REVSTAT – Statistical Journal
Volume 20, Number 4, July 2022, 405–426
https://doi.org/10.57805/revstat.v20i4.379

Time Series Analysis for Longitudinal Survey Data under Informative Sampling and Nonignorable Missingness

Authors: Zhan Liu ✉
– Hubei Key Laboratory of Applied Mathematics, School of Mathematics and Statistics, Hubei University,
Wuhan, 430062, China
20170066@hubu.edu.cn

Chun Yip Yau
– Department of Statistics, Chinese University of Hong Kong,
Hong Kong, China
cyyau@sta.cuhk.edu.hk

Received: January 2020 Revised: June 2020 Accepted: August 2020

Abstract:

- The analysis of longitudinal survey data is often complicated when informative sampling or nonignorable missing data exists. Existing methods that can handle both informative sampling and nonignorable missing data are only limited to the situation of no time dependence in the data. In this paper, we develop a sample likelihood based approach for estimation of time series model in longitudinal survey data under informative sampling and nonignorable missingness. In particular, some informative sampling models and a response model are proposed to describe the mechanisms of informative sampling and nonignorable missingness. A sample likelihood is derived based on the conditional distribution of the observed measurements. Also, an effective computation algorithm is developed to compute the sample likelihood. Simulation studies are carried out to investigate the performance of the proposed estimator. A real data example based on data from AIDS Clinical Trial Group 193A Study is presented to illustrate the proposed method.

Keywords:

- *autoregressive model; exponential model; probit model; logistic model; sample likelihood.*

AMS Subject Classification:

- 62D05, 62F10.

✉ Corresponding author.

1. INTRODUCTION

Longitudinal surveys are designed to measure a sample of respondents repeatedly over time, and have been extensively applied in various fields such as clinical studies, biological research and social sciences. Longitudinal surveys are prevalence in studying human's behaviors, health, and mortality because they provide efficient means to estimate the change in the population, evaluate interventions, test causal hypotheses, and reduce the cost of data collection [35]. Since longitudinal surveys are conducted at different points of time, the serial observations obtained from a given unit usually show time dependence. Therefore, a time series model can be employed to analyze longitudinal survey data [12].

Informative sampling, which refers to sampling design in which the sampling probabilities are correlated with the response variable (conditional on covariates), is often encountered in longitudinal surveys, see, e.g., Fuller [15]. However, studies ignoring informative sampling can lead to seriously biased results (Pfeffermann [27], [26]; Eideh and Nathan [12]; Eideh [9]; Sverchkov and Pfeffermann [33]). To handle informative sampling, Pfeffermann *et al.* [25] derived the sample distribution from the population distribution and the sampling probabilities under informative sampling, which can permit the use of classical inference methods. Chambers and Skinner [7], and Pfeffermann and Sverchkov [24] discussed the sample likelihood approach, the pseudo-likelihood approach and the estimating equations approach for fitting generalized linear models under informative sampling, based on the sample distribution of Pfeffermann *et al.* [25]. In fact, the sample likelihood approach has been explored in many different directions including small area estimation (Pfeffermann and Sverchkov [22]; Eideh and Nathan [11]; Verret *et al.* [37]), general linear modelling (Chambers and Skinner [7]; Pfeffermann and Sverchkov [22]; Eideh [9]), and multi-level model analysis (Pfeffermann *et al.* [23]; Cai [6]). Recently, Bonnery *et al.* [4] established the asymptotic properties of the sample likelihood approach under informative sampling. Other proposed methods include the inverse probability weighting method (Boudreau and Lawless [5]; Kim and Skinner [17]) and calibration adjustments (Moser *et al.* [20]). However, most of the above studies explored the informative sampling problem in the non-longitudinal survey context. Informative sampling in longitudinal surveys was considered in Eideh and Nathan [12], [13], and Eideh [9]. Eideh and Nathan [12], [13] discussed the sample likelihood and pseudo-likelihood methods in fitting time series models for longitudinal survey data under informative sampling. Eideh [9] explored further the sample likelihood, pseudo-likelihood likelihood and estimating equations methods in fitting general linear model for longitudinal survey data under informative sampling.

In addition to informative sampling, another major issue in longitudinal surveys is the missing data problem. Following Little and Rubin [18], the mechanisms of missing data can be classified into three types: missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). In particular, missing completely at random and missing at random are called ignorable missingness, whereas not missing at random is called nonignorable missingness. Under nonignorable missingness, the missing probability depends on the response variable, and thus will lead to unreliable estimation results (Eideh [9]; Schlomer *et al.* [30]; Taisir and Islam [34]). A solution to this problem is the modeling of nonignorable missing data, which has been applied to general linear models (Bahari *et al.* [2]), generalized linear mixed models (Stubbendick and Ibrahim [32]; Sabry *et al.* [29];

Almohisen *et al.* [1]), quantile regression models (Yuan and Yin [38]), latent random effects models (Tseng *et al.* [36]; Bhuyan [3]), and Markov chain models (Cole *et al.* [8]; Taisir and Islam [34]).

When informative sampling and nonignorable missingness occur in longitudinal surveys simultaneously, the joint treatment of the two problems becomes a key issue. Pfeffermann [21] proposed a unified approach to handle the two problems by combining the observed data model with the missing data model and the target population model based on the Bayes theorem. Sverchkov and Pfeffermann [33] extended the approach in Pfeffermann and Sverchkov [22] in small area estimation under informative sampling to the case that both informative sampling and nonignorable missingness exist. However, these approaches only considered data measured at a certain time point and are not applicable to longitudinal data. Eideh and Nathan [10], and Farahania *et al.* [14] considered methods to handle informative sampling and nonignorable missingness simultaneously in longitudinal data analysis. However, their discussions focus mainly on general regression models.

In this paper, we study time series modeling for longitudinal survey data under informative sampling and nonignorable missingness. Treating informative sampling and nonignorable missingness simultaneously becomes especially challenging in time series models due to the serial correlation of the response variable at various time points. We consider models to explore the effect of each of informative sampling and nonignorable missingness. For informative sampling, a variety of models, including exponential, probit, and logistic models are considered to capture the dependence between the selection probability and the response variable. For nonignorable missingness, we consider a logistic model to relate the response probability to the response variables. Based on these models, we derive a sample likelihood for parameter estimation under informative sampling and nonignorable missingness. To compute the sample likelihood function efficiently, an approximation to the integrals in the sample likelihood based on series expansions is proposed. Simulation studies and real data application are provided to illustrate the effectiveness of the proposed method.

The remainder of the paper is organized as follows. Section 2 describes time series models and parameter estimation methods for longitudinal survey data. Section 3 discusses informative sampling and nonignorable missingness in longitudinal surveys. In Section 4, the sample likelihood is derived for conducting time series analysis in longitudinal survey data under informative sampling and nonignorable missingness. Simulations studies and real data analysis are performed in Sections 5 and 6, respectively.

2. TIME SERIES MODEL FOR LONGITUDINAL SURVEY DATA

Let $U=\{1,...,N\}$ be the index set of a finite population U of size N. Let $y_{i,t}$ ($i=1,...,N$, $t=1,...,T$) be the value of a response variable y of unit i at time t, and x_i be the values of the covariates of unit i, which are always observed and remain constant over time. A random sample S of size n is then selected from the finite population at time 1 ($t=1$) and measured independently from time 1 to time T. Suppose that $y_{i,t}$ is correlated with the past values $y_{i,t'}$, $1 \leq t' < t \leq T$, for each T. A time series model can then be fitted to analyze this longitudinal survey data. Typically, time series models with short-range dependence are often applied in decision-making and policymaking [12]. For simplicity, we consider the first-order autoregressive (AR(1)) model

$$y_{i,t} - \mu = \phi(y_{i,t-1} - \mu) + \varepsilon_{i,t}\,, \qquad i = 1, ..., N, t = 1, ..., T\,, \tag{2.1}$$

where μ is the mean level of the data, the errors $\varepsilon_{i,t} \overset{\text{iid}}{\sim} N(0, \sigma^2)$, and $|\phi| < 1$. The model parameter $\theta = (\mu, \phi, \sigma)$ is of our interest. Note that unit i in the AR(1) model will fall into the set $\{1, ..., n\}$ when the sample data is used to estimate the model parameters.

Usually, the maximum likelihood estimation approach is employed to obtain the model parameter estimators. Let $\mathbf{y}_i = (y_{i,1}, ..., y_{i,T})'$ be the vector of T measurements on unit i $(i = 1, ..., N)$. Then, the density function of $\mathbf{y}_i$ can be expressed as $f(\mathbf{y}_i; \theta) = f(y_{i,1}; \theta) \cdot \cdot \prod_{t=2}^{T} f(y_{i,t}|y_{i,t-1}; \theta)$. For the AR(1) model, we have $y_{i,1} \sim N(\mu, \sigma^2/(1 - \phi^2))$ and $f(y_{i,t}|y_{i,t-1}; \theta) = (2\pi\sigma^2)^{-1/2} \exp\{-[y_{i,t} - \phi(y_{i,t-1} - \mu) - \mu]^2/(2\sigma^2)\}$. Thus, the log-likelihood function of θ can be written as

$$\log L(\theta) \;=\; \sum_{i=1}^{n} \log f(y_{i,1}; \theta) + \sum_{i=1}^{n} \sum_{t=2}^{T} \log f(y_{i,t}|y_{i,t-1}; \theta)\,. \tag{2.2}$$

It follows that the maximum likelihood estimator of θ can be obtained by maximizing the log-likelihood function in (2.2).

3. INFORMATIVE SAMPLING AND NONIGNORABLE MISSINGNESS IN LONGITUDINAL SURVEYS

3.1. Informative sampling

Analytic inference from longitudinal survey data usually fails to account for the complex sampling design, such as informative sampling. A sampling design is called informative when the sample selection probabilities are related to the response variable y, even after conditioning on the covariates. In practice, selection probabilities may be correlated with the response variable, the covariates and possibly, design variables used for sampling. For simplicity, we consider the case that selection probabilities depend only on the response variable.

