Title	Kernel-Profile Efficient Estimation in Generalized Partially Linear Models with Missing Outcomes in Longitudinal Studies
Manuscript ID	SS-2024-0380
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0380
Complete List of Authors	Zhongzhe Ouyang, Chang Wang and Lu Wang
Corresponding Authors	Lu Wang
E-mails	luwang@umich.edu
Notice: Accepted author version.	

Statistica Sinica

Kernel-Profile Efficient Estimation in Generalized Partially Linear Models With Missing Outcomes in Longitudinal Studies

Zhongzhe Ouyang[1], Chang Wang[1], and Lu Wang[1]

[1]*Department of Biostatistics, University of Michigan*

Abstract: We consider a generalized partially linear model with missing outcomes in longitudinal studies. Our proposed method, the longitudinal augmented inverse probability weighted kernel-profile estimating equations, employs kernel estimating equations for the nonparametric part and profile estimating equations for the parametric part. Auxiliary variables are used to model both the missingness and the conditional mean. The resulting estimators for both the parametric and nonparametric parts are doubly robust. To further understand these estimators, we derive the semiparametric efficiency bound and the asymptotic properties of the proposed estimators. We find that the estimator for the parametric part attains the semiparametric efficiency bound under the multivariate normal assumption. We demonstrate the empirical performance of the proposed method through simulation studies and an application to CD4 count data.

Key words and phrases: Correlated data; Double robustness; Augmented inverse probability weighting; Profile-kernel estimating equations; Semiparametric efficiency

1. Introduction

Our work is motivated by the Latency and Early Neonatal Provision of Antiretroviral Drugs Clinical Trial (LEOPARD) study (Kuhn et al. 2020; Yates and Kuhn 2022). When investigating the dynamic progression of CD4+ T cell count and its predictors, it is well known in the literature that the relationship between CD4 count and time is nonlinear. Additionally, the dataset also includes viral load as a time-varying variable, which is known to be strongly associated with both CD4 count and with the likelihood of missingness. However, including it in the main model could potentially distort the association between CD4 count and other infant characteristics. Consequently, we treat viral load as an auxiliary variable, which is not our primary interest but helps to explain the missingness of CD4 count in the dataset. This motivates the development of an estimation method for the generalized partially linear model tailored for longitudinal data with missing outcomes, where auxiliary variables exist.

Partially linear models extend linear models, offering g r eater flexibility in modeling complex relationships between the response variable and covariates (Engle et al., 1986; Zhang et al., 2011; Härdle et al., 2012). In these models, the response variable Y is characterized by two components: a parametric part that entails a linear function of predictors with a finitedimensional parameter $\boldsymbol{\beta}$, and a nonparametric part, which is a nonparametric function $\theta(\cdot)$ of a scalar variable. In longitudinal studies, a common approach to estimating partially linear models is through the use of the profile-kernel method (Severini and Staniswalis 1994; Lin and Carroll 2001a; Lin and Carroll 2001b). However, for repeated measures data, the conventional profile-kernel method yields a $\sqrt{n}$-inconsistent estimator of $\boldsymbol{\beta}$ unless either a working independence correlation structure is used or an undersmoothed kernel method is adopted. To address this issue, Wang et al. (2005) modified the profile-kernel method by substituting the conventional kernel estimator with the iterative kernel estimator, which is capable of accounting for the within-cluster correlation (Wang 2003; Lin et al. 2004). Nevertheless, these methods are primarily designed for complete data.

Early work on partially linear models with missing data primarily focused on cross-sectional settings, where missingness arises either in outcomes or covariates. A variety of approaches have been proposed to address these problems, including inverse probability weighting, imputation, and kernel-based estimators (Wang et al. 2004; Liang et al. 2007; Liang 2008; Qin et al. 2012).

Subsequent work put more emphasis on longitudinal settings. For example, Zhang and Zhu (2011) proposed kernel-profile estimators to handle both missing outcomes and error-prone covariates in longitudinal studies. Chen and Zhou (2013) developed inverse probability weighted kernel-profile estimating equations to handle missing outcomes under the assumption of missing at random (MAR, Rubin 1976), incorporating population-level information using a pseudo-empirical likelihood-based method. To enhance robustness, Lin et al. (2017) introduced doubly robust estimators to handle missing outcomes that remain consistent if either the linear conditional mean model or the dropout model is correctly specified. Although they included past outcomes as auxiliary variables to achieve double robustness and improve efficiency, their reliance on the assumptions of linear conditional mean and constant conditional variance limits their method, as it fails to explore the correlation between missing outcomes and covariates. Some research also addresses the case of missing not at random (Shao and Wang 2022; Du et al. 2023). However, these methods did not consider auxiliary variables, and estimating the dropout model remains challenging.

Recently, Wang et al. (2024) introduced a semiparametric efficient and doubly robust estimator for cross-sectional data with missing outcomes, incorporating auxiliary variables to improve robustness and efficiency. Despite these advancements, it is limited to independent, cross-sectional settings and does not account for within-subject correlation present in longitudinal data. Extending their framework to longitudinal data poses several key challenges. First, it requires a re-derivation of the semiparametric efficiency bound with the longitudinal data structure, and then, we need to reconstruct appropriate estimating equations based on the efficient score. Second, in contrast to the cross-sectional setting, estimating the nonparametric function $\theta(\cdot)$ in longitudinal data presents additional challenges. Standard kernel smoothing methods tend to ignore a certain dependency structure, leading to a paradox where working independence or undersmoothing may yield greater efficiency than using the true correlation structure. This calls for a modified kernel estimation approach. The main contributions of this paper that distinguish it from previous work are as follows: i) Unlike Wang et al. (2005) and Wang et al. (2024), we additionally allow the outcomes to have a monotone missing pattern in longitudinal studies; ii) We derive the semiparametric efficiency bound and efficient score in the generalized partially linear models with missing outcomes in longitudinal studies. iii) By leveraging auxiliary variable information, the proposed method in this paper achieves superior robustness and efficiency relative to those in litera-ture, such as Chen and Zhou (2013) and Lin et al. (2017); and iv) Motivated by the iterative kernel estimator (Wang et al., 2005), we modified the kernel estimating equations in longitudinal setting so that the efficiency of $\hat{\boldsymbol{\beta}}$ does not require a working independence correlation structure or an undersmoothed kernel estimator, and the efficient estimation of $\boldsymbol{\beta}$ accounts for within-subject correlation appropriately.

Augmented inverse probability weighting (AIPW) methods are widely used for handling missing data problems. These methods combine outcome regression models and inverse probability weighting to improve the robustness of the estimator, while incorporating an additional augmentation term to enhance estimation efficiency (Robins et al., 1994; Bang and Robins, 2005). Building upon this idea, we propose the longitudinal augmented inverse probability weighted (LAIPW) kernel-profile estimating equations (LAIPW-KPEE), an estimation method for the generalized partially linear model tailored for longitudinal data with missing outcomes and auxiliary variables. Observing that an infant with a high viral load is more likely to be lost to follow-up at the next visit, we utilize this auxiliary variable to explain the missing data. Obviously, both the dropout probability and the conditional mean of the CD4 count depend on the history of viral load, where higher past viral load levels are typically associated with larger dropout probability and lower CD4 counts. We prove that with auxiliary variables, our method can improve the efficiency of estimating both $\boldsymbol{\beta}$ and $\theta(\cdot)$. The LAIPW kernel-profile estimators of $\boldsymbol{\beta}$ and $\theta(\cdot)$ are doubly robust in that consistency is guaranteed if either the missing data model or the conditional mean model is correctly specified, but not necessarily both. Although it is difficult to construct a closed-form expression for semiparametrically efficient score in a more general case, the LAIPW kernel-profile estimator of $\boldsymbol{\beta}$ attains the semiparametric efficiency bound under the multivariate normal assumption. Our approach is specifically designed for scenarios where auxiliary variables are available and play a key role in explaining the missing data.

This paper is organized as follows: Section 2 outlines the background of the problem and introduces the proposed LAIPW kernel-profile estimating equations. Section 3 investigates the semiparametric efficiency bounds under the multivariate normal assumption. Section 4 explores the asymptotic properties of the estimators of $\boldsymbol{\beta}$ and $\theta(\cdot)$. A simulation study, conducted in Section 5, assesses the performance of the LAIPW-KPEE method. An application of the LAIPW-KPEE method in the LEOPARD study is described in Section 6. Section 7 concludes with a discussion.

