Statistica Sinica Preprint No: SS-2015-0317

Title	A mixed-effects estimating equation approach to nonignorable missing longitudinal data with refreshment samples
Manuscript ID	SS-2015-0317
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0317
Complete List of Authors	Xuan Bi and Annie Qu
Corresponding Author	Annie Qu
E-mail	anniequ@illinois.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

A mixed-effects estimating equation approach to nonignorable missing longitudinal data with refreshment samples

Xuan Bi and Annie Qu

University of Illinois at Urbana-Champaign

Abstract

Nonignorable missing data occur frequently in longitudinal studies and can cause biased estimations. Refreshment samples which draw new subjects randomly in subsequent waves from the original population could mitigate the bias. In this paper, we introduce a mixed-effects estimating equation approach that enables one to incorporate refreshment samples and recover informative missing information from the measurement process. We show that the proposed method achieves consistency and asymptotic normality for fixed-effect estimation under shared-parameter models, and we extend it to a more general nonignorable-missing framework. Our finite sample simulation studies show the effectiveness and robustness of the proposed method under different missing mechanisms. In addition, we apply our method to election poll longitudinal survey data with refreshment samples from the 2007-2008 Associated Press–Yahoo! News.

Keywords: Missing not at random; Non-monotone missing pattern; Quadratic inference function; Survey data; Shared-parameter model

Address for correspondence: Annie Qu, Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820. E-mail: anniequ@illinois.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

1 Introduction

Missing data are often encountered in longitudinal studies. Among all the missing mechanisms, missing not at random (MNAR; Rubin (1976)) is the most challenging one to handle. For example, in a public survey, participants with lower socioeconomic status may have a lower probability to release their annual income (Kim and Yu (2011)); and in AIDS clinical trials, subjects with a lower CD4 level may drop out prematurely due to death or pessimism about treatment. Estimation and inference procedures ignoring non-random missing mechanisms may lead to misleading and biased conclusions.

Existing literature on analyzing the MNAR mechanism for longitudinal data includes, but is not limited to, Diggle and Kenward (1994), Little (1994, 1995), Hogan and Laird (1997), Molenberghs et al. (1997), Ibrahim et al. (2001), Roy (2003), Stubbendick and Ibrahim (2003, 2006), Lin et al. (2004), Vansteelandt et al. (2007), Zhou et al. (2010), Spagnoli et al. (2011), and Shao and Zhang (2015). Most of these methods are built under certain MNAR assumptions. However, the MNAR assumption is typically difficult to verify in practice, since the information required for such a test is also missing (Van Buuren (2012)). Consequently it is challenging to assess model effectiveness and robustness under a general MNAR setting, and a sensitivity analysis (Rotnitzky et al. (1998); Robins et al. (2000)) might be required.

An alternative strategy for handling nonignorable missing data is to introduce refreshment samples as part of the experimental design (Ridder (1992)), which recruits new subjects randomly from the same population in subsequent waves over time. Hirano et al. (2001) demonstrate that implementing refreshment samples can mitigate the effect of data attrition, and Deng et al. (2013) further show that refreshment samples are useful for adjusting for bias. However, studies of statistical properties are still limited in application to refreshment samples, partially because baseline values from refreshment samples are typically missing. In addition, the existing methods are restricted to few waves with a

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

small longitudinal cluster size, as it can be computationally intensive to handle refreshment samples with a large number of repeated measurements. Furthermore, the MNAR could still exist even after recruitment of refreshment samples.

We propose a mixed-effects estimating equation approach (MEEE) that preserves the advantages of estimating equations in addressing refreshment samples. The key idea is to reduce the estimation bias through utilizing unspecified random effects for MNAR data. In addition, our theoretical properties confirm that the fixed-effects estimators solved through the MEEE are consistent and asymptotically normal under two MNAR mechanisms. The proposed method has practical advantages as it is able to utilize a large number of repeated measurements from the same subject to achieve higher estimation accuracy for the random effects. This is in contrast to traditional methods which can be problematic if the cluster size of longitudinal data is large (Lipsitz et al. (2009)).

The idea of unspecified random effects has also been considered in the existing literature under the likelihood framework (for example, Tsonaka et al. (2009, 2010); Li et al. (2012); Maruotti (2015)). Our method has several advantages compared to the existing approaches. First, it does not require random effects to follow a discrete distribution with finite support points. In fact, our random effects are solved by estimating equations and their values are not restricted to a certain set. For existing approaches, the selection of the number of support points for the unspecified random-effects distribution remains an open question (Tsonaka et al. (2009)). Second, the proposed method is not restricted to shared-parameter models (SPM) (e.g., Wu and Carroll (1988); Wu and Bailey (1989); Follmann and Wu (1995)) where the response variable and the missing process are linked through random effects. We demonstrate that it can be applied under both SPMs and an extended SPM where the missing process is related to observed responses in addition to random effects. Third, the proposed method does not require baseline observations. This is especially useful for longitudinal survey studies with refreshment samples. In such a case, existing methods requiring baseline observations are difficult to implement, while the proposed method is

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

still applicable.

The rest of this article is organized as follows. Section 2 introduces notation and the shared-parameter model assumption. Section 3 illustrates how to construct unbiased estimating equations under MNAR mechanisms and provides theoretical properties. Section 4 demonstrates the performance of the proposed method through simulation studies. Section 5 applies the proposed method to election poll survey data provided by the 2007-2008 Associated Press–Yahoo! News. The last section presents concluding remarks and a brief discussion. All technical proofs are in the Appendix.

2 Notation and Basic Assumptions

In this section, we introduce notation and basic assumptions for longitudinal missing data.

Let y_{it} denote the tth observation from the ith subject, with $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ p-dimensional fixed-effects and q-dimensional random-effects covariates, respectively, $i = 1, \ldots, n$, $t = 1, \ldots, T$. We assume that the responses and covariates are linked through a known inverse link function μ:

$$\mathrm{E}(y_{it}|\mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{b}_i) = \mu(\mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_{it}\mathbf{b}_i),$$

where $\boldsymbol{\beta}$ is a p-dimensional fixed-effects vector and $\mathbf{b}_i$ is a q-dimensional random-effects vector, $i = 1, \ldots, n$. Suppose $\mathbf{b}_1, \ldots, \mathbf{b}_n$ are true realizations of an unknown stochastic process, and write $\mathbf{b} = (\mathbf{b}_1', \ldots, \mathbf{b}_n')'$. We do not impose any distribution assumption on $\mathbf{b}$.

Let δ_{it} be y_{it}'s missing indicator, $\delta_{it} = 1$ if y_{it} is observed and $\delta_{it} = 0$ otherwise. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{iT})'$, and $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})'$, where $\mu_{it} = \mu(\mathbf{x}_{it}\boldsymbol{\beta}+\mathbf{z}_{it}\mathbf{b}_i)$. Let $n_i = \sum_{t=1}^{T} \delta_{it}$ be the number of measurements observed from the ith subject. We define a missing indicator matrix Δ_i as an $n_i \times T$-dimensional matrix corresponding to the rows of identity matrix I_T for which $\mathbf{y}_i$ is observed.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

The idea of creating Δ_i is to transform the hypothetical complete response $\mathbf{y}_i$ into an observed vector $\Delta_i \mathbf{y}_i$ (e.g., Paik (1997)). For example, if $\boldsymbol{\delta}_i = (1, 0, 1, 0, 1)'$ and $\mathbf{y}_i = (y_{i1}^o, y_{i2}^m, y_{i3}^o, y_{i4}^m, y_{i5}^o)'$ where the superscript "o" and "m" indicate "observed" and "missing" respectively, then $\Delta_i \mathbf{y}_i = (y_{i1}^o, y_{i3}^o, y_{i5}^o)'$. Each $\boldsymbol{\delta}_i$ determines a unique Δ_i, and thus Δ_i is a function of $\boldsymbol{\delta}_i$. We take $\mathbf{y}_i^o = \Delta_i \mathbf{y}_i$ and $\boldsymbol{\mu}_i^o = \Delta_i \boldsymbol{\mu}_i$.

