IMPROVING SEMIPARAMETRIC ESTIMATION OF LONGITUDINAL DATA WITH COVARIANCE FUNCTION

by

Fang Fang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2018

Approved by:

Dr. Yanqing Sun

Dr. Jiancheng Jiang

Dr. Shaoyu Li

Dr. Weidong Tian

ABSTRACT

FANG FANG. Improving Semiparametric Estimation of Longitudinal Data with Covariance Function. (Under the direction of DR. YANQING SUN)

In this dissertation, we aim to improve efficiency of estimation in longitudinal data under generalized semi-parametric varying-coefficient models.

First, we investigate a profile weighted least square approach for model estimation by utilizing within subject correlations. Several methods for incorporating the within subject correlations are explored, including quasi-likelihood approach(QL), minimum generalized variance approach (MGV), the quadratic inference function approach (QIF) and newly proposed weighted least square approach (WLS). Our simulation study shows that the covariance assisted estimation is more efficient then working independence approach(WI).

Second, we apply the above methods to more complex generalized semiparametric varying-coefficient models that not only describe time-constant effects and time-varying effects as above but also model covariate-varying effects. The asymptotic properties of the estimators are derived theoretically. The simulation study demonstrates similar results as above.

The proposed estimation methods are applied to two real data sets. One is ACTG 244 clinical trial, the other is the STEP study with MITT cases. Both show that our methods by using correlation structure in estimation give more efficient estimation and provide more information about the data, and have broad applications in biomedical studies where within subject correlation often exists.

ACKNOWLEDGMENTS

First and foremost I want to thank my advisor Dr. Yanqing Sun. Every Friday afternoon, she would spent hours with me discussing about my research. Her rigorous academic attitude and ways of thinking have contributed immensely to my personal and professional life.

Second, I would like to thank the committee members, Drs. Jiancheng Jiang, Shaoyu Li, and Weidong Tian for their constructive comments and valuable suggestions. Also a thank you to all the honorable professors in our Mathematics and Statistics Department. They convincingly conveyed a spirit of adventure in regard to research and an excitement in regard to teaching. Owing to their guidance in courses like Real Analysis, Probability Theory and Theory of Statistics, I am able to understand the theoretical proofs in literature and apply them into my research. Not forgetting to thank Professor Runze Li in Penn State University for sharing his simulation programs.

Third, I also would like to thank Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center for providing the data and sharing his expertise in the data applications. This research was partially supported by NIAID NIH award number R37AI054165, and National Science Foundation grant DMS-1513072.

Last but not least, I am deeply grateful for my family's support especially my parents-in-law who help me take care of my two little boys.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

In biomedical, epidemiological, social and economical research, where data are often collected at different time points for different subjects, we need to use longitudinal data or more generally clustered data. However, for longitudinal data or clustered data, there is often serial correlation between repeated measurements for the same subject, how to incorporate such with-subject correlation into estimation procedure is a major problem. In AIDS clinical trials, for example, viral loads and CD4 counts are measured repeatedly during the course of studies. These biomarkers within the same subject are often correlated positively. An important objective of the AIDS clinical trials is to examine treatment effectiveness on these longitudinal biomarkers.

In this dissertation, we use a generalized semiparametric varying-coefficient model to study treatment effects in such clinical trials, and improve the estimation efficiency by utilizing within-subject correlation that often exists in longitudinal data.

1.1 Motivation

A motivating example is a historical case study of antiretroviral treatment regimens, ACTG 244. Zidovudine (ZDV) was the first drug approved for treatment of HIV infection. Initial approval was based on evidence of a short-term survival advantage over placebo when zidovudine was given to patients with advanced HIV disease. Shortly after that, zidovudine resistance was associated with disease progression mea-

sured by a rise in plasma virus and decline in CD4 cell counts in both children and adults receiving zidovudine monotherapy (Larder et al., 1991). Subsequent studies suggested benefits of switching patients to treatments that combined ZDV with didanosine (ddI) or with ddI plus nevirapine (NVP). ACTG 244 enrolled subjects receiving ZDV monotherapy and monitored their HIV in plasma bi-monthly for the T215Y/F mutation. When a subject's viral population developed the 215 mutation, the subject was randomized to continue ZDV, add ddI or add ddI plus NVP.

An important question is whether and how the treatment switching has any beneficial effects in treating the HIV infected patients. We investigate the treatment switching problem under the generalized semiparametric varying-coeffcient model for longitudinal data. Varying-coefficient models, first proposed by Hastie and Tibshirani (1993), are appealing in longitudinal studies as they help to explore the extent to which covariates effects change with time comparing with parametric models. It allows effective simultaneous modeling of parametric effects for some covariates and varying coefficient effects for other covariates. For the varying-coefficient effects, some may depend on time called time-varying effects, some may depend on other covariates other than time, which called covariate-varying effects. Generalized varying-coefficient models provide a rich family of models with different link functions, including categorical response data where little work has been done. Here we apply generalized varying coefficient model to consider such treatment switching effects, which represents a covariate-varying covariate effect with the exposure modifying variable the time since treatment switching. What's more, associated with inferences are the standard errors of the estimated coefficients. To make predictions

or inferences more accurate, we want the standard errors to be as small as possible. The estimators obtained when considering correlation structure in estimating equations will reduce the standard error significantly in general while maintain the consistency.

These methods further have broad applications since treatment switching is common in medical studies, including switching antiretroviral therapies in response to results of viral load and drug resistance testing (Grabar et al., 2000), and, very generally, switching therapies for chronic diseases based on biomarker response results.

1.2 Literature Review

For longitudinal data, to explore the possible time-dependent effects of some covariates, we have to use time-varying coefficient model which was proposed by Hastie and Tibshirani (1993). It allows some covariates' effect on responses changes with time. In more general setting, the time in time-varying coefficient model does not have to be time, but can be any time-dependent covariate. By relaxing assumptions of time-varying coefficient model, varying-coefficient models can explore how covariates affect responses depend on other time-dependent covariates, thus to widen time-varying coefficient model's applicability.

Cai et al. (2000) applied varying-coefficient models to generalized linear models, which works both for continuous and discrete responses under a generalized linear model framework. The conditional distribution of response belongs to an exponential family and a known transform of the underlying regression function is linear. More recently, generalized semiparametric varying-coefficient models are often used in lon-

gitudinal study in econometrics, biomedical research, and epidemiology. As shown in Model (2.1), the mean of outcome variable Y depends on some covariates Z parametrically and some covariates X nonparametrically. In many instances, this model is desirable since it retains the flexibility of nonparametric modeling and parsimony and ease of interpretation of parametric modeling. We can make inference of Z while making minimal assumptions on the effects of X when we don't know the effects of X are linear or nonlinear. Moreover, with appropriately chosen link function, this model can be applied to both continuous and discrete responses. For instance, if Y is continuous, we use identity link, if Y is bernoulli we use logit link and if Y follows a counting process we can choose logarithm link.

$$\mu_i(t) = E\{Y_i(t)|X_i(t), Z_i(t)\} = g^{-1}\{\alpha^T(t)X_i(t) + \beta^T Z_i(t)\}, \tag{1.1}$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters. We assume that,

$$E\{Y_i(t)|X_i(t), Z_i(t), X_i, Z_i\} = E\{Y_i(t)|X_i(t), Z_i(t)\} = \mu_i(t) \tag{1.2}$$

(see Pepe and Couper (1997) for a discussion of this assumption).

