A Mixed-Effects Estimating Equation Approach to Nonignorable Missing Longitudinal Data

Xuan Bi and Annie Qu*

Department of Statistics, University of Illinois at Urbana-Champaign

Census of Bureau, Washington, D.C., Nov. 13, 2015

Missing Data Mechanisms

- Missing mechanisms can be classified as (Rubin, 1976):
  - Missing completely at random (MCAR)
  - Missing at random (MAR)
    - *Missingness depends on observed values*
    - Weighted GEE (Robins et al., 1995; Rotnitzky et al., 1998)
    - Multiple imputation (Rubin, 1987; Paik, 1997; Fitzmaurice, Laird and Ware, 2011)
    - Both are consistent under MAR
  - **Missing not at random (MNAR)**
    - *Missingness depends on un-observed values*

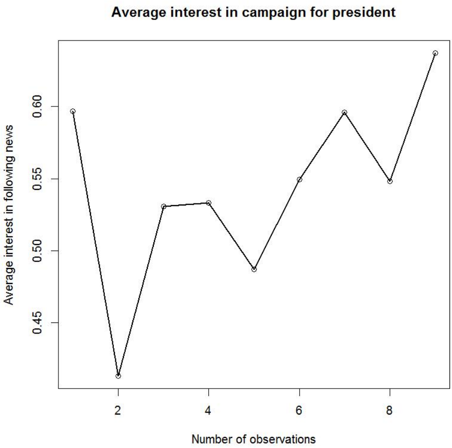
Existing Approaches for Nonignorable Missing Data

- When missing is nonignorable, additional assumptions or sensitivity analysis might be required (Robins, 1997)
    - Likelihood methods (Diggle and Kenward, 1994; Ibrahim et al., 2001)
    - Semiparametric methods (Scharfstein et al., 1999; Kim and Yu, 2011)
    - Mixed-effects models (Tsonaka et al., 2009; Shao and Zhang, 2015)

A Motivating Example: Presidential Election Survey Data

- 4719 survey respondents were measured their interests in presidential election for 9 waves (2007-2008, Associated Press-Yahoo! News Poll)
- Responses are **missing intermittently** with a missing rate of **49.7%**
- 1990 respondents were **refreshment samples** recruited in wave 3, 5, 6, or 9
- Predictors include Time, Gender, Race, Age, Education, Income, Marital Status and Location

A Motivating Example: Presidential Election Survey Data



A Motivating Example: Presidential Election Survey Data

- Respondents with higher interest tend to participate more
- Last-wave recruits show a significantly higher interest (p-value$= 8.56 \times 10^{-10}$)
- Data could be missing not at random
- **Goals:** Correct estimation bias

Shared Parameter Models for Nonignorable Missingness

- Introduce shared parameter model (Wu and Carroll, 1988)
- **Model Assumption:**
  - The response $\mathbf{y}$ and the missing indicator $\boldsymbol{\delta}$ are independent given the random effect $\mathbf{b}$:

$$\mathbf{y} \perp\!\!\!\perp \boldsymbol{\delta} | \mathbf{b}$$

- Parametric assumptions on $\mathbf{b}$ difficult to verify
- Existing approaches require a full or partial likelihood formulation
- Restrictive: requiring a dropout missing pattern
- Intensive computation involving high-dimensional integration or sampling procedures

The Proposed Method

- **Basic Ideas**:
  - Estimating equations utilizing **unspecified** random effects
- **Properties**:
  - No parametric assumptions on random effects
  - Non-monotone missing pattern (without baseline observations)
  - Correlated errors (serial correlation)

Notations

- $\mathbf{y}_i$ is an $n_i \times 1$ observed response vector, $i = 1, \dots, N$
- X_i is an $n_i \times p$ fixed-effects covariates matrix
- $\boldsymbol{\beta}$ is a $p \times 1$ fixed-effects parameter
- Z_i is an $n_i \times q$ random-effects covariates matrix; usually a subset of X_i
- $\mathbf{b} = (\mathbf{b}_1', \dots, \mathbf{b}_N')'$ is the random-effects parameter, with each $\mathbf{b}_i$ a $q \times 1$ random effect for subject i
- $\mathrm{E}(\mathbf{y}_i | \mathbf{b}_i) = \boldsymbol{\mu}(X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i) = \boldsymbol{\mu}_i$

