PLSC 504: Fall 2020
Panel Data for Non-Continuous Responses (including GEEs)

October 14, 2020

One-way unit effects (logit):

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

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:

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

Chamberlain's Fixed Effects (continued)

Intuition:

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

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

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

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. Beck & Katz (2001).

Random Effects

Model is:

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

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?

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

Practical Things

- $\hat{\rho} =$ proportion of the variance due to the α_is.

- Implementation: Gauss-Hermite quadrature, or (better) MCMC.

- Best with N large and T small.

- Critically requires $\text{Cov}(\mathbf{X}, \alpha) = 0$ [but see: Chamberlain's "correlated random effects" (CRE) Estimator].

Unit Effects in Practice - Some Simulations

Start with:

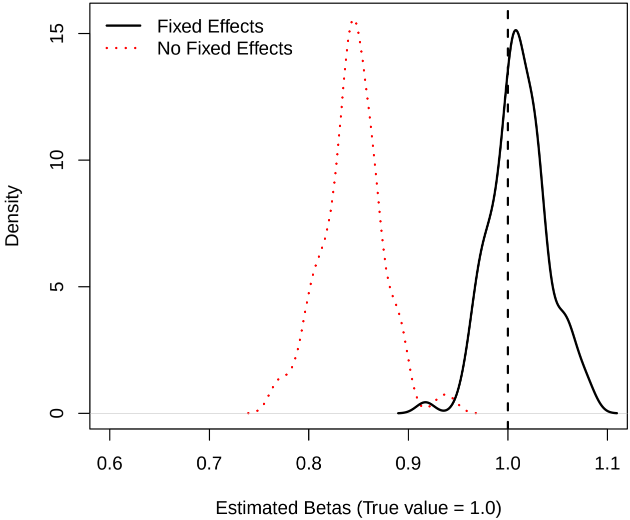
$$\begin{aligned} Y_{it}^* &= 0 + (1 \times X_{it}) + (1 \times D_{it}) + (1 \times \alpha_i) + u_{it} \\ Y_{it} \in \{0,1\} &= f(Y_{it}^*) \end{aligned}$$

where:

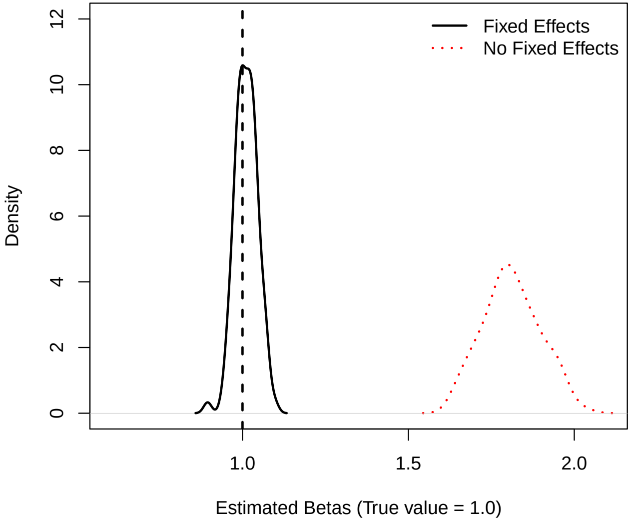
- $\alpha_i \sim N(0,1)$
- $X_{it} \sim N(0, \sigma_X^2)$
- $D_{it} \in \{0,1\}$
- $\text{Cov}(X_{it}, \alpha_i) = \{0, 0.69\}$
- $\text{Cov}(D_{it}, \alpha_i) = 0$
- $f(\cdot) = \{\text{logit, probit}\}$ (as appropriate)

and $N = T = 100$.

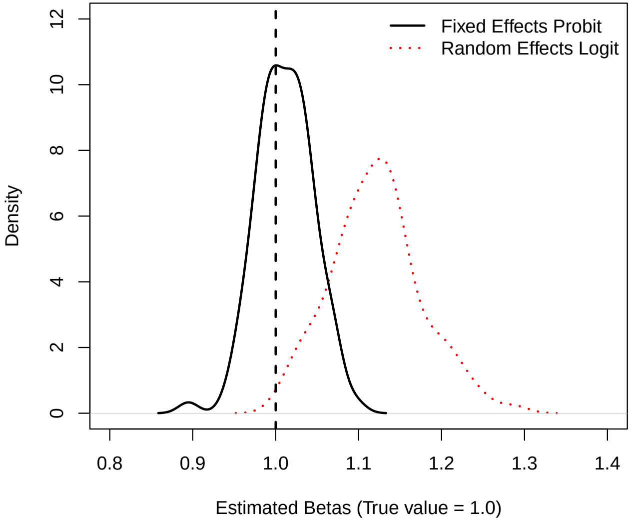
Logit $\hat{\beta}_X$s for $\mathrm{Cov}(X_{it}, \alpha_i) = 0$



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



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



Software

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).

