GSERM - St. Gallen 2022
Analyzing Panel Data

June 10, 2022

Logit/Probit Redux

Start with:

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + u_i$$

$$\begin{aligned} Y_i &= 0 \text{ if } Y_i^* < 0 \\ Y_i &= 1 \text{ if } Y_i^* \geq 0 \end{aligned}$$

So:

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* \geq 0) \\ &= \Pr(\mathbf{X}_i\boldsymbol{\beta} + u_i \geq 0) \\ &= \Pr(u_i \geq -\mathbf{X}_i\beta) \\ &= \Pr(u_i \leq \mathbf{X}_i\beta) \\ &= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} f(u)du \end{aligned}$$

Logit

"Standard logistic" PDF:

$$\Pr(u) \equiv \lambda(u) = \frac{\exp(u)}{[1 + \exp(u)]^2}$$

CDF:

$$\begin{aligned}
\Lambda(u) &= \int \lambda(u) du \\
&= \frac{\exp(u)}{1 + \exp(u)} \\
&= \frac{1}{1 + \exp(-u)}
\end{aligned}$$

Logistic $\rightarrow$ "Logit"

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\
&= \Pr(u_i \leq \mathbf{X}_i\boldsymbol{\beta}) \\
&= \Lambda(\mathbf{X}_i\boldsymbol{\beta}) \\
&= \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}
\end{aligned}
$$

$$
\text{(equivalently)} = \frac{1}{1 + \exp(-\mathbf{X}_i\boldsymbol{\beta})}
$$

Likelihoods

$$L_i = \left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]^{1-Y_i}$$

$$L = \prod_{i=1}^{N}\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]^{1-Y_i}$$

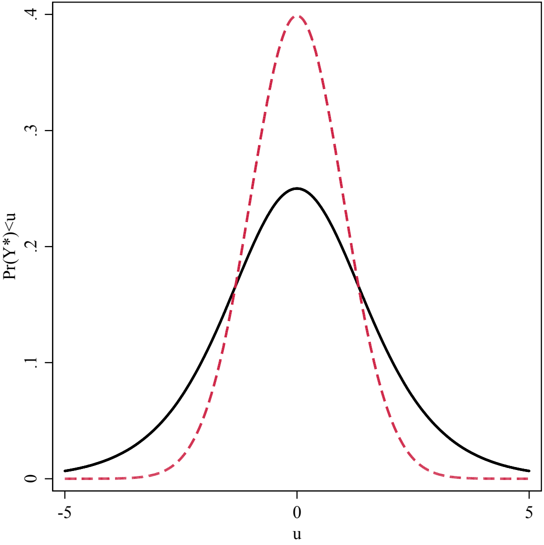
$$\begin{aligned}\ln L \quad = \quad & \sum_{i=1}^{N} Y_i \ln\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right) + \\ & (1-Y_i)\ln\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]\end{aligned}$$

Probit...

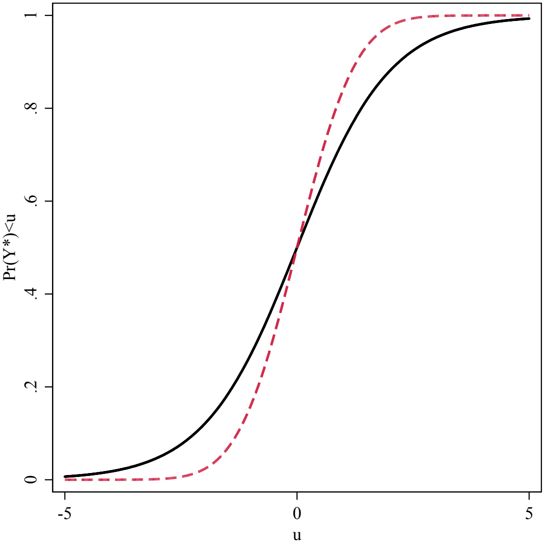
$$\Pr(u) \equiv \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

Standard Normal and Logistic PDFs



Standard Normal and Logistic CDFs



Probit (continued...)

