Pseudo-empirical Likelihood-Based Method Using Calibration for Longitudinal Data with Drop-Out

Baojiang Chen

Department of Biostatistics, University of Nebraska Medical Center, Omaha, NE, US 68198
baojiang.chen@unmc.edu

Xiao-Hua Zhou

Department of Biostatistics, University of Washington, Seattle, Washington, US 98195
Northwest HSR&D Center of Excellence, VA Puget Sound Health Care System, Metropolitan Park West, 1100 Olive Way #1400, Seattle, WA 98101
azhou@u.washington.edu

Kwun Chuen Gary Chan

Department of Biostatistics, University of Washington, Seattle, Washington, US 98195
kcgchan@u.washington.edu

Summary. In observational studies, interest mainly lies in estimation of the population-level relationship between the explanatory variables and dependent variables, and the estimation is often undertaken using a sample of longitudinal data. In some situations, the longitudinal data sample features biases and loss of estimation efficiency due to non-random drop-out. However, inclusion of population-level information can increase estimation efficiency. In this paper we propose an empirical likelihood-based method to incorporate population-level information in a longitudinal study with drop-out. The population-level information is incorporated via constraints on functions of the parameters, and non-random drop-out bias is corrected by using a weighted generalized estimating equations method. We provide a three-step estimation procedure that makes computation easier. Some commonly used methods are compared in simulation studies, which demonstrate that our proposed method can correct the non-random drop-out bias and increase the estimation efficiency, especially for small sample size or when the missing proportion is high. In some situations, the efficiency improvement is substantial. Finally, we apply this method to an Alzheimer's disease study.

Keywords: Calibration; drop-out; empirical likelihood; longitudinal data.

1. Introduction

In observational studies, interest often lies in estimation of the population-level relationship between the explanatory variables and dependent variables. In practice, the estimation is undertaken using sample survey data since it is hard to collect all the required information for the whole population. To improve the estimation efficiency, sources of population-level data such as census data can be used, since these kind of data are comparatively free of sampling error and are often independent of the model parameters. Thus, incorporating this auxiliary information can lead to more accurate estimates and better inferences (Chaudhuri, Handcock, and Rendall, 2008). The motivating example in this paper is the Alzheimer's disease (AD) study introduced in Section 2. The goal is to investigate risk factors that affect the occurrence of dementia. To study this, a sample containing 2313 subjects observed from 2005 to 2008 is used. Data from the National Alzheimer's Coordinating Center

(NACC) Uniform Data Set (UDS) contain population-level information on the dementia proportions from 2005 to 2008. We want to combine the sample data and the population-level information to improve the estimation efficiency in the AD study.

In many situations sample survey data are scheduled to be observed repeatedly, say annually, leading to longitudinal data. During follow-up, subjects may drop out of the study due to, for example, their disease status. For instance, in the AD study, at each clinic visit, a patient may drop out of the study due to developing dementia. Analyses based only on individuals with complete data can lead to invalid inferences if the mechanism leading to the missing data depends on the response. Under a missing completely at random (MCAR) mechanism (Little and Rubin, 2002), analyses based on generalized estimating equations (GEE) (Liang and Zeger, 1986) yield consistent estimators of the regression parameters. However, when data are missing at random (MAR) or missing not at random (MNAR) (Little and Rubin, 2002), analyses based on GEE generally give inconsistent estimators. Robins, Rotnitzky and Zhao (1995) developed a class of inverse probability weighted generalized estimating equations (IPWGEE) which can yield consistent estimators when data are MAR. The weights are obtained from models for the missing data process, and these models must be correctly specified for the resulting estimators to be consistent. However, for the IPWGEE method, it is hard to incorporate population-level auxiliary information, and the estimator is generally sensitive to misspecification of the missing data model when the missing data mechanism is MAR.

Chen and Qin (1993) adapted the empirical likelihood method (Owen, 1988, 1990) to incorporate available auxiliary information in finite population sampling. Qin and Lawless (1994) demonstrated that the empirical likelihood method can be used to solve estimating equations when the number of estimating equations exceeds the number of parameters, and they further defined a profile empirical likelihood of the parameters of interest and obtained estimates by maximizing it directly. Chen and Sitter (1999) introduced a pseudo-empirical likelihood approach to the effective use of auxiliary information in complex surveys. Chaudhuri, Handcock, and Rendall (2008) introduced a two-step estimation procedure for generalized linear models, incorporating population-level information via the empirical likelihood approach, which overcomes the numerical and implementation difficulties of direct constrained maximization. These articles focus on cross-sectional data. For longitudinal data, You et al. (2006) constructed a block empirical likelihood method using independent working correlation matrix; Xue and Zhu (2007) proposed another block empirical likelihood method through centering longitudinal data, in which they do not consider the within-subject correlation structure either; Wang, Qian and Carroll (2010) considered the element-wise and subject-wise empirical likelihood to incorporate the within-subject correlation structure of the repeated measurement, where they consider only the continuous response variable. All the literature on the empirical likelihood method for the longitudinal data does not consider incorporating the auxiliary information and does not consider the missing data problem.

Recently, there has been considerable literature on analysis of missing data using the empirical likelihood method (e.g., Wang and Rao, 2002a, 2002b; Liang, Wang and Carroll, 2007; Stute, Xue and Zhu, 2007; Qin and Zhang, 2007; Qin, Zhang and Leung, 2010). These methods consider only cross-sectional studies, and little work has been devoted to longitudinal studies with missing data. In this paper, we propose a pseudo-empirical likelihood-based method to incorporate the population-level information for longitudinal data with drop-out. This approach is appealing in that it not only extends the previous proposed methods for dealing with incomplete cross-sectional data to the incomplete longitudinal data, but it can improve estimation efficiency through incorporating the population-level information, especially for small sample size or when the missing proportion is high, and sometimes, the efficiency improvement is substantial compared to the IPWGEE method.

The remainder of this paper is organized as follows. In Section 2, we introduce an Alzheimer's

disease study. In Section 3, we introduce notation and models. In Section 4, we give the forms of the estimating equations and provide details on estimation and inference. Simulation studies are described in Section 5. Data arising from an Alzheimer's disease study are analyzed in Section 6. Concluding remarks are made in Section 7.

2. An Alzheimer's disease study

There are approximately 5 million people in the U.S and more than 37 million people worldwide affected by Alzheimer's disease. With Alzheimer's disease, a person's memory and ability to learn and carry out daily activities such as talking and eating are gradually destroyed. As the disease progresses, individuals may also experience changes in personality and behavior. Unfortunately, there exist no cures for Alzheimer's disease nor ways to predict how fast someone will progress through the stages of the disease. However, early Alzheimer's diagnosis and treatment can slow the progression of Alzheimer's symptoms. So, it is desirable to identify the risk factors that affect the occurrence of dementia.

To study the Alzheimer's disease, 29 Alzheimer's Disease Centers (ADCs) have been funded by the National Institute on Aging (NIA), and the National Alzheimer's Coordinating Center (NACC) is responsible for developing and maintaining a database of participant information collected from these ADCs. A standardized clinical data set, called the Uniform Data Set (UDS), was constructed. The goal of the UDS is to provide ADC researchers a standard set of assessment procedures, collected longitudinally, to better characterize ADC participants with mild Alzheimer's disease and mild cognitive impairment in comparison with nondemented controls (Beekly *et al.*, 2007).

A challenge of AD studies is the non-random drop-out. The UDS was implemented from September 2005 as initial visits, and follow-up visits based on Clinical Task Force definitions were developed annually, with up to four clinical visits in 2008. During this study, patients may drop out of the study according to their previously observed degree of disease activity, for example. This creates the drop-out that is missing at random. A sample of 2313 subjects enrolled in 2005 was used to study the relationship between covariates and responses (for more details, see Section 6). The missing proportion is 35.5%. Often, the easily implemented method such as the complete case analysis gives biased inferences.