Model (2.1) is very flexible since it can be reduced to many important statistical models in special settings. For example, when $\beta = 0$, it becomes a generalized nonparametric model for repeated measurements (Lin and Carroll, 2000; Wang, 2003). When $p_1 = 1$ and $X_i \equiv 1$, it becomes a generalized partially linear model for clustered

data (Lin and Carroll, 2001b; He et al., 2002, 2005; Wang et al., 2005; Huang et al., 2007). Furthermore if we let $g(\cdot)$ be an identity link, it becomes a partially linear model (Härdle et al., 2012).

In model (2.1), T can be multiple dimensional to include multiple nonparametric curves. Due to the curse of dimensionality, here we assume T is univariate for simplicity. The first component of X is set to be 1 to include a nonparametric baseline function.

Several authors have applied above model to independent data (Carroll et al., 1997; Hastie and Tibshirani, 1990; Severini and Staniswalis, 1994), and they used kernel method to estimate $\alpha(t)$ and the profile likelihood-based method to estimate β. A lot of work has been done to estimate the unknown nonparametric function $\alpha(t)$ and parametric parameter β. Zeger and Diggle (1994), Severini and Staniswalis (1994) and Lin and Carroll (2001a,b) estimated $\alpha(t)$ using a kernel method by ignoring the within-subject correlation while estimating β using weighted least squares by accounting for the within-subject correlation, and Lin and Carroll (2000) showed that by ignoring the within-subject correlation will achieve the most efficient estimation of $\alpha(t)$ when standard kernel methods are used. To approximate $\alpha(t)$, the pth local polynomial kernel smoothing can be used. When $p = 1$, it's local linear kernel smoothing; when $p = 0$, it becomes local average kernel method. Zeger and Diggle (1994) considered a semiparametric mixed model, where they decomposed the error into a random effect part for the serial correlation within subjects and a random measurement error part, and used a back-fitting algorithm and local-average kernel smoothing for estimation. Severini and Staniswalis (1994) also used local average

kernel to approximate $\alpha(t)$ and profile-kernel GEEs to estimate β.

Lin and Carroll (2001a) and Lin and Carroll (2001b) both use local linear kernel. The former showed that when nonparametric part is only a smooth function of a subject-level covariate (i.e., $T_i(t)$ doesn't change with t), the most efficient estimators of $\alpha(t)$ and β are obtained by incorporating the actual correlation matrix into both the estimation equations. However, Lin and Carroll (2001b) showed that when $T_i(t)$ and $Z_i(t)$ are both time-varying covariates that changes with t, to achieve a consistent estimate of β, we need either assume working independence or undersmooth $\alpha(t)$(i.e. choosing a bandwidth smaller than the one selected by cross validation) when using the actual correlation matrix. Such consistent estimator is still not semiparametric efficient. To estimate $\alpha(t)$, when covariate $X_i(t)$ is time-invariant, we need to use the actual correlation matrix to achieve a more efficient estimate of $\alpha(t)$; when covariate $X_i(t)$ is time-varying, then we need to assume working independence.

Chen and Jin (2006) avoided the above undersmoothing problem through a (non-smooth) piecewise local polynomial approximation of $\alpha(t)$, and Huang et al. (2006) used (smooth) spline approximations to approximate the nonparametric part and then apply GEE to estimate the parameters. Both Chen and Jin (2006) and Huang et al. (2006) achieved the semiparametric efficiency by accounting for with-subject correlation and only assumed conditional moment restrictions without the strict multivariate Gaussian error assumption. Wang (2003) and Wang et al. (2005) revised the original kernel GEE estimating equation in Lin and Carroll (2001b) to incorporate the true correlation matrix for both $\alpha(t)$ and β and achieved semiparametric efficiency in Gaussian case. However the improvement of efficiency from Lin and Carroll (2001b)'s

estimator to Wang et al. (2005)'s is not significant comparing with the improvement from independence case to Lin and Carroll (2001b)'s.

When using profile-kernel method to estimate $\alpha(t)$ and β, we need to estimate the true covariance function of Y and iterate with estimation of $\alpha(t)$ and β until convergence. A more efficient estimation of $\alpha(t)$ and β depends on the precision of estimation of the covariance function of Y. How to estimate the covariance function is another key problem in model (2.1). All the work above assume a simple structure of covariance function that can be decomposed into marginal variance of Y and symmetric correlation matrix. The symmetric correlation matrix depends on a nuisance parameter θ that can be estimated by method of moments in parametric generalized estimating equations(GEE) method (Liang and Zeger, 1986). Fan et al. (2007) proposed another two methods to estimate θ, one is optimizing the quasi-likelihood (QL), and the other one is minimizing the generalized variance of β (MGV).

The quasi-likelihood (QL) method proposed in Fan et al. (2007) constructs the likelihood based on a normal distribution, since their model is a special case of our model with an identity link for continuous responses, however in the generalized linear model with non-identity link, the error part doesn't follow a normal distribution but possibly an Bernoulli or a Poisson distribution for discrete responses, thus we propose Weighted least squares (WLS) method to circumvent this restriction.

Either using method of moments (Liang and Zeger, 1986), or quasi-likelihood method (QL) and minimizing generalized variance of β (MGV) in Fan et al. (2007) or our weighted least square method(WLS), the existence of $\hat{\theta}$ will not be guaranteed in some situations and $\hat{\theta}$ may not be close to the true value when the working

correlation structure is misspecified. Qu et al. (2000) introduced the quadratic inference function method(QIF) to avoid the above issues. They represent the inverse of working correlation matrix by a linear combination of basis matrices, and then using a quadratic inference function which follows the generalized method of moments (GMM) in Hansen (1982) to construct estimating equations. In this way, they don't need to estimate the nuisance parameters in correlation matrix, and can achieve an asymptotically efficient estimator for β even the correlation structure is misspecified. Qu and Song (2004) showed a better robustness property of QIF method compared to GEE. Qu and Li (2006) applied QIF method to varying-coefficient models for longitudinal data and used penalized splines to approximate the varying-coefficient part. Bai et al. (2008) generalized QIF method to partial linear models for longitudinal data and used B-splines to approximate the nonparametric part.

Qi et al. (2016) has done some research assume working independence within subjects for this model, however, for repeated measurements of longitudinal data, correlation within subjects often exists and is helpful to be used to improve the estimation efficiency and prediction. Incorporation of the within-cluster correlation has been applied to various models along the lines of generalized estimating equations (see,e.g., Lin and Carroll, 2001a,b; Chen and Jin, 2006; Wang et al., 2005; Huang et al., 2007). However, they focus on either a partially linear regression model with an identity link under the multivariate Gaussian assumption, or a generalized partial linear model with only a time trajectory as the nonparametric part. Our model contains not only a time-varying effect on some covariates but also a covariate-varying effect on other covariates with a general link, thus can characterize the treatment effects and

treatment switching effects.

CHAPTER 2: SEMIPARAMETRIC ESTIMATION OF LONGITUDINAL DATA—A COMPARATIVE STUDY

In this chapter we apply different methods to generalized semiparametric varying-coefficient models for longitudinal data and compare their performance in different settings. In Section 2.2 and 2.3, we describe the above methods in detail. In Section 2.4, we assess the finite sample performance of different methods with Monto Carlo simulation and illustrate the strength and limitations of them for continuous responses. In Section 2.5, we do similar empirical study for discrete responses. In Section 2.6, we apply these methods to a real data problem. Concluding remarks are given in Section 2.7.