Example: Segal (1986) Search & Seizure Cases

$Y = 1$ (search allowed)

- `warrant`: Whether (=1) or not (=0) a warrant was issued,
- `house`: Whether (=1) or not (=0) the search was of a private home,
- `person`: Whether (=1) or not (=0) the search was of a person,
- `business`: Whether (=1) or not (=0) the search was of a business,
- `car`: Whether (=1) or not (=0) the search was of an automobile,
- `us`: Whether (=1) or not (=0) the U.S. government was the petitioner,
- `except`: The number of "exceptions" outlined by the Court under which the search fell, and
- `justideo`: The justice's Segal-Cover (1989) ideology score, ranging from zero (most conservative) to 1 (most liberal).

$N = 14, \ \bar{T} = 74.1.$

```
> summary(Segal)
     justid          caseid          year          vote          warrant    
 Min.   : 1.0   Min.   :  1   Min.   :63   Min.   :0.00   Min.   :0.00  
 1st Qu.: 6.0   1st Qu.: 34   1st Qu.:69   1st Qu.:0.00   1st Qu.:0.00  
 Median : 8.0   Median : 64   Median :73   Median :1.00   Median :0.00  
 Mean   : 8.1   Mean   : 64   Mean   :73   Mean   :0.53   Mean   :0.15  
 3rd Qu.:11.0   3rd Qu.: 94   3rd Qu.:78   3rd Qu.:1.00   3rd Qu.:0.00  
 Max.   :14.0   Max.   :123   Max.   :81   Max.   :1.00   Max.   :1.00  
     house          person          business          car            us       
 Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   :0.0   Min.   :0.00  
 1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.:0.0   1st Qu.:0.00  
 Median :0.00   Median :0.00   Median :0.00   Median :0.0   Median :0.00  
 Mean   :0.23   Mean   :0.31   Mean   :0.15   Mean   :0.2   Mean   :0.45  
 3rd Qu.:0.00   3rd Qu.:1.00   3rd Qu.:0.00   3rd Qu.:0.0   3rd Qu.:1.00  
 Max.   :1.00   Max.   :1.00   Max.   :1.00   Max.   :1.0   Max.   :1.00  
     except         justideo   
 Min.   :0.00   Min.   :0.05  
 1st Qu.:0.00   1st Qu.:0.17  
 Median :0.00   Median :0.73  
 Mean   :0.35   Mean   :0.59  
 3rd Qu.:1.00   3rd Qu.:0.88  
 Max.   :3.00   Max.   :1.00  
```

```
> SegalLogit<-glm(vote~warrant+house+person+business+car+us+
                     except+justideo,data=Segal,family="binomial")
> summary(SegalLogit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3147  -0.9405   0.3898   0.9348   1.9032

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.9419     0.2799   6.938 3.97e-12 ***
warrant       0.5335     0.2083   2.561 0.010440 *
house        -1.0840     0.2756  -3.934 8.36e-05 ***
person       -0.9438     0.2569  -3.674 0.000239 ***
business     -1.4722     0.2975  -4.949 7.46e-07 ***
car          -1.0066     0.2816  -3.574 0.000351 ***
us            0.4824     0.1482   3.254 0.001136 **
except        0.8640     0.1384   6.243 4.29e-10 ***
justideo     -2.4026     0.2158 -11.134  < 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: 1434.9  on 1036  degrees of freedom
Residual deviance: 1196.7  on 1028  degrees of freedom
AIC: 1214.7

Number of Fisher Scoring iterations: 4
```

```
> library(bife)
> SegalFEL<-bife(vote~warrant+house+person+business+car+us+
+               except | justid,data=Segal,
+               model="logit")
> summary(SegalFEL)
binomial - logit link

vote ~ warrant + house + person + business + car + us + except |
    justid

Estimates:
         Estimate Std. error z value Pr(> |z|)
warrant     0.599      0.228    2.63   0.00866 **
house      -1.473      0.305   -4.82   1.4e-06 ***
person     -1.124      0.282   -3.99   6.7e-05 ***
business   -1.837      0.326   -5.63   1.8e-08 ***
car        -1.202      0.308   -3.90   9.6e-05 ***
us          0.537      0.162    3.32   0.00091 ***
except      1.093      0.155    7.03   2.1e-12 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

residual deviance= 1048,
null deviance= 1435,
nT= 1037, N= 14

Number of Fisher Scoring Iterations: 5

Average individual fixed effect= 0.665
```

