Statistica Sinica Preprint No: SS-2022-0226	
Title	Modeling the Population Mean Outcome Trajectory in Observational Studies With Varying Time to Intervention
Manuscript ID	SS-2022-0226
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0226
Complete List of Authors	Hyunkeun Cho and Seonjin Kim
Corresponding Authors	Hyunkeun Cho
E-mails	hyunkeun-cho@uiowa.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Statistica Sinica

Modeling the population mean outcome trajectory in observational studies with varying time to intervention

HYUNKEUN RYAN CHO and SEONJIN KIM

University of Iowa and Miami University

Abstract: A repeatedly measured outcome in longitudinal studies allows researchers to monitor how the outcome changes over time. When an intervention affects the outcome and subjects initiate the intervention at different times during the course of a study, it is essential to account for the varying time to intervention (TTI) in models of such changes. In this study, we develop a piecewise polynomial regression model with TTI-varying coefficients that describes the population mean outcome over time. The TTI-varying coefficients in the model enable us to capture the population mean outcome trajectory, affected by both the intervention and the varying TTI. In observational studies, other covariates can confound these effects, leading to estimation bias if not properly accounted for. To mitigate this, we propose a double-weighted estimation procedure based on a kernel function and a generalized propensity score. The proposed estimation procedure effectively corrects the estimation bias of the TTI-varying coefficients and provides valid statistical inferences about the coefficients. We apply our approach to assess changes in the population mean of an inflammation biomarker for HIV-infected adults in Haiti who initiate antiretroviral therapy following the World Health Organization guideline.

Key words and phrases: Causal inference; Generalized propensity score; Kernel smoothing; Longitudinal data; Piecewise polynomial regression; Varying coefficients model.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

2. Introduction

Modeling changes in an outcome over time is essential for patient assessment in biomedical studies. An analysis of longitudinal data in which the outcome is measured repeatedly for a subject can successfully control extraneous, but unavoidable sources of variability among subjects. However, interventions that affect the changes in the outcome can occur at different times during the course of a longitudinal study. When the effect of the intervention depends on the time to intervention (TTI), it is crucial to adjust for the TTI when modeling the longitudinal outcome trajectory; see Wu and Tian (2008), Xing and Ying (2012), Liu et al. (2018), and Cho et al. (2020).

For example, an inflammation biomarker is one of the risk factors for adults infected with human immunodeficiency virus (HIV). Because inflammation is a risk factor for other disease progression, monitoring changes of the inflammation biomarker over time is essential. Antiretroviral therapy (ART) has proved effective in reducing inflammation, and is recommended for HIV-infected adults (Kanters et al. (2016)). However, owing to limited resources, HIV-infected adults in Haiti have initiated ART following the World Health Organization (WHO) guideline, leading to these adults initiating ART at different times. As a result, it is important to study how the effect of ART on changes in the inflammation biomarker is influenced by different ART initiation times.

It is straightforward to evaluate the TTI-varying effect of the intervention on the outcome by assessing the population mean outcome trajectory if the data are observed in either of the following circumstances: 1) the TTI is assigned randomly to subjects, or 2) all subjects undergo the intervention at the same time. In observational studies in which the intervention is initiated following a guideline, strategy, or other factors, neither of these circumstances are feasible. In particular, if

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

factors that affect the TTI exist, estimating the population mean outcome trajectory is challenging. For instance, suppose we wish to evaluate the intervention effect on the outcome when the TTI is a specific value s. One approach is to estimate the mean outcome trajectory using subjects who intervened at time s or nearby. However, it is likely that the subsample does not represent the study population well in the presence of potential confounders. As a result, an estimated mean outcome trajectory would be biased, unless the confounders are properly controlled.

In this study, we develop a piecewise polynomial regression model with TTI-varying coefficients that describes the marginal mean outcome over time. The proposed model smoothly connects the polynomial functions before and after the intervention. The TTI-varying coefficients allow us to explore the population mean outcome trajectory with respect to different times to intervention. Therefore, the proposed marginal mean model captures both dynamic longitudinal changes in the population mean outcome over time and the varying effect of the intervention along with the times to intervention. A hypothesis test is proposed to select the most parsimonious model that correctly specifies the population mean outcome pattern. If the intervention affects changes on the outcome over time, the pattern of the repeated outcome is altered after the TTI. Therefore, we develop another hypothesis test that investigates whether or not the intervention at a specific time is effective.

We propose a double-weighted estimation procedure to estimate the TTI-varying coefficients, while accounting for potential confounders that can cause selection bias under the weighted generalized estimating equations framework (Robins et al. (1994); Chen et al. (2010); Qu et al. (2011)). Because the proposed approach contains two weights, that is, a kernel function and a

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

generalized propensity score (GPS, Hirano and Imbens (2004)), we call it the double-weighted estimation method. The kernel function up-weights subjects who initiate the intervention at a specific time, or nearby. The GPS links the TTI and the potential confounders. We propose a simple and easy implementation to predict the GPS using the definition of the probability density function. Using the predicted GPS, the proposed procedure corrects the estimation bias effectively. Our simulation studies show that an estimation procedure that does not control for the confounders yields a biased estimator of the TTI-varying coefficients. In contrast, the double-weighted procedure successfully corrects the bias, and provides valid statistical inferences about the TTI-varying coefficients. We prove that the double-weighted estimator asymptotically follows a multivariate normal distribution with a mean vector of the true coefficients under regularity conditions on the GPS and the kernel function.

Repeated measures within each subject are likely to be correlated, and the degree of correlations can vary with the TTI. The proposed estimation approach accommodates the within-subject correlations, and improves the estimation efficiency of the TTI-varying coefficients. In addition, the approach accounts for heterogeneous correlations across TTIs, without estimating nuisance parameters associated with the working correlation structure that varies with the TTIs (Kim et al. (2019)). When the repeated outcome is not continuous, specifying its full likelihood under a marginal regression framework is challenging. The proposed estimation approach is readily applied to analyze repeated discrete outcome, because it requires only the first two moments.

The remainder of the paper proceeds as follows. In Section 2, we develop the piecewise polynomial regression model with the TTI-varying coefficients. In Section 3, we propose the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

double-weighted estimation procedure and present statistical inferences about the TTI-varying coefficients. In Section 4, we implement the proposed approach by selecting a parsimonious model and predicting the GPS. In Section 5, we apply the proposed approach to data from the aforementioned HIV study, and explore changes in the population mean inflammation biomarker at two different ART initiation times. Simulation studies and closing remarks are given in Sections 6 and 7, respectively.

3. Modeling the population mean outcome trajectory

For a typical framework of longitudinal studies with a varying TTI variable, we denote by T a real-valued variable of time, $\mathcal{T}$ a bounded subset of $(0, \infty)$ such that $T \in \mathcal{T}$, Y_T a real-valued response variable at time T, $\mathbf{Z}$ a vector of q real-valued covariates, and $S \in \mathcal{T}$ a real-valued TTI variable. Suppose that n subjects are drawn randomly from a population of interest and Y_T is measured repeatedly during the course of the study. The longitudinal random sample of $\{Y_T, T, S, \mathbf{Z}\}$ is denoted by $\{Y_{ij}, T_{ij}, S_i, \mathbf{Z}_i : \ i = 1, \ldots, n; \ j = 1, \ldots, n_i\}$, where the TTI S_i, covariate vector $\mathbf{Z}_i$, and n_i outcomes, $Y_{i1}, \ldots, Y_{in_i}$, at time points $T_{i1}, \ldots, T_{in_i}$, respectively, are measured for subject i. We call $\mathbf{Z}_i$ preintervention covariates, because it consists of covariates measured prior to an intervention.

When a longitudinal study is designed, the same visit or assessment schedule is normally planned for all individuals. In this regard, we assume that there exist no potential factors that confound the associations between the measurement time T and the outcome Y_T. In practice, the assessment times are likely not the same across all individuals for various reasons, including

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

missed visits or simply visit times falling outside the predefined windows. This results in different values of n_i and $T_{i1}, \ldots, T_{in_i}$ across individuals. In addition, the intervention can be initiated at any time during the follow-up period; that is, there exist individuals such that $T_{ij} \neq S_i$, for all j.

Under the stable unit treatment value assumption (SUTVA, Imbens and Rubin (2015)) that only one version of the intervention is used and no interference between subjects exists, we define a potential outcome measured at time T if an intervention is initiated at time S since the baseline, and denote it by $Y(T, S)$. This follows from the unconfoundedness assumption between the measurement and the outcome, and the consistency assumption that a potential outcome for subject i at time T_{ij} is observed as $Y_{ij} = Y(T_{ij}, S_i)$. We are interested in estimating the average outcome trajectory of individuals who intervened at $S \in \mathcal{T}$, denoted by $\mu(T, S) = E\{Y(T, S)\}$. Therefore, we develop a marginal mean regression model that assesses changes in the population mean outcome based on generalized linear models for longitudinal data (Liang and Zeger (1986)).

Suppose that $\mu(T, S)$ depends on T and S through a known link function of $\jmath(\cdot)$ (e.g., the logit link function for Bernoulli random response variables, or the log link function for count response variables). Assuming that the transformed mean response changes linearly over time T, but that the rate of the change is altered by the intervention at time S, we formulate the following marginal mean regression model with TTI-varying coefficients for the potential outcome $Y(T, S)$:

$$\jmath\{\mu(T, S)\} = \begin{cases} \beta_0(S) + \beta_1(S)T & T \leq S \\ \alpha_0(S) + \alpha_1(S)T & T > S, \end{cases} \tag{3.1}$$

where $\beta_0(S)$, $\beta_1(S)$, $\alpha_0(S)$, and $\alpha_1(S)$ are unknown smooth functions of the TTI S. Because

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

$\beta_1(S)$ and $\alpha_1(S)$ are the rates of the change before and after the intervention, respectively, the difference between $\beta_1(S)$ and $\alpha_1(S)$ is the expected rate of the change due to the intervention at time S. As a result, the TTI-varying effects of the intervention on the changes in $\jmath\{\mu(T,S)\}$ can be obtained by accessing $\beta_1(S)$ and $\alpha_1(S)$ with respect to the TTI S.