2. Methods

2.1 Generalized Partially Linear Models

Consider a longitudinal study with n subjects and an equal number of post-baseline measurements m. Let $\boldsymbol{X}_i = (\boldsymbol{X}_{i0}, \ldots, \boldsymbol{X}_{im})^T$ and $\boldsymbol{T}_i = (T_{i0}, \ldots, T_{im})^T$, where $(\boldsymbol{X}_{ij}^T, T_{ij})^T$ is a vector of covariates collected from subject i $(i = 1, \ldots, n)$ at time j $(j = 1, \ldots, m)$ with $\boldsymbol{X}_{ij}$ being a p-dimensional vector and T_{ij} a scalar. In addition to $\boldsymbol{X}_{ij}$ and T_{ij}, auxiliary variables $\boldsymbol{U}_i = (\boldsymbol{U}_{i0}, \ldots, \boldsymbol{U}_{im})^T$ are also considered. Here, time 0 indexes the baseline measurement prior to the start of follow-up. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ denote the outcomes measured after the baseline. We assume that $\boldsymbol{X}_{ij}$ and T_{ij} are always observed, whereas Y_{ij} can be subject to dropout. This can occur, for example, when $\boldsymbol{X}_{ij}$ consists of deterministic functions of the baseline variables or external time-dependent variables, and T_{ij} is a deterministic function of time. Define $\boldsymbol{W}_{ij} = (\boldsymbol{U}_{ij}^T, Y_{ij})^T$, $j = 1, \ldots, m$ and $\boldsymbol{W}_{i0} = \boldsymbol{U}_{i0}$. We use $\overline{\boldsymbol{W}}_{ij} = \{\boldsymbol{X}_i, \boldsymbol{T}_i, \boldsymbol{W}_{i0}, \boldsymbol{W}_{i1}, \ldots, \boldsymbol{W}_{ij}\}$ to denote the observed data for subject i at time j. The inclusion of $\boldsymbol{X}_i$ and $\boldsymbol{T}_i$ simplifies the notation when conditioning on the observed data. Dropout means that once patients leave the study, they will not return. Let $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{im})^T$ denote a vector of missing indicator with $R_{ij} = 1$

if $\boldsymbol{U}_{ij}$ and Y_{ij} are observed at time j and $R_{ij} = 0$ otherwise. Throughout this paper, we assume $\{\overline{\boldsymbol{W}}_{im}, \boldsymbol{R}_i\}_{i=1,\ldots,n}$ are independent and identically distributed.

Consider the generalized partially linear models

$E(Y_{ij}\,|\boldsymbol{X}_{ij}\,, T_{ij}\,) = \mu\{\boldsymbol{X}_{ij}^T\boldsymbol{\beta} + \theta(T_{ij}\,)\}, j = 1, \ldots, m,$ (2.1) where $\mu(\cdot)$ is a known monotonic link function, $\boldsymbol{\beta}$ is a p-dimensional vector, and $\theta(\cdot)$ is an unknown smooth function. It is worth noting that the esti-mation of interest is $E(Y_{ij}\,|\boldsymbol{X}_{ij}\,, T_{ij}\,)$ instead of $E(Y_{ij}\,|\boldsymbol{X}_{ij}\,, T_{ij}\,, \boldsymbol{U}_{ij}\,)$, which is why we consider $\boldsymbol{U}_i$ as auxiliary variables.

We assume the missing data process satisfies

$$P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}, \boldsymbol{Y}_i) = P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}). \tag{2.2}$$

Under assumption (2.2), among subjects observed at time $j-1$, the probability of remaining in the study at time t depends only on the observed history $\overline{\boldsymbol{W}}_{i(j-1)}$ and is unrelated to the future outcomes. This assumption is weaker than the MAR assumption (Rubin, 1976), which is equivalent to

$$P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(T+1)}) = P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}).$$

In addition, we assume there exists a constant c such that

$$P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}) > c > 0, \tag{2.3}$$

which is known as the positivity assumption. We suppose that the missing probability $\pi_{ij} = P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)})$ is a known function of $\overline{\boldsymbol{W}}_{i(j-1)}$ up to a unknown finite dimension vector $\boldsymbol{\tau}$; i.e.,

$$P(R_{ij} = 1|R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}) = \pi_{ij}(\boldsymbol{\tau}), \tag{2.4}$$

where $\pi_{ij}(\cdot)$ is a known smooth function. For example, we can assume a logistic model and estimate $\boldsymbol{\tau}$ by maximizing the partial likelihood,

$$\prod_{i=1}^{n}\prod_{j=1}^{m}\left[\{\pi_{ij}(\boldsymbol{\tau})\}^{R_{ij}}\{1-\pi_{ij}(\boldsymbol{\tau})\}^{1-R_{ij}}\right]^{R_{i(j-1)}}.$$

Let $\overline{\pi}_{ij}(\boldsymbol{\tau}) = \pi_{i1}(\boldsymbol{\tau}) \times \cdots \times \pi_{ij}(\boldsymbol{\tau})$. When model (2.4) is correctly specified, $\overline{\pi}_{ij}(\boldsymbol{\tau}) = P(R_{ij} = 1|\overline{\boldsymbol{W}}_{i(j-1)})$. Thus, $\overline{\pi}_{ij}(\hat{\boldsymbol{\tau}})$ is a maximum partial likelihood estimator of the probability of subject i remaining in the study at time j given the observed history $\overline{\boldsymbol{W}}_{i(j-1)}$. Throughout this paper, we refer to model (2.4) as the missing data model.

Let $\boldsymbol{\Sigma}_i$ be the true covariance matrix of $\mathbf{Y}_i$, i.e., $\boldsymbol{\Sigma}_i = Var(\mathbf{Y}_i|\boldsymbol{X_i}, \boldsymbol{T}_i) = Var(\boldsymbol{\varepsilon}_i|\boldsymbol{X_i}, \boldsymbol{T}_i)$, where $\boldsymbol{\varepsilon}_i = \mathbf{Y}_i - \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{\theta}(\boldsymbol{T}_i)\}$. Here, $\boldsymbol{\mu}(\boldsymbol{z}) = (\mu(z_1), \cdots, \mu(z_m))^T$ is a matrix notation and we can similarly define other functions. Similarly to Liang and Zeger (1986), we define the working covariance matrix of $\boldsymbol{Y}_i$ as $\boldsymbol{V}_i(\boldsymbol{\zeta}) = \boldsymbol{A}_i^{-1/2}\boldsymbol{R}_i(\boldsymbol{\zeta})\boldsymbol{A}_i^{-1/2}$, where $\boldsymbol{A}_i$ is a diagonal matrix consisting of a scale parameter and a variance function of the mean and $\boldsymbol{R}_i(\boldsymbol{\zeta})$ is an invertible working correlation matrix depending on a finite nuisance parameter $\boldsymbol{\zeta}$.

In the presence of missing data, we estimate $\boldsymbol{V}_i(\boldsymbol{\zeta})$ from the second moment of the modified residual,

$$\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i} \equiv R_{im}\overline{\pi}_{im}^{-1}\boldsymbol{\varepsilon}_i - \sum_{k=1}^{m}(R_{ik} - \pi_{ik}R_{i(k-1)})\overline{\pi}_{ik}^{-1}[\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{i(k-1)}) - \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\theta}(\boldsymbol{T}_i)\}], \tag{2.5}$$

where $\boldsymbol{\phi}_k(\cdot)$ is a user-defined vector function in $\mathbb{R}^m$. We will provide further insight into this modified residual in Section 3.