$$\begin{aligned}\Pr(Y_i = 1) &= \Phi(\mathbf{X}_i\boldsymbol{\beta}) \\ &= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{X}_i\boldsymbol{\beta})^2}{2}\right) d\mathbf{X}_i\boldsymbol{\beta}\end{aligned}$$

$$L = \prod_{i=1}^{N} [\Phi(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i} [1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})]^{(1-Y_i)}$$

$$\ln L = \sum_{i=1}^{N} Y_i \ln\Phi(\mathbf{X}_i\boldsymbol{\beta}) + (1 - Y_i) \ln[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})]$$

Panel / TSCS: What Can Go Wrong?

Suppose:

$$\begin{aligned} X_{it} &= \rho_X \mathbf{X}_{it-1} + \nu_{it} \\ u_{it} &= \rho_u u_{it-1} + \epsilon_{it} \end{aligned}$$

For high values of ρ, logit/probit:

- $\hat{\boldsymbol{\beta}}$s are consistent, but s.e.s are biased, inefficient (Poirier and Ruud 1988);
- $\rightarrow$ underestimate Var(β) by up to 50 percent (Beck and Katz 1997).

One-way unit effects:

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

for logit only, so:

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

Incidental Parameters:

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

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

- Chamberlain:

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

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

Points:

- Fixed effects = no estimates for $\boldsymbol{\beta}_b$
- Interpretation: per logit, but $\mid \hat{\alpha}_i$.
- BTSCS in IR: Green et al. (2001) v. B&K (2001).

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 $\sigma_\alpha^2 = \left(\frac{\rho}{1-\rho}\right)$.

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).

Unit Effects in Practice - Some Simulations

Start with:

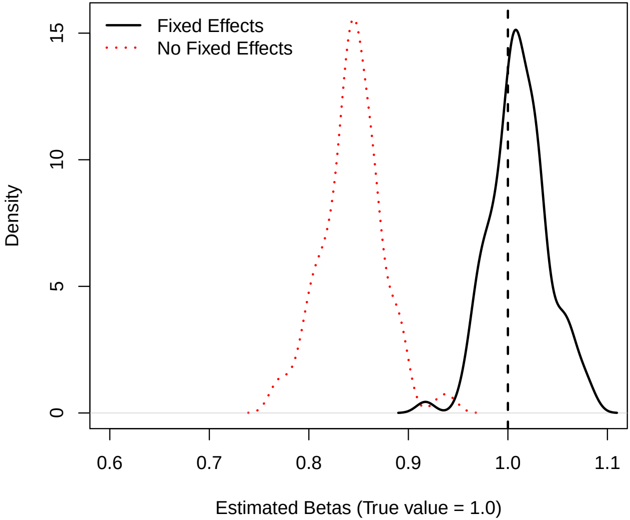
$$\begin{aligned} Y_{it}^* &= 0 + (1 \times X_{it}) + (1 \times D_{it}) + (1 \times \alpha_i) + u_{it} \\ Y_{it} \in \{0, 1\} &= f(Y_{it}^*) \end{aligned}$$

where:

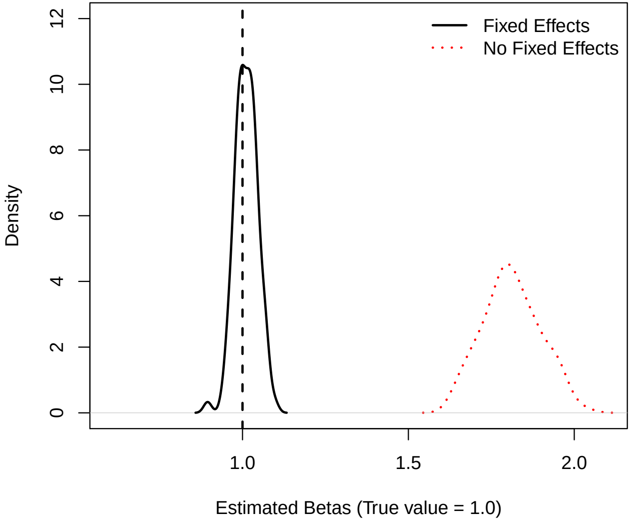
- $\alpha_i \sim N(0, 1)$
- $X_{it} \sim N(0, \sigma_X^2)$
- $D_{it} \in \{0, 1\}$
- $\text{Cov}(X_{it}, \alpha_i) = \{0, 0.69\}$
- $\text{Cov}(D_{it}, \alpha_i) = 0$
- $f(\cdot) = \{\text{logit, probit}\}$ (as appropriate)

and $N = T = 100$.