Under the continuity assumption of $\mu(T,S)$ in time T, we combine two segments in (3.1), and propose the following TTI-varying coefficient piecewise linear model:

$$\jmath\{\mu(T,S)\} = \beta_0(S) + \beta_1(S)T + \beta_2(S)(T-S)_+, \tag{3.2}$$

where $\beta_2(S) = \alpha_1(S) - \beta_1(S)$, and $(T-S)_+ = (T-S)I(T > S)$ is a truncated term with a fixed knot of S and an indicator function $I(T > S)$ being one if $T > S$ and zero otherwise. As a result, the change of the population mean outcome due to the intervention at time S is reflected in the last term in model (3.2). For example, $\beta_2(S) = 0$ indicates that the intervention at time S does not alter the rate of the change in the outcome, because the linear pattern of the time-varying outcome remains the same before and after the intervention. If $\beta_2(S)$ is a nonzero constant over $S \in \mathcal{T}$, the effect of the intervention remains the same, regardless of the TTIs.

The piecewise linearity assumption between the time and the transformed mean response can be relaxed by developing a piecewise polynomial regression model with TTI-varying coefficients, as follows:

$$\jmath\{\mu(T,S)\} = \beta_0(S) + \beta_1(S)T + \ldots + \beta_{p_s}(S)T^{p_s} + \beta_{p_s+1}(S)(T-S)_+^{p_s}, \tag{3.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

where $(T-S)_+^{p_s} = (T-S)^{p_s} I(T > S)$ is the p_s degree truncated polynomial term, and the degree of nonlinearity p_s can vary with S. Note that model (3.3) smoothly connects two different polynomial curves with a p_s degree of the polynomial, under the restriction that their first $p_s - 1$ derivatives are continuous in time T (Gallant and Fuller (1973)).

4. Statistical inference

In this section, we propose estimation procedures and discuss statistical inferences about the TTI-varying coefficients in model (3.3) in observational studies in which the preintervention covariates confound the associations between the TTI and the repeated outcomes.

4.1 Double-weighted estimation procedure

In order to control for covariates that could cause an estimation bias in $\mu(T, S)$, we propose a double-weighted estimation procedure based on the inverse probability weighting scheme (Horvitz and Thompson (1952)). Following Hirano and Imbens (2004) on propensity score analysis, we assume that the TTI is independent of the potential outcome, conditional on the covariates, denoted by $Y(T, s) \perp S|\mathbf{Z}$, for $s \in \mathcal{T}$. This assumption rules out any systematic selection into the TTI based on unobservable covariates, and is called the "no hidden bias assumption." The assumption is a natural extension of the unconfoundedness assumption commonly used for binary treatments (Rosenbaum and Rubin (1983); Heckman et al. (1998); Imbens (2000)).

Rosenbaum and Rubin (1983) show that adjusting for differences in the propensity score (i.e., the probability of receiving the treatment conditioning on the preintervention covariates) removes the selection bias between treated and untreated individuals under the unconfoundedness assump-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

tion. We define a GPS, denoted by $f(s|\mathbf{Z})$, that is the conditional density of the TTI variable given the covariates. We assume that every individual has a nonzero density of intervening at any time point in $\mathcal{T}$; that is, $f(s|\mathbf{Z}) > 0$, for $s \in \mathcal{T}$. Hirano and Imbens (2004) show that the GPS exhibits the following properties: 1) within strata with the same value of $f(s|\mathbf{Z})$, the occurrence of the event $S = s$ does not depend on the value of $\mathbf{Z}$, that is, $\mathbf{Z} \perp I\{S = s\}|f(s|\mathbf{Z})$; and 2) the TTI is unconfounded, given the GPS and the aforementioned unconfoundedness assumption. The second property enables us to remove selection bias by using the GPS on the estimation of $\mu(T, s)$. In particular, we can identify the causal parameter $E\{Y(T, s)\}$ from the observed data as $\mu(T, s) = E[E\{Y(T, s)|f(s|\mathbf{Z}) = c\}] = E[E\{Y_T|S = s, f(s|\mathbf{Z}) = c\}]$.

For the estimation of $\beta_0(S), \ldots, \beta_{p_s+1}(S)$ in model (3.3) at given values of TTI s and $f(s|\mathbf{Z}_i)$, we propose the following double-weighted generalized estimating equations:

$$\sum_{i=1}^{n} \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2} \mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1} \mathbf{A}_i^{-1/2} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(s)\} \frac{K_i(s)}{f(s|\mathbf{Z}_i)} = 0, \tag{4.4}$$

where $\dot{\boldsymbol{\mu}}_i = \partial \boldsymbol{\mu}_i(s)/\partial \boldsymbol{\beta}(s)$, $\boldsymbol{\mu}_i(s) = \{\mu(T_{i1}, s), \ldots, \mu(T_{in_i}, s)\}^\top$, $\mu(T_{ij}, s) = \mathrm{J}^{-1}\{\mathbf{X}_{ij}\boldsymbol{\beta}(s)\}$, $\mathbf{X}_{ij} = \left(1, T_{ij}, \ldots, T_{ij}^{p_s}, (T_{ij} - s)_+^{p_s}\right)^\top$, $\boldsymbol{\beta}(s) = \{\beta_0(s), \beta_1(s), \ldots, \beta_{p_s}(s), \beta_{p_s+1}(s)\}^\top$, $\mathbf{A}_i$ is a diagonal variance matrix of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^\top$, $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}$ is a working correlation structure of $\mathbf{Y}_i$ with a nuisance smoothing function vector of $\boldsymbol{\rho}(s)$, and $K_i(s) = K\{(S_i - s)/b(s)\}$ is a kernel function with a local bandwidth $b(s)$. The kernel function up-weights subjects whose TTI is closer to the given value of s for a consistent estimation of $\boldsymbol{\beta}(s)$. The GPS $f(s|\mathbf{Z}_i)$ is used as an inverse weight to eliminate the disparity between the study population and the sampling population (i.e., the group of subjects who intervened at s). Within-subject correlations are considered, while al-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

lowing the degree of the correlations $\boldsymbol{\rho}(s)$ to vary with the TTIs. When the working correlation structure is specified correctly, the efficient estimator can be obtained by solving (4.4), but this requires estimating the unknown nuisance parameter vector $\boldsymbol{\rho}(s)$.

An alternative is to approximate $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1}$ in (4.4) as $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1} = \sum_{d=1}^{D} \eta_d(s)\mathbf{B}_{id}$, where $\mathbf{B}_{i1}, \ldots, \mathbf{B}_{iD}$ are basis matrices and $\eta_1(s), \ldots, \eta_D(s)$ are unknown varying coefficients. The choice of a set of the basis matrices depends on the type of working correlation structure. For example, if the compound symmetry structured is assumed, then $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1} = \eta_1(s)\mathbf{B}_{i1} + \eta_2(s)\mathbf{B}_{i2}$, where $\mathbf{B}_{i1}$ is an identity matrix, and $\mathbf{B}_{i2}$ is a matrix with zero on the diagonal, and one elsewhere. If the first-order autoregressive, denoted by AR(1), structure is assumed, then $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1} \approx \eta_1(s)\mathbf{B}_{i1} + \eta_2(s)\mathbf{B}_{i2}$, where $\mathbf{B}_{i1}$ is an identity matrix, and $\mathbf{B}_{i2}$ is a matrix with one on the sub-diagonals, and zero elsewhere (Qu et al. (2000)). If $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}$ is unstructured, a set of basis matrices can be obtained using an eigenvector decomposition method; see Zhou and Qu (2012) and Cho and Qu (2015).

After extending (4.4) to a score vector of $\mathbf{g}_i\{\boldsymbol{\beta}(s)\} = \mathbf{h}_i\{\boldsymbol{\beta}(s)\}K_i(s)/f(s|\mathbf{Z}_i)$, with

$$\mathbf{h}_i\{\boldsymbol{\beta}(s)\} = \begin{pmatrix} \dot{\boldsymbol{\mu}}_i^{\top}\mathbf{A}_i^{-1/2}\mathbf{B}_{i1}\mathbf{A}_i^{-1/2}\{\mathbf{Y}_i - \boldsymbol{\mu}_i(s)\} \\ \vdots \\ \dot{\boldsymbol{\mu}}_i^{\top}\mathbf{A}_i^{-1/2}\mathbf{B}_{iD}\mathbf{A}_i^{-1/2}\{\mathbf{Y}_i - \boldsymbol{\mu}_i(s)\} \end{pmatrix}, \tag{4.5}$$

an estimator of $\boldsymbol{\beta}(s)$ is obtained by minimizing the quadratic inference function (QIF, Qu et al.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

(2000))

$$Q\{\boldsymbol{\beta}(s)\} = nb(s)\ \mathbf{G}\{\boldsymbol{\beta}(s)\}^{\top}\mathbf{V}\{\boldsymbol{\beta}(s)\}^{-1}\mathbf{G}\{\boldsymbol{\beta}(s)\}, \tag{4.6}$$

where $\mathbf{G}\{\boldsymbol{\beta}(s)\} = \sum_{i=1}^{n} \mathbf{g}_i\{\boldsymbol{\beta}(s)\}/nb(s)$ and $\mathbf{V}\{\boldsymbol{\beta}(s)\} = \sum_{i=1}^{n} \mathbf{g}_i\{\boldsymbol{\beta}(s)\}\mathbf{g}_i\{\boldsymbol{\beta}(s)\}^{\top}/nb(s)$. This accounts for within-subject correlations, without needing to estimate the varying nuisance parameter vector $\boldsymbol{\rho}(s)$ in $\mathbf{R}_i\{\boldsymbol{\rho}(s)\}$. As a result, the estimator is more efficient than the one obtained under the working independent correlation structure. In addition, it is the most efficient of the estimators obtained from the same set of estimating equations in (4.5), because $Q\{\boldsymbol{\beta}(s)\}$ optimally combines the extended scores by taking the inverse of their variability. With $\widehat{\boldsymbol{\beta}}(s) = \text{argmin}_{\boldsymbol{\beta}(s)}\, Q\{\boldsymbol{\beta}(s)\}$, the mean outcome trajectory for the study population intervened at time s is estimated as $\widehat{\mu}(T, s) = \jmath^{-1}\{\widehat{\beta}_0(s) + \widehat{\beta}_1(s)T + \ldots + \widehat{\beta}_{p_s}(s)T^{p_s} + \widehat{\beta}_{p_s+1}(s)(T-s)_+^{p_s}\}$.