Let I_i be the sample indicator variable, taking values of 1 if unit $i \in U$ is selected to the sample S and 0 if otherwise. The selection probabilities can then be denoted by $\pi_i = P(I_i = 1|y_i)$. Let $f_s(y_i)$ and $f_p(y_i)$ denote the sample density and the population density of y_i, respectively. In fact, the density functions $f(y_{i,1}; \theta)$ and $f(y_{i,t}|y_{i,t-1}; \theta)$ in Section 2 are the population densities, which can also be denoted by $f_p(y_{i,1}; \theta)$ and $f_p(y_{i,t}|y_{i,t-1}; \theta)$, respectively. Following Pfeffermann *et al.* [25] as well as Sikov and Stern [31], the sample density $f_s(y_i)$ is given by

$$\begin{aligned} f_s(y_i) \;&=\; f(y_i|I_i = 1) \;=\; \frac{f(y_i, I_i = 1)}{P(I_i = 1)} \\ &=\; \frac{P(I_i = 1|y_i) f_p(y_i)}{P(I_i = 1)} \;=\; \frac{E_p(\pi_i|y_i) f_p(y_i)}{E_p(\pi_i)}\,, \end{aligned} \tag{3.1}$$

where $\pi_i = P(I_i = 1|y_i)$, $E_p(\pi_i|y_i) = \int P(I_i = 1|y_i, \pi_i) f_p(\pi_i|y_i) d\pi_i = P(I_i = 1|y_i)$, and $E_p(\pi_i) = \int P(I_i = 1|y_i) f_p(y_i) dy_i = P(I_i = 1)$. Under informative sampling, the selection probability $\pi_i = P(I_i = 1|y_i)$ depends on y_i. Hence, $E_p(\pi_i|y_i) \neq E_p(\pi_i)$ and $P(I_i = 1|y_i) \neq P(I_i = 1)$, yielding $f_s(y_i) \neq f_p(y_i)$ in general. That is, the sample distribution is different from the population distribution. However, the sample distribution is viewed as the same as the population distribution in many analysis under informative sampling, which have resulted in false inferences (Pfeffermann [27], [26]).

In order to access the sample density, $E_p(\pi_i|y_i) = P(I_i = 1|y_i)$ can be modeled to explore the relationship between the selection probabilities π_i and the response variable values y_i. Pfeffermann *et al.* [25] and Eideh and Nathan [12] considered

$$\text{Exponential model:} \quad E_p(\pi_i|y_i) = \exp(a_0 + a_1 y_i)\,, \tag{3.2}$$

where a_0 and a_1 are unknown model parameters. Besides, the probit model and logistic model, which are less common in longitudinal surveys under informative sampling, can also be explored to explain the informative sampling mechanism:

$$\text{Probit model:} \quad E_p(\pi_i|y_i) = \Phi(b_0 + b_1 y_i)\,, \tag{3.3}$$

$$\text{Logistic model:} \quad E_p(\pi_i|y_i) = \frac{\exp(c_0 + c_1 y_i)}{1 + \exp(c_0 + c_1 y_i)}\,, \tag{3.4}$$

where b_0, b_1, c_0, c_1 are unknown model parameters.

3.2. Nonignorable missingness

Missing data is another problem which often arises in longitudinal surveys. Here, we assume that there exists nonignorable missingness in longitudinal surveys. In particular, the values $y_{i,1}$ at time 1 are complete and some of $y_{i,2}, ..., y_{i,T}$ suffer from missingness for $i = 1, ..., n$. Denote the response indicator variable by

$$\delta_{i,t} = \begin{cases} 1 & \text{if } y_{i,t} \text{ is observed}\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{3.5}$$

The nonignorable missingness implies that missingness depends on the response variable. In other words, the response probability is related to the response variable. Under the AR(1) model, we model the response mechanism using a logistic model

$$\begin{aligned} P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}) &=: \pi(x_i, y_{i,t-1}, y_{i,t}; \eta) \\ &= \frac{\exp(\eta_1 x_i + \eta_2 y_{i,t-1} + \eta_3 y_{i,t})}{1 + \exp(\eta_1 x_i + \eta_2 y_{i,t-1} + \eta_3 y_{i,t})}\,, \end{aligned} \tag{3.6}$$

where $\eta = (\eta_1, \eta_2, \eta_3)$ is the unknown parameter. Equation (3.6) asserts that the response probability $P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t})$ at time t depends not only on the value $y_{i,t}$ at time t and the covariate x_i, but also on its past value $y_{i,t-1}$. Clearly, the response mechanism is nonignorable missingness. Note that (3.6) extends the nonignorable response mechanism in Qin *et al.* [28] by incorporating the effect of past observations into the response probability.

For notational simplicity, only one covariate x is considered in the response model. The extension to multiple covariates $x_1, ..., x_p$ in the response model is straightforward.

If we ignore the informative sampling and nonignorable missingness, using the complete case (CC) analysis (Farahania *et al.* [14]), the log-likelihood function of θ in the AR(1) model based on the observed data is rewritten as

$$
\begin{aligned}
(3.7) \qquad \log L(\theta) &= \sum_{i=1}^{n} \log f(y_{i,1};\theta) + \sum_{t=2}^{T}\sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t} \log f(y_{i,t}|y_{i,t-1};\theta) \\
&= \sum_{i=1}^{n}\left\{-\frac{1}{2}\log\left(\frac{2\pi\sigma^2}{1-\phi^2}\right) - \frac{(1-\phi^2)(y_{i,1}-\mu)^2}{2\sigma^2}\right\} \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t}\left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[y_{i,t} - \phi(y_{i,t-1}-\mu) - \mu]^2\right\}.
\end{aligned}
$$

Then, we can get the maximum likelihood estimator $\hat{\theta}$ of θ via maximizing the log-likelihood function in (3.7). However, the obtained estimator $\hat{\theta}$ is obviously biased because it ignores the informative sampling and nonignorable missingness (Pfeffermann *et al.* [25]; Little and Rubin [18]; Farahania *et al.* [14]). In fact, the observed sample distribution is different from the population distribution under both informative sampling and nonignorable missingness, which cannot guarantee that the log-likelihood function in (3.7) gives the correct estimates.

4. SAMPLE LIKELIHOOD AND ESTIMATION UNDER INFORMATIVE SAMPLING AND NONIGNORABLE MISSINGNESS

4.1. Sample likelihood under informative sampling

The sample distribution differs from the population distribution under informative sampling. Therefore, the sample likelihood will be different from the general likelihood under noninformative sampling. Because the sample is only selected from the finite population at time 1 in longitudinal surveys, the sample distribution at time 1 can be obtained by replacing y_i in (3.1) with $y_{i,1}$ in longitudinal surveys. In what follows, the sample density function $f_s(\mathbf{y}_i)$ of $\mathbf{y}_i$ in longitudinal surveys under informative sampling can be expressed as

$$
\begin{aligned}
(4.1) \qquad f_s(\mathbf{y}_i) &= f_s(y_{i,1};\theta)\prod_{t=2}^{T} f_p(y_{i,t}|y_{i,t-1};\theta) \\
&= \frac{E_p(\pi_i|y_{i,1})f_p(y_{i,1};\theta)}{E_p(\pi_i)}\prod_{t=2}^{T} f_p(y_{i,t}|y_{i,t-1};\theta).
\end{aligned}
$$

Then, the log-likelihood function becomes

$$
\begin{aligned}
(4.2) \qquad \log L &= \sum_{i=1}^{n}\log E_p(\pi_i|y_{i,1}) - \sum_{i=1}^{n}\log E_p(\pi_i) \\
&\quad + \sum_{i=1}^{n}\log f(y_{i,1};\theta) + \sum_{i=1}^{n}\sum_{t=2}^{T}\log f(y_{i,t}|y_{i,t-1};\theta).
\end{aligned}
$$

4.2. Sample likelihood under informative sampling and nonignorable missingness

When nonignorable missingness also exists in longitudinal surveys under informative sampling, $\sum_{i=1}^{n}\sum_{t=2}^{T}\log f(y_{i,t}|y_{i,t-1};\theta)$ in (4.2) needs to be modified since $f(y_{i,t}|y_{i,t-1};\theta)$ is not available when $y_{i,t}$ or $y_{i,t-1}$ is missing. Taking the response mechanism (3.6) into account, we propose to replace $f(y_{i,t}|y_{i,t-1};\theta)$ by the conditional densities based on the observed response, namely $f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)$ or $f(y_{i,t}|x_i,y_{i,t-1},\delta_{i,t-1}=1,\delta_{i,t}=1)$, depending on whether $y_{i,t-1}$ is missing or not. It follows that the log-likelihood function under informative sampling and nonignorable missingness can be rewritten as

$$
\begin{aligned}
\log L =& \sum_{i=1}^{n}\log E_p(\pi_i|y_{i,1}) - \sum_{i=1}^{n}\log E_p(\pi_i) + \sum_{i=1}^{n}\log f(y_{i,1};\theta)\\
&+\sum_{i=1}^{n}\sum_{t=2}^{T}\delta_{i,t-1}\delta_{i,t}\log f(y_{i,t}|x_i,y_{i,t-1},\delta_{i,t-1}=1,\delta_{i,t}=1)\\
&+\sum_{i=1}^{n}\sum_{t=2}^{T}(1-\delta_{i,t-1})\delta_{i,t}\log f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)\,.
\end{aligned}
\tag{4.3}
$$

Next, we derive the expressions for $f(y_{i,t}|x_i,y_{i,t-1},\delta_{i,t-1}=1,\delta_{i,t}=1)$ and $f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)$ in the following lemma. The proof is given in the Appendix.

