Statistics and Its Interface Volume 11 (2018) 265–279

Analysis of longitudinal data under nonignorable nonmonotone nonresponse

Puying Zhao, Lei Wang*, and Jun Shao

We consider identification and estimation in a longitudinal study with nonignorable nonmonotone nonresponse in responses. To handle the identifiability issue, we use a baseline covariates named as nonresponse instrument that can be excluded from the nonresponse propensity conditional on other observed covariates and the variables subject to nonresponse. The generalized method of moments is applied to estimate the parameters in the nonresponse propensity. Marginal response means and the parameters defined via regression models between responses and baseline covariates can be estimated by inverse probability weighting using the estimated propensity. Alternatively, we derive an augmented inverse probability weighting estimator and apply the importance sampling technique for its computation. Consistency and asymptotic normality of the proposed estimators are established under possibly misspecified models. Simulations are performed to evaluate the finite sample performance of the estimators. Also, a real data example is presented to demonstrate the proposed methodology.

AMS 2000 subject classifications: Primary 62H12, 62G05; secondary 62G20.
Keywords and phrases: Generalized method of moments, Identifiability, Instrument, Misspecified models, Nonignorable nonmonotone nonresponse, Robustness.

1. INTRODUCTION

In survey sampling, social science, epidemiology studies and many other statistical problems, data are often collected from every sampled subject at multiple time points, which are referred to as longitudinal data. Missing data are often encountered in longitudinal studies, due to drop-out, mistimed measurements, subjects being too sick to visit the clinic (Ibrahim and Molenberghs, 2009). Missingness is ignorable or at random if the missing data mechanism/propensity depends on observed data only (Little and Rubin, 2002), and is monotone if a datum is missing at time t implies that all data at time $s > t$ are missing. When missing data are ignorable and/or monotone, Laird (1988) and Little (1995) present some well-established methods. In practice, however, missing data are often nonignorable and nonmonotone, i.e., the propensity depends on missing values and subjects move in and out of the study as time progresses, which presents great challenges for statistical analysis. Although some effort has been devoted to handle nonignorable nonmonotone nonresponse (see, e.g., Troxel et al., 1998; Vansteelandt et al., 2007; Xu et al., 2008; Shao et al., 2012), there is no general well-established method and the research on this topic is far from complete. The major issue is that the propensity is not identifiable due to nonignorable nonresponse and is very complicated due to nonmonotone nonresponse (Robins and Ritov, 1997; Scharfstein et al., 1999). Assumptions must be imposed to propensity in order to derive useful results, but these assumptions are difficult or impossible to verify.

Vansteelandt et al. (2007) proposed a class of occasion-specific tilted models for nonresponse propensity, but they assumed that the part in propensity causing nonignorable nonresponse is known to avoid the identifiability issue. A sensitivity analysis was proposed to deal with the situation where one does not know that information, but sensitivity analysis is ad hoc and has limited application scope.

Different from the sensitivity analysis in Vansteelandt et al. (2007), we propose to estimate the unknown nonignorable nonresponse propensity by imposing a parametric model on the propensity that does not involve a baseline covariate used to identify the parameters in the propensity model. Using a covariate not involved in the propensity to deal with the identifiability issue has been studied in Wang et al. (2014) and Shao and Wang (2016) and such a covariate is called nonresponse instrument. Details are given in Section 2, where we apply the generalized method of moments (GMM, Hansen, 1982) to estimate the propensity and establish the consistency and asymptotic normality of the estimators.

Our second achievement is to derive three consistent and asymptotically normal estimators of marginal means of the responses or parameters in a regression between the responses and baseline covariates in the longitudinal study. The main technique we use is the inverse probability weighting (IPW, Robins et al., 1994). Alternatively, we construct augmented IPW (AIPW) estimators by making use of an identity relating the conditional density for nonrespondents and the quantities that are functions of observed data. The AIPW estimators are model-assisted estimators in the sense that they are efficient when the working model used to form augmented data is correct, but they are still consistent and asymptotically normal when the working model is misspecified. Details are given in Sections 3 and 4.

*Corresponding author. The first two authors contributed equally to this work.

In addition, we consider in Section 5 an algorithm for the computation of the AIPW estimators using the importance sampling technique.

Simulation studies and a data example for illustration are presented in Section 6. Section 7 concludes with a discussion. Technical details are given in the Appendix.

2. IDENTIFICATION AND ESTIMATION OF PROPENSITY

Consider a longitudinal study with a unit having time-dependent response Y_t and covariate vector V_t at time $t = 0, 1, ..., T$. At the baseline $t = 0$, $S_0 = (Y_0, V_0)$ together with an additional time-independent covariate vector X are observed. When $t \geq 1$, Y_t may be missing although V_t is observed. Let δ_t be the response indicator that equals 1 if Y_t is observed and equals 0 otherwise. We consider nonmonotone nonresponse so that $\delta_t = 0$ does not imply that $\delta_{t+1} = 0$. For the nonresponse propensity, we assume that X can be decomposed as $X = (U, Z)$, where U is continuous but Z can be any kind, such that for $t = 1, ..., T$,

$$\begin{aligned} & \Pr(\delta_t = 1 | \delta_{-t}, Y, V, X) \\ = & \Pr(\delta_t = 1 | S_0, \delta_1 S_1, ..., \delta_{t-1} S_{t-1}, S_t, U) \\ = & \pi_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0), \end{aligned} \tag{1}$$

where $\delta_{-t} = (\delta_1, ..., \delta_{t-1}, \delta_{t+1}, ..., \delta_T)$, $Y = (Y_0, Y_1, ..., Y_T)$, $V = (V_0, V_1, ..., V_T)$, $\overrightarrow{S}_{t-1} = (U, S_0, \delta_1 S_1, ..., \delta_{t-1} S_{t-1})$ is the observed history of S_j, $j < t$, including the baseline covariate U, π_t is a known function, and α_t^0 is an unknown parameter vector. The first equation in assumption (1) means that the nonresponse of components of S_t does not depend on future values but depends on S_t so that nonresponse is nonignorable; given S_t, $\overrightarrow{S}_{t-1}$, and the baseline covariate U, the baseline covariate Z can be excluded from the nonresponse propensity, which is used to create more estimation equations for estimating the propensity and ensures that the propensity is identifiable. Such a Z is referred to as a nonresponse instrument (Wang et al., 2014). The second equation in assumption (1) imposes a parametric model on the propensity.

To estimate the propensity, we consider an independent and identically distributed sample of n units. For each $i = 1, ..., n$, values of U, Z, S_t, δ_t, and $\overrightarrow{S}_t$ previously described are denoted by U_i, Z_i, S_{it}, δ_{it}, and $\overrightarrow{S}_{it}$, respectively.

First, we consider a continuous r-dimensional Z. Define estimating functions

$$g_t(Y, V, X, \delta, \alpha_t) = \overrightarrow{\xi}_t \left\{ \frac{\delta_t}{\pi_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t)} - 1 \right\}, \tag{2}$$

where $\delta = (\delta_1, ..., \delta_T)$ and $\overrightarrow{\xi}_t$ is a known vector-valued function of Z and $\overrightarrow{S}_{t-1}$ with dimension $K \geq$ the dimension of α_t. Throughout the paper we denote $\mathbb{E}_n(\mathbb{Y}) = n^{-1} \sum_{i=1}^n \mathbb{Y}_i$ for any random element $\mathbb{Y}$. If α_t^0 is the true parameter value, $G_t(\alpha_t^0) = E\{g_t(Y, V, X, \delta, \alpha_t^0)\} = 0$. Then, when $K =$ the dimension of α_t, we estimate α_t^0 by solving the sample equations $\mathbb{E}_n\{g_t(Y, V, X, \delta, \alpha_t)\} = 0$. When $K >$ the dimension of α_t, the estimating equations are over-identified and we apply the generalized method of moments (GMM) approach as follows. Let $G_{nt}(\alpha_t) = \mathbb{E}_n\{g_t(Y, V, X, \delta, \alpha_t)\}$,

$$\tilde{\alpha}_t = \arg\min_{\alpha_t} G_{nt}(\alpha_t)^\top G_{nt}(\alpha_t)$$

be the first-step GMM estimator, where $a^\top$ is the transpose of a. Let $\hat{W}_{nt} = \mathbb{E}_n\{g_t(Y, V, X, \delta, \tilde{\alpha}_t) g_t(Y, V, X, \delta, \tilde{\alpha}_t)^\top\}$ and the second-step GMM estimator of α_t^0 is

$$\hat{\alpha}_t = \arg\min_{\alpha_t} G_{nt}(\alpha_t)^\top \hat{W}_{nt}^{-1} G_{nt}(\alpha_t). \tag{3}$$

Note that (2) is for α_t at time t only, although data at $t = 0, 1, ..., t$ are used. We may construct estimating equations for $(\alpha_1^\top, ..., \alpha_t^\top)^\top$, but the dimension of $(\alpha_1^\top, ..., \alpha_t^\top)^\top$ may be too large to numerically handle.

Consider in general $Z = (Z_c, Z_d)$, where Z_c is a continuous r-dimensional covariate vector and Z_d is a discrete covariate taking values $1, ..., J$. Then, the GMM estimation of the propensity is based on estimating equations: $E\{g_t(Y, V, X, \delta, \alpha_t^0)\} = 0$, where $g_t(Y, V, X, \delta, \alpha_t)$ is defined the same as that in (2) with $\overrightarrow{\xi}_t$ being a function of ζ, Z_c, and $\overrightarrow{S}_{t-1}$, ζ is the J-dimensional vector whose lth component is $I(Z_d = l)$, and $I(\cdot)$ is the indicator function.

Throughout this paper, expectations (i.e., notation E) are taken in respect to the true distribution. For notational simplicity, let $g_t(\alpha_t) = g_t(Y, V, X, \delta, \alpha_t)$ and $\pi_t(\alpha_t) = \pi_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t)$ for $t = 1, ..., T$. For any matrix A, let $\|A\| = \sqrt{\text{trace}(A^\top A)}$ and $A^{\otimes 2} = AA^\top$. Let $\xrightarrow{\mathcal{L}}$ denote convergence in distribution and C denote a generic positive constant which may vary depending on the context. The following theorem presents some asymptotic properties of the GMM estimator $\hat{\alpha}_t$ for every t. A sketch of the proof is in the Appendix.

Theorem 2.1. *Assume the following regularity conditions hold:*

- *C1. (a) The parameter space for α_t, $\mathcal{A}$, is a compact set and $\alpha_t^* \in \mathcal{A}$ is the unique solution to $G_t(\alpha_t) = 0$; (b) $\sup_{\alpha_t} \|G_t(\alpha_t)\| < \infty$; (c) uniformly for all $\alpha_t \in \mathcal{A}$, the matric $\Gamma_t(\alpha_t) = -E\{\overrightarrow{\xi}_t \pi_t(\alpha_t^0)[\pi_t(\alpha_t)^{-1} - 1]\Xi(\alpha_t)^\top\}$ is of full rank, where and $\Xi(\alpha_t) = \partial \text{logit}\{\pi_t(\alpha_t)\}/\partial \alpha_t$ with $\text{logit}(u) = \log\{u/(1-u)\}$; (d) The matrix $W_t(\alpha_t) = E\{g_t(\alpha_t)^{\otimes 2}\}$ is positive definite.*
- *C2. (a) The propensity model $\pi_t(\alpha_t)$ is twice differentiable with respect to α_t; (b) $\pi_t(\alpha_t^0) \geq C > 0$ for all $i = 1, ..., n$, and $t = 1, ..., T$; (c) $\partial \pi_t(\alpha_t)/\partial \alpha_t$ is uniformly bounded.*

Then, as $n \to \infty$,

$$n^{1/2}(\hat\alpha_t - \alpha_t^*) \xrightarrow{\mathcal{L}} N(0, \Sigma_t^*),$$

where $\Sigma_t^* = \{\Gamma_t^{*\top} W_t^{*-1}\Gamma_t^*\}^{-1}$ *with* $\Gamma_t^* = \Gamma_t(\alpha_t^*)$ *and* $W_t^* = W_t(\alpha_t^*)$.

In Theorem 2.1, large sample properties of the two-step GMM estimator $\hat\alpha_t$ are established for the case where the propensity model (1) may be misspecified so that α_t^* may not be the same as α_t^0. In the presence of misspecification, the proposed GMM procedure consistently estimates α_t^* minimizing the population version of the empirical GMM discrepancy. If the propensity model (1) is correctly specified, then $\hat\alpha_t$ is consistent for the true parameter vector α_t^0, which is summarized in the following corollary.

Corollary 2.1. *Assume that the regularity conditions C1 and C2 given in Theorem 2.1 hold, and the propensity model (1) is correctly specified. Then, as $n \to \infty$,*

$$n^{1/2}(\hat\alpha_t - \alpha_t^0) \xrightarrow{\mathcal{L}} N(0, \Sigma_t),$$

where $\Sigma_t = \{\Gamma_t^\top W_t^{-1}\Gamma_t\}^{-1}$ *with* $\Gamma_t = \Gamma_t(\alpha_t^0)$ *and* $W_t = W_t(\alpha_t^0)$.

3. ESTIMATION OF RESPONSE MEANS

Once a consistent estimator $\hat\alpha_t$ of α_t^0 is obtained, we can make inference on the marginal distribution of Y_t or the conditional distribution of Y_t given X. In this section we consider the estimation of response means $\mu_t^0 = E(Y_t)$, $t = 1, ..., T$. For a fixed t, μ_t^0 can be estimated by inverse probability weighting (IPW), i.e.,

$$\hat\mu_t^{\text{ipw1}} = \mathbb{E}_n\Big\{\frac{\delta_t Y_t}{\pi_t(\hat\alpha_t)}\Big\}, \tag{4}$$

$$\hat\mu_t^{\text{ipw2}} = \mathbb{E}_n\Big\{\frac{\delta_t Y_t}{\pi_t(\hat\alpha_t)}\Big\}\Big/\mathbb{E}_n\Big\{\frac{\delta_t}{\pi_t(\hat\alpha_t)}\Big\},$$

where $\pi_t(\hat\alpha_t) = \pi_t(\overrightarrow{S}_{t-1}, S_t, \hat\alpha_t)$. Since estimators $\hat\mu_t^{\text{ipw1}}$ and $\hat\mu_t^{\text{ipw2}}$ are constructed based on complete observations only, their estimation efficiency might be improved by imputation. To proceed, we first consider an artificial situation where

$$m_{0t}(\overrightarrow{S}_{t-1}) = E[Y_t|\overrightarrow{S}_{t-1}, \delta_t = 0], \quad t = 1, ..., T, \tag{5}$$

are observed statistics. An augmented inverse probability weighting (AIPW) estimator is

$$\hat\mu_t^{\text{aipw}} = \mathbb{E}_n\Big\{\frac{\delta_t Y_t}{\pi_t(\hat\alpha_t)} - \frac{\delta_t - \pi_t(\hat\alpha_t)}{\pi_t(\hat\alpha_t)} m_{0t}(\overrightarrow{S}_{t-1})\Big\}. \tag{6}$$

The second term on the right-hand side of (6) is used for a possible efficiency improvement over $\hat\mu_t^{\text{ipw1}}$. In real applications, however, $m_{0t}(\overrightarrow{S}_{t-1})$ in (5) is unknown and has to be estimated. Under assumption (1), after some algebraic manipulations, we obtain that

$$\frac{\Pr(S_t \in B|\overrightarrow{S}_{t-1}, \delta_t = 0)}{\Pr(S_t \in B|\overrightarrow{S}_{t-1}, \delta_t = 1)} = \frac{\Pr(\delta_t = 0|S_t \in B, \overrightarrow{S}_{t-1})/\Pr(\delta_t = 1|S_t \in B, \overrightarrow{S}_{t-1})}{\Pr(\delta_t = 0|\overrightarrow{S}_{t-1})/\Pr(\delta_t = 1|\overrightarrow{S}_{t-1})}, \tag{7}$$

for each t and any Borel set B, and

$$E\{\delta_t \mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0)|\overrightarrow{S}_{t-1}\} = \Pr(\delta_t = 0|\overrightarrow{S}_{t-1}),$$

where $\mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0) = \{\pi_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0)\}^{-1} - 1$ denotes the conditional odds of nonresponse. These results imply that

$$f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 0) = \frac{f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)\mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0)}{E\{\mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0)|\overrightarrow{S}_{t-1}, \delta_t = 1\}}, \tag{8}$$

where $f(A|B)$ denotes the conditional density of A given B. It follows from (8) that the joint conditional density $f(S_t|\overrightarrow{S}_{t-1})$ is identifiable under assumption (1). Moreover, (8) relates the joint conditional density $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 0)$ for nonrespondents to the quantities calculated based on observed data. As a result,

$$\begin{aligned} m_{0t}(\overrightarrow{S}_{t-1}) &= \int Y_t f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 0) dS_t \\ &= \frac{\int Y_t \mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0) f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1) dS_t}{\int \mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0) f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1) dS_t}. \end{aligned}$$

Thus, in order to estimate $m_{0t}(\overrightarrow{S}_{t-1})$, it suffices to estimate $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)$, since $\pi_t(\alpha_t^0)$ and hence $\mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \alpha_t^0)$ has already been estimated (Section 2). The estimation of $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)$ can be done using observed data.