Another statistical challenge is to use auxiliary information efficiently to improve estimation efficiency. As is generally known, efficiency will be lost due to the smaller sample size caused by missing data. However, using auxiliary information can boost efficiency. For instance, in the NACC UDS, we are interested in investigating risk factors that influence the occurrence of dementia. A random sample of 2313 subjects will be used to study this problem. Besides this sample, population-level auxiliary information, the proportions of dementia in the UDS are known for each year of follow up from 2005 to 2008. Incorporating this auxiliary information can lead to more accurate estimates and better inference, since these auxiliary data are comparatively free of sampling error and are often independent of the model parameters.

The overall goal of this paper is to develop new and effective statistical methods for investigating risk factors that influence the occurrence of dementia in the presence of auxiliary information and non-random drop-out.

3. Notation and models

Consider a sample involving n individuals in which each individual is to be examined at J assessment times. (In the Alzheimer's Disease study, $n = 2313$ and $J = 4$.) Let Y_{ij} denote the response

for subject i at the j th assessment (in the Alzheimer's Disease study, Y_{ij} is binary) and $\mathbf{X}_{ij}$ denote the corresponding covariate vector. For convenience we let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iJ})^T$, and $\mathbf{X}_i = (\mathbf{X}_{i1}^T, \mathbf{X}_{i2}^T, \dots, \mathbf{X}_{iJ}^T)^T$, $i = 1, \dots, n$.

Define $\mu_{ij} = E(Y_{ij}|\mathbf{X}_i)$, and let $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iJ})^T$ denote the full vector of means. We suppose the mean of Y_{ij} depends on the covariate vector for subject i at time j through a model of the form

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta}, \quad (1)$$

for $j = 1, \dots, J$, $i = 1, \dots, n$, where $g(\cdot)$ is a monotone differentiable link function, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients of interest. The variance is expressed as $v_{ij} = \text{var}(Y_{ij}|\mathbf{X}_i) = h(\mu_{ij})$, where $h(\cdot)$ is the variance function. We also make the implicit assumption that $E(Y_{ij}|\mathbf{X}_i) = E(Y_{ij}|\mathbf{X}_{ij})$ and $\text{var}(Y_{ij}|\mathbf{X}_i) = \text{var}(Y_{ij}|\mathbf{X}_{ij})$ as is commonly used (e.g., Pepe and Anderson, 1994).

One challenge for this kind of study is the incomplete data problem. To indicate the availability of data we let $R_{ij} = 1$ if subject i is observed at time point j , and 0 otherwise. Here we consider a monotone missing pattern where $R_{ij} = 0$ indicates that $R_{ik} = 0$ for $k \geq j$. We consider a missing at random (MAR) mechanism for dropout, where we assume

$$P(R_{ij} = 1 | R_{i,j-1} = 1, \mathbf{Y}_i, \mathbf{X}_i) = P(R_{ij} = 1 | R_{i,j-1} = 1, \bar{\mathbf{Y}}_{i,j-1}, \bar{\mathbf{X}}_{i,j-1}), \quad (2)$$

where $\bar{\mathbf{Y}}_{ij} = \{Y_{i1}, \dots, Y_{ij}\}$ and $\bar{\mathbf{X}}_{ij} = \{\mathbf{X}_{i1}, \dots, \mathbf{X}_{ij}\}$. Let $\lambda_{ij} = P(R_{ij} = 1 | R_{i,j-1} = 1, \bar{\mathbf{Y}}_{i,j-1}, \bar{\mathbf{X}}_{i,j-1})$, we then can build a generalized linear model for λ_{ij} , say, logistic regression, as follows:

$$\text{logit } \lambda_{ij} = \mathbf{u}_{ij}^T \boldsymbol{\alpha}, \quad (3)$$

where $\mathbf{u}_{ij}$ is a function of $\{\bar{\mathbf{Y}}_{i,j-1}, \bar{\mathbf{X}}_{i,j-1}\}$, and $\boldsymbol{\alpha}$ is an unknown parameter vector. Let $\pi_{ij} = \prod_{k=1}^j \lambda_{ik}$ be the marginal probability that subject i at time j is observed.

4. Pseudo-empirical likelihood for longitudinal data in the presence of population-level information

4.1. Weighted generalized estimating equation

If there is no missing data, we can build a generalized estimating equation to solve for the parameter of interest, $\boldsymbol{\beta}$, as follows:

$$\sum_{i=1}^n D_i V_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (4)$$

where $D_i = \partial \boldsymbol{\mu}_i^T / \partial \boldsymbol{\beta}$, V_i is a working covariance matrix which is often expressed as $V_i = \phi \text{diag}(v_{ij}, j = 1, \dots, J)^{1/2} C_i \text{diag}(v_{ij}, j = 1, \dots, J)^{1/2}$, ϕ is a dispersion parameter, and C_i is a working correlation matrix.

With a monotone missing data pattern, we may employ the IPWGEE method in the same spirit of Robins, Rotnitzky and Zhao (1995). Specifically, the IPWGEE is given by $\sum_{i=1}^n \mathbf{g}_1(\mathbf{Y}_i, \mathbf{X}_i) = \mathbf{0}$ where $\mathbf{g}_1(\mathbf{Y}_i, \mathbf{X}_i) = D_i V_i^{-1} \Delta_i^w (\mathbf{Y}_i - \boldsymbol{\mu}_i)$, $\Delta_i^w = \text{diag}(R_{ij}/\pi_{ij}(\boldsymbol{\alpha}), j = 1, \dots, J)$, and $\pi_{ij}(\boldsymbol{\alpha}) = \prod_{k=1}^j \lambda_{ik}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a parameter from the missing data model. When $\boldsymbol{\alpha}$ is unknown, estimation can be obtained by maximizing the log-likelihood for the parameter $\boldsymbol{\alpha}$: $\sum_{i=1}^n \ell_i(\boldsymbol{\alpha})$, where

$$\ell_i(\boldsymbol{\alpha}) = \left\{ \sum_{j=1}^{t_i-1} \log(\lambda_{ij}) + \log(1 - \lambda_{it_i}) \right\},$$

and $t_i = \min\{j : R_{ij} = 0\}$. Equivalently, we solve the estimating equation $\mathbf{S}(\boldsymbol{\alpha}) = \sum_{i=1}^n \mathbf{S}_i(\boldsymbol{\alpha}) = \mathbf{0}$, where $\mathbf{S}_i(\boldsymbol{\alpha}) = \partial \ell_i(\boldsymbol{\alpha}) / \partial \boldsymbol{\alpha}^T$. This procedure can be implemented by using available software. Denote the estimator $\hat{\boldsymbol{\alpha}}$. Thus we can obtain the estimated marginal probability $\hat{\pi}_{ij}$.

In practice, the working correlation matrix is usually unknown and must be estimated. Liang and Zeger (1986) showed that it can be estimated in the iterative fitting process by using the current value of the parameter vector to compute appropriate functions of the Pearson residual $e_{ij} = (y_{ij} - \mu_{ij}) / \sqrt{v_{ij}/w_{ij}}$, where $w_{ij} = r_{ij}/\pi_{ij}$. For example, with the unstructured working correlation matrix $C_i = (c_{jk})_{J \times J}$, $\hat{c}_{jk} = 1$ if $j = k$, and $\hat{c}_{jk} = 1/\{(n-p)\hat{\phi}\} \sum_{i=1}^n e_{ij}e_{ik}$, where $\hat{\phi} = 1/(nJ-p) \sum_{i=1}^n \sum_{j=1}^J e_{ij}^2$. For the other correlation structure estimates, please see Liang and Zeger (1986). Besides this simple averages of observed cross-products estimates, there are other estimates for the correlations such as the estimates by taking the variance of the cross-products into account proposed by Prentice (1988), the estimates by using the multivariate normal (Gaussian) estimating equations (Lee, Laird and Johnston, 1999; Hall and Severini, 1998; Wang and Carey, 2003). The latter approach (based on Gaussian estimating equations) has been shown to have less bias because it ensures a positive definite estimate of V_i ; whereas the first two methods do not.