Lemma 4.1. *The conditional density $f(y_{i,t}|x_i,y_{i,t-1},\delta_{i,t-1}=1,\delta_{i,t}=1)$ satisfies*

$$
f(y_{i,t}|x_i,y_{i,t-1},\delta_{i,t-1}=1,\delta_{i,t}=1) = \frac{\pi(x_i,y_{i,t-1},y_{i,t})f(y_{i,t}|y_{i,t-1})}{\int\pi(x_i,y_{i,t-1},y_t)f(y_t|y_{i,t-1})dy_t}\,,
\tag{4.4}
$$

and $f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)$ satisfies

$$
\begin{aligned}
&f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)\\
&=\frac{\iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i,y_{t-1},y_{i,t})[1-\pi(x_i,y_{t-2},y_{t-1})]dy_{t-2}dy_{t-1}}{f(x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)}.
\end{aligned}
\tag{4.5}
$$

Substituting (4.4) and (4.5) into (4.3) yields the following log-likelihood function under informative sampling and nonignorable missingness

$$
\begin{aligned}
\log L =& \sum_{i=1}^{n}\log E_p(\pi_i|y_{i,1}) - \sum_{i=1}^{n}\log E_p(\pi_i) + \sum_{i=1}^{n}\log f(y_{i,1};\theta)\\
&+\sum_{t=2}^{T}\sum_{i=1}^{n}\delta_{i,t-1}\delta_{i,t}\Big\{\log f(y_{i,t}|y_{i,t-1};\theta)+\log\pi(x_i,y_{i,t-1},y_{i,t};\eta)\\
&\quad-\log\int\pi(x_i,y_{i,t-1},y_t;\eta)f(y_t|y_{i,t-1};\theta)dy_t\Big\}\\
&+\sum_{t=2}^{T}\sum_{i=1}^{n}(1-\delta_{i,t-1})\delta_{i,t}\Big\{\log\iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i,y_{t-1},y_{i,t})\\
&\quad\cdot[1-\pi(x_i,y_{t-2},y_{t-1})]dy_{t-2}dy_{t-1}-\log f(x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)\Big\}\,.
\end{aligned}
\tag{4.6}
$$

Using (3.2), (3.3) and (3.4), the log-likelihood functions under nonignorabe missingness and the three informative sampling models can be expressed as

Exponential model:

$$
\begin{aligned}
(4.7)\quad & \log L(\theta, \eta, a_1) \\
&= a_1 \sum_{i=1}^{n} y_{i,1} - n[a_1\mu + \sigma^2 a_1^2/(2(1-\phi^2))] + \sum_{i=1}^{n} \log f(y_{i,1};\theta) \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t}\Big\{\log f(y_{i,t}|y_{i,t-1};\theta) + \log \pi(x_i, y_{i,t-1}, y_{i,t};\eta) \\
&\quad - \log \int \pi(x_i, y_{i,t-1}, y_t;\eta) f(y_t|y_{i,t-1};\theta)dy_t\Big\} \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n}(1-\delta_{i,t-1})\delta_{i,t}\Big\{\log \iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i, y_{t-1}, y_{i,t}) \\
&\quad \cdot [1-\pi(x_i, y_{t-2}, y_{t-1})]dy_{t-2}dy_{t-1} - \log f(x_i, \delta_{i,t-1}=0, \delta_{i,t}=1)\Big\},
\end{aligned}
$$

Probit model:

$$
\begin{aligned}
(4.8)\quad & \log L(\theta, \eta, b_0, b_1) \\
&= \sum_{i=1}^{n} \log \Phi(b_0 + b_1 y_{i,1}) - \sum_{i=1}^{n} \log \int \Phi(b_0 + b_1 y_{i,1}) f(y_{i,1}) dy_{i,1} + \sum_{i=1}^{n} \log f(y_{i,1};\theta) \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t}\Big\{\log f(y_{i,t}|y_{i,t-1};\theta) + \log \pi(x_i, y_{i,t-1}, y_{i,t};\eta) \\
&\quad - \log \int \pi(x_i, y_{i,t-1}, y_t;\eta) f(y_t|y_{i,t-1};\theta)dy_t\Big\} \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n}(1-\delta_{i,t-1})\delta_{i,t}\Big\{\log \iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i, y_{t-1}, y_{i,t}) \\
&\quad \cdot [1-\pi(x_i, y_{t-2}, y_{t-1})]dy_{t-2}dy_{t-1} - \log f(x_i, \delta_{i,t-1}=0, \delta_{i,t}=1)\Big\},
\end{aligned}
$$

Logistic model:

$$
\begin{aligned}
(4.9)\quad & \log L(\theta, \eta, c_0, c_1) \\
&= -\sum_{i=1}^{n} \log[1+\exp(-c_0 - c_1 y_{i,1})] - \sum_{i=1}^{n} \log \int [1+\exp(-c_0 - c_1 y_{i,1})]^{-1} f(y_{i,1}) dy_{i,1} \\
&\quad + \sum_{i=1}^{n} \log f(y_{i,1};\theta) + \sum_{t=2}^{T}\sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t}\Big\{\log f(y_{i,t}|y_{i,t-1};\theta) \\
&\quad + \log \pi(x_i, y_{i,t-1}, y_{i,t};\eta) - \log \int \pi(x_i, y_{i,t-1}, y_t;\eta) f(y_t|y_{i,t-1};\theta)dy_t\Big\} \\
&\quad + \sum_{t=2}^{T}\sum_{i=1}^{n}(1-\delta_{i,t-1})\delta_{i,t}\Big\{\log \iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i, y_{t-1}, y_{i,t}) \\
&\quad \cdot [1-\pi(x_i, y_{t-2}, y_{t-1})]dy_{t-2}dy_{t-1} - \log f(x_i, \delta_{i,t-1}=0, \delta_{i,t}=1)\Big\}.
\end{aligned}
$$

Therefore, the maximum likelihood estimators of θ, η, a_1, b_0, b_1, c_0, and c_1 can be obtained by maximizing the log-likelihood functions in (4.7), (4.8) or (4.9).

4.3. Computations of the likelihood function

Note that computing the log-likelihood functions in (4.7), (4.8) and (4.9) involves the density $f(x_i, \delta_{i,t-1} = 0, \delta_{i,t} = 1)$, as well as the integrals $\int \pi(x_i, y_{i,t-1}, y_t; \eta) f(y_t|y_{i,t-1}; \theta) dy_t$, $\int [1 + \exp(-c_0 - c_1 y_{i,1})]^{-1} f(y_{i,1}) dy_{i,1}$, $\iint f(y_{t-2}) f(y_{t-1}|y_{t-2}) f(y_{i,t}|y_{t-1}) \pi(x_i, y_{t-1}, y_{i,t}) \cdot$ $\cdot [1 - \pi(x_i, y_{t-2}, y_{t-1})] dy_{t-2} dy_{t-1}$ and $\int \Phi(b_0 + b_1 y_{i,1}) f(y_{i,1}) dy_{i,1}$. In this section we discuss effective computations for these quantities.

First, $f(x_i, \delta_{i,t-1} = 0, \delta_{i,t} = 1)$ can be approximated by the empirical distribution

$$f(x_i, \delta_{i,t-1} = 0, \delta_{i,t} = 1) \approx \sum_{\substack{i,\delta_{i,t}=1;\\ \delta_{i,t-1}=0}} (1 - \delta_{i,t-1}) \delta_{i,t} \Big/ n \,.$$

Next, the following lemma provides a series expansion for the integral $\int \pi(x_i, y_{i,t-1}, y_t; \eta) \cdot$ $\cdot f(y_t|y_{i,t-1}; \theta) dy_t$. The proof is provided in the Appendix.

Lemma 4.2. *The integral $\int \pi(x_i, y_{i,t-1}, y_t) f(y_t|y_{i,t-1}) dy_t$ satisfies*

$$\int \pi(x_i, y_{i,t-1}, y_t) f(y_t|y_{i,t-1}) dy_t$$

$$= \begin{cases} \displaystyle\sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2) \Phi(\gamma - \beta k) \\ \displaystyle\qquad + \frac{1}{c} \sum_{k=0}^{\infty} \left(-\frac{1}{c}\right)^k \exp(\beta^2 (k+1)^2/2) [1 - \Phi(\gamma + \beta k + \beta)] \,, & \beta > 0 \,, \\ \displaystyle\sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2) [1 - \Phi(\gamma - \beta k)] \\ \displaystyle\qquad + \frac{1}{c} \sum_{k=0}^{\infty} \left(-\frac{1}{c}\right)^k \exp(\beta^2 (k+1)^2/2) \Phi(\gamma + \beta k + \beta) \,, & \beta < 0 \,, \\ \displaystyle\frac{1}{1+c} \,, & \beta = 0 \,, \end{cases} \tag{4.10}$$

where $c = \exp[-(\eta_1 x_i + \eta_2 y_{i,t-1} + \eta_3 \tilde{\mu})]$, $\tilde{\mu} = \mu + \phi(y_{i,t-1} - \mu)$, $\beta = -\eta_3 \sigma$, $\gamma = -\log c/\beta$ *and* Φ *is the distribution function of standard normal distribution.*

In practice, the infinite series in (4.10) has to be approximated by a finite truncated sum. Simulation studies show that the truncation of summing up to $k = 10$ gives a good approximation to the infinite series in most cases.

Based on Lemma 4.2, the following corollary gives a similar series expansion for the integral $\int [1 + \exp(-c_0 - c_1 y_{i,1})]^{-1} f(y_{i,1}) dy_{i,1}$ in (4.9). The proof is presented in the Appendix.

Corollary 4.1. *The integral* $\int [1+\exp(-c_0-c_1 y_{i,1})]^{-1} f(y_{i,1}) dy_{i,1}$ *satisfies*

$$\int [1+\exp(-c_0-c_1 y_{i,1})]^{-1} f(y_{i,1}) dy_{i,1}$$

$$(4.11) \qquad = \begin{cases} \sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2)\Phi(\gamma-\beta k) \\ \quad + \frac{1}{c}\sum_{k=0}^{\infty}\left(-\frac{1}{c}\right)^k \exp(\beta^2(k+1)^2/2)[1-\Phi(\gamma+\beta k+\beta)]\,, & \beta>0\,, \\ \sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2)[1-\Phi(\gamma-\beta k)] \\ \quad + \frac{1}{c}\sum_{k=0}^{\infty}\left(-\frac{1}{c}\right)^k \exp(\beta^2(k+1)^2/2)\Phi(\gamma+\beta k+\beta)\,, & \beta<0\,, \\ \frac{1}{1+c}\,, & \beta=0\,, \end{cases}$$

where $c=\exp(-c_0-c_1\mu)$, $\beta=-c_1\sigma/\sqrt{1-\phi^2}$, $\gamma=-\log c/\beta$ *and* Φ *is the distribution function of standard normal distribution.*

Lastly, for the double integral $\iint f(y_{t-2}) f(y_{t-1}|y_{t-2}) f(y_{i,t}|y_{t-1}) \pi(x_i, y_{t-1}, y_{i,t}) \cdot \cdot [1-\pi(x_i, y_{t-2}, y_{t-1})] dy_{t-2} dy_{t-1}$, the series expansion approach is not applicable. Thus, it is necessary to consider other numerical methods for computing the double integral. Here, we adopt the Gauss-Hermite quadrature (Liu and Pierce [19]) to approximate it. Similarly, the Gauss-Hermite quadrature can also be employed to approximate the integral $\int \Phi(b_0+b_1 y_{i,1}) f(y_{i,1}) dy_{i,1}$ in (4.8). In `R`, the function `gauss.quad` under the package `statmod` can be employed. Simulations show that the choice of 9 nodes gives satisfactory performance. In summary, the computation of maximum likelihood function based on Lemma 4.2, Corollary 4.1 and the Gauss-Hermite quadrature has higher efficiency than that based on direct integration.