Penalized Conditional Quasilikelihood

- The conditional quasi-likelihood of y given the random effects b is $l_q^b = -\frac{1}{2\phi}\sum_{i=1}^N d_i(y_i, \mu_i^b)$, where $d_i(y, u) = -2\int_y^u \frac{y-u}{a_i v(u)} du$
- Impose a constraint to ensure identifiability: $P_A b = 0$
- P_A is the projection matrix on the null space of $(I - P_X)Z$ where X and Z are the design matrices for fixed and random effects respectively
- Penalized conditional quasilikelihood (Jiang, 1999)

$$l_q = -\frac{1}{2\phi}\sum_{i=1}^N d_i(y_i, \mu_i^b) - \frac{1}{2}\lambda|P_A b|^2$$

Conditional Extended Score Corresponding for β and b

- Take the derivatives of the penalized conditional quasilikelihood l_q corresponding to β and b
- The quasi-score equation corresponding to the fixed effect β is

$$\sum_{i=1}^{N}(\frac{\partial \mu_i}{\partial \beta})'(\mathrm{W}_i)^{-1}(y_i - \mu_i^b) = 0.$$

- The quasi-score equation corresponding to the random effects b is

$$\begin{pmatrix} h_1 & = & (\frac{\partial \mu_1}{\partial b_1})'(\mathrm{W}_1^b)^{-1}(y_1 - \mu_1^{b_1}) - \lambda \frac{\partial P_A b}{\partial b_1} P_A b = 0 \\ & & \vdots \\ h_N & = & (\frac{\partial \mu_i}{\partial b_N})'(\mathrm{W}_N^b)^{-1}(y_N - \mu_N^{b_N}) - \lambda \frac{\partial P_A b}{\partial b_N} P_A b = 0 \end{pmatrix}$$

Moment Conditions for Fixed Effects

- Construct the moment conditions for fixed-effects β conditional on $\mathbf{b}$,

$$\mathbf{G}_N^f = \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i^f(\beta) = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^N (\frac{\partial\mu_i}{\partial\beta})' A_i^{-1/2} M_1 A_i^{-1/2}(y_i-\mu_i) \\ \vdots \\ \sum_{i=1}^N (\frac{\partial\mu_i}{\partial\beta})' A_i^{-1/2} M_m A_i^{-1/2}(y_i-\mu_i) \end{pmatrix}.$$

  where A_i is marginal variance of $\mathbf{y}_i$, and M_j's are basis matrix representations of the empirical correlations

- Conditional on $\mathbf{b}$,

$$\hat{\beta} = \arg\min(\bar{\mathbf{G}}_N^f)'(\bar{C}_N^f)^{-1}(\bar{\mathbf{G}}_N^f)$$

  where $\bar{C}_N^f = (1/N)\sum \mathbf{g}_i^f(\beta)\mathbf{g}_i^f(\beta)'$

Moment Conditions for Random Effects

- The random effect $\mathbf{b}$ is considered as a realization of a random process

- Construct the moment conditions for $\mathbf{b}$:

$$\mathbf{G}^r(\mathbf{b}) = \{(\mathbf{g}_1^r)', \ldots, (\mathbf{g}_N^r)', \lambda_1 \mathbf{b}', \lambda_2 (\mathbf{P}_{\mathbf{A}} \mathbf{b})'\}',$$

  where $\mathbf{g}_i^r = (\frac{\partial \mu_i}{\partial b_i})' \mathrm{C}^{-1}(y_i - \mu_i)$ and $C = \hat{\mathrm{Var}}(\mathbf{y}|\mathbf{b})$

- Estimate β and $\mathbf{b}$ by iteratively minimizing $(\bar{\mathbf{G}}_N^f)'(\bar{C}_N^f)^{-1}(\bar{\mathbf{G}}_N^f)$ and $(\mathbf{G}^r)'(\mathbf{G}^r)$

Tuning Parameters Selection

- λ_1 penalizes the variance of $\mathbf{b}$ to control the variance magnitude and assure convergence
- λ_2 penalizes the mean of $\mathbf{b}$ to assure identifiability
- Use a generalized cross validation to tune λ_1
- λ_2 is not critical and is fixed to be $\log(n)$

Theoretical Properties

- If $\mathbf{y} \perp\!\!\!\perp \boldsymbol{\delta}|\mathbf{b}$ then $\mathrm{E}(\mathbf{g}_i^f|\mathbf{b}) = \mathbf{0}$
- Consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$
- No need to estimate $\boldsymbol{\delta}$
- However, $\mathbf{y} \perp\!\!\!\perp \boldsymbol{\delta}|\hat{\mathbf{b}}$ is not necessarily true as $\hat{\mathbf{b}}$ might not be a consistent estimator of $\mathbf{b}$
- If $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbf{g}_i^f(\boldsymbol{\beta}|\hat{\mathbf{b}}) \to \mathbf{0}$, then we have consistency and asymptotic normality of $\boldsymbol{\beta}$, given $\hat{\mathbf{b}}$.