Under the MAR mechanism, if the missing data model is correctly specified, we can obtain a consistent estimator for β by solving the IPWGEE, provided the mean of the response μ_i is correctly specified.

4.2. Population-level auxiliary information

In a finite population study, inference is often based on a sample of the finite population. Besides this sample, some population quantities (population-level auxiliary information) are often known. A disadvantage of the IPWGEE is that it does not incorporate finite population-level auxiliary information into the estimation. Suppose that there is a known vector function $\mathbf{h}_{ij}$ at each time point j , for which $E_j[\mathbf{h}_{ij}(Y_{ij}, \mathbf{X}_{ij}, \mathbf{A}_j)] = \mathbf{0}$ with known parameter $\mathbf{A}_j$, $j = 1, \dots, J$, where $\mathbf{A}_j$ represents the finite population-level information at time point j . For example, in the Alzheimer's disease study in Section 6, we know the finite population means of the response $A_j = E(Y_{ij})$, the proportions of dementia in the UDS from 2005 to 2008, which are 69.6%, 63.1%, 63.1%, and 66.4%, respectively. Then we have $h_{ij} = Y_{ij} - A_j$, where $A_1 = 69.6\%$, $A_2 = 63.1\%$, $A_3 = 63.1\%$, and $A_4 = 66.4\%$. Let $\mathbf{h}_i(\mathbf{Y}_i, \mathbf{X}_i) = (\mathbf{h}_{i1}^T, \dots, \mathbf{h}_{iJ}^T)^T$. The auxiliary information can be summarized by the moment conditions $E(\mathbf{g}_2(\mathbf{Y}_i, \mathbf{X}_i)) = \mathbf{0}$ where $\mathbf{g}_2(\mathbf{Y}_i, \mathbf{X}_i) = E(\Delta_i^w \mathbf{h}_i(\mathbf{Y}_i, \mathbf{X}_i))$.

4.3. Pseudo-empirical likelihood

Since the estimating functions $(\mathbf{g}_1^T, \mathbf{S}^T)^T$ define a just-identified system of moment conditions, the estimating functions $\mathbf{g} = (\mathbf{g}_1^T, \mathbf{g}_2^T, \mathbf{S}^T)^T$ defines an overidentified system of moment conditions. To solve an overidentified system efficiently, one can employ the empirical likelihood method (Qin and Lawless, 1994). Empirical likelihood attains the minimum asymptotic variance in the class of all estimating equations which are linear combinations of $\mathbf{g}_1$, $\mathbf{g}_2$ and $\mathbf{S}$ (Qin and Lawless, 1994), which makes it at least as efficient as the IPWGEE estimator which solves $(\mathbf{g}_1^T, \mathbf{S}^T)^T$. Using empirical likelihood, including the population-level auxiliary information is always at least as efficient as without using it. Although empirical likelihood has such a nice theoretical property that motivates the inclusion of population-level information, solving an arbitrary overidentified system of estimating estimations may be a non-convex problem and can pose serious computational issues. From a practical and computational perspective, we adopt a pseudo-empirical likelihood method in the same spirit of Chen and Sitter (1999) and a three-step procedure in the same spirit of Chaudhri, Handcock, and Rendall (2008).

In the absence of the missing data, Wang, Qian and Carroll (2010) proposed an element-wise empirical likelihood method, where they maximize the element-wise empirical likelihood

$$L = \prod_{j=1}^J \prod_{i=1}^n p_{ij}, \quad (5)$$

with $p_{ij} \geq 0$, $\sum_{j=1}^J \sum_{i=1}^n p_{ij} = 1$ and other constraints. We comment that we can incorporate more complex constraints by using the element-wise empirical probabilities p_{ij} . For example, if we know the cross time constraint $E[\sum_{j=1}^J h_{ij}(Y_{ij}, \mathbf{X}_{ij}, \mathbf{A}_j)] = 0$, then the constraint in the element-wise empirical likelihood is $\sum_{i=1}^n \sum_{j=1}^J p_{ij} h_{ij}(Y_{ij}, \mathbf{X}_{ij}, \mathbf{A}_j) = 0$. We now extend this method to the missing data problem as follows.

Let n_j denote the number of subjects who are observed up to (and including) j but missing after j , $j = 1, \dots, J$. For simplicity, we assume the first n_j subjects are observed at time j . Define p_{ij} as the probability mass for subject i at time j for $j = 1, \dots, J$, and $i = 1, \dots, n$. We define a modified pseudo-empirical likelihood (Chen and Sitter, 1999) for the observed longitudinal data as

$$L = \prod_{j=1}^J \prod_{i=1}^n p_{ij}^{r_{ij}/\pi_{ij}}. \quad (6)$$

The motivation for (6) is from the inverse probability weighting method (Robins, Rotnitzky and Zhao, 1995). The idea behind this method is that the weight has eliminated the bias, by “reconstructing” the log-empirical likelihood by upweighting the data from subjects who have a small chance of being observed. To be specific, by taking expectation of the log-pseudo-empirical likelihood (6) with respect to the missing data probability, we can obtain the log-empirical likelihood $\sum_{j=1}^J \sum_{i=1}^n \log(p_{ij})$, by noting that $E(r_{ij}/\pi_{ij}) = 1$ if the missing data model is correctly specified.

From the properties of a distribution function it follows that $p_{ij} \geq 0$ and $\sum_{j=1}^J \sum_{i=1}^n p_{ij} = 1$. However, with missing data, p_{ij} ($i > n_j$) are not estimable, so we must add constraints to the observed mass p_{ij} . Again, motivated by the inverse probability weighting method, we upweight the data from subjects who have a small chance of being observed. Therefore, we would rather use the following constraints

$$p_{ij} \geq 0, \sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} = 1. \quad (7)$$

Similarly, for the population-level information, we add one more constraint that

$$\sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} h_{ij}(Y_{ij}, \mathbf{X}_{ij}, \mathbf{A}_j) = 0, \quad (8)$$

for $j = 1, \dots, J$.

In the same spirit of Chaudhuir, Handcock, and Rendall (2008), we propose a three-step estimation procedure as follows:

- (a) In the first step, we estimate the missing data parameter α via a maximum likelihood method, then calculate the estimated weights $\hat{\pi}_{ij}$.
- (b) In the second step, we maximize the pseudo-empirical likelihood (6) on the simplex, under the constraints (7) and (8), with respect to p_{ij} . Denote the estimated value $\hat{p}_{ij}$.

(c) In the third step we solve the following weighted GEE for the parameter β ,

$$\sum_{i=1}^n D_i \Delta_i V_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \quad (9)$$

with the weights $\Delta_i = \text{diag}(p_{ij} R_{ij} / \pi_{ij}, j = 1, \dots, J)$, replaced by $\hat{\Delta}_i = \text{diag}(\hat{p}_{ij} R_{ij} / \hat{\pi}_{ij}, j = 1, \dots, J)$. Similarly, we can estimate the working correlation matrix by using the Pearson residual $e_{ij} = (y_{ij} - \mu_{ij}) / \sqrt{v_{ij} / w_{ij}}$, where $w_{ij} = p_{ij} r_{ij} / \pi_{ij}$. For example, with the unstructured working correlation matrix $C_i = (c_{jk})_{J \times J}$, $\hat{c}_{jk} = 1$ if $j = k$, and $\hat{c}_{jk} = 1 / \{(n-p)\hat{\phi}\} \sum_{i=1}^n e_{ij} e_{ik}$, where $\hat{\phi} = 1 / (nJ - p) \sum_{i=1}^n \sum_{j=1}^J e_{ij}^2$. For the other correlation structure estimates, please see Liang and Zeger (1986).

