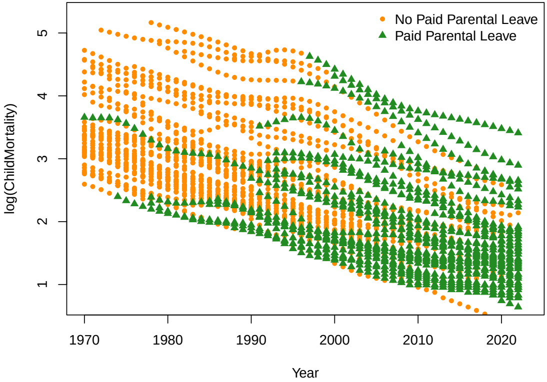
*p<0.1; **p<0.05; ***p<0.01

Another Approach: RDD

Intuition: Compare the child mortality "trajectories" of countries before and after they implement paid parental leave policies.

The model is:

$$\begin{aligned} \text{Child Mortality}_{it} &= \beta_0 + \beta_1(\text{Paid Parental Leave}_{it}) + \beta_2(\text{Time}_t) + \\ &= \beta_3(\text{Paid Parental Leave}_{it} \times \text{Time}_t) + (\text{confounders}) + u_{it} \end{aligned}$$



RDD Regressions

Table 3: RDD Models of log(Child Mortality)

	OLS #1	OLS #2	One-Way FE #1	One-Way FE #2	Two-Way FE #1
(Intercept)	4.3322***	11.5069***			
	(0.0554)	(0.2989)			
Paid Parental Leave	−0.7039***	0.0889	−0.0456	0.1515+	−8.7686**
	(0.1299)	(0.1916)	(0.0412)	(0.0792)	(2.7133)
Time (1950=0)	−0.0390***	−0.0229***	−0.0421***	−0.0430***	
	(0.0013)	(0.0018)	(0.0004)	(0.0011)	
Paid Parental Leave x Time	0.0097***	−0.0036	0.0011	−0.0028*	0.1741***
	(0.0024)	(0.0034)	(0.0007)	(0.0014)	(0.0496)
ln(GDP Per Capita)		−0.6948***		−0.2036***	
		(0.0178)		(0.0249)	
ln(Net Aid Received)		−0.0579***		0.0122*	
		(0.0113)		(0.0054)	
Government Expenditures		−0.0292***		0.0094***	
		(0.0034)		(0.0017)	
Num.Obs.	2785	778	2785	778	2785
R2	0.416	0.746	0.906	0.930	0.005
R2 Adj.	0.415	0.744	0.904	0.927	−0.035

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

Challenges:

- Multiple periods (years) per unit (country), both before and after "treatment"
- "Staggered" treatment timing (adoption of *Paid Parental Leave*)

One approach:

Callaway, Brantley, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225:200-230.

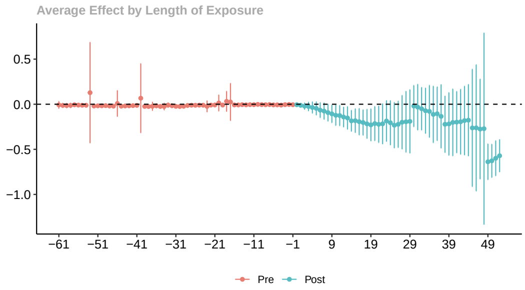
Details:

- Deals with the issues related above
- Flexibly fit / interpreted using the `did` package

Simple bivariate model (no controls):

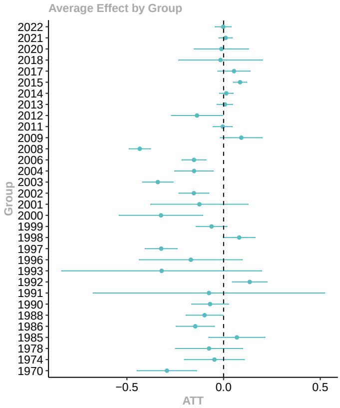
```
> DiD.fit1<-att_gt(yname = "lnCM",gname = "YearPPL",idname = "ID",
+                 tname = "YearNumeric",allow_unbalanced_panel = TRUE,
+                 xformla = ~1,data = WDI,est_method = "reg")

> # Event study object:
>
> DiD.ev1 <- aggte(DiD.fit1,type="dynamic",na.rm=TRUE)
```

Plot the event study results:



ATTs by "Group"

```
> DiD.grp1<-aggte(DiD.fit1,type="group",na.rm=TRUE)
> summary(DiD.grp1)

Call:
aggte(MP = DiD.fit1, type = "group", na.rm = TRUE)

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics, Vol.
225, No. 2, pp. 200-230, 2021.


Overall summary of ATT's based on group/cohort aggregation:
     ATT    Std. Error     [ 95%  Conf. Int.]
 -0.0894         0.017    -0.123     -0.0561 *


Group Effects:
 Group Estimate Std. Error [95% Simult.  Conf. Band]
  1970  -0.2932     0.0378       -0.4465     -0.1398 *
  1974  -0.0475     0.0382       -0.2024      0.1074
  1978  -0.0755     0.0428       -0.2489      0.0980
  1985   0.0686     0.0357       -0.0763      0.2135
  1986  -0.1466     0.0244       -0.2456     -0.0476 *
  1988  -0.0986     0.0235       -0.1941     -0.0032 *
  1990  -0.0700     0.0235       -0.1651      0.0251
  1991  -0.0757     0.1477       -0.6743      0.5230
 .
 .
 .
  2020  -0.0113     0.0346       -0.1515      0.1290
  2021   0.0104     0.0086       -0.0245      0.0453
  2022  -0.0021     0.0102       -0.0434      0.0393
---
Signif. codes: '*' confidence band does not cover 0

Control Group:  Never Treated,  Anticipation Periods:  0
Estimation Method:  Outcome Regression
```

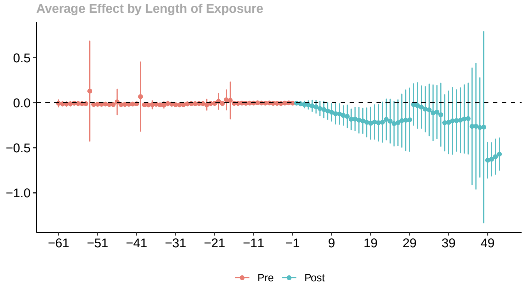


Differences-in-Differences with Controls

Adding control variables:

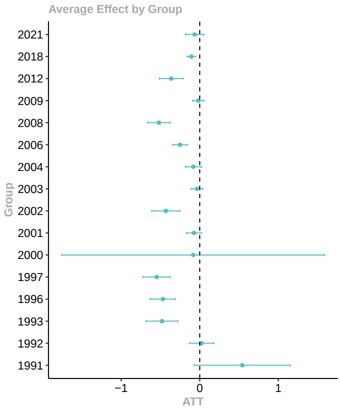
```
> DiD.fit2<-att_gt(yname = "lnCM",gname = "YearPPL",idname = "ID",
+                  tname = "YearNumeric",allow_unbalanced_panel = TRUE,
+                  xformla = ~lnGDPPerCap+lnNetAidReceived+GovtExpenditures,
+                  data = WDI, est_method = "reg")
>
> # Event study object:
>
> DiD.ev2 <- aggte(DiD.fit2,type="dynamic",na.rm=TRUE)
```

Plot the event study results:



ATTs by "Group" (with controls)

```
> DiD.grp2<-aggte(DiD.fit2,type="group",na.rm=TRUE)
> summary(DiD.grp2)

Call:
aggte(MP = DiD.fit2, type = "group", na.rm = TRUE)

Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.  "Difference-in-
Differences with Multiple Time Periods." Journal of Econometrics,
Vol. 225, No. 2, pp. 200-230, 2021.

Overall summary of ATT's based on group/cohort aggregation:
    ATT    Std. Error     [ 95%  Conf. Int.]
 -0.169        0.0423     -0.252     -0.0857 *

Group Effects:
 Group Estimate Std. Error [95% Simult.  Conf. Band]
  1991   0.5416     0.1729        0.0788      1.0044 *
  1992   0.0236     0.0568       -0.1284      0.1757
  1993  -0.4799     0.0768       -0.6854     -0.2743 *
  1996  -0.4694     0.0607       -0.6319     -0.3069 *
  1997  -0.5490     0.0617       -0.7141     -0.3838 *
  2000  -0.0837     0.3987       -1.1510      0.9835
  2001  -0.0730     0.0377       -0.1739      0.0280
  2002  -0.4316     0.0684       -0.6146     -0.2486 *
  2003  -0.0351     0.0266       -0.1064      0.0361
  2004  -0.0821     0.0362       -0.1789      0.0148
  2006  -0.2511     0.0350       -0.3448     -0.1575 *
  2008  -0.5189     0.0533       -0.6614     -0.3763 *
  2009  -0.0188     0.0251       -0.0861      0.0485
  2012  -0.3636     0.0552       -0.5115     -0.2158 *
  2018  -0.1067     0.0199       -0.1600     -0.0534 *
  2021  -0.0651     0.0470       -0.1910      0.0608
---
Signif. codes: '*' confidence band does not cover 0

Control Group:  Never Treated,  Anticipation Periods:  0
Estimation Method:  Outcome Regression
```