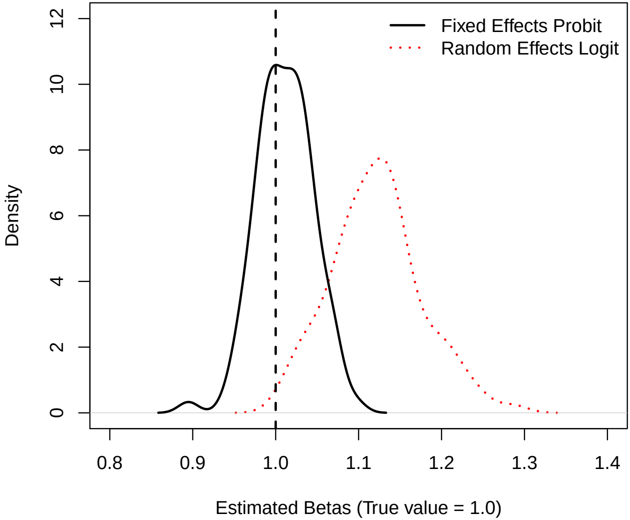
Logit $\hat{\beta}_X$s for $\mathrm{Cov}(X_{it}, \alpha_i) = 0$



Logit $\hat{\beta}_X$s for $\mathrm{Cov}(X_{it}, \alpha_i) \approx 0.69$



Logit $\hat{\boldsymbol{\beta}}_X$s for $\text{Cov}(X_{it}, \alpha_i) \approx 0.69$



R

- `pglm` (panel GLMs) (maximum likelihood + quadrature)
- `bife` (fixed-effects logit / probit only)
- `glmer` (general mixed-effects models, including RE)
- `glmmML` (via Gauss-Hermite quadrature)
- `MCMCpack` (`MCMChlogit`)
- Various user-generated functions (e.g., here).

Stata

- `xtprobit`, `xtlogit`, `xtcloglog`
- Plus `xttrans` (transition probabilities), `quadchk` (quadrature checking), `xtrho` / `xtrhoi` (estimation of within-unit covariances)

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    min        max       range       se
ISO3*              1 13392    108.50      62.36   1.00      216.0      215.00     0.54
Year*              2 13392     31.50      17.90   1.00       62.0       61.00     0.15
country*           3 13330    108.00      62.07   1.00      215.0      214.00     0.54
CivilWar           4  9052      0.13       0.34   0.00        1.0        1.00     0.00
OnsetCount         5  9394      0.05       0.24   0.00        4.0        4.00     0.00
LandArea           6 12906 613525.38 1766486.19   2.03 16389950.0 16389947.97 15549.43
PopMillions        7 13073     24.64     103.13   0.00     1410.9     1410.93     0.90
UrbanPopulation    8 13045     51.39      25.74   2.08      100.0       97.92     0.23
GDPPerCapita       9  9582  11685.74   18675.05 144.20   181709.3   181565.14   190.78
GDPPerCapGrowth   10  9598      1.89       6.21 -64.99      140.4      205.36     0.06
PostColdWar       11 13330      0.52       0.50   0.00        1.0        1.00     0.00
POLITY            12  8279      5.55       3.71   0.00       10.0       10.00     0.04
POLITYSquared     13  8279     44.57      40.24   0.00      100.0      100.00     0.44
```

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)       -1.03275    0.52731   -1.96  0.05017 .
log(LandArea)      0.01085    0.03246    0.33  0.73815
log(PopMillions)   0.66364    0.03696   17.96  < 2e-16 ***
UrbanPopulation    0.01090    0.00335    3.26  0.00113 **
log(GDPPerCapita) -0.50128    0.06128   -8.18  2.8e-16 ***
GDPPerCapGrowth   -0.04029    0.00644   -6.26  3.9e-10 ***
PostColdWar       -0.31102    0.08588   -3.62  0.00029 ***
POLITY             0.67438    0.06122   11.02  < 2e-16 ***
POLITYSquared     -0.06526    0.00579  -11.27  < 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: 5843.6  on 6996  degrees of freedom
Residual deviance: 4624.8  on 6988  degrees of freedom
  (6395 observations deleted due to missingness)
AIC: 4643
```