Random Effects Logit

```
> SegalRE<-glmmML(vote~warrant+house+person+business+car+us+
                  except+justideo,data=Segal,family="binomial",
                  cluster=justid)
> summary(SegalRE)

Call:  glmmML(formula = vote ~ warrant + house + person + business +
car + us + except + justideo, family = "binomial", data = Segal, cluster = justid)

              coef se(coef)     z Pr(>|z|)
(Intercept)  2.016    0.565  3.57  3.6e-04
warrant      0.594    0.226  2.63  8.5e-03
house       -1.434    0.303 -4.73  2.2e-06
person      -1.104    0.280 -3.95  7.9e-05
business    -1.799    0.324 -5.56  2.7e-08
car         -1.181    0.306 -3.86  1.1e-04
us           0.531    0.160  3.31  9.3e-04
except       1.070    0.154  6.95  3.6e-12
justideo    -2.344    0.737 -3.18  1.5e-03

Scale parameter in mixing distribution:  0.926 gaussian
Std. Error:                              0.195

        LR p-value for H_0: sigma = 0:  4.63e-24

Residual deviance: 1100 on 1027 degrees of freedom  AIC: 1120
```

$\hat{\beta}$ Comparisons

```
> Bs
                 Logit     FEs     REs
(Intercept)  1.9419      NA  2.0164
warrant      0.5335  0.5992  0.5942
house       -1.0840 -1.4733 -1.4340
person      -0.9438 -1.1236 -1.1041
business    -1.4722 -1.8367 -1.7991
car         -1.0066 -1.2021 -1.1805
us           0.4824  0.5369  0.5312
except       0.8640  1.0926  1.0704
justideo    -2.4026      NA -2.3445
```

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

- No "incidental parameters" problem (see e.g. Cameron and Trivedi, pp. 281-2)
- Means "brute force" approach also works
- Can be fit via:
  - `pglm` (in `pglm`)
  - `feglm` (in `fixest`)
  - `glmmML`

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

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

Panel Models: Software

- Tobit = `censReg`
- Poisson (random effects) = `glmmML` or `glmer` or `pglm`
- Poisson (fixed effects) = `glmmML` or `pglm` or `fixest` or "brute force"
- Negative binomial = `pglm`

Example: State Failure Task Force

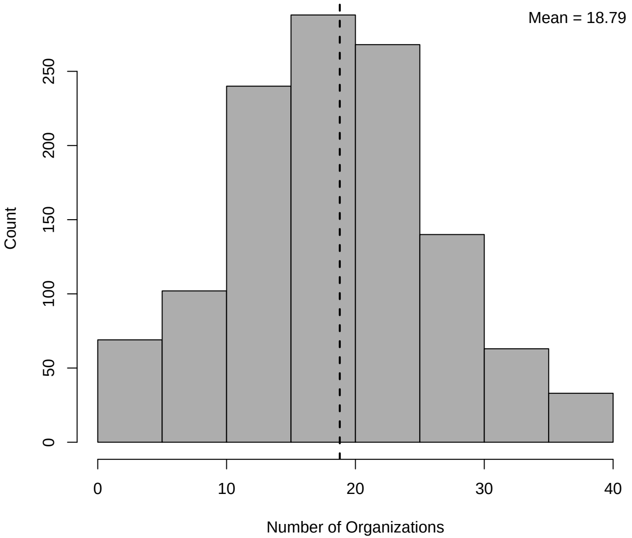
Project examining "state failure" (begun early 1990s)...