Note that when extreme propensity scores are present, a stabilized weight $\ell(s)$ can be used as $\mathbf{g}_i\{\boldsymbol{\beta}(s)\} = \mathbf{h}_i\{\boldsymbol{\beta}(s)\}K_i(s)\ell(s)/f(s|\mathbf{Z}_i)$, where $\ell(s)$ is an arbitrary function of S evaluated at the TTI s, although a marginal density of the TTI is commonly used. The stabilized inverse probability avoids obtaining an estimator of $\mu(T, s)$ that is dominated by repeated outcomes of individuals with an extremely small value of $f(s|\mathbf{Z}_i)$.

For statistical inferences about $\widehat{\boldsymbol{\beta}}(s)$, we demonstrate asymptotic properties of $\widehat{\boldsymbol{\beta}}(s)$. Note that with undersmoothing, $nb(s)^5 \to 0$, Wilks' phenomenon holds for the QIF. Therefore, we use the QIF with undersmoothing to build a goodness-of-fit statistic to select the best degree of polynomial in model (3.3), and to build a hypothesis test statistic to check whether or not the intervention at time s is effective.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Theorem 1. *Let s be a fixed interior point in $\mathcal{T}$ and $\boldsymbol{\beta}_0(s)$ be a true parameter vector. Under the causal inference conditions discussed in Section 3.1, the regularity conditions in the Appendix, $nb(s) \to \infty$, and $nb(s)^5 \to 0$, we have:*

(i) $\sqrt{nb(s)}\left(\widehat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}_0(s)\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \varphi_K \int \frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})} d\mathbf{z} \{\boldsymbol{\Phi}(s)^\top \boldsymbol{\Sigma}(s) \boldsymbol{\Phi}(s)\}^{-1}\right)$, *where* $\xrightarrow{d}$ *denotes convergence in distribution,* $\varphi_K = \int K^2(u)du$, $f_{\mathbf{Z}}(\cdot)$ *is the density function of* $\mathbf{Z}$, $\boldsymbol{\Phi}(s) = E[\partial \mathbf{h}_i\{\boldsymbol{\beta}(S_i)\}/\partial \boldsymbol{\beta}(S_i) | S_i = s]$, *and* $\boldsymbol{\Sigma}(s) = E[\mathbf{h}_i\{\boldsymbol{\beta}(S_i)\}\mathbf{h}_i\{\boldsymbol{\beta}(S_i)\}^\top | S_i = s]$;

(ii) $Q\{\widehat{\boldsymbol{\beta}}(s)\} \xrightarrow{d} \chi^2_{(p_s+2)(D-1)}$ *if model (3.3) is specified correctly, where* $\chi^2_{(p_s+2)(D-1)}$ *is a chi-squared distribution with* $(p_s+2)(D-1)$ *degrees of freedom.*

It is well known that the kernel-based estimator is consistent, but biased (Li and Racine (2007)), and that the bias term is $O(b(s)^2)$; see the proof of Lemma 1 in the Appendix for details. Because the bias term contains first- and second-order derivatives, which are not easy to estimate in practice, it is common practice to either ignore it or to undersmooth it with a slightly smaller bandwidth than the optimal bandwidth satisfying $nb(s)^5 \to 0$, as shown in Theorem 1. Given the optimal local bandwidth $O(n^{-1/5})$, Theorem 1 (i) shows that the resultant estimator asymptotically follows a multivariate normal distribution with a mean vector of the true coefficients at the specific value of the TTI s.

4.2 Inference about the TTI-varying coefficients

It is of particular interest to perform a statistical inference about the last term $\beta_{p_s+1}(s)$ in model (3.3), because $\beta_{p_s+1}(s)$ quantifies the effect of the intervention at time s on the change of the mean outcome pattern. Given that $Q\{\boldsymbol{\beta}_0(s)\}$ is an analog to the negative twice loglikelihood, a hypothe-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

sis test for $H_0 : \beta_{p_s+1}(s) = 0$ against $H_a : \beta_{p_s+1}(s) \neq 0$ is conducted by comparing $Q\{\widetilde{\boldsymbol{\beta}}(s)\}$ with $Q\{\widehat{\boldsymbol{\beta}}(s)\}$, where $\widetilde{\boldsymbol{\beta}}(s)$ and $\widehat{\boldsymbol{\beta}}(s)$ are estimators obtained under H_0 and H_a, respectively.

Theorem 2. *Let s be a fixed interior point in $\mathcal{T}$. Under the causal inference conditions in Section 3.1, the regularity conditions in the Appendix, $nb(s) \to \infty$, and $nb(s)^5 \to 0$, if the null hypothesis is true, $Q\{\widetilde{\boldsymbol{\beta}}(s)\}$ is as small as $Q\{\widehat{\boldsymbol{\beta}}(s)\}$, and the test statistic $\mathcal{W}(s) = Q\{\widetilde{\boldsymbol{\beta}}(s)\} - Q\{\widehat{\boldsymbol{\beta}}(s)\}$ asymptotically follows a chi-squared distribution with one degree of freedom.*

Theorem 2 indicates that the intervention at time s is effective if the test statistic $\mathcal{W}(s)$ is larger than the $(1-\alpha)$th percentile of the chi-squared distribution with one degree of freedom at a significance level of α. The hypothesis test is an analog to the traditional likelihood ratio test that compares two nested models, because the null model is nested within the alternative model. The test is useful and easy to implement, because estimating the limiting variance of $\beta_{p_s+1}(s)$ in Theorem 1 (i) is difficult in practice, but not required in the proposed test.

4.3 Kernel-weighted estimation procedure

When no confounders exist (e.g., the TTI is randomized to subjects in a population), the following kernel-weighted GEE and QIF can be used to estimate $\boldsymbol{\beta}(s)$:

$$\sum_{i=1}^{n} \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2} \mathbf{R}_i\{\boldsymbol{\rho}(s)\}^{-1} \mathbf{A}_i^{-1/2} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(s)\} K_i(s) = 0 \tag{4.7}$$

and

$$Q_K\{\boldsymbol{\beta}(s)\} = nb(s)\, \mathbf{G}_K\{\boldsymbol{\beta}(s)\}^\top \mathbf{V}_K\{\boldsymbol{\beta}(s)\}^{-1} \mathbf{G}_K\{\boldsymbol{\beta}(s)\}, \tag{4.8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

where $\mathbf{G}_K\{\boldsymbol{\beta}(s)\} = \sum_{i=1}^{n} \mathbf{g}_{K,i}\{\boldsymbol{\beta}(s)\}/nb(s)$, $\mathbf{g}_{K,i}\{\boldsymbol{\beta}(s)\} = \mathbf{h}_i\{\boldsymbol{\beta}(s)\}K_i(s)$, $\mathbf{h}_i\{\boldsymbol{\beta}(s)\}$ is defined in (4.5), and $\mathbf{V}_K\{\boldsymbol{\beta}(s)\} = \sum_{i=1}^{n} \mathbf{g}_{K,i}\{\boldsymbol{\beta}(s)\}\mathbf{g}_{K,i}\{\boldsymbol{\beta}(s)\}^\top/nb(s)$. An estimator of $\boldsymbol{\beta}(s)$ can be obtained by minimizing $Q_K\{\boldsymbol{\beta}(s)\}$, denoted by $\widehat{\boldsymbol{\beta}}_K(s) = \text{argmin}_{\boldsymbol{\beta}(s)} Q_K\{\boldsymbol{\beta}(s)\}$. Note that (4.7) is a special case of (4.4) with a constant value of $f(s|\mathbf{Z}_i)$. Thus, none of the preintervention covariates are related to the link between the TTI variable and the potential outcome. As a result, the asymptotic properties in Theorems 1 and 2 can be used to perform a statistical inference about $\boldsymbol{\beta}(s)$ based on $Q_K\{\boldsymbol{\beta}(s)\}$. In Section 6, we show that the kernel-weighted estimation procedure leads to a valid statistical inference about $\boldsymbol{\beta}(s)$, when no confounders exist.

5. Implementation

5.1 Selection of a parsimonious model

Choosing the best degree of polynomial p_s in model (3.3) is essential to select a parsimonious model that specifies the time-varying population mean outcome correctly. At the given value of the TTI s, we provide an iterative two-step procedure that selects the local bandwidth $b(s)$ and polynomial degree p_s in model (3.3).

1. Given a predetermined value of p_s, we modify a leave-one-subject-out cross-validation method (Rice and Silverman (1991)) and select the local bandwidth $b'(s)$ by minimizing the kernel-weighted least squares

$$b'(s) = \underset{b(s)>0}{\text{argmin}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{Y_{ij} - \widehat{\mu}^{(-i)}(T_{ij}, S_i)\right\}^2 K\left\{\frac{(S_i-s)}{b(s)}\right\}}{\sum_{i=1}^{n}\sum_{j=1}^{n_i} K\left\{\frac{(S_i-s)}{b(s)}\right\}}, \tag{5.9}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

where $\widehat{\mu}^{(-i)}(T_{ij}, S_i)$ is an estimate of the population mean at time T_{ij} intervened at time S_i, with the bandwidth $b(s)$ obtained from all data except the ith subject. The cross-validation obtains a local bandwidth effectively by using the kernel-based weights accounting for the distance between the data S_i and the TTI of interest s. To hold the asymptotic properties in Theorem 1, we obtain the optimal bandwidth $b^*(s)$ by undersmoothing $b'(s)$ in (5.9) as $b^*(s) = b'(s)n^{-1/20}$. Because $b'(s)$ and $b^*(s)$ are $O(n^{-1/5})$ and $O(n^{-1/4})$, respectively, the condition for undersmoothing $nb^*(s)^5 \to 0$ in Theorem 1 is fulfilled.

2. Given the selected local bandwidth $b^*(s)$, $\widehat{\boldsymbol{\beta}}(s)$ is obtained by minimizing $Q\{\boldsymbol{\beta}(s)\}$. Following Theorem 1 (ii), we select p_s as the best polynomial degree if $Q\{\widehat{\boldsymbol{\beta}}(s)\}$ is no greater than the $(1-\alpha)$th percentile of the chi-squared distribution with $(p_s+2)(D-1)$ degrees of freedom at a significance level α.

In practice, we let the initial value of p_s be one, which is the piecewise linear model (3.2), and repeat Steps 1 and 2 by increasing p_s by one until the criterion in Step 2 is met. This iterative procedure enables us to choose the most parsimonious model, based on Theorem 1 (ii) that $Q\{\widehat{\boldsymbol{\beta}}(s)\}$ converges in distribution to $\chi^2_{(p_s+2)(D-1)}$ under the correctly specified model.