Fixed Effects

```
> 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)     -4.00079    6.80808   -0.59    0.5568
log(PopMillions)   0.79303    0.29847    2.66    0.0079 **
UrbanPopulation    0.01179    0.01228    0.96    0.3368
log(GDPPerCapita) -0.33859    0.17226   -1.97    0.0493 *
GDPPerCapGrowth   -0.04960    0.00833   -5.96   2.6e-09 ***
PostColdWar       -0.21475    0.17822   -1.20    0.2282
POLITY             0.70692    0.09365    7.55   4.4e-14 ***
POLITYSquared     -0.07382    0.00890   -8.29   < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

residual deviance= 2846,
null deviance= 4422,
nT= 3971, N= 83

( 6395 observation(s) deleted due to missingness )
( 3026 observation(s) deleted due to perfect classification )

Number of Fisher Scoring Iterations: 6

Average individual fixed effect= 48.24
```

```
> 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, 18 iterations
Return code 2: successive function values within tolerance limit (tol)
Log-Likelihood: -1634
10  free parameters
Estimates:
                  Estimate Std. error t value Pr(> t)
(Intercept)       -4.08609    1.02028   -4.00 6.2e-05 ***
log(LandArea)      0.15120    0.05920    2.55 0.01065 *
log(PopMillions)   1.20067    0.08537   14.06 < 2e-16 ***
UrbanPopulation    0.01973    0.00598    3.30 0.00097 ***
log(GDPPerCapita) -0.61681    0.11732   -5.26 1.5e-07 ***
GDPPerCapGrowth   -0.04979    0.00816   -6.10 1.1e-09 ***
PostColdWar       -0.38811    0.12189   -3.18 0.00145 **
POLITY             0.68171    0.08400    8.12 4.9e-16 ***
POLITYSquared     -0.07368    0.00811   -9.08 < 2e-16 ***
sigma              2.29777    0.11784   19.50 < 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	-1.03		-4.09^{*}
	(0.53)		(1.02)
ln(Land Area)	0.01	-4.00	0.15^{*}
	(0.03)	(6.81)	(0.06)
ln(Population)	0.66^{*}	0.79^{*}	1.20^{*}
	(0.04)	(0.30)	(0.09)
Urban Population	0.01^{*}	0.01	0.02^{*}
	(0.00)	(0.01)	(0.01)
ln(GDP Per Capita)	-0.50^{*}	-0.34^{*}	-0.62^{*}
	(0.06)	(0.17)	(0.12)
GDP Growth	-0.04^{*}	-0.05^{*}	-0.05^{*}
	(0.01)	(0.01)	(0.01)
Post-Cold War	-0.31^{*}	-0.21	-0.39^{*}
	(0.09)	(0.18)	(0.12)
POLITY	0.67^{*}	0.71^{*}	0.68^{*}
	(0.06)	(0.09)	(0.08)
POLITY Squared	-0.07^{*}	-0.07^{*}	-0.07^{*}
	(0.01)	(0.01)	(0.01)
Estimated Sigma			2.30^{*}
			(0.12)
AIC	4642.76		3287.00
BIC	4704.44		
Log Likelihood	-2312.38	-1422.95	-1633.50
Deviance	4624.76	2845.89	
Num. obs.	6997	3971	

$^{*}p < 0.05$

Censoring and Event Counts

"Lower" censored Y:

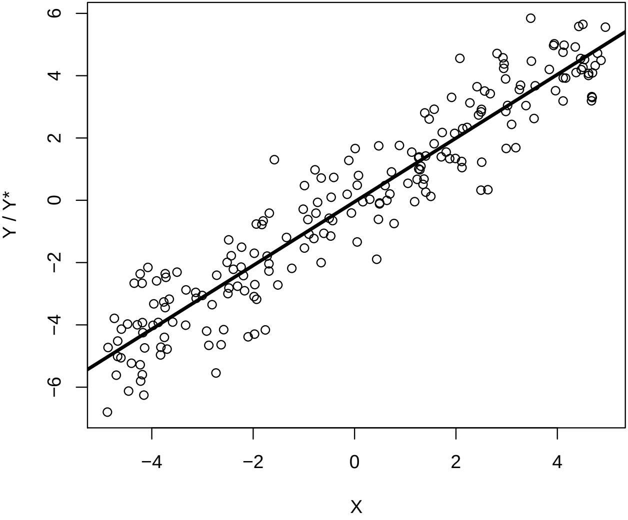
$$
\begin{aligned}