- $N \approx 170$ countries, measured at five-year intervals, 1957-1997 (so $T = 9$)
- Response variable: `CIOB`: The number of "type B" (universal-membership) intergovernmental organizations that country was a member of in that year ($\bar{Y} = 19$, range $=$ $\{0{,}38\}$).
- Predictors:
  - POLITY (autocracy / democracy) score (range [-10,10])
  - Percent of the population that is urban (`unuurbpc`)
  - The *durability* of the country's POLITY score: How many years has the POLITY score been the same? (`poldurab`)
  - A *trend* variable (1900=0)

State Failure Task Force Data

```
> summary(SFTF)
   countryid        year         sftprev        sftpeth         sftpreg     
 AFG    :   9   Min.   :1957   Min.   :0.0   Min.   :0.00   Min.   :0.00  
 ALB    :   9   1st Qu.:1967   1st Qu.:0.0   1st Qu.:0.00   1st Qu.:0.00  
 ARG    :   9   Median :1977   Median :0.0   Median :0.00   Median :0.00  
 AUL    :   9   Mean   :1979   Mean   :0.1   Mean   :0.13   Mean   :0.12  
 AUS    :   9   3rd Qu.:1992   3rd Qu.:0.0   3rd Qu.:0.00   3rd Qu.:0.00  
 BEL    :   9   Max.   :1997   Max.   :1.0   Max.   :1.00   Max.   :1.00  
 (Other):1149                                                             
    sftpgen        poldurab     unuurbpc         ciob          cioc     
 Min.   :0.00   Min.   : 0   Min.   :  2   Min.   : 0   Min.   : 0.0  
 1st Qu.:0.00   1st Qu.: 4   1st Qu.: 23   1st Qu.:14   1st Qu.: 2.0  
 Median :0.00   Median :12   Median : 41   Median :19   Median : 5.0  
 Mean   :0.08   Mean   :21   Mean   : 43   Mean   :19   Mean   : 5.6  
 3rd Qu.:0.00   3rd Qu.:30   3rd Qu.: 62   3rd Qu.:24   3rd Qu.: 8.0  
 Max.   :1.00   Max.   :97   Max.   :100   Max.   :38   Max.   :24.0  
                NA's   :5    NA's   :57                                
     POLITY        SumEvents  
 Min.   :-10.0   Min.   : 0  
 1st Qu.: -7.0   1st Qu.: 0  
 Median : -4.0   Median : 0  
 Mean   : -0.7   Mean   : 6  
 3rd Qu.:  8.0   3rd Qu.: 5  
 Max.   : 10.0   Max.   :61  
 NA's   :14      NA's   :9   

> pdim(SFTF)
Unbalanced Panel: n=170, T=1-9, N=1203
```

Distribution of Y



Basic Poisson

```
> Poisson<-glm(ciob~POLITY+unuurbpc+poldurab+I(year-1900),
+               data=SFTF,family="poisson")
> summary(Poisson)

Call:
glm(formula = ciob ~ POLITY + unuurbpc + poldurab + I(year -
    1900), family = "poisson", data = SFTF)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-7.204  -0.723   0.141   0.888   3.872

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.727987   0.046533   37.13  < 2e-16 ***
POLITY         0.010356   0.000982   10.55  < 2e-16 ***
unuurbpc       0.004864   0.000320   15.20  < 2e-16 ***
poldurab       0.002025   0.000295    6.87  6.5e-12 ***
I(year - 1900) 0.011826   0.000569   20.78  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 4453.9  on 1131  degrees of freedom
Residual deviance: 2906.6  on 1127  degrees of freedom
  (71 observations deleted due to missingness)
AIC: 8129

Number of Fisher Scoring iterations: 4
```

Poisson with Fixed Effects

```
> Poisson.FE<-pglm(ciob~POLITY+unuurbpc+poldurab+I(year-1900),
+                  data=SFTF,family="poisson",effect="individual",
+                  model="within",index="countryid")

> summary(Poisson.FE)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 3 iterations
Return code 1: gradient close to zero
Log-Likelihood: -2558
4  free parameters
Estimates:
                Estimate Std. error t value Pr(> t)
POLITY         -0.007437   0.001939   -3.84 0.00013 ***
unuurbpc        0.005011   0.001580    3.17 0.00151 **
poldurab       -0.000477   0.000749   -0.64 0.52386
I(year - 1900)  0.018411   0.001115   16.51 < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
--------------------------------------------
```