To improve efficiency, if the dimension of the covariates α is not high, we may combine the first and second step and estimate the parameter α and p_{ij} jointly, but we expect the gain in efficiency is small and computation burden can increase greatly. So, we do not further consider this case in this paper.

It is easy to verify that there is a unique maximizing $\hat{\mathbf{p}} = (\hat{p}_{ij}, j = 1, \dots, J, i = 1, \dots, n_j)$ under the constraints (7) and (8). However, it is possible that there is no β satisfying the parametric constraints for weights $\hat{\Delta}_i$ (Small and Wang, 2003), unless the regularity condition (d) in the Appendix are satisfied. This condition means that for each subject i , the probability of remaining in the study at each time is bounded away from 0, which guarantees the existence of the $n^{1/2}$ consistency estimate of β (Robins, Rotnitzky and Zhao, 1995). This is the same limitation in existing methods (e.g., Robins, Rotnitzky and Zhao, 1995). In that situation, the three-step procedures may fail. When it happens, people may consider ignoring the information at that time point as there is almost no information at that time point. Empirical evidence suggests that the three-step procedures work well for most cases (Chaudhuri, Handcock and Rendall, 2008).

4.4. Estimation and inference

The first step involves estimating the missing data model parameter α by solving $\sum_{i=1}^n \mathbf{S}_i(\alpha) = \mathbf{0}$ as introduced in Section 4.1, which can be implemented using available software.

Note that the second step requires constrained maximization. For computational purposes, we maximize the log pseudo-empirical likelihood

$$\ell r = \sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} \log(p_{ij}), \quad (10)$$

subject to the constraints (7) and (8).

By introducing the Lagrange multipliers η_0 and $\boldsymbol{\eta}_j$, $j = 1, \dots, J$, maximization (10) subject to the constraints (7) and (8) can be translated into maximization of an unconstrained one with objective function

$$\ell u = \sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} \log(p_{ij}) - \eta_0 \left(\sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} - 1 \right) - N \sum_{j=1}^J \boldsymbol{\eta}_j^T \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} \mathbf{h}_{ij}, \quad (11)$$

where $N = \sum_{i=1}^n \sum_{j=1}^J r_{ij} / \pi_{ij}$. Taking the first derivative of (11) with respect to p_{ij} and setting it to zero, we obtain

$$\frac{r_{ij}}{\pi_{ij} p_{ij}} - \eta_0 \frac{r_{ij}}{\pi_{ij}} - N \frac{r_{ij}}{\pi_{ij}} \boldsymbol{\eta}_j^T \mathbf{h}_{ij} = 0. \quad (12)$$

Multiplying p_{ij} on both sides of (12) and taking summation with respect to i and j , we obtain

$$\sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} - \eta_0 \sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} - N \sum_{j=1}^J \boldsymbol{\eta}_j^T \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} \mathbf{h}_{ij} = 0,$$

which leads to $\eta_0 = N$ by noting that $\sum_{j=1}^J \sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} = 1$ and $\sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} \mathbf{h}_{ij} = 0$. By plugging $\eta_0 = N$ into (12) and solving (12) for p_{ij} , we can obtain

$$p_{ij} = \frac{1}{N(1 + \boldsymbol{\eta}_j^T \mathbf{h}_{ij})},$$

for $r_{ij} = 1$. As p_{ij} is defined only on $r_{ij} = 1$, for expression simplicity, we set $p_{ij} = 0$ if $r_{ij} = 0$. So, we have

$$p_{ij} = \frac{r_{ij}}{N(1 + \boldsymbol{\eta}_j^T \mathbf{h}_{ij})}, j = 1, \dots, J, i = 1, \dots, n. \quad (13)$$

Since the Lagrange multipliers $\boldsymbol{\eta}_j$ are unknown, we need to estimate them. By plugging (13) into (8), we obtain the estimating equations

$$\sum_{i=1}^n \frac{r_{ij} \mathbf{h}_{ij}}{N \pi_{ij} (1 + \boldsymbol{\eta}_j^T \mathbf{h}_{ij})} = \mathbf{0}, j = 1, \dots, J. \quad (14)$$

Let $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \dots, \boldsymbol{\eta}_J^T)^T$, and $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\eta}}_1^T, \dots, \hat{\boldsymbol{\eta}}_J^T)^T$ be the estimator which solves estimating equations (14) for $\boldsymbol{\eta}$.

Then the estimated probability (13) is

$$\hat{p}_{ij} = \frac{r_{ij}}{\hat{N}(1 + \hat{\boldsymbol{\eta}}_j^T \mathbf{h}_{ij})}, j = 1, \dots, J, i = 1, \dots, n,$$

where $\hat{N} = \sum_{i=1}^n \sum_{j=1}^J r_{ij} / \hat{\pi}_{ij}$.

The third step consists of the usual estimation of the parameters for the weighted GEE (9) with $\hat{\Delta}_i = \text{diag}(\hat{p}_{ij} R_{ij} / \hat{\pi}_{ij}, j = 1, \dots, J)$, which can also be implemented by using existing software, such as the function `geeglm` in the R package `geepack`. Denote the estimator of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}$.

We now develop the asymptotic properties of $\hat{\boldsymbol{\beta}}$. Let $\mathbf{U}_i(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) = D_i \Delta_i V_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i)$,

$$\mathbf{H}_{ij}(\boldsymbol{\eta}_j, \boldsymbol{\alpha}) = \frac{r_{ij} \mathbf{h}_{ij}}{\pi_{ij} (1 + \boldsymbol{\eta}_j^T \mathbf{h}_{ij})},$$

and

$$\mathbf{H}_i(\boldsymbol{\eta}, \boldsymbol{\alpha}) = (\mathbf{H}_{i1}^T(\boldsymbol{\eta}_1, \boldsymbol{\alpha}), \dots, \mathbf{H}_{iJ}^T(\boldsymbol{\eta}_J, \boldsymbol{\alpha}))^T.$$

Define

$$\begin{aligned} \Gamma(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) &= E[\partial \mathbf{U}_i^T(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\beta}], \\ I_{12}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) &= E[\partial \mathbf{U}_i^T(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\eta}], \\ I_{13}(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) &= E[\partial \mathbf{U}_i^T(\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\alpha}], \\ I_2(\boldsymbol{\eta}, \boldsymbol{\alpha}) &= E[\partial \mathbf{H}_i^T(\boldsymbol{\eta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\eta}], \\ I_{23}(\boldsymbol{\eta}, \boldsymbol{\alpha}) &= E[\partial \mathbf{H}_i^T(\boldsymbol{\eta}, \boldsymbol{\alpha}) / \partial \boldsymbol{\alpha}], \end{aligned}$$

$$I_3(\alpha) = E[\partial \mathbf{S}_i^T(\alpha) / \partial \alpha],$$

$$\begin{aligned} \mathbf{Q}_i(\beta, \eta, \alpha) &= \mathbf{U}_i(\beta, \eta, \alpha) - I_{12}(\beta, \eta, \alpha) I_2^{-1}(\eta, \alpha) \mathbf{H}_i(\eta, \alpha) \\ &\quad + I_{12}(\beta, \eta, \alpha) I_2^{-1}(\eta, \alpha) I_{23}(\eta, \alpha) I_3^{-1}(\alpha) \mathbf{S}_i(\alpha) - I_{13}(\beta, \eta, \alpha) I_3^{-1}(\alpha) \mathbf{S}_i(\alpha). \end{aligned}$$