5. SIMULATION STUDIES

To evaluate the performance of the estimators obtained by dealing with informative sampling and nonignorable missingness in longitudinal surveys, we conduct a simulation study to compare the estimators under informative sampling and/or nonignorable missingness. In the simulation, $N=1000$ univariate normal values of $y_{i,1}$ are independently generated from $y_1 \sim N(\mu, \sigma^2/(1-\phi^2))$ for the first time period ($t=1$), where $\mu=0.8$, $\phi=0.3$ and $\sigma=0.5$. Then, we generate $N=1000$ population values of $y_{i,t}$ ($i=1,...,N$) at time $t=2,...,T$ with $T=10,20$ and 40 from the AR(1) model, $y_{i,t}-\mu=\phi(y_{i,t-1}-\mu)+\varepsilon_{i,t}$, where $\varepsilon_{i,t}\sim N(0,1)$ is independent error term. The AR(1) model parameters μ, ϕ and σ are of our interest.

For the sample selection, samples of size $n=10$, 20 and 40 are selected from the population via probability proportional to size (PPS) systematic sampling with size variable z. The size variable z values are generated in the following ways, which produce various sampling methods:

(**1**) Exponential sampling: $z_i = \exp(0.9 + 0.3y_{i,1} + \mu_i)$, $\mu_i \sim U(0, 1)$.

(**2**) Probit sampling: $z_i = \Phi(0.72 + 0.09y_{i,1} + \mu_i)$, $\mu_i \sim U(0, 2)$.

(**3**) Logistic sampling: $z_i = [1 + \exp(0.6 - 0.5y_{i,1} - \mu_i)]^{-1}$, $\mu_i \sim U(0, 5)$.

(**4**) Noninformative sampling: $z_i = \exp(1.5\mu_i)$, $\mu_i \sim U(0, 4)$.

Note that exponential sampling, probit sampling and logistic sampling are informative. Under the above sampling approaches, selection probabilities are defined as $\pi_i = nz_i \Big/ \sum_{i=0}^{N} z_i$.

For the missingness mechanism, the population value of the covariate is generated from $x_i \sim N(0, 1)$, $i = 1, ..., N$. We assume that the covariate x_i and the response variable $y_{i,1}$ at time $t = 1$ are always observed, but $y_{i,t}$ at time $t = 2, ..., T$ may subject to missingness. The response or missing indicator $\delta_{i,t}$ of $y_{i,t}$ are independently generated from a Bernoulli distribution with the response probabilities $\pi_{it}(\eta) = P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}; \eta)$ specified by $\pi_{it}(\eta) = [1 + \exp(-\eta_1 x_i - \eta_2 y_{i,t-1} - \eta_3 y_{i,t})]^{-1}$, where $\eta_1 = 0.2, \eta_2 = 0.4, \eta_3 = -0.5$. The average response rates under exponential sampling, probit sampling, logistic sampling and non-informative sampling are about 50% for the above nonignorable missing mechanism.

For samples under exponential sampling, probit sampling and logistic sampling, we compute the model parameter estimates by maximizing the sample likelihood under informative sampling and nonignorable missingness. For the sample under noninformative sampling, the model parameter estimators is obtained by maximizing the following log-likelihood function.

$$
\begin{aligned}
\log L &= \sum_{i=1}^{n} \log f(y_{i,1}; \theta) \\
&\quad + \sum_{t=2}^{T} \sum_{i=1}^{n} \delta_{i,t-1}\delta_{i,t} \Big\{ \log \pi(x_i, y_{i,t-1}, y_{i,t}; \eta) + \log f(y_{i,t}|y_{i,t-1}; \theta) \\
&\quad - \log \int \pi(x_i, y_{i,t-1}, y_t; \eta) f(y_t|y_{i,t-1}; \theta) dy_t \Big\} \\
&\quad + \sum_{t=2}^{T} \sum_{i=1}^{n} (1 - \delta_{i,t-1})\delta_{i,t} \\
&\quad \cdot \Big\{ \log \iint f(y_{t-2}) f(y_{t-1}|y_{t-2}) f(y_{i,t}|y_{t-1}) \pi(x_i, y_{t-1}, y_{i,t}) \\
&\quad \cdot [1 - \pi(x_i, y_{t-2}, y_{t-1})] dy_{t-2} dy_{t-1} - \log f(x_i, \delta_{i,t-1} = 0, \delta_{i,t} = 1) \Big\}.
\end{aligned} \tag{5.1}
$$

For comparison, we also compute the naive estimators, which ignore informative sampling and nonignorable missingness, and are obtained by maximizing the log-likelihood function (3.7). Moreover, the estimators obtained by ignoring informative sampling or nonignorable missingness under exponential sampling, probit sampling and logistic sampling are computed. The estimation procedure is repeated $B = 500$ times. For each estimator, the Monte Carlo biases (Bias), standard deviations (SD) under various n and T are presented. Besides, we also compute the estimation error $\|\hat{\theta} - \theta\|_2$ of the parameter $\theta = (\mu, \phi, \sigma)$, denoted by ER, and the standard deviation of ER to further measure the performance of θ. The results are provided in Tables 1, 2 and 3.

Table 1: Monte Carlo biases, standard deviations and estimation errors of the point estimators under $n = 10$ and $T = 10$.

Sampling	Estimate	Naive		Proposed		Ignore Sampling		Ignore Missingness	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
Exponential	$\widehat{\mu}$	−0.0103	0.1129	−0.0001	0.0735	0.0186	0.0732	−0.0947	0.1712
	$\widehat{\phi}$	−0.0241	0.2032	−0.0093	0.1196	−0.0096	0.1139	−0.0256	0.2003
	$\widehat{\sigma}$	−0.0232	0.0648	−0.0032	0.0507	−0.0003	0.0521	−0.0355	0.0638
Missing	$\widehat{\eta}_1$			−0.0135	0.0592	−0.0103	0.0636		
46.28%	$\widehat{\eta}_2$			0.0425	0.0548	0.0461	0.0527		
	$\widehat{\eta}_3$			0.0206	0.0526	0.0131	0.0559		
	$\widehat{a}_1$			0.0203	0.0657			0.4578	1.0580
	ER (SD)	0.2134 (0.1176)		0.1147 (0.0958)		0.1147 (0.0912)		0.2529 (0.1425)	
Probit	$\widehat{\mu}$	−0.7105	14.8879	0.0011	0.0743	0.0035	0.0777	−0.0919	0.2255
	$\widehat{\phi}$	−0.0420	0.4807	−0.0163	0.1061	−0.0170	0.1226	−0.0029	0.2371
	$\widehat{\sigma}$	−0.0184	0.1993	−0.0043	0.0500	−0.0051	0.0505	−0.0190	0.0725
Missing	$\widehat{\eta}_1$			−0.0092	0.0504	−0.0061	0.0616		
46.76%	$\widehat{\eta}_2$			0.0337	0.0462	0.0475	0.0645		
	$\widehat{\eta}_3$			0.0179	0.0452	0.0260	0.0509		
	$\widehat{b}_0$			0.0210	0.0517			8.3838	135.7257
	$\widehat{b}_1$			0.0176	0.0485			5.4545	114.8679
	ER (SD)	0.9055 (14.8865)		0.1061 (0.0910)		0.1175 (0.1007)		0.2855 (0.1988)	
Logistic	$\widehat{\mu}$	−0.0412	0.1105	−0.0021	0.0492	0.0032	0.0764	−0.0625	0.1091
	$\widehat{\phi}$	−0.0361	0.2065	0.0113	0.0460	−0.0150	0.1152	0.0188	0.0801
	$\widehat{\sigma}$	−0.0300	0.0612	0.0015	0.0425	−0.0033	0.0510	−0.0118	0.0555
Missing	$\widehat{\eta}_1$			−0.0055	0.0323	−0.0048	0.0600		
46.55%	$\widehat{\eta}_2$			0.0134	0.0252	0.0454	0.0536		
	$\widehat{\eta}_3$			0.0061	0.0217	0.0228	0.0561		
	$\widehat{c}_0$			0.0190	0.0241			0.0478	0.0497
	$\widehat{c}_1$			0.0289	0.0223			0.0656	0.0569
	ER (SD)	0.2183 (0.1213)		0.0631 (0.0499)		0.1145 (0.0938)		0.1335 (0.0892)	
Noninform	$\widehat{\mu}$	−0.0404	0.1103	0.0032	0.0779				
	$\widehat{\phi}$	−0.0411	0.2348	−0.0230	0.1312				
	$\widehat{\sigma}$	−0.0258	0.0660	−0.0029	0.0516				
Missing	$\widehat{\eta}_1$			−0.0052	0.0645				
46.29%	$\widehat{\eta}_2$			0.0516	0.0721				
	$\widehat{\eta}_3$			0.0213	0.0680				
	ER (SD)	0.2327 (0.1463)		0.1249 (0.1042)					

From Table 1, it can be seen that the proposed method that deals with informative sampling and nonignorable missingness simultaneously generally has smaller biases in comparison with the others under the four sampling mechanisms. As expected, the parameter estimation error of the proposed method is the smallest among all methods under various sampling schemes, followed by the estimators handling nonignorable missingness but ignoring informative sampling, whereas the estimation errors of the naive estimators and the estimators dealing with informative sampling but ignoring nonignorable missingness are relatively large among the four methods under exponential sampling, probit sampling and logistic sampling. Moreover, it is obvious that the proposed estimators of the parameters μ, ϕ, σ in AR(1) model have smaller biases than the naive estimators when the sampling design is noninformative.