Y_i &= Y_i^* \text{ if } Y_i^* > L \\
&= L \text{ if } Y_i^* \leq L
\end{aligned}
$$

"Upper-censored":

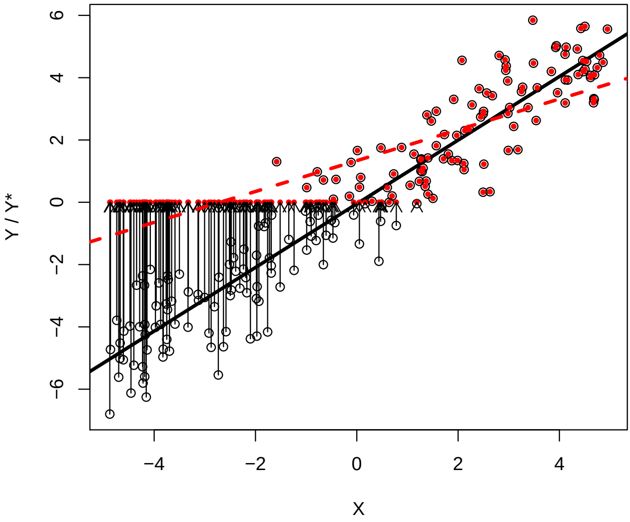
$$
\begin{aligned}
Y_i &= Y_i^* \text{ if } Y_i^* > L \\
&= U \text{ if } Y_i^* \geq L
\end{aligned}
$$

$\rightarrow$ bias in $\hat{\beta}$ (toward zero) + inconsistency...

Censoring Bias



Censoring Bias



In the lower-censoring case, for $Y^* > L$, we have:

$$\mathbf{L}_1(\boldsymbol{\beta}, \sigma^2 | Y, L) = \prod_{Y_i > L} \phi(Y_i^* | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2).$$

and for $Y^* \leq L$:

$$\begin{aligned}
\Pr(Y_i = L) &= \Pr(Y_i^* \leq L) \\
&= \int_{-\infty}^{L} \phi(Y_i^* | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) \, dY^* \\
&= \Phi(L | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2).
\end{aligned}$$

which implies:

$$\mathbf{L}_2(\boldsymbol{\beta}, \sigma^2 | Y, L) = \prod_{Y_i = L} \Phi(L | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2).$$

Combined likelihood:

$$\mathbf{L}(\boldsymbol{\beta}, \sigma^2 | Y, L) = \prod_{Y_i > L} \phi(Y_i^* | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2) \prod_{Y_i = L} \Phi(L | \mathbf{X}_i, \boldsymbol{\beta}, \sigma^2).$$

One-way unit effects:

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

Models:

- No fixed-effects conditioning (a la logit) $\rightarrow$ inconsistency.
- Generally use random effects (via `survival` or `xttobit`).

Properties:

- Discrete / integer-values
- Non-negative
- "Cumulative"

Motivation:

$$\text{Arrival Rate} = \lambda$$

$$\Pr(\text{Event})_{t,t+h} = \lambda h$$

$$\Pr(\text{No Event})_{t,t+h} = 1 - \lambda h$$

$$\begin{aligned}
\Pr(Y_t = y) &= \frac{\exp(-\lambda h)\lambda h^y}{y!} \\
&= \frac{\exp(-\lambda)\lambda^y}{y!}
\end{aligned}$$

Poisson: Assumptions and Motivations

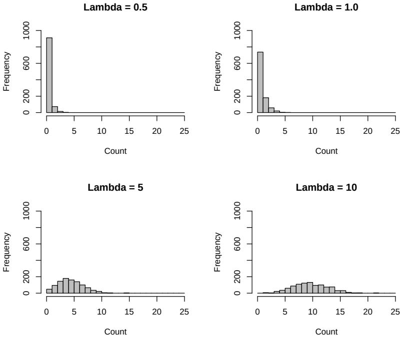
- No Simultaneous Events
- Constant Arrival Rate
- Independent Event Arrivals

Another motivation: For M independent Bernoulli trials with (sufficiently small) probability of success π and where $M\pi \equiv \lambda > 0$,

$$\begin{aligned}
\Pr(Y_i = y) &= \lim_{M\to\infty}\left[\binom{M}{y}\left(\frac{\lambda}{M}\right)^{y}\left(1-\frac{\lambda}{M}\right)^{M-y}\right] \\
&= \frac{\lambda^y \exp(-\lambda)}{y!}
\end{aligned}$$

Poisson: Characteristics

- Discrete
- $\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$
- Is not preserved under affine transformations...
- For $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$, $Z = X + Y \sim \text{Poisson}(\lambda_{X+Y})$ iff X and Y are independent but