Consistency under Shared Parameter Model Assumption

Theorem

If:

1) $\mathbf{y} \perp\!\!\!\perp \delta | \mathbf{b}$
2) $\frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i^f(\boldsymbol{\beta}|\hat{\mathbf{b}}) \to \mathbf{0}$ *as* $N \to \infty$,

then conditional on $\hat{\mathbf{b}}$,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(\frac{1}{\sqrt{N}}),$$

and

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(0, \Sigma).$$

Assumption Relaxation

- Shared parameter model assumption $\mathbf{y} \perp\!\!\!\perp \boldsymbol{\delta}|\mathbf{b}$ is a concept of "conditional MCAR"
- However, random effects alone may not capture the missing mechanism
- "Conditional MAR": allow the observed response to carry out information

$$\boldsymbol{\delta}|(\mathbf{b}, \mathbf{y}) = \boldsymbol{\delta}|\mathbf{b} \Rightarrow \boldsymbol{\delta}|(\mathbf{b}, \mathbf{y}) = \boldsymbol{\delta}|(\mathbf{b}, \mathbf{y}^o)$$

where $\mathbf{y} = (\mathbf{y}^o, \mathbf{y}^m)$, with $\mathbf{y}^o$ the observed values and $\mathbf{y}^m$ the missing values

Assumption Relaxation

- Linear Conditional Mean assumption (Qu et al., 2010)

$$E(\mathbf{y}^m|\mathbf{b}, \mathbf{y}^o) \text{ is linear in } \mathbf{y}^o$$

- An idea of first-order expansion
- Holds true for normal responses, and approximately true for binary or ordinal responses

Consistency without Shared Parameter Model Assumption

Theorem

If:

1) $E(\mathbf{y}^m|\mathbf{b}, \mathbf{y}^o)$ *is linear in* $\mathbf{y}^o$
2) $\frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i^f(\boldsymbol{\beta}|\hat{\mathbf{b}}) \to \mathbf{0}$ *as* $N \to \infty$,

then conditional on $\hat{\mathbf{b}}$,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = O_p(\frac{1}{\sqrt{N}}),$$

and

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(0, \Sigma).$$

Simulation Study 1

- Sample size $N = 100$, the largest cluster size $T = 10$
- Fixed effects: intercept, treatment, time and interaction, $\boldsymbol{\beta} = (-1, 1, 1, 0.8)'$
- Random effects: $\mathbf{z}_{it} = (1, x_{i3})$, $b_{i1} \overset{iid}{\sim} 2 \times \text{Beta}(0.5, 0.5) - 1$ and $b_{i2} \overset{iid}{\sim} \text{N}(0, 0.5^2)$; b_{i1} is bimodal
- The missing pattern is **intermittent missing**
- The missing rate is about **45%**

Simulation Study 1

- **The measurement process**:
  - y_{it}'s are binary responses, with probability of being 1:

$$\text{logit}(p_{it}^{\mu}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i$$

  - $\text{Cor}(\mathbf{y}_i|\mathbf{b}_i)$ is AR-1 with $\rho = 0.2, 0.6$

- **The missing process**:
  - $\delta_{it} = 1$ if subject i is observed at time t, and 0 otherwise
  - $\text{logit}\{\text{P}(\delta_{it} = 1)\} = \boldsymbol{\gamma}'\mathbf{b}_i - t/T + 0.5$, where $\boldsymbol{\gamma} = (2, 1.5)'$

Simulation Study 1

Table: Absolute bias and RMSE of fixed-effects estimation when responses are binary with an AR1 correlation structure