Because the estimation of $m_{0t}(\overrightarrow{S}_{t-1})$ is for the purpose of improving efficiency, and because the AIPW estimator $\hat\mu_t^{\text{aipw}}$ in (6) is a model-assisted estimator in the sense that as long as the propensity estimator $\pi_t(\hat\alpha_t)$ is consistent, $\hat\mu_t^{\text{aipw}}$ is consistent regardless of whether $m_{0t}(\overrightarrow{S}_{t-1})$ is consistent or not, we propose to apply a parametric method, i.e., we consider a working parametric model $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1) = f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^0)$, where f_t is a known function and γ_t^0 is an unknown s_t-dimensional parameter, and estimate γ_t^0 by the maximum likelihood estimator $\hat\gamma_t$ based on observed data, which is a solution of

$$\mathbb{E}_n\big\{\delta_t \partial \log f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t)/\partial\gamma_t\big\} = 0. \tag{9}$$

Then, $m_{0t}(\overrightarrow{S}_{t-1})$ can be estimated as

$$\hat m_{0t}(\hat\alpha_t, \hat\gamma_t) = \frac{\int Y_t \mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \hat\alpha_t) f_t(S_t|\overrightarrow{S}_{t-1}, \hat\gamma_t) dS_t}{\int \mathcal{O}_t(\overrightarrow{S}_{t-1}, S_t, \hat\alpha_t) f_t(S_t|\overrightarrow{S}_{t-1}, \hat\gamma_t) dS_t}. \tag{10}$$

Our proposed AIPW estimator $\hat{\mu}_t^{\text{aipw}}$ is given by equation (6) with $m_{0t}(\overrightarrow{S}_{t-1})$ replaced by its estimator $\hat{m}_{0t}(\hat{\alpha}_t, \hat{\gamma}_t)$.

The following proposition presents some asymptotic properties of $\hat{\gamma}_t$ defined by (9) for the case where the working model $f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^0)$ may be misspecified.

Proposition 3.1. *Assume the following regularity conditions:*

C3. (a) The parameter space for γ_t^0, Υ_t, is a compact set; (b) $E\{\log f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)\}$ exists, $t = 1, ..., T$; (c) the Kullback-Leibler Information Criterion defined as $E\{\log [f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)/f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t)]\}$ has a unique minimum at $\gamma_t^* \in \Upsilon_t$.

C4. (a) The assumed parametric model $f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t)$ satisfies the regularity conditions of maximum likelihood estimation of misspecified models, see conditions A1–A6 in White (1982); (b) uniformly for all $\gamma_t \in \Upsilon_t$, the matrix $A_t(\gamma_t) = E\{\partial^2 \log f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t)/\partial\gamma_t\partial\gamma_t^\top\}$ is of full rank; (c) the matrix $B_t(\gamma_t^*) = E\{[\partial \log f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^*)/\partial\gamma_t]^{\otimes 2}\}$ is nonsingular.

Then, as $n \to \infty$,

$$n^{1/2}(\hat{\gamma}_t - \gamma_t^*) \xrightarrow{\mathcal{L}} N(0, \omega_t^{-1}\mathcal{C}(\gamma_t^*)),$$

where $\mathcal{C}(\gamma_t) = A_t(\gamma_t)^{-1}B_t(\gamma_t)A_t(\gamma_t)^{-1}$ and $\omega_t = \Pr(\delta_t = 1)$.

Let $\Delta_{1t} = E\{Y_t[1 - \pi_t(\alpha_t^0)]\Xi(\alpha_t^0)\}$, $\Delta_{2t} = E\{(Y_t - \mu_t^0)[1 - \pi_t(\alpha_t^0)]\Xi(\alpha_t^0)\}$, $\eta_t(\alpha_t^0) = [\delta_t - \pi_t(\alpha_t^0)]\Xi(\alpha_t^0)$, $\Lambda_t = (\Gamma_t^\top W_t^{-1}\Gamma_t)^{-1}\Gamma_t^\top W_t^{-1}$, $\Phi_t = \text{Cov}[\delta_t\pi_t(\alpha_t^0)^{-1}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}, \eta_t(\alpha_t^0)]$,

$$H_t(\alpha_t, \gamma_t) = \frac{\delta_t Y_t}{\pi_t(\alpha_t)} - \frac{\delta_t - \pi_t(\alpha_t)}{\pi_t(\alpha_t)} m_{0t}(\alpha_t, \gamma_t), \tag{11}$$

where $m_{0t}(\alpha_t, \gamma_t)$ is given by (10) with $\hat{\alpha}_t$ and $\hat{\gamma}_t$ replaced by α_t and γ_t, respectively. The following theorem presents some asymptotic properties of the proposed IPW estimators $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$ and AIPW estimator $\hat{\mu}_t^{\text{aipw}}$. A sketch of the proof is given in the Appendix.

Theorem 3.1. *Assume that the regularity conditions C1 and C2 given in Theorem 2.1 hold, and the propensity model (1) is correct. As $n \to \infty$,*
(i) for $\kappa = 1, 2$,

$$n^{1/2}(\hat{\mu}_t^{\text{ipw}\kappa} - \mu_t^0) \xrightarrow{\mathcal{L}} N(0, V_{\kappa t}),$$

where $V_{1t} = \text{Var}\{\delta_t Y_t \pi_t(\alpha_t^0)^{-1} + \Delta_{1t}^\top \Lambda_t g_t(\alpha_t^0)\}$ and $V_{2t} = \text{Var}\{\delta_t \pi_t(\alpha_t^0)^{-1}(Y_t - \mu_t^0) + \Delta_{2t}^\top \Lambda_t g_t(\alpha_t^0)\}$;
(ii)

$$n^{1/2}(\hat{\mu}_t^{\text{aipw}} - \mu_t^0) \xrightarrow{\mathcal{L}} N(0, V_{3t}),$$

where $V_{3t} = \text{Var}\{H_t(\alpha_t^0, \gamma_t^*) + \Phi_t^\top \Lambda_t g_t(\alpha_t^0)\}$.

The second part of Theorem 3.1 shows that the proposed AIPW estimator $\hat{\mu}_t^{\text{aipw}}$ is consistent and asymptotically normal even if model $f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^0)$ is misspecified. This is important in our problem because it is difficult to impose an exactly correct model for $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)$ due to the fact that data are longitudinal with nonignorable nonmonotone nonresponse. Furthermore, the result indicates that the asymptotic distribution of $\hat{\mu}_t^{\text{aipw}}$ does not depend on the efficiency of $\hat{\gamma}_t$. Simulation results in Section 6 show that $\hat{\mu}_t^{\text{aipw}}$ improves the IPW estimators when the model $f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^0)$ is nearly but not exactly correct.

4. ESTIMATION OF REGRESSION MODELS

We now consider the estimation of a parameter vector β^0 defined via T regression models between Y_t and the baseline covariate vector X:

$$E(Y_t|X) = \mathcal{G}_t(X; \beta^0), t = 1, ..., T, \tag{12}$$

where $\mathcal{G}_t(X; \beta^0)$ is a known continuously differentiable function and β^0 is an unknown p-dimensional parameter vector. Without missing data, suppose that β^0 is estimated by solving $E\{\mathcal{D}_t(X, \beta^0)[Y_t - \mathcal{G}_t(X; \beta^0)]\} = 0$, where $\mathcal{D}_t(X, \beta)$ is a vector function with the same dimension as β^0. We assume that the nonresponse propensity model is of the same form as in (1), even though the regression models (12) are defined without involving time-dependent variables V_t. Here V_t can be treated as additional auxiliary variables. With the nonresponse and propensity assumption (1), the estimating equation becomes

$$E\Big\{\frac{\delta_t \mathcal{D}_t(X, \beta^0)[Y_t - \mathcal{G}_t(X; \beta^0)]}{\pi_t(\alpha_t^0)}\Big\} = 0.$$

Thus, with α_t^0 estimated by $\hat{\alpha}_t$ in Section 2, a consistent IPW estimator $\hat{\beta}^{\text{ipw}}$ of β^0 can be obtained by solving

$$\hat{U}_I(\beta, \hat{\alpha}_t) = \sum_{t=1}^{T} \mathbb{E}_n\Big\{\frac{\delta_t \mathcal{D}_t(X, \beta)[Y_t - \mathcal{G}_t(X; \beta)]}{\pi_t(\hat{\alpha}_t)}\Big\} = 0. \tag{13}$$

The following theorem presents large sample properties of the IPW estimator $\hat{\beta}^{\text{ipw}}$ for the case that either the propensity model (1) or the marginal regression model (12) is misspecified. A sketched proof is in the Appendix. Define $\mathscr{D}_t(\beta) = \partial\mathcal{D}_t(X, \beta)/\partial\beta^\top$,

$$\begin{aligned}
U_I(\beta, \alpha_t) =& E\{\sum_{t=1}^{T} \delta_t \pi_t(\alpha_t)^{-1}\mathcal{D}_t(X, \beta)[Y_t - \mathcal{G}_t(X; \beta)]\},\\
\mathcal{J}_1(\beta, \alpha_t) =& E\{\sum_{t=1}^{T} \pi(\alpha_t^0)\pi(\alpha_t)^{-1}\mathscr{D}_t(X, \beta)[Y_t - \mathcal{G}_t(X; \beta)]\},\\
\mathcal{J}_2(\beta, \alpha_t) =& E\{\sum_{t=1}^{T} \pi(\alpha_t^0)\pi(\alpha_t)^{-1}\mathcal{D}_t(X, \beta)\mathcal{D}_t(X, \beta)^\top\},\\
\mathcal{J}_3(\beta, \alpha_t) =& E\{\sum_{t=1}^{T} \pi_t(\alpha_t^0)[\pi_t(\alpha_t)^{-1} - 1]\mathcal{D}_t(X, \beta)\\
&\times [Y_t - \mathcal{G}_t(X; \beta)]\Xi(\alpha_t)^\top\}.
\end{aligned}$$

Theorem 4.1. *Assume the regularity conditions in Theorem 2.1 and the following regularity condition holds:*

C5. (a) The parameter space for β^0, $\mathcal{B}$, is a compact set and $\beta_I^* \in \mathcal{B}$ is the unique solution to $U_I(\beta, \alpha_t^*) = 0$; (b) $\sup_{\beta\in\mathcal{B}} \|\mathcal{G}_t(X;\beta)\| < \infty$; (c) uniformly for all $\alpha_t \in \mathcal{A}$ and $\beta \in \mathcal{B}$, the matrices $\mathcal{J}_2(\beta,\alpha_t)$ and $\mathcal{J}_1(\beta,\alpha_t) - \mathcal{J}_2(\beta,\alpha_t)$ are of full rank; (d) $V_I^* = \mathrm{Var}\{\sum_{t=1}^T \delta_t[\pi_t(\alpha_t^*)^{-1}\mathcal{D}_t(X,\beta_I^*)[Y_t - \mathcal{G}_t(X;\beta_I^*)] + \mathcal{J}_3^*\Lambda_t^* g_t(\alpha_t^*)]\}$ is nonsingular, where $\Lambda_t^* = [\Gamma_t^{*\top} W_t^{*-1}\Gamma_t]^{*-1}\Gamma_t^{*\top} W_t^{*-1}$, $\mathcal{J}_1^* = \mathcal{J}_1(\beta_I^*, \alpha_t^*)$, $\mathcal{J}_2^* = \mathcal{J}_2(\beta_I^*, \alpha_t^*)$, and $\mathcal{J}_3^* = \mathcal{J}_3(\beta_I^*, \alpha_t^*)$.

Then, as $n \to \infty$,

$$n^{1/2}(\hat\beta^{\mathrm{ipw}} - \beta_I^*) \xrightarrow{\mathcal{L}} N(0, \Sigma_I^*),$$

where $\Sigma_I^* = (\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1} V_I^* (\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1}$.

It follows from Theorem 4.1 that the proposed IPW estimator is a consistent and asymptotically normal estimator of the pseudo-true value β_I^*, which is the solution to the population version of the empirical IPW discrepancy. If both the propensity model and the marginal regression model are correctly specified, we have the following corollary.

Corollary 4.1. *Assume the conditions in Theorem 4.1 and that both (1) and (12) are correct. Then, as $n \to \infty$,*

$$n^{1/2}(\hat\beta^{\mathrm{ipw}} - \beta^0) \xrightarrow{\mathcal{L}} N(0, \Sigma_I),$$

where $V_I = \mathrm{Var}\{\sum_{t=1}^T [\delta_t \pi_t(\alpha_t^0)^{-1}\mathcal{D}_t(X,\beta^0)[Y_t - \mathcal{G}_t(X;\beta^0)] + \mathcal{J}_3\Lambda_t g_t(\alpha_t^0)]\}$, $\mathcal{J}_2 = \mathcal{J}_2(\beta^0, \alpha_t^0)$, $\mathcal{J}_3 = \mathcal{J}_3(\beta^0, \alpha_t^0)$, $\Sigma_I = \mathcal{J}_2^{-1} V_I \mathcal{J}_2^{-1}$.