2.2 LAIPW Kernel-Profile Estimating Equations

We focus on the local linear kernel estimator of $\theta(\cdot)$. Let $K_h(t) = h^{-1}K(t/h)$, where $K(\cdot)$ is a mean-zero symmetric density function. Without loss of generality, we assume $\int s^2K(s)ds = 1$. Define $\boldsymbol{G}_{ij}(t)$ as a $m \times 2$ matrix with the j-th row $(1, (T_{ij} - t)/h)$ and other rows as $(0, 0)$. We propose the following iterative LAIPW kernel-profile estimation procedure:

Step 1: Let $\tilde{\theta}(\cdot)$ to be the current estimator of $\theta(\cdot)$. For the fixed $\hat{\boldsymbol{\beta}}$ and any given t, solve the following LAIPW kernel estimating equations for $\boldsymbol{\alpha} = \boldsymbol{\alpha}(t, \hat{\boldsymbol{\beta}}) = (\alpha_0, \alpha_1)^T$,

$$\sum_{i=1}^{n}\sum_{j=1}^{m} K_h(T_{ij} - t)\mu^{(1)}_{ij,t}(\boldsymbol{\alpha})\boldsymbol{G}^T_{ij}(t)\boldsymbol{V}_i^{-1}\Bigg[R_{im}\overline{\pi}_{im}^{-1}(\hat{\boldsymbol{\tau}})\{\boldsymbol{Y}_i - \boldsymbol{\mu}^*_{ij,t}(\boldsymbol{\alpha})\} - \sum_{k=1}^{m}(R_{ik} - \pi_{ik}(\hat{\boldsymbol{\tau}})R_{i(k-1)})\overline{\pi}_{ik}^{-1}(\hat{\boldsymbol{\tau}})\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{i(k-1)}) - \boldsymbol{\mu}^*_{ij,t}(\boldsymbol{\alpha})\}\Bigg] = 0, \tag{2.6}$$

where $\boldsymbol{V}_i$ is a short for $\boldsymbol{V}_i(\hat{\boldsymbol{\zeta}})$, $\mu_{ij,t}^{(1)}(\boldsymbol{\alpha})$ is the first derivative of the function $\mu(\cdot)$ evaluated at $\boldsymbol{X}_{ij}^T\boldsymbol{\beta}+\alpha_0+\alpha_1(T_{ij}-t)/h$, and the l-th element of $\boldsymbol{\mu}_{ij,t}^{*}(\boldsymbol{\alpha})$ is

$$\mu\left\{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}+I(l=j)[\alpha_0+\alpha_1(T_{ij}-t)/h]+I(l\neq j)\tilde{\theta}(T_{il})\right\}.$$

Then, the first component of the vector $\hat{\boldsymbol{\alpha}}$ is the estimator of $\theta(t)$, which we denote as $\hat{\theta}(t,\hat{\boldsymbol{\beta}})$.

Step 2: For a fixed $\hat{\theta}(\cdot)$, we solve the following LAIPW profile estimating equations for $\boldsymbol{\beta}$,

$$\sum_{i=1}^{n}\left\{\frac{\partial\boldsymbol{\mu}[\boldsymbol{X}_i\boldsymbol{\beta}+\hat{\boldsymbol{\theta}}(\boldsymbol{T}_i,\beta)]}{\partial\boldsymbol{\beta}}\right\}^T\boldsymbol{V}_i^{-1}\Bigg[R_{im}\overline{\pi}_{im}^{-1}(\hat{\boldsymbol{\tau}})\{\boldsymbol{Y}_i-\boldsymbol{\mu}[\boldsymbol{X}_i\beta+\hat{\theta}(\boldsymbol{T}_i,\boldsymbol{\beta})]\}$$
$$-\sum_{k=1}^{m}(R_{ik}-\pi_{ik}(\hat{\boldsymbol{\tau}})R_{i(k-1)})\overline{\pi}_{ik}^{-1}(\hat{\boldsymbol{\tau}})\{\boldsymbol{\phi}_k(\overline{\boldsymbol{W}}_{i(k-1)})-\boldsymbol{\mu}[\boldsymbol{X}_i\beta+\hat{\boldsymbol{\theta}}(\boldsymbol{T}_i,\boldsymbol{\beta})]\}\Bigg]=0.$$

(2.7)

Repeat Step 1 and Step 2 until the relative change in the norm of the estimators for both $\theta(\cdot)$ and $\boldsymbol{\beta}$ falls below a pre-specified threshold. The initial values of $\boldsymbol{\beta}$ and $\theta(\cdot)$ can be obtained using various methods. In our implementation, we set the threshold to 10^{-4}, estimate the initial $\boldsymbol{\beta}$ using a generalized linear model with a linear component of T_{ij} , and obtain the initial $\theta(\cdot)$ using a LAIPW kernel method with a working independence correlation structure. The final estimates are defined as $\{\hat{\theta}(t),\hat{\boldsymbol{\beta}}\}$, where $\hat{\theta}(t)$ $=\hat{\theta}(t,\hat{\boldsymbol{\beta}})$. The second term within the summation in both (2.6) and

(2.7) are augmentation terms, which are derived from projecting the first term within the summation onto the nuisance tangent space. Removing the augmentation terms from (2.6) and (2.7) leads to the longitudinal inverse probability weighted kernel-profile estimating equations (LIPW-KPEE).

We postulate a parametric model

$$E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}] = \phi_{kl}(\overline{\boldsymbol{W}}_{i(k-1)}; \boldsymbol{\eta}), \tag{2.8}$$

which we refer to as the conditional mean model. Then, $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}(\cdot)$ are doubly robust. This double robustness property provides an additional opportunity to carry out the valid inference for the estimators in that $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}(\cdot)$ are consistent as long as either the missing data model (2.4) or the conditional mean model (2.8) is correctly specified. The choice of $\phi_{kl}(\cdot)$ is analogous to the optimal augmentation function considered in Robins and Rotnitzky (1992), where "optimal" refers to the function that leads to the most efficient estimator. In the Appendix, we verify that our specification of $\phi_{kl}(\cdot)$ is also optimal in our setting.

However, model (2.8) cannot be directly fitted in the presence of missing data. Noting that $E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}] = Y_{il}$ when $l < k$, we only need to consider the case where $l \geq k$. Robins and Rotnitzky (1992) proved that $E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}] = E[Y_{il}|R_{i(k-1)} = 1, \overline{\boldsymbol{W}}_{i(k-1)}]$ under the monotone missing pattern. In general, $E[Y_{il}|R_{il} = 1, \overline{\boldsymbol{W}}_{i(k-1)}] \neq E[Y_{il}|R_{i(k-1)} = 1, \overline{\boldsymbol{W}}_{i(k-1)}]$, so simply regressing Y_{il} on a function of $\overline{\boldsymbol{W}}_{k-1}$ among subjects who are still observed at time l will yield a biased estimator unless the missing mechanism is missing completely at random. To address this issue, we recommend estimating $E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}]$ with the sequential imputation method proposed by Paik (1997). Alternative methods are also available (Robins and Rotnitzky 1995, Bang and Robins 2005, Tsiatis et al. 2011, van der Laan and Gruber 2012).

3. Semiparametric Efficiency Bound

In this section, we present the semiparametrically efficient score and semiparametric efficiency bound under the multivariate normal assumption. In the case of full data, Wang et al. (2005) showed that the semiparametrically efficient score under the multivariate normal assumption is

$$\boldsymbol{S}_{eff}^{full} = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon},$$

where $\boldsymbol{\Sigma}$ is the conditional variance of $\boldsymbol{\varepsilon}$. The corresponding semiparametric efficiency bound is

$$\mathcal{V}_{eff}^{full} = E\left[\{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T \boldsymbol{\Sigma}^{-1} \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}\right],$$

where $\boldsymbol{\varphi}_{eff}(\cdot)$ satisfies

$$\sum_{j=1}^{m}\sum_{l=1}^{m} E\left[\sigma^{jl}\{\boldsymbol{X}_l - \boldsymbol{\varphi}_{eff}(T_l)\}|T_j = t\right] f_j(t) = 0$$

with σ^{jl} being the (j, l)-th element of $\boldsymbol{\Sigma}^{-1}$.

In the presence of missing data, we define $\boldsymbol{\varepsilon}_i^* = R_{im}\overline{\pi}_{im}^{-1}\boldsymbol{\varepsilon}_i - \sum_{j=1}^{m}(R_{ij} - \pi_{ij}R_{i(j-1)})\overline{\pi}_{ij}^{-1}E[\boldsymbol{\varepsilon}_i|\overline{\boldsymbol{W}}_{i(k-1)}]$, a modified residual using the optimal $\boldsymbol{\phi}_k(\cdot)$. Let $\boldsymbol{\Sigma}^*$ denote the conditional variance of $\boldsymbol{\varepsilon}^*$. Robins and Rotnitzky (1992) and Robins et al. (1994) developed theories that establish a connection between semiparametric regression on the full-data law and the observed-data law. This helps us to derive the observed-data semiparametrically efficient score. In the Appendix, we show that the observed-data semiparametrically efficient score under the multivariate normal assumption is

$$\boldsymbol{S}_{eff} = \{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T\boldsymbol{\Sigma}^{*,-1}\boldsymbol{\varepsilon}^*. \tag{3.9}$$

It then follows that the semiparametric efficiency bound $\mathcal{V}_{eff} = E[\boldsymbol{S}_{eff}\boldsymbol{S}_{eff}^T]^{-1}$ under the multivariate normal assumption is given by

$$\mathcal{V}_{eff} = E\left[\{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}^T\boldsymbol{\Sigma}^{*,-1}\{\boldsymbol{X} - \boldsymbol{\varphi}_{eff}(\boldsymbol{t})\}\right]. \tag{3.10}$$

It is noteworthy that the full-data semiparametric efficiency bound and the observed-data semiparametric efficiency bound only diverge in the conditional variance of the residual. This result is straightforward since $\boldsymbol{\varepsilon}$ is a biased estimator in the presence of the missing data, and we need a modified residual $\boldsymbol{\varepsilon}^*$ to correct this bias.