The shared-parameter model assumption (e.g., Follmann and Wu (1995)) is

$$\mathbf{y}_i \perp\!\!\!\perp \boldsymbol{\delta}_i | \mathbf{b}_i, \tag{1}$$

which assumes that the missing process $\boldsymbol{\delta}_i$ and the measurement process $\mathbf{y}_i$ share the same random effect $\mathbf{b}_i$. The missing mechanism satisfying (1) must be MNAR. We further discuss this assumption and provide an extension in Section 3.1.

3 The General Method

In this section, we propose an unbiased estimating equation approach to estimate the fixed effect $\boldsymbol{\beta}$ and the random effect $\mathbf{b}$. Throughout, we use unbiasedness to denote the conditional unbiasedness of an estimating equation given latent random effects.

A standard GEE can be formulated as:

$$\sum_{i=1}^{n} \dot{\boldsymbol{\mu}}_i' A_i^{-\frac{1}{2}} R^{-1} A_i^{-\frac{1}{2}} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{2}$$

where $\dot{\boldsymbol{\mu}}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$, A_i is a diagonal matrix of marginal variance of $\mathbf{y}_i$, and R is a working correlation matrix that contains fewer nuisance parameters than an unspecified correlation matrix.

The unbiasedness of (2) leads to the consistency and asymptotic normality of the fixed-effect estimator $\hat{\boldsymbol{\beta}}$ (Liang and Zeger (1986); Robins et al. (1994); Rotnitzky et al. (1998)).

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Our goal is to build unbiased estimating equations in the presence of nonignorable missing data.

3.1 Construction of Unbiased Estimating Equations

In this subsection, we demonstrate the unbiasedness of the proposed MEEE by incorporating unspecified random effects. Specifically, if either the SPM assumption or a relaxed version of the SPM assumption is satisfied, we have conditionally unbiased estimating equations that do not rely on a specification of the missing process.

Let $A_i^o = \Delta_i A_i \Delta_i'$, $R_i^o = \Delta_i R \Delta_i'$ and

$$\begin{aligned}\bar{\mathbf{G}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_i &= \frac{1}{n}\sum_{i=1}^{n}(\dot{\boldsymbol{\mu}}_i^o)'(A_i^o)^{-\frac{1}{2}}(R^o)^{-1}(A_i^o)^{-\frac{1}{2}}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)\\ &= \frac{1}{n}\sum_{i=1}^{n}(\dot{\boldsymbol{\mu}}_i)'\Delta_i'(\Delta_i A_i \Delta_i')^{-1/2}(\Delta_i R \Delta_i')^{-1}(\Delta_i A_i \Delta_i')^{-1/2}\Delta_i\,(\mathbf{y}_i - \boldsymbol{\mu}_i)\,.\end{aligned}$$

We define a $T \times T$ weighting matrix conditional on the random effect $\mathbf{b}_i$

$$K_i = K_i(\boldsymbol{\delta}_i|\mathbf{b}_i) = \Delta_i'(\Delta_i A_i \Delta_i')^{-1/2}(\Delta_i R \Delta_i')^{-1}(\Delta_i A_i \Delta_i')^{-1/2}\Delta_i, i = 1, \ldots, n.$$

Then $\mathbf{G}_i = (\dot{\boldsymbol{\mu}}_i)'K_i(\mathbf{y}_i - \boldsymbol{\mu}_i)$.

If (1) holds, then $\mathrm{E}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} = \mathbf{0}$ implies that

$$\mathrm{E}(\mathbf{G}_i|\mathbf{b}_i) = \mathrm{E}\{\dot{\boldsymbol{\mu}}_i'K_i(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} = \dot{\boldsymbol{\mu}}_i'\mathrm{E}\{K_i|\mathbf{b}_i\}\mathrm{E}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} = \mathbf{0}. \quad (3)$$

This decomposition does not restrict the weighting matrix $K_i(\boldsymbol{\delta}_i|\mathbf{b}_i)$, which contains the information of the MNAR mechanism. Compared to existing SPM approaches, the distribution formulation of the missing process $\boldsymbol{\delta}_i$ is not needed here.

The formulation of (3) still requires (1). This assumption does not hold if the missingness is a function of past or current responses directly. Several methods have been

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

developed to weaken the SPM assumption. Among them, Henderson et al. (2000) and Rizopoulos et al. (2008) introduce different but correlated random effects for the measurement process and the missing process, where the two processes no longer "share" the same random effects. Little (2008) and Yuan and Little (2009) propose a mixed-effects hybrid model that models the dropout process directly. Nevertheless, these works still require parametric assumptions on the random effects, and some are only applicable for the drop-out missing mechanism.

We propose to relax assumption (1) through strengthening the association between $\mathbf{y}_i$ and $\boldsymbol{\delta}_i$ if the random effect itself cannot completely capture the missing information. Our relaxation does not require any parametric distribution on the random effects nor does it restrict the missing patterns. In particular, the relaxed assumption is applicable to the estimating equation framework where only the first two moments are known. We introduce a new missing mechanism termed *conditionally missing at random (CMAR)*.

Definition 1. *A missing mechanism is conditionally missing at random if missingness does not depend on unobserved data, given the observed data and the random effects.*

Mathematically, (1) states that $\boldsymbol{\delta}_i|\mathbf{b}_i, \mathbf{y}_i \stackrel{d}{=} \boldsymbol{\delta}_i|\mathbf{b}_i$, where $\stackrel{d}{=}$ denotes "equivalent in distribution." Then Definition 1 generalizes (1) to

$$\boldsymbol{\delta}_i|\mathbf{b}_i, \mathbf{y}_i \stackrel{d}{=} \boldsymbol{\delta}_i|\mathbf{b}_i, \mathbf{y}_i^o. \tag{4}$$

It can be shown that the CMAR mechanism is still a MNAR mechanism. This generalization is analogous to the generalization from MCAR to MAR, as we allow the observed response $\mathbf{y}^o$ to carry information of the missing mechanism as well. The new assumption offers more flexibility, since it no longer requires the random effect $\mathbf{b}$ to capture all information associating the missing process with the measurement process. In the following we show that the estimating equation (2) remains unbiased.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Let $W_i = A_i^{\frac{1}{2}} R A_i^{\frac{1}{2}}$, then for each subject i

$$\begin{aligned}\mathbf{0} &= \mathrm{E}\{\dot{\boldsymbol{\mu}}_i' W_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} \\ &= \mathrm{E}\{\dot{\boldsymbol{\mu}}_i' W_i^{-1}\mathrm{E}(\mathbf{y}_i - \boldsymbol{\mu}_i|\mathbf{b}_i, \mathbf{y}_i^o, \boldsymbol{\delta}_i)|\mathbf{b}_i\}.\end{aligned}$$

If $\mathbf{y}_i = (\mathbf{y}_i^o, \mathbf{y}_i^m)'$ and $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_i^o, \boldsymbol{\mu}_i^m)'$ then, based on the CMAR assumption in (4), $\mathrm{E}(\mathbf{y}_i^m - \boldsymbol{\mu}_i^m|\mathbf{b}_i, \mathbf{y}_i^o, \boldsymbol{\delta}_i) = \mathrm{E}(\mathbf{y}_i^m - \boldsymbol{\mu}_i^m|\mathbf{b}_i, \mathbf{y}_i^o)$ is no longer a function of $\boldsymbol{\delta}_i$, and thus can be modeled by available information through random effects. This follows similarly as the MAR definition.