All of these indicate that the proposed method has a good performance in handling nonignorable missingness. Besides, the proposed method generally yields the smallest standard deviations of the four methods for the estimation of the parameters μ, ϕ, σ under different sampling approaches. Similar results can be found in Table 2 and 3 which focus on different sample sizes. From Tables 1, 2 and 3, it can be seen as well that the estimation error of the proposed method decreases with the increase in the sample size n and the time period T for the four sampling schemes. It is noteworthy that the differences between the estimation errors of the proposed estimators and the estimators ignoring informative sampling but handling nonignorable missingness become smaller under various sampling schemes as n and T increase. This is reasonable because the sampling at time 1 may have a smaller effect on the estimation of the AR(1) model parameters as the time period T becomes larger. In conclusion, the proposed method performs best in the estimation of parameters.

Table 2: Monte Carlo biases, standard deviations and estimation errors of the point estimators under $n = 20$ and $T = 20$.

Sampling	**Estimate**	**Naive**		**Proposed**		**Ignore Sampling**		**Ignore Missingness**	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
Exponential	$\widehat{\mu}$	−0.0374	0.0625	0.0043	0.0439	0.0140	0.0395	−0.0904	0.0735
	$\widehat{\phi}$	−0.0069	0.0976	−0.0139	0.0710	−0.0115	0.0671	−0.0077	0.0963
	$\widehat{\sigma}$	−0.0053	0.0332	0.0039	0.0269	0.0048	0.0261	−0.0113	0.0331
	$\widehat{\eta}_1$			−0.0322	0.0507	−0.0319	0.0507		
Missing	$\widehat{\eta}_2$			0.0533	0.1096	0.0541	0.0399		
49.36%	$\widehat{\eta}_3$			0.0117	0.1068	0.0162	0.0381		
	$\widehat{a}_1$			0.0263	0.0452			0.3980	0.5521
	ER (SD)	0.1148 (0.0529)		0.0669 (0.0585)		0.0701 (0.0467)		0.1419 (0.0629)	
Probit	$\widehat{\mu}$	−0.0612	0.0629	0.0008	0.0381	0.0004	0.0400	−0.0880	0.0806
	$\widehat{\phi}$	−0.0020	0.0977	−0.0116	0.0621	−0.0063	0.0658	0.0061	0.1022
	$\widehat{\sigma}$	−0.0082	0.0341	0.0037	0.0262	0.0035	0.0273	−0.0052	0.0365
	$\widehat{\eta}_1$			−0.0337	0.0508	−0.0308	0.0543		
Missing	$\widehat{\eta}_2$			0.0422	0.0319	0.0553	0.0406		
49.40%	$\widehat{\eta}_3$			0.0157	0.0335	0.0232	0.0386		
	$\widehat{b}_0$			0.0218	0.0311			−1.4934	14.4169
	$\widehat{b}_1$			0.0197	0.0358			10.9502	95.3024
	ER (SD)	0.1233 (0.0571)		0.0636 (0.0457)		0.0675 (0.0465)		0.1472 (0.0662)	
Logistic	$\widehat{\mu}$	−0.0570	0.0617	−0.0010	0.0288	0.0036	0.0395	−0.0661	0.0638
	$\widehat{\phi}$	−0.0035	0.1012	0.0095	0.0331	−0.0083	0.0688	0.0170	0.0585
	$\widehat{\sigma}$	−0.0074	0.0329	0.0056	0.0209	0.0031	0.0251	0.0001	0.0301
	$\widehat{\eta}_1$			−0.0190	0.0322	−0.0285	0.0489		
Missing	$\widehat{\eta}_2$			0.0178	0.0198	0.0555	0.0426		
49.27%	$\widehat{\eta}_3$			0.0095	0.0215	0.0199	0.0366		
	$\widehat{c}_0$			0.0193	0.0191			0.0438	0.0394
	$\widehat{c}_1$			0.0308	0.0191			0.0601	0.0382
	ER (SD)	0.1222 (0.0592)		0.0433 (0.0246)		0.0677 (0.0491)		0.0984 (0.0579)	
Noninform	$\widehat{\mu}$	−0.0699	0.0622	0.0014	0.0423				
	$\widehat{\phi}$	0.0012	0.0985	0.0002	0.0641				
	$\widehat{\sigma}$	−0.0093	0.0331	0.0020	0.0258				
Missing	$\widehat{\eta}_1$			−0.0391	0.0540				
49.39%	$\widehat{\eta}_2$			0.0575	0.0398				
	$\widehat{\eta}_3$			0.0216	0.0393				
	ER (SD)	0.1253 (0.0625)		0.0681 (0.0438)					

Table 3: Monte Carlo biases, standard deviations and estimation errors of the point estimators under $n = 40$ and $T = 40$.

Sampling	**Estimate**	**Naive**		**Proposed**		**Ignore Sampling**		**Ignore Missingness**	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
Exponential	$\widehat{\mu}$	−0.0614	0.0342	0.0014	0.0297	0.0078	0.0282	−0.0890	0.0377
	$\widehat{\phi}$	0.0012	0.0504	−0.0097	0.0677	−0.0090	0.0613	0.0023	0.0504
	$\widehat{\sigma}$	−0.0040	0.0194	0.0039	0.0160	0.0051	0.0175	−0.0066	0.0190
Missing	$\widehat{\eta}_1$			−0.0746	0.0666	−0.0745	0.0660		
50.69%	$\widehat{\eta}_2$			0.0891	0.1987	0.1009	0.1384		
	$\widehat{\eta}_3$			−0.0025	0.1928	−0.0081	0.1423		
	$\widehat{a}_1$			0.0344	0.0425			0.3001	0.3354
	ER (SD)	0.0828 (0.0318)		0.0462 (0.0608)		0.0485 (0.0517)		0.1048 (0.0360)	
Probit	$\widehat{\mu}$	−0.0742	0.0316	0.00081	0.0224	0.0027	0.0346	−0.0904	0.0390
	$\widehat{\phi}$	0.0002	0.0474	−0.0085	0.0373	−0.0107	0.0813	0.0045	0.0503
	$\widehat{\sigma}$	−0.0051	0.0180	0.0043	0.0151	0.0052	0.0206	−0.0038	0.0189
Missing	$\widehat{\eta}_1$			−0.0788	0.0548	−0.0769	0.0784		
50.67%	$\widehat{\eta}_2$			0.0663	0.0372	0.1129	0.2268		
	$\widehat{\eta}_3$			0.0098	0.0325	−0.0103	0.2049		
	$\widehat{b}_0$			0.0301	0.0282			−0.7884	7.4087
	$\widehat{b}_1$			0.0261	0.0303			5.5026	38.9418
	ER (SD)	0.0905 (0.0302)		0.0394 (0.0256)		0.0501 (0.0765)		0.1059 (0.0372)	
Logistic	$\widehat{\mu}$	−0.0716	0.0344	−0.0029	0.0190	0.0029	0.0332	−0.0693	0.0358
	$\widehat{\phi}$	0.0061	0.0496	0.0094	0.0298	−0.0073	0.0741	0.0191	0.0392
	$\widehat{\sigma}$	−0.0040	0.0166	0.0070	0.0112	0.0062	0.0173	−0.0004	0.0160
Missing	$\widehat{\eta}_1$			−0.0389	0.0391	−0.0704	0.0632		
50.58%	$\widehat{\eta}_2$			0.0278	0.0179	0.1123	0.2690		
	$\widehat{\eta}_3$			0.0080	0.0243	−0.0142	0.2514		
	$\widehat{c}_0$			0.0245	0.0175			0.0494	0.0320
	$\widehat{c}_1$			0.0357	0.0170			0.0524	0.0323
	ER (SD)	0.0892 (0.0337)		0.0340 (0.0190)		0.0481 (0.0684)		0.0830 (0.0368)	
Noninform	$\widehat{\mu}$	−0.0754	0.0347	0.0013	0.0256				
	$\widehat{\phi}$	0.0004	0.0477	−0.0079	0.0465				
	$\widehat{\sigma}$	−0.0043	0.0169	0.0049	0.0160				
Missing	$\widehat{\eta}_1$			−0.0725	0.0551				
50.56%	$\widehat{\eta}_2$			0.0939	0.0724				
	$\widehat{\eta}_3$			0.0049	0.0746				
	ER (SD)	0.0915 (0.0331)		0.0451 (0.0335)					

6. REAL DATA ANALYSIS

The longitudinal data examined in this section comes from AIDS Clinical Trial Group 193A Study (Henry *et al.* [16]). It concerns AIDS patients with advanced immune suppression which is measured with CD4 counts. A total of 1309 patients were randomized to one of the four treatment groups including (1) 600mg zidovudine alternating monthly with 400mg didanosine, (2) 600mg zidovudine plus 2.25mg of zalcitabine, (3) 600mg zidovudine plus 400mg of didanosine, and (4) 600mg zidovudine plus 400mg of didanosine plus 400mg of nevirapine. The numbers of patients in the four treatment groups are $n = 325$, 324, 330 and 330, respectively. Treatments started at the time of week 0 (baseline), and were measured before the treatments and every 8 weeks. That is, data is collected on the 0, 8, 16, 24, 32, 40th weeks. Here, we denote the six follow-up time points by $t = 1, 2, 3, 4, 5, 6$. The measured outcome variable log(CD4 count + 1) is of our interest, whose values in six time intervals $(0, 4]$, $(4, 12]$, $(12, 20]$, $(20, 28]$, $(28, 36]$, $(36, 40]$ are viewed as y_t for $t = 1, 2, 3, 4, 5, 6$.