5.2 Prediction of the GPS

Modeling the GPS $f(s|\mathbf{Z}_i)$ plays an important role in providing an accurate estimator of $\boldsymbol{\beta}(s)$. We need to predict the GPS prior to minimizing $Q\{\boldsymbol{\beta}(s)\}$ when estimating $\boldsymbol{\beta}(s)$. We propose a simple and easy implementation for predicting $f(s|\mathbf{Z}_i)$ based on the definition of the probability density function that $f(s|\mathbf{Z}_i) = \lim_{d\to 0}\{F(s+d|\mathbf{Z}_i) - F(s-d|\mathbf{Z}_i)\}/2d = \lim_{d\to 0} E\{I(s-d < S \leq$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

$s+d)|\mathbf{Z}_i\}/2d$, where $F(s|\mathbf{Z}_i) = P(S \le s|\mathbf{Z}_i)$ is the conditional distribution function of s given $\mathbf{Z}_i$. Given a small positive value of d^*, the GPS for subject i can be approximated as

$$f(s|\mathbf{Z}_i) \approx E\{M_i(s)|\mathbf{Z}_i\}/2d^*, \tag{5.10}$$

where $M_i(s) = I(s - d^* < S_i \le s + d^*)$ is a Bernoulli random variable that indicates whether or not subject i initiates the intervention at time s, or nearby. To avoid the curse of dimensionality due to the multivariate covariate $\mathbf{Z}_i$, we model $E\{M_i(s)|\mathbf{Z}_i\}$ in (5.10) under the following generalized linear model with the logit link function:

$$E\{M_i(s)|\mathbf{Z}_i\} = P\{M_i(s) = 1|\mathbf{Z}_i\} = \frac{\exp\{\gamma_0(s) + \sum_{j=1}^{q} \gamma_j(s) Z_{ij}\}}{1 + \exp\{\gamma_0(s) + \sum_{j=1}^{q} \gamma_j(s) Z_{ij}\}}, \tag{5.11}$$

where $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{iq})^\top$ and $\gamma_0(s), \ldots, \gamma_q(s)$ are unknown TTI-varying coefficients. If the logistic regression model in (5.11) is specified correctly, Fan et al. (1996) show that as $d \to 0$, $E\{M_i(s)|\mathbf{Z}_i\}/2d \xrightarrow{p} f(s|\mathbf{Z}_i)$, under Condition (C4) in the Appendix.

We need to choose an optimal value of d^* to predict $f(s|\mathbf{Z}_i)$. We propose a cross-validation approach based on an integrated squared error (ISE, Fan and Yim (2004))

$$\begin{aligned}
\text{ISE} &= \int \{\widehat{f}_d(s|\mathbf{z}) - f(s|\mathbf{z})\}^2 f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} ds \\
&= \int \widehat{f}_d^2(s|\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} ds - 2\int \widehat{f}_d(s|\mathbf{z}) f(s|\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} ds + \int f^2(s|\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} ds \\
&= \int f_{\mathbf{Z}}(\mathbf{z}) \left\{ \int \widehat{f}_d^2(s|\mathbf{z}) ds \right\} d\mathbf{z} - 2\int \widehat{f}_d(s|\mathbf{z}) f(s, \mathbf{z}) d\mathbf{z} ds + \int f^2(s|\mathbf{z}) f_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z} ds,
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

where $\widehat{f}_d(s|\mathbf{z})$ estimates $f(s|\mathbf{z})$ by conducting a logistic regression in (5.11) with a bandwidth d. Because the last term of the quadratic expansion does not depend on d, a cross-validation is driven by the first two terms in ISE as

$$CV(d) = \frac{1}{n}\sum_{i=1}^{n}\int \widehat{f}^2_{d,-i}(s|\mathbf{Z}_i)ds - \frac{2}{n}\sum_{i=1}^{n}\widehat{f}_{d,-i}(S_i|\mathbf{Z}_i), \tag{5.12}$$

where $\widehat{f}^2_{d,-i}(s|\mathbf{Z}_i)$ is obtained from all data, except subject i, using the approximation in (5.10). The optimal bandwidth is selected by minimizing $CV(d)$ as $d^* = \text{argmin}_{d>0}\, CV(d)$, and is $O(n^{-1/5})$, like the bandwidth $b'(s_0)$ in (5.9) (Fan et al. (1996)). However, undersmoothing is not required for this bandwidth, because undersmoothing uses a smaller bandwidth in order to eliminate the bias asymptotically more quickly, while losing some efficiency. Because this results in a slower convergence rate, the bandwidth d^* with the order $O(n^{-1/5})$ is fast enough to achieve the consistency of the estimator of $f(s|\mathbf{Z}_i)$ and the asymptotic results in Theorems 1 and 2.

6. Analysis of a guideline-based intervention study

In HIV-infected subjects, inflammations often result in other disease progression, such as cardiovascular disease and chronic anemia. Because ART is effective for reducing inflammations (Kanters et al. (2016)), it is often recommended for HIV-infected adults. However, owing to limited resources, such adults in Haiti have initiated ART following the WHO guideline of the early 2010s, that is, ART is initiated when the CD4 cell count is below 200 or AIDS has developed. This leads to different times of ART initiation, without knowing how these differing times would affect their level of inflammation over time.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

As part of a clinical trial of an HIV study conducted in 2010 (Severe et al. (2010)), 816 HIV-infected adults in Haiti who meet the following baseline criteria are enrolled: older than 18; CD4 counts between 200 and 350; AIDS-free; and no prior ART. A clinician meets all participants monthly and starts the intervention when the WHO guideline is met. Interleukin 6 (IL-6) is an inflammation biomarker of interest, and is collected approximately every six months over a three-year period. Note that the value of IL-6 correlates positively with clinical severity in HIV-infected adults. We apply the double-weighted estimating procedure and assess the TTI-varying coefficients in model (3.3) with the identity link function at a specific value of the TTI $s = 0.75$ (i.e., ART is initiated 9 months since the baseline) and 1.5 (i.e., ART is initiated 18 months since the baseline). Note that five covariates at the baseline are observed prior to ART initiation. In Table 1, we provide the kernel-weighted average of the observed baseline covariates, including gender, CD4 cell counts, hemoglobin, BMI, and age at $s = 0.75$ and 1.5. The results show that some covariates, particularly on gender and CD4 cell counts, are not balanced between the two TTIs.

In order to adjust for differences in the covariates, we predict the GPS using the logistic regression analysis at each value of s, conditioning on the preintervention covariates as $\log\left[P\{M_i(s) = 1|\mathbf{Z}_i\}/P\{M_i(s) = 0|\mathbf{Z}_i\}\right] = \gamma_0(s)+\gamma_1(s)\text{gender}_i+\gamma_2(s)\text{CD4}_i+\gamma_3(s)\text{hemoglobin}_i+\gamma_4(s)\text{BMI}_i+\gamma_5(s)\text{age}_i$, where $M_i(s)$ is one if $|S_i - s| < d^*$, and is zero otherwise. The optimal value of $d^* = 0.30$ is selected using the cross-validation in (5.12). Based on the predicted GPS, the double-weighted average of the baseline covariates is reported in Table 1. We confirm that the differences in the covariates of gender and CD4 cell counts are reduced substantially using the GPS analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Given the predicted GPS, we apply the iterative two-step procedure in Section 4.1 using the Epanechnikov kernel function, and select $b^*(s) = 0.46$ and $p_s = 2$ at $s = 0.75$ (i.e., $Q\{\widehat{\boldsymbol{\beta}}(0.75)\} = 2.248 < \chi^2_{0.95,4} = 9.49$), and $b^*(s) = 0.31$ and $p_s = 2$ at $s = 1.5$ (i.e., $Q\{\widehat{\boldsymbol{\beta}}(1.5)\} = 3.09 < \chi^2_{0.95,4}$). For $p_s = 1$, $Q\{\widehat{\boldsymbol{\beta}}(0.75)\} = 12.08$ and $Q\{\widehat{\boldsymbol{\beta}}(1.5)\} = 10.42$ are both greater than $\chi^2_{0.95,3} = 7.81$. We estimate $\boldsymbol{\beta}(s) = (\beta_0(s), \beta_1(s), \beta_2(s), \beta_3(s))^\top$ in

$$\mu(T, s) = \beta_0(s) + \beta_1(s)T + \beta_2(s)T^2 + \beta_3(s)(T - s)^2_+, \tag{6.13}$$

and provide the fitted mean IL-6 trajectories, $\widehat{\mu}(T, s)$ at $s = 0.75$ and 1.5 in Figure 1.

Figure 1 shows that $\widehat{\mu}(T, 0.75)$ and $\widehat{\mu}(T, 1.5)$ remain the same at the baseline, and increase with a very similar rate of the change before ART initiation. These results indicate that the sample selection biases in the two subsamples at $s = 0.75$ and 1.5 are corrected properly and thus, the two fitted mean trajectories are very comparable before ART initiation. After ART is initiated, $\widehat{\mu}(T, 0.75)$ and $\widehat{\mu}(T, 1.5)$ decrease, but the rate of the decrease lessens over time during the post-treatment period. The fitted population mean no longer decreases at the end of the follow-up. As a result, the greater the delay before initiating ART, the higher is the mean of IL-6. Of note is that the population mean pattern appears to be comparable. Nevertheless, faster initiation of ART in the study population had a more positive effect on reducing IL-6. We also check whether ART is effective at $s = 0.75$ and 1.5 by testing $H_0 : \beta_3(s) = 0$ against $H_a : \beta_3(s) \neq 0$ in model (6.13). The test statistic $\mathcal{W}(s)$ in Theorem 2 is 9.49, with a p-value of 0.002, at $s = 0.75$, and 7.21, with a p-value of 0.007, at $s = 1.5$. These results confirm that ART is significantly effective in reducing the population mean of IL-6 in both cases at a significance level of 0.05.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

To demonstrate the importance of adjusting for confounders in the analysis of the guideline-based intervention study, we fit model (6.13) at $s = 0.75$ and 1.5 using the kernel-weighted estimation procedure in Section 3.3, and denote the fitted mean outcome trajectory by $\widehat{\mu}_K(T, s)$. The right panel of Figure 1 shows that during the pre-treatment period, $\widehat{\mu}_K(T, 0.75)$ increases more rapidly than $\widehat{\mu}_K(T, 1.5)$. Although the initiation of ART is delayed in HIV-infected adults, the fitted mean IL-6 of the population is lower during most of the follow-up period. Comparing the two estimation procedures, $\widehat{\mu}_K(T, 0.75)$ obtains using the kernel-weighted approach is larger than that of the double-weighted approach in the smaller delay population over the follow-up period. This phenomenon reverses in the greater delay population. The different fitted mean outcome trajectories can be explained by the fact that the distribution of the baseline covariates is not balanced across times to ART initiation, as shown in Table 1. Therefore, the distribution of participants initiated at time s deviates from that of all participants drawn randomly from the study population. As a result, it is likely that the kernel-weighted estimation procedure yields a biased estimate of $\mu(T, s)$, whereas the double-weighted estimation procedure corrects the bias successfully.