		MEEE	PQL	WGEE	MI
$\rho = 0.2$	β_0	$0.39_{0.42}$	$0.30_{0.35}$	$1.04_{1.77}$	$0.55_{0.56}$
	β_1	$0.21_{0.26}$	$0.24_{0.32}$	$2.04_{3.44}$	$0.57_{0.59}$
	β_2	$0.16_{0.21}$	$0.30_{0.39}$	$1.22_{2.03}$	$0.21_{0.24}$
	β_3	$0.22_{0.29}$	$0.28_{0.36}$	$1.94_{3.23}$	$0.51_{0.53}$
$\rho = 0.6$	β_0	$0.38_{0.43}$	$0.50_{0.65}$	$0.92_{1.41}$	$0.48_{0.50}$
	β_1	$0.29_{0.36}$	$0.99_{1.38}$	$0.87_{1.25}$	$0.53_{0.57}$
	β_2	$0.18_{0.23}$	$1.39_{1.82}$	$0.85_{1.66}$	$0.23_{0.26}$
	β_3	$0.27_{0.34}$	$1.02_{0.42}$	$0.95_{1.34}$	$0.52_{0.55}$

MEEE, the proposed mixed-effects estimating equation; PQL, penalized quasi-likelihood (Breslow and Clayton, 1993); WGEE, weighted generalized estimating equation (Robins et al., 1995); MI, multiple imputation (Fitzmaurice et al., 2011); Monotonized data are used for WGEE.

Simulation Study 2

- Responses are **count data**
- Shared parameter model assumption is **violated**
- The missing pattern is **dropout**
- The missing rate is about **35%**

Simulation Study 2

- **The measurement process**:
  - Each $\mathbf{y}_i$ follows a correlated Poisson distribution with $\rho = 0, 0.4, 0.7$ and:

$$\log(\lambda_{it}) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{it},$$

where $x_{it} \overset{iid}{\sim} \mathrm{N}(0,1)$, $(\beta_0, \beta_1)' = (0, 0.8)'$ and $b_{ij} \overset{iid}{\sim} \mathrm{N}(0, 0.2^2)$

- **The missing process**:
  - $\delta_{it} = 1$ if subject i is observed at time t, and 0 otherwise.
  - $\text{logit}(p_{it}^{\delta}) = \begin{cases} (y_{i,t-1} + y_{it} + y_{i,t+1}) - 1.5 & \text{if } t < T \\ 1.5 \cdot (y_{i,t-1} + y_{it}) - 3 & \text{if } t = T \end{cases}$
  - $N = 100$ and $T = 5$, assume $\delta_{it} = \cdots = \delta_{iT} = 0$ if $\delta_{i,t-1} = 0$

Simulation Study 2

Table: MSE of fixed-effects estimation when responses are count data with an AR1 correlation structure

		MEEE	PQL	WGEE	MI	DR
$\rho = 0.0$	β_0	0.163	0.178	0.234	0.339	0.563
	β_1	0.085	0.062	0.103	0.163	0.459
$\rho = 0.4$	β_0	0.192	0.334	0.293	0.344	0.494
	β_1	0.126	0.196	0.130	0.099	0.370
$\rho = 0.7$	β_0	0.202	0.663	0.623	0.467	1.011
	β_1	0.185	0.357	0.247	0.126	0.461

MEEE, the proposed mixed-effects estimating equation; PQL, penalized quasi-likelihood (Breslow and Clayton, 1993); WGEE, weighted generalized estimating equation (Robins et al., 1995); MI, multiple imputation (Fitzmaurice et al., 2011); DR, doubly-robust generalized estimating equation (Seaman and Copas, 2009).

Presidential Election Survey Data

- A random intercept model:

$$\text{logit}(\mu_{it}) = X_{it}\boldsymbol{\beta} + b_i, t = 1, \ldots, n_i, i = 1, \ldots, N$$

- X_{it} is a p-dimensional covariate
- $N = 4719$, $p = 11$, and the largest cluster size is $T = 9$

Presidential Election Survey Data

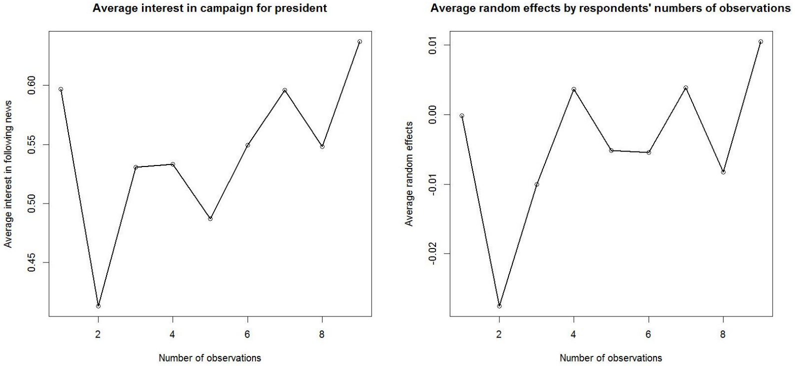
Table: Estimates and *p*-value of fixed effects