To improve the efficiency and obtain a model-assisted estimator of β^0, we follow the idea in Section 3 and construct an AIPW estimator $\hat\beta^{\mathrm{aipw}}$ of β^0 by solving the following augmented estimating equations:

$$\begin{aligned} &\hat U_A(\beta, \hat\alpha_t, \hat\gamma_t) \\ (14) \quad &= \sum_{t=1}^T \mathbb{E}_n\{\mathcal{D}_t(X,\beta)[H_t(\hat\alpha_t, \hat\gamma_t) - \mathcal{G}_t(X;\beta)]\} = 0. \end{aligned}$$

Define

$$\begin{aligned} U_A(\beta, \alpha_t, \gamma_t) =& E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta)[H_t(\alpha_t,\gamma_t) - \mathcal{G}_t(X;\beta)]\Big\}, \\ \mathcal{I}_1(\beta, \alpha_t, \gamma_t) =& E\Big\{\sum_{t=1}^T \mathscr{D}_t(X,\beta_I^*)[H_t(\alpha_t,\gamma_t) - \mathcal{G}_t(X;\beta)]\Big\}, \\ \mathcal{I}_2(\beta, \alpha_t, \gamma_t) =& E\Big\{\sum_{t=1}^T \pi_t(\alpha_t^0)[\pi_t(\alpha_t)^{-1} - 1]\mathcal{D}_t(X,\beta) \\ & \times \{Y_t - m_{0t}(\alpha_t,\gamma_t)\}\Xi(\alpha_t)^\top\Big\}, \\ \mathcal{I}_3(\beta, \alpha_t, \gamma_t) =& E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta)[1 - \delta_t\pi_t(\alpha_t)^{-1}] \\ & \times \frac{\partial m_{0t}(\alpha_t,\gamma_t)}{\partial \alpha_t^\top}\Big\}, \\ \mathcal{I}_4(\beta, \alpha_t, \gamma_t) =& E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta)[1 - \delta_t\pi_t(\alpha_t)^{-1}] \\ & \times \frac{\partial m_{0t}(\alpha_t,\gamma_t)}{\partial \gamma_t^\top}\Big\}. \end{aligned}$$

We now establish the asymptotic normality of the proposed AIPW estimator $\hat\beta^{\mathrm{aipw}}$. The proof of the following theorem is in the Appendix.

Theorem 4.2. *Assume the regularity conditions in Theorem 2.1 and the following regularity condition:*

C6. (a) $\beta_A^* \in \mathcal{B}$ is the unique solution to $U_A(\beta, \alpha_t^*, \gamma_t^*) = 0$; (b) uniformly for all $\alpha_t \in \mathcal{A}$, $\beta \in \mathcal{B}$ and $\gamma_t \in \Upsilon_t$, the matrix $\mathcal{I}_1(\beta,\alpha_t,\gamma_t) - \mathcal{J}_2(\beta,\alpha_t)$ is of full rank; (c) the matrix $V_A^* = \mathrm{Var}\{\sum_{t=1}^T[\varphi_{1t}^* + \varphi_{2t}^* + \varphi_{3t}^*]\}$ is nonsingular, where $\varphi_{1t}^* = \mathcal{D}_t(X,\beta_A^*)[H_t(\alpha_t^*,\gamma_t^*) - \mathcal{G}_t(X;\beta_A^*)]$, $\varphi_{2t}^* = (\mathcal{I}_2^* - \mathcal{I}_3^*)\Lambda_t^* g_t(\alpha_t^*)$, $\varphi_{3t}^* = -\mathcal{I}_4^*\{\omega_0 A_t(\gamma_t^*)\}^{-1}\delta_t \partial \log f_t(S_t | \overrightarrow{S}_{t-1}, \gamma_t^*)/\partial\gamma_t$, $\mathcal{I}_1^* = \mathcal{I}_1(\beta_A^*, \alpha_t^*, \gamma_t^*)$, $\mathcal{I}_2^* = \mathcal{I}_2(\beta_A^*, \alpha_t^*, \gamma_t^*)$, $\mathcal{J}_2^* = \mathcal{J}_2(\beta_A^*, \alpha_t^0)$, $\mathcal{I}_3^* = \mathcal{I}_3(\beta_A^*, \alpha_t^*, \gamma_t^*)$, and $\mathcal{I}_4^* = \mathcal{I}_4(\beta_A^*, \alpha_t^*, \gamma_t^*)$.

Then, as $n \to \infty$,

$$n^{1/2}(\hat\beta^{\mathrm{aipw}} - \beta_A^*) \xrightarrow{\mathcal{L}} N(0, \Sigma_A^*),$$

where $\Sigma_A^* = (\mathcal{I}_1^* - \mathcal{J}_2^*)^{-1} V_A^* (\mathcal{I}_1^* - \mathcal{J}_2^*)^{-1}$.

In Theorem 4.2, large sample properties of the AIPW estimator $\hat\beta^{\mathrm{aipw}}$ are established for the case where models may be misspecified. The pseudo-true value β_A^* is the solution to the population version of the empirical AIPW discrepancy. If models are correctly specified, we have the following corollary.

Corollary 4.2. *Assume that the regularity conditions in Theorem 4.2 hold, and all the models used are correctly specified. Then, as $n \to \infty$,*

$$n^{1/2}(\hat\beta^{\mathrm{aipw}} - \beta^0) \xrightarrow{\mathcal{L}} N(0, \Sigma_A),$$

where $\Sigma_A = \mathcal{J}_2^{-1} V_A \mathcal{J}_2^{-1}$, $V_A = \mathrm{Var}\{\sum_{t=1}^T[\varphi_{1t} + \varphi_{2t}]\}$, $\varphi_{1t} = \mathcal{D}_t(X,\beta^0)[H_t(\alpha_t^0,\gamma_t^0) - \mathcal{G}_t(X;\beta^0)]$, $\varphi_{2t} = \mathcal{I}_2\Lambda_t g_t(\alpha_t^0)$, $\mathcal{I}_2 = \mathcal{I}_2(\beta^0, \alpha_t^0, \gamma_t^0)$.

In conclusion, as long as the propensity model (1) and the regression model (12) are correct, the AIPW estimator $\hat\beta^{\mathrm{aipw}}$ is consistent. It can improve the IPW estimator if the working model is nearly correct.

5. COMPUTATION

It is easy to apply some numerical optimization methods to compute the proposed two-step GMM estimator $\hat\alpha_t$ of α_t^0.

Once $\hat{\alpha}_t$ is obtained, the computation of the proposed IPW estimators $\hat{\mu}_t^{\text{ipw1}}$, $\hat{\mu}_t^{\text{ipw2}}$ and $\hat{\beta}^{\text{ipw}}$ is not a difficult task. For the AIPW estimators $\hat{\mu}_t^{\text{aipw}}$ and $\hat{\beta}^{\text{aipw}}$, we need compute the conditional expectation $m_{0t}(\hat{\alpha}_t, \hat{\gamma}_t)$ in (10), which involves intractable integration and can be computationally challenging. Here, we focus on the calculation of $\hat{\beta}^{\text{aipw}}$, whereas the discussion for $\hat{\mu}_t^{\text{aipw}}$ is similar.

Following Riddles et al. (2016), we employ the idea of importance sampling to approximate $\hat{m}_{0t}(\hat{\alpha}_t, \hat{\gamma}_t)$ for subject i by

$$\hat{m}_{0t}(\hat{\alpha}_t, \hat{\gamma}_t) = \sum_{j:\delta_{jt}=1} \mathcal{W}_{ijt}(\hat{\alpha}_t, \hat{\gamma}_t) Y_{jt}, \tag{15}$$

where

$$\mathcal{W}_{ijt}(\alpha_t, \gamma_t) = \frac{\mathcal{O}_t(\overrightarrow{S}_{i(t-1)}, S_{jt}, \hat{\alpha}_t) f_{ijt}(\hat{\gamma}_t)/\hat{f}_t(S_{jt})}{\sum_{k:\delta_{kt}=1} \mathcal{O}_t(\overrightarrow{S}_{i(t-1)}, S_{jt}, \hat{\alpha}_t) f_{ikt}(\hat{\gamma}_t)/\hat{f}_t(S_{kt})}, \tag{16}$$

$f_{ijt}(\hat{\gamma}_t) = f_t(S_{jt}|\overrightarrow{S}_{i(t-1)}, \hat{\gamma}_t)$, $\hat{f}_t(S_{jt}) = n_{1t}^{-1} \sum_{i:\delta_{it}=1} f_{ijt}(\hat{\gamma}_t)$ is a consistent estimator of the marginal density $f_t(S_t) = f(S_t|\delta_t = 1)$ using the respondents evaluated at $Y_t = Y_{jt}$ and $V_t = V_{jt}$, and $n_{1t} = \sum_{i=1}^n \delta_{it}$. Therefore, the computation of $\hat{\beta}^{\text{aipw}}$ has the following four steps:

(i) Applying a numerical optimization method to compute the two-step GMM estimator $\hat{\alpha}_t$ by numerically minimizing the criterion function given in (3) defined in Section 2.

(ii) Specify a working parametric model $f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t)$ for the conditional density $f(S_t|\overrightarrow{S}_{t-1}, \delta_t = 1)$. Using the completely observed data, obtain $\hat{\gamma}_t$ by maximizing $\prod_{\delta_{it}=1} f_t(S_{it}|\overrightarrow{S}_{i(t-1)}, \gamma_t)$ over γ_t.

(iii) Plugging $\hat{\alpha}_t$ and $\hat{\gamma}_t$ in (16) to compute the weight $\mathcal{W}_{ijt}(\hat{\alpha}_t, \hat{\gamma}_t)$ and then construct the augmented data

$$\hat{H}_{it}(\hat{\alpha}_t, \hat{\gamma}_t) = \frac{\delta_t Y_{it}}{\pi_{it}(\hat{\alpha}_t)} - \frac{\delta_t - \pi_{it}(\hat{\alpha}_t)}{\pi_{it}(\hat{\alpha}_t)} \sum_{j:\delta_{jt}=1} \mathcal{W}_{ijt}(\hat{\alpha}_t, \hat{\gamma}_t) Y_{jt},$$

for $i = 1, ..., n$ and $t = 1, ..., T$. Here $\pi_{it}(\hat{\alpha}_t) = \pi(\overrightarrow{S}_{i(t-1)}, S_{it}, \hat{\alpha}_t)$.

(iv) Compute $\hat{\beta}^{\text{aipw}}$ by solving the augmented equation

$$\sum_{t=1}^T \mathbb{E}_n\{\mathcal{D}_t(X, \beta)[\hat{H}_t(\hat{\alpha}_t, \hat{\gamma}_t) - \mathcal{G}_t(X; \beta)]\} = 0.$$

In Step (iii) introduced above, the conditional expectation $\hat{m}_{0t}(\hat{\alpha}_t, \hat{\gamma}_t)$ is estimated based on the weighted empirical distribution, and the weights $\mathcal{W}_{ijt}(\hat{\alpha}_t, \hat{\gamma}_t)$ in (16) can be viewed as the fractional weights assigned to the imputed values. Such a computational technique is very attractive because it does not introduce additional variability due to Monte Carlo approximation. See Kim (2011) for detailed discussion on parametric fractional imputation for missing data analysis. Our empirical studies in Section 6 suggest that the results based on approximation (15) are less sensitive to the choice of the parametric model for the conditional distribution for respondents.

6. SIMULATION STUDIES AND DATA ANALYSIS

In this section, we conduct two simulation studies and a real data analysis to examine the finite-sample performances of the proposed estimators of unconditional response means, μ_t, $t = 1, ..., T$, and regression parameter β. In simulations, we obtain the simulated absolute bias (AB) and standard deviation (SD) of estimators of μ_t and β, the standard error (SE) obtained by the bootstrap with replication size 100, and the coverage probability (CP) of the confidence intervals at the nominal level 95% based on asymptotic normality and bootstrap SE. All results are based on 1,000 simulation replications and the sample sizes $n = 200$ and 500.

6.1 Simulation 1: response means

We consider the case where there is no time-dependent covariate V ($S_t = Y_t$) and $X = Z$ ($U = 0$). For $t = 1, ..., 4$, we consider that Y_{it}'s, $i = 1, ..., n$, are independently generated from one linear regression model and one nonlinear regression model described as follows:

A1. $Y_{i1} = 1 + Y_{i0} + Z_i + \varepsilon_{i1}$, $Y_{i2} = Z_i + Y_{i0} + Y_{i1} + \varepsilon_{i2}$, $Y_{i3} = Z_i + Y_{i0} + Y_{i2} + \varepsilon_{i3}$, $Y_{i4} = Z_i + Y_{i0} + Y_{i3} + \varepsilon_{i4}$;

A2. $Y_{i1} = Y_{i0} + Z_i^2 + \varepsilon_{i1}$, $Y_{i2} = 1 + 2Y_{i0} + Z_i^2 + Z_i \exp(-Y_{i1}^2) + \varepsilon_{i2}$, $Y_{i3} = 2 + 3Y_{i0} + Z_i^2 + 2Y_{i1} \exp(-Y_{i2}^2/2) + \varepsilon_{i3}$, $Y_{i4} = 3 + 4Y_{i0} + Z_i^2 + 4Y_{i2} \exp(-Y_{i3}^2/4) + \varepsilon_{i4}$,

where $Y_{i0} \sim N(0, 1)$, $Z_i \sim N(1, 2)$, ε_{i1}, ε_{i2}, ε_{i3}, and ε_{i4} are independently generated from $N(0, 1)$, and Y_{i0}, Z_i and ε_{it}'s are independent. We also consider a multivariate standard normal vector $Z_i = (Z_{i1}, Z_{i2}, Z_{i3})^\top$ and replace Z_i in A1 by $(Z_{i1} + Z_{i2} + Z_{i3})/3$.

We generate δ_{it} independently from the Bernoulli distribution with π_t under assumption (1), and we consider four choices of π_t:

M1. $\pi_t = \{1 + \exp[\alpha_{0t} + \alpha_{1t} Y_0 + I(t > 1)(\alpha_{2t} \delta_{t-1} Y_{t-1})]\}^{-1}$;
M2. $\pi_t = \{1 + \exp\{\alpha_{0t} + \alpha_{1t} Y_0 + I(t > 1)(\alpha_{2t} \delta_{t-1} Y_{t-1}) + \alpha_{3t} Y_t]\}^{-1}$;
M3. $\pi_t = \{1 + \exp\{\alpha_{0t} + \alpha_{1t} Y_0 + I(t > 1)(\alpha_{2t} \delta_{t-1} Y_{t-1}^2) + \alpha_{3t} Y_t]\}^{-1}$;
M4. $\pi_t = 1 - \Phi\{\alpha_{0t} + \alpha_{1t} Y_0 + I(t > 1)(\alpha_{2t} \delta_{t-1} Y_{t-1}) + \alpha_{3t} Y_t\}$;

where $\alpha_{1t} = -0.6$, $\alpha_{2t} = -0.3$, $\alpha_{3t} = 0.1$ in scenarios A1, $\alpha_{1t} = -0.6$, $\alpha_{2t} = 0.1$, $\alpha_{3t} = -0.1$ in scenarios A2, and $\alpha_{01} = -1$, $\alpha_{02} = -0.6$, $\alpha_{03} = -0.4$, $\alpha_{04} = -0.2$, $\Phi(\cdot)$ is the standard normal distribution function. While M1 is an ignorable missing data case, i.e., the propensity does not depend on Y_t, M2–M4 represent three different nonignorable missing data cases. For $t = 1, ..., 4$, the unconditional means μ_t are 2, 3, 4, 5 in scenarios A1, and 3, 4.07, 5.08, 6.48 in scenario A2; for A1, the unconditional missing percentages for four time points are about 28.3%, 29.4%, 30.5%, 30.1% in scenario M1, 32.1%, 34.5%, 37.3%, 37.7% in scenario M2, 32.1%, 26.9%, 22.6%, 18.7% in scenario M3; 23.9%, 27.5%, 31.2%, 33.0% in scenario M4; for A2, the unconditional missing percentages for four time points are about 28.3%, 41.4%, 46.0%, 52.1% in scenario M1, 24.1%, 33.9%, 36.9%, 39.8% in scenario M2, 24.1%, 50.3%, 45.6%, 60.7% in scenario M3, 15.5%, 27.8%, 32.3%, 36.3% in scenario M4.