4. Asymptotic Properties of the LAIPW Kernel-profile Estimator

We use $\boldsymbol{\beta}_0$ and $\theta_0(\cdot)$ to denote the true values of $\boldsymbol{\beta}$ and $\theta(\cdot)$, respectively. Let $\boldsymbol{\tau}^*$, $\boldsymbol{\eta}^*$, and $\boldsymbol{\zeta}^*$ be the probability limit of $\widehat{\boldsymbol{\tau}}$, $\widehat{\boldsymbol{\eta}}$, and $\widehat{\boldsymbol{\zeta}}$, v_i^{jl} be the (j,l)-th element of the matrix $\boldsymbol{V}_i^{-1}$, $f_j(t)$ be the marginal density of T_{ij}, and $f_{jl}(t_j,t_l)$ be the joint density of T_{ij} and T_{il} for any i. Let $\boldsymbol{\Delta}$ be a diagonal matrix with the j-th diagonal element being the first derivative of the function $\mu(\cdot)$ evaluated at $\boldsymbol{X}_j^T\boldsymbol{\beta}_0+\theta_0(t)$.

We first study the asymptotic property of $\hat{\theta}(t,\boldsymbol{\beta}_0)$.

Theorem 1. *Under the assumption of* (2.2) *and* (2.3), *if either model* (2.4) *or model* (2.8) *is correctly specified, as* $h\to 0$, $n\to\infty$, *and* $\log(n)/nh\to 0$, *the asymptotic bias of* $\hat{\theta}(t,\boldsymbol{\beta}_0)$ *is* $b^*(t)h^2/2$, *where* $b^*(t)$ *satisfies*

$$b^*(t)=\theta_0^{(2)}(t)-W_2^{-1}(t)\sum_{j=1}^{m}\sum_{l\neq j}^{m}E[\Delta_{jj}v^{jl}\Delta_{ll}b^*(T_l)|T_j=t]f_j(t), \qquad (4.11)$$

with $W_2(t)=\sum_{j=1}^{m}E[\Delta_{jj}^2v^{jj}|T_j=t]f_j(t)$.

Theorem 1 establishes the double robustness property of $\hat{\theta}(t,\boldsymbol{\beta}_0)$. Specifically, $\hat{\theta}(t,\boldsymbol{\beta}_0)$ is asymptotically unbiased if either the missing data model or the conditional mean model is correctly specified. Compared to the LIPW-KPEE method, our LAIPW-KPEE method allows the missing data model to be misspecified as long as the conditional mean model is correctly specified.

Next, we aim to study the asymptotic property of $\hat{\boldsymbol{\beta}}$. Before that, we define $\hat{\boldsymbol{\varphi}}(t, \boldsymbol{\beta}) = -\partial\hat{\boldsymbol{\theta}}(\boldsymbol{t}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$ and $\boldsymbol{\varphi}(t)$ as the probability limit of $\hat{\boldsymbol{\varphi}}(t, \boldsymbol{\beta}_0)$. The following Lemma presents the form of $\boldsymbol{\varphi}(t)$.

Lemma 1. *Assume $\hat{\boldsymbol{\theta}}(t, \cdot)$ is continuously differentiable, and its derivative is bounded in a neighborhood of $\boldsymbol{\beta}_0$. Under the assumption of* (2.2) *and* (2.3)*, if either model* (2.4) *or model* (2.8) *is correctly specified, then $\boldsymbol{\varphi}(t)$ solves the following equations,*

$$\sum_{j=1}^{m}\sum_{l=1}^{m} E\left[\Delta_{jj}v^{jl}\Delta_{ll}\{\boldsymbol{X}_l - \boldsymbol{\varphi}(T_l)\}|T_j = t\right] f_j(t) = 0, \tag{4.12}$$

where Δ_{jj} is the (j, j)-th element of the matrix $\boldsymbol{\Delta}$. Furthermore, if the working model $\boldsymbol{V}(\boldsymbol{\zeta})$ is correctly specified, we have $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}_{eff}(t)$.

Lemma 1 shows that $\boldsymbol{\varphi}(t)$ happens to be $\boldsymbol{\varphi}_{eff}(t)$ if the working correlation model and either the missing data model or the conditional mean model are correctly specified. This result builds a bridge between the LAIPW-KPEE profile estimating equations (2.7) and semiparametrically efficient score (3.9). After some calculations, equation (4.12) can be written as the

Fredholm integral equation of the second kind,

$$\boldsymbol{\varphi}(t) = \boldsymbol{q}(t) + \int \boldsymbol{H}(t,s)\boldsymbol{\varphi}(s)ds, \tag{4.13}$$

where

$$\boldsymbol{H}(t,s) = \frac{\sum_j \sum_{l \neq j} E\left[\Delta_{jj} v^{jl} \Delta_{ll} | T_l = s, T_j = t\right] f_{lj}(s,t)}{\sum_{j=1}^m E\left[v^{jj}\Delta_{jj}^2 | T_j = t\right] f_j(t)}$$

and

$$q(t) = \frac{\sum_{j=1}^m \sum_{l=1}^m E\left[\Delta_{jj} v^{jl} \Delta_{ll} \boldsymbol{X}_l | T_j = t\right] f_j(t)}{\sum_{j=1}^m E\left[v^{jj}\Delta_{jj}^2 | T_j = t\right] f_j(t)}.$$

When $\boldsymbol{H}(t,s)$ is square-integrable and -1 is not a eigenvalue of (4.13), the solution of (4.13) is unique. However, a general closed-form solution does not exist, and numerical techniques are typically required to approximate the solution. We will not delve into the specifics of these numerical methods here and refer readers to relevant literature (e.g., Atkinson 1997, Pearson 2012, Sastry 2012). Alternatively, one may assume a working independence correlation structure, leading to the result that

$$\boldsymbol{\varphi}(t) = \frac{\sum_{j=1}^m \sum_{l=1}^m E\left[\Delta_{jj} v^{jl} \Delta_{ll} \boldsymbol{X}_l | T_j = t\right]}{\sum_{j=1}^m E\left[v^{jj}\Delta_{jj}^2 | T_j = t\right]}.$$

With the forms of $\boldsymbol{\varphi}(t)$, we can straightforwardly derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. Let $\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_i - \boldsymbol{\varphi}(\boldsymbol{T}_i)$,

$$\begin{aligned}\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi},i}(\boldsymbol{\tau},\boldsymbol{\eta},\boldsymbol{\beta},\theta) = {} & R_{im}\overline{\pi}_{im}^{-1}(\boldsymbol{\tau})[\boldsymbol{Y}_i - \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta} + \theta(t)\}] \\ & - \sum_{j=1}^m (R_{ij} - \pi_{ij}(\boldsymbol{\tau})R_{i(j-1)})\overline{\pi}_{ij}^{-1}(\boldsymbol{\tau})[\boldsymbol{\phi}^*_j(\overline{\boldsymbol{W}}_{i(j-1)};\boldsymbol{\eta}) - \boldsymbol{\mu}\{\boldsymbol{X}_i\boldsymbol{\beta} + \theta(t)\}]\end{aligned}$$

and $\boldsymbol{D}(\boldsymbol{\beta},\boldsymbol{\zeta}) = \{\partial\boldsymbol{\mu}[\boldsymbol{X}\boldsymbol{\beta}+\hat{\boldsymbol{\theta}}(\boldsymbol{T},\boldsymbol{\beta})]/\partial\boldsymbol{\beta}^T\}\boldsymbol{V}^{-1}(\boldsymbol{\zeta})$. We use $\boldsymbol{S}(\boldsymbol{R},\boldsymbol{W}_{obs};\boldsymbol{\tau})$ and $\boldsymbol{l}(\boldsymbol{W}_{obs};\boldsymbol{\eta})$ to denote the estimating functions for $\boldsymbol{\tau}$ and $\boldsymbol{\tau}$, where $\boldsymbol{W}_{obs}$ denote the observed data.