Resources

- Good references:
  - Freedman (2012)*
  - Shalizi (someday)*
  - Morgan and Winship (2014)
  - Pearl et al. (2016)
  - Peters et al. (2017)
- Courses / syllabi (a sampling):
  - Eggers (2019)
  - Frey (2023)
  - Imai (2023)
  - Munger (2023)
  - Xu (2018, 2023).
  - Yamamoto (2022)
- Other useful things:
  - The CRAN task view on causal inference
  - The Causal Inference Book
  - Some useful notes

* I really like this one.

Appendix

Event Counts: Unit Effects

$$Y_{it} \sim \text{Poisson}(\mu_{it} = \alpha_i \lambda_{it})$$

with $\lambda_{it} = \exp(\mathbf{X}_{it}\boldsymbol{\beta})$ implies:

$$\begin{aligned} \text{E}(Y_{it} \mid \mathbf{X}_{it}, \alpha_i) &= \mu_{it} \\ &= \alpha_i \exp(\mathbf{X}_{it}\boldsymbol{\beta}) \\ &= \exp(\delta_i + \mathbf{X}_{it}\boldsymbol{\beta}) \end{aligned}$$

where $\delta_i = \ln(\alpha_i)$.

Fixed-Effects Poisson:

- ...has no "incidental parameters" problem (see e.g. Cameron and Trivedi, pp. 281-2)
- This means "brute force" approach works
- Fitted via `glmmML` in R, `xtpoisson` (and `xtnbreg`) in Stata

Random-Effects Models

$$\begin{aligned}
\Pr(Y_{i1} = y_{i1}, \ldots Y_{iT} = y_{iT}) &= \int_0^{\infty} \Pr(Y_{i1} = y_{i1}, \ldots Y_{iT} = y_{iT}) f(\alpha_i) d\alpha_i \\
&= \int_0^{\infty} \left[\prod_{t=1}^{T} \Pr(Y_{it} \mid \alpha_i) \right] f(\alpha_i) d\alpha_i
\end{aligned}$$

- Simplest to assume $\alpha_i \sim \Gamma(\theta)$
- Yields a model with $\mathrm{E}(Y_{it}) = \lambda_{it}$ and $\mathrm{Var}(Y_{it}) = \lambda_{it} + \frac{\lambda_{it}^2}{\theta}$
- Via `glmmML` or `glmer` in R, or `xtpois, re` in Stata
- $\exists$ random effects negative binomial too...

R:

- Tobit = `censReg` (in **censReg**)
- Poisson (random effects) = `glmmML` in **glmmML** or `glmer` in **lme4**
- Poisson (fixed effects) = `glmmML` or "brute force"
- All of the above = `pglm`

Stata:

- Tobit = `xttobit` (`re` only)
- Poisson / negative binomial = `xtpoisson`, `xtnbreg` (both with `fe`, `re` options)

Conflict Onsets: Pooled Poisson

```
> xtabs(~DF$OnsetCount)

DF$OnsetCount
   0    1    2    3    4
8981  375   30    7    1

> Poisson<-glm(OnsetCount~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+              GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared,data=DF,family="poisson")

> summary(Poisson)

Call:
glm(formula = OnsetCount ~ log(LandArea) + log(PopMillions) +
    UrbanPopulation + log(GDPPerCapita) + GDPPerCapGrowth + PostColdWar +
    POLITY + POLITYSquared, family = "poisson", data = DF)

Coefficients:
                  Estimate Std. Error z value   Pr(>|z|)
(Intercept)       -2.38261    0.72320   -3.29    0.00099 ***
log(LandArea)      0.06936    0.04693    1.48    0.13941
log(PopMillions)   0.42571    0.04569    9.32    < 2e-16 ***
UrbanPopulation    0.00603    0.00472    1.28    0.20106
log(GDPPerCapita) -0.42991    0.08086   -5.32 0.00000011 ***
GDPPerCapGrowth   -0.03595    0.00641   -5.61 0.00000002 ***
PostColdWar        0.27202    0.12002    2.27    0.02343 *
POLITY             0.32968    0.08289    3.98 0.00006961 ***
POLITYSquared     -0.03636    0.00793   -4.59 0.00000449 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2390.6  on 6996  degrees of freedom
Residual deviance: 1949.8  on 6988  degrees of freedom
  (6395 observations deleted due to missingness)
AIC: 2704

Number of Fisher Scoring iterations: 6
```