Note that the last record of the variable log(CD4 count + 1) in the interval is adopted as y_t if there are more than one values of log(CD4 count + 1) in a time interval. The covariates related to the response variable include Age (years) and Gender (Male=1, Female=0). Details on the data set can be found at `https://content.sph.harvard.edu/fitzmaur/ala/cd4.txt`.

In the longitudinal survey, the covariates are completely observed, whereas the response variable y_t (CD4 counts) is subject to missingness due to skipping visits or dropouts. In fact, a low CD4 count implies that HIV has damaged a patient's immune system to an extent that they are at risk of serious illnesses or even deaths. Thus, a lower CD4 count increases the chance of dropouts due to serious illnesses or deaths. As the patients' dropouts are related to the CD4 count, the missing process is potentially nonignorable. The missing rates under the four treatments are approximately 37.79%, 37.19%, 37.93% and 35.86%, respectively. Let $\delta_{i,t}$ be the indicator variable for $y_{i,t}$. Define

$$\delta_{i,t} = \begin{cases} 1 & \text{if } y_{i,t} \text{ is observed}\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{6.1}$$

for $i = 1, 2, ..., n$ and $t = 1, 2, 3, 4, 5, 6$. We are interested in estimating the response probability $P(\delta_{i,t} = 1 | x_i, y_{i,t-1}, y_{i,t})$. We fit the response model using the age variable x_1 and the gender variable x_2 in the following logistic model:

$$P(\delta_{i,t} = 1 | x_{i1}, x_{i2}, y_{i,t-1}, y_{i,t}) = \frac{\exp(\eta_1 x_{i1} + \eta_2 x_{i2} + \eta_3 y_{i,t-1} + \eta_4 y_{i,t})}{1 + \exp(\eta_1 x_{i1} + \eta_2 x_{i2} + \eta_3 y_{i,t-1} + \eta_4 y_{i,t})}\,, \tag{6.2}$$

where $\eta_1, \eta_2, \eta_3, \eta_4$ are the unknown parameters. This missing mechanism is obviously non-ignorable. For comparison, we also consider the following working model for the response probability under ignorable missing mechanism:

$$P(\delta_{i,t} = 1 | x_{i1}, x_{i2}) = \frac{\exp(\eta'_1 x_{i1} + \eta'_2 x_{i2})}{1 + \exp(\eta'_1 x_{i1} + \eta'_2 x_{i2})}\,, \tag{6.3}$$

where η'_1 and η'_2 are the unknown parameters. The response probability in equation (6.3) only depends on the covariates x_1 and x_2, implying that the missing mechanism is ignorable.

Assume that the sampling design is exponential sampling, probit sampling and logistic sampling, respectively. For comparison, we consider two models, the AR(1) model (2.1) and the following mean model.

$$y_{i,t} = \mu + \varepsilon_{i,t}\,, \quad i = 1, ..., n, \quad t = 1, ..., 6\,, \tag{6.4}$$

where $\varepsilon_{i,t} \sim N(0, \sigma^2)$. In fact, the mean model has no time dependence and been considered by Zhao *et al.* [39]. The estimates of model parameters μ, ϕ, σ under different missing models, sampling schemes and treatments, together with the mean squares of the model residuals (MSE), are presented in Tables 4 and 5.

As shown in Tables 4 and 5, Treatment 4 presents greater estimated values of μ than other Treatments regardless of models, missing mechanisms or sampling approaches. Also, the estimates of μ under Treatment 1 are the lowest among all treatments for all sampling methods and two missing models. That is, patients under Treatment 4 are superior to those under other Treatments in terms of the average number of CD4 counts, and the average number of patients' CD4 counts under Treatment 1 is relatively low. In fact, a high CD4 counts indicates a strong immune system, which suggests that the patient lives longer. This may reduce the possibility to drop outs for patients, which in turn reduces the differences between the parameter estimates under nonignorable missingness and ignorable missingness.

Table 4: Estimates for the AIDS clinical trial group 193A study data under nonignorable missingness.

Sampling	Estimate	Treatment 1		Treatment 2		Treatment 3		Treatment 4	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
		AR(1) Model	Mean Model	AR(1) Model	Mean Model	AR(1) Model	Mean Model	AR(1) Model	Mean Model
Exponential	$\widehat{\mu}$	2.5268	2.7442	2.6766	2.7326	2.6609	2.7989	2.8167	2.8772
	$\widehat{\phi}$	0.7124		0.6561		0.7228		0.7730	
	$\widehat{\sigma}$	0.7076	0.9504	0.7618	1.0893	0.7739	1.1018	0.7203	1.1377
	MSE	0.5539	0.6781	0.5848	0.8760	0.7674	1.0883	0.9008	1.3400
Probit	$\widehat{\mu}$	2.9169	2.7406	2.8934	2.8550	2.8528	2.9042	2.9490	3.1211
	$\widehat{\phi}$	0.6963		0.7092		0.7470		0.7591	
	$\widehat{\sigma}$	0.7202	0.9300	0.7641	1.0827	0.7526	1.1261	0.7392	1.1644
	MSE	0.5265	0.6784	0.5761	0.8511	0.7504	1.0439	0.8805	1.2657
Logistic	$\widehat{\mu}$	2.6969	2.7452	2.9060	2.7831	2.8900	2.7952	2.9263	2.9543
	$\widehat{\phi}$	0.6276		0.6951		0.7671		0.7809	
	$\widehat{\sigma}$	0.7544	0.9577	0.7740	1.0982	0.7597	1.1136	0.7288	1.1028
	MSE	0.5182	0.6780	0.5717	0.8621	0.7538	1.0903	0.8903	1.3036

Table 5: Estimates for the AIDS clinical trial group 193A study data under ignorable missingness.

Sampling	Estimate	Treatment 1		Treatment 2		Treatment 3		Treatment 4	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
		AR(1) Model	Mean Model	AR(1) Model	Mean Model	AR(1) Model	Mean Model	AR(1) Model	Mean Model
Exponential	$\widehat{\mu}$	2.5349	2.6818	2.6518	2.7504	2.7867	2.9202	3.1894	3.0855
	$\widehat{\phi}$	0.6718		0.6961		0.7288		0.7639	
	$\widehat{\sigma}$	0.7002	0.9481	0.7563	1.0625	0.7701	1.1311	0.7490	1.1440
	MSE	0.5428	0.6880	0.5938	0.8705	0.7523	1.0391	0.8573	1.2691
Probit	$\widehat{\mu}$	2.7210	2.7339	2.7974	2.7847	2.8407	2.8982	3.2598	3.1054
	$\widehat{\phi}$	0.6775		0.6994		0.7286		0.7692	
	$\widehat{\sigma}$	0.7065	0.9519	0.7614	1.0698	0.7728	1.1334	0.7365	1.1449
	MSE	0.5289	0.6792	0.5806	0.8617	0.7461	1.0458	0.8535	1.2669
Logistic	$\widehat{\mu}$	2.8759	2.7102	2.8401	2.7172	2.7586	2.9382	2.8827	2.9391
	$\widehat{\phi}$	0.6661		0.7182		0.7373		0.7777	
	$\widehat{\sigma}$	0.7525	0.9815	0.7634	1.0796	0.7772	1.0921	0.7411	1.1267
	MSE	0.5184	0.6825	0.5822	0.8812	0.7579	1.0343	0.8941	1.3098

This point is in line with the fact that the estimates of the key model parameter ϕ under nonignorable missingness are very close to those under ignorable missingness in the same Treatment 4 for various sampling approaches, whereas there is a clear difference between the parameter estimates of ϕ under nonignorable missingness and ignorable missingness in Treatment 1 for different sampling schemes. Moreover, the estimator of ϕ in the AR(1) model under Treatment 4 is the largest among all treatments under each informative sampling model for each missing mechanism, suggesting that the number of CD4 counts of Treatment 4 keeps decreasing more slowly in comparison with the others. Therefore, we conclude that Treatment 4 has better effect on the AIDS disease than other treatments. Besides, in terms of the variance estimators $\hat{\sigma}^2$ of residuals and MSE, the AR(1) model yields lower $\hat{\sigma}^2$ and MSE than the mean model. Thus, it seems very reasonable to use the AR(1) model over the mean model to analyze this data set.

A. APPENDIX

Proof of Lemma 4.1: First, the conditional density $f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1}=1, \delta_{i,t}=1)$ can be obtained, similar to Pfeffermann *et al.* [25], as

$$\text{(A.1)} \quad f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1, \delta_{i,t} = 1) = \frac{P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}, \delta_{i,t-1} = 1) f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1)}{\int P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_t, \delta_{i,t-1} = 1) f(y_t|x_i, y_{i,t-1}, \delta_{i,t-1} = 1) dy_t}.$$

The term $P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}, \delta_{i,t-1} = 1)$ on the right side of (A.1) can be written as

$$\text{(A.2)} \quad \begin{aligned} P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}, \delta_{i,t-1} = 1) &= \frac{P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}) P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}, y_{i,t}, \delta_{i,t} = 1)}{P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}, y_{i,t})} \\ &= P(\delta_{i,t} = 1|x_i, y_{i,t-1}, y_{i,t}) \\ &= \pi(x_i, y_{i,t-1}, y_{i,t}). \end{aligned}$$

The term $f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1)$ on the right side of (A.1) can be written as

$$\text{(A.3)} \quad f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1) = \frac{P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}, y_{i,t}) f(y_{i,t}|y_{i,t-1})}{P(\delta_{i,t-1} = 1|x_i, y_{i,t-1})},$$

where $f(y_{i,t}|y_{i,t-1}) = \exp\{-[y_{i,t} - \mu - \phi(y_{i,t-1} - \mu)]^2/2\sigma^2\}/\sqrt{2\pi}\sigma$.