7. Simulation studies

In this section, we use simulation studies to determine whether the double-weighted estimation procedure effectively removes estimation bias problems when confounding factors exist. We also show that the proposed method performs valid statistical inferences, including the parsimonious model selection and the hypothesis test for the last term $\beta_2(S)$ in the piecewise linear regression model, $\mu(T, S) = \beta_0(S) + \beta_1(S)T + \beta_2(S)(T - S)_+$, where $\beta_0(S) = \exp(0.1S - 0.15)$, $\beta_1(S) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

$\cos(\pi S)/25 - 1$, and $\beta_2(S) = 0.03 - 0.02S$.

To simulate a longitudinal sample, we first generate the ith subject's TTI S_i, for $i = 1, \ldots, 1000$, independently from a uniform distribution $\mathcal{U}(0.5, 2.5)$. Given the value of S_i, we generate five repeated outcomes for the subject i as

$$Y_{ij} = \mu(T_{ij}, S_i) + \epsilon_{ij} = \beta_0(S_i) + \beta_1(S_i)T_{ij} + \beta_2(S_i)(T_{ij} - S_i)_+ + \epsilon_{ij}, \tag{7.14}$$

where $T_{i1} = 0$ indicates the time at the baseline, $T_{ij} = j + \mathcal{U}(-2, -1)$, for $j = 2, \ldots, 5$, resulting in unequally spaced measurement time points between zero and four across subjects. The random error is modeled with three confounders, denoted by Z_{i1}, Z_{i2}, and Z_{i3}, as

$$\epsilon_{ij} = Z_{i1} + Z_{i2}T_{ij} + Z_{i3}(T_{ij} - S_i)_+ + e_{ij}, \tag{7.15}$$

where Z_{i1}, Z_{i2}, Z_{i3}, and e_{ij} are generated independently from a normal distribution with mean zero and standard deviation 0.5, where $E(\epsilon_{ij}) = 0$ still holds. We then drop Y_{ij} if an indicator is zero. This indicator is generated independently at visits $j = 2, \ldots, 5$ for subject i from a Bernoulli distribution with the success probability of 0.8. This results in a different number of repeated outcomes across subjects.

To examine the performance of the double-weighted estimation approach, an indicator variable of $M_i(S_i)$ is generated independently from a Bernoulli distribution with success probability $P\{M_i(S_i) = 1 | Z_{i1}, Z_{i2}, Z_{i3}\} = \exp(0.2Z_{i1} + 0.3Z_{i2} - 0.5Z_{i3}) / \{1 + \exp(0.2Z_{i1} + 0.3Z_{i2} - 0.5Z_{i3})\}$. From the 1000 subjects in the simulated sample, we drop subjects with $M_i(S_i) = 0$ from the sam-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

ple, and consider the subsample in our analysis. Subjects in the subsample are likely to have a higher value of Z_{i1} or Z_{i2} or a smaller value of Z_{i3}, resulting in the sample selection problems. As a result, Z_{i1}, Z_{i2}, and Z_{i3} are confounders that affect the TTI and the population mean outcome.

At a given TTI value of $s = 1.5$, we assess $\boldsymbol{\beta}(1.5) = (\beta_0(1.5), \beta_1(1.5), \beta_2(1.5))$ in (7.14) from 1000 simulated subsamples using $Q\{\boldsymbol{\beta}(s)\}$ and $Q_K\{\boldsymbol{\beta}(s)\}$ under the independence, AR(1), and compound symmetry working correlation structures. Table 2 shows the average bias and mean squared error (MSE) of the estimates. The results show that the kernel-weighted approach using $Q_K\{\boldsymbol{\beta}(s)\}$ yields biased estimates, and that the direction and amount of the bias depends on the level of confounder. Specifically, the average bias of $\widehat{\beta}_2(1.5)$ is negative and largest, because the random error depends on Z_{i3}, and Z_{i3} has the largest effect on the indicator $M_i(S_i)$ and the formation of the subset. In contrast, the double-weighted approach using $Q\{\boldsymbol{\beta}(s)\}$ decreases the average bias substantially.

We check the performance of the parsimonious model selection, and confirm that at a significance level of 0.05, the proportion of rejecting the piecewise linear model is 0.057 and 0.056 under the AR(1) and compound symmetry structures, respectively. We also conduct a statistical inference about $\beta_2(1.5)$ in (7.14). Because the true value of $\beta_2(s)$ is zero at $s = 1.5$, we test $H_0 : \beta_2(1.5) = 0$ against $H_a : \beta_2(1.5) \neq 0$ using the test statistic $\mathcal{W}(s)$ in Theorem 2 when the null hypothesis is true. At a significance level of 0.05, the proportion of rejection is 0.047 and 0.054 under the AR(1) and compound symmetry structures, respectively. According to Q-Q plots for the chi-squared distribution with one degree of freedom versus $\mathcal{W}(1.5)$ in Figure 2, the Q-Q plots follow the identity line sufficiently well, because the null hypothesis is true. We conduct a similar test based upon

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

$Q_K\{\boldsymbol{\beta}(s)\}$, but the rejection rate increases to 0.119 and 0.124 under the AR(1) and compound symmetry structures, respectively. Moreover, the test statistic no longer follows the chi-squared distribution.

The repeated outcomes, $Y_{i1}, \ldots, Y_{i5}$, are correlated, and Table 2 shows that the estimation efficiency of $\boldsymbol{\beta}(s)$ improves by accommodating the within-subject correlations (i.e., the AR(1) and compound symmetry) as compared with ignoring the correlations (i.e., the independence structure). In particular, the true covariance matrix of $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i5})^\top$ is $0.25(\mathbf{X}_i\mathbf{X}_i^\top + \mathbf{I}_5)$, where $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{i5})^\top$ with $\mathbf{X}_{ij} = \big(1, T_{ij}, (T_{ij} - 2)_+\big)^\top$, and $\mathbf{I}_5$ is the 5×5 identity matrix. The AR(1) approximates the true correlation structure better than the compound symmetry does. This is aligned with the smallest MSE under the AR(1) structure in all cases under consideration.

We also compare the performance of the two estimation approaches in cases where Z_{i1}, Z_{i2}, and Z_{i3} are no longer confounding factors. We generate $M_i(S_i)$ independently from a Bernoulli distribution with $P\{M_i(S_i) = 1 | Z_{i1}, Z_{i2}, Z_{i3}\} = 0.5$, and include subjects with $M_i(S_i) = 1$ in a subsample. From the 1000 subsamples, we assess $\boldsymbol{\beta}(1.5)$ in (7.14) using $Q_K\{\boldsymbol{\beta}(s)\}$ and $Q\{\boldsymbol{\beta}(s)\}$ under the three working correlation structures. Table 2 confirms that both the kernel-weighted and double-weighted estimation procedures perform well in terms of both small biases and MSEs. In sum, our simulation studies confirm that the proposed procedure performs well, regardless of the presence of confounding factors.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

8. Conclusion

An intervention can occur at different times across individuals in studies where the effect of the TTI on the repeated outcome remains uncertain. Wu and Tian (2008), Xing and Ying (2012), Liu et al. (2018), and Cho et al. (2020) have proposed longitudinal models that account for the varying TTI effect on the repeated outcome when no confounders exist. In observational studies, the intervention is rather initiated based on other factors that confound the TTI effect. Controlling for plausible confounders is a crucial, but challenging part of observational data analysis. This becomes more critical and inevitable in longitudinal observational studies when the entire set of repeatedly measured outcomes and the TTI are confounded as a whole.

As an example, the two fitted time-varying mean outcomes show discernibly different patterns in the analysis of repeated IL-6 outcomes in Section 5. This shows that confounding factors can occur in the WHO guideline-based intervention study, and the fitted mean without controlling for the confounding factors deviates substantially from the population mean over time. In contrast, the proposed double-weighted estimation procedure reduces the risk of estimation bias and achieves a consistent estimator of the population mean outcome trajectory from samples in observational studies.

The piecewise polynomial regression model with the TTI-varying coefficients and double-weighted estimation procedure can also be applied to other types of longitudinal observational studies with an event for which the timing is not controllable, such as physiological phenomena or natural phenomena. The event would occur to subjects at different times during the follow-up period, and could have a significant and different effect on the outcome. The proposed methodology

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

provides a way to assess the population mean outcome trajectory accurately and to examine the effect of times to event on the outcome time-varying population mean outcome effectively.

Supplementary Material

Supplementary materials available online include an implementation algorithm and R codes of simulation studies.

References

Chen, B., Yi, G. Y. and Cook, R. J. (2010). Weighted generalized estimating functions for longitudinal response and covariate data which are missing at random. *Journal of the American Statistical Association* **105**, 336–353.

Cho, H., Kim, S. and Lee, M. (2020). Adjusting a subject-specific time of event in longitudinal studies. *Statistical Methods in Medical Research* **29**, 1787–1798.

Cho, H. and Qu, A. (2015). Efficient estimation for longitudinal data by combining large-dimensional moment conditions. *Electronic Journal of Statistics* **9**, 1315–1334.

Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83**, 189–206.

Fan, J. and Yim, Y. H. (2004). A crossvalidation method for estimating conditional densities. *Biometrika* **91**, 819–834.

Gallant, A. R. and Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association* **68**, 144–147.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica* **66**, 1017–1098.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Wiley & Sons.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.

Imbens, G. W. (2000). The role of the propensity score in estimating dose response functions. *Biometrika* **87**, 706–710.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: an Introduction*. Cambridge University Press.

Kanters, S., Vitoria, M., Doherty, M., Socias, M. E., Ford, N., Forrest, J. I., Popoff, E., Bansback, N., Nsanzimana, S., Thorlund, K. and Mills, E. J. (2016). Comparative efficacy and safety of first-line antiretroviral therapy for the treatment of HIV infection: a systematic review and network meta-analysis. *The Lancet HIV* **3**, e510–e520.