2.1 Model

In this chapter we study a generalized time-varying coefficient model for longitudinal data, to explore the possible time-dependent effects of some covariates. It allows some covariates' effect on responses changes with time. In more general setting, the time in time-varying coefficient model does not have to be time, but can be any time-dependent covariate. In Chapter 3, we will turn to a generalized varying-coefficient model by relaxing assumptions of time-varying coefficient model, which can explore how covariates affect responses depend on other time-dependent covariates.

Model (2.1) is a generalized time-varying coefficient model, where the conditional expectation of response Y depends on some covariates Z parametrically and the other

covariates X nonparametrically.

$$\mu_i(t) = E\{Y_i(t)|X_i(t), Z_i(t)\} = g^{-1}\{\alpha^T(t)X_i(t) + \beta^T Z_i(t)\}, \tag{2.1}$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters. We assume that,

$$E\{Y_i(t)|X_i(t), Z_i(t), X_i, Z_i\} = E\{Y_i(t)|X_i(t), Z_i(t)\} = \mu_i(t) \tag{2.2}$$

(see Pepe and Couper (1997) for a discussion of this assumption).

In model (2.1), T can be multiple dimensional to include multiple nonparametric curves. Due to the curse of dimensionality, here we assume T is univariate for simplicity. The first component of X is set to be 1 to include a nonparametric baseline function.

2.2 Profile-kernel Estimation of Regression Coefficients

To estimate the coefficients in above model (2.1), we use profile-kernel method to estimate nonparametric function $\alpha(t)$ and parametric coefficient vector β in an iterative nature.

First, given β, one can use *pth* local polynomial smoothing to approximate $\alpha(t)$ (Lin and Carroll, 2000), here we use local linear smoothing (i.e. p=1)(Lin and Carroll, 2001a,b; Wang et al., 2005; Fan et al., 2007). Local linear smoothing has some nice properties: good boundary behavior (Fan and Gijbels, 1996), high statistical efficiency and design adaptation (Fan, 1993). At each t_0, let $\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t -$

$t_0)^2)$ be the first order Taylor expansion of $\alpha(\cdot)$ for $t \in \mathcal{N}_{t_0}$, a neighborhood of t_0, where $\dot{\alpha}(t_0)$ is the derivative of $\alpha(t)$ at $t = t_0$. Denote $\alpha^*(t_0) = (\alpha^T(t_0), \dot{\alpha}^T(t_0))^T$, $X_i^*(t, t - t_0) = X_i(t) \otimes (1, t - t_0)^T$, where $\otimes$ is the Kronecker product.

For $t \in \mathcal{N}_{t_0}$, model (2.1) can be approximated by

$$\tilde{\mu}(t, t_0, \alpha^*(t_0), \beta | X_i, Z_i) = g^{-1}\{\alpha^{*T}(t_0) X_i^*(t, t - t_0) + \beta^T Z_i(t)\}, \tag{2.3}$$

At each t_0 and for fixed β, we propose the following local linear estimating function for $\alpha^*(t_0)$:

$$U_\alpha(\alpha^*; \beta, t_0) = \sum_{i=1}^{n} X_i^*(t_0)^T \Delta_i(t_0) K_{ih}^{1/2}(t_0) V_{1i}^{-1}(t_0) K_{ih}^{1/2}(t_0) \left[Y_i - \tilde{\mu}_i^*\right], \tag{2.4}$$

where for simplicity we denote $Y_i(T_{ij}) = Y_{ij}$, $\mu_i(T_{ij}) = \mu_{ij}$, $X_i(T_{ij}) = X_{ij}$ and $Z_i(T_{ij}) = Z_{ij}$, let $Y_i = (Y_{i1}, \cdots, Y_{iJ_i})^T$, $\mu_i^* = (\mu_{i1}^*, \cdots, \mu_{iJ_i}^*)^T$, $\mu_{ij}^* = \mu(T_{ij}, \alpha^*(t_0), \beta | X_{ij}, Z_{ij})$, $\Delta_i = diag\{\dot{\mu}_{ij}\}$, $\dot{\mu}(\cdot)$ is the first derivative of $\mu(\cdot)$, $V_{1i}^{-1}(t_0)$ is a nonnegative weight process, $K(\cdot)$ is a kernel function, $h = h_n > 0$ is a bandwidth parameter and $K_h(\cdot) = K(\cdot/h)/h$, $K_{ih}(t_0) = diag\{K_h(T_{ij} - t_0)\}$. The solution to the equation $U_\alpha(\alpha^*; \beta, t_0) = 0$ is denoted by $\tilde{\alpha}^*(t_0, \beta)$. Let $\tilde{\alpha}(t_0, \beta)$ be the first p_1 components of $\tilde{\alpha}^*(t_0, \beta)$.

Second, given estimated $\alpha(t)$ derived above, the profile weighted least-squares estimator $\hat{\beta}$ is obtained by minimizing the following profile least-squares function:

$$\ell_\beta(\beta) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \hat{\mu}_i(\beta)]^T V_{2i}^{-1} [Y_i - \hat{\mu}_i(\beta)], \tag{2.5}$$

where V_{2i}^{-1} is the inverse of working covariance matrix, and Y_i, μ_i are defined the same way as those in (2.4) except that they are evaluated at $\hat{\mu}_{ij}(\beta) = g^{-1}\{\tilde{\alpha}^T(T_{ij}, \beta) X_{ij} +$

$\beta^T Z_{ij}\}$.

How to choose the weight process $V_{1i}^{-1}(t_0)$ in equation (2.4) and the working covariance matrix in V_{2i}^{-1} in equation (2.5)? There are several work addressing this issue in literature (Lin and Carroll, 2001a,b; Wang et al., 2005; Fan et al., 2007). They all assume the working covariance function in equation (2.4) denoted by V_{ki}, where $k = 1, 2$ is decomposed as below for simplicity:

$$V_{ki} = A_i R_{ki}(\theta) A_i \tag{2.6}$$

where $A_i = diag\{\sigma(T_{ij})\}$ is the square root of marginal variance of Y_i, $R_{ki}(\theta)$ is the working correlation matrix. All of these work agreed on that a more efficient estimation of nonparametric part $\alpha(t)$ is achieved by ignoring the correlation within subjects, while the most efficient estimation for β is obtained by choosing the working covariance matrix as the inverse of true covariance matrix of response Y. Thus we let $R_{1i}(t)$ in $V_{1i}(t)$ as defined in equation (2.2) be an identity matrix in equation (2.4), and V_{2i}^{-1} be the inverse of covariance matrix of response Y in equation (2.5). Here we assume the covariance function of Y is known. In the following section we will discuss how to estimate this covariance function of Y using different methods.