We study the performance of the following five estimators of μ_t: the proposed estimators $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$ defined in (4) and $\hat{\mu}_t^{\text{aipw}}$ defined in (6); $\hat{\mu}_t^{\text{cc}} = \sum_{i=1}^n \delta_{it} Y_{it} / \sum_{i=1}^n \delta_{it}$, the sample mean of the observed Y_{it}'s; and $\bar{Y}_t = \sum_{i=1}^n Y_{it}/n$, the sample mean when there is no missing data, which is used as a standard.

For the choice of π_t in assumption (1), we do not assume that we know exactly the form of the propensity. Instead, we use the working propensity model

$$\pi_t(\alpha_t) = \{1 + \exp[\alpha_{0t} + \alpha_{1t} Y_0 + I(t > 1)(\alpha_{2t}\delta_{t-1}Y_{t-1}) + \gamma_t Y_t]\}^{-1}, \tag{17}$$

for all scenarios M1–M4. Under M1 and M2, the working model (17) is correct. Under M3 and M4, however, the working model (17) is misspecified so that we can see the robustness of the proposed estimators $\hat{\mu}_t^{\text{ipw1}}$, $\hat{\mu}_t^{\text{ipw2}}$ and $\hat{\mu}_t^{\text{aipw}}$ against propensity model misspecification.

Furthermore, we use the working model $f_t(Y_t | \overrightarrow{S}_{t-1}, \gamma_t) \sim N(\overrightarrow{S}_{t-1}^{\top}\gamma_t^{(1)}, \gamma_t^{(2)})$, with unknown parameter vector $\gamma_t^{\top} = ((\gamma_t^{(1)})^{\top}, \gamma_t^{(2)})$. Note that this working model is always incorrect.

Simulation results are presented in Tables 1-3, for model A1, model A1 with multivariate Z, and model A2, respectively. A few conclusions can be drawn from the simulation results.

(i) Bias. The proposed estimators, $\hat{\mu}_t^{\text{ipw1}}$, $\hat{\mu}_t^{\text{ipw2}}$ and $\hat{\mu}_t^{\text{aipw}}$, have negligible biases in most of the cases. Among these three estimators, $\hat{\mu}_t^{\text{aipw}}$ perform better than $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$. On the other hand, $\hat{\mu}_t^{\text{cc}}$ is biased.

(ii) Standard deviation. $\hat{\mu}_t^{\text{aipw}}$ performs better than $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$ and the improvement is substantial when $t = 4$. The SDs of $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$ are comparable, smaller than the SD of $\hat{\mu}_{cc}$, and become smaller when the mean response rate or the sample size is larger.

(iii) Standard error. The bootstrap variance estimator works well under all cases.

(iv) Coverage probability. When the working model (17) is correct, the coverage probabilities based on $\hat{\mu}_t^{\text{ipw1}}$, $\hat{\mu}_t^{\text{ipw2}}$ and $\hat{\mu}_t^{\text{aipw}}$ are all close to the nominal level 0.95, and are quite comparable with the method based on $\bar{Y}_t$ assuming no missing data. The main price paid for missing data is the increased standard deviation so that the confidence intervals are longer due to non-ignorable nonmonotone missing. When the working model (17) is incorrect, coverage probabilities based on proposed methods are still acceptable in most cases. The confidence interval based on $\hat{\mu}_t^{\text{cc}}$ does not perform well in most of cases, because of the bias in $\hat{\mu}_t^{\text{cc}}$.

Table 1. Absolute bias (AB), standard deviation (SD), standard error (SE) and coverage probability (CP) values for simulation 1 with model A1

Method		$n=200$				$n=500$			
		$t=1$	$t=2$	$t=3$	$t=4$	$t=1$	$t=2$	$t=3$	$t=4$
		scenario M1							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.018	0.017	0.024	0.009	0.003	0.001	0.010	0.001
	SD	0.153	0.279	0.402	0.525	0.098	0.182	0.264	0.365
	SE	0.157	0.290	0.414	0.539	0.101	0.181	0.261	0.351
	CP	0.936	0.945	0.942	0.923	0.954	0.948	0.939	0.948
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.006	0.004	0.030	0.129	0.001	0.007	0.011	0.036
	SD	0.152	0.279	0.406	0.552	0.098	0.182	0.266	0.382
	SE	0.156	0.295	0.436	0.578	0.101	0.182	0.266	0.388
	CP	0.937	0.948	0.953	0.946	0.956	0.951	0.938	0.958
$\hat{\mu}_t^{\text{aipw}}$	AB	0.004	0.004	0.012	0.016	0.002	0.006	0.009	0.013
	SD	0.151	0.276	0.401	0.522	0.098	0.181	0.263	0.350
	SE	0.154	0.281	0.403	0.534	0.100	0.178	0.256	0.339
	CP	0.935	0.940	0.943	0.939	0.950	0.948	0.952	0.953
$\bar{Y}_t$	AB	0.001	0.002	0.001	0.010	0.001	0.004	0.005	0.007
	SD	0.139	0.261	0.384	0.507	0.092	0.173	0.256	0.338
	SE	0.140	0.262	0.383	0.500	0.089	0.166	0.243	0.320
	CP	0.937	0.934	0.935	0.925	0.945	0.944	0.947	0.947
$\hat{\mu}_t^{\text{cc}}$	AB	0.151	0.664	1.341	2.194	0.158	0.672	1.349	2.219
	SD	0.164	0.300	0.433	0.536	0.106	0.194	0.284	0.361
	SE	0.164	0.301	0.433	0.545	0.103	0.187	0.269	0.339
	CP	0.832	0.416	0.142	0.025	0.148	0.003	0	0
		scenario M2							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.021	0.037	0.087	0.150	0.007	0.019	0.040	0.111
	SD	0.161	0.304	0.445	0.607	0.101	0.177	0.261	0.377
	SE	0.165	0.303	0.451	0.612	0.100	0.181	0.276	0.393
	CP	0.955	0.955	0.949	0.924	0.949	0.940	0.946	0.949
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.008	0.012	0.022	0.003	0.002	0.010	0.020	0.053
	SD	0.160	0.300	0.436	0.622	0.100	0.176	0.258	0.367
	SE	0.161	0.297	0.448	0.646	0.098	0.179	0.268	0.386
	CP	0.948	0.954	0.948	0.946	0.948	0.941	0.951	0.951
$\hat{\mu}_t^{\text{aipw}}$	AB	0.002	0.001	0.001	0.001	0.001	0.005	0.001	0.014
	SD	0.158	0.295	0.421	0.564	0.099	0.176	0.251	0.339
	SE	0.158	0.284	0.416	0.569	0.098	0.176	0.257	0.340
	CP	0.953	0.943	0.940	0.926	0.949	0.940	0.940	0.943
$\bar{Y}_t$	AB	0.001	0.007	0.016	0.020	0.001	0.002	0.003	0.006
	SD	0.147	0.271	0.397	0.516	0.091	0.167	0.242	0.318
	SE	0.141	0.263	0.384	0.497	0.089	0.166	0.243	0.320
	CP	0.940	0.943	0.940	0.917	0.942	0.943	0.943	0.940
$\hat{\mu}_t^{\text{cc}}$	AB	0.062	0.373	0.808	1.582	0.060	0.364	0.816	1.540
	SD	0.178	0.334	0.493	0.646	0.108	0.207	0.302	0.395
	SE	0.170	0.323	0.481	0.608	0.108	0.204	0.303	0.393
	CP	0.936	0.785	0.585	0.299	0.896	0.557	0.274	0.026

Table 1. Continued

Method		n = 200 t = 1	t = 2	t = 3	t = 4	n = 500 t = 1	t = 2	t = 3	t = 4
		scenario M3							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.029	0.047	0.076	0.110	0.008	0.017	0.040	0.036
	SD	0.156	0.280	0.410	0.534	0.101	0.179	0.266	0.354
	SE	0.164	0.286	0.409	0.538	0.099	0.177	0.264	0.352
	CP	0.941	0.927	0.921	0.905	0.949	0.943	0.932	0.936
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.017	0.003	0.044	0.093	0.003	0.008	0.043	0.101
	SD	0.155	0.283	0.420	0.547	0.101	0.178	0.265	0.353
	SE	0.159	0.292	0.428	0.545	0.098	0.177	0.269	0.357
	CP	0.940	0.929	0.937	0.882	0.954	0.947	0.940	0.935
$\hat{\mu}_t^{\text{aipw}}$	AB	0.012	0.018	0.021	0.040	0.002	0.002	0.005	0.008
	SD	0.154	0.277	0.396	0.518	0.100	0.176	0.252	0.335
	SE	0.156	0.285	0.425	0.526	0.097	0.177	0.269	0.341
	CP	0.937	0.929	0.929	0.904	0.947	0.949	0.956	0.953
$\bar{Y}_t$	AB	0.006	0.018	0.026	0.040	0.001	0.002	0.001	0.004
	SD	0.140	0.260	0.382	0.504	0.091	0.171	0.247	0.326
	SE	0.139	0.259	0.374	0.490	0.089	0.165	0.242	0.313
	CP	0.928	0.923	0.909	0.891	0.941	0.944	0.938	0.918
$\hat{\mu}_t^{\text{cc}}$	AB	0.052	0.376	0.541	0.575	0.057	0.395	0.573	0.605
	SD	0.163	0.311	0.448	0.577	0.108	0.196	0.285	0.368
	SE	0.169	0.310	0.441	0.563	0.107	0.197	0.284	0.361
	CP	0.928	0.740	0.714	0.776	0.911	0.483	0.483	0.595
		scenario M4							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.019	0.041	0.119	0.130	0.007	0.022	0.070	0.117
	SD	0.153	0.297	0.461	0.547	0.095	0.179	0.286	0.471
	SE	0.161	0.308	0.462	0.564	0.098	0.189	0.304	0.470
	CP	0.954	0.954	0.929	0.881	0.947	0.949	0.952	0.936
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.005	0.012	0.017	0.145	0.002	0.010	0.030	0.027
	SD	0.152	0.291	0.458	0.558	0.095	0.175	0.272	0.498
	SE	0.157	0.307	0.482	0.573	0.096	0.185	0.297	0.520
	CP	0.948	0.959	0.943	0.890	0.953	0.947	0.956	0.936
$\hat{\mu}_t^{\text{aipw}}$	AB	0.002	0.006	0.047	0.070	0.001	0.006	0.030	0.053
	SD	0.151	0.280	0.442	0.534	0.095	0.172	0.257	0.411
	SE	0.158	0.296	0.443	0.550	0.097	0.181	0.276	0.421
	CP	0.952	0.950	0.932	0.904	0.954	0.944	0.953	0.933
$\bar{Y}_t$	AB	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.002
	SD	0.137	0.259	0.382	0.510	0.087	0.164	0.244	0.324
	SE	0.140	0.261	0.374	0.470	0.089	0.166	0.243	0.313
	CP	0.944	0.949	0.926	0.870	0.942	0.947	0.937	0.915
$\hat{\mu}_t^{\text{cc}}$	AB	0.070	0.466	1.021	1.763	0.077	0.478	1.033	1.797
	SD	0.158	0.305	0.467	0.568	0.100	0.192	0.010	0.375
	SE	0.160	0.300	0.434	0.547	0.102	0.192	0.284	0.366
	CP	0.923	0.666	0.368	0.160	0.878	0.312	0.060	0.004

Table 2. Absolute bias (AB), standard deviation (SD), standard error (SE) and coverage probability (CP) values for simulation 1 with model A1 and multivariate Z

Method		n = 200 t = 1	t = 2	t = 3	t = 4	n = 500 t = 1	t = 2	t = 3	t = 4
		scenario M1							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.042	0.064	0.094	0.112	0.023	0.031	0.048	0.065
	SD	0.148	0.223	0.302	0.379	0.095	0.139	0.196	0.240
	SE	0.164	0.244	0.321	0.400	0.100	0.145	0.195	0.246
	CP	0.947	0.948	0.935	0.939	0.943	0.953	0.941	0.943
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.013	0.008	0.001	0.009	0.003	0.004	0.001	0.001
	SD	0.147	0.223	0.307	0.385	0.094	0.139	0.196	0.243
	SE	0.163	0.243	0.326	0.415	0.098	0.143	0.194	0.246
	CP	0.959	0.961	0.953	0.962	0.948	0.952	0.939	0.953
$\hat{\mu}_t^{\text{aipw}}$	AB	0.008	0.010	0.013	0.021	0.001	0.004	0.005	0.007
	SD	0.147	0.226	0.300	0.375	0.094	0.138	0.195	0.239
	SE	0.155	0.229	0.305	0.384	0.095	0.139	0.188	0.238
	CP	0.953	0.954	0.947	0.954	0.944	0.952	0.939	0.943
$\bar{Y}_t$	AB	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	SD	0.108	0.189	0.273	0.352	0.069	0.122	0.175	0.227
	SE	0.108	0.190	0.271	0.352	0.068	0.120	0.172	0.224
	CP	0.945	0.947	0.945	0.948	0.942	0.947	0.947	0.946
$\hat{\mu}_t^{\text{cc}}$	AB	0.031	0.071	0.307	0.654	0.033	0.078	0.311	0.652
	SD	0.130	0.228	0.331	0.414	0.085	0.148	0.216	0.280
	SE	0.130	0.227	0.324	0.425	0.082	0.143	0.206	0.270
	CP	0.935	0.928	0.829	0.669	0.917	0.901	0.674	0.340
		scenario M2							
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.041	0.067	0.090	0.128	0.018	0.024	0.044	0.067
	SD	0.151	0.244	0.321	0.409	0.094	0.149	0.193	0.251
	SE	0.165	0.260	0.344	0.431	0.103	0.157	0.209	0.268
	CP	0.954	0.945	0.946	0.929	0.962	0.962	0.954	0.951
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.004	0.010	0.022	0.036	0.003	0.009	0.007	0.008
	SD	0.155	0.243	0.321	0.429	0.095	0.148	0.194	0.257
	SE	0.169	0.271	0.368	0.480	0.102	0.158	0.213	0.277
	CP	0.965	0.960	0.962	0.955	0.965	0.962	0.962	0.956
$\hat{\mu}_t^{\text{aipw}}$	AB	0.012	0.018	0.030	0.048	0.001	0.001	0.007	0.010
	SD	0.156	0.237	0.337	0.398	0.096	0.146	0.192	0.244
	SE	0.161	0.247	0.328	0.410	0.100	0.151	0.200	0.253
	CP	0.956	0.955	0.943	0.936	0.962	0.956	0.953	0.949
$\bar{Y}_t$	AB	0.001	0.001	0.002	0.004	0.005	0.005	0.007	0.008
	SD	0.109	0.191	0.271	0.356	0.067	0.120	0.171	0.221
	SE	0.107	0.189	0.269	0.349	0.068	0.121	0.173	0.225
	CP	0.945	0.938	0.939	0.930	0.950	0.946	0.948	0.952
$\hat{\mu}_t^{\text{cc}}$	AB	0.111	0.386	0.755	1.273	0.115	0.394	0.752	1.271
	SD	0.129	0.232	0.334	0.432	0.081	0.148	0.213	0.266
	SE	0.130	0.229	0.333	0.422	0.082	0.147	0.214	0.270
	CP	0.864	0.607	0.373	0.170	0.721	0.239	0.062	0.001

6.2 Simulation 2: marginal regression

We consider that

$$Y_{it} = \beta_1 t + \beta_2 Y_{i0} + \beta_3 (tZ_i) + \varepsilon_{it}, \quad t = 1, 2, 3, 4,$$

where $Y_{i0} \sim N(1,1)$, $Z_i \sim N(1,1)$, ε_{it}'s are independently generated from $N(0,1)$, Z_i, Y_{i0} and ε_{it}'s are independent. The true parameter vector $\beta = (\beta_1, \beta_3, \beta_3)^\top = (0.1, 0.1, 0.1)^\top$.