Theorem 2. *Under the assumption of* (2.2) *and* (2.3)*, if either model* (2.4) *or model* (2.8) *is correctly specified, as* $h\to 0$*,* $n\to\infty$*,* $nh^8\to 0$*, and* $nh/\log(1/h)\to\infty$*, we have*

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\to N(\boldsymbol{0},\boldsymbol{\Omega}_{\boldsymbol{\phi}}(\boldsymbol{V})), \tag{4.14}$$

where $\boldsymbol{\Omega}_{\boldsymbol{\phi}}(\boldsymbol{V}) = \boldsymbol{A}^{-1}(\boldsymbol{V})\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V})\boldsymbol{A}^{-1}(\boldsymbol{V})$*,* $\boldsymbol{A}(\boldsymbol{V}) = E[\tilde{\boldsymbol{X}}^T\boldsymbol{\Delta}V^{-1}\boldsymbol{\Delta}\tilde{\boldsymbol{X}}]$*, and*

$$\begin{aligned}\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V}) = Var\Big\{&\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi}}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\\ &- E\left[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\frac{\partial}{\partial\boldsymbol{\tau}}\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi}}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right]E\left[\frac{\partial}{\partial\boldsymbol{\tau}}\boldsymbol{S}(\boldsymbol{R},\boldsymbol{W}_{obs};\boldsymbol{\tau}^*)\right]^{-1}\boldsymbol{S}(\boldsymbol{R},\boldsymbol{W}_{obs};\boldsymbol{\tau}^*)\\ &- E\left[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\frac{\partial}{\partial\boldsymbol{\eta}}\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi}}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)\right]E\left[\frac{\partial}{\partial\boldsymbol{\eta}}\boldsymbol{l}(\boldsymbol{W}_{obs};\boldsymbol{\eta}^*)\right]^{-1}\boldsymbol{l}(\boldsymbol{W}_{obs};\boldsymbol{\eta}^*)\Big\}.\end{aligned}$$

It is worth noting that $E[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\partial\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi}}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)/\partial\boldsymbol{\eta}] = 0$ if the missing data model is correctly specified and $E[\boldsymbol{D}(\boldsymbol{\beta}_0,\boldsymbol{\zeta}^*)\partial\boldsymbol{\varepsilon}^*_{\boldsymbol{\phi}}(\boldsymbol{\tau}^*,\boldsymbol{\eta}^*,\boldsymbol{\beta}_0,\theta_0)/\partial\boldsymbol{\tau}] = 0$ if the conditional mean model is correctly specified. Thus, $\boldsymbol{B}_{\boldsymbol{\phi}}(\boldsymbol{V})$ can be simplified if the missing data model is correctly specified, the conditional mean model is correctly specified, or both are correctly specified. Theorem 2 presents the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, which also possesses the property of double robustness. Furthermore, the efficiency of $\hat{\boldsymbol{\beta}}$ from our

LAIPW-KPEE method does not require a working independence correlation structure or undersmooth estimation of $\hat{\theta}(t)$. Following the law of large numbers, the covariance of $\hat{\boldsymbol{\beta}}$ can be consistently estimated by evaluating all quantities at $\hat{\boldsymbol{\beta}}$, $\hat{\theta}(t)$, $\hat{\boldsymbol{\tau}}$, $\hat{\boldsymbol{\eta}}$, and $\hat{\boldsymbol{\zeta}}$. For the estimation of the variance of $\hat{\theta}(t)$, if necessary, we recommend using a bootstrap method.

Let $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\phi}}(\boldsymbol{V})$ denote the asymptotic variance of $\hat{\boldsymbol{\beta}}$ if the true π_{ij} are known and used in LAIPW-KPEE method. In addition, let $\boldsymbol{\phi}_{opt}$ denote that the conditional mean model is correctly specified. Then, the following Corollary compares the asymptotic variance of $\hat{\boldsymbol{\beta}}$ under the scenario of estimated $\pi_{ij}(\hat{\boldsymbol{\tau}})$ and true π_{ij} and discusses the choice of $\boldsymbol{\phi}$ and $\boldsymbol{V}$.

Corollary 1. *Under the assumption of* (2.2) *and* (2.3)*, if either model* (2.4) *or model* (2.8) *is correctly specified, then we have*

(a) $\boldsymbol{\Omega}_{\boldsymbol{\phi}}(\boldsymbol{V}) \leq \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\phi}}(\boldsymbol{V})$*;*

(b) $\boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{V}) = \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{V})$*;*

(c) For any $\boldsymbol{\phi}$, $\boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{V}) \leq \boldsymbol{\Omega}_{\boldsymbol{\phi}}(\boldsymbol{V})$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{V}) \leq \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\phi}}(\boldsymbol{V})$*;*

(d) For any $\boldsymbol{V}$, $\boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{V}) \geq \boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{\Sigma}^*)$*. Furthermore,* $\boldsymbol{\Omega}_{\boldsymbol{\phi}_{opt}}(\boldsymbol{\Sigma}^*) = \boldsymbol{A}(\boldsymbol{\Sigma}^*)^{-1}$.

Part (a) of Corollary 1 states that estimating π_{ij} helps improve the efficiency of $\hat{\boldsymbol{\beta}}$ even π_{ij} is known. This phenomenon is also found in other semiparametric estimating procedures (Robins et al. 1995; Wang et al. 2024).

However, an exception is noted in part (b) of Corollary 1. It reveals that we no longer have this efficiency gain if $\boldsymbol{\phi}_{opt}$ is adopted. Part (c) of Corollary 1 shows that the optimal choice of ϕ_{kl} is $E[Y_{il}|\overline{\boldsymbol{W}}_{i(k-1)}]$. Part (d) of Corollary 1 claims that the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is minimized when the missing data model, the conditional mean model, and the working correlation model are correctly specified simultaneously. In addition, it reaches the semiparametric efficiency bound, i.e., $\boldsymbol{\Omega}_{\phi_{opt}}(\boldsymbol{\Sigma}^*) = \mathcal{V}_{eff}$, under the multivariate normal assumption. Thus, $\hat{\boldsymbol{\beta}}$ is the most efficient estimator under the multivariate normal assumption.

Since local linear smoothing is employed, the choice of bandwidth plays a crucial role in balancing the bias-variance tradeoff. A small bandwidth reduces bias but increases variance due to overfitting, while a larger bandwidth lowers variance but leads to higher bias. A common method for selecting bandwidth is the empirical bias bandwidth selection (EBBS) method (Ruppert, 1997). In fact, Theorem 2 shows that $\hat{\boldsymbol{\beta}}$ is robust to the choice of bandwidth, provided it is appropriately selected.

5. Simulation Results

In this section, we conduct a simulation study to evaluate and compare the finite-sample performances of three estimators: the LAIPW kernel-profile estimator, the LIPW kernel-profile estimator, and the naive estimator. Specifically, the LIPW kernel-profile estimator is obtained from the LIPW kernel-profile estimating equations, where the augmentation terms in (2.6) and (2.7) are omitted. The naive estimator is derived based on kernel-profile estimating equations using only complete cases, as described in Wang et al. (2005). We generate 100 datasets with sample size $n = 500$ and number of post-baseline measurements $m = 3$. In total, four covariates are included: T_{ij}, X_{1ij}, X_{2i}, and an auxiliary covariate U_{ij}, where X_{2i} is a time-fixed variable while the others are time-varying variables. Specifically, $T_{ij} \sim Z_{1i} + e_{1ij}$, $X_{1ij} \sim Z_{1i} + e_{2ij}$, $X_{2i} \sim \text{Bernoulli}(0.5)$, and $U_{ij} \sim Z_{2i} + \text{Normal}(T_{ij}, 0.05) + X_{2i}$, where $Z_{1i} \sim \text{Uniform}(-1, 1)$ and $Z_{2i} \sim \text{Uniform}(0, 4)$ are two independent time-fixed variables while $e_{1ij} \sim \text{Uniform}(-1, 1)$ and $e_{2ij} \sim \text{Uniform}(-1, 1)$ are two independent time-varying variables. This data generating process makes T_{ij}, X_{2ij}, and U_{ij} correlated. Let $\boldsymbol{X}_{ij} = (X_{1ij}, X_{2i})$. The response variable Y_{ij} is generated from a normal distribution with mean

$$E[Y_{ij}|\boldsymbol{X}_{ij}, T_{ij}, U_{ij}] = m(T_{ij}) + \boldsymbol{X}_{ij}^T\tilde{\boldsymbol{\beta}} + U_{ij}$$

and exchangeable covariance matrix with marginal variance 1 and correlation coefficient 0.6, where $m(T_{ij}) = \sin(2T_{ij})$ and $\tilde{\boldsymbol{\beta}} = (1, 3)^T$. Recall that our primary interest is $E[Y_{ij}|\boldsymbol{X}_{ij}, T_{ij}]$ instead of $E[Y_{ij}|\boldsymbol{X}_{ij}, T_{ij}, U_{ij}]$. It is easy to show that