Taking the derivative of $\ell_\beta(\beta)$ with respect to β leads to score function for β.

$$U_\beta(\beta) = \sum_{i=1}^{n} \{\frac{\partial \tilde{\alpha}(T_i, \beta)}{\partial \beta} X_i + Z_i\}^T \Delta_i V_{2i}^{-1} \left[Y_i - \hat{\mu}_i(\beta)\}\right] \tag{2.7}$$

where $\frac{\partial \tilde{\alpha}(T_i,\beta)}{\partial \beta} X_i = \{\frac{\partial \tilde{\alpha}^T(T_{i1},\beta)}{\partial \beta} X_i(T_{i1}), \cdots, \frac{\partial \tilde{\alpha}^T(T_{iJ_i},\beta)}{\partial \beta} X_i(T_{iJ_i})$. Here $\frac{\partial \tilde{\alpha}(t,\beta)}{\partial \beta}$ be the first p_1 components of $\frac{\partial \alpha^*(t,\beta)}{\partial \beta}$ which can be expressed in terms of the partial derivatives of $U_\alpha(\alpha^*; \beta, t)$ at $\alpha^* = \tilde{\alpha}^*(t, \beta)$. Specifically, since $U_\alpha(\tilde{\alpha}^*(t, \beta); \beta, t) \equiv \mathbf{0}_{2p_1}$, it follows

that $\tilde{\alpha}^*(t,\beta)$ satisfies

$$\left\{\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \alpha^*}\frac{\partial \tilde{\alpha}^*(t,\beta)}{\partial \beta}+\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \beta}\right\}\Big|_{\alpha^*=\tilde{\alpha}^*(t,\beta)}=\mathbf{0}_{2p_1}.$$

Therefore,

$$\frac{\partial \tilde{\alpha}^*(t,\beta)}{\partial \beta}=-\left\{\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \alpha^*}\right\}^{-1}\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \beta}\Big|_{\alpha^*=\tilde{\alpha}^*(t,\beta)}, \tag{2.8}$$

where

$$\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \alpha^*}=-\sum_{i=1}^{n}X_i^*(t)^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\Delta_i(t)X_i^*(t), \tag{2.9}$$

and

$$\frac{\partial U_\alpha(\alpha^*;\beta,t)}{\partial \beta}=-\sum_{i=1}^{n}Z_i^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\Delta_i(t)X_i^*(t). \tag{2.10}$$

When we choose identity link for model (2.1), the score function (2.4) and (2.7) will reduced to an explicit form, otherwise they will not have closed form for other links of $g(\cdot)$. The Newton-Raphson iterative method can be used to solve the above score functions.

To ensure convergence in above iteration between estimation of $\alpha(t)$ and β, we use the independence case as the starting value, that is, let weight matrix V_{2i}^{-1} in (2.7) be identity matrix to get the initial estimation of $\alpha_0(t)$ and β_0. No more than five iterations, we achieve convergence in above profile-kernel algorithm. Estimation of β is not very sensitive of bandwidth h, as long as it is not too large to cause biases. We use k fold cross-validation to choose bandwidth. Fan and Li (2004) mentioned that $\alpha(t)$ cannot be estimated well at some tail of the observation times because of

sparsity. We adopted their suggestion to eliminate 5% of the data at the tail.

A more efficient estimation will be achieved when the weight matrix chosen in function (2.7) is close to the inverse of true covariance matrix of Y (Fan et al., 2007), thus in the following section we will discuss how to estimate the covariance function of Y.

2.3 Estimation of Covariance Function of Responses

To estimate covariance function of Y in model (2.1), there are different methods, such as generalized estimating equation(GEE) approach proposed by Liang and Zeger (1986), quasi-likelihood(QL) approach and minimum generalized variance(MGV) approach, both proposed by Fan et al. (2007), and weighted least squares(WLS) approach newly proposed here. Also there is another way to deal with weight matrix V_{2i}^{-1}. Qu et al. (2000) proposed quadratic inference function(QIF) approach which circumvent estimating covariance function of Y by approximating the weight matrix by a family of basis matrices. We will introduce these approaches briefly in this section.

The covariance function of Y_i denoted as Σ_i is decomposed as below for simplicity:

$$\Sigma_i = A_i R_i(\theta) A_i \tag{2.11}$$

where $A_i = diag\{\sigma(T_{ij})\}$ is the square root of marginal variance of Y_i, $R_i(\theta)$ is the correlation matrix. In the following we will discuss how to estimate the marginal variance and correlation matrix of Y_i respectively.

2.3.1 Estimation of Marginal Variance

To estimate $A_i = diag\{\sigma(T_{ij})\}$ in equation (2.3), we need to estimate marginal variance of Y_i. In GEE approach, it assumes that the marginal density of $Y_{ij} = y$ is from an exponential family, i.e.,

$$f(y) = exp[\{y\eta - a(\eta) + b(y)\}\phi], \tag{2.12}$$

and the first two moments of Y_{ij} are given by

$$\mu_{ij} = E(Y_{ij}) = a'(\eta_{ij}), \quad var(Y_{ij}) = a''(\eta_{ij})/\phi. \tag{2.13}$$

where $\eta_{ij} = \alpha^T(T_{ij})X_{ij} + \beta^T Z_{ij}$, $a'(\cdot) = g^{-1}(\cdot)$ and ϕ is the scale parameter. Given estimation of $\alpha(t)$ and β in Section 2.2, we obtain $\hat{\eta}_{ij} = \hat{\alpha}^T(T_{ij})X_{ij} + \hat{\beta}^T Z_{ij}$. The current Pearson residual is defined by $\hat{r}_{ij} = \{Y_{ij} - a'(\hat{\eta}_{ij})\}/\{sqrt(a''(\hat{\eta}_{ij}))\}$, and the scale parameter ϕ can be estimated by $\hat{\phi}^{-1} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\hat{r}_{ij}^2/(N-p_1)$ where $N = \sum J_i$ is the total number of observations and p_1 is the dimension of parameter β.

When Y is following a normal distribution, then $a'(\hat{\eta}_{ij}) = \hat{\mu}_{ij} = \hat{\alpha}^T(T_{ij})X_{ij} + \hat{\beta}^T Z_{ij}$, $a''(\hat{\eta}_{ij}) = 1$ and the marginal variance formula of Y is:

$$\widehat{var}(Y_{ij}) = \hat{\phi}^{-1} = \sum_{i=1}^{n}\sum_{j=1}^{J_i}\hat{r}_{ij}^2/(N-p_1). \tag{2.14}$$

When Y is a Poisson process, we use a log link in model $a'(\hat{\eta}_{ij}) = \hat{\mu}_{ij} = exp\{\hat{\alpha}^T(T_{ij})X_{ij} + \hat{\beta}^T Z_{ij}\}$, $\hat{\phi} = 1$, and the marginal variance formula of Y is:

$$\widehat{var}(Y_{ij}) = a''(\hat{\eta}_{ij}) = \hat{\mu}_{ij} = exp\{\hat{\alpha}^T(T_{ij})X_{ij} + \hat{\beta}^T Z_{ij}\}. \tag{2.15}$$

When Y is from a Bernoulli distribution, $a'(\hat{\eta}_{ij}) = \mu_{ij} = 1/1 + exp\{-[\hat{\alpha}^T(T_{ij})X_{ij} + \hat{\beta}^T Z_{ij}]\}$,

$\hat{\phi} = 1$, and the marginal variance formula of Y is:

$$\widehat{var}(Y_{ij}) = a''(\hat{\eta}_{ij}) = \hat{\mu}_{ij}(1 - \hat{\mu}_{ij}). \tag{2.16}$$

Fan et al. (2007) used univariate kernel smoothing to estimate marginal variance of normal random variable Y by the following formula:

$$\widehat{var}(Y_{ij}) = \hat{\sigma}^2(t) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i} \hat{r}_{ij}^2 K_{h1}(t - T_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_{h1}(t - T_{ij})} \tag{2.17}$$

The above formula is in the same spirit of GEE's method of moments approach, but adding a kernel weight when taking average which is more accurate for time-dependent and nonstationary random error process.