Theorem. Suppose that $(\mathbf{Y}_1^T, \mathbf{X}_1^T)^T, \dots, (\mathbf{Y}_n^T, \mathbf{X}_n^T)^T$ are independent and identically distributed random vectors from the population space governed by model (1). Assuming the regularity conditions stated in the Appendix hold and the missing data model is correctly specified, we have the following asymptotic results: As $n \rightarrow \infty$,

$$n^{1/2}(\hat{\beta} - \beta_0) \rightarrow N(0, \Gamma^{-1}(\beta_0, \mathbf{0}, \alpha_0) \Sigma [\Gamma^{-1}(\beta_0, \mathbf{0}, \alpha_0)]^T),$$

where β_0 and α_0 are the true values of β and α , respectively, and

$$\Sigma = E[\mathbf{Q}_i(\beta_0, \mathbf{0}, \alpha_0) \mathbf{Q}_i^T(\beta_0, \mathbf{0}, \alpha_0)].$$

The proof is given in the Appendix. To make inferences, we consistently estimate the matrix Γ by

$$\hat{\Gamma} = n^{-1} \sum_{i=1}^n [\partial \mathbf{U}_i^T(\hat{\theta}) / \partial \beta],$$

and Σ by $\hat{\Sigma} = n^{-1} \sum_{i=1}^n [\hat{\mathbf{Q}}_i \hat{\mathbf{Q}}_i^T]$, where $\hat{\theta} = (\hat{\beta}^T, \hat{\eta}^T, \hat{\alpha}^T)^T$, and $\hat{\mathbf{Q}}_i$ is $\mathbf{Q}_i$ with $\hat{\theta}$ replaced.

5. Simulations

We first consider a longitudinal study with binary Y and with $J = 3$, where we suppose there is a superpopulation with sample size $N = 10000$, which is generated from the model

$$\text{logit } \mu_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{ij2},$$

where $\mu_{ij} = E(Y_{ij} | X_{i1}, X_{ij2})$, X_{i1} is a time-independent covariate generated from $\text{Bin}(1, 0.5)$, and $X_{ij2} \sim N(X_{i1}, 1)$. We assume the true correlation matrix is exchangeable with correlation coefficient ρ . The true value is $\beta_0 = \log(1.5)$, $\beta_1 = 1$, and $\beta_2 = 1$. The study sample of size n is randomly sampled from the superpopulation.

The missing data model is

$$\text{logit } \lambda_{ij} = \alpha_0 + \alpha_1 Y_{i,j-1}, \quad (15)$$

where $\lambda_{ij} = P(R_{ij} = 1 | R_{i,j-1} = 1, Y_{i,j-1})$, and the true value $\alpha_0 = 0$; we vary α_1 to represent the different missing proportions. We assume all the subjects are observed at time 1.

Suppose we know the population mean at each time point $A_j = E(Y_{ij})$, $j = 1, \dots, J$; the constraints are

$$\sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij} (Y_{ij} - A_j) = 0, \quad j = 1, 2, 3.$$

We compare the following methods: (a) The proposed method by using the unstructured working correlation matrix; (b) The IPWGEE method based on Robins, Rotnitzky and Zhao (1995) by using weight matrix $\Delta_i^w = \text{diag}(R_{ij}/\pi_{ij}, j = 1, \dots, 3)$ with the unstructured working correlation matrix; (c) The available data GEE method with the unstructured working correlation matrix; and (d) the available data Gaussian estimating equation (Gaussian EE) method with the unstructured

working correlation matrix, in which the correlation matrix is estimated by using the multivariate normal (Gaussian) estimating equations (Lee, Laird and Johnston, 1999; Hall and Severini, 1998; Wang and Carey, 2003).

Table 1 lists the results with correctly specified missing data model (15), where SD is the standard deviation for the 2000 simulations, and CP is the coverage probability for the nominal level 95%. It can be seen that, as expected, the proposed and IPWGEE methods give consistent estimates and good coverage probabilities under all situations, but the available data GEE method give biased estimates for the intercept (β_0) and the time invariant coefficient (β_1). However, in this situation, the available GEE method yields small biases for the time-varying coefficient (β_2). The Gaussian EE method also gives big biases, especially for β_0 and β_1 . The proposed method yields more efficient estimates than the IPWGEE method, especially for small sample sizes; when the sample size increases, the efficiency gain decreases; as the missing proportion increases (α_1 decreases), the efficiency gain increases; the efficiency gain is also dependent on the response correlation ρ : as ρ increases, the efficiency gain increases. All indicate that incorporating auxiliary information will yield more efficient estimates.

Insert Tables 1 here

Next, we consider a continuous longitudinal data where we assume Y is normally distributed and with $J = 3$, where we suppose there is a superpopulation with sample size $N = 10000$, which is generated from the mean model

$$\mu_{ij} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{ij2},$$

with standard deviation 1, where X_{i1} is a time-independent covariate generate from $Bin(1, 0.5)$, and $X_{ij2} \sim N(X_{i1}, 1)$. We also assume the true correlation matrix is exchangeable with correlation coefficient ρ . The true value is $\beta_0 = \log(1.5)$, $\beta_1 = 1$, and $\beta_2 = 1$. The study sample of size n is randomly sampled from the superpopulation. The missing data model is the same as (15). We consider the population mean at each time point, A_j , $j = 1, \dots, J$, is known, and the constraints are

$$\sum_{i=1}^n \frac{r_{ij}}{\pi_{ij}} p_{ij}(Y_{ij} - A_j) = 0, \quad j = 1, 2, 3.$$

Insert Table 2 here

Table 2 lists the results with correctly specified missing data model (15). Similarly, the proposed and IPWGEE methods give consistent estimates and good coverage probabilities under all situations, but the available data GEE method gives biased estimates. The Gaussian EE method also gives big biases when $n = 100$, $\rho = 0.9$, and $\alpha_1 = -4$. For the other cases, the Gaussian EE method yields small biases. The proposed methods yield more efficient estimates than the IPWGEE method and the Gaussian EE method, especially for small sample sizes; when the sample size increases, the efficiency gain decreases; as the missing proportion increases (α_1 decreases), the efficiency gain increases; the efficiency gain is also dependent on the response correlation ρ : as ρ increases, the efficiency gain increases. All indicate that incorporating auxiliary information will yield more efficient estimates.

In conclusion, the available data GEE analysis, in general, gives biased estimates and poor coverage probabilities when data are missing at random, but the proposed and IPWGEE methods yield negligible biases and good coverage probabilities when the missing data model is correctly specified. In all cases in this simulation study, the proposed method yields more efficient estimates than the IPWGEE method. This is expected since auxiliary information is incorporated in the analysis in the proposed method. Also, in this simulation study, the available data Gaussian estimating equation method yields biased estimates when the outcome is binary with high correlations; however, it performs better when the outcome is continuous.

6. Application to an Alzheimer's disease study

We apply the proposed method to the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS). One of the goals of the study is to investigate the risk factors that influence the occurrence of dementia. The response of interest is the clinical diagnosis for dementia (Yes/No), because definitive diagnosis of dementia is possible only through neuropathological examination, which is not available for the majority of subjects. The covariates that may influence the occurrence of dementia include sex, family history of dementia (fhdem, yes/no), diabetes (yes/no), behavioral assessment (depression or dysphoria, yes/no), hypertension (yes/no), education (years), and age, in which diabetes, behavioral assessment, and hypertension are time-varying, and the others are not. The proportions (i.e., $E(Y_{ij})$) of dementia in the UDS from 2005 to 2008 are 69.6%, 63.1%, 63.1%, 66.4%, respectively.

A sample of 2313 subjects enrolled in 2005 is used to study the association between the responses and covariates. Follow-up visits for subjects are scheduled at approximately one-year intervals, with up to four clinical visits until 2008. During the study, patients may drop out of the study. We think the MAR mechanism may be reasonable because, for example, a patient's visit to a clinic may depend on his/her previous observed diagnosis of disease status: if s/he was diagnosed dementia last year, s/he may be likely to attend the clinic visit this year in order to control the disease efficiently. The missing proportions in 2006, 2007 and 2008 are 29.8%, 35.4% and 39.6%, respectively.