Kim, S., Cho, H. and Zhang, X. (2019). Initial severity-dependent longitudinal model with application to a randomized controlled trial of women with depression. *Statistics in Medicine* **38**, 1687–1689.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 12–22.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Liu, T., Wu, C. O., Li, Z. and Li, Y. (2018). Semiparametric random-effects conditional density models for longitudinal analysis with concomitant intervention. *Statistica Sinica* **28**, 1333–1349.

Qu, A., Lindsay, B. G. and Li, B. (2000). Improving generalized estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Qu, A., Yi, G. Y., Song, P. S. and Wang, P. (2011). Assessing the validity of weighted generalized estimating equations. *Biometrika* **98**, 215–224.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B* **53**, 233–243.

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

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

Severe, P., Juste, M. A. J., Ambroise, A., Eliacin, L., Marchand, C., Apollon, S., Edwards, A., Bang, H., Nicotera, J., Godfrey, C., Gulick, R. M., Johnson, W. D., Pape, J. W. and Fizgerald, D. W. (2010). Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *The New England Journal of Medicine* **363**, 257–265.

Wu, C. O. and Tian, X. (2008). A varying-coefficient model for the evaluation of time-varying concomitant intervention effects in longitudinal studies. *Statistics in Medicine* **27**, 3042–3056.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Xing, H. and Ying Z. (2012). A semiparametric change-point regression model for longitudinal observations. *Journal of the American Statistical Association* **107**, 1625–1637.

Zhou, J. and Qu, A. (2012). Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistics Association* **107**, 701–710.

Appendix

The following conditions are imposed to study the asymptotic properties of the estimator $\widehat{\boldsymbol{\beta}}(s) = \text{argmin}_{\boldsymbol{\beta}(s)} Q\{\boldsymbol{\beta}(s)\}$.

(C1) There exists a $\boldsymbol{\beta}_0(s)$ such that $E[\mathbf{h}_i\{\boldsymbol{\beta}(s)\}K_i(s)/f(s|\mathbf{Z}_i)] \to 0$ for all i if and only if $\boldsymbol{\beta}(s) = \boldsymbol{\beta}_0(s)$.

(C2) The matrix $\boldsymbol{\Sigma}(s)$ is positive definite, and $\boldsymbol{\Phi}(s)$ is of full rank. $\boldsymbol{\Sigma}(\cdot)$ and $\boldsymbol{\Phi}(\cdot)$ are twice continuously differentiable in a neighborhood of s.

(C3) $\boldsymbol{\beta}(\cdot)$ is third continuous differentiable in a neighborhood of s. The inverse link function $\jmath^{-1}(\cdot)$ is strictly monotone and has a continuous third derivative. Thus, $\mu(T, \cdot)$ is third continuous differentiable in a neighborhood of s,

(C4) The density functions $f_{S,\mathbf{Z}}(\cdot, \cdot)$ and $f_{\mathbf{Z}}(\cdot)$ are bounded, positive and third continuous differentiable.

Lemma 1. *Under the regularity conditions (C1)-(C4), the causal inference conditions in Section*

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

3.1, $nb(s)^5 \to 0$*, and* $n \to \infty$*, we have*

$$\frac{1}{\sqrt{nb(s)}} \sum_{i=1}^{n} \mathbf{g}_i\{\boldsymbol{\beta}_0(s)\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(s)\varphi_K \int \frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})} d\mathbf{z}\right).$$

Proof of Lemma 1. Recall $\mathbf{g}_i\{\boldsymbol{\beta}(s)\} = \mathbf{h}_i\{\boldsymbol{\beta}(s)\}K_i(s)/f(s|\mathbf{Z}_i)$, $K_i(s) = K\{(S_i - s)/b(s)\}$, $\mu(T,S) = E\{Y(T,S)\}$ and $\boldsymbol{\mu}_i(S_i) = (\mu(T_{i1}, S_i), \ldots, \mu(T_{in_i}, S_i))^\top$. We define $\boldsymbol{\epsilon}_i(S_i) = (\epsilon(T_{i1}, S_i), \ldots, \epsilon(T_{in_i}, S_i))^\top$ where $\epsilon(T,S) = Y(T,S) - E\{Y(T,S)\}$. We can decompose $\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\} = \mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}K_i(s)/f(s|\mathbf{Z}_i) + \boldsymbol{\ell}_i$, where

$$\mathbf{h}_i\{\boldsymbol{\beta}(s)\} = \begin{pmatrix} \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{i1}\mathbf{A}_i^{-1/2}\{\mathbf{Y}_i - \boldsymbol{\mu}_i(S_i)\} \\ \vdots \\ \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{iD}\mathbf{A}_i^{-1/2}\{\mathbf{Y}_i - \boldsymbol{\mu}_i(S_i)\} \end{pmatrix} = \begin{pmatrix} \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{i1}\mathbf{A}_i^{-1/2}\boldsymbol{\epsilon}_i(S_i) \\ \vdots \\ \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{iD}\mathbf{A}_i^{-1/2}\boldsymbol{\epsilon}_i(S_i) \end{pmatrix}$$

and

$$\boldsymbol{\ell}_i = \begin{pmatrix} \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{i1}\mathbf{A}_i^{-1/2}\{\boldsymbol{\mu}_i(S_i) - \boldsymbol{\mu}_i(s)\}\frac{K_i(s)}{f(s|\mathbf{Z}_i)} \\ \vdots \\ \dot{\boldsymbol{\mu}}_i^\top \mathbf{A}_i^{-1/2}\mathbf{B}_{iD}\mathbf{A}_i^{-1/2}\{\boldsymbol{\mu}_i(S_i) - \boldsymbol{\mu}_i(s)\}\frac{K_i(s)}{f(s|\mathbf{Z}_i)} \end{pmatrix}.$$

Consider the second term $\boldsymbol{\ell}_i$ above first. Since $K(\cdot)$ has bound support, it is sufficient to consider v such that $|v - s| = O\{b(s)\}$. Define $\dot{\mu}(s) = \partial\mu(T,s)/\partial s$, $\ddot{\mu}(s) = \partial^2\mu(T,s)/\partial s^2$ and $\dot{f}_S(s,\mathbf{z}) = \partial f_{S,\mathbf{Z}}(s,\mathbf{z})/\partial s$. By Taylor's expansion and the symmetry of kernel $\int u^\gamma K(u)du = 0$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

for $\gamma = 1, 3, 5, \ldots$, we have, for any $t \in \mathcal{T}$,

$$
\begin{aligned}
& E\left[\frac{\{\mu(t, S_i) - \mu(t, s)\}K_i(s)}{f(s|\mathbf{Z}_i)}\right] \\
= & \int\int \frac{[(v-s)\dot{\mu}(s) + (v-s)^2\ddot{\mu}(s)/2 + O\{(v-s)^3\}]K\{(v-s)/b(s)\}}{f(s|\mathbf{z})} f_{S,\mathbf{Z}}(v, \mathbf{z})dvd\mathbf{z} \\
& \text{by } \; u = \frac{(v-s)}{b(s)} \\
= & \int\int \frac{[ub(s)\dot{\mu}(s) + \{ub(s)\}^2\ddot{\mu}(s)/2 + u^3 O\{b(s)^3\}]K(u)}{f(s|\mathbf{z})} f_{S,\mathbf{Z}}\{s + ub(s), \mathbf{z}\}b(s)dud\mathbf{z} \\
= & \int\int \frac{[ub(s)\dot{\mu}(s) + \{ub(s)\}^2\ddot{\mu}(s)/2]K(u)}{f(s|\mathbf{z})} \{f_{S,\mathbf{Z}}(s, \mathbf{z}) + ub(s)\dot{f}_S(s, \mathbf{z})\}b(s)dud\mathbf{z} + O\{b(s)^5\} \\
= & \int\int \frac{u^2 b(s)^3 \dot{\mu}(s)\dot{f}_S(s, \mathbf{z})K(u)}{f(s|\mathbf{z})} dud\mathbf{z} + \int\int \frac{u^2 b(s)^3 \ddot{\mu}(s) f_{S,\mathbf{Z}}(s, \mathbf{z})K(u)}{2f(s|\mathbf{z})} dud\mathbf{z} + O\{b(s)^5\} \\
= & \; b(s)^3 \mu_K \left\{ \dot{\mu}(s) \int \frac{\dot{f}_S(s, \mathbf{z})}{f(s|\mathbf{z})} d\mathbf{z} + \frac{\ddot{\mu}(s)}{2} \int \frac{f_{S,\mathbf{Z}}(s, \mathbf{z})}{f(s|\mathbf{z})} d\mathbf{z} \right\} + O\{b(s)^5\} \\
= & \; b(s)^3 \mu_K \left\{ \dot{\mu}(s) \int \frac{\dot{f}_S(s, \mathbf{z})}{f(s|\mathbf{z})} d\mathbf{z} + \frac{\ddot{\mu}(s)}{2} \right\} + O\{b(s)^5\}. \qquad \text{(A.1)}
\end{aligned}
$$

and thus $E[\{nb(s)\}^{-1/2} \sum_{i=1}^n \boldsymbol{\ell}_i] = O\{\sqrt{nb(s)^5}\}$. Note that $\int f_{S,\mathbf{Z}}(s, \mathbf{z})/f(s|\mathbf{z})d\mathbf{z} = \int\{f(s|\mathbf{z})f_{\mathbf{Z}}(\mathbf{z})\}/f(s|\mathbf{z})d\mathbf{z} = \int f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z} = 1$. Since $\boldsymbol{\ell}_i$, $i = 1, \ldots, n$, are independent, it is easy to show that $Var[\{nb(s)\}^{-1/2} \sum_{i=1}^n \boldsymbol{\ell}_i] \to 0$ as $nb(s) \to \infty$. Therefore, $\{nb(s)\}^{-1/2} \sum_{i=1}^n \boldsymbol{\ell}_i = o_p(1)$ can be shown under the assumption of undersmoothing (i.e., $nb(s)^5 \to 0$).