2.3.2 Estimation of Correlation Coefficient

To estimate correlation coefficient vector θ in correlation matrix $R(\theta)$, when the correlation matrix is unstructured, the GEE approach is using methods of moments, i.e.,

$$\hat{R}_{kl}(\theta) = \sum_{i=1}^{n} \hat{r}_{ik}(\theta)\hat{r}_{il}(\theta)/(N - p_1). \tag{2.18}$$

When the correlation structure is defined as special cases such as one-dependent, exchangeable and AR(1) structure, Liang and Zeger (1986) have derived formulas to calculate $\hat{\theta}$ for each specific case.

The above GEE approach in Liang and Zeger (1986) is based on a parametric model, but it can also be extended to kernel GEE approach to be applied to our semiparametric model (Lin and Carroll, 2001a,b; Wang, 2003; Wang et al., 2005). Kernel GEE approach is using local linear kernel function to approximate the non-

parametric part $\alpha(t)$ in model (2.1), thus to transform the model to a parametric form for each time point t, then above method of moments can be applied to estimate the covariance function as well. As shown in Lin and Ying (2001)'s simulation study, semiparametric estimators have an extremely high relative efficiencies relative to Parametric GEE method, being greater than 95% in all cases.

Fan et al. (2007) proposed quasi-likelihood (QL) approach to estimate correlation coefficient vector θ, that is, to optimize the following quasi-likelihood function.

$$\hat{\theta} = \arg\max_{\theta}\big(-\frac{1}{2}\sum_{i=1}^{n}\{log|R_i(\theta)| + \hat{r}_i^T A_i^{-1} R_i(\theta)^{-1} A_i^{-1}\hat{r}_i\}\big), \tag{2.19}$$

where R_i and A_i are defined the same as in equation (2.3), $\hat{r}_i = \{\hat{r}_{i1}, \ldots, \hat{r}_{iJ_i}\}$ is estimator for vector ϵ_i and $\hat{r}_{ij}$ is the same as the current Pearson residual defined in section (2.3.1). Fan et al. (2007) focused on continuous responses, but it can be extended to discrete responses such as counting process or binary responses as well.

In Fan et al. (2007), they also proposed minimum generalized covariance (MGV) approach to estimate correlation parameter vector θ by minimizing the generalized variance of $\hat{\beta}$, that is,

$$\hat{\theta} = \arg\min_{\theta} |\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)|, \tag{2.20}$$

where $\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)$ is the estimated covariance matrix of $\hat{\beta}$ that depends on estimated marginal variance $\hat{\sigma}^2$ and correlation parameter vector θ. The generalized covariance of $\hat{\beta}$ was defined as the determinant of $\Sigma_{\hat{\beta}}(\hat{\sigma}^2, \theta)$.

Based on Fan et al. (2007)'s quasi-likelihood(QL) approach, we propose the Weighted-Least-Square (WLS) approach to estimate θ by minimizing the weighted least squares,

that is,

$$\hat{\theta} = \arg\min_{\theta} \big(\hat{r}_i^T A_i^{-1} R_i(\theta)^{-1} A_i^{-1} \hat{r}_i\big), \tag{2.21}$$

For responses that don't follow a Gaussian process, we expect Weighted-Least-Square (WLS) approach will give a more robust estimation than Quasi-likelihood(QL)approach.

2.3.3 Quadratic Inference Function

It has been shown in literature that as long as $\hat{\theta}$ is $n^{1/2}$ consistent, it will not affect the asymptotic distribution of $\hat{\alpha}(t)$ and $\hat{\beta}$ in model(2.1). However, when we try to estimate the covariance function for Y using above GEE, QL, MGV or WLS approaches, the estimates of θ may be nonexistent or inconsistent, and the estimated correlation matrix may not be positive definite when the correlation structure is misspecified. Qu et al. (2000) proposed quadratic inference function (QIF) approach to circumvent the above problems. The inverse of the working correlation matrix was represented as a linear combination of basis matrices, that is

$$R^{-1} \approx a_1 M_1 + a_2 M_2 + \cdots + a_m M_m, \tag{2.22}$$

where M_1 is an identity matrix, and $M_2, \cdots, M_m$ are symmetric basis matrices which are determined by the structure of $R(\theta)$, and $a_1, \cdots, a_m$ are constant coefficients. The advantage of this approach is that it does not require estimation of nuisance parameters a_i's. They further stack above basis matrices of the correlation matrix to

construct a 'extended score' of β:

$$g_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta) = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^{n}\{\frac{\partial\tilde{\alpha}(T_i,\beta)}{\partial\beta}X_i + Z_i\}^T\Delta_i A_i^{-1}M_1A_i^{-1}\left[Y_i - \hat{\mu}_i(\beta)\right]\} \\ \vdots \\ \sum_{i=1}^{n}\{\frac{\partial\tilde{\alpha}(T_i,\beta)}{\partial\beta}X_i + Z_i\}^T\Delta_i A_i^{-1}M_mA_i^{-1}\left[Y_i - \hat{\mu}_i(\beta)\right]\} \end{pmatrix} \tag{2.23}$$

The quadratic inference function is constructed to be,

$$Q_n(\beta) = g_n' C_n^{-1} g_n \tag{2.24}$$

where $C_n = (1/n^2)\sum_{i=1}^{n} g_i(\beta)g_i(\beta)'$. Given estimated $\alpha(t)$ in section 2.2, the QIF estimator of β is

$$\hat{\beta} = \arg\min_{\beta} Q_n(\beta) \tag{2.25}$$

Qu et al. (2000) mentioned that QIF performs as well as GEE with the true correlation structure, however when the correlation structure is misspecified, QIF is still optimal among the family where the misspecified working correlation structure can be represented by the chosen basis matrices. Thus QIF is more robust and more efficient than GEE method.

2.3.4 Mixed correlation structure

When the correlation structure is known or defined as one specific form like exchangeable, $AR(1)$ or $ARMA(1,1)$ structure, then we can directly apply the methods discussed in section 2.3.2 to estimate θ since the correlation matrix $R(\theta)$ is a known function of coefficient vector θ. However when the correlation structure is unknown or a complex form, we used a mixed correlation structure to approximate the unknown

$R(\theta)$.

Borrowing the idea of Fan et al. (2007), we embed a given working correlation into a family of correlation matrix $\rho_0(s,t,\theta_0),\ldots,\rho_m(s,t)$, and approximate the true correlation matrix by a combination of the members in this family, that is,

$$\rho(s,t,\theta) = b_0\rho_0(s,t;\theta_0) + b_1\rho_1(s,t,\theta_1) + \cdots + b_m\rho_m(s,t,\theta_m), \tag{2.26}$$

where $\theta = (\theta_0, b_0, \theta_1, b_1, \ldots, b_m, \theta_m)$ and $b_0 + \cdots + b_m = 1$ with all $b_i \geq 0$.