Consider the following regression model for the response process

$$\text{logit } \mu_{ij} = \mathbf{u}_{ij}^T \boldsymbol{\beta},$$

where $\mathbf{u}_{ij}$ is the covariate vector at time point j , which includes sex, fhdem, diabetes, depression, hypertension, education, age, and the indicator variables for centers.

For the drop-out process, we build the following regression models:

$$\text{logit } \lambda_{ij} = \mathbf{v}_{ij}^T \boldsymbol{\alpha},$$

where $\mathbf{v}_{ij}$ include sex, fhdem, diabetes, hypertension, education, age, and previous response.

In line with the simulation study, here we also use four methods to analyze the data. The first method is the proposed method based on the weight matrix Δ_i with the unstructured working correlation matrix; the second method is the IPWGEE method based on the weight matrix Δ_i^w with the unstructured working correlation matrix; the third method is the available data GEE analysis with the unstructured working correlation matrix; and the fourth method is the available data Gaussian estimating equation method. The results are reported in Table 3. Here we only list the effects for the risk factors and omit the center effects though some of them are highly significant. It can be seen that the proposed and the IPWGEE methods give close estimates, and they give different results from the available data GEE and Gaussian estimating equation analyses, since the missing

completely at random mechanism is not feasible (see Table 4), so the GEE and Gaussian estimating equation methods may yield biased estimates as seen in the simulation studies. The proposed method yields more efficient estimate than the IPWGEE method, though they provide similar coefficient estimates. Both methods reveal that SEX, family history of dementia, diabetes, hypotension, and education have no significant effect on the occurrence of dementia; depression and age have positive effects on the occurrence of dementia.

For the missing data model, we carry out standard diagnostic tests for the fit of regression models by comparing a model with an expanded model to do a model selection. Here, we only list the results for the final model. Significance of sex, depression, education, age, and the previous diagnosis of dementia, indicating that the missing completely at random mechanism is not feasible. So, we can draw the conclusions that the complete case analysis can draw a wrong conclusion about a relationship between a risk factor and the occurrence of dementia.

Insert Tables 3 and 4 here

7. Discussion

In this paper, we propose an empirical likelihood-based method to incorporate population-level information in a longitudinal study with non-random drop-out. This method has been proved to be robust to misspecification of the working correlation matrix under the missing at random mechanism. Theoretical and simulation studies demonstrate that incorporating the population-level information can improve the estimation efficiency, especially when the sample size is small or when the missing proportion is high. In some situations, the efficiency improvement is substantial. Application to an Alzheimer's disease study demonstrates the usefulness of the proposed method.

We focussed here primarily on estimation and inference regarding drop-out. Intermittent missing data are very common in practice. The proposed method can be extended to this situation by properly modeling the missing data model. However, this is a challenging problem. Another popular method for handling the intermittent missing data is the multiple imputation (Rubin, 1987). The proposed method can apply to the imputed data by adding an appropriate weight obtained from the auxiliary information. This will be under investigation in the future.

This paper considers only the missing at random mechanism. In practice, we often face data that are missing not at random. Another future research direction will be to extend this method to the missing not at random mechanism.

Since the Alzheimer's disease study is of the elderly, there is a challenge related to the reason for dropout. Dropout due to death is very different from dropout due to other causes. In our example in this paper, we only consider dropout due to other causes (subjects were still alive). A future study may focus on developing methods for the dropout due to death and other reasons.

Acknowledgements

The authors thank the Editor, Associated Editor and two referees for their constructive comments and suggestions. Dr. Xiao-Hua Zhou, Ph.D., is presently a Core Investigator and Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. Dr. Zhou was supported in part by Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Research Career Scientist

Award RCS 05-196. Both Drs. Zhou and Chen were supported in part by National Institute on Aging grant U01AG016976. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service.

Appendix: Regularity conditions and proof of Theorem*Regularity conditions:*

- (a) $(\mathbf{Y}_i^T, \mathbf{X}_i^T)^T, i = 1, \dots, n$ are independent and identically distributed random variable vectors governed by model (1);
- (b) $\beta_0, \mathbf{0}$, and α_0 lie in the interior of compact sets of β, η and α ;
- (c) $E[\mathbf{U}_i^*(\beta, \eta, \alpha)] \neq 0$ if $(\beta, \eta, \alpha) \neq (\beta_0, \mathbf{0}, \alpha_0)$, where

$$\mathbf{U}_i^*(\beta, \eta, \alpha) = (\mathbf{U}_i^T(\beta, \eta, \alpha), \mathbf{H}_i^T(\eta, \alpha), \mathbf{S}_i^T(\alpha))^T;$$

- (d) $P(R_{ij} = 1 | \mathbf{Y}_i, \mathbf{X}_i)$ are bounded away from zero;
- (e) $\text{var}[\mathbf{U}_i^*(\beta_0, \mathbf{0}, \alpha_0)]$ is finite and positive definite;
- (f) The matrices I_2, I_3 and Γ are non-singular;
- (g) $\mathbf{U}_i, \mathbf{H}_i, \mathbf{S}_i, \partial \mathbf{U}_i^T / \partial(\beta, \eta, \alpha), \partial \mathbf{H}_i^T / \partial \eta$ and $\partial \mathbf{S}_i^T / \partial \alpha$ are bounded by integrable function in a neighborhood of $(\beta_0, \mathbf{0}, \alpha_0)$;
- (h) $E[\partial \mathbf{U}_i^*(\beta_0, \mathbf{0}, \alpha_0) / \partial(\beta^T, \eta^T, \alpha^T)^T]$ exists and is invertible;
- (i) $E[\sup \|\mathbf{U}_i^*(\beta, \eta, \alpha)\|], E[\sup \|\partial \mathbf{U}_i^*(\beta, \eta, \alpha) / \partial(\beta^T, \eta^T, \alpha^T)^T\|], E[\sup \|\mathbf{U}_i^*(\beta, \eta, \alpha) \mathbf{U}_i^*(\beta, \eta, \alpha)^T\|]$, are all finite in the Cartesian product of β, η and α , where $\|\mathbf{A}\| \equiv [\sum_{ij} A_{ij}^2]^{1/2}$ for any matrix A with element A_{ij} ;
- (j) $L_i(\beta, \eta, \alpha)$ is a regular parametric model for the observed data, where $L_i(\beta, \eta, \alpha)$ is a density that differs from the true density only in that (β, η, α) replaces $(\beta_0, \mathbf{0}, \alpha_0)$;
- (k) For all $(\beta^*, \eta^*, \alpha^*)$ in a neighborhood N of $(\beta_0, \mathbf{0}, \alpha_0)$, $E_{(\beta^*, \eta^*, \alpha^*)}[\mathbf{U}_i^*(\beta, \eta, \alpha)]$ and $E_{(\beta^*, \eta^*, \alpha^*)}[\sup_{(\beta, \eta, \alpha) \in N} \|\mathbf{U}_i^*(\beta, \eta, \alpha) \mathbf{U}_i^*(\beta, \eta, \alpha)^T\|]$ are bounded, where $E_{(\beta^*, \eta^*, \alpha^*)}$ refers to expectation with respect to $L_i(\beta^*, \eta^*, \alpha^*)$.