Fixed Effects Poisson

```
> FEPoisson<-pglm(OnsetCount~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+                GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared,data=DF,family="poisson",
+               effect="individual",model="within")

> summary(FEPoisson)

--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 3 iterations
Return code 8: successive function values within relative tolerance limit (reltol)
Log-Likelihood: -1021
8  free parameters
Estimates:
                 Estimate Std. error t value    Pr(> t)
log(LandArea)    -1.67100    2.83168   -0.59    0.55512
log(PopMillions)  0.61473    0.32126    1.91    0.05568 .
UrbanPopulation  -0.04603    0.01335   -3.45    0.00056 ***
log(GDPPerCapita) -0.09145    0.14421   -0.63    0.52600
GDPPerCapGrowth  -0.02637    0.00654   -4.03 0.00005499 ***
PostColdWar       0.48566    0.19617    2.48    0.01330 *
POLITY            0.52507    0.10791    4.87 0.00000114 ***
POLITYSquared    -0.05379    0.01060   -5.07 0.00000039 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
--------------------------------------------
```

Alternative Fixed Effects Poisson (using `feglm`)

```
> FEPoisson2<-feglm(OnsetCount~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+                   GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared|ISO3,data=DF,family="poisson")

NOTES: 6,395 observations removed because of NA values (LHS: 3,998, RHS: 6,395).
       67 fixed-effects (2,499 observations) removed because of only 0 outcomes.

> summary(FEPoisson2,cluster="ISO3")

GLM estimation, family = poisson, Dep. Var.: OnsetCount
Observations: 4,498
Fixed-effects: ISO3: 93
Standard-errors: Clustered (ISO3)
                  Estimate Std. Error t value     Pr(>|t|)
log(LandArea)     -1.67100   2.159264 -0.7739 0.4390039115
log(PopMillions)   0.61473   0.340011  1.8080 0.0706106957 .
UrbanPopulation   -0.04603   0.019252 -2.3911 0.0167991301 *
log(GDPPerCapita) -0.09145   0.151293 -0.6045 0.5455437492
GDPPerCapGrowth   -0.02637   0.006008 -4.3900 0.0000113372 ***
PostColdWar        0.48566   0.293791  1.6531 0.0983179526 .
POLITY             0.52507   0.112045  4.6862 0.0000027826 ***
POLITYSquared     -0.05379   0.011709 -4.5937 0.0000043554 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Log-Likelihood: -1,156.1   Adj. Pseudo R2: 0.094671
           BIC:  3,163.5     Squared Cor.: 0.162849
```

Random Effects Poisson

```
> REPoisson<-glmer(OnsetCount~log(LandArea)+log(PopMillions)+UrbanPopulation+log(GDPPerCapita)+
+                    GDPPerCapGrowth+PostColdWar+POLITY+POLITYSquared+(1|ISO3),data=DF,family="poisson")

> summary(REPoisson)

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: poisson  ( log )
Formula: OnsetCount ~ log(LandArea) + log(PopMillions) + UrbanPopulation +
    log(GDPPerCapita) + GDPPerCapGrowth + PostColdWar + POLITY +
    POLITYSquared + (1 | ISO3)
   Data: DF

     AIC      BIC   logLik deviance df.resid
    2602     2670    -1291     2582     6987

Scaled residuals:
   Min     1Q Median     3Q    Max
-0.945 -0.227 -0.144 -0.086 17.093

Random effects:
 Groups Name        Variance Std.Dev.
 ISO3   (Intercept) 0.588    0.767
Number of obs: 6997, groups:  ISO3, 160

Fixed effects:
                  Estimate Std. Error z value     Pr(>|z|)
(Intercept)       -4.33127    1.09253   -3.96 0.0000735687 ***
log(LandArea)      0.07661    0.07524    1.02        0.309
log(PopMillions)   0.42058    0.08230    5.11 0.0000003215 ***
UrbanPopulation   -0.00756    0.00649   -1.16        0.244
log(GDPPerCapita) -0.16788    0.10506   -1.60        0.110
GDPPerCapGrowth   -0.03182    0.00660   -4.82 0.0000014481 ***
PostColdWar        0.29773    0.12970    2.30        0.022 *
POLITY             0.49337    0.09700    5.09 0.0000003649 ***
POLITYSquared     -0.05419    0.00942   -5.75 0.0000000089 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation of Fixed Effects:
            (Intr) lg(LA) lg(PM) UrbnPp l(GDPP GDPPCG PstClW POLITY
log(LandAr) -0.774
lg(PpMllns)  0.395 -0.656
UrbanPopltn  0.364 -0.043 -0.033
lg(GDPPrCp) -0.589  0.020  0.022 -0.737
GDPPrCpGrwt  0.041  0.066 -0.106  0.126 -0.165
PostColdWar -0.112  0.186 -0.245 -0.218  0.035 -0.053
POLITY      -0.278  0.006 -0.001 -0.075  0.214  0.066 -0.255
POLITYSqurd  0.261  0.028 -0.038  0.052 -0.241 -0.065  0.208 -0.968
optimizer (Nelder_Mead) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.116002 (tol = 0.002, component 1)
Model is nearly unidentifiable: very large eigenvalue
 - Rescale variables?
```