- ...same is not true for differences.
- $\lambda \to \infty \iff Y \sim N$



Poisson Regression

Suppose

$$E(Y_i) \equiv \lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$$

then

$$\Pr(Y_i = y|\mathbf{X}_i, \boldsymbol{\beta}) = \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^y}{y!}$$

with likelihood:

$$L = \prod_{i=1}^{N} \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i}}{Y_i!}$$

and log-likelihood:

$$\ln L = \sum_{i=1}^{N}[-\exp(\mathbf{X}_i\boldsymbol{\beta}) + Y_i\mathbf{X}_i\boldsymbol{\beta} - \ln(Y_i!)]$$

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

<u>Fixed-Effects Poisson</u>:

- ...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

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

- 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"

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

$^*p < 0.05$

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

Generalized Estimating Equations

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:

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

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

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

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

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

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, $\hat{\boldsymbol{\beta}}_M \approx \frac{\hat{\boldsymbol{\beta}}_C}{\sqrt{1+0.35\sigma_\eta^2}}$, where $\sigma_\eta^2 > 0$ 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`)

GEEs: Software Tips

- 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 War Redux... GEE: Independence

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

> summary(GEE.ind)

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

 Coefficients:
                  Estimate Std.err  Wald Pr(>|W|)
(Intercept)        -1.0327  1.9726  0.27  0.60059
log(LandArea)       0.0109  0.1234  0.01  0.92992
log(PopMillions)    0.6636  0.1568 17.90 0.000023 ***
UrbanPopulation     0.0109  0.0137  0.64  0.42538
log(GDPPerCapita)  -0.5013  0.2454  4.17  0.04106 *
GDPPerCapGrowth    -0.0403  0.0128  9.88  0.00167 **
PostColdWar        -0.3110  0.2594  1.44  0.23049
POLITY              0.6744  0.2105 10.26  0.00136 **
POLITYSquared      -0.0653  0.0194 11.34  0.00076 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation structure = independence
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.803   0.291
Number of clusters:   160  Maximum cluster size: 57
```