There are several methods to impute missing values based on observed values. For example, Paik (1997) applies mean imputation for longitudinal data. Seaman and Copas (2009) combine mean imputation with a weighting strategy to construct a doubly robust estimator. Qu et al. (2010) propose to impute missing values through utilizing the linear conditional mean method (LCM). Here we adopt the LCM under the context of the mixed-effects model for simplicity:

$$\mathrm{E}(\mathbf{y}_i^m - \boldsymbol{\mu}_i^m|\mathbf{b}_i, \mathbf{y}_i^o) = W_i^{21}(W_i^{11})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o), \tag{5}$$

where $W_i^{21} = \mathrm{Cov}(\mathbf{y}_i^m, \mathbf{y}_i^o|\mathbf{b}_i)$ and $W_i^{11} = \mathrm{Var}(\mathbf{y}_i^o|\mathbf{b}_i)$.

Given (5) and the fact that

$$W_i^{-1}\begin{pmatrix} I \\ W_i^{21}(W_i^{11})^{-1}\end{pmatrix} = \begin{pmatrix}(W_i^{11})^{-1} \\ 0\end{pmatrix},$$

we have

$$\begin{aligned}\mathbf{0} = \mathrm{E}\{\dot{\boldsymbol{\mu}}_i' W_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} &= \mathrm{E}\{(\dot{\boldsymbol{\mu}}_i^o)'(W_i^{11})^{-1}(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o)|\mathbf{b}_i\} \\ &= \mathrm{E}\{\dot{\boldsymbol{\mu}}_i' K_i(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\}.\end{aligned}$$

This indicates that, once the LCM method is valid, the estimating equation in (2) is

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

unbiased without requiring (1). Under either the SPM or the CMAR assumption, one can check the conditional unbiasedness of the estimating equation $\mathrm{E}\{\dot{\boldsymbol{\mu}}_i' K_i(\mathbf{y}_i - \boldsymbol{\mu}_i)|\mathbf{b}_i\} = \mathbf{0}$ through a chi-square test (Hansen (1982); Qu et al. (2000)) to test the null hypothesis for the mean zero assumption of the estimating functions.

The LCM imputation method (5) is based on first-order linear approximation. The imputed values are valid if the conditional distribution $\mathbf{y}_i|\mathbf{b}_i$ is multivariate normal, or bivariate binary (Qu et al. (2010)). In addition, for a multivariate binary distribution, the LCM is valid if it belongs to the conditional linear family (Qaqish (2003)) that assumes zero for the second and higher-order terms in Bahadur's representation (Bahadur (1961)). For such circumstances as multivariate count data, the LCM provides an approximate estimation with accuracy similar to linear regression. Nevertheless, more complicated imputation methods should be considered if one believes high-order approximations are necessary for the observed data.

3.2 Estimation of Mixed Effects

In this subsection we discuss how to solve the proposed MEEE and estimate both fixed effects and unspecified random effects. When the sample size is small or the missing rate is high, the empirical correlation matrix might be unstable or non-positive definite, In which case, we avoid the estimation of the matrix. Specifically, we formulate estimating functions based on the observed data as

$$\bar{\mathbf{g}}_n^f = \frac{1}{n}\sum_{i=1}^n \mathbf{g}_i^f(\boldsymbol{\beta}|\mathbf{b}_i) = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n (\dot{\boldsymbol{\mu}}_i^o)'(A_i^o)^{-1/2} M_{i1} (A_i^o)^{-1/2}\,(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o) \\ \vdots \\ \sum_{i=1}^n (\dot{\boldsymbol{\mu}}_i^o)'(A_i^o)^{-1/2} M_{im} (A_i^o)^{-1/2}\,(\mathbf{y}_i^o - \boldsymbol{\mu}_i^o) \end{pmatrix},$$

where $M_{ij} = \Delta_i M_j \Delta_i'$ and $\{M_j\}_{j=1}^m$ is a matrix representation of R^{-1} satisfying $R^{-1} = \sum_{j=1}^m a_j M_j$. Here M_j is a basis matrix containing only 0's and 1's. See more details on

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

selection of M_j's in Qu et al. (2000), and the number of basis matrices m in Zhou and Qu (2012).

The equality $M_{ij} = \Delta_i M_j \Delta_i'$ entails the assumption $(\Delta_i R \Delta_i')^{-1} = \Delta_i R^{-1} \Delta_i'$, which simplifies the matrix representation for R^{-1} of each subject. This representation does not affect the consistency of estimation when misspecified, and provides better efficiency compared to using an independence structure.

We take

$$K_{ij} = K_{ij}(\boldsymbol{\delta}_i|\mathbf{b}_i) = \Delta_i'(\Delta_i A_i \Delta_i')^{-1/2} M_{ij} (\Delta_i A_i \Delta_i')^{-1/2} \Delta_i, i = 1, \ldots, n; j = 1 \ldots, m.$$

Then solving $\bar{\mathbf{g}}_n^f = \mathbf{0}$ is equivalent to solving

$$\bar{\mathbf{g}}_n^f(\boldsymbol{\beta}|\mathbf{b}) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n (\dot{\boldsymbol{\mu}}_i)' K_{i1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n (\dot{\boldsymbol{\mu}}_i)' K_{im} (\mathbf{y}_i - \boldsymbol{\mu}_i) \end{pmatrix} = \mathbf{0}.$$

The relation between K_i and K_{ij} is $K_i = \sum_{j=1}^m a_j K_{ij}$.

For the fixed-effects estimation, since there are more estimating functions than parameters, we estimate $\boldsymbol{\beta}$ by applying the generalized method of moments (Hansen (1982)) conditional on $\mathbf{b}$:

$$\hat{\boldsymbol{\beta}} = \arg\min (\bar{\mathbf{g}}_n^f)' (\bar{C}_n^f)^{-1} (\bar{\mathbf{g}}_n^f), \tag{6}$$

where $\bar{C}_n^f = \dfrac{1}{n}\sum_{i=1}^n (\mathbf{g}_i^f)(\mathbf{g}_i^f)'$.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

For the random-effects estimation, we solve

$$\bar{\mathbf{g}}_n^r = \begin{pmatrix} (\frac{\partial \boldsymbol{\mu}_1}{\partial \mathbf{b}_1})' K_1(\mathbf{y}_1 - \boldsymbol{\mu}_1) \\ \vdots \\ (\frac{\partial \boldsymbol{\mu}_n}{\partial \mathbf{b}_n})' K_n(\mathbf{y}_n - \boldsymbol{\mu}_n) \\ \lambda P_A \mathbf{b} \end{pmatrix} = \mathbf{0},$$

where P_A is the projection matrix on the null space of $(I - P_X)Z$, P_X the projection matrix on X, and λ is a tuning parameter. The term $\lambda P_A \mathbf{b}$ is to ensure the identifiability of $\hat{\mathbf{b}}$. The random-effect estimator $\hat{\mathbf{b}}$ is obtained as

$$\hat{\mathbf{b}} = \arg\min\{(\bar{\mathbf{g}}_n^r)'(\bar{\mathbf{g}}_n^r) + \lambda_1^2 \mathbf{b}'\mathbf{b}\}, \tag{7}$$

where $\lambda_1^2 \mathbf{b}'\mathbf{b}$ is an L_2-penalty term to control the magnitude of $\mathrm{Var}(\mathbf{b})$ in order to ensure the convergence in optimization. We estimate $\boldsymbol{\beta}$ and $\mathbf{b}$ by solving (6) and (7) iteratively. In Section 3.4 we discuss in detail how the tuning parameters λ and λ_1 are selected.