Alternative RE Poisson (using `pglm`)

```
> REPoisson2<-pglm(OnsetCount~log(LandArea)+log(PopMillions)+
+                  UrbanPopulation+log(GDPPerCapita)+
+                  GDPPerCapGrowth+PostColdWar+POLITY+
+                  POLITYSquared,data=DF,family="poisson",
+                effect="individual",model="random")

> summary(REPoisson2)

--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 4 iterations
Return code 8: successive function values within relative tolerance limit (reltol)
Log-Likelihood: -1292
10  free parameters

Estimates:
                  Estimate Std. error t value     Pr(> t)
(Intercept)       -3.67347    1.05113   -3.49     0.00047 ***
log(LandArea)      0.05547    0.07325    0.76     0.44888
log(PopMillions)   0.44374    0.08003    5.54 0.000000030 ***
UrbanPopulation   -0.00613    0.00637   -0.96     0.33518
log(GDPPerCapita) -0.19283    0.10268   -1.88     0.06038 .
GDPPerCapGrowth   -0.03201    0.00655   -4.88 0.000001044 ***
PostColdWar        0.29663    0.12891    2.30     0.02139 *
POLITY             0.47529    0.09584    4.96 0.000000708 ***
POLITYSquared     -0.05274    0.00929   -5.68 0.000000014 ***
sigma              1.70087    0.41233    4.12 0.000037074 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
--------------------------------------------
```

Table!

Panel Event Count Models

	Poisson	FE Poisson	RE Poisson	Neg. Bin.	FE N.B.	RE N.B.
Intercept	-2.38^{*}		-4.33^{*}	-2.41^{*}	-62.39	-4.32^{*}
	(0.72)		(1.09)	(0.74)		(1.09)
ln(Land Area)	0.07	-1.67	0.08	0.07	6.56	0.08
	(0.05)	(2.83)	(0.08)	(0.05)		(0.08)
ln(Population)	0.43^{*}	0.61	0.42^{*}	0.42^{*}	1.25	0.42^{*}
	(0.05)	(0.32)	(0.08)	(0.05)	(1.46)	(0.08)
Urban Population	0.01	-0.05^{*}	-0.01	0.01	-0.10	-0.01
	(0.00)	(0.01)	(0.01)	(0.00)	(0.08)	(0.01)
ln(GDP Per Capita)	-0.43^{*}	-0.09	-0.17	-0.42^{*}	3.26^{*}	-0.17
	(0.08)	(0.14)	(0.11)	(0.08)	(1.25)	(0.11)
GDP Growth	-0.04^{*}	-0.03^{*}	-0.03^{*}	-0.04^{*}	-0.07^{*}	-0.03^{*}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)
Post-Cold War	0.27^{*}	0.49^{*}	0.30^{*}	0.27^{*}	-0.57	0.30^{*}
	(0.12)	(0.20)	(0.13)	(0.12)	(1.15)	(0.13)
POLITY	0.33^{*}	0.53^{*}	0.49^{*}	0.32^{*}	1.29^{*}	0.49^{*}
	(0.08)	(0.11)	(0.10)	(0.09)	(0.59)	(0.10)
POLITY Squared	-0.04^{*}	-0.05^{*}	-0.05^{*}	-0.04^{*}	-0.10^{*}	-0.05^{*}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.05)	(0.01)
Estimated Sigma				0.06		
				(0.03)		
AIC	2704.01	2057.19	2601.46	2699.78	-1271.03	2603.46
BIC	2765.69		2670.00			2678.84
Log Likelihood	-1343.01	-1020.59	-1290.73	-1339.89	644.51	-1290.73
Deviance	1949.83					
Num. obs.	6997		6997			6997
Num. groups: ISO3			160			160
Var: ISO3 (Intercept)			0.59			0.59

$^{*}p < 0.05$