Proof of Theorem:

Since $n^{-1/2} \sum_{i=1}^n [\mathbf{U}_i(\hat{\beta}, \hat{\eta}, \hat{\alpha})] = \mathbf{0}$, the Taylor expansion in the neighborhood of $(\beta_0, \mathbf{0}, \alpha_0)$ gives

$$\begin{aligned} \mathbf{0} &= n^{-1/2} \sum_{i=1}^n \mathbf{U}_i(\beta_0, \mathbf{0}, \alpha_0) + \Gamma(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\beta} - \beta_0) + I_{12}(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\eta} - \mathbf{0}) \\ &\quad + I_{13}(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\alpha} - \alpha_0) + o_p(1). \end{aligned} \quad (16)$$

Note that $n^{-1/2} \sum_{i=1}^n [\mathbf{H}_i(\hat{\eta}, \hat{\alpha})] = \mathbf{0}$, the Taylor expansion in the neighborhood of $(\mathbf{0}, \alpha_0)$ gives

$$\mathbf{0} = n^{-1/2} \sum_{i=1}^n \mathbf{H}_i(\mathbf{0}, \alpha_0) + I_2(\mathbf{0}, \alpha_0) n^{1/2} (\hat{\eta} - \mathbf{0}) + I_{23}(\mathbf{0}, \alpha_0) n^{1/2} (\hat{\alpha} - \alpha_0) + o_p(1),$$

so we have

$$n^{1/2} (\hat{\eta} - \mathbf{0}) = -I_2^{-1}(\mathbf{0}, \alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{H}_i(\mathbf{0}, \alpha_0) + I_2^{-1}(\mathbf{0}, \alpha_0) I_{23}(\mathbf{0}, \alpha_0) n^{1/2} (\hat{\alpha} - \alpha_0) + o_p(1), \quad (17)$$

Replacing (17) into (16), we obtain

$$\begin{aligned} \mathbf{0} &= n^{-1/2} \sum_{i=1}^n \mathbf{U}_i(\beta_0, \mathbf{0}, \alpha_0) + \Gamma(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\beta} - \beta_0) \\ &\quad - I_{12}(\beta_0, \mathbf{0}, \alpha_0) I_2^{-1}(\mathbf{0}, \alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{H}_i(\mathbf{0}, \alpha_0) + I_{12}(\beta_0, \mathbf{0}, \alpha_0) I_2^{-1}(\mathbf{0}, \alpha_0) I_{23}(\mathbf{0}, \alpha_0) \\ &\quad n^{1/2} (\hat{\alpha} - \alpha_0) + I_{13}(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\alpha} - \alpha_0) + o_p(1). \end{aligned} \quad (18)$$

By the standard Taylor series expansion arguments in the neighborhood of α_0 we have that

$$n^{1/2} (\hat{\alpha} - \alpha_0) = -I_3^{-1}(\alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{S}_i(\alpha_0) + o_p(1). \quad (19)$$

Plugging (19) into (18), and after some algebra, we obtain

$$\mathbf{0} = n^{-1/2} \sum_{i=1}^n \mathbf{Q}_i(\beta_0, \mathbf{0}, \alpha_0) + \Gamma(\beta_0, \mathbf{0}, \alpha_0) n^{1/2} (\hat{\beta} - \beta_0) + o_p(1).$$

If $\Gamma(\beta_0, \mathbf{0}, \alpha_0)$ is nonsingular, we have

$$n^{1/2} (\hat{\beta} - \beta_0) = -\Gamma^{-1}(\beta_0, \mathbf{0}, \alpha_0) n^{-1/2} \sum_{i=1}^n \mathbf{Q}_i(\beta_0, \mathbf{0}, \alpha_0) + o_p(1).$$

Then the asymptotic distribution of $n^{1/2} (\hat{\beta} - \beta_0)$ follows by the Slutsky's theorem and the central limit theorem.

References

- Beekly, D. L., Ramos, E. M., Lee, W. W., Deitrich, W. D., Jacka, M. E., Wu, J., Hubbard, J. L., Koepsell, T. D., Morris, J. C., Kukull, W. A., and The NIA Alzheimer's Disease Centers. (2007). The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set. *Alzheimer Disease and Associated Disorders*, **21**, 249–258.
- Chaudhui, S., Handcock, M. S. and Rendall, M. S. (2008). Generalized linear models incorporating population level information: an empirical-likelihood-based approach. *Journal of the Royal Statistical Society Series B*, **70**, 311–328.
- Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, **80**, 107–116.
- Chen, J. and Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistic sinica*, **9**, 385–406.
- Hall, D. B. and Severini T. A. (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association*, **93**, 1365–1375.
- Lee, H., Laird, N. M. and Johnston, G. (1999). Combining GEE and REML for estimation of generalized linear models with incomplete multivariate data. Unpublished Manuscript.

- Liang, H., Wang, S. and Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika*, **94**, 185–198.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.) New York: John Wiley.
- Owen, A. (1988). Empirical likelihood ratio for confidence intervals for a single function. *Biometrika*, **75**, 237–249.
- Owen, A. (1990). Empirical likelihood ratio for confidence regions. *The Annals of Statistics*, **18**, 90–120.
- Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation*, **23**, 939–951.
- Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.
- Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, **22**, 300–325.
- Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society Series B*, **69**, 101–122.
- Qin, J., Zhang, B. and Leung, D. H. Y. (2010). Empirical likelihood in missing data problem. *Journal of the American Statistical Association*, **104**, 1492–1503.
- Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106–121.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.
- Small, C. G. and Wang, J. (2003). *Numerical Methods for Nonlinear Estimating Equations*. New York: Oxford University Press.
- Stute, W., Xue, L. and Zhu, L. (2007). Empirical likelihood inference in nonlinear errors-in-covariates models with validation data. *Journal of the American Statistical Association*, **102**, 332–346.
- Wang, Y. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: implications for generalised estimating equations performance. *Biometrika*, **90**, 29–41.
- Wang, S., Qian, L. and Carroll, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika*, **97**, 79–93.

- Wang, Q. and Rao, J. N. K. (2002a). Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics*, **30**, 896–924.
- Wang, Q. and Rao, J. N. K. (2002b). Empirical likelihood-based inference in linear errors-in-covariates models with validation data. *Biometrika*, **89**, 345–358.
- Xue, L. and Zhu, L. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika*, **94**, 921–937.
- You, J., Chen, G. and Zhou, Y. (2006). Block empirical likelihood for longitudinal partially linear regression models. *Canadian Journal of Statistics*, **34**, 79–96.

Table 1. Empirical biases and standard deviations for binary longitudinal data with mean constraints of responses: the missing data model is correctly specified