Now consider the first term $\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}K_i(s)/f(s|\mathbf{Z}_i)$. By the fact that $E\{K(s)|\mathbf{Z} = \mathbf{z}\} = f(s|\mathbf{z}) + o\{b(s)\}$ (Fan et al. (1996)) and $\boldsymbol{\epsilon}(S_i)$ and $(\mathbf{Z}_i, K_i(s))$ are independent, we have

$$
\begin{aligned}
E\left\{\frac{\boldsymbol{\epsilon}_i(S_i)K_i(s)}{f(s|\mathbf{Z}_i)}\right\} &= E\left[E\left\{\frac{\boldsymbol{\epsilon}_i(S_i)K_i(s)}{f(s|\mathbf{Z}_i)}\middle|\boldsymbol{\epsilon}_i(S_i), \mathbf{Z}_i\right\}\right] \\
&= E\left[\frac{\boldsymbol{\epsilon}_i(S_i)}{f(s|\mathbf{Z}_i)} E\{K_i(s)|\boldsymbol{\epsilon}_i(S_i), \mathbf{Z}_i\}\right] \\
&= E\left[\frac{\boldsymbol{\epsilon}_i(S_i)}{f(s|\mathbf{Z}_i)} E\{K_i(s)|\mathbf{Z}_i\}\right]
\end{aligned}
$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

$$= \int\int \frac{\mathbf{u}}{f(s|\mathbf{z})}[f(s|\mathbf{z}) + o\{b(s)\}]f_{\boldsymbol{\epsilon},\mathbf{Z}}(\mathbf{u},\mathbf{z})d\mathbf{u}d\mathbf{z} = 0,$$

and thus $E[\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}K_i(s)/f(s|\mathbf{Z}_i)] = 0$. In addition, by a similar manner in (A.1) and the independence of $\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}$, $i = 1, \ldots, n$, we can show

$$\begin{aligned} Var\left[\frac{1}{\sqrt{nb(s)}}\sum_{i=1}^{n}\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}\frac{K_i(s)}{f(s|\mathbf{Z}_i)}\right] &= E\left[\frac{1}{nb(s)}\sum_{i=1}^{n}\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}\mathbf{h}_i\{\boldsymbol{\beta}_0(S_i)\}^\top\frac{K_i^2(s)}{f^2(s|\mathbf{Z}_i)}\right] \\ &= \boldsymbol{\Sigma}(s)\varphi_K\int\frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})}d\mathbf{z} + O\{b(s)\}. \end{aligned} \tag{A.2}$$

Then, the following result is obtained by the central limit theorem and Slutsky's theorem, as $nb(s)^5 \to 0$ and $n \to \infty$, we have

$$\frac{1}{\sqrt{nb(s)}}\sum_{i=1}^{n}\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}(s)\varphi_K\int\frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})}d\mathbf{z}\right).$$

♢

Lemma 2. *Under the regularity conditions (C1)-(C4), the causal inference conditions in Section 3.1, and* $1/nb(s) + b(s) \to 0$*, we have:*

$$\begin{aligned} \frac{1}{nb(s)}\sum_{i=1}^{n}\frac{\partial\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}}{\partial\boldsymbol{\beta}(s)} &\xrightarrow{p} \boldsymbol{\Phi}(s); \\ \frac{1}{nb(s)}\sum_{i=1}^{n}\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}^\top &\xrightarrow{p} \boldsymbol{\Sigma}(s)\varphi_K\int\frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})}d\mathbf{z}. \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Proof. By the symmetry of the kernel function $K(\cdot)$ and Taylor's expansion, we have

$$\begin{aligned}
& E\left[\frac{1}{nb(s)}\sum_{i=1}^{n}\frac{\partial \mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}}{\partial \boldsymbol{\beta}(s)}\right] \\
= \quad & E\left[\frac{1}{nb(s)}\sum_{i=1}^{n}\frac{\partial \mathbf{h}_i\{\boldsymbol{\beta}_0(s)\}}{\partial \boldsymbol{\beta}(s)}\frac{K_i(s)}{f(s|\mathbf{Z}_i)}\right] \\
= \quad & \frac{1}{b(s)}E\left\{\frac{\boldsymbol{\Phi}(s)K(s)}{f(s|\mathbf{Z})}\right\} = \frac{\boldsymbol{\Phi}(s)}{b(s)}\int\int K\Big(\frac{v-s}{b}\Big)\frac{f_{S,\mathbf{Z}}(v,\mathbf{z})}{f(s|\mathbf{z})}dvd\mathbf{z} \\
\overset{u=(s-s)/b(s)}{=} \quad & \frac{\boldsymbol{\Phi}(s)}{b(s)}\int\int K(u)f_{S,\mathbf{Z}}\{s+ub(s),\mathbf{z}\}b(s)/f(s|\mathbf{z})dud\mathbf{z} \\
= \quad & \boldsymbol{\Phi}(s)\int\int K(u)[f_{S,\mathbf{Z}}(s,\mathbf{z})+uO\{b(s)\}]/f(s|\mathbf{z})dud\mathbf{z} \\
= \quad & \boldsymbol{\Phi}(s)\int\frac{K(u)f_{S,\mathbf{Z}}(s,\mathbf{z})}{f(s|\mathbf{z})}dud\mathbf{z}+O\{b(s)^2\} \\
= \quad & \boldsymbol{\Phi}(s)\int K(u)du\int\frac{f(s|\mathbf{z})f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})}d\mathbf{z}+O\{b(s)^2\} = \boldsymbol{\Phi}(s)+O\{b(s)^2\}
\end{aligned}$$

and it can be shown that variance of each element in $\{nb(s)\}^{-1}\sum_{i=1}^{n}\partial \mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}/\partial\boldsymbol{\beta}(s)$ and $\{nb(s)\}^{-1}\sum_{i=1}^{n}\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}\mathbf{g}_i\{\boldsymbol{\beta}_0(s)\}^{\top}$ is of order $\{nb(s)\}^{-1}$. Then by the result in (A.2), the desired results are proven. ♢

Proof of Theorem 1. By Taylor's expansion, we have

$$\mathbf{G}\{\widehat{\boldsymbol{\beta}}(s)\} = \mathbf{G}\{\boldsymbol{\beta}_0(s)\} + \dot{\mathbf{G}}\{\breve{\boldsymbol{\beta}}(s)\}\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\},$$

where $\mathbf{G}\{\boldsymbol{\beta}(s)\} = \sum_{i=1}^{n}\mathbf{g}_i\{\boldsymbol{\beta}(s)\}/nb(s)$, $\dot{\mathbf{G}}\{\boldsymbol{\beta}(s)\} = \partial\mathbf{G}\{\boldsymbol{\beta}(s)\}/\partial\boldsymbol{\beta}(s)$, and $\breve{\boldsymbol{\beta}}(s)$ lies between $\widehat{\boldsymbol{\beta}}(s)$ and $\boldsymbol{\beta}_0(s)$. Since $\dot{\mathbf{G}}\{\widehat{\boldsymbol{\beta}}(s)\}^{\top}\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\mathbf{G}\{\widehat{\boldsymbol{\beta}}(s)\} = 0$, we have

$$\dot{\mathbf{G}}\{\widehat{\boldsymbol{\beta}}(s)\}^{\top}\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\mathbf{G}\{\boldsymbol{\beta}_0(s)\} + \dot{\mathbf{G}}\{\widehat{\boldsymbol{\beta}}(s)\}^{\top}\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\dot{\mathbf{G}}\{\breve{\boldsymbol{\beta}}(s)\}\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\} = 0.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

It is rewritten as $\sqrt{nb(s)}\left\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\right\}=$

$$-\left[\dot{\mathbf{G}}\{\widehat{\boldsymbol{\beta}}(s)\}^\top\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\dot{\mathbf{G}}\{\breve{\boldsymbol{\beta}}(s)\}\right]^{-1}\dot{\mathbf{G}}\{\widehat{\boldsymbol{\beta}}(s)\}^\top\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\sqrt{nb(s)}\mathbf{G}\{\boldsymbol{\beta}_0(s)\}.$$

It follows from Lemmas 1 and 2 and Slutsky's theorem that

$$\sqrt{nb(s)}\left\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\right\}\xrightarrow{d}\mathcal{N}\Big(\mathbf{0},\varphi_K\int\frac{f_{\mathbf{Z}}(\mathbf{z})}{f(s|\mathbf{z})}d\mathbf{z}\{\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)\boldsymbol{\Phi}(s)\}^{-1}\Big).$$

Now, we show that $Q\{\widehat{\boldsymbol{\beta}}(s)\}$ converges to the chi-squared distribution with $(p_s+2)(D-1)$ degrees of freedom. By Taylor's expansion and Lemma 2, we have

$$\begin{aligned}\mathbf{G}\{\widehat{\boldsymbol{\beta}}(s)\}&=\mathbf{G}\{\boldsymbol{\beta}_0(s)\}+\dot{\mathbf{G}}\{\boldsymbol{\beta}_0(s)\}\left\{\widehat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}_0(s)\right\}+o_p(1)\\&=[\mathbf{I}_{(p_s+2)D}-\boldsymbol{\Phi}(s)\{\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)\boldsymbol{\Phi}(s)\}^{-1}\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)^{-1}]\mathbf{G}\{\boldsymbol{\beta}_0(s)\}+o_p(1).\end{aligned}\tag{A.3}$$

By plugging (A.3) in $Q\{\widehat{\boldsymbol{\beta}}(s)\}=nb(s)\mathbf{G}\{\widehat{\boldsymbol{\beta}}(s)\}^\top\mathbf{V}\{\widehat{\boldsymbol{\beta}}(s)\}^{-1}\mathbf{G}\{\widehat{\boldsymbol{\beta}}(s)\}$ and Lemma 2, $Q\{\widehat{\boldsymbol{\beta}}(s)\}$ is rewritten as $Q\{\widehat{\boldsymbol{\beta}}(s)\}=\mathbf{H}\{\boldsymbol{\beta}_0(s)\}^\top\mathbf{S}_1(s)\mathbf{H}\{\boldsymbol{\beta}_0(s)\}+o_p(1)$, where $\mathbf{S}_1(s)=\mathbf{I}_{(p_s+2)D}-\boldsymbol{\Sigma}(s)^{-1/2}\boldsymbol{\Phi}(s)\{\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)^{-1}\boldsymbol{\Phi}(s)\}^{-1}\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)^{-1/2}$ and $\mathbf{H}\{\boldsymbol{\beta}(s)\}=\sqrt{nb(s)}\boldsymbol{\Sigma}(s)^{-1/2}\mathbf{G}\{\boldsymbol{\beta}_0(s)\}$. Following Lemma 1, $\mathbf{H}_n\{\boldsymbol{\beta}_0(s)\}$ converges to the standard multivariate normal distribution and $\mathbf{S}_1(s)$ is an idempotent and symmetric matrix with trace equal to $(p_s+2)(D-1)$. Consequently, $Q\{\widehat{\boldsymbol{\beta}}(s)\}$ converges to the chi-squared distribution with $(p_s+2)(D-1)$ degrees of freedom. $\diamondsuit$