GEE: Exchangeable

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

> summary(GEE.exc)

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

 Coefficients:
                  Estimate  Std.err  Wald Pr(>|W|)
(Intercept)       -2.91574  2.05337  2.02  0.15561
log(LandArea)      0.05297  0.15494  0.12  0.73245
log(PopMillions)   0.55323  0.16035 11.90  0.00056 ***
UrbanPopulation    0.00533  0.01165  0.21  0.64714
log(GDPPerCapita) -0.21791  0.17470  1.56  0.21229
GDPPerCapGrowth   -0.03530  0.00904 15.23 0.000095 ***
PostColdWar       -0.14044  0.23285  0.36  0.54641
POLITY             0.54979  0.17023 10.43  0.00124 **
POLITYSquared     -0.05610  0.01664 11.36  0.00075 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation structure = exchangeable
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.725   0.185
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha     0.34   0.112
Number of clusters:   160  Maximum cluster size: 57
```

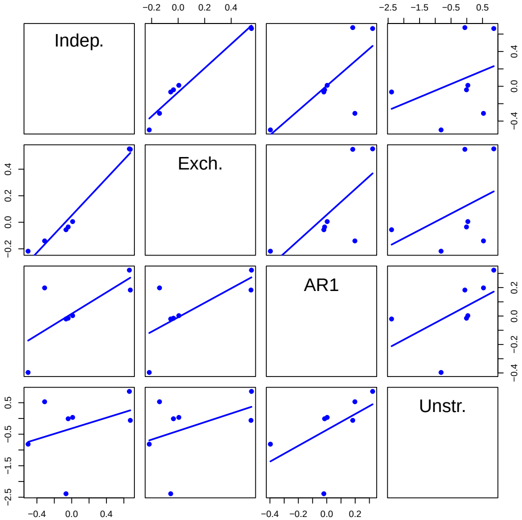
GEE: AR(1)

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

> summary(GEE.ar1)

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

 Coefficients:
                  Estimate  Std.err Wald Pr(>|W|)
(Intercept)       -2.11808  2.41377 0.77    0.380
log(LandArea)      0.17430  0.18542 0.88    0.347
log(PopMillions)   0.32266  0.19145 2.84    0.092 .
UrbanPopulation    0.00279  0.01595 0.03    0.861
log(GDPPerCapita) -0.39669  0.23482 2.85    0.091 .
GDPPerCapGrowth   -0.01526  0.00728 4.40    0.036 *
PostColdWar        0.19787  0.24491 0.65    0.419
POLITY             0.18284  0.12351 2.19    0.139
POLITYSquared     -0.02066  0.01320 2.45    0.117
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation structure = ar1
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.825   0.352
  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha     0.92  0.0404
Number of clusters:   160  Maximum cluster size: 57
```

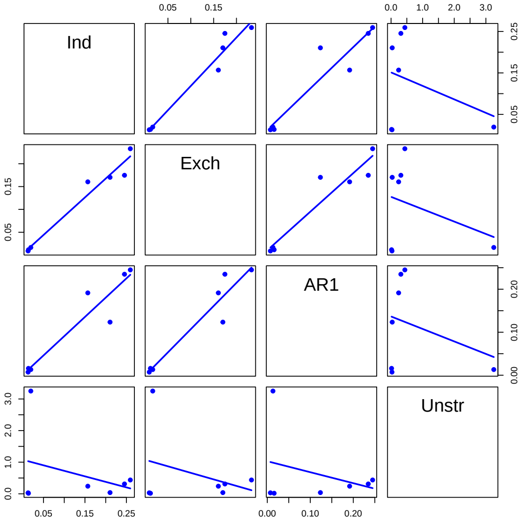
GEE: Unstructured (2013-2017)

```
> 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)       -2.38896  3.25077  0.54  0.46241
log(LandArea)      0.16453  0.19119  0.74  0.38949
log(PopMillions)   0.85836  0.24080 12.71  0.00036 ***
UrbanPopulation    0.03406  0.01715  3.95  0.04699 *
log(GDPPerCapita) -0.81577  0.31150  6.86  0.00882 **
GDPPerCapGrowth   -0.00896  0.03066  0.09  0.77000
POLITY             0.53049  0.43746  1.47  0.22526
POLITYSquared     -0.06053  0.03800  2.54  0.11119
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation structure = unstructured
Estimated Scale Parameters:

            Estimate Std.err
(Intercept)    0.658   0.783
  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2    0.380   0.471
alpha.1:3    0.393   0.489
alpha.1:4    0.356   0.447
alpha.1:5    0.296   0.372
alpha.2:3    0.748   0.851
alpha.2:4    0.289   0.369
alpha.2:5    0.466   0.541
alpha.3:4    0.407   0.517
alpha.3:5    0.677   0.795
alpha.4:5    0.446   0.558
Number of clusters:   159  Maximum cluster size: 5
```