We generate δ_i from the Bernoulli distribution with four choices of π_t, i.e., M1–M4, defined in Section 5.1. We consider that $\alpha_{01} = -0.8$, $\alpha_{02} = -0.6$, $\alpha_{03} = -0.4$, $\alpha_{04} = -0.2$, $\alpha_{1t} = -0.5$, $\alpha_{2t} = 0.2$, $\alpha_{3t} = -0.5$ and the unconditional missing percentages for four time points are about 22.6%, 25.4%, 31.5%, 36.2% in scenario M1, 21.6%, 24.9%, 28.0%, 31.3% in scenario M2, 17.9%, 22.7%, 25.4%, 28.5% in scenario M3; 11.8%, 15.7%, 19.5%, 23.7% in scenario M4.

Table 2. Continued

Method		$n=200$				$n=500$			
		$t=1$	$t=2$	$t=3$	$t=4$	$t=1$	$t=2$	$t=3$	$t=4$
scenario M3									
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.020	0.024	0.066	0.121	0.003	0.007	0.033	0.064
	SD	0.168	0.234	0.310	0.397	0.096	0.142	0.200	0.257
	SE	0.170	0.242	0.316	0.393	0.099	0.144	0.200	0.264
	CP	0.953	0.953	0.937	0.911	0.955	0.949	0.937	0.952
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.001	0.011	0.007	0.016	0.006	0.006	0.008	0.006
	SD	0.167	0.235	0.314	0.409	0.096	0.143	0.201	0.254
	SE	0.170	0.246	0.332	0.422	0.099	0.145	0.201	0.270
	CP	0.950	0.957	0.952	0.936	0.952	0.944	0.939	0.957
$\hat{\mu}_t^{\text{aipw}}$	AB	0.003	0.003	0.021	0.054	0.004	0.003	0.001	0.011
	SD	0.169	0.237	0.310	0.400	0.096	0.142	0.198	0.246
	SE	0.168	0.243	0.322	0.406	0.099	0.144	0.200	0.263
	CP	0.950	0.953	0.946	0.919	0.955	0.952	0.945	0.961
$\bar{Y}_t$	AB	0.002	0.002	0.004	0.008	0.001	0.001	0.001	0.001
	SD	0.111	0.193	0.277	0.356	0.069	0.122	0.175	0.226
	SE	0.107	0.189	0.270	0.347	0.068	0.120	0.172	0.224
	CP	0.932	0.936	0.939	0.922	0.940	0.942	0.953	0.948
$\hat{\mu}_t^{\text{cc}}$	AB	0.109	0.445	0.721	0.871	0.111	0.451	0.724	0.879
	SD	0.131	0.224	0.311	0.399	0.082	0.143	0.198	0.253
	SE	0.129	0.221	0.307	0.386	0.082	0.140	0.197	0.250
	CP	0.854	0.466	0.355	0.400	0.723	0.119	0.046	0.067
scenario M4									
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.014	0.024	0.059	0.106	0.006	0.006	0.044	0.079
	SD	0.151	0.226	0.302	0.389	0.087	0.144	0.204	0.269
	SE	0.155	0.233	0.309	0.380	0.095	0.146	0.208	0.269
	CP	0.958	0.939	0.922	0.906	0.962	0.951	0.952	0.930
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.008	0.016	0.026	0.054	0.003	0.007	-0.01	0.005
	SD	0.151	0.231	0.313	0.405	0.086	0.145	0.203	0.266
	SE	0.155	0.242	0.335	0.419	0.094	0.148	0.214	0.288
	CP	0.959	0.944	0.949	0.920	0.961	0.951	0.960	0.953
$\hat{\mu}_t^{\text{aipw}}$	AB	0.001	0.006	0.025	0.053	0.001	0.004	0.007	0.028
	SD	0.152	0.234	0.307	0.397	0.087	0.143	0.195	0.259
	SE	0.157	0.240	0.323	0.403	0.095	0.150	0.212	0.273
	CP	0.960	0.943	0.938	0.915	0.964	0.959	0.964	0.944
$\bar{Y}_t$	AB	0.003	0.001	0.001	0.004	0.001	0.002	0.002	0.002
	SD	0.109	0.191	0.273	0.355	0.068	0.122	0.174	0.224
	SE	0.107	0.188	0.267	0.342	0.068	0.121	0.172	0.221
	CP	0.943	0.933	0.924	0.898	0.949	0.946	0.941	0.935
$\hat{\mu}_t^{\text{cc}}$	AB	0.133	0.453	0.707	0.873	0.135	0.450	0.705	0.879
	SD	0.123	0.211	0.301	0.392	0.077	0.134	0.189	0.242
	SE	0.121	0.207	0.292	0.374	0.077	0.133	0.189	0.241
	CP	0.790	0.419	0.341	0.374	0.585	0.083	0.040	0.058

We use the same working propensity model and $f_t(Y_t|\overrightarrow{S}_{t-1},\gamma_t)$ as in Section 5.1, and study the performance of the following four estimators of β: the proposed estimators $\hat{\beta}^{\text{ipw}}$ and $\hat{\beta}^{\text{aipw}}$ defined in (13) and (14), respectively; the least square estimator $\hat{\beta}$ when there is no missing data, which is defined as the root of the following equations $\sum_{t=1}^{T}\mathbb{E}_n\{\mathcal{D}_t(X,\beta)[Y_t-\mathcal{G}_t(X;\beta)]\}=0$; the least square estimator $\hat{\beta}_{cc}$ only using the observed data, which is defined as the root of the following equations $\sum_{t=1}^{T}\mathbb{E}_n\{\delta_t\mathcal{D}_t(X,\beta)[Y_t-\mathcal{G}_t(X;\beta)]\}=0$.

Table 3. Absolute bias (AB), standard deviation (SD), standard error (SE) and coverage probability (CP) values for simulation 1 with model A2

Method		$n=200$				$n=500$			
		$t=1$	$t=2$	$t=3$	$t=4$	$t=1$	$t=2$	$t=3$	$t=4$
scenario M1									
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.022	0.059	0.131	0.292	0.012	0.026	0.053	0.122
	SD	0.305	0.343	0.421	0.638	0.198	0.216	0.250	0.344
	SE	0.307	0.386	0.469	0.734	0.192	0.225	0.262	0.404
	CP	0.933	0.954	0.948	0.940	0.935	0.947	0.947	0.959
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.005	0.034	0.074	0.171	0.006	0.016	0.030	0.077
	SD	0.306	0.334	0.405	0.679	0.199	0.215	0.247	0.373
	SE	0.307	0.366	0.448	0.706	0.192	0.219	0.254	0.382
	CP	0.934	0.947	0.949	0.951	0.936	0.944	0.945	0.959
$\hat{\mu}_t^{\text{aipw}}$	AB	0.010	0.034	0.055	0.128	0.007	0.012	0.028	0.060
	SD	0.311	0.337	0.429	0.636	0.200	0.222	0.268	0.329
	SE	0.308	0.350	0.428	0.644	0.193	0.217	0.261	0.356
	CP	0.930	0.944	0.941	0.945	0.933	0.943	0.939	0.946
$\bar{Y}_t$	AB	0.005	0.006	0.008	0.002	0.001	0.001	0.003	0.007
	SD	0.300	0.319	0.371	0.420	0.196	0.207	0.237	0.272
	SE	0.297	0.317	0.366	0.419	0.189	0.202	0.233	0.266
	CP	0.932	0.942	0.939	0.931	0.941	0.934	0.943	0.933
$\hat{\mu}_t^{\text{cc}}$	AB	0.161	0.039	0.360	0.487	0.152	0.062	0.330	0.452
	SD	0.357	0.362	0.476	0.577	0.230	0.246	0.312	0.363
	SE	0.348	0.369	0.474	0.559	0.222	0.235	0.301	0.358
	CP	0.934	0.937	0.882	0.850	0.899	0.917	0.792	0.752
scenario M2									
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.014	0.017	0.065	0.137	0.002	0.002	0.017	0.049
	SD	0.307	0.337	0.412	0.542	0.193	0.206	0.243	0.327
	SE	0.303	0.333	0.429	0.564	0.190	0.207	0.249	0.328
	CP	0.937	0.935	0.951	0.946	0.942	0.952	0.949	0.952
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.001	0.006	0.015	0.043	0.005	0.007	0.001	0.019
	SD	0.309	0.341	0.411	0.549	0.193	0.206	0.245	0.332
	SE	0.305	0.339	0.432	0.574	0.191	0.208	0.248	0.333
	CP	0.938	0.943	0.956	0.954	0.942	0.953	0.947	0.955
$\hat{\mu}_t^{\text{aipw}}$	AB	0.008	0.001	0.015	0.026	0.001	0.005	0.006	0.008
	SD	0.309	0.338	0.413	0.506	0.193	0.207	0.248	0.306
	SE	0.304	0.332	0.404	0.511	0.191	0.208	0.246	0.311
	CP	0.934	0.936	0.944	0.947	0.942	0.952	0.941	0.948
$\bar{Y}_t$	AB	0.005	0.001	0.005	0.002	0.004	0.006	0.013	0.006
	SD	0.303	0.329	0.381	0.445	0.190	0.201	0.231	0.273
	SE	0.298	0.318	0.367	0.424	0.187	0.201	0.230	0.265
	CP	0.932	0.923	0.936	0.933	0.942	0.953	0.947	0.937
$\hat{\mu}_t^{\text{cc}}$	AB	0.555	0.807	0.877	1.389	0.563	0.796	0.889	1.381
	SD	0.356	0.402	0.464	0.518	0.223	0.243	0.276	0.305
	SE	0.352	0.383	0.441	0.497	0.222	0.243	0.280	0.313
	CP	0.686	0.454	0.495	0.214	0.267	0.088	0.104	0.007

Simulation results are presented in Table 4. It can be seen that the proposed estimators $\hat{\beta}^{\text{ipw}}$ and $\hat{\beta}^{\text{aipw}}$ have negligible biases in all cases, but the biases of $\hat{\beta}^{\text{cc}}$ for β_1 and β_2 are large in M2–M4. In terms of coverage probabilities, the two strong competitors are $\hat{\beta}^{\text{ipw}}$ and $\hat{\beta}^{\text{aipw}}$, and the CPs of $\hat{\beta}^{\text{cc}}$ for β_1 and β_2 do not perform well in M2–M4, especially when $n=500$.

Table 3. Continued

Method		n = 200 t = 1	t = 2	t = 3	t = 4	n = 500 t = 1	t = 2	t = 3	t = 4
					scenario M3				
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.015	0.020	0.068	0.148	0.003	0.015	0.064	0.232
	SD	0.305	0.330	0.382	0.541	0.189	0.206	0.244	0.328
	SE	0.300	0.325	0.385	0.552	0.192	0.208	0.246	0.325
	CP	0.944	0.928	0.931	0.900	0.950	0.952	0.951	0.901
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.002	0.003	0.010	0.024	0.003	0.005	0.049	0.206
	SD	0.306	0.336	0.392	0.560	0.189	0.207	0.246	0.347
	SE	0.303	0.334	0.414	0.556	0.193	0.211	0.254	0.363
	CP	0.945	0.935	0.942	0.927	0.949	0.956	0.955	0.926
$\hat{\mu}_t^{\text{aipw}}$	AB	0.008	0.018	0.053	0.115	0.001	0.014	0.048	0.174
	SD	0.306	0.331	0.382	0.501	0.190	0.207	0.244	0.312
	SE	0.302	0.327	0.390	0.505	0.193	0.209	0.246	0.321
	CP	0.946	0.934	0.934	0.912	0.949	0.950	0.945	0.928
$\bar{Y}_t$	AB	0.002	0.002	0.004	0.006	0.001	0.003	0.002	0.006
	SD	0.301	0.322	0.363	0.425	0.186	0.201	0.229	0.262
	SE	0.296	0.316	0.358	0.408	0.190	0.203	0.232	0.265
	CP	0.941	0.931	0.931	0.905	0.952	0.960	0.952	0.952
$\hat{\mu}_t^{\text{cc}}$	AB	0.557	0.856	1.178	1.366	0.565	0.862	1.178	1.368
	SD	0.364	0.368	0.397	0.424	0.220	0.231	0.254	0.265
	SE	0.349	0.356	0.389	0.404	0.225	0.229	0.252	0.264
	CP	0.656	0.323	0.143	0.088	0.272	0.029	0	0.001
					scenario M4				
$\hat{\mu}_t^{\text{ipw1}}$	AB	0.022	0.013	0.083	0.088	0.012	0.007	0.035	0.058
	SD	0.291	0.321	0.413	0.572	0.190	0.208	0.260	0.357
	SE	0.299	0.318	0.441	0.582	0.191	0.207	0.269	0.363
	CP	0.949	0.914	0.937	0.938	0.942	0.929	0.953	0.947
$\hat{\mu}_t^{\text{ipw2}}$	AB	0.005	0.011	0.026	0.017	0.006	0.001	0.020	0.033
	SD	0.293	0.326	0.417	0.616	0.190	0.210	0.266	0.394
	SE	0.305	0.331	0.455	0.600	0.193	0.211	0.276	0.381
	CP	0.952	0.928	0.947	0.939	0.946	0.937	0.955	0.945
$\hat{\mu}_t^{\text{aipw}}$	AB	0.021	0.005	0.025	0.014	0.012	0.003	0.009	0.006
	SD	0.292	0.322	0.398	0.521	0.190	0.208	0.257	0.321
	SE	0.301	0.320	0.412	0.525	0.192	0.206	0.255	0.334
	CP	0.950	0.923	0.936	0.950	0.945	0.929	0.947	0.955
$\bar{Y}_t$	AB	0.016	0.015	0.016	0.014	0.009	0.006	0.005	0.012
	SD	0.287	0.308	0.361	0.418	0.189	0.202	0.230	0.266
	SE	0.294	0.305	0.362	0.416	0.187	0.197	0.229	0.265
	CP	0.949	0.915	0.933	0.933	0.942	0.928	0.947	0.952
$\hat{\mu}_t^{\text{cc}}$	AB	0.462	0.830	1.093	1.750	0.468	0.837	1.107	1.755
	SD	0.312	0.354	0.415	0.445	0.209	0.228	0.267	0.278
	SE	0.324	0.345	0.414	0.445	0.206	0.222	0.262	0.286
	CP	0.732	0.352	0.234	0.025	0.380	0.032	0.007	0

6.3 ACTG 193A data analysis

For illustration, we apply the proposed estimators to a longitudinal data from the AIDS Clinical Trial Group (ACTG) 193A (Henry et al., 1998), which was a study of HIV-AIDS patients with advanced immune suppression. In this study, patients were randomized to one of the four daily regimens containing 600mg of zidovudine considered as four treatments: (I) zidovudine alternating monthly with 400mg didanosine, (II) zidovudine plus 2.25mg of zalcitabine, (III) zidovudine plus 400mg of didanosine, and (IV) zidovudine plus 400mg of didanosine plus 400mg of nevirapine.