We propose a chi-square test to test the validity of the LCM imputation method. We construct two sets of estimating equations: one contains subjects with no missing response, and the other has missing responses imputed using the LCM method. Since the first set of estimating equations is unbiased, a chi-square test can be conducted to test whether or not the second set of estimating equations is unbiased. Let

$$\mathcal{C} = \left\{ i \in \{1, \ldots, n\} : \sum_{t=1}^{T} \delta_{it} = T \right\}$$

be the set of complete subjects. Write $\Phi_1 = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} (\dot{\boldsymbol{\mu}}_i)' V^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)$, where $|\cdot|$ denotes the cardinality of a set and V the covariance matrix calculated based on subjects with

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

completed responses and

$$\Phi_2 = \begin{pmatrix} \frac{1}{n-|\mathcal{C}|}\sum_{i\notin\mathcal{C}}(\dot{\boldsymbol{\mu}}_i)' K_{i1}\,(\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \frac{1}{n-|\mathcal{C}|}\sum_{i\notin\mathcal{C}}(\dot{\boldsymbol{\mu}}_i)' K_{im}\,(\mathbf{y}_i - \boldsymbol{\mu}_i) \end{pmatrix}$$

be the estimating functions using subjects with missing values. Let $\Phi = (\Phi_1', \Phi_2')'$. According to Theorem 1 of Qu et al. (2011), under the null hypothesis H_0: $\mathrm{E}(\Phi) = \mathbf{0}$, $\Phi'\mathrm{Var}^{-1}(\Phi)\Phi \sim \chi^2_{mp}$.

3.3 Asymptotic Properties

In this subsection, we investigate fixed-effects estimation consistency and asymptotic normality. Lemma 1 provides the asymptotic property when $\mathbf{b}$ is known or is consistently estimated, and Theorem 1 shows that the desirable properties still hold under certain conditions even if $\mathbf{b}$ is unspecified. Hereafter we use $\boldsymbol{\beta}_0$ and $\mathbf{b}_0 = (\mathbf{b}_{01}', \ldots, \mathbf{b}_{0n}')'$ to denote the true fixed effect and the true random effect, respectively.

Lemma 1. *Given (1) and conditional on $\mathbf{b}_0$, $\hat{\boldsymbol{\beta}}$ solved by (6) satisfies $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(\frac{1}{\sqrt{n}})$, and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, \Sigma_0)$, where Σ_0 is derived in the proof.*

The conclusions in Lemma 1 still hold if $\mathbf{b}_0$ is replaced by a consistent estimator $\hat{\mathbf{b}}$, but the consistency of $\hat{\mathbf{b}}$ requires that the cluster size T goes to infinity, which may be too restrictive in practice. A weaker condition retains the properties stated in Lemma 1 are still valid: we assume that $\hat{\mathbf{b}}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_i^f(\boldsymbol{\beta}_0|\hat{\mathbf{b}}) \to \mathbf{0} \text{ as } n \to \infty. \tag{8}$$

This condition implies that, conditional on $\hat{\mathbf{b}}$, the sample mean of estimating equations for the fixed effect converges to 0 when the sample size n goes to infinity while the cluster size

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

T is fixed.

Condition (8) is weaker than the consistency of $\hat{\mathbf{b}}$ since, if $\hat{\mathbf{b}}$ is consistent, then (8) holds true for a large T. For a counterexample where (8) does not imply the consistency of $\hat{\mathbf{b}}$, suppose $y_{it} = \beta_0 + b_i + \varepsilon_{it}$, where $\beta_0 = 0$, $\mathrm{E}(b_i) = 0$, $\mathrm{E}(\varepsilon_{it}) = 0$, and $\mathrm{Corr}(\varepsilon_{it}, \varepsilon_{it'}) = 1$, for $i = 1, \ldots, n$ and $t, t' = 1, \ldots, T$. Then $y_{it} = y_{it'}$ with a probability of 1 and the corresponding quasi-likelihood equation is

$$g_i^f(\beta_0|\hat{b}) = \dot{\boldsymbol{\mu}}_i'(\mathbf{y}_i - \boldsymbol{\mu}_i) = \sum_{t=1}^{T}(y_{it} - \hat{b}_i).$$

If (8) is satisfied, then $\hat{b}_i = \dfrac{1}{T}\sum_{t=1}^{T} y_{it} = y_{i1}$, but $\hat{b}_i = y_{i1}$ is not a consistent estimator of b_i as $T \to \infty$.

Theorem 1. *If (1) and (8) hold then, conditional on $\hat{\mathbf{b}}$, $\hat{\boldsymbol{\beta}}$ solved by (6) satisfies $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(\frac{1}{\sqrt{n}})$, and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, \Sigma)$, where Σ is derived in the proof.*

If $\hat{\mathbf{b}}$ is a consistent estimator of $\mathbf{b}_0$, then we have $\Sigma = \Sigma_0$ (Wang et al. (2012)). Next, we relax the regular shared-parameter model assumption, and show that the fixed-effects estimator is still consistent and asymptotically normal under the CMAR mechanism described in Definition 1.

Corollary 1. *If (4) and (8) hold and the LCM imputation method (5) is valid then, conditional on $\hat{\mathbf{b}}$, $\hat{\boldsymbol{\beta}}$ solved by (6) satisfies $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(\frac{1}{\sqrt{n}})$, and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, \Sigma)$.*

Corollary 1 can be shown similarly as was done in the derivation of Theorem 1 and is therefore omitted.

Remark: Given $\boldsymbol{\beta}_0$ or its consistent estimator $\hat{\boldsymbol{\beta}}$, the penalized random-effects estimator $\hat{\mathbf{b}}$ is consistent as the cluster size T goes to infinity, as discussed in Cho et al. (2016), given that regularity conditions are satisfied. Under the nonignorable missing data framework, the consistency property holds as long as either the SPM or the CMAR assumption

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

is satisfied. The proof is quite similar to Cho et al. (2016), and is omitted here. One notable condition is the L_2-mixingale condition (McLeish (1975)) that controls the serial correlation $\mathrm{Cor}(\mathbf{y}_i|\mathbf{b}_i)$ to achieve consistency, $\mathrm{Cor}(y_{it}, y_{i,t+s})$ should be sufficiently small with an increase of s.

3.4 Tuning Parameter Selection

We discuss the selection of the tuning parameters λ and λ_1 in (7). A large value of λ guarantees that the random-effects estimation is identifiable. However, a very large value of λ does not enhance identifiability significantly, and might result in slower convergence or non-convergence of the algorithm. In our numerical studies, we find that $\lambda = \log(n)$ is sufficiently large to balance the needs of estimation identifiability and algorithm convergence.