Next, the two conditional probabilities of $\delta_{i,t-1}$ in (A.3) can be expressed as

$$\text{(A.4)} \quad \begin{aligned} P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}, y_{i,t}) &= \int P(\delta_{i,t-1} = 1|x_i, y_{t-2}, y_{i,t-1}) f(y_{t-2}|y_{i,t-1}, y_{i,t}) dy_{t-2} \\ &= \int \pi(x_i, y_{t-2}, y_{i,t-1}) f(y_{t-2}|y_{i,t-1}, y_{i,t}) dy_{t-2}, \end{aligned}$$

and

$$\text{(A.5)} \quad \begin{aligned} P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}) &= \int P(\delta_{i,t-1} = 1|x_i, y_{t-2}, y_{i,t-1}) f(y_{t-2}|y_{i,t-1}) dy_{t-2} \\ &= \int \pi(x_i, y_{t-2}, y_{i,t-1}) f(y_{t-2}|y_{i,t-1}) dy_{t-2}, \end{aligned}$$

respectively, where $\pi(x_i, y_{t-2}, y_{i,t-1})$ is defined in (3.6).

According to the AR(1) model, we can easily prove $f(y_{t-2}|y_{i,t-1}, y_{i,t}) = f(y_{t-2}|y_{i,t-1})$. Then, we have $P(\delta_{i,t-1} = 1|x_i, y_{i,t-1}, y_{i,t}) = P(\delta_{i,t-1} = 1|x_i, y_{i,t-1})$. Moreover, $f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1) = f(y_{i,t}|y_{i,t-1})$ holds. Thus, the conditional density in (A.1) can be written as

$$\text{(A.6)} \quad f(y_{i,t}|x_i, y_{i,t-1}, \delta_{i,t-1} = 1, \delta_{i,t} = 1) = \frac{\pi(x_i, y_{i,t-1}, y_{i,t}) f(y_{i,t}|y_{i,t-1})}{\int \pi(x_i, y_{i,t-1}, y_t) f(y_t|y_{i,t-1}) dy_t}.$$

Therefore, (4.4) in Lemma 4.1 holds.

Now we derive the results for $f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)$. Based on the definition of the conditional density, we have

$$f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1) = \frac{f(x_i,y_{i,t},\delta_{i,t-1}=0,\delta_{i,t}=1)}{f(x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)}, \tag{A.7}$$

where $f(x_i,y_{i,t},\delta_{i,t-1}=0,\delta_{i,t}=1)$ can be given by

$$\begin{aligned} &f(x_i,y_{i,t},\delta_{i,t-1}=0,\delta_{i,t}=1)\\ &= \iint f(x_i,y_{t-2},y_{t-1},y_{i,t})f(\delta_{i,t-1}=0,\delta_{i,t}=1|x_i,y_{t-2},y_{t-1},y_{i,t})dy_{t-2}dy_{t-1}\\ &= \iint f(x_i,y_{t-2})f(y_{t-1}|x_i,y_{t-2})f(y_{i,t}|x_i,y_{t-2},y_{t-1})P(\delta_{i,t}=1|x_i,y_{t-2},y_{t-1},y_{i,t})\\ &\quad \cdot P(\delta_{i,t-1}=0|x_i,y_{t-2},y_{t-1},y_{i,t},\delta_{i,t}=1)dy_{t-2}dy_{t-1}\\ &= \iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i,y_{t-1},y_{i,t})[1-\pi(x_i,y_{t-2},y_{t-1})]dy_{t-2}dy_{t-1}. \end{aligned} \tag{A.8}$$

Thus, we can obtain

$$\begin{aligned} &f(y_{i,t}|x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)\\ &= \frac{\iint f(y_{t-2})f(y_{t-1}|y_{t-2})f(y_{i,t}|y_{t-1})\pi(x_i,y_{t-1},y_{i,t})[1-\pi(x_i,y_{t-2},y_{t-1})]dy_{t-2}dy_{t-1}}{f(x_i,\delta_{i,t-1}=0,\delta_{i,t}=1)}. \end{aligned} \tag{A.9}$$

It follows that (4.5) in Lemma 4.1 holds. □

Proof of Lemma 4.2: According to $\pi(x_i,y_{i,t-1},y_{i,t}) = \exp(\eta_1x_i+\eta_2y_{i,t-1}+\eta_3y_{i,t})/[1+\exp(\eta_1x_i+\eta_2y_{i,t-1}+\eta_3y_{i,t})] = 1/[1+\exp(-\eta_1x_i-\eta_2y_{i,t-1}-\eta_3y_{i,t})]$ and $f(y_{i,t}|y_{i,t-1}) = (2\pi\sigma^2)^{-1/2}\exp\{-[y_{i,t}-\phi(y_{i,t-1}-\mu)-\mu]^2/(2\sigma^2)\}$, we have

$$\begin{aligned} &\int \pi(x_i,y_{i,t-1},y_t)f(y_t|y_{i,t-1})dy_t\\ &= \frac{1}{\sqrt{2\pi}\sigma}\int \frac{1}{1+\exp[-(\eta_1x_i+\eta_2y_{i,t-1}+\eta_3y_t)]}\exp\left\{-\frac{[y_t-\phi(y_{i,t-1}-\mu)-\mu]^2}{2\sigma^2}\right\}dy_t. \end{aligned} \tag{A.10}$$

Let $\tilde{\mu} = \mu+\phi(y_{i,t-1}-\mu)$ and $c = \exp[-(\eta_1x_i+\eta_2y_{i,t-1}+\eta_3\tilde{\mu})]$, we can obtain

$$\begin{aligned} &\int \pi(x_i,y_{i,t-1},y_t)f(y_t|y_{i,t-1})dy_t\\ &\quad = \frac{1}{\sqrt{2\pi}\sigma}\int \frac{1}{1+\exp\{-[\eta_1x_i+\eta_2y_{i,t-1}+\eta_3\tilde{\mu}+\eta_3(y_t-\tilde{\mu})]\}}\exp\left[-\frac{(y_t-\tilde{\mu})^2}{2\sigma^2}\right]dy_t\\ &\quad = \frac{1}{\sqrt{2\pi}\sigma}\int \frac{1}{1+c\cdot\exp(-\eta_3x)}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx\\ &\quad = \frac{1}{\sqrt{2\pi}}\int \frac{1}{1+c\cdot\exp(\beta y)}\exp\left(-\frac{y^2}{2}\right)dy, \end{aligned} \tag{A.11}$$

where $\beta = -\eta_3\sigma$.

When $\beta > 0$ and $0 < c \cdot \exp(\beta y) < 1$, we have $y < \gamma = -\log c/\beta$. Further, we can write

$$
\begin{aligned}
\text{(A.12)}\quad & \int \pi(x_i, y_{i,t-1}, y_t) f(y_t | y_{i,t-1}) dy_t \\
= & \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\gamma} \frac{1}{1 + c \cdot \exp(\beta y)} \exp\left(-\frac{y^2}{2}\right) dy + \int_{\gamma}^{\infty} \frac{1}{1 + c \cdot \exp(\beta y)} \exp\left(-\frac{y^2}{2}\right) dy \right] \\
= & \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\gamma} \sum_{k=0}^{\infty} [-c \cdot \exp(\beta y)]^k \exp\left(-\frac{y^2}{2}\right) dy \right. \\
& \left. + \frac{\exp(\beta^2/2)}{c} \int_{\gamma}^{\infty} \sum_{k=0}^{\infty} [-1/(c \cdot \exp(\beta y))]^k \exp\left[-\frac{(y+\beta)^2}{2}\right] dy \right] \\
= & \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\gamma} \sum_{k=0}^{\infty} (-c)^k \exp\left(\frac{\beta^2 k^2}{2}\right) \exp\left[-\frac{(y-\beta k)^2}{2}\right] dy \right. \\
& \left. + \frac{1}{c} \int_{\gamma}^{\infty} \sum_{k=0}^{\infty} \left(-\frac{1}{c}\right)^k \exp\left[\frac{\beta^2 (k+1)^2}{2}\right] \exp\left\{-\frac{[y + \beta(k+1)]^2}{2}\right\} dy \right] \\
= & \sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2) \Phi(\gamma - \beta k) + \frac{1}{c} \sum_{k=0}^{\infty} (-\frac{1}{c})^k \exp[\beta^2 (k+1)^2/2][1 - \Phi(\gamma + \beta k + \beta)] .
\end{aligned}
$$

Similarly, when $\beta < 0$ and $0 < c \cdot \exp(\beta y) < 1$, we have $y > \gamma = -\log c/\beta$. Then we can obtain

$$
\begin{aligned}
\text{(A.13)}\quad & \int \pi(x_i, y_{i,t-1}, y_t) f(y_t | y_{i,t-1}) dy_t \\
= & \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\gamma} \frac{1}{1 + c \cdot \exp(\beta y)} \exp\left(-\frac{y^2}{2}\right) dy + \int_{\gamma}^{\infty} \frac{1}{1 + c \cdot \exp(\beta y)} \exp\left(-\frac{y^2}{2}\right) dy \right] \\
= & \frac{1}{\sqrt{2\pi}} \left[\int_{\gamma}^{\infty} \sum_{k=0}^{\infty} (-c)^k \exp\left(\frac{\beta^2 k^2}{2}\right) \exp\left[-\frac{(y-\beta k)^2}{2}\right] dy \right. \\
& \left. + \frac{1}{c} \int_{-\infty}^{\gamma} \sum_{k=0}^{\infty} \left(-\frac{1}{c}\right)^k \exp\left[\frac{\beta^2 (k+1)^2}{2}\right] \exp\left\{-\frac{[y + \beta(k+1)]^2}{2}\right\} dy \right] \\
= & \sum_{k=0}^{\infty} (-c)^k \exp(\beta^2 k^2/2)[1 - \Phi(\gamma - \beta k)] + \frac{1}{c} \sum_{k=0}^{\infty} \left(-\frac{1}{c}\right)^k \exp[\beta^2 (k+1)^2/2] \Phi(\gamma + \beta k + \beta) .
\end{aligned}
$$

Specially, when $\beta = 0$, we get

$$
\int \pi(x_i, y_{i,t-1}, y_t) f(y_t | y_{i,t-1}) dy_t = \frac{1}{\sqrt{2\pi}} \int \frac{1}{1+c} \exp\left(-\frac{y^2}{2}\right) dy = \frac{1}{1+c} .
$$

