# Marginal Linear Regression Models

*Brady T. West*

# Marginal Modeling Approaches

- Researcher may not care about estimating between-cluster variance in slopes or intercepts; only *interested in marginal fixed effects across all higher level clusters*!
- Researcher only *interested in overall marginal relationships*, possibly for different subgroups
- Alternative methods for fitting **marginal models** ~ focus on **Generalized Estimating Equations (GEE)**

# Marginal Linear Regression Models

- When fitting marginal models to **normal** dependent variables, estimate parameters defined by this model ***(no random effects)***

$$y_{ti} = \beta_0 + \beta_1 x_{1ti} + \ldots + \beta_p x_{pti} + e_{ti}$$

$$y_i = (y_{1i}, y_{2i}, \ldots, y_{n_i i})' \sim N(X_i \beta, V_i)$$

- General variance-covariance matrix $V_i$: many possible structures
- **Do not explicitly estimate variances and covariances of random cluster effects *(because there are none)*!**

# Generalized Estimating Equations

When fitting models to dependent data using GEE → seek estimates of parameters that solve **score function** (or **estimating equation**):

$$S(\beta) = \sum_{i=1}^{n} D_i^T V_i^{-1} (y_i - \mu_i) = 0$$

mean

$D_i^T$ is $n_i$ x $p$ matrix with $(i,j)$th elements $\dfrac{\partial \mu_{ti}}{\partial \beta}$

$V_i$ is $n_i$ x $n_i$ variance-covariance matrix for observations on cluster $i$

*(user-specified)*

# Generalized Estimating Equations

$$S(\beta) = \sum_{i=1}^{n} D_i^T V_i^{-1} (y_i - \mu_i) = 0$$

$y_i$ = vector *(column)* of outcome measures collected for *i*-th cluster

$\mu_i$ = vector of expected means for outcome measures **based on specified model** *(<u>fixed effects</u> enter here … regression function defines mean!)*

Equations solved: Iteratively weighted least squares or Fisher scoring algorithms

# Generalized Estimating Equations

- GEE methodology first introduced by Liang and Zeger in seminal 1986 paper
- Seldom know true variance-covariance matrix for observations
  In practice → **working correlation matrix** *("plausible guess" for true structure)*
- Typically specify correlations and ***not*** covariances …
  for non-normal outcomes, variances defined by mean structure of model
- **Allows methodology to be easily adapted to non-normal dependent variables** (Poisson, Binomial, etc.)

# Generalized Estimating Equations

Variances of parameter estimates computed using **sandwich estimator**
*(Liang and Zeger, 1986)*

- Estimator based on specified working correlation matrix (e.g., exchangeable), and variance-covariance matrix of observations based on expected means from fitted model

- Estimators of fixed effect parameters are **consistent** even if specified working correlation matrix is incorrect; *bad choices affect standard errors*

# Generalized Estimating Equations

- Choices for working correlation matrix in models fitted using GEE:
  - **Independence** (zero correlation, independent observations)
  - **Exchangeable** (constant correlation of observations in the same cluster)
  - **AR(1)** (first-order auto-regressive, decaying correlation over time)
  - **Unstructured** (completely general correlations of observations)

- Estimators of correlations proposed by Liang and Zeger (1986)
- Focus on fixed effects, ***not*** inferences about "nuisance" correlations

# Generalized Estimating Equations

- Inference "robust" to possible variance-covariance matrix misspecification possible when using **Wald Tests** based on sandwich estimates of standard errors
- Adapted information criterion (QIC) to compare fits of competing models
- Good specification of mean structure *(interactions, non-linear terms, etc.)* important!
- With large data sets, alternative choices of working correlation matrix do not make large difference, but should still be considered

# Generalized Estimating Equations

- Poor choices of working correlation matrix affect standard errors and inferences $\longrightarrow$ try to choose reasonable model based on information criteria
- **QIC** *(popular and widely implemented)* has important shortcomings in choosing "correct" correlation structure
- Westgate and Burchett (2017) describe **better alternatives** for identifying correct correlation structure, and provide software to both fit GEE models and compute advocated information criteria

# Revisiting the ESS Model

- Interested in relationship of trust in police (TRSTPLC, IV) with a person's attitude about whether people generally try to help others (PPLHLP, DV)

- Observations clustered by interviewer, ignored dependency in data in Week 2! **GEE can account for this.**

- Marginal linear regression model using GEE (with **exchangeable** correlation structure, assuming constant correlation of observations within interviewer) to make inference about **overall, marginal relationship** between two variables!

# Revisiting the ESS Model

GEE estimate of fixed effect of TRSTPLC is positive (0.04) and **not significant** ($p = 0.054$); estimated intercept is 4.47 ($p < 0.01$)

Estimates suggest similar direction for relationship, but marginal estimate is 1/3 of the multilevel model (0.12), when explicitly controlling for random interviewer effects

Multilevel model:
"**for a given interviewer**
one-unit increase in trust in police
leads to 0.12 increase in helpfulness"

Marginal model:
"**across all interviewers**
one-unit increase …
0.04 increase in helpfulness"

# Model Diagnostics

- Assuming constant correlation within interviewers, the “nuisance” estimate of the correlation was 0.05; QIC = 6790.61
- Unstructured and first-order autoregressive correlation structures don’t make sense; no time ordering of cross-sectional observations within each interviewer

**What about independence?** QIC = 6791.55

**There is slight evidence of a better fit for the exchangeable model ~** important to account for correlation!

# Conclusions from the Example

- Marginally, when looking at overall relationship across interviewers, did not find as much evidence of relationship as did when controlling for interviewer effects explicitly
- Accounting for dependency *(rather than assuming independence of observations within each interviewer)* did seem to improve model fit slightly; **comparing models is important!**

**Remember**

When fitting marginal models, can no longer make inference about between-cluster variance!

# What's Next?

- Marginal **logistic** regression models or binary outcomes
  ~ easily fitted using GEE.
- **Revisit** example of **predicting probability** of **ever smoking 100 cigarettes in one's lifetime**
  ~ see what changes compared to multilevel modeling approach