This mixed correlation structure can be applied to above QL, MGV or WLS approaches by substituting the mixed correlation matrix in objective function (2.19), (2.20) and (2.21). In this way, we may estimate complicated unknown correlation structure. This can also be applied to QIF approach. We stack possible basic matrices of the unknown correlation matrix to construct a general quadratic inference function. In equation (2.22), M_1 is an identity matrix, and $M_2, \cdots, M_m$ are symmetric basis matrices which are combination of the basic matrices of the members in the correlation family.

2.4 Simulation Study for Continuous Response

First, we examine the performance of the various methods for continuous longitudinal responses for model (2.1) with identity link function.

$$Y_i(t) = \alpha^T(t)X_i(t) + \beta^T Z_i(t) + \epsilon_i(t), \tag{2.27}$$

where $0 \leq t \leq \tau$, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters, and $\epsilon_i(t)$ is a mean zero process. In this section, we conduct a simulation study to assess the performance of

above methods for estimation of model (2.27) in various model settings for different scenarios.

2.4.1 Simulation Models and Performance Comparisons

For longitudinal data, there are many possible formulations of how data were collected. The time points $\{t_{ij}\}$ are a random sample from a certain population. In Lin and Ying (2001), they proposed four different cases of the observation times: (1)Observation times are independent of covariates; (2) Observation times depend on covarites; (3)Observation times are fixed; (4) Observation times are scheduled but can be randomly missed.

Based on their discussion, we consider two categories: independent observation times (Observation I) and dependent observation times (Observation II). One example of independent observation times is that the measurements are taken at scheduled time points and any deviation from the schedule occurs in a completely random fashion. We use the following setting in Fan et al. (2007)'s simulation study for independent observation times.

(A1) Every individual has the same scheduled time points, {0,1,2,...,12}, each of which has a 20% probability of being skipped except time 0. Then a uniform $[0, 1]$ random variable is added to the nonskipped scheduled time points. Thus every subject has approximately $7 \sim 13$ observations with an average 11 and different time points. $X_1(t) \equiv 1$, $(X_2(t), Z_1(t))^T$ follows a bivariate normal distribution with mean 0, variance 1 and correlation 0.5 for a given t, and $Z_2(t)$ is a Bernoulli random variable with success probability 0.5 for each subject i and doesn't change within subjects to

mimic a treatment effect. For coefficients, $\beta = (1, 2)^T$, and $\alpha(t) = (\alpha_1(t), \alpha_2(t)) = (\sqrt{t/12}, \sin(2\pi t/12))$. The error part $\epsilon(t)$ is a Gaussian process with mean 0, variance changing with time $\sigma^2(t) = 0.5exp(t/12)$ and the correlation structure is $ARMA(1, 1)$, i.e., $corr(\epsilon(s), \epsilon(t)) = \gamma\rho^{|t-s|}$ for $s \neq t$. We let $\theta = (\gamma, \rho) = (0.85, 0.9)$, $\theta = (\gamma, \rho) = (0.85, 0.6)$, $\theta = (\gamma, \rho) = (0.85, 0.3)$ to consider strong, moderate and weak correlation respectively.

However the only difference here from Fan et al. (2007)'s model in their simulation example $Z_{2i}(t)$ does change within subject, i.e., $Z_{2i}(t)$ is a Bernoulli random variable with probability 0.5 within each subject. Here we consider Z_{2i} as a treatment effect just to diversify covariates' setting. Sometimes the observation time may not be random and may depend on covariates. To consider dependent observation times, we use the following setting in Lin and Ying (2001) and Sun et al. (2013).

(A2) Observation time T_i for each subject i is a Poisson process with the proportional mean rate $h_i = 0.6\exp(0.7Z_{2i})$, which depends on covariate Z_{2i}. Let C_i be the end of follow-up time or censoring time, whichever comes first. The responses for subject i can only be observed at time points before C_i. The censoring times C_i are generated from a uniform distribution on $[1.5, 8]$ and $0 \leq t \leq \tau$ with $\tau = 3.5$. There are approximately three observations per subject on $[0, \tau]$ and about 30% of subjects are censored before $\tau = 3.5$. Let $\alpha_0(t) = 0.5\sqrt{t}$, $\alpha_1(t) = 0.5\sin(t)$ and $\beta = (0.5, 1)^T$, $Z_{1i}(t)$ is an uniform random variable on$[0, 1]$, Z_{2i} is a Bernoulli random variable with success probability of 0.5 for each subject i, $X_{1i}(t) \equiv 1$, and $X_{2i}(t)$ is a Bernoulli random variable with success probability of 0.5 at each time point t within each subject i. The error $\epsilon_i(t)$ has a normal distribution with mean ϕ_i and variance ν^2, and

ϕ_i is $N(0,1)$. We can show that when $\nu = 0.5$, the true covariance matrix of $\epsilon_i(t)$ is exchangeable structure with correlation coefficient $\theta = 0.8$ and constant variance $\sigma^2 = 1.25$. We also consider moderate and weak correlation letting $\nu = 1$ and $\nu = 2$ such that correlation coefficient $\theta = 0.5$ and $\theta = 0.2$ respectively.

The simulation result of parameter estimation in Observation I with strong correlation ($(\gamma, \rho) = (0.85, 0.9)$) is shown in Table 1. Each entry of the table is calculated based on 1000 repetitions for 200 subjects. It summarizes the Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 by different approaches. Bias is calculated as the mean of 1000 estimates minus the true value of β, we expect it to be close to zero if the estimators are consistent. SEE is the sample standard error of 1000 estimates which can be regarded as the true standard errors except for Monte Carlo error. ESE is the sample mean of the estimated standard errors using the asymptotic standard error formula derived theoretically. We can test the accuracy of the standard error formula by comparing SEE and ESE. CP is the 95% empirical coverage probability which indicates how accuracy of the confidence interval. Median is the median value of 1000 estimates minus the true value of β, and MAD is the median absolute deviation of the 1000 estimates divided by a factor of 0.6745, which is a robust estimator of standard error to exclude the effects of extreme values.

We have four blocks in the table, the first block is assuming a working independence correlation structure(WI), i.e., letting weight matrix V_{2i}^{-1} in estimating equation (2.7) be an identity matrix. This is our baseline case. We compare all other cases to WI case. The second block is assuming an exchangeable correlation structure, which is a misspecification scenario for Observation I since the true correlation is an $ARMA(1,1)$

process. The third block is assuming an $ARMA(1,1)$ correlation structure which is the true scenario for Observation I. The fourth block is assuming a mixed correlation structure as we discussed in section 2.3.4. In each scenario we compare different approaches. The kernel function is taken as the Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2)_+$ and bandwidth is chosen by $k - fold$ cross validation where $k = 10$.

From Table 1, firstly, we can see the estimates from all methods are consistent, the theoretical standard error formula is correct since SEE and ESE are close for each case. The coverage probabilities are close to the 95% nominal level for most cases except for QL method when estimating β_2 by assuming true or mixed correlation structure. MAD is close to SEE which means there are rarely any effects from outliers. Secondly estimation of QL, MGV and WLS have a smaller standard error than that of WI, which means the methods consider correlation can improve the efficiency of estimation comparing with working independence case that ignores the correlation within subjects. Thirdly, the efficiency improved is more evident for β_1 than β_2. Fourthly, assuming a true or mixed correlation structure will be more efficient than the scenario where correlation structure is misspecified.