Table 4. Absolute bias (AB), standard deviation (SD), standard error (SE) and coverage probability (CP) values for simulation 2

Method		n = 200 β1	β2	β3	n = 500 β1	β2	β3
				scenario M1			
$\hat{\beta}^{\text{ipw}}$	AB	0.004	0.001	0.003	0.002	0.001	0.002
	SD	0.047	0.052	0.018	0.030	0.034	0.010
	SE	0.047	0.052	0.018	0.032	0.035	0.011
	CP	0.952	0.959	0.933	0.950	0.950	0.951
$\hat{\beta}^{\text{aipw}}$	AB	0.004	0.001	0.003	0.002	0.001	0.002
	SD	0.048	0.052	0.018	0.030	0.034	0.010
	SE	0.046	0.052	0.019	0.031	0.034	0.011
	CP	0.943	0.961	0.945	0.947	0.952	0.959
$\hat{\beta}$	AB	0.001	0.000	0.000	0.000	0.000	0.001
	SD	0.022	0.032	0.013	0.013	0.021	0.008
	SE	0.021	0.033	0.013	0.013	0.021	0.008
	CP	0.938	0.944	0.931	0.948	0.954	0.952
$\hat{\beta}^{\text{cc}}$	AB	0.001	0.001	0.001	0.002	0.001	0.001
	SD	0.029	0.041	0.017	0.018	0.026	0.010
	SE	0.028	0.040	0.017	0.018	0.025	0.010
	CP	0.947	0.945	0.911	0.943	0.954	0.933
				scenario M2			
$\hat{\beta}^{\text{ipw}}$	AB	0.004	0.004	0.001	0.001	0.000	0.001
	SD	0.045	0.049	0.017	0.027	0.032	0.011
	SE	0.044	0.049	0.017	0.029	0.033	0.011
	CP	0.938	0.942	0.944	0.952	0.958	0.955
$\hat{\beta}^{\text{aipw}}$	AB	0.004	0.004	0.002	0.000	0.000	0.001
	SD	0.046	0.048	0.017	0.027	0.033	0.010
	SE	0.043	0.049	0.018	0.029	0.033	0.011
	CP	0.930	0.939	0.949	0.950	0.959	0.961
$\hat{\beta}$	AB	0.001	0.000	0.001	0.000	0.001	0.000
	SD	0.021	0.033	0.013	0.014	0.021	0.008
	SE	0.021	0.034	0.013	0.014	0.021	0.008
	CP	0.937	0.935	0.933	0.956	0.946	0.946
$\hat{\beta}^{\text{cc}}$	AB	0.033	0.015	0.002	0.032	0.013	0.001
	SD	0.028	0.039	0.016	0.018	0.025	0.010
	SE	0.027	0.039	0.016	0.017	0.025	0.010
	CP	0.766	0.921	0.938	0.563	0.924	0.945

For the HIV study, the CD4 cell count is of prime interest which decreases as HIV progresses. The CD4 counts were collected from 1,280 patients before the treatments were applied (baseline measurements) and we use their records in the analysis. After the treatments were applied, the CD4 count was scheduled to be collected from each patient every 8 weeks. We consider the first four follow-up times, $8, 16, 24, 32$, as four time points $t = 1, 2, 3, 4$, and use the CD4 counts in four time intervals, $(4, 12]$, $(12, 20]$, $(20, 28]$, $(28, 36]$, as the study variable Y_t for $t = 1, 2, 3, 4$, because the realized follow-up time points might be a little different from the scheduled time points. A few patients had more than one measurement in one time interval, in which case

Table 4. Continued

Method		$n=200$ $\beta1$	$\beta2$	$\beta3$	$n=500$ $\beta1$	$\beta2$	$\beta3$
		scenario M3					
$\hat\beta^{\text{ipw}}$	AB	0.002	0.006	0.004	0.001	0.001	0.002
	SD	0.043	0.047	0.016	0.028	0.029	0.010
	SE	0.042	0.045	0.017	0.029	0.030	0.010
	CP	0.954	0.943	0.943	0.958	0.964	0.947
$\hat\beta^{\text{aipw}}$	AB	0.001	0.003	0.004	0.005	0.009	0.002
	SD	0.046	0.050	0.016	0.028	0.032	0.010
	SE	0.043	0.049	0.017	0.029	0.033	0.011
	CP	0.946	0.948	0.954	0.955	0.951	0.951
$\hat\beta$	AB	0.001	0.000	0.000	0.001	0.001	0.000
	SD	0.022	0.034	0.013	0.014	0.021	0.008
	SE	0.021	0.033	0.013	0.014	0.021	0.008
	CP	0.951	0.946	0.948	0.948	0.949	0.939
$\hat\beta^{\text{cc}}$	AB	0.030	0.018	0.001	0.029	0.018	0.001
	SD	0.026	0.038	0.016	0.017	0.024	0.010
	SE	0.026	0.037	0.016	0.017	0.023	0.010
	CP	0.784	0.914	0.939	0.558	0.880	0.942
		scenario M4					
$\hat\beta^{\text{ipw}}$	AB	0.013	0.010	0.002	0.004	0.002	0.001
	SD	0.040	0.046	0.016	0.027	0.031	0.011
	SE	0.040	0.046	0.016	0.029	0.031	0.010
	CP	0.940	0.935	0.952	0.952	0.943	0.949
$\hat\beta^{\text{aipw}}$	AB	0.012	0.010	0.002	0.003	0.001	0.002
	SD	0.041	0.047	0.016	0.028	0.031	0.010
	SE	0.040	0.046	0.017	0.028	0.031	0.010
	CP	0.932	0.937	0.959	0.946	0.938	0.951
$\hat\beta$	AB	0.000	0.001	0.000	0.000	0.000	0.000
	SD	0.022	0.033	0.013	0.014	0.021	0.008
	SE	0.021	0.033	0.013	0.013	0.021	0.008
	CP	0.953	0.938	0.955	0.955	0.951	0.933
$\hat\beta^{\text{cc}}$	AB	0.043	0.030	0.001	0.043	0.031	0.002
	SD	0.026	0.038	0.015	0.017	0.024	0.010
	SE	0.026	0.037	0.015	0.016	0.024	0.010
	CP	0.614	0.861	0.949	0.239	0.714	0.931

we use the last record in that interval as Y_t at time t.

To apply the proposed method, we use the baseline measurement as the instrument Z. Because the baseline measurements were taken before the treatments were applied, it is reasonable to assume that the propensity at time t does not depend on Z given $\overrightarrow{S}_{t-1}$ and Y_t. We fit propensity model (17) to the data set. The estimates and their standard errors based on the bootstrap are reported in Table 5.

It can be seen that the number of CD4 counts of Treatment I keeps decreasing rapidly compared with other treatments, which indicates that Treatments II, III and IV have good performance on the HIV-AIDS disease. Under Treatments I, II and III, all three estimates, $\hat\mu_t^{\text{ipw1}}$, $\hat\mu_t^{\text{ipw2}}$ and $\hat\mu_t^{\text{aipw}}$, and their standard errors are almost identical. However, these estimates are different from $\hat\mu_t^{\text{cc}}$. Specifically, for $t=1$ under Treatments I and IV, $\hat\mu_t^{\text{cc}}$ may be too conservative since it is smaller than $\hat\mu_t^{\text{ipw1}}$, $\hat\mu_t^{\text{ipw2}}$ and $\hat\mu_t^{\text{aipw}}$; for $t=3,4$ under Treatment I and $t=2,3,4$ under Treatments II, III and IV, $\hat\mu_t^{\text{cc}}$ may be too optimistic since it is larger than $\hat\mu_t^{\text{ipw1}}$, $\hat\mu_t^{\text{ipw2}}$ and $\hat\mu_t^{\text{aipw}}$.

Table 5. Estimates (with standard errors in parentheses) for the ACTG 193A data

Estimator	$t=1$	$t=2$	$t=3$	$t=4$
	Treatment I			
$\hat\mu_t^{\text{ipw1}}$	26.62 (2.51)	32.36 (1.91)	14.96 (2.45)	14.11 (1.26)
$\hat\mu_t^{\text{ipw2}}$	26.63 (2.48)	32.23 (1.84)	16.68 (2.52)	15.06 (1.32)
$\hat\mu_t^{\text{aipw}}$	26.59 (2.46)	32.42 (1.89)	17.26 (2.68)	14.80 (1.44)
$\hat\mu_t^{\text{cc}}$	26.21 (1.59)	23.12 (1.69)	19.25 (1.27)	18.64 (1.15)
Missing (%)	30.6	26.8	47.3	39.1
	Treatment II			
$\hat\mu_t^{\text{ipw1}}$	34.49 (3.59)	26.28 (3.68)	17.82 (1.71)	17.00 (1.40)
$\hat\mu_t^{\text{ipw2}}$	34.48 (3.46)	26.94 (3.64)	18.01 (1.76)	17.47 (1.43)
$\hat\mu_t^{\text{aipw}}$	34.58 (3.43)	26.73 (3.61)	18.42 (1.98)	17.49 (1.41)
$\hat\mu_t^{\text{cc}}$	31.19 (2.18)	27.77 (1.71)	22.21 (1.64)	23.06 (1.68)
Missing (%)	31.5	23.2	44.3	43.6
	Treatment III			
$\hat\mu_t^{\text{ipw1}}$	38.78 (5.04)	31.57 (4.21)	23.19 (3.19)	19.41 (1.93)
$\hat\mu_t^{\text{ipw2}}$	38.92 (5.01)	31.58 (4.23)	23.10 (3.22)	20.13 (2.05)
$\hat\mu_t^{\text{aipw}}$	39.78 (4.96)	31.58 (4.23)	23.81 (3.07)	19.53 (1.96)
$\hat\mu_t^{\text{cc}}$	42.20 (3.65)	36.24 (2.68)	30.12 (2.89)	30.70 (2.79)
Missing (%)	32.0	22.5	46.5	41.5
	Treatment IV			
$\hat\mu_t^{\text{ipw1}}$	48.03 (3.89)	42.06 (4.71)	30.56 (3.74)	31.97 (3.49)
$\hat\mu_t^{\text{ipw2}}$	48.17 (4.05)	42.09 (4.94)	31.50 (3.63)	32.23 (3.48)
$\hat\mu_t^{\text{aipw}}$	48.50 (4.00)	42.05 (4.70)	32.51 (4.23)	32.89 (3.77)
$\hat\mu_t^{\text{cc}}$	42.86 (3.23)	48.80 (4.32)	38.77 (3.75)	40.59 (4.36)
Missing (%)	28.4	23.5	45.7	37.0

7. DISCUSSION

Handling longitudinal data with nonignorable nonmonotone nonresponse is a challenging problem, mainly due to the issue of identifiability of the nonresponse propensity. Assumptions on propensity must be imposed to develop useful methods but they cannot be verified due to nonignorable nonresponse. We use a parametric propensity model and a GMM approach making use of a nonresponse instrument to identify unknown parameters in the propensity. Our asymptotic results for the proposed estimators are established under situations where the propensity model can be misspecified. Alternatively, we construct a model-assisted AIPW estimator that depends on an estimator of a conditional mean for nonrespondents under a working parametric model. Although the working parametric model is often misspecified, the AIPW estimator is still consistent and asymptotically normal and has good empirical performance in our simulation studies.

A nonresponse instrument Z plays a crucial role in our method, which is assumed to be given in this paper. In applications, baseline covariates are good candidates for instruments, as our empirical study of ACTG 193A indicated, because baseline covariates were obtained before the treatments were applied and it is reasonable to assume that the propensity at time t does not depend on baseline measurements given longitudinal observations $\overrightarrow{S}_{t-1}$ defined in (1) and S_t. Nonresponse instrument selection in general is very difficult and challenging, which is a topic of our future research.

The result on IPW in Section 4 can be extended to the situation where $\mathcal{G}_t(X;\beta^0)$ in (12) is replaced by $\mathcal{G}_t(X,V_t;\beta_t^0)$.

In some problems V_t may have missing components, in addition to missing Y_t values. Although our method can be applied by changing δ_t to the indicator of completely observing (Y_t,V_t), it discards incomplete data and hence is not efficient. To develop method producing efficient estimators, however, is very challenging because the missingness of (Y_t,V_t) may have many patterns that are hard to model. Further research is needed.

ACKNOWLEDGEMENTS

We are grateful to the Editor, the Associate Editor and two anonymous referees for their insightful comments and suggestions on this article, which have led to significant improvements. This work was supported by the National Natural Science Foundation of China (11501208), Postdoctoral Science Foundation of China (2014M560317), the Chinese 111 Project grant B14019, and the U.S. National Science Foundation grants DMS-1305474 and DMS-1612873.

APPENDIX

Proof of Theorem 2.1. Recall $g_t(\alpha_t) = g_t(Y,V,X,\boldsymbol{\delta},\alpha_t)$, $G_{nt}(\alpha_t) = \mathbb{E}_n\{g_t(\alpha_t)\}$ and $G_t(\alpha_t) = E\{g_t(\alpha_t)\}$. By the law of large number (LLN), it can be shown that $G_{nt}(\alpha_t) - G_t(\alpha_t) = o_p(1)$ for all $\alpha_t \in \mathcal{A}$. Since both $g_t(\alpha_t)$ and $G_{nt}(\alpha_t)$ are continuous at each $\alpha_t \in \mathcal{A}$,

$$\sup_{\alpha_t\in\mathcal{A}} \|G_{nt}(\alpha_t) - G_t(\alpha_t)\| = o_p(1).$$

This, coupled with GMM identification (i.e., Lemma 2.3 of Newey and McFadden, 1994), shows that $\tilde{\alpha}_t = \alpha_t^* + o_p(1)$. By the LLN, it can be shown that $\hat{W}_{nt} = W_t^* + o_p(1)$. Let $Q_t(\alpha_t) = G_t(\alpha_t)^\top W_t^{*-1} G_t(\alpha_t)$ and $Q_n(\alpha_t) = G_{nt}(\alpha_t)^\top \hat{W}_{nt}^{-1} G_{nt}(\alpha_t)$. Based on Lemma 2.3 and Theorem 2.1 of Newey and McFadden (1994), to prove $\hat{\alpha}_t - \alpha_t^* = o_p(1)$, it is enough to show that $\sup_{\alpha_t\in\mathcal{A}} |Q_{nt}(\alpha_t) - Q_t(\alpha_t)| = o_p(1)$. This is true because by the triangle and Cauchy-Schwartz inequalities,

$$\begin{aligned}
&\sup_{\alpha_t\in\mathcal{A}} |Q_{nt}(\alpha_t) - Q_t(\alpha_t)| \\
\le &\sup_{\alpha_t\in\mathcal{A}} \left|\{G_{nt}(\alpha_t) - G_t(\alpha_t)\}^\top \hat{W}_{nt}^{-1}\{G_{nt}(\alpha_t) - G_t(\alpha_t)\}\right| \\
&+ \sup_{\alpha_t\in\mathcal{A}} \left|G_t(\alpha_t)^\top (\hat{W}_{nt}^{-1} + (\hat{W}_{nt}^{-1})^\top)\{G_{nt}(\alpha_t) - G_t(\alpha_t)\}\right| \\
&+ \sup_{\alpha_t\in\mathcal{A}} \left|G_t(\alpha_t)^\top (\hat{W}_{nt}^{-1} - W_t^{*-1}) G_t(\alpha_t)\right| \\
\le &\sup_{\alpha_t\in\mathcal{A}} \left\|G_{nt}(\alpha_t) - G_t(\alpha_t)\}\right\|^2 \|\hat{W}_{nt}^{-1}\| \\
&+ 2\sup_{\alpha_t\in\mathcal{A}} \left\|G_t(\alpha_t)\right\|\left\|G_{nt}(\alpha_t) - G_t(\alpha_t)\right\| \|\hat{W}_t^{*-1}\| \\
&+ \sup_{\alpha_t\in\mathcal{A}} \left\|G_t(\alpha_t)\right\|^2 \|\hat{W}_{nt}^{-1} - W_t^{*-1}\| = o_p(1).
\end{aligned}$$

This proves $\hat{\alpha}_t = \alpha_t^* + o_p(1)$.