n	ρ	α_1	Parameter	Proposed Method			IPWGEE			GEE			Gaussian EE		
				BIAS	SD	CP	BIAS	SD	CP	BIAS	SD	CP	BIAS	SD	CP
100	0.9	-4	β_0	0.028	0.315	0.940	-0.041	0.661	0.941	-0.377	0.319	0.656	-0.152	0.424	0.881
		-4	β_1	0.052	0.622	0.950	0.093	1.263	0.957	0.158	0.676	0.887	0.222	0.699	0.929
		-4	β_2	0.054	0.281	0.940	0.040	0.553	0.949	-0.055	0.295	0.907	0.007	0.287	0.929
		-2	β_0	0.031	0.306	0.940	-0.040	0.627	0.936	-0.293	0.348	0.727	-0.128	0.307	0.937
		-2	β_1	0.057	0.628	0.954	0.040	1.238	0.948	0.135	0.666	0.902	0.162	0.670	0.937
		-2	β_2	0.063	0.286	0.940	-0.002	0.526	0.946	-0.048	0.298	0.892	-0.061	0.287	0.899
	0.6	-4	β_0	0.041	0.327	0.950	-0.043	0.657	0.938	-0.256	0.330	0.796	-0.105	0.353	0.959
		-4	β_1	0.079	0.650	0.944	0.080	1.260	0.946	0.139	0.704	0.906	0.129	0.662	0.959
		-4	β_2	0.082	0.302	0.938	0.081	0.576	0.948	0.005	0.302	0.940	0.028	0.411	0.944
		-2	β_0	0.033	0.288	0.944	-0.034	0.579	0.956	-0.211	0.321	0.832	0.009	0.292	0.961
		-2	β_1	0.044	0.553	0.940	0.026	1.206	0.946	0.106	0.661	0.902	0.062	0.576	0.963
		-2	β_2	0.059	0.278	0.944	0.051	0.555	0.954	-0.007	0.299	0.922	-0.024	0.283	0.918
200	0.9	-4	β_0	-0.002	0.200	0.958	-0.012	0.602	0.939	-0.404	0.226	0.449	-0.105	0.206	0.925
		-4	β_1	0.067	0.417	0.944	0.037	1.133	0.937	0.156	0.477	0.884	0.106	0.428	0.938
		-4	β_2	0.011	0.231	0.944	-0.058	0.498	0.944	-0.097	0.188	0.896	-0.049	0.247	0.850
		-2	β_0	0.017	0.218	0.940	-0.081	0.426	0.940	-0.303	0.215	0.590	-0.112	0.261	0.883
		-2	β_1	0.007	0.349	0.964	0.085	0.807	0.952	0.081	0.413	0.908	0.126	0.355	0.968
		-2	β_2	0.022	0.189	0.940	-0.098	0.391	0.942	-0.093	0.197	0.880	-0.088	0.194	0.872
	0.6	-4	β_0	0.033	0.217	0.950	-0.033	0.638	0.942	-0.272	0.230	0.670	-0.029	0.223	0.943
		-4	β_1	0.020	0.417	0.944	0.041	1.100	0.939	0.143	0.464	0.895	0.044	0.424	0.951
		-4	β_2	0.026	0.182	0.946	0.055	0.565	0.945	-0.037	0.204	0.922	-0.035	0.191	0.947
		-2	β_0	-0.007	0.200	0.946	-0.058	0.455	0.964	-0.246	0.227	0.706	-0.039	0.201	0.961
		-2	β_1	0.065	0.379	0.950	0.085	0.880	0.940	0.108	0.444	0.882	0.075	0.393	0.965
		-2	β_2	0.017	0.176	0.956	0.001	0.383	0.946	-0.047	0.182	0.934	-0.048	0.195	0.911

Table 2. Empirical biases and standard deviations for continuous longitudinal data with mean constraints of responses: the missing data model is correctly specified

n	ρ	α_1	Parameter	Proposed Method			IPWGEE			GEE			Gaussian EE		
				BIAS	SD	CP	BIAS	SD	CP	BIAS	SD	CP	BIAS	SD	CP
100	0.9	-4	β_0	-0.001	0.137	0.949	-0.065	0.243	0.940	-0.307	0.162	0.312	-0.104	0.161	0.868
		-4	β_1	-0.009	0.216	0.959	0.087	0.395	0.946	0.231	0.256	0.728	0.102	0.234	0.896
		-4	β_2	-0.006	0.066	0.943	0.007	0.131	0.940	-0.001	0.089	0.952	-0.016	0.073	0.844
		-2	β_0	-0.002	0.140	0.942	-0.077	0.296	0.960	-0.276	0.163	0.372	-0.080	0.163	0.876
		-2	β_1	-0.005	0.220	0.946	-0.071	0.471	0.946	0.176	0.251	0.804	0.077	0.229	0.928
		-2	β_2	0.002	0.064	0.948	0.001	0.168	0.936	-0.005	0.092	0.932	-0.017	0.071	0.858
	0.6	-4	β_0	-0.008	0.140	0.942	-0.076	0.261	0.936	-0.195	0.157	0.562	-0.054	0.154	0.884
		-4	β_1	0.010	0.222	0.940	0.071	0.365	0.948	0.147	0.235	0.844	0.049	0.228	0.920
		-4	β_2	-0.006	0.084	0.942	-0.007	0.134	0.948	-0.003	0.091	0.942	-0.009	0.090	0.888
		-2	β_0	-0.012	0.140	0.946	-0.033	0.256	0.942	-0.186	0.145	0.610	-0.047	0.155	0.892
		-2	β_1	0.010	0.210	0.942	-0.027	0.479	0.944	0.127	0.230	0.870	0.033	0.225	0.924
		-2	β_2	-0.006	0.082	0.944	0.005	0.181	0.940	0.001	0.091	0.934	-0.004	0.089	0.894
200	0.9	-4	β_0	-0.006	0.101	0.946	-0.064	0.201	0.942	-0.311	0.122	0.112	-0.091	0.117	0.840
		-4	β_1	-0.004	0.147	0.946	0.056	0.344	0.956	0.218	0.185	0.632	0.096	0.160	0.884
		-4	β_2	0.002	0.041	0.942	-0.003	0.125	0.932	0.002	0.067	0.934	-0.017	0.048	0.900
		-2	β_0	-0.001	0.099	0.950	-0.044	0.245	0.942	-0.279	0.118	0.152	-0.090	0.111	0.842
		-2	β_1	-0.019	0.150	0.942	-0.014	0.383	0.942	0.184	0.176	0.698	0.086	0.156	0.896
		-2	β_2	0.008	0.040	0.940	-0.004	0.150	0.952	-0.001	0.068	0.926	-0.008	0.046	0.912
	0.6	-4	β_0	-0.003	0.096	0.948	-0.091	0.235	0.948	-0.206	0.113	0.368	-0.053	0.110	0.884
		-4	β_1	-0.006	0.155	0.944	0.070	0.325	0.948	0.154	0.170	0.768	0.059	0.161	0.910
		-4	β_2	0.003	0.056	0.944	-0.004	0.142	0.940	-0.001	0.064	0.952	-0.012	0.059	0.930
		-2	β_0	-0.007	0.096	0.948	-0.016	0.233	0.948	-0.191	0.104	0.416	-0.058	0.102	0.908
		-2	β_1	0.001	0.144	0.948	-0.063	0.414	0.948	0.126	0.155	0.840	0.064	0.152	0.918
		-2	β_2	0.002	0.053	0.946	-0.012	0.152	0.944	0.000	0.061	0.964	-0.012	0.059	0.944

Table 3. The dementia study: response model.

Para.	Proposed Method			IPWGEE			GEE			Gaussian EE		
	EST	SE	P	EST	SE	P	EST	SE	P	EST	SE	P
(Intercept)	0.684	0.178	<0.001	0.681	0.187	<0.001	0.378	0.178	0.034	0.460	0.162	0.005
Sex(F)	-0.126	0.111	0.255	-0.130	0.114	0.256	-0.026	0.106	0.804	-0.015	0.107	0.892
Depression(Yes)	0.595	0.089	<0.001	0.611	0.109	<0.001	0.372	0.083	<0.001	0.254	0.080	0.002
Fhdem(Yes)	0.173	0.098	0.078	0.175	0.108	0.105	0.092	0.082	0.262	0.075	0.070	0.280
Diabete(Yes)	0.219	0.157	0.163	0.215	0.172	0.210	0.251	0.143	0.079	0.072	0.149	0.631
Hypertension(Yes)	-0.095	0.105	0.366	-0.097	0.115	0.398	0.025	0.101	0.802	-0.240	0.099	0.016
Education(Years)	0.001	0.005	0.841	0.001	0.006	0.884	0.000	0.006	0.991	-0.008	0.006	0.151
Age-65	0.025	0.007	<0.001	0.026	0.008	0.002	0.019	0.008	0.013	0.027	0.008	0.001

Table 4. The dementia study: missing model

Parameter	EST	SE	P
(Intercept)	0.614	0.138	<0.001
Sex(F)	-0.388	0.088	<0.001
Depression(Yes)	1.114	0.110	<0.001
Fhdem(Yes)	-0.162	0.092	0.078
Diabete(Yes)	0.159	0.137	0.248
Hypertension(Yes)	-0.108	0.089	0.229
Education(Years)	-0.012	0.004	0.005
Age-65	0.036	0.006	<0.001
Prev. dementia(Yes)	4.181	0.504	<0.001