Poisson with Random Effects

```
> Poisson.RE<-glmer(ciob~POLITY+unuurbpc+poldurab+I(year-1900)+
+                     (1|countryid),data=SFTF,family="poisson")
> summary(Poisson.RE)
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: poisson  ( log )
Formula:
ciob ~ POLITY + unuurbpc + poldurab + I(year - 1900) + (1 | countryid)
   Data: SFTF

     AIC      BIC   logLik deviance df.resid
    6811     6841    -3399     6799     1126

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.569 -0.279  0.080  0.391  2.681

Random effects:
 Groups    Name        Variance Std.Dev.
 countryid (Intercept) 0.159    0.399
Number of obs: 1132, groups:  countryid, 160

Fixed effects:
                Estimate Std. Error z value    Pr(>|z|)
(Intercept)     1.200274   0.063085   19.03     < 2e-16 ***
POLITY         -0.003484   0.001812   -1.92       0.055 .
unuurbpc        0.005996   0.001064    5.64 0.000000017 ***
poldurab        0.001167   0.000672    1.74       0.082 .
I(year - 1900)  0.016385   0.000855   19.16     < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation of Fixed Effects:
            (Intr) POLITY unrbpc poldrb
POLITY       0.354
unuurbpc    -0.075 -0.139
poldurab     0.224  0.348 -0.087
I(yer-1900) -0.628 -0.273 -0.589 -0.313
convergence code: 0
Model is nearly unidentifiable: very large eigenvalue
 - Rescale variables?
```

Alternative Poisson w/ Random Effects

```
> Poisson.RE3<-pglm(ciob~POLITY+unuurbpc+poldurab+I(year-1900),
+                         data=SFTF,effect="individual",
+                         model="random",family="poisson",
+                         index="countryid")
> summary(Poisson.RE3)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 7 iterations
Return code 2: successive function values within tolerance limit
Log-Likelihood: -3391
6  free parameters
Estimates:
                Estimate Std. error t value Pr(> t)
(Intercept)     1.267122   0.060660   20.89 < 2e-16 ***
POLITY         -0.003455   0.001794   -1.93   0.054 .
unuurbpc        0.006147   0.001039    5.92 3.3e-09 ***
poldurab        0.001147   0.000660    1.74   0.082 .
I(year - 1900)  0.016342   0.000831   19.66 < 2e-16 ***
sigma           7.037765   0.891293    7.90 2.9e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
--------------------------------------------
```

$\hat{\beta}$ Comparisons: Poisson Models

```
> BPs <- data.frame(Poisson = Poisson$coefficients,
+                   FEs = c(NA,Poisson.FE$estimate),
+                   REs = Poisson.RE@beta,
+                   row.names = names(Poisson$coefficients))

> BPs
                 Poisson        FEs       REs
(Intercept)     1.727987         NA  1.200274
POLITY          0.010356 -0.0074369 -0.003484
unuurbpc        0.004864  0.0050111  0.005996
poldurab        0.002025 -0.0004774  0.001167
I(year - 1900) 0.011826  0.0184112  0.016385
```

$\hat{\beta}$ Comparisons: Negative Binomial Models (not shown)

```
> BNBs <- data.frame(NegBin = c(NB$coefficients,NB$theta),
+                   NegBinFEs = c(NA,NB.FE$coefficients),
+                   NegBinREs = c(NB.RE@beta,getME(NB.RE,"glmer.nb.theta")),
+                   row.names = c(names(NB$coefficients),"theta"))

> BNBs
                  NegBin    NegBinFEs      NegBinREs
(Intercept)      1.68694           NA        1.20027
POLITY           0.00941    -0.007442       -0.00348
unuurbpc         0.00525     0.005011        0.00600
poldurab         0.00180    -0.000477        0.00117
I(year - 1900)   0.01218     0.018416        0.01639
theta           12.59048 10000.000000 1082684.03067
```

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*
- `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

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

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

$\mathbf{V}_i = \text{Var}(Y_{it}|\mathbf{X}_{it}, \beta)$ 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 $\hat{\boldsymbol{S}}_i = (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)'$.

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

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

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

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)

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).