Next, we derive the asymptotic normality of $\hat{\alpha}_t$. With probability approaching one, we have the first-order condition $2\Gamma_{nt}(\hat{\alpha}_t)\hat{W}_{nt}^{-1} G_{nt}(\hat{\alpha}_t) = 0$, where $\Gamma_{nt}(\alpha_t) = \partial G_{nt}(\alpha_t)/\partial\alpha_t$. Expanding $G_{nt}(\hat{\alpha}_t)$ around α_t^*, multiplying through by $n^{1/2}$, and solving gives

$$\begin{aligned}
&n^{1/2}(\hat{\alpha}_t - \alpha_t^*) \\
= &-[\Gamma_{nt}^\top(\hat{\alpha}_t)\hat{W}_{nt}^{-1}\Gamma_{nt}(\bar{\alpha}_t)]^{-1}\Gamma_{nt}^\top(\hat{\alpha}_t)\hat{W}_{nt}^{-1} n^{1/2} G_{nt}(\alpha_t^*),
\end{aligned}$$

where $\bar{\alpha}_t$ is between $\hat{\alpha}_t$ and α_t^*. By simply calculation and the LLN, for all $\alpha_t \in \mathcal{A}$,

$$\begin{aligned}
&\Gamma_{nt}(\alpha_t) \\
= &-\mathbb{E}_n\{\overrightarrow{\xi}_t \delta_t \pi_t(\alpha_t)^{-1}[1 - \pi_t(\alpha_t)]\Xi(\alpha_t)^\top\} \\
= &-E\{\overrightarrow{\xi}_t \pi_t(\alpha_t^0)[\pi_t(\alpha_t)^{-1} - 1]\Xi(\alpha_t)^\top\} + o_p(1).
\end{aligned}$$

This, together with $\hat{W}_{nt} = W_t^* + o_p(1)$ and $\hat{\alpha}_t = \alpha_t^* + o_p(1)$, implies that

$$\begin{aligned}
&[\Gamma_{nt}^\top(\hat{\alpha}_t)\hat{W}_{nt}^{-1}\Gamma_{nt}(\bar{\alpha}_t)]^{-1}\Gamma_{nt}^\top(\hat{\alpha}_t)\hat{W}_{nt}^{-1} \\
= &[\Gamma_t^{*\top} W_t^{*-1}\Gamma_t^*]^{-1}\Gamma_t^{*\top} W_t^{*-1} + o_p(1).
\end{aligned}$$

By the Slutzky theorem, we can show $n^{1/2}(\hat{\alpha}_t - \alpha_t^*) \xrightarrow{\mathcal{L}} N(0, (\Gamma_t^{*\top} W_t^{*-1}\Gamma_t^*)^{-1})$. This completes the proof. □

Proof of Corollary 2.1. When the intermittent propensity model is correctly specified, $\pi_t(\alpha_t^*) = \pi_t(\alpha_t^0)$, $\hat{W}_{nt} = W_t + o_p(1)$ and $\hat{\alpha}_t = \alpha_t^0 + o_p(1)$. Then the proof for the Corollary 2.1 follows directly from these facts. □

Proof of Theorem 3.1. We first derive the asymptotic properties of the proposed Horvitz-Thompson type estimators $\hat{\mu}_t^{\text{ipw1}}$ and $\hat{\mu}_t^{\text{ipw2}}$. By simply calculation and the LLN, we can prove that $\partial\mathbb{E}_n[\delta_t Y_t/\pi_t(\alpha_t^0)]/\partial\alpha_t = -\Delta_{1t} + o_p(1)$, where $\Delta_{1t} = E\{Y_t[1 - \pi_t(\alpha_t^0)]\Xi(\alpha_t^0)\}$. Similarly, we have $\partial\mathbb{E}_n[\delta_t\pi_t(\alpha_t^0)^{-1}(Y_t - \mu_t^0)]/\partial\alpha_t = -\Delta_{2t} + o_p(1)$, where $\Delta_{2t} = E\{(Y_t - \mu_t^0)[1 - \pi_t(\alpha_t^0)]\Xi(\alpha_t^0)\}$. From the proof of Theorem 2.1, we have $\hat{\alpha}_t - \alpha_t^0 = -\Lambda_t G_{nt}(\alpha_t^0)$, where $\Lambda_t = [\Gamma_t^\top W_t^{-1}\Gamma_t]^{-1}\Gamma_t^\top W_t^{-1}$. Then, for $\hat{\mu}_t^{\text{ipw1}}$, we have

$$\hat{\mu}_t^{\text{ipw1}} - \mu_t^0$$

$$= \mathbb{E}_n[\delta_t Y_t/\pi_t(\hat\alpha_t)] - \mu_t^0$$
$$= \mathbb{E}_n[\delta_t Y_t/\pi_t(\alpha_t^0)] - \mu_t^0 - \Delta_{1t}^\top(\hat\alpha_t - \alpha_t^0) + o_p(n^{-1/2})$$
$$= \mathbb{E}_n[\delta_t Y_t/\pi_t(\alpha_t^0) + \Delta_{1t}^\top \Lambda_t g_t(\alpha_t^0)] - \mu_t^0 + o_p(n^{-1/2}),$$

and for $\hat\mu_t^{\mathrm{ipw2}}$, we have

$$\begin{aligned}
&\hat\mu_t^{\mathrm{ipw2}} - \mu_t^0 \\
=\ & \mathbb{E}_n[\delta_t Y_t/\pi_t(\hat\alpha_t)]/\mathbb{E}_n[\delta_t/\pi_t(\hat\alpha_t)] - \mu_t^0 \\
=\ & \{\mathbb{E}_n[\delta_t/\pi_t(\hat\alpha_t)]\}^{-1}\mathbb{E}_n[\delta_t \pi_t(\hat\alpha_t)^{-1}(Y_t - \mu_t^0)] \\
=\ & \mathbb{E}_n[\delta_t \pi_t(\alpha_t^0)^{-1}(Y_t - \mu_t^0)] - \Delta_{2t}^\top(\hat\alpha_t - \alpha_t^0) + o_p(n^{-1/2}) \\
=\ & \mathbb{E}_n[\delta_t \pi_t(\alpha_t^0)^{-1}(Y_t - \mu_t^0) + \Delta_{2t}^\top \Lambda_t g_t(\alpha_t^0)] + o_p(n^{-1/2}).
\end{aligned}$$

Applying the central limit theorem (CLT), $n^{1/2}(\hat\mu_t^{\mathrm{ipw}\kappa} - \mu_t^0) \xrightarrow{\mathcal{L}} N(0, V_{\kappa t})$, where $V_{1t} = \mathrm{Var}\{\delta_t Y_t/\pi_t(\alpha_t^0) + \Delta_{1t}^\top \Lambda_t g_t(\alpha_t^0)\}$ and $V_{2t} = \mathrm{Var}\{\delta_t \pi_t(\alpha_t^0)^{-1}(Y_t - \mu_t^0) + \Delta_{2t}^\top \Lambda_t g_t(\alpha_t^0)\}$.

Next, we consider the asymptotic distribution of the AIPW estimator $\hat\mu_t^{\mathrm{aipw}}$. Note that $n^{-1}\mathbb{E}_n(\delta_t) = \Pr(\delta_t = 1) + o_p(1)$. Let $\omega_t = \Pr(\delta_t = 1)$. Using the arguments of White (1982, Theorem 3.1 and Theorem 3.2), we can show that $\hat\gamma_t - \gamma_t^* = o_p(1)$ and

$$\begin{aligned}
&\hat\gamma_t - \gamma_t^* \\
&= -\{\omega_t A_t(\gamma_t^*)\}^{-1}\mathbb{E}_n\Big\{\delta_t \partial \log f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^*)/\partial\gamma_t\Big\} \\
&\quad + o_p(n^{-1/2}).
\end{aligned}$$

We consider $\partial\mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^*)\}/\partial\alpha_t$, $\partial\mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^*)\}/\partial\gamma_t$. By calculation, we obtain

$$\begin{aligned}
&\frac{\partial\mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^*)\}}{\partial\alpha_t} \\
&= -\mathbb{E}_n[\delta_t \pi_t(\alpha_t^0)^{-2}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}\partial\pi_t(\alpha_t^0)/\partial\alpha_t] \\
&\quad + \mathbb{E}_n[\{1 - \delta_t \pi_t(\alpha_t^0)^{-1}\}\partial m_{0t}(\alpha_t^0, \gamma_t^*)/\partial\alpha_t] \\
&=: T_{n1} + T_{n2}.
\end{aligned}$$

Under propensity model (1), it is easy to show that $T_{n2} = o_p(1)$. Note that

$$\begin{aligned}
&E\Big\{\frac{\delta_t}{\pi_t(\alpha_t^0)}\{1 - \pi_t(\alpha_t^0)\}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}\Xi(\alpha_t^0)\Big\} \\
=\ & E\Big\{\frac{\delta_t}{\pi_t(\alpha_t^0)}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}\{\delta_t - \pi_t(\alpha_t^0)\}\Xi(\alpha_t^0)\Big\} \\
=\ & E\Big\{\frac{\delta_t}{\pi_t(\alpha_t^0)}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}\eta_t(\alpha_t^0)\Big\} =: \Phi_t.
\end{aligned}$$

Note that we also have $\Phi_t = \mathrm{Cov}[\delta_t \pi_t(\alpha_t^0)^{-1}\{Y_t - m_{0t}(\alpha_t^0, \gamma_t^*)\}, \eta_t(\alpha_t^0)]$. Then, we have $T_{n1} = -\Phi_t + o_p(1)$. Applying the LLN and under nonresponse assumption (1), it can be shown that $\partial\mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^*)\}/\partial\gamma_t = o_p(1)$. Therefore, for $\hat\mu_t^{\mathrm{aipw}}$, we have that

$$\begin{aligned}
&\hat\mu_t^{\mathrm{aipw}} - \mu_t^0 \\
=\ & \mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^0)\} - \mu_t^0 + \Phi_t^\top \Lambda_t G_{nt}(\alpha_t^0) + o_p(n^{-1/2}) \\
=\ & \mathbb{E}_n\{H_t(\alpha_t^0, \gamma_t^0) + \Phi_t^\top \Lambda_t g_t(\alpha_t^0)\} - \mu_t^0 + o_p(n^{-1/2}).
\end{aligned}$$

Applying the CLT, we have that $n^{1/2}(\hat\mu_t^{\mathrm{aipw}} - \mu_t^0) \xrightarrow{\mathcal{L}} N(0, V_{3t})$, where $V_{3t} = \mathrm{Var}\{H_t(\alpha_t^0, \gamma_t^0) + \Phi_t^\top \Lambda_t g_t(\alpha_t^0)\}$. The proof of Theorem 3.1 is completed. □

Proof of Theorem 4.1. We first prove $\hat\beta^{\mathrm{ipw}} \xrightarrow{p} \beta_I^*$. It is sufficient to verify the conditions of theorem 5.9 of van der Vaart (1998). According to Lemma 2.4 in Newey and Mcfadden (1994), together with the continuity of $\hat U_I(\beta, \hat\alpha_t)$ in β, we can show that

$$\sup_{\beta\in\mathcal{B}} \|\hat U_I(\beta, \hat\alpha_t) - U_I(\beta, \alpha_t^*)\| \xrightarrow{p} 0.$$

Next, we show $\inf_{\beta:\|\beta-\beta_I^*\|\geq\epsilon}\|U_I(\beta, \alpha_t^*)\| > 0$, for any $\epsilon > 0$ and $\boldsymbol{\beta} \in \mathbb{R}^d$. This is true because

$$\begin{aligned}
&\inf_{\beta:\|\beta-\beta_I^*\|\geq\epsilon}\|U_I(\beta, \alpha_t^*)\| \\
=\ & \inf_{\beta:\|\beta-\beta_I^*\|\geq\epsilon}\|U_I(\beta, \alpha_t^*) - U_I(\beta_I^*, \alpha_t^*)\| \\
\geq\ & \inf_{\beta:\|\beta-\beta_I^*\|\geq\epsilon}\|[\mathcal{J}_1(\tilde\beta, \alpha_t^*) - \mathcal{J}_2(\tilde\beta, \alpha_t^*)](\beta - \beta_I^*)\|,
\end{aligned}$$

which is strictly positive under Assumption C3. Here $\tilde\beta$ is between β and β_I^*. Therefore, all underlying conditions of Theorem 5.9 of Van der Vaart (1998) hold, and this proves the consistency of $\hat\beta^{\mathrm{ipw}}$.

Using simple algebraic manipulations,

$$\begin{aligned}
&\frac{\partial\hat U_I(\beta_I^*, \alpha_t^*)}{\partial\beta^\top} \\
=\ & \sum_{t=1}^T \mathbb{E}_n\Big\{\frac{\delta_t}{\pi(\alpha_t^*)}\mathscr{D}_t(X, \beta_I^*)[Y_t - \mathcal{G}_t(X;\beta_I^*)]\Big\} \\
&- \sum_{t=1}^T \mathbb{E}_n\Big\{\frac{\delta_t}{\pi(\alpha_t^*)}\mathcal{D}_t(X, \beta_I^*)\mathcal{D}_t(X, \beta_I^*)^\top\Big\} \\
=\ & E\Big\{\sum_{t=1}^T \pi(\alpha_t^0)\pi(\alpha_t^*)^{-1}\mathscr{D}_t(X, \beta_I^*)[Y_t - \mathcal{G}_t(X;\beta_I^*)]\Big\} \\
&- E\Big\{\sum_{t=1}^T \pi(\alpha_t^0)\pi(\alpha_t^*)^{-1}\mathcal{D}_t(X, \beta_I^*)\mathcal{D}_t(X, \beta_I^*)^\top\Big\} + o_p(1) \\
=: & \mathcal{J}_1^* - \mathcal{J}_2^* + o_p(1),
\end{aligned}$$

and

$$\begin{aligned}
&-\frac{\partial\hat U_I(\beta_I^*, \alpha_t^*)}{\partial\alpha_t^\top} \\
&= E\Big\{\sum_{t=1}^T \pi_t(\alpha_t^0)[\pi_t(\alpha_t^*)^{-1} - 1]\mathcal{D}_t(X, \beta_I^*) \\
&\quad \times [Y_t - \mathcal{G}_t(X;\beta_I^*)]\Xi(\alpha_t^*)^\top\Big\} + o_p(1)
\end{aligned}$$

$$=: \mathcal{J}_3^* + o_p(1).$$

Now using Taylor expansion, $\hat{U}_I(\hat\beta^{\text{ipw}}, \hat\alpha_t) = \hat{U}_I(\beta_I^*, \alpha_t^*) + (\mathcal{J}_1^* - \mathcal{J}_2^*)(\hat\beta^{\text{ipw}} - \beta_I^*) - \mathcal{J}_3^*(\hat\alpha_t - \alpha_t^*) + o_p(n^{-1/2})$, which leads to

$$\hat\beta^{\text{ipw}} - \beta_I^* = -(\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1}\{\hat{U}_I(\beta_I^*, \alpha_t^*) - \mathcal{J}_3^*(\hat\alpha_t - \alpha_t^*)\} + o_p(n^{-1/2}).$$