The estimation is more sensitive to the choice of λ_1, since a larger value of λ_1 leads to a smaller variance of $\hat{\mathbf{b}}$ that could affect the estimation of $\boldsymbol{\beta}$. The term $\lambda_1 \mathbf{b}'\mathbf{b}$ is essentially an L_2-penalty, which controls the bias-variance trade-off of $\mathbf{b}$. As a special case, when $\boldsymbol{\mu}$ is an identity mean function, $\hat{\mathbf{b}}$ is equivalent to a ridge regression estimator. We use a cross-validation method to select λ_1. Each subject has a unique random effect, and hence a classical K-fold cross-validation is not applicable. We propose a longitudinal K-fold cross-validation with $K = T$. Thus, we remove measurements observed at time t from all subjects ($t = 1, \ldots, T$), and minimize the objective function

$$L(\lambda_1) = \frac{1}{\sum_{i=1}^n n_i} \sum_{i=1}^n (\mathbf{y}_i^o - \hat{\mathbf{y}}_i^o)'(W_i^o)^{-1}(\mathbf{y}_i^o - \hat{\mathbf{y}}_i^o),$$

where $\hat{\mathbf{y}}_i^o = \Delta_i(\hat{y}_{i1}^{(-1)}, \ldots, \hat{y}_{iT}^{(-T)})'$ and $\hat{y}_{it}^{(-t)}$ is the predicted value of y_{it} without using information from time t.

To calculate $\hat{y}_{it}^{(-t)}$, Hastie et al. (2009, Chapter 7.2) suggest $\hat{y}_{it} = \arg\max_y f_{it,\lambda_1}(y)$, where f_{it,λ_1} is the probability density function or the probability mass function of y_{it} with

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

parameter λ_1. For example, if y_{it} is normally distributed, then $\hat{y}_{it}^{(-t)} = \mathbf{x}_{it}\hat{\boldsymbol{\beta}}^{(-t)}$; and if y_{it} follows a Poisson distribution, then $\hat{y}_{it}^{(-t)} = [\hat{\mu}_{it}^{(-t)}]$, where $[\hat{\mu}_{it}^{(-t)}]$ denotes the largest integer not greater than $\hat{\mu}_{it}^{(-t)}$.

4 Simulation Studies

We conducted simulation studies to examine the performance of the proposed method. To make a fair comparison to existing approaches using the SPM, we compared the proposed MEEE with generalized linear mixed-effects models (GLMMs), where the estimating equations are unbiased under the SPM assumption (1). The difference is that the GLMM assumes normality of the random effects and independence of repeated measurements, given that random effects are taken into account. The GLMM can be implemented through the penalized quasi-likelihood (PQL; Breslow and Clayton (1993)) or the adaptive Gaussian-Hermite quadrature approach (GHQ; Anderson and Aitkin (1985)). We also compared the proposed method with two marginal estimating equation approaches: the weighted generalized estimating equations (WGEE; Robins et al. (1995)), and the multiple imputation method for longitudinal data (MI; Fitzmaurice et al. (2011)). Although the WGEE and multiple imputation have the marginal interpretation and are valid under only MAR, they are benchmark methods under the estimating equation framework for longitudinal missing data.

The PQL and the GHQ are carried out using the R functions “glmmPQL” and “glmer”, respectively, while the WGEE is obtained by assigning weights to the R function “geem.” For the MI approach, we imputed missing data following Fitzmaurice et al. (2011, Chapter 18.2) when the missing pattern is monotone, or applied the R package “MICE” to impute intermittent missing data utilizing the chained equation (Van Buuren (2007)). For WGEE, we tailored the responses to a monotone pattern of missingness in order to apply the weighting strategy.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

We also conducted a chi-square test (Qu et al. (2000)) to test the unbiasedness of the fixed-effects estimating equations $\mathrm{E}(\bar{\mathbf{g}}_n^f) = \mathbf{0}$, and the validity of the LCM imputation method indicated in Section 3.2.

4.1 Study 1: Count Responses under the SPM Assumption

We chose the sample size $n = 150$ and the cluster size $T = 3$. The fixed-effects covariates $\mathbf{x}_{it} = (1, \mathrm{trt}, \mathrm{time}, \mathrm{trt} \times \mathrm{time})'$, where "trt" (treatment) was assigned to 1 if $i \leq n/2$ and 0 otherwise, "time" was the standardized time effect, and "trt $\times$ time" was the interaction effect of treatment and time. The fixed-effects parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)' = (-0.5, 0.5, 0.2, 0.2)'$. The random-effects covariates $\mathbf{z}_{it} = (x_{it1}, x_{it3})'$, and the random effects $b_{ij} \overset{iid}{\sim} \mathrm{Unif}(-0.2, 0.2)$ for $j = 1, 2$. Each $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ was sampled from a multivariate Poisson distribution with mean $\boldsymbol{\lambda}_i$ satisfying

$$\log(\lambda_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i, \ t = 1, \ldots, T,$$

and the correlation structure for the repeated measurements was exchangeable with correlation parameter $\rho = 0, 0.4$ or 0.7. The correlated Poisson data were generated by the R package "corcounts."

For the missing process, let $p_{it}^{\delta} = \mathrm{P}(\delta_{it} = 1)$ and the logistic model of p_{it}^{δ} be

$$\mathrm{logit}(p_{it}^{\delta}) = 0.1 b_{i1} - 0.3t/T + 0.1, \ t = 2, \ldots, T,$$

where the assumption (1) is satisfied. The term $-t/T$ ensures that the missing rate is higher toward the end of the study, mimicking the real data. We assumed a monotone pattern of missingness, and $\delta_{it} = \cdots = \delta_{iT} = 0$ if $\delta_{i,t-1} = 0$. The overall missing rate was about 45%.

Table 1 provides simulation results based on 200 simulation runs. For the unbiasedness

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

test, we rejected the null hypothesis 8, 5 and 10 times out of 200 replications at a significance level of 0.05 when the serial correlation $\rho = 0$, $\rho = 0.4$ and $\rho = 0.7$, respectively. This indicates that the estimating equations are unbiased, and fixed-effects estimates are consistent. For the test of the validity of the LCM, we rejected the null hypothesis 101, 63 and 19 times out of 200 replications at a level of 0.05 for $\rho = 0$, 0.4 and 0.7, respectively. This agrees with the theory that the LCM imputation is only an approximation when the response variable is count data. Nevertheless, the proposed method satisfies the SPM assumption (1), and performs the best, even when the LCM condition is violated.

Overall the proposed estimators are less biased and have smaller standard errors compared to the WGEE and the MI approaches. In addition, the improvement of the proposed method is more significant when the correlation parameter ρ increases, in general. The PQL does not converge due to a small cluster size T, and the GHQ is not applicable here since the dimension of parameters is greater than the number of data points.

4.2 Study 2: Binary Responses under the CMAR Assumption

In the second simulation study, we evaluated the performance of the proposed estimator when (1) is violated but (4) is satisfied.

We generated data with sample size $n = 80$ and cluster size $T = 6$. The fixed-effect covariates and the random-effect covariates were the same as in the simulation study 1. The fixed-effect was $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)' = (-0.5, 1, 0.8, 0.8)'$. The random-effect covariate was $\mathbf{z}_{it} = (x_{it1}, x_{it3})'$, and the random effects $b_{ij} \overset{iid}{\sim} \text{Unif}(-0.2, 0.2)$ for $j = 1, 2$. The response $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ followed a multivariate Bernoulli distribution with mean function μ_{it} satisfying

$$\text{logit}(\mu_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{z}_{it}'\mathbf{b}_i,\ t = 1, \ldots, T,$$

and an AR-1 correlation structure with correlation parameter $\rho = 0.2$ or 0.6, generated by the R package "MultiOrd." The correlation structure of $\rho > 0.6$ cannot be generated due

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

to infeasibility of this mean function (Chaganty and Joe (2006)).