Method	MEEE	PQL	MI	WGEE*	DR*
Intercept	$-4.35_{0.00}$	$-5.37_{0.00}$	$-2.40_{0.00}$	$-3.19_{0.00}$	$-2.86_{0.01}$
Time	$0.11_{0.00}$	$0.23_{0.00}$	$0.11_{0.00}$	$0.11_{0.00}$	$0.11_{0.00}$
Age	$0.04_{0.00}$	$0.05_{0.00}$	$0.02_{0.00}$	$0.03_{0.00}$	$0.03_{0.04}$
Education	$0.55_{0.00}$	$0.62_{0.00}$	$0.29_{0.00}$	$0.37_{0.00}$	$0.36_{0.12}$
Gender	$0.19_{0.00}$	$0.22_{0.01}$	$0.11_{0.04}$	$0.03_{0}.67$	$0.06_{0.89}$
Income	$0.05_{0.00}$	$0.06_{0.00}$	$0.03_{0.00}$	$0.03_{0.00}$	$0.03_{0.59}$
Marital Status	$-0.01_{0.87}$	$-0.03_{0.76}$	$-0.01_{0.85}$	$-0.02_{0.81}$	$-0.08_{0.86}$
Location	$0.15_{0.07}$	$0.13_{0.26}$	$0.05_{0.48}$	$0.07_{0.46}$	$0.03_{0.95}$
Black	$0.60_{0.00}$	$0.70_{0.00}$	$0.33_{0.00}$	$0.43_{0.00}$	$0.22_{0.79}$
Other	$-0.23_{0.07}$	$-0.26_{0.14}$	$-0.11_{0.30}$	$-0.43_{0.01}$	$-0.20_{0.82}$
Hispanic	$0.05_{0.66}$	$0.08_{0.65}$	$0.02_{0.87}$	$-0.12_{0.38}$	$-0.06_{0.94}$

*Data are monotonized before analysis; Refreshment samples are not used

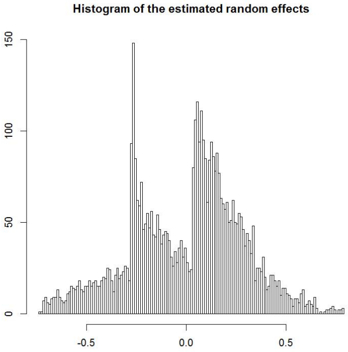
Presidential Election Survey Data

- Large random effects $\Rightarrow$ High interest in election



Presidential Election Survey Data

- The estimated random effects: showing a bi-modal pattern



Concluding Remarks

- Construct an unbiased estimating equation
- Utilizing unspecified random effects to quantify nonignorable missing data
- Consistency and asymptotic normality of fixed-effects estimators
- Can handle intermittent non-response data occurred in refreshment samples

References

- Bi, X. and Qu, A. (2015+). A mixed-effects estimating equation approach to nonignorable missing longitudinal data. *Submitted*.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9-25.
- Diggle, P., Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *Appl. Stat.* **43**, 49-93.
- Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2011). *Applied longitudinal analysis.* Wiley, New York.
- Ibrahim, J. G., Chen, M. H. and Lipsitz, S. R. (2001). Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika* **88**, 551-564.
- Kim, J. K. and Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *J. Amer. Statist. Assoc.* **106**, 157-165.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**, 106-121.

References

- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.
- Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.* **94**, 1096-1120.
- Shao, J. and Zhang, J. (2015). A transformation approach in linear mixed-effects models with informative missing responses. *Biometrika*, in press.
- Seaman, S. and Copas, A. (2009). Doubly robust generalized estimating equations for longitudinal data. *Statistics in Medicine* **28**, 937-955.
- Tsonoka, R., Verbeke, G. and Lesaffre, E. (2009). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Biometrics* **65**, 81-87.
- Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175-188.

Thank You!