Let $\Lambda_t^* = [\Gamma_t^{*\top} W_t^{*-1}\Gamma_t]^{*-1}\Gamma_t^{*\top}W_t^{*-1}$. From Theorem 2.1, we have $\hat\alpha_t - \alpha_t^* = -\Lambda_t^* G_{nt}(\alpha_t^*)$. Therefore, we obtain

$$\begin{aligned}&\hat\beta^{\text{ipw}} - \beta_I^* \\ &= -(\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1}\sum_{t=1}^T \mathbb{E}_n\Big\{\delta_t\pi_t(\alpha_t^*)^{-1}\mathcal{D}_t(X,\beta_I^*) \\ &\quad \times [Y_t - \mathcal{G}_t(X;\beta_I^*)] + \mathcal{J}_3^*\Lambda_t^* g_t(\alpha_t^*)\Big\} + o_p(n^{-1/2}).\end{aligned}$$

Applying the CLT, we have that $n^{1/2}(\hat\beta^{\text{ipw}} - \beta_I^*) \xrightarrow{\mathcal{L}} N(0, \Sigma_I^*)$, where $\Sigma_I^* = (\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1}V_I^*(\mathcal{J}_1^* - \mathcal{J}_2^*)^{-1}$ with $V_I^* = \text{Var}\{\sum_{t=1}^T \delta_t\pi_t(\alpha_t^*)^{-1}\mathcal{D}_t(X,\beta_I^*)[Y_t - \mathcal{G}_t(X;\beta_I^*)] + \mathcal{J}_3^*\Lambda_t^* g_t(\alpha_t^*)\}$. □

Proof of Corollary 4.1. When the intermittent propensity model and the marginal regression model are correctly specified, $\hat\alpha_t = \alpha_t^0 + o_p(1)$, $\pi_t(\alpha_t^*) = \pi_t(\alpha_t^0)$, $\mathcal{G}_t(X;\beta_I^*) = \mathcal{G}_t(X;\beta^0)$, and

$$\frac{\partial \hat{U}_I(\beta^0, \alpha_t^0)}{\partial \beta^\top} = -E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta^0)\mathcal{D}_t(X,\beta^0)^\top\Big\} + o_p(1);$$

$$\begin{aligned}-\frac{\partial \hat{U}_I(\beta^0, \alpha_t^0)}{\partial \alpha_t^\top} =& E\Big\{\sum_{t=1}^T \pi_t(\alpha_t^0)[\pi_t(\alpha_t^0)^{-1} - 1]\mathcal{D}_t(X,\beta^0) \\ &\times [Y_t - \mathcal{G}_t(X;\beta^0)]\Xi(\alpha_t^0)^\top\Big\} + o_p(1).\end{aligned}$$

Then the proof for the Corollary 4.1 follows directly from these facts. □

Proof of Theorem 4.2. By the LLN, together with the facts $\hat\alpha_t = \alpha_t^* + o_p(1)$ and $\hat\gamma_t = \gamma_t^* + o_p(1)$, it can be shown that $\hat{U}_A(\beta, \hat\alpha_t, \hat\gamma_t) = U_A(\beta, \alpha_t^*, \gamma_t^*) + o_p(1)$ for all $\beta \in \mathcal{B}$. Since $\mathcal{G}_t(X;\beta)$ is continuous at each $\beta \in \mathcal{B}$, we further have that

$$\sup_{\beta\in\mathcal{B}} \|\hat{U}_A(\beta, \hat\alpha_t, \hat\gamma_t) - U_A(\beta, \alpha_t^*, \gamma_t^*)\| \xrightarrow{p} 0.$$

Additionally,

$$\begin{aligned}&\inf_{\beta:\|\beta-\beta_A^*\|\ge\epsilon} \|U_A(\beta, \alpha_t^*, \gamma_t^*)\| \\ =& \inf_{\beta:\|\beta-\beta_I^*\|\ge\epsilon} \|U_A(\beta, \alpha_t^*, \gamma_t^*) - U_A(\beta_A^*, \alpha_t^*, \gamma_t^*)\| \\ \ge& \inf_{\beta:\|\beta-\beta_I^*\|\ge\epsilon} \|[\mathcal{I}_1(\tilde\beta, \alpha_t^*, \gamma_t^*) - \mathcal{J}_2(\tilde\beta, \alpha_t^0)](\beta - \beta_A^*)\|,\end{aligned}$$

which is strictly positive under Assumption C6. Here $\tilde\beta$ is between β and β_A^*. Then, using the arguments of Van der Vaart (1998, Theorem 5.9), we can show that $\hat\beta^{\text{aipw}} \xrightarrow{p} \beta_A^*$.

Using simple algebraic manipulations and LLN,

$$\begin{aligned}&\frac{\partial \hat{U}_A(\beta_A^*, \alpha_t^*, \gamma_t^*)}{\partial\beta^\top} \\ =& \sum_{t=1}^T \mathbb{E}_n\Big\{\mathscr{D}_t(X,\beta_I^*)[H_t(\alpha_t^*, \gamma_t^*) - \mathcal{G}_t(X;\beta_I^*)]\Big\} \\ &- \sum_{t=1}^T \mathbb{E}_n\Big\{\mathcal{D}_t(X,\beta_I^*)\mathcal{D}_t(X,\beta_I^*)^\top\Big\} \\ =& E\Big\{\sum_{t=1}^T \mathscr{D}_t(X,\beta_I^*)[H_t(\alpha_t^*, \gamma_t^*) - \mathcal{G}_t(X;\beta_I^*)]\Big\} \\ &- E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta_I^*)\mathcal{D}_t(X,\beta_I^*)^\top\Big\} + o_p(1) \\ &=: \mathcal{I}_1^* - \mathcal{J}_2^* + o_p(1),\end{aligned}$$

$$\begin{aligned}&\frac{\partial \hat{U}_A(\beta_A^*, \alpha_t^*, \gamma_t^*)}{\partial\alpha_t^\top} \\ &= -\sum_{t=1}^T \mathbb{E}_n\Big\{\mathcal{D}_t(X,\beta_A^*)\delta_t\pi_t(\alpha_t^*)^{-2}\{Y_t - m_{0t}(\alpha_t^*, \gamma_t^*)\} \\ &\quad \times \frac{\partial\pi_t(\alpha_t^*)}{\partial\alpha_t^\top} - \mathcal{D}_t(X,\beta_A^*)[1 - \delta_t\pi_t(\alpha_t^*)^{-1}]\frac{\partial m_{0t}(\alpha_t^*, \gamma_t^*)}{\partial\alpha_t^\top}\Big\} \\ &= -E\Big\{\sum_{t=1}^T \mathcal{D}_t(X,\beta_A^*)\delta_t\pi_t(\alpha_t^*)^{-2}\{Y_t - m_{0t}(\alpha_t^*, \gamma_t^*)\} \\ &\quad \times \frac{\partial\pi_t(\alpha_t^*)}{\partial\alpha_t^\top} - \mathcal{D}_t(X,\beta_A^*)[1 - \delta_t\pi_t(\alpha_t^*)^{-1}]\frac{\partial m_{0t}(\alpha_t^*, \gamma_t^*)}{\partial\alpha_t^\top}\Big\} \\ &\quad + o_p(1) \\ &=: -\mathcal{I}_2^* + \mathcal{I}_3^* + o_p(1),\end{aligned}$$

$$\begin{aligned}&\frac{\partial \hat{U}_A(\beta_A^*, \alpha_t^*, \gamma_t^*)}{\partial\gamma^\top} \\ =& \sum_{t=1}^T \mathbb{E}_n\Big\{\mathcal{D}_t(X,\beta_A^*)[1 - \delta_t\pi_t(\alpha_t^*)^{-1}]\frac{\partial m_{0t}(\alpha_t^*, \gamma_t^*)}{\partial\gamma^\top}\Big\} \\ =& E\Big\{\mathcal{D}_t(X,\beta_A^*)[1 - \delta_t\pi_t(\alpha_t^*)^{-1}]\frac{\partial m_{0t}(\alpha_t^*, \gamma_t^*)}{\partial\gamma^\top}\Big\} + o_p(1) \\ &=: \mathcal{I}_4^* + o_p(1).\end{aligned}$$

Applying Taylor expansion, together with asymptotic expansions for $\hat\alpha_t$ and $\hat\gamma_t$, we show that

$$\begin{aligned}&\hat{U}_A(\hat\beta^{\text{aipw}}, \hat\alpha_t, \hat\gamma_t) \\ =& \hat{U}_A(\beta_A^*, \alpha_t^*, \gamma_t^*) + (\mathcal{I}_1^* - \mathcal{J}_2^*)(\hat\beta^{\text{aipw}} - \beta_A^*) \\ &- (\mathcal{I}_2^* - \mathcal{I}_3^*)(\hat\alpha_t - \alpha_t^*) + \mathcal{I}_4^*(\hat\gamma_t - \gamma_t^*) + o_p(n^{-1/2}) \\ =& \hat{U}_A(\beta_A^*, \alpha_t^*, \gamma_t^*) + (\mathcal{I}_1^* - \mathcal{J}_2^*)(\hat\beta^{\text{aipw}} - \beta_A^*) \\ &- \mathcal{I}_4^*\{\omega_t A_t(\gamma_t^*)\}^{-1}\mathbb{E}_n\Big\{\delta_t\partial\log f_t(S_t|\overrightarrow{S}_{t-1}, \gamma_t^*)/\partial\gamma_t\Big\} \\ &+ (\mathcal{I}_2^* - \mathcal{I}_3^*)\Lambda_t^* G_{nt}(\alpha_t^*) + o_p(n^{-1/2})\end{aligned}$$

$$= \sum_{t=1}^{T} \mathbb{E}_n\{\Psi_t(\beta_A^*, \alpha_t^*, \gamma_t^*)\} + (\mathcal{I}_1^* - \mathcal{J}_2^*)(\hat{\beta}^{\text{aipw}} - \beta_A^*) + o_p(n^{-1/2}),$$

where $\Psi_t(\beta_A^*, \alpha_t^*, \gamma_t^*) = \varphi_{1t}(\beta_A^*, \alpha_t^*, \gamma_t^*) + \varphi_{2t}(\alpha_t^*, \gamma_t^*) + \varphi_{3t}(\alpha_t^*, \gamma_t^*)$, $\varphi_{1t}(\beta_A^*, \alpha_t^*, \gamma_t^*) = \mathcal{D}_t(X, \beta_A^*)[H_t(\alpha_t^*, \gamma_t^*) - \mathcal{G}_t(X; \beta_A^*)]$, $\varphi_{2t}(\alpha_t^*, \gamma_t^*) = (\mathcal{I}_2^* - \mathcal{I}_3^*)\Lambda_t^* g_t(\alpha_t^*)$, $\varphi_{3t}(\alpha_t^*, \gamma_t^*) = -\mathcal{I}_4^*\{\omega_t A_t(\gamma_t^*)\}^{-1}\delta_t \partial \log f_t(S_t | \overrightarrow{S}_{t-1}, \gamma_t^*)/\partial\gamma_t$. Then,

$$\hat{\beta}^{\text{aipw}} - \beta_A^* = -(\mathcal{I}_1^* - \mathcal{J}_2^*)^{-1} \sum_{t=1}^{T} \mathbb{E}_n\{\Psi_t(\beta_A^*, \alpha_t^*, \gamma_t^*)\} + o_p(n^{-1/2}).$$

Applying the CLT, we have that $n^{1/2}(\hat{\beta}^{\text{aipw}} - \beta_A^*) \xrightarrow{\mathcal{L}} N(0, \Sigma_A^*)$, where $\Sigma_A^* = (\mathcal{I}_1^* - \mathcal{J}_2^*)^{-1} V_A^* (\mathcal{I}_1^* - \mathcal{J}_2^*)^{-1}$ with $V_A^* = \text{Var}\{\sum_{t=1}^{T} \Psi_t(\beta_A^*, \alpha_t^*, \gamma_t^*)\}$. □

Proof of Corollary 4.2. When all the models used are correctly specified, $\hat{\alpha}_t = \alpha_t^0 + o_p(1)$, $\hat{\gamma}_t = \gamma_t^0 + o_p(1)$, and

$$\frac{\partial \hat{U}_A(\beta^0, \alpha_t^0, \gamma_t^0)}{\partial \beta^\top} = -E\Big\{\sum_{t=1}^{T} \mathcal{D}_t(X, \beta^0)\mathcal{D}_t(X, \beta^0)^\top\Big\} + o_p(1),$$

$$\frac{\partial \hat{U}_A(\beta^0, \alpha_t^0, \gamma_t^0)}{\partial \alpha_t^\top} = -E\Big\{\sum_{t=1}^{T} \pi_t(\alpha_t^0)[\pi_t(\alpha_t^0)^{-1} - 1]\mathcal{D}_t(X, \beta^0) \times \{Y_t - m_{0t}(\alpha_t^0, \gamma_t^0)\}\Xi(\alpha_t^0)^\top\Big\} + o_p(1),$$

$$\frac{\partial \hat{U}_A(\beta^0, \alpha_t^0, \gamma_t^0)}{\partial \gamma^\top} = o_p(1).$$

Then the proof for the Corollary 4.2 follows directly from the above facts. □

Received 15 December 2016

REFERENCES

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.

Henry, K., Erice, A., Tierney, C., Balfour, H. H. J., Fischl, M. A., Kmack, A., Liou, S. H., Kenton, A., Hirsch, M. S., Phair, J., Martinez, A., Kahn, J. O. 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** 339–349. MR2495958

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test* **18** 1–43. MR2804214

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika* **98** 119–132.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7** 305–315. MR1354029

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90** 1112–1121. MR1925014

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd ed.), Wiley, New York. MR1315971

Newey, W. and Mcfadden, D. (1994). *Large Sample Estimation and Hypothesis Testing*. Springer, New York.

Riddles, M. K., Kim, J. K. and Im, J. (2016). A propensity-score-adjustment method for nonignorable nonresponse. *Journal of Survey Statistics and Methodology* **4** 215–245.

Robins, J. M. and Ritov, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine* **16** 285–319. MR1294730

Robins, J. M., Rotnitzky, A. and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89** 846–866. MR1731478

Scharfstein, D., Rotnitzky, A. and Robins, J. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94** 1096–1120.

Shao, J., Klein, M. and Xu, J. (2012). Imputation for nonmonotone nonresponse in the survey of industrial research and development. *Survey Methodology* **38** 143–155. MR3465829

Shao, J. and Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika* **103** 175–187. MR1665826

Troxel, A. B., Lipsitz, S. R. and Harrington, D. P. (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika* **85** 661–672. MR2416795

Vansteelandt, S., Rotnitzky, A. and Robins, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94** 841–860. MR1652247

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge. MR3241279

Wang, S., Shao, J. and Kim, J. K. (2014). An instrument variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica* **24** 1097–1116. MR0640163

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50** 1–25.

Xu, J., Shao, J., Palta, M. and Wang, L. (2008). Imputation for longitudinal data with last-value-dependent non-monotone missing values. *Survey Methodology* **34** 153–162.

Puying Zhao
Department of Statistics and Actuarial Science
University of Waterloo
Canada
E-mail address: pyzhao@live.cn

Lei Wang
LPMC and Institute of Statistics
Nankai University
Tianjin
China
Department of Statistics
University of Wisconsin-Madison
U.S.A.
E-mail address: leiwang.stat@gmail.com

Jun Shao
Department of Statistics
University of Wisconsin-Madison
U.S.A.
E-mail address: shao@stat.wisc.edu