We generated the missing process through the logistic model

$$\text{logit}(p_{it}^{\delta}) = 0.1y_{i1} + 0.2b_{i1} - 0.5t/T + 0.5,\ t = 2, \ldots, T.$$

In this setting, (4) is satisfied. The missing pattern is intermittent, and the overall missing rate is about 40%.

Table 2 provides the simulation results from 200 replications when $\rho = 0.2$ or 0.6. For the unbiasedness test, we rejected the null 72 and 80 times out of 200 simulations at a significance level of 0.05 when the serial correlation $\rho = 0.2$ and $\rho = 0.6$, respectively. Thus, the unbiasedness of the estimating equations was mildly violated. In addition, since (1) is violated, we conducted the chi-square test proposed in Section 3.2 to test the validity of the LCM, which leds us to reject the null hypothesis 14 and 37 times out of 200 replications for $\rho = 0.2$ and 0.6, respectively. The LCM is mildly violated in this setting. However, the MEEE still outperforms other approaches and achieves the smallest absolute bias and standard error with its coverage probability around 95%. The improvement of the proposed method is more evident when the correlation parameter $\rho = 0.6$, where the PQL and GHQ deteriorate drastically with higher bias, standard errors, and much lower coverage probability of the confidence interval.

5 Application

IWe analyzed data collected from the 2007-2008 Associated Press-Yahoo! News Poll. This survey was intended to evaluate changes in nationwide attitude and opinion toward the presidential election in 2008. It was an eleven-wave survey with the first nine waves conducted during the year prior to the 2008 general election.

Respondents were invited to participate in all follow-up waves, regardless of their re-

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

sponses to the previous waves, so the missing pattern is non-monotone. However, this survey suffered greatly from data attrition, where only 63% of the first wave respondents participated in wave 9. To offset the high percentage of the missing rate, the survey recruited new participants as refreshment samples in waves 3, 5, 6, and 9.

We chose one of the survey questions as a response variable: "How much interest do you have in following news about the campaign for president?" Following the Pew Research Center (2010) and Deng et al.'s (2013) strategy, we dichotomized the 5-level response: 1 for answers "a great deal" or "quite a bit", and 0 otherwise. We analyzed all available data collected in the 9 waves before the election, and the total sample size was 4719. The response measurements from the same subject were correlated, with an approximately exchangeable correlation structure and an average correlation around 0.6. The overall missing rate of the response variable was 49.7%. The predictors were all observed, including time, age, education, gender, household income, marital status, whether living in a metropolitan statistical area (MSA Status), and race/ethnicity.

The missingness of the response variable is likely to be nonignorable. This is reflected by the left panel of Figure 1 showing that respondents are more interested in the presidential election if they stay in the survey longer; the missing probability depends highly on the measurement process. In addition, the missing mechanisms occurring in the refreshment samples are also likely to be MNAR. A two-sample t-test shows that in the last wave before the election, new respondents collected in the last wave have significantly higher interest in the presidential campaign than the respondents recruited in earlier waves, indicating that the earlier measurements from refreshment samples are missing nonignorably. This is also indicated by the left panel of Figure 1, in that the responses from subjects with only one observation have a higher average interest in the presidential election, as these subjects are mainly recruited in the last wave.

We assumed a random intercept model for the MEEE, the PQL and the GHQ, and compared them with three marginal approaches: GEE, WGEE, and MI, for which estima-

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

tions, standard errors and p-values are provided in Table 3. We conducted the unbiasedness test of the estimating equations and the validity test of the LCM for the proposed method. Both tests rejected the null, indicating that these assumptions are violated. However, the MEEE approach agrees with most of the other methods that as the election time gets closer, older people with higher education level and higher household income were more interested in the presidential election. In addition, except for the MI with monotonized data, the other six methods showed that "Black and Non-Hispanic" people were more interested in the presidential election than "White and Non-Hispanic." The most interesting finding here is that methods incorporating refreshment samples such as MEEE, PQL, GHQ, GEE and MI with all available data were able to detect a significant difference in interest between males and females, which coincides with the finding of the Pew Research Center (2010). This implies that refreshment samples may contain important information which should not be ignored. In addition, the MEEE had smaller standard errors for estimators regarding "MSA status" and "Other Non-Hispanic", with more significant p-values.

The right panel of Figure 1 plots the average estimated random effects versus the number of observations, which agrees with the left panel in that a large value of random effect implies high interest in the election. Figure 2 is a histogram of the estimated random effects given by the MEEE, which shows a bimodal pattern. A Shapiro-Wilk test indicates that the normality assumption for random effects is severely violated (p-value $< 10^{-15}$). Existing approaches that impose the normality assumption may result in estimation bias and misleading inference.

6 Discussion

In this paper, we propose a mixed-effects model to correct estimation bias for nonignorable missing data. Mainly, we construct unbiased estimating equations with unspecified random effects under a shared-parameter model, and extend it to a more general nonignorable-missing framework. We show that consistency of the fixed-effects parameter estimation can still be achieved under the more general framework. To our knowledge, most existing methods in the shared-parameter model framework require either a parametric distribution assumption or finite support points for the random effects, while our method allows unspecified random effects that do not have such restrictions. In addition, the proposed method imposes no restriction on the missing pattern, and hence it can be effectively applied to refreshment samples where baseline observations are subject to missingness.

For future research, it would be worthwhile to develop a method for handling missing covariates and responses simultaneously (e.g., Lee and Tang (2006); Chen et al. (2010)). In our framework, since neither the SPM assumption in (1) nor the relaxed assumption in (4) imposes constraints on covariates, we can treat a covariate with missing values as a new response variable and apply the MEEE.

Acknowledgment

This research was partially supported by the National Science Foundation (DMS-1308227). The authors thank the Editor, An associate editor, and two reviewers for helpful comments and suggestions.

Supplementary material

Supplementary material includes regularity conditions and the proofs of Lemma 1 and Theorem 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

References

Anderson, D. A. and Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *J. R. Statist. Soc. B* **47**, 203–10.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. *Studies in Item Analysis and Prediction*, 158-168. Stanford University Press.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.

Chaganty, N. R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93**, 197–206.

Chen, B., Yi, G. Y. and Cook, R. J. (2010). Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *J. Am. Statist. Assoc.* **105**, 336–53.

Cho, H., Wang, P. and Qu, A. (2016). Personalized treatment for longitudinal data using unspecified random-effects model. *Statist. Sinica*, In press.

Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y. and Zheng, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. *Statistical Science* **28**, 238–56.

Diggle, P. J. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Applied Statist.* **43**, 49–93.

Fitzmaurice, G. M., Laird, N. M. and Ware, J. H. (2011). *Applied Longitudinal Analysis*, 2nd ed. Wiley, New York.

Follmann, D. A. and Wu, M. C. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–68.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–54.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed. Springer, New York.

Henderson, R., Diggle, P. J., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465–80.

Hirano, K., Imbens, G. W., Ridder, G., and Rubin, D. B. (2001). Combining panel data sets with attrition and refreshment samples. *Econometrica* **69**, 1645-59.

Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statist. Med.* **16**, 239–57.

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**, 551–64.