Table 2 is the simulation result for the same model but with a moderate correlation ($(\gamma, \rho) = (0.85, 0.6)$), and Table 3 is for a weak correlation ($(\gamma, \rho) = (0.85, 0.3)$), from which we can see the stronger is the correlation, the more efficiency will be obtained. MGV approach has large variance when correlation is misspecified for estimation of β_2.

Results for Observation II are shown in Table 4, Table 5 and Table 6 for strong, moderate and weak correlation respectively, from which we can draw similar results

as those from Observation I. Again MGV method is not robust, as there are outliers in either misspecified correlation scenario or mixed correlation scenario.

Lin and Carroll (2001b) showed that conventional profile-kernel method does not yield an efficient estimator of β when the parametric covariate Z is dependent of the nonparametric covariate T. Thus we only consider Z that are independent of T. For these time-independent covariates Z, we compare the different methods' performance between time-invariant covariates and time-variant covariates. Time-invariant means that the covariate doesn't change with time for each subject, such as gender and treatment effects. While time-variant means that the covariate changes with time for each subject such as biomarker CD4 level for each patient during a clinical trial. In Observation I settings, Z_{2ij} is a time-variant covariate while Z_{2i} is a time-invariant covariate and is a Bernoulli random variable with success probability of 0.5 for each i. Here we didn't see much improvement in efficiency for time-invariant covariate Z_{2i}, which also appears in the simulation studies in Lin and Carroll (2001b) and Wang et al. (2005). They both showed that no gain in efficiency is realized in estimation of coefficients of time-invariant covariates which are also independent of T_{ij} and the design is balanced with respect to the corresponding covariates, but for time-variant covariates, the methods considering correlation reduce the variance by more than 50%.

We draw several conclusions based on these simulation results. For time-variant covariate Z_1, WLS, QL, MGV and QIF all perform well and improve the efficiency comparing with working independence(WI) method, and there is good agreement between the estimated and empirical standard errors. The difference between SEE and

ESE is less than half of a standard deviation of the 1000 estimated standard errors. This implies that standard error formula is accurate. Comparing Table (2), Table (3) with Table (1), the stronger the correlation is, the efficiency obtained will be greater. However, for time-invariant covariate Z_2, the efficiency doesn't improve much. The estimates are unbiased and there is good agreement between the estimated and empirical standard errors for time-variant covariate Z_1 while the estimated standard errors by sandwich formula still tend to underestimate the covariance for time-invariant covariate Z_2. And compare Observation I with Observation II, whether observation time depend on covariates or not doesn't affect the performance of all the methods.

When assuming exchangeable correlation structure (misspecified correlation structure), the estimation error are slighted higher than assuming $ARMA(1,1)$ correlation structure (true correlation structure) and assuming mixed correlation structure(a family of possible correlation structures), but still less than working independence case. The performance of assuming $ARMA(1,1)$ correlation structure (true correlation structure) and assuming mixed correlation structure(a family of possible correlation structures) are close. Thus assuming a mixed correlation structure when we don't know the true correlation structure performs as well as when we know the correlation structure in advance.

2.4.2 More Simulation Study of Robustness

To see whether the different methods still perform well under misspecified correlation structures, we consider the following scenarios of correlation structures which are not exchangeable, ARMA(1,1) or mixture of both.

(B1) Random intercept model. We add an random intercept part $b_i \sim N(0, 0.2^2)$ to the original error in Observation I model in section 2.4.1 where it follows a multivariate standard normal with an $ARMA(1,1)$ correlation structure and correlation coefficients $(0.85, 0.6)$. Now the new error becomes $\epsilon(t) = b_i + ARMA(1,1)$. Since b_i is a subject level covariate shared by same subject i, this transformation will make the correlation stronger within subjects.

(B2) Random slope model. We add both a random intercept and random slope to the original $ARMA(1,1)$ error, now $b_i = b_{i0} + b_{i1}W_{ij}(t)$ where b_{i0} and b_{i1} are subject level covariate which doesn't change within subject i, but W_{ij} is a time-varying covariate. In the new model we let the error part be $\epsilon_i(t) = [b_{i0} + b_{i1}\cos(2\pi t/14) + b_{i2}\sin(2\pi t/14) + e_i(t)]/3$ where b_{i0}, b_{i1}, b_{i2} are independent normal with zero mean and standard deviation $2/5, 1/5, 1/5$ respectively and $e_i(t) \sim N(0,1)$ with $ARMA(1,1)$ correlation structure and correlation coefficients $(0.85, 0.6)$.

(B3) Unimodal mixnormal model. Now instead adding a perturbation term to the original error, we substitute the error part by a unimodal mixnormal error with the same ARMA(1,1) correlation structure and correlation coefficients $(0.85, 0.9)$. Let $\epsilon(t) \sim 0.25N(-0.75, 0.364^2) + 0.75N(0.25, 0.364^2)$. The histogram of the error shows a skewed and unimodal shape.

(B4) Bimodal mixnormal model. We change the above unimodal mixnormal to bimodal mixnormal for $\epsilon(t)$ with the same $ARMA(1,1)$ correlation structure and correlation coefficients $(0.85, 0.9)$. Let $\epsilon(t) \sim 0.25N(-1.05, 0.1803^2) + 0.75N(0.35, 0.1803^2)$. The histogram of the error shows a skewed and bimodal shape.

The simulation for QL, MGV, WLS and QIF methods under (B1), (B2), (B3) and

(B4) by assuming EX, ARMA(1,1) and EX+ARMA(1,1) are conducted. The results are illustrated in Table 7 for random intercept model and Table 8 for random slope model respectively. We can see the similar result as before, and assuming ARMA(1,1) structure still performs well even adding a random effect to the model. The estimation results of Unimodal mixnormal model are shown in Table 9. The estimation results of Bimodal mixnormal model are shown in Table 10. Under skewed non-normal distribution, we still get similar results.

2.5 Simulation Study for discrete response

In this section we examine the performance of the various methods for model (2.1) with logarithm and logistic link functions when the longitudinal response is discrete.

(D1) Bernoulli I model. We generate independent observation time with random missing similar as Observation I model. Every individual has the same scheduled time points, {0,1,2,...,8}, each of which has a 20% probability of being skipped except time 0. Then a uniform $[0, 1]$ random variable is added to the nonskipped scheduled time points. The error part $\epsilon(t_{ij})$ is a Gaussian process with mean 0, constant variance 1 and the correlation structure is exchangeable structure with parameter θ, i.e. $corr(\epsilon(s), \epsilon(t)) = \theta$ for $s \neq t$. We let $\theta = 0.5$. For the covariates, $X_{ij} \equiv 1$, $Z_{ij} = (Z_{1ij}, Z_{2ij})^T$ for a given t_{ij}, Z_{1ij} and Z_{2ij} are independent standard normal random variables. For coefficients, $\beta = (0.01, 0.01)^T$, $\alpha(t) = \sin(\pi t/30) - 0.5$. We use the methods in Macke et al. (2009) to generate bernoulli random variable Y_{ij} with the given mean μ_{ij} such that $logit(\mu_{ij}) = \alpha(T_{ij})X_{ij} + \beta^T Z_{ij}$ and given correlation structure.

(D2) Poisson model. We generate the Poisson random variable Y_{ij} with mean μ such that $log(\mu) = \alpha(T_{ij})X_{ij} + \beta^T Z_{ij}$ and exchangeable correlation structure with $\theta = 0.5$. The other covariates and coefficients are the same as the above model.