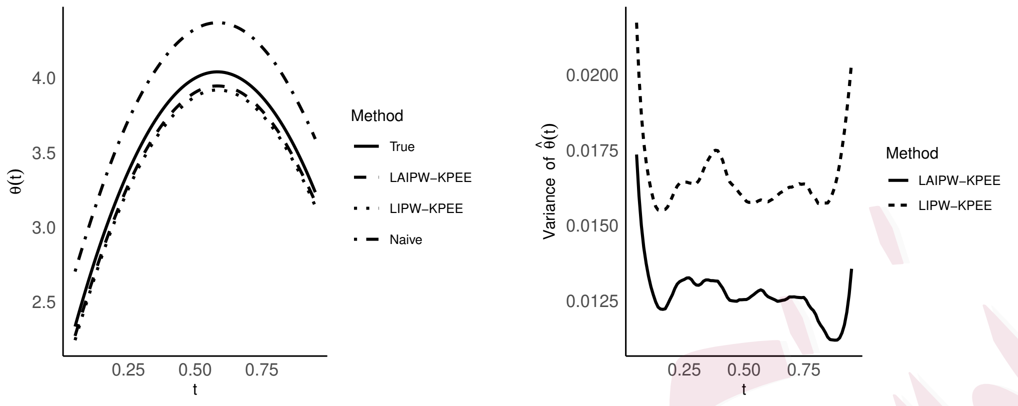
$$E[Y_{ij}|\boldsymbol{X}_{ij}, T_{ij}] = \theta(T_{ij}) + \boldsymbol{X}_{ij}^T\boldsymbol{\beta},$$

where $\theta(T_{ij}) = \sin(2T_{ij}) + T_{ij}$ and $\boldsymbol{\beta} = (1, 4)^T$. The missing probability at time j follows a logistic model

$$\text{logit}\{\pi_{ij}\} = (U_{i(j-1)} - 0.5)I(0.5 < U_{i(j-1)} \leq 3.5) + I(U_{i(j-1)} > 3.5). \tag{5.15}$$

This missing mechanism indicates that missingness depends only on the observed variable $U_{i(j-1)}$, and consequently assumption (2.2) holds. The Monte Carlo mean of the missing percentage is 19%. Throughout the simulation, we adopt the Epanechnikov kernel function and select bandwidth with the leave-one-out cross-validation method.

Figure 1(a) displays the empirical mean of the estimated $\hat{\theta}(\cdot)$ over 100 replications when both the missing data model and the conditional mean model are correctly specified. The curves for both the LAIPW kernel-profile estimator and the LIPW kernel-profile estimator closely follow the true curve, with the LAIPW kernel-profile estimator demonstrating an even smaller point-wise bias. In contrast, the curve for the naive estimator significantly deviates from the true curve. Figure 1(b) visualizes the corresponding empirical variance of $\hat{\theta}(\cdot)$ for the LAIPW kernel-profile estimator



(a) (b)

Figure 1: (a) Comparison of the true $\theta(t)$ and point-wise empirical mean of the nonparametric functions $\hat{\theta}(t)$. (b) Comparison of point-wise empirical variance of the nonparametric functions $\hat{\theta}(t)$.

and the LIPW kernel-profile estimator, highlighting the higher efficiency of the LAIPW kernel-profile estimator over the LIPW kernel-profile estimator at every point of t.

Table 1 summarizes the bias, estimated standard error, empirical standard error, and empirical mean squared error (MSE) of $\boldsymbol{\beta}$, along with the empirical mean integrated mean squared error (MISE) of $\theta(\cdot)$ for each method over 100 replications. Here, empirical MISE is defined as $\int\{\hat{\theta}(t,\boldsymbol{\beta})-\theta_0(t)\}^2 dF(t)$, where $F(t)$ is the cumulative distribution function of t. Apart from the scenarios where π and ϕ are known or estimated with correctly specified models, we further consider the following three scenarios: i) The conditional mean model is correctly specified, but the missing data model is misspecified with model $\mathrm{logit}\{\pi_{ij}(\tau')\} = \tau'$; ii) The missing data model is correctly specified, but the conditional mean model is misspecified with model $\phi_{kl}(\overline{\boldsymbol{W}}_{i(k-1)}; \boldsymbol{\eta}') = \eta_0' + \eta_1' T_{il} + \eta_2' X_{1il}$; iii) Both the missing data model and the conditional mean model are misspecified with the models in i) and ii). In Table 1, we observe a significant bias of the naive estimator in β_2. This bias arises from the collinearity between covariate X_{2ij} and auxiliary variable U_{ij}. When either π and ϕ are known, or both the missing data model and the conditional mean model are correctly specified, the LIPW kernel-profile estimator and the LAIPW kernel-profile estimator effectively reduce the estimation bias. The double robustness property of the LAIPW kernel-profile estimator is observed from the fact that the empirical bias of $\hat{\beta}_1$ and $\hat{\beta}_2$ are small if either the missing data model or the conditional mean model is correctly specified. In contrast, the LIPW kernel-profile estimator shows a large bias when the missing data model is misspecified. Furthermore, we observe an efficiency gain of the LAIPW kernel-profile estimator in terms of both standard error and MSE relative to the LIPW kernel-profile estimator. Even under the scenario of both the missing data model and the conditional mean model being misspecified, the

LAIPW kernel-profile estimator still has a comparable performance relative to the LIPW kernel-profile estimator and naive estimator. In terms of the empirical MISE of $\hat{\theta}(\cdot)$, if one of the models is correctly specified, the naive estimator has the largest MISE, while that of the LAIPW kernel-profile estimator is generally smaller than that of the LIPW kernel-profile estimator. In summary, the LAIPW-KPEE method is more efficient and less sensitive to model misspecification.

In practice, the missing pattern may depend on the unobserved values themselves, which is known as missing not at random (MNAR, Rubin 1976). To explore the robustness of the LAIPW-KPEE method when the missingness assumption (2.4) is violated, we conduct a sensitivity analysis under a similar setting. Specifically, we allow the missingness to depend on the outcome at the current time point,

$$\mathrm{logit}\{\pi_{ij}\} = (U_{i(j-1))} - 0.5)I(0.5 < U_{i(j-1))} \leq 3.5) + I(U_{i(j-1))} > 3.5) + \gamma Y_{ij},$$

where γ is a coefficient that controls the degree of deviation from the MAR mechanism. Clearly, this missing data generating process implies an MNAR mechanism, except when $\gamma = 0$, which corresponds to a MAR mechanism. The missing data model is constructed based on (5.15) for the estimation of π_{ij}, leading to a biased estimate.

Based on the range of the outcome Y_{ij}, we select γ to vary from -0.5

Table 1: Comparison of naive, LIPW-KPEE, and LAIPW-KPEE estimators in terms of bias, estimated standard error (EST S.E.), empirical standard error (EMP S.E.), and empirical mean squared error (EMP MSE) of $\hat{\boldsymbol{\beta}}$ and empirical mean integrated mean squared error (EMP MISE) of $\hat{\theta}(\cdot)$ based on 100 replications of the simulation.