Kim, J. K. and Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *J. Am. Statist. Assoc.* **106**, 157–65.

Lee, S.-Y. and Tang, N.-S. (2006). Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statist. Sinica* **16**, 1117–41.

Li, N., Elashoff, R. M., Li, G., and Tseng, C.-H. (2012). Joint analysis of bivariate longitudinal ordinal outcomes and competing risks survival times with nonparametric distributions for random effects. *Stat. Med.* **31**, 1707–21.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalised linear models. *Biometrika* **73**, 13–22.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Lin, H., McCulloch, C. E., and Rosenheck, R. A. (2004a). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60**, 295–305.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Sinha, D., Parzen, M., and Lipshultz, S. (2009). Joint generalized estimating equations for multivariate longitudinal binary outcomes with missing data: An application to acquired immune deficiency syndrome data. *J. R. Statist. Soc. A* **172**, 3–20.

Little, R. J. A. (1994). A class of pattern-mixture models for normal missing data. *Biometrika* **81**, 471–83.

Little, R. J. A. (1995). Modeling the drop-out mechanism in longitudinal studies. *J. Am. Statist. Assoc.* **90**, 1112–21.

Little, R. J. A. (2008). Selection and pattern-mixture models. In *Advances in Longitudinal Data Analysis* (G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs eds). CRC Press, London.

Maruotti, A. (2015). Handling non-ignorable dropouts in longitudinal data: A conditional model based on a latent Markov heterogeneity structure. *Test* **24**, 84–109.

McLeish, D. L. (1975). A maximal inequality and dependent strong laws. *The Annals of Probability* **3**, 829–39.

Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* **84**, 33–44.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *J. Am. Statist. Assoc.* **92**, 1320–29.

Pew Research Center (2010). Four years later Republicans faring better with men, whites, independents, and seniors (press release). Available at http://www.people-press.org/files/legacy-pdf/643.pdf.

Qaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika* **90**, 455-63.

Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–36.

Qu, A., Lindsay, B. G., and Lu, L. (2010). Highly efficient aggregate unbiased estimating functions approach for correlated data with missing at random. *J. Am. Statist. Assoc.* **105**, 194–204.

Qu, A., Yi, G. Y., Song, P. X.-K., and Wang, P. (2011). Assessing the validity of weighted generalized estimating equations. *Biometrika* **98**, 215–24.

Ridder, G. (1992). An empirical evaluation of some models for non-random attrition in panel data. *Structural Change and Economic Dynamics* **3**, 337-55.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95**, 63–74.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–66.

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. Am. Statist. Assoc.* **90**, 106–21.

Robins, J. M., Rotnitzky, A. and Scharfstein, D. O. (2000) Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials.* Springer, New York.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *J. Am. Statist. Assoc.* **93**, 1321–39.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* **59**, 829–36.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–92.

Seaman, S. and Copas, A. (2009). Doubly robust generalized estimating equations for longitudinal data. *Statist. Med.* **28**, 937–55.

Shao, J. and Zhang, J. (2015). A transformation approach in linear mixed-effects models with informative missing responses. *Biometrika* in press.

Spagnoli, A., Henderson, R., Boys, R. J., and Houwing-Duistermaat, J. J. (2011). A hidden Markov model for informative dropout in longitudinal response data with crisis states. *Statistics and Probability Letters* **81**, 730–38.

Stubbendick, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* **59**, 1140–50.

Stubbendick, A. L. and Ibrahim, J. G. (2006). Likelihood-based inference with nonignorable missing responses and covariates in models for discrete longitudinal data. *Statist. Sinica* **16**, 1143–67.

Tsonaka, R., Verbeke, G., and Lesaffre, E. (2009). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Biometrics* **65**, 81–87.

Tsonaka, R., Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2010). Nonignorable models for intermittently missing categorical longitudinal responses. *Biometrics* **66**, 834–44.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* **16**, 219–42.

Vansteelandt, S., Rotnitzky, A., and Robins, J. M. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841–60.

Wang, P., Tsai, G. F., and Qu, A. (2012). Conditional inference functions for mixed-effects models with unspecified random-effects distribution. *J. Am. Statist. Assoc.* **107**, 725–36.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45**, 939–55.

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**, 175–88.

Yuan, Y. and Little, R. J. A. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics* **65**, 478-86.

Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *J. Am. Statist. Assoc.* **107**, 701-10.

Zhou, Y., Little, R, J. A., and Kalbfleisch, J. D. (2010). Block-conditional missing at random models for missing data. *Statist. Science* **25**, 517–32.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Table 1: The absolute bias, standard error and coverage probability of the fixed-effect estimation from 200 replications for the count responses with an exchangeable correlation structure of parameter ρ.

			MEEE	WGEE	MI
$\rho = 0$	β_1	Abs. Bias	0.106	0.112	0.123
		Std. Error	0.130	0.135	0.148
		CP	0.975	0.930	0.945
	β_2	Abs. Bias	0.139	0.143	0.157
		Std. Error	0.168	0.176	0.191
		CP	0.965	0.930	0.920
	β_3	Abs. Bias	0.122	0.143	0.156
		Std. Error	0.155	0.179	0.198
		CP	0.985	0.950	0.955
	β_4	Abs. Bias	0.158	0.176	0.198
		Std. Error	0.201	0.230	0.249
		CP	0.960	0.935	0.915
$\rho = 0.4$	β_1	Abs. Bias	0.118	0.125	0.137
		Std. Error	0.152	0.166	0.179
		CP	0.980	0.935	0.955
	β_2	Abs. Bias	0.150	0.157	0.179
		Std. Error	0.192	0.203	0.233
		CP	0.965	0.930	0.950
	β_3	Abs. Bias	0.107	0.128	0.156
		Std. Error	0.139	0.163	0.196
		CP	0.955	0.920	0.945
	β_4	Abs. Bias	0.134	0.157	0.205
		Std. Error	0.171	0.202	0.260
		CP	0.955	0.930	0.945
$\rho = 0.7$	β_1	Abs. Bias	0.126	0.128	0.178
		Std. Error	0.157	0.162	0.222
		CP	0.980	0.920	0.975
	β_2	Abs. Bias	0.155	0.167	0.260
		Std. Error	0.205	0.218	0.326
		CP	0.950	0.900	0.935
	β_3	Abs. Bias	0.092	0.101	0.272
		Std. Error	0.117	0.130	0.562
		CP	0.945	0.915	0.990
	β_4	Abs. Bias	0.110	0.141	0.645
		Std. Error	0.138	0.177	1.131
		CP	0.960	0.910	0.985

MEEE: mixed-effects estimating equation; PQL: penalized quasi-likelihood; GHQ: adaptive Gaussian-Hermite quadrature; WGEE: weighted generalized estimating equation; MI: multiple imputation; Abs. Bias: absolute bias; Std. Error: standard error; CP: coverage probability. The PQL does not converge due to a small cluster size T, and the GHQ is not applicable since the dimension of parameters is greater than the number of data points.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Table 2: The absolute bias, standard error and coverage probability of the fixed-effect estimation from 200 replications for the binary responses with an AR-1 correlation structure of parameter ρ, where monotonized responses are used for the WGEE.