Example: President Bush (41) Approval

```
> url <- getURL("https://raw.githubusercontent.com/PrisonRodeo/PLSC504-2017-git/master/Data/Bush/
> Bush <- read.csv(text = url)
> summary(Bush)
      idno            year         approval          partyid            perfin
 Min.   :  1.0   Min.   :1990   Min.   :-2.0000   Min.   :-3.0000   Min.   :-2.00000
 1st Qu.:156.8   1st Qu.:1990   1st Qu.:-1.2500   1st Qu.:-2.0000   1st Qu.:-1.00000
 Median :312.5   Median :1991   Median : 1.0000   Median : 1.0000   Median : 0.00000
 Mean   :312.5   Mean   :1991   Mean   : 0.2302   Mean   : 0.3793   Mean   : 0.02724
 3rd Qu.:468.2   3rd Qu.:1992   3rd Qu.: 2.0000   3rd Qu.: 2.0000   3rd Qu.: 1.00000
 Max.   :624.0   Max.   :1992   Max.   : 2.0000   Max.   : 3.0000   Max.   : 2.00000
     nateco             age             educ            class          nonwhite
 Min.   :-2.0000   Min.   :18.00   Min.   :1.000   Min.   :1.000   Min.   :0.0000
 1st Qu.:-2.0000   1st Qu.:32.00   1st Qu.:3.000   1st Qu.:1.000   1st Qu.:0.0000
 Median :-1.0000   Median :41.00   Median :4.000   Median :4.000   Median :0.0000
 Mean   :-0.9797   Mean   :45.34   Mean   :4.048   Mean   :3.002   Mean   :0.1378
 3rd Qu.: 0.0000   3rd Qu.:59.00   3rd Qu.:6.000   3rd Qu.:4.000   3rd Qu.:0.0000
 Max.   : 2.0000   Max.   :85.00   Max.   :7.000   Max.   :6.000   Max.   :1.0000
     female
 Min.   :0.0000
 1st Qu.:0.0000
 Median :1.0000
 Mean   :0.5192
 3rd Qu.:1.0000
 Max.   :1.0000

 > pdim(Bush)
Balanced Panel: n=624, T=3, N=1872
```

GEE: Independence

```
> library(geepack)
> GEE.IND<-geeglm(approval~partyid+perfin+nateco+age+educ+class+nonwhite+female,
  data=Bush,id=idno,family=gaussian,corstr="independence")
> summary(GEE.IND)

 Coefficients:
             Estimate   Std.err    Wald Pr(>|W|)
(Intercept)  1.118752  0.165415  45.742 1.35e-11 ***
partyid     -0.317251  0.017570 326.032  < 2e-16 ***
perfin       0.118223  0.032527  13.211 0.000278 ***
nateco       0.360036  0.039828  81.719  < 2e-16 ***
age         -0.001526  0.002270   0.452 0.501292
educ        -0.048732  0.026603   3.355 0.066982 .
class       -0.035451  0.024571   2.082 0.149078
nonwhite    -0.287660  0.112827   6.500 0.010786 *
female      -0.011875  0.076408   0.024 0.876493
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)    1.839 0.05423

Correlation: Structure = independenceNumber of clusters: 624  Maximum cluster size: 3
```

Identical to GLM

```
> GLM <- glm(approval~partyid+perfin+nateco+age+educ+class+nonwhite+female,
             data=Bush,family=gaussian)

> # Coefficients:
> cbind(GEE.IND$coefficients,GLM$coefficients)
                [,1]     [,2]
(Intercept)  1.11875  1.11875
partyid     -0.31725 -0.31725
perfin       0.11822  0.11822
nateco       0.36004  0.36004
age         -0.00153 -0.00153
educ        -0.04873 -0.04873
class       -0.03545 -0.03545
nonwhite    -0.28766 -0.28766
female      -0.01188 -0.01188

> # Standard Errors:
> cbind(sqrt(diag(GEE.IND$geese$vbeta.naiv)),sqrt(diag(vcov(GLM))))
               [,1]    [,2]
(Intercept) 0.13827 0.13861
partyid     0.01615 0.01619
perfin      0.02963 0.02970
nateco      0.03857 0.03866
age         0.00193 0.00194
educ        0.02148 0.02153
class       0.02066 0.02071
nonwhite    0.09477 0.09500
female      0.06356 0.06371
```

```
> GEE.EXC<-geeglm(approval~partyid+perfin+nateco+age+educ+class+nonwhite+female,
  data=Bush,id=idno,family=gaussian,corstr="exchangeable")
> summary(GEE.EXC)

 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  1.14375  0.16592  47.52  5.4e-12 ***
partyid     -0.31881  0.01738 336.60  < 2e-16 ***
perfin       0.10193  0.03195  10.18   0.0014 **
nateco       0.32912  0.03964  68.94  < 2e-16 ***
age         -0.00262  0.00228   1.32   0.2512
educ        -0.05096  0.02669   3.65   0.0562 .
class       -0.03311  0.02471   1.80   0.1803
nonwhite    -0.29156  0.11374   6.57   0.0104 *
female      -0.01596  0.07687   0.04   0.8356
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     1.84  0.0542

Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.232  0.0275
Number of clusters:   624   Maximum cluster size: 3
```