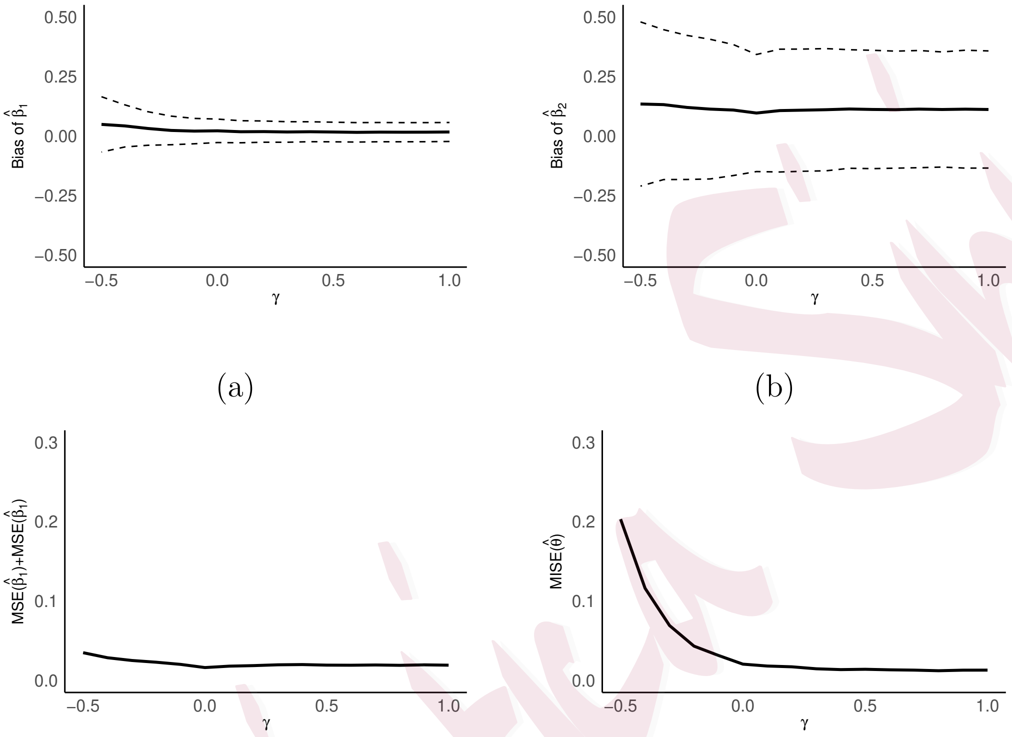
	$\boldsymbol{\beta_1 = 1}$				$\boldsymbol{\beta_2 = 4}$				$\boldsymbol{\theta(\cdot)}$
	Bias of $\hat{\beta}_1$	EST S.E.	EMP S.E.	EMP MSE	Bias of $\hat{\beta}_2$	EST S.E.	EMP S.E.	EMP MSE	EMP MISE
Naive	0.020	0.017	0.024	0.001	0.427	0.136	0.139	0.201	0.129
LIPW-KPEE									
True π	0.024	0.026	0.029	0.001	0.117	0.174	0.152	0.023	0.037
Consistent $\hat{\pi}$	0.025	0.025	0.029	0.001	0.112	0.156	0.141	0.020	0.030
Wrong $\hat{\pi}$	0.021	0.024	0.025	0.001	0.458	0.158	0.155	0.236	0.182
LAIPW-KPEE									
True π and ϕ	0.014	0.020	0.017	0.001	0.098	0.129	0.126	0.016	0.023
Consistent $\hat{\pi}$ and $\hat{\phi}$	0.022	0.025	0.027	0.001	0.096	0.135	0.125	0.016	0.021
Wrong $\hat{\pi}$	0.020	0.026	0.025	0.001	0.107	0.139	0.133	0.018	0.021
Wrong $\hat{\phi}$	0.021	0.026	0.025	0.001	0.101	0.144	0.127	0.016	0.025
Both wrong	0.027	0.0333	0.032	0.001	0.358	0.139	0.130	0.145	0.204

to 1. The average bias of $\hat{\beta}_1$ and $\hat{\beta}_2$ with 95% confidence interval over 100 replications are shown in Figure 2 (a) and (b). We also display the changes of the MSE of $\hat{\beta}_1$ plus the MSE of $\hat{\beta}_2$, as well as the changes of MISE of $\hat{\theta}$ in Figure 2 (c) and (d). We only observe a minor difference between MAR and MNAR in terms of bias and MSE when $\gamma > 0$. However, a larger bias and MSE is found when $\gamma < 0$. This is not entirely attributed to the MNAR mechanism, as the sample size also decreases rapidly as γ decreases. To some extend, the LAIPW-KPEE method still has a comparable performance even under MNAR.

Additional simulation results are provided in the Appendix, including a comparison of computational time across different estimators and a simulation study under a smaller sample size setting ($n = 100$).

6. Application

To illustrate the validity of the LAIPW-KPEE method in practice, we applly our method to the LEOPARD study. The objective of the LEOPARD study is to examine the dynamic progress of early initiation of antiretroviral treatment (ART) in perinatally HIV-infected infants. This study includes a cohort of 122 perinatally infected infants enrolled at the Rahima Moosa Mother and Child Hospital in Johannesburg, South Africa, between 2014



(a) (b)

(c) (d)

Figure 2: Sensitivity analysis of the LAIPW-KPEE method under MNAR. (a), (b) display the bias of $\hat{\beta}_1$ and $\hat{\beta}_2$ with 95% confidence interval as γ changes. (c), (d) display the sum of MSE of $\hat{\beta}_1$ and $\hat{\beta}_2$, and MISE of $\hat{\theta}$ as γ changes.

and 2017.

In this paper, we are interested in identifying the predictors of CD4 count changes over days of age among the infants with initiation of ART. Considering that the majority of infants initiated ART within two weeks of birth (median age: 3 days; interquartile range (IQR): 1–7 days) and the long time span of the study, it is reasonable to approximate the time to initiation of ART by age. Preliminary analysis indicates a non-linear relationship between CD4 count and age (in days). This motivates us to include age (in days) in the nonparametric part of the LAIPW-KPEE method. In addition, we also include sex, mean-centered birth weight, preterm, delivery mode, maternal prenatal ART history, mother's CD4 count, and breastfeeding status as the covariates in the parametric part. We use the $\log_{10}$ transformed viral load (VL) as the auxiliary variable, which, compared to the CD4 count, can be obtained more quickly and easily from a blood sample. Due to limited data, we restrict our analysis to the first two years. By utilizing observations with dropout missing and excluding infants with missing covariates, we finally obtain a dataset consisting of 110 infants and 553 observations with a 21.9% missing rate on the CD4 count. With 10-fold cross-validation, we select the bandwidth that minimizes the average MSE.

The additional details and results for the missing data model and the conditional mean model are summarized in the Appendix. Furthermore, we conduct model diagnostics to assess the fit and validity of both models and find that they are reasonable.

Figure 3 shows the curve of $\theta(\text{age})$ estimated by the naive method, LIPW-KPEE method, and LAIPW-KPEE method, adjusting for other covariates. Notably, we observe a change point around 250 days. This indicates the effective protection of ART on the immune system for over half a year, after which its effect starts to diminish. When comparing the curves produced by the three methods, we observe an upward shift in the curve of the naive method. This can be attributed to selection bias, where lower CD4 counts typically reflect more advanced HIV progression and are associated with a higher risk of dropout. Consequently, individuals who remain in the study tend to be healthier. Since the naive method only relies on complete cases, it may overestimate the CD4 count trajectory. In contrast, the curves from the LIPW-KPEE and LAIPW-KPEE methods align closely, as these methods correct for selection bias by recovering additional information from incomplete cases.

Table 2 presents the $\hat{\boldsymbol{\beta}}$ derived from the naive method, LIPW-KPEE method, and LAIPW-KPEE method. The results of LIPW kernel-profile estimates and LAIPW kernel-profile estimates reveal a significant influence

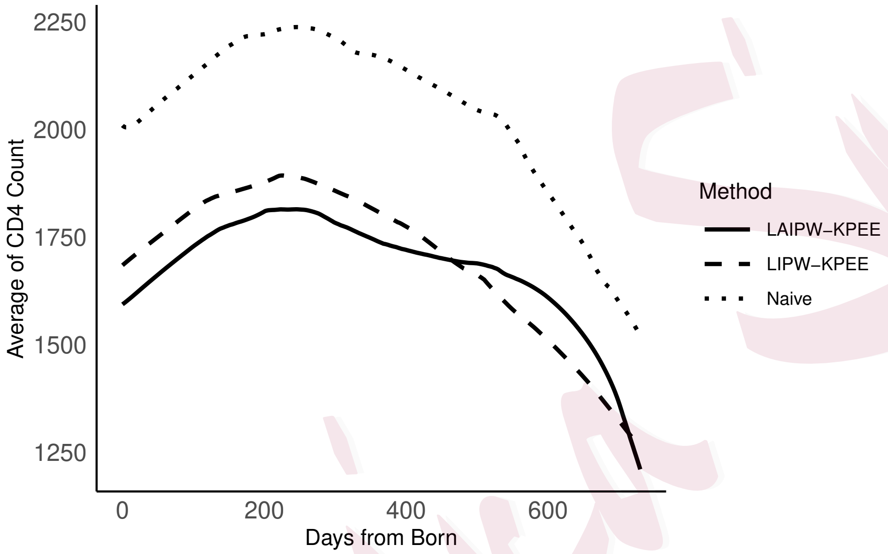


Figure 3: Average CD4 count over time, adjusting for other covariates, in the LEOPARD study derived from naive method, LIPW-KPEE method, and LAIPW-KPEE method.

