

GSERM - Oslo 2018

Generalized Estimating Equations

January 17, 2018 (morning session)

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

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

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

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

Define:

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

→ “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

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

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

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

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

Independent:

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

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

Exchangeable:

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

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

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

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

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

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

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

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

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

Score equations:

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

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):

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

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

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

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

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

Inference (aka, magic!)

- $\hat{\Sigma}_{\text{Model}}$
 - 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.

Use $\hat{\Sigma}_{\text{Robust}}$.

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 β s,
- Account for within-unit correlation in an informed way, but also
- Provide 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, $\hat{\beta}_M \approx \frac{\hat{\beta}_C}{\sqrt{1+0.35\sigma_\eta^2}}$, where $\sigma_\eta^2 > 0$ is the variance of the unit effects.

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

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

Substantive interest in $\mathbf{R}_i(\boldsymbol{\alpha})$ (e.g., Prentice 1988)?

Add:

$$\mathbf{U}_{GEE}(\boldsymbol{\alpha}) = \sum_{i=1}^N \mathbf{E}_i' \mathbf{W}_i^{-1} (\mathbf{Z}_i - \boldsymbol{\eta}_i)$$

where

- $\mathbf{E}_i = \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\alpha}}$,
- \mathbf{W}_i is the “working” VCV matrix for the \mathbf{Z}_i s,
- $\mathbf{Z}_i' = (Z_{i12}, Z_{i13}, \dots, Z_{iT-1, T-1})$ are the $\frac{T(T-1)}{2}$ observed sample pairwise correlations for i , and
- $\boldsymbol{\eta}_i$ is a vector of expected values for \mathbf{Z}_i *which may include covariates*.

Independently from $U_{GEE}(\beta)$:

$$\mathbf{U}_{GEE}(\alpha, \beta) = \sum_{i=1}^N \begin{pmatrix} \mathbf{D}'_i & \mathbf{0} \\ \mathbf{0} & \mathbf{E}'_i \end{pmatrix} \begin{pmatrix} \mathbf{V}_i^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_i^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \mu_i \\ \mathbf{Z}_i - \eta_i \end{pmatrix}$$

or allowing the two to covary:

$$\mathbf{U}_{GEE}(\alpha, \beta) = \sum_{i=1}^N \begin{pmatrix} \mathbf{D}'_i & \mathbf{0} \\ \mathbf{F}'_i & \mathbf{E}'_i \end{pmatrix} \begin{pmatrix} \mathbf{V}_i^{-1} & \text{Cov}(\mathbf{Y}_i, \mathbf{W}_i) \\ \text{Cov}(\mathbf{W}_i, \mathbf{Y}_i) & \mathbf{W}_i^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \mu_i \\ \mathbf{Z}_i - \eta_i \end{pmatrix}$$

where $\mathbf{F}_i = \frac{\partial \alpha_i}{\partial \beta}$.

- Allows simultaneous modeling of first and second moments.
- Conditional on proper specification, $\hat{\beta}_{GEE2S}$ are somewhat more efficient than $\hat{\beta}_{GEES}$.
- Model (1) requires specification of third and fourth moments.
- Many (e.g. Diggle et al.) suggest using $\mathbf{W}_i = \frac{\mathbf{I}}{m \times m}$.
- Biggest drawback: *Requires correct specification of $\mathbf{R}_i(\alpha)$ for consistent estimates of $\hat{\beta}$.*
- Software is somewhat limited (EE, MAREG/WinMAREG, geepack, orth, possibly SASTM).

Software	Command(s)/Package(s)
Stata	xtgee / xtlogit / xtprobit / xtpois / etc.
R	gee / geepack / multgeeB / orth / repolr
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).

Example: President Bush (41) Approval

```
> url <- getURL("https://raw.githubusercontent.com/PrisonRodeo/GSERM-Oslo-2018-git/master/1")
> 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: Exchangeable

```
> 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.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

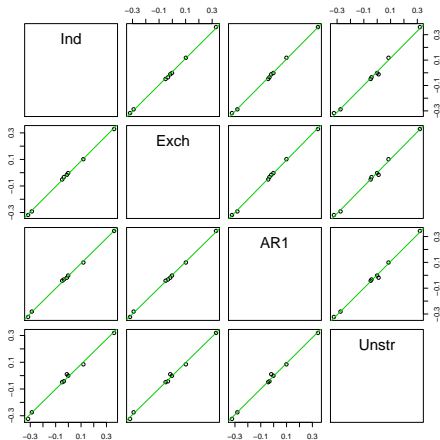
Estimated Correlation Parameters:

	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

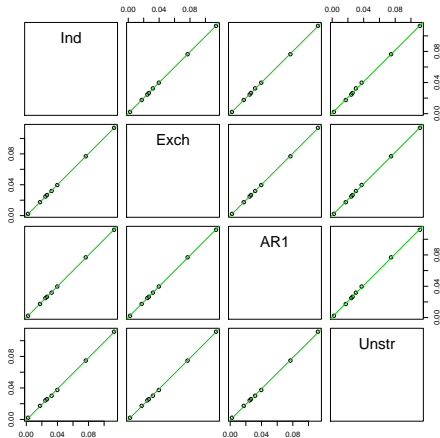
Comparing $\hat{\beta}$ s

```
> 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")
```



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

```
> 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