Comparing $\hat{\beta}$s



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



GEEs are:

- Robust
- Flexible
- Extensible beyond panel/TSCS context

Addendum: Survival Analysis

- Models for *time-to-event data*.
- Roots in biostats/epidemiology, plus engineering, sociology, economics.
- Examples...
  - Political careers, confirmation durations, position-taking, bill cosponsorship, campaign contributions, policy innovation/adoption, etc.
  - Cabinet/government durations, length of civil wars, coalition durability, etc.
  - War duration, peace duration, alliance longevity, length of trade agreements, etc.
  - Strike durations, work careers (including promotions, firings, etc.), criminal careers, marriage and child-bearing behavior, etc.

Time-To-Event Data

Characteristics:

- *Discrete* events (i.e., not continuous),
- Take place over *time*,
- May not (or never) experience the event (i.e., possibility of *censoring*).

Terminology:

$$
\begin{aligned}
Y_i &= \text{the duration until the event occurs,} \\
Z_i &= \text{the duration until the observation is ``censored''} \\
T_i &= \min\{Y_i, Z_i\}, \\
C_i &= \text{0 if observation } i \text{ is } \textit{censored}\text{, 1 if it is not.}
\end{aligned}
$$

Density:

$$f(t) = \Pr(T_i = t)$$

CDF:

$$\Pr(T_i \leq t) \equiv F(t) = \int_0^t f(t)\,dt$$

Survival function:

$$\begin{aligned} \Pr(T_i \geq t) \equiv S(t) &= 1 - F(t) \\ &= 1 - \int_0^t f(t)\,dt \end{aligned}$$

Hazard:

$$\begin{aligned} \Pr(T_i = t | T_i \geq t) \equiv h(t) &= \frac{f(t)}{S(t)} \\ &= \frac{f(t)}{1 - \int_0^t f(t)\,dt} \end{aligned}$$

Grouped-Data Survival Approaches

Model:

$$\Pr(C_{it} = 1) = f\left(\mathbf{X}_{it}\beta\right)$$

Advantages:

- Easily estimated, interpreted and understood
- Natural interpretations:
  - $\hat{\beta}_0 \;\approx$ "baseline hazard"
  - Covariates shift this up or down.
- Can incorporate data on time-varying covariates
- Lots of software

Potential Disadvantages:

- Requires time-varying data
- Must deal with time dependence explicitly

Temporal Issues in Grouped-Data Models

(Implicit) "Baseline" hazard:

$$h_0(t) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$$

$\longrightarrow$ No temporal dependence / "flat" hazard

Time trend:

$$\Pr(Y_{it} = 1) = f(\mathbf{X}_{it}\beta + \gamma T_{it})$$

- $\hat{\gamma} > 0 \rightarrow$ rising hazard
- $\hat{\gamma} < 0 \rightarrow$ declining hazard
- $\hat{\gamma} = 0 \rightarrow$ "flat" (exponential) hazard

Variants/extensions: Polynomials...

$$\Pr(Y_{it} = 1) = f(\mathbf{X}_{it}\beta + \gamma_1 T_{it} + \gamma_2 T_{it}^2 + \gamma_3 T_{it}^3 + ...)$$

Temporal Issues in Grouped-Data Models

"Time dummies":

$$\Pr(Y_{it} = 1) = f[\mathbf{X}_{it}\beta + \alpha_1 I(T_{i1}) + \alpha_2 I(T_{i2}) + ... + \alpha_{t_{\max}} I(T_{it_{\max}})]$$

$\longrightarrow$ BKT's cubic splines; might also use:

- Fractional polynomials
- Smoothed duration
- Loess/lowess fits
- Other splines (B-splines, P-splines, natural splines, etc.)