Table 2: Partially linear model coefficients $\boldsymbol{\beta}$ on CD4 count in the LEOPARD study, estimated by naive method, LIPW-KPEE method, and LAIPW-KPEE method.

	LAIPW-KPEE			**LIPW-KPEE**			**Naive**		
Predictors	$\hat{\beta}$	S.E.	p-value	$\hat{\beta}$	S.E.	p-value	$\hat{\beta}$	S.E.	p-value
male vs. female	-79.23	162.46	0.63	-160.81	161.78	0.32	-130.29	153.37	0.40
birth weight	0.22	0.17	0.20	0.18	0.14	0.20	0.21	0.15	0.16
preterm vs. full term	268.25	362.68	0.46	226.28	249.68	0.36	250.26	253.68	0.32
maternal prenatal ART history									
No ART up until delivery		*reference*			*reference*			*reference*	
during pregnancy, $<$ 12 weeks	463.88	168.24	0.01	489.57	180.49	0.01	225.08	171.67	0.19
during pregnancy, 12+ weeks	662.64	176.58	$<$ 0.01	721.69	195.87	$<$ 0.01	490.91	174.70	$<$ 0.01
before pregnancy and continued	521.03	269.67	0.05	564.73	210.51	0.01	364.67	206.81	0.08
normal delivery vs. caesarean section	-108.67	186.31	0.56	-173.19	191.30	0.37	-221.22	160.92	0.17
some breastfeeding vs. no breastfeeding	-200.50	187.13	0.28	-202.25	155.29	0.19	-52.52	147.43	0.72
mother's CD4 count	0.36	0.40	0.36	0.58	0.33	0.07	0.46	0.30	0.13

of a mother's prenatal ART history on an infant's CD4 count. In contrast, the naive estimates find an association only in cases where mothers initiate ART more than 12 weeks prior to pregnancy, resulting in a higher CD4 count in infants.

7. Discussion

In this paper, we introduce a semiparametric regression model tailored for longitudinal studies with monotone missing outcomes. Our proposed LAIPW-KPEE method is specifically crafted to address the nuances of longitudinal studies. In the context of cross-sectional studies, our method converges to the AIPW method as proposed by Wang et al. (2024). The LAIPW kernel-profile estimator leverages auxiliary variables, which, while not the primary focus of the model, aid in explaining the missing data.

We demonstrate that the LAIPW kernel-profile estimator is consistent if either the missing data model for $P(R_{ij} = 1 | R_{i(j-1)} = 1, \overline{\boldsymbol{W}}_{i(j-1)}) = \pi_{ij}(\boldsymbol{\tau})$ or the conditional mean model for $E[Y_{il} | \overline{\boldsymbol{W}}_{i(k-1)}]$ is correctly specified. This doubly robust property is attributed to the incorporation of weights and augmentation terms in both the kernel estimating equations for $\theta(\cdot)$ and the profile estimating equations for $\boldsymbol{\beta}$. The modified kernel estimator facilitates the efficiency of $\boldsymbol{\beta}$ without necessitating assumptions regarding a working independence correlation structure or employing an under-smoothed kernel estimator. Furthermore, the LAIPW-KPEE method generates a kernel estimator of $\theta(\cdot)$ with reduced variation, while preserving $\boldsymbol{\beta}$ as the most efficient estimator under the multivariate normal assumption.

Throughout this paper, several assumptions are imposed to ensure that the proposed methodology possesses desirable theoretical properties. For ease of notation, we primarily focus on the setting where the number of measurements is equal across subjects. Nevertheless, our method can be easily extended to unbalanced data by incorporating subject-specific dimension of the working covariance structures in the estimating equations, which does not affect the consistency or asymptotic properties of the proposed estimators. In addition, we assume that the missingness only occurs in response variable and auxiliary variables, which are either both observed or both missing at any given time point. However, our results remain valid even when auxiliary variables are always observed. Our method is not applicable when missingness occurs in the covariates and we refer interested readers to other literature (e.g., Liang 2008; Qin et al. 2012) for further extensions, which would be an interesting topic for future research. Meanwhile, we assume the missing data pattern is monotone. For more general missing patterns, one may "artificially" impose a monotone structure by discarding all subsequent measurements after the first missing time point. It can be shown that such intentional deletion does not affect the efficiency of our proposed method. A more detailed discussion of this extension can be found in Robins et al. (1995) and Robins and Rotnitzky (1995). Another adjustable assumption is that the scalar T in the nonparametric part can be extended to a set of time-varying covariates with an additive model. To enhance computational efficiency and streamline derivation, we employ the local linear kernel estimator of $\theta(\cdot)$ in the kernel estimating equations, but actually this can be extended to the local polynomial kernel estimator and the asymptotic results will be similar.

References

Atkinson, K. E. (1997). *The numerical solution of integral equations of the second kind*, Volume 4. Cambridge university press.

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics 61*(4), 962–973.

Chen, B. and X.-H. Zhou (2013). Generalized partially linear models for incomplete longitudinal data in the presence of population-level information. *Biometrics 69*(2), 386–395.

Du, J., Y. Li, and X. Cui (2023). Identification and estimation of generalized additive partial linear models with nonignorable missing response. *Communications in Mathematics and Statistics*, 1–44.

Engle, R. F., C. W. Granger, J. Rice, and A. Weiss (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association 81*(394), 310–320.

Härdle, W., H. Liang, and J. Gao (2012). *Partially linear models*. Springer Science & Business

Media.

Kuhn, L., R. Strehlau, S. Shiau, F. Patel, Y. Shen, K.-G. Technau, M. Burke, G. Sherman, A. Coovadia, G. M. Aldrovandi, et al. (2020). Early antiretroviral treatment of infants to attain hiv remission. *EClinicalMedicine 18*.

Liang, H. (2008). Generalized partially linear models with missing covariates. *Journal of multivariate analysis 99*(5), 880–895.

Liang, H., S. Wang, and R. J. Carroll (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika 94*(1), 185–198.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Lin, H., B. Fu, G. Qin, and Z. Zhu (2017). Doubly robust estimation of generalized partial linear models for longitudinal data with dropouts. *Biometrics 73*(4), 1132–1139.

Lin, X. and R. J. Carroll (2001a). Semiparametric regression for clustered data. *Biometrika 88*(4), 1179–1185.

Lin, X. and R. J. Carroll (2001b). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association 96*(455), 1045–1056.

Lin, X., N. Wang, A. H. Welsh, and R. J. Carroll (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika 91*(1), 177–193.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association 92*(440), 1320–1329.

Pearson, C. (2012). *Handbook of applied mathematics: selected results and methods*. Springer Science & Business Media.

Qin, G., Z. Zhu, and W. K. Fung (2012). Robust estimation of the generalised partial linear model with missing covariates. *Journal of Nonparametric Statistics 24*(2), 517–530.

Robins, J. M. and A. Rotnitzky (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pp. 297–331. Springer.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*(429), 122–129.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the american statistical association 90*(429), 106–121.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association 92*(439), 1049–1062.

Sastry, S. S. (2012). *Introductory methods of numerical analysis*. PHI Learning Pvt. Ltd.

Severini, T. A. and J. G. Staniswalis (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association 89*(426), 501–511.

Shao, Y. and L. Wang (2022). Generalized partial linear models with nonignorable dropouts. *Metrika 85*(2), 223–252.

Tsiatis, A. A., M. Davidian, and W. Cao (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics 67*(2), 536–545.

van der Laan, M. J. and S. Gruber (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The international journal of biostatistics 8*(1).

Wang, L., Z. Ouyang, and X. Lin (2024). Doubly robust estimation and semiparametric efficiency in generalized partially linear models with missing outcomes. *Stats 7*(3), 924.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika 90*(1), 43–52.

Wang, N., R. J. Carroll, and X. Lin (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association 100*(469), 147–157.

Wang, Q., O. Linton, and W. Härdle (2004). Semiparametric regression analysis with missing response at random. *Journal of the American Statistical Association 99*(466), 334–345.

Yates, A. and L. Kuhn (2022). Healthy dynamics of CD4 T cells may drive HIV resurgence in perinatally-infected infants on antiretroviral therapy. `https://doi.org/10.3886/E167981V1`. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Zhang, H. H., G. Cheng, and Y. Liu (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association 106*(495), 1099–1112.

Zhang, T. and Z. Zhu (2011). Empirical likelihood inference for longitudinal data with missing response variables and error-prone covariates. *Communications in Statistics-Theory and Methods 40*(18), 3230–3244.