Thus, Lemma 4.2 holds. □

Proof of Corollary 4.1: Note that the results in Lemma 4.2 can also be used to compute the integral $\int[1+\exp(-c_0-c_1y_{i,1})]^{-1}f(y_{i,1})dy_{i,1}$ in (4.9). Similar to the proof of Lemma 4.2, the integral $\int[1+\exp(-c_0-c_1y_{i,1})]^{-1}f(y_{i,1})dy_{i,1}$ can be written as

$$\begin{aligned}\text{(A.14)}\qquad &\int[1+\exp(-c_0-c_1y_{i,1})]^{-1}f(y_{i,1})dy_{i,1}\\ &\quad=\frac{\sqrt{1-\phi^2}}{\sqrt{2\pi}\sigma}\int\frac{1}{1+\exp(-c_0-c_1y_{i,1})}\exp\left\{-\frac{(1-\phi^2)(y_{i,1}-\mu)^2}{2\sigma^2}\right\}dy_{i,1}\,.\end{aligned}$$

Let $y=\sqrt{1-\phi^2}(y_{i,1}-\mu)/\sigma$, we have

$$\begin{aligned}\text{(A.15)}\qquad &\int[1+\exp(-c_0-c_1y_{i,1})]^{-1}f(y_{i,1})dy_{i,1}\\ &\quad=\frac{1}{\sqrt{2\pi}}\int\frac{1}{1+\exp[-c_0-c_1(\sigma y/\sqrt{1-\phi^2}+\mu)])}\exp\left(-\frac{y^2}{2}\right)dy\\ &\quad=\frac{1}{\sqrt{2\pi}}\int\frac{1}{1+c\cdot\exp(\beta y)}\exp\left(-\frac{y^2}{2}\right)dy\,,\end{aligned}$$

where $c=\exp(-c_0-c_1\mu)$ and $\beta=-c_1\sigma/\sqrt{1-\phi^2}$. Thus, we can compute the integral $\int[1+\exp(-c_0-c_1y_{i,1})]^{-1}f(y_{i,1})dy_{i,1}$ by replacing $c=\exp[-(\eta_1x_i+\eta_2y_{i,t-1}+\eta_3\tilde{\mu})]$ and $\beta=-\eta_3\sigma$ in Lemma 4.2 with $c=\exp(-c_0-c_1\mu)$ and $\beta=-c_1\sigma/\sqrt{1-\phi^2}$. It follows that Corollary 4.1 holds. □

ACKNOWLEDGMENTS

The authors thank the Editor, the Associate Editor and referees for their constructive comments. The collaborative work described in this paper was supported by HKSAR-RGC-GRF Nos. 14305517, 14601015 and 14302719 (Yau) and National Social Science Foundation of China, No. 18BTJ022 (Liu).

REFERENCES

[1] Almohisen, A.; Henderson, R. and Alshingiti, A.M. (2019). An alternative sensitivity approach for longitudinal analysis with dropout, *Journal of Probability and Statistics*, `https://doi.org/10.1155/2019/1019303`.

[2] Bahari, F.; Parsi, S. and Ganjali, M. (2019). Empirical likelihood inference in general linearmodel with missing values in response and covariates by MNAR mechanism, *Statistical Papers*, `https://doi.org/10.1007/s00362-019-01103-0`.

[3] Bhuyan, P. (2019). Estimation of random-effects model for longitudinal data with nonignorable missingness using Gibbs sampling, *Computational Statistics*, `https://doi.org/10.1007/s00180-019-00887-x`.

[4] Bonnery, D.; Breidt, F.J. and Coquet, F. (2018). Asymptotics for the maximum sample likelihood estimator under informative selection from a finite population, *Bernoulli*, **24**(2), 929–955.

[5] Boudreau, C. and Lawless, J.F. (2006). Survival analysis based on the proportional hazards model and survey data, *The Canadian Journal of Statistics*, **34**(2), 203–216.

[6] Cai, T. (2013). Investigation of ways to handle sampling weights for multilevel model analyses, *Sociological Methodology*, **43**(1), 178–219.

[7] Chambers, R.L. and Skinner, C.J. (2003). *Analysis of Survey Data*. John Wiley and Sons, Ltd., New York.

[8] Cole, B.F.; Bonetti, M.; Zalavasky, A.M. and Gelber, R.D. (2005). A multistate markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missingness, *Statistics in Medicine*, **24**(15), 2317–2334.

[9] Eideh, A.A.H. (2010). Fitting general linear model for longitudinal survey data under informative sampling, *Statistics in Transition-new series*, **11**(3), 517–538.

[10] Eideh, A.A.H. and Nathan, G. (2009a). *Joint treatment of nonignorable dropout and informative sampling for longitudinal survey data*. In "Methodology of Longitudinal Surveys" (P. Lynn, Ed.), pp. 251-264, John Wiley and Sons, Ltd., New York.

[11] Eideh, A.A.H. and Nathan, G. (2009b). Two-stage informative cluster sampling — estimation and prediction with applications for small-area models, *Journal of Statistical Planning and Inference*, **139**(9), 3088–3101.

[12] Eideh, A.A.H. and Nathan, G. (2006a). Fitting time series models for longitudinal survey data under informative sampling, *Journal of Statistical Planning and Inference*, **136**(9), 3052–3069.

[13] Eideh, A.A.H. and Nathan, G. (2006b). The analysis of data from sampling surveys under informative sampling, *Acta et Commentationes Universitatis Tartuensis de Mathematica*, **10**, 41–51.

[14] Farahania, Z.S.M.; Khorrama, E.; Ganjalib, M. and Baghfalakic, T. (2019). Longitudinal data analysis in the presence of informative sampling: weighted distribution or joint modelling, *Journal of Applied Statistics*, **46**(12), 2111–2127.

[15] Fuller, W.A. (2009). *Sampling Statistics*, John Wiley and Sons, Inc., New Jersey.

[16] Henry, K.; Erice, A.; Tierney, C.; Balfour, J.H.; Fischl, M.A.; Kmack, A.; Liou, S.H.; Kenton, A.; Hirsch, M.S.; Phair, J.; Martinez, A. and for the AIDS Clinical Trial Group 193A Study Team (1998). A randomized, controlled, double-blind study comparing the survival benefit of four different reverse transcriptase inhibitor therapies (three-drug, two-drug, and alternating drug) for the treatment of advanced AIDS, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **19**(4), 339–349.

[17] Kim, J.K. and Skinner, C.J. (2013). Weighting in survey analysis under informative sampling, *Biometrika*, **100**(2), 385–398.

[18] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis With Missing Data* (Second Edition), John Wiley and Sons, Inc., New Jersey.

[19] Liu, Q. and Pierce, D.A. (1994). A note on Gauss-Hermite quadrature, *Biometrika*, **81**(3), 624–629.

[20] Moser, A.; Bopp, M.; Zwahlen, M. and Swiss National Cohort Study Group (2018). Calibration adjustments to address bias in mortality analyses due to informative sampling — a census-linked survey analysis in Switzerland, `https://doi.org/10.7717/peerj.4376`.

[21] Pfeffermann, D. (2017). Bayes-based non-bayesian inference on finite populations from non-representative samples: a unified approach, *Calcutta Statistical Association Bulletin*, **69**(1), 35–63.

[22] Pfeffermann, D. and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas, *Journal of the American Statistical Association*, **102**(480), 1427–1439.

[23] Pfeffermann, D.; Moura, F.A.D.S. and Silva, P.L.D.N. (2006). Multi-level modelling under informative sampling, *Biometrika*, **93**(4), 943–959.

[24] Pfeffermann, D. and Sverchkov, M. (2003). *Fitting generalized linear models under informative sampling*. In "Analysis of Survey Data" (Southampton, 1999), Wiley Series in Survey Methodology, pp. 175–195, Wiley, Chichester.

[25] Pfeffermann, D.; Krieger, A.M. and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling, *Statistica Sinica*, **8**, 1087–1114.

[26] Pfeffermann, D. (1996). The use of sampling weights for survey data analysis, *Statistical Methods in Medical Research*, **5**(3), 239–261.

[27] Pfeffermann, D. (1993). The role of sampling weights when modeling survey data, *International Statistical Review*, **61**, 317–337.

[28] Qin, J.; Leung, D. and Shao, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling, *Journal of the American Statistical Association*, **97**(457), 193–200.

[29] Sabry, M.Y.; Kholy, R.B.E. and Gad, A.M. (2016). Generalized linear mixed models for longitudinal data with missing values: a monte carlo EM approach, *International Journal of Probability and Statistics*, **5**(3), 82–88.

[30] Schlomer, G.L.; Bauman, S. and Card, N.A. (2010). Best practices for missing data management in counseling psychology, *Journal of Counseling Psychology*, **57**(1), 1–10.

[31] Sikov, A. and Stern, J.M. (2019). Application of the full Bayesian significance test to model selection under informative sampling, *Statistical Papers*, **60**(1), 89–104.

[32] Stubbendick, A.L. and Ibrahim, J.G. (2006). Likelihood-based inference with nonignorable missing responses and covariates in models for discrete longitudinal data, *Statistica Sinica*, **16**, 1143–1167.

[33] Sverchkov, M. and Pfeffermann, D. (2018). Small area estimation under informative sampling and not missing at random non-response, *Journal of the Royal Statistical Society: Series A*, **181**(4), 981–1008.

[34] Taisir, R. and Islam, M.A. (2014). EM algorithm for longitudinal data with non-ignorable missing values: an application to health data, *Bangladesh Journal of Scientific Research*, **27**(2), 133–142.

[35] Thompson, M.E. (2015). Using longitudinal complex survey data, *Annual Review of Statistics and Its Application*, **2**, 305–320.

[36] Tseng, C.H.; Elashoff, R.; Li, N. and Li, G. (2016). Longitudinal data analysis with non-ignorable missing data, *Statistical Methods in Medical Research*, **25**(1), 205–220.

[37] Verret, F.; Rao, J.N.K. and Hidiroglou, M.A. (2015). Model-based small area estimation under informative sampling, *Survey Methodology*, **41**(2), 333–347.

[38] Yuan, Y. and Yin, G.S. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data, *Biometrics*, **66**, 105–114.

[39] Zhao, P.Y.; Wang, L. and Shao, J. (2018). Analysis of longitudinal data under nonignorable nonmonotone nonresponse, *Statistics and Its Interface*, **11**(2), 265–279.