Proof of Theorem 2. We let $\boldsymbol{\beta}(s)=(\beta_0,\ldots,\beta_{p_s},\beta_{p_s+1})^\top=(\boldsymbol{\beta}^*(s)^\top,\beta_{p_s+1})^\top$. Under $H_0:\beta_{p_s+1}=0$, the true parameter vector of $\boldsymbol{\beta}(s)$ is $\boldsymbol{\beta}_0(s)=(\boldsymbol{\beta}_0^*(s)^\top,0)^\top$. By a similar manner

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

in (A.1), the test statistic $\mathcal{W}(s) = Q\{(\widetilde{\boldsymbol{\beta}}_0^*(s)^\top, 0)\} - Q\{(\widehat{\boldsymbol{\beta}}^*(s)^\top, \widehat{\beta}_{p_s})\}$ is written as

$$\mathcal{W}(s) = \mathbf{H}\{\boldsymbol{\beta}_0(s)\}^\top \boldsymbol{\Sigma}(s)^{-1/2}\{\mathbf{S}_2(s) - \mathbf{S}_3(s)\}\boldsymbol{\Sigma}(s)^{-1/2}\mathbf{H}\{\boldsymbol{\beta}_0(s)\} + o_p(1),$$

where $\mathbf{S}_2(s) = \boldsymbol{\Phi}(s)\{\boldsymbol{\Phi}(s)^\top\boldsymbol{\Sigma}(s)^{-1}\boldsymbol{\Phi}(s)\}^{-1}\boldsymbol{\Phi}(s)^\top$ and $\mathbf{S}_3(s) = \boldsymbol{\Phi}^*(s)\{\boldsymbol{\Phi}^*(s)^\top\boldsymbol{\Sigma}(s)^{-1}\boldsymbol{\Phi}^*(s)\}^{-1}\boldsymbol{\Phi}^*(s)^\top$ with $\boldsymbol{\Phi}^*(s) = E[\partial \mathbf{h}_i\{\boldsymbol{\beta}_0^*(s)\}/\partial\boldsymbol{\beta}^*(s)\}]$. Since $\mathbf{S}_2(s)$ and $\mathbf{S}_3(s)$ are idempotent and symmetric matrices with trace equal to $(p_s + 2)$ and $(p_s + 1)$, respectively. As consequence, $\mathcal{W}(s)$ converges to the chi-squared distribution with one degree of freedom. ♢

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Table 1: Kernel-weighted average and double-weighted average of five baseline covariates at $s = 0.75$ and 1.5, respectively.

	Kernel weight		Double weight	
Covariate	$s = 0.75$	$s = 1.5$	$s = 0.75$	$s = 1.5$
Female (%)	56.6	73.9	58.1	61.1
CD4	262.9	268.1	265.3	265.1
Hemoglobin	11.30	11.42	11.38	11.34
BMI	22.08	21.33	21.71	21.46
Age	37.79	37.34	38.10	37.64

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Table 2: Averages of bias and mean squared error (MSE) of estimates under the AR(1), compound symmetry (CS), and independent (IN) working correlation structures when confounders exist (top) and do not (bottom).

Confounders Exist		BIAS		MSE	
		$Q\{\boldsymbol{\beta}(s)\}$	$Q_K\{\boldsymbol{\beta}(s)\}$	$Q\{\boldsymbol{\beta}(s)\}$	$Q_K\{\boldsymbol{\beta}(s)\}$
AR(1)	$\beta_0(s)$	-0.0016	0.0208	0.0028	0.0038
	$\beta_1(s)$	0.0052	0.0419	0.0041	0.0060
	$\beta_2(s)$	-0.0051	-0.0661	0.0089	0.0145
CS	$\beta_0(s)$	-0.0005	0.0207	0.0031	0.0039
	$\beta_1(s)$	0.0064	0.0451	0.0047	0.0069
	$\beta_2(s)$	-0.0085	-0.0719	0.0111	0.0161
IN	$\beta_0(s)$	-0.0080	0.0130	0.0034	0.0039
	$\beta_1(s)$	0.0198	0.0652	0.0081	0.0116
	$\beta_2(s)$	-0.0388	-0.1020	0.0235	0.0310
Confounders Not exist		BIAS		MSE	
		$Q\{\boldsymbol{\beta}(s)\}$	$Q_K\{\boldsymbol{\beta}(s)\}$	$Q\{\boldsymbol{\beta}(s)\}$	$Q_K\{\boldsymbol{\beta}(s)\}$
AR(1)	$\beta_0(s)$	-0.0052	-0.0051	0.0031	0.0033
	$\beta_1(s)$	0.0072	0.0078	0.0039	0.0040
	$\beta_2(s)$	-0.0067	-0.0075	0.0088	0.0090
CS	$\beta_0(s)$	-0.0054	-0.0053	0.0031	0.0033
	$\beta_1(s)$	0.0074	0.0078	0.0048	0.0050
	$\beta_2(s)$	-0.0089	-0.0092	0.0108	0.0110
IN	$\beta_0(s)$	-0.0118	-0.0117	0.0035	0.0038
	$\beta_1(s)$	0.0231	0.0238	0.0069	0.0072
	$\beta_2(s)$	-0.0399	-0.0401	0.0218	0.0220

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

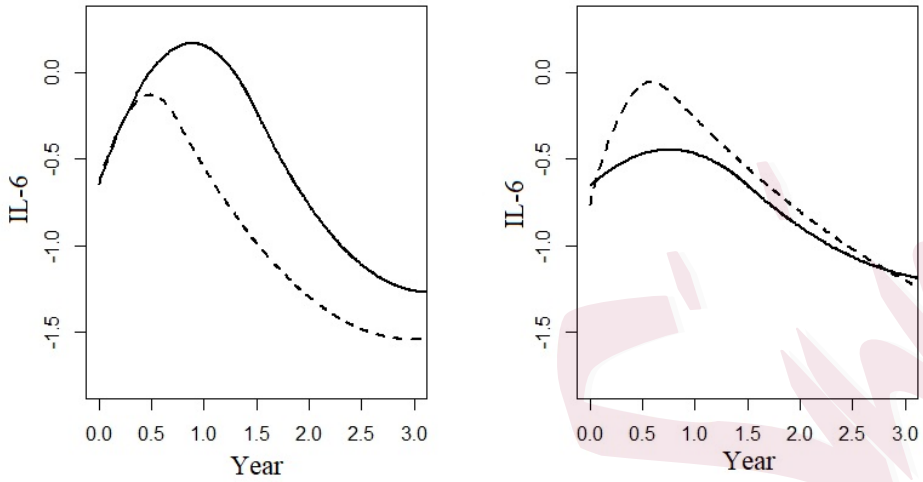

Figure 1: Fitted mean IL-6 trajectories using $Q\{\boldsymbol{\beta}(s)\}$ (left) and $Q_K\{\boldsymbol{\beta}(s)\}$ (right) at $s = 0.75$ (dashed curves) and $s = 1.5$ (solid curves), respectively, under the AR(1) working correlation structure.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

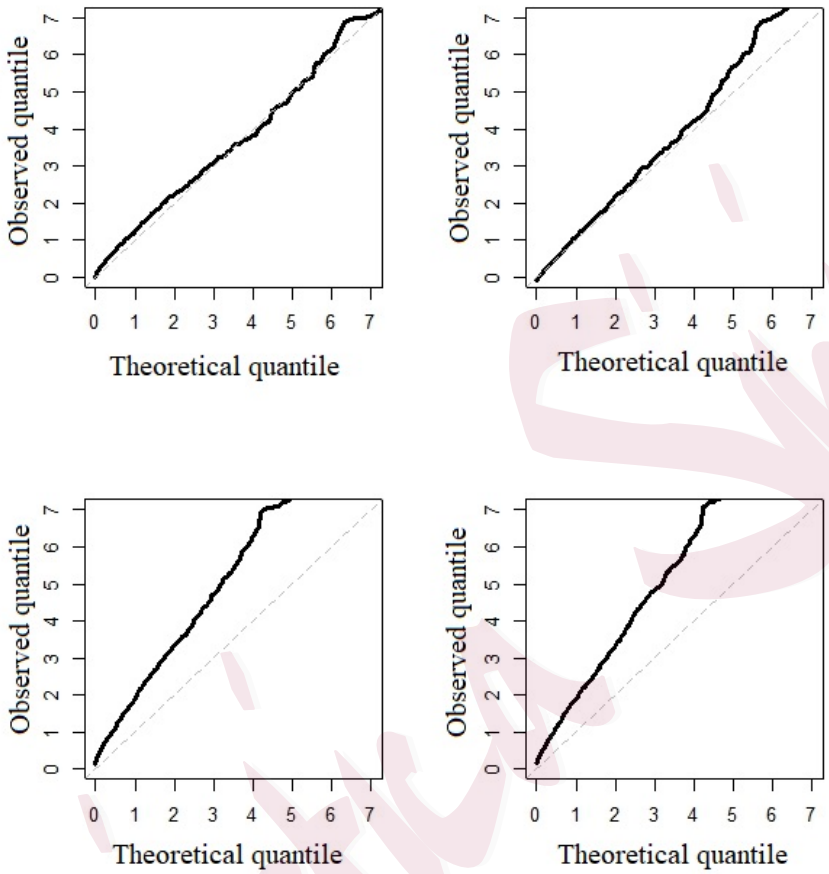

Figure 2: Quantile–quantile plots for the chi-squared distribution with one degree of freedom versus the test statistic based on $Q\{\boldsymbol{\beta}(s)\}$ (top) and $Q_K\{\boldsymbol{\beta}(s)\}$ (bottom), when the null hypothesis is true under the AR(1) (left) and compound symmetry (right) structures.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0226

Department of Biostatistics, University of Iowa, Iowa City, IA 52242

E-mail: (hyunkeun-cho@uiowa.edu)

Department of Statistics, Miami University, Oxford, OH 45056

E-mail: (kims20@miamioh.edu)