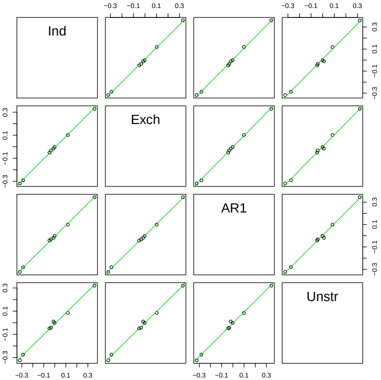
GEE: AR(1)

```
> GEE.AR1<-geeglm(approval~partyid+perfin+nateco+age+educ+class+nonwhite+female,
  data=Bush,id=idno,family=gaussian,corstr="ar1")
> summary(GEE.AR1)

 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  1.03609  0.16610  38.91  4.4e-10 ***
partyid     -0.32297  0.01736 346.07  < 2e-16 ***
perfin       0.09890  0.03186   9.64   0.0019 **
nateco       0.34337  0.03967  74.94  < 2e-16 ***
age         -0.00191  0.00229   0.70   0.4038
educ        -0.04255  0.02658   2.56   0.1094
class       -0.03270  0.02488   1.73   0.1888
nonwhite    -0.28120  0.11208   6.29   0.0121 *
female      -0.01873  0.07690   0.06   0.8075
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     1.84  0.0543

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.285  0.0303
Number of clusters:   624   Maximum cluster size: 3
```

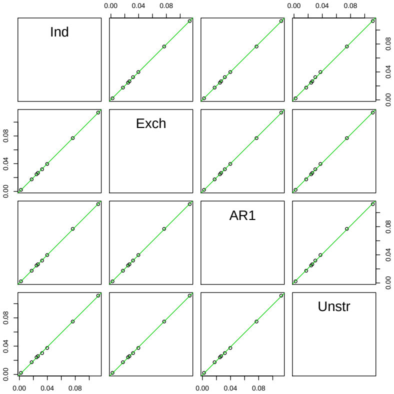
GEE: Unstructured

```
> GEE.UNSTR<-geeglm(approval~partyid+perfin+nateco+age+educ+class+nonwhite+female,
  data=Bush,id=idno,family=gaussian,corstr="unstructured")
> summary(GEE.UNSTR)

 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept)  1.00139  0.16016  39.09    4e-10 ***
partyid     -0.32372  0.01724 352.37   <2e-16 ***
perfin       0.08457  0.03017   7.86   0.0051 **
nateco       0.31947  0.03741  72.94   <2e-16 ***
age         -0.00111  0.00220   0.26   0.6135
educ        -0.04884  0.02586   3.57   0.0589 .
class       -0.04235  0.02421   3.06   0.0803 .
nonwhite    -0.27429  0.11139   6.06   0.0138 *
female       0.01041  0.07479   0.02   0.8893
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     1.85  0.0542

Correlation: Structure = unstructured  Link = identity

Estimated Correlation Parameters:
          Estimate Std.err
alpha.1:2  0.51573  0.0371
alpha.1:3  0.18614  0.0407
alpha.2:3  0.00277  0.0400
Number of clusters:   624   Maximum cluster size: 3
```

```
> betas<-cbind(GEE.IND$coefficients,GEE.EXC$coefficients,GEE.AR1$coefficients,
  GEE.UNSTR$coefficients)
> library(car)
> scatterplotMatrix(betas[-1,],smooth=FALSE,var.labels=c("Ind","Exch","AR1","Unstr"),
  diagonal="none")
```



```
> ses<-cbind(sqrt(diag(GEE.IND$geese$vbeta)),sqrt(diag(GEE.EXC$geese$vbeta)),
  sqrt(diag(GEE.AR1$geese$vbeta)),sqrt(diag(GEE.UNSTR$geese$vbeta)))
> scatterplotMatrix(ses[-1,],smooth=FALSE,var.labels=c("Ind","Exch","AR1","Unstr"),
  diagonal="none")
```



GEEs are:

- Robust
- Flexible
- Extensible beyond panel/TSCS context