			MEEE	PQL	GHQ	WGEE	MI
$\rho = 0.2$	β_1	Abs. Bias	0.150	0.214	0.204	0.435	0.172
		Std. Error	0.194	0.272	0.253	0.621	0.173
		CP	0.970	0.933	0.952	0.896	0.935
	β_2	Abs. Bias	0.246	0.401	0.345	0.806	0.413
		Std. Error	0.301	0.483	0.390	1.127	0.225
		CP	0.985	0.867	0.959	0.891	0.715
	β_3	Abs. Bias	0.171	0.311	0.304	0.527	0.148
		Std. Error	0.211	0.341	0.316	0.696	0.181
		CP	0.955	0.860	0.932	0.891	0.980
	β_4	Abs. Bias	0.269	0.379	0.356	0.828	0.405
		Std. Error	0.333	0.471	0.419	1.164	0.230
		CP	0.975	0.927	0.959	0.896	0.790
$\rho = 0.6$	β_1	Abs. Bias	0.217	1.100	5.361	0.515	0.192
		Std. Error	0.280	1.148	7.913	0.658	0.243
		CP	0.935	0.805	0.727	0.873	0.948
	β_2	Abs. Bias	0.333	2.452	8.488	0.750	0.362
		Std. Error	0.424	2.186	11.953	1.050	0.328
		CP	0.960	0.605	0.695	0.923	0.907
	β_3	Abs. Bias	0.177	2.532	7.689	0.577	0.153
		Std. Error	0.234	2.024	10.605	0.737	0.191
		CP	0.910	0.305	0.609	0.845	0.979
	β_4	Abs. Bias	0.306	2.429	4.998	0.862	0.365
		Std. Error	0.402	2.044	6.361	1.246	0.280
		CP	0.975	0.558	0.781	0.901	0.845

MEEE: mixed-effects estimating equation; PQL: penalized quasi-likelihood; GHQ: adaptive Gaussian-Hermite quadrature; WGEE: weighted generalized estimating equation; MI: multiple imputation; Abs. Bias: absolute bias; Std. Error: standard error; CP: coverage probability.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

Table 3: The estimates, standard errors and p-values of fixed effects on respondents' interest in following news about the presidential campaign

Predictor	Statistics	MEEE	PQL	GHQ	GEE	MI	MI*	WGEE*
Intercept	Estimate	-4.352	-5.370	-7.368	-3.181	-2.395	-1.966	-3.187
	Std. Error	0.210	0.236	0.336	0.152	0.149	0.569	0.211
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Time	Estimate	0.112	0.228	0.250	0.111	0.111	0.012	0.110
	Std. Error	0.006	0.007	0.010	0.005	0.005	0.196	0.007
	p-value	0.000	0.000	0.000	0.000	0.000	0.951	0.000
Age	Estimate	0.043	0.047	0.068	0.029	0.021	0.019	0.031
	Std. Error	0.002	0.003	0.004	0.002	0.002	0.003	0.002
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Education	Estimate	0.546	0.616	0.869	0.379	0.292	0.283	0.366
	Std. Error	0.036	0.047	0.064	0.029	0.028	0.036	0.041
	p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gender	Estimate	0.190	0.220	0.313	0.147	0.113	0.063	0.032
	Std. Error	0.063	0.086	0.114	0.053	0.055	0.065	0.075
	p-value	0.003	0.011	0.006	0.006	0.038	0.333	0.668
Household Income	Estimate	0.047	0.058	0.087	0.036	0.025	0.021	0.031
	Std. Error	0.009	0.012	0.015	0.007	0.008	0.008	0.010
	p-value	0.000	0.000	0.000	0.000	0.001	0.007	0.002
Marital Status	Estimate	-0.011	-0.028	-0.001	-0.018	-0.011	-0.034	-0.019
	Std. Error	0.067	0.093	0.123	0.057	0.056	0.062	0.081
	p-value	0.872	0.760	0.997	0.759	0.846	0.577	0.813
MSA Status	Estimate	0.150	0.133	0.199	0.076	0.051	0.030	0.072
	Std. Error	0.083	0.117	0.156	0.071	0.072	0.082	0.097
	p-value	0.069	0.256	0.200	0.284	0.478	0.713	0.457
Black, Non-Hispanic	Estimate	0.598	0.702	1.000	0.438	0.331	0.117	0.428
	Std. Error	0.137	0.167	0.218	0.101	0.106	0.156	0.134
	p-value	0.000	0.000	0.000	0.000	0.002	0.453	0.001
Other, Non-Hispanic	Estimate	-0.226	-0.261	-0.339	-0.130	-0.105	-0.093	-0.428
	Std. Error	0.124	0.176	0.234	0.112	0.102	0.129	0.160
	p-value	0.068	0.138	0.148	0.244	0.300	0.471	0.007
Hispanic	Estimate	0.053	0.078	0.014	0.053	0.016	-0.028	-0.124
	Std. Error	0.120	0.169	0.222	0.104	0.096	0.112	0.142
	p-value	0.657	0.645	0.950	0.613	0.867	0.802	0.383

*Monotonized responses are used, where all follow-ups are deleted once the first missing datum occurs. MEEE: mixed-effects estimating equation; PQL: penalized quasi-likelihood; GHQ: adaptive Gaussian-Hermite quadrature; GEE: generalized estimating equation; MI: multiple imputation; WGEE: weighted generalized estimating equation; Std: Error, standard error.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0317

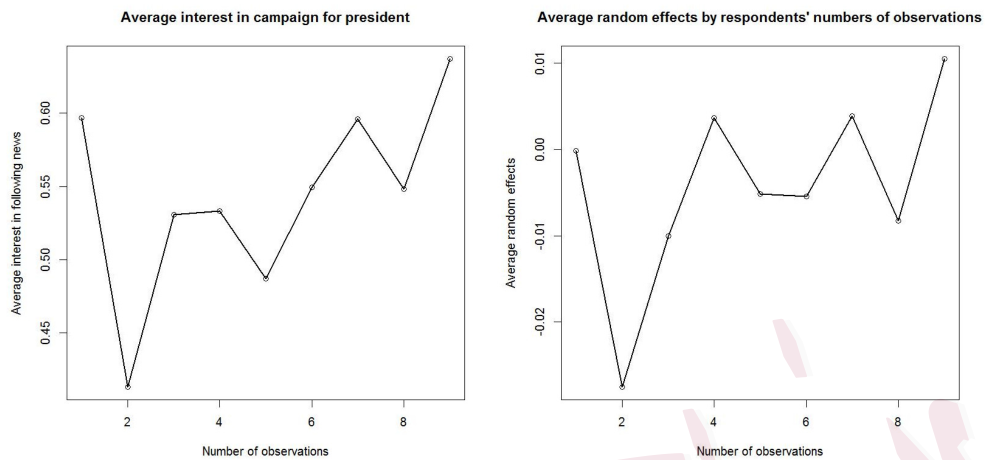

Figure 1: A comparison of the respondents' average interest in the presidential campaign and the average of the estimated random effects by MEEE, both plotted against respondents' number of observed occasions; the right panel is plotted using the same model but without time as a predictor.

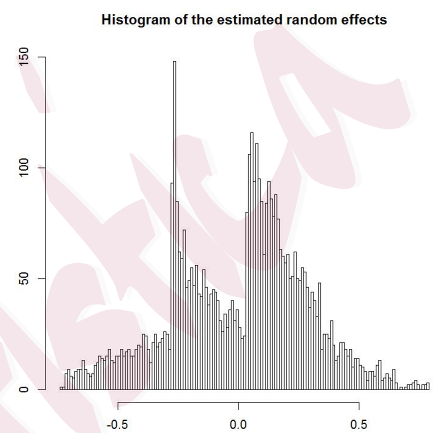

Figure 2: Histogram of the estimated random effects by MEEE.