(D3) Bernoulli II model. We use the same model setting in Study 1 in simulation part of He et al. (2005). We generate a bernoulli response with mean μ such that $logit(\mu) = \alpha(T_{ij})X_{ij} + \beta Z_{ij}$, where $X_{ij} \equiv 1$, $Z_{ij} \sim Uniform[-1, 1]$, $\beta = 0.4$, $\alpha(t) = \cos(\pi t/2)$ and $T_{ij} \sim Uniform[-1, 1]$ and is independent from Z_{ij}. This is a balanced design with cluster size $n_i = 3$, let $Y_{i1} = b_{i1}w_{i0}+(1-b_{i1})w_{i1}$, $Y_{i2} = b_{i2}w_{i0}+(1-b_{i2})w_{i2}$, and $Y_{i3} = b_{i3}w_{i0} + (1 - b_{i3})w_{i3}$, where (b_{i1}, b_{i2}, b_{i3}) are independent Bernoulli variables with mean 0.5 and w_{i0}, w_{i1}, w_{i2} and w_{i3} are independent Bernoulli variables with means $\mu_{i1}, \mu_{i1}, 2\mu_{i2} - \mu_{i1}$ and $2\mu_{i3} - \mu_{i1}$. Thus Y_{ij} has an exchangeable covariance structure for each i, but the covariances vary from cluster to cluster.

The estimation results of these three models are summarized in Table 11 for Bernoulli I model, Table 12 for Poisson model and Table 13 for Bernoulli II model. Under each model setting we generate 1000 datasets, each consisting of $n = 200$ subjects. We obtain similar results as in continuous responses and is summarized in Section 2.7.

2.6 Real Data Application

In this section, we apply the above methods to a real data example. We demonstrate how various methods that consider correlation structure improves estimation efficiency than working independence case. We consider the analysis of a HIV-1 RNA data set from an AIDS clinical trial. In this study, all subjects received a single

protease inhibitor(PI) while others received a double-PI antiretroviral regimens in treating HIV-infected patients. HIV-1 RNA levels in plasma was measured repeatedly during the follow-up. The scheduled visit times were at weeks 0, 2, 4, 8, 16 and 24. But the actual visit times of individuals may vary around the scheduled visiting times. Some patients had prior antiviral treatment with non-nucleoside analogue reverse transcriptase inhibitors(NNRTI) and others did not have prior NNRTI treatment. The prior NNRTI treatment is considered to be a factor that affects the antiviral responses to the antiretroviral regimens in the current study.

A total of 481 patients were enrolled in the study, with 2626 total visits. Owing to technical limitations, 175 measurements of HIV-1 RNA levels were censored below the detection limit, and three were censored above the detection limit. We restrict our analysis to those responses within detection range. This data set has been analyzed by Sun et al. (2013) and Sun and Wu (2005). Here we use the same transformed time scale $t = log_{10}(\text{day of actual visit}+40) - log_{10}(32)$ of actual visits so that the transformed sampling time points are more evenly distributed suitable for bandwidth selection. The maximum of the transformed sampling times is $\tau = 0.88$. The response variable $Y_i(t)$ is the change of HIV-1 RNA level using a log_{10} scale at time $t \in [0, \tau]$ from the baseline. Let $X = 1$ denote the patients who received a double-PI treatment and $X = 0$ for patients who received a single-PI treatment. Let Z be the indicator of the prior antiviral treatment with NNRTI, with 1 for having had NNRTI and 0 for having not received NNRTI.

Analysis of Sun and Wu (2005) shows that the effect of treatment(double-PI versus single-PI) is time-varying after adjusting for the prior NNRTI antiviral treatment

experience under the semiparametric additive regression model, and Sun et al. (2013) show that the identity link will give a smaller prediction error than logarithm link function. Thus we consider to fit the following model

$$Y_i(t) = \alpha_0(t) + \alpha_1(t)X_i + \beta Z_i + \epsilon_i(t), \tag{2.28}$$

where $0 \leq t \leq \tau$, $\alpha_0(t)$ is the baseline function, $\alpha_1(t)$ is the time-varying treatment effect, β is a fixed parameter denoting the effect of prior antiviral treatment with NNRTI, and $\epsilon_i(t)$ is a mean zero process.

We used the K-fold cross-validation method with $K = 50$ to get the bandwidth $h_{cv} = 0.06$. The estimates of β by different methods is list in Table (14). The estimates of $\alpha_0(t)$ and $\alpha_1(t)$ and the confidence intervals by different methods are plotted in Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5, Fig. 6. The double-PI antiretroviral regimens works better than the single PI regimens in reducing viral load in treating HIV-infected patients and this effect becomes stronger over time during the course of the study. The patients who had prior antiviral treatment with NNRTI tend to have higher level of viral load than those who did not have the prior treatment.

2.7 Concluding remarks

Based on above analysis, we get the following conclusions:

1. The estimates by all methods are consistent, the coverage probabilities are close to the 95% nominal level for most cases.

2. WLS, QL, MGV and QIF all perform well and improve the efficiency comparing with working independence(WI) method, and there is good agreement between

the estimated and empirical standard errors for time-variant covariate while the estimated standard errors by sandwich formula still tend to underestimate the covariance for time-invariant covariate.

3. The efficiency improved is more evident for time-variant covariate then time-invariant covariate. We didn't see much improvement in efficiency for time-invariant covariate, but for time-variant covariates, the methods considering correlation reduce the variance and in strong correlation case by QL method the reduction can be more than 50%.

4. Assuming a true or mixed correlation structure will be more efficient than the scenario where correlation structure is misspecified, i.e., when assuming a misspecified correlation structure, the estimation error are slighted higher than assuming the true correlation structure or assuming a mixed correlation structure, but still less than working independence case. Assuming a mixed correlation structure when we don't know the true correlation structure performs as well as when we know the correlation structure in advance.

5. Estimation of correlation parameter from QL method is closer to the true value than that of MGV or WLS method. MGV and WLS method tend to underestimate the correlation parameters.

6. The stronger is the correlation, the more efficiency will be obtained.

7. MGV method is not robust, as there are outliers in either misspecified correlation scenario or mixed correlation scenario.

8. Whether observation time depend on covariates or not doesn't affect the performance of all the methods. Under skewed non-normal distribution such as perturbation of normal distribution or discrete responses that follow a Poisson or Bernoulli distribution, all above results stand.

Table 1: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.8$ based on 1000 simulations under model (A1) with strong correlation coefficients $(0.85, 0.9)$.

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0013	.0244	.0235	.938	.0013	.0242	.0059	.1016	.1011	.944	.0040	.0998
Assuming ARMA correlation structure(True)												
QL	.0005	.0130	.0127	.936	.0009	.0132	.0030	.0955	.0848	.914	.0024	.0961
MGV	.0008	.0147	.0140	.929	.0010	.0153	.0052	.0936	.0918	.943	.0043	.0917
QIF	.0005	.0143	.0136	.931	.0001	.0142	.0051	.0926	.0877	.939	.0053	.0905
WLS	.0006	.0142	.0136	.935	.0005	.0141	.0050	.0926	.0897	.942	.0041	.0906
Assuming Mixed correlation structure												
QL	.0005	.0131	.0127	.937	.0009	.0132	.0030	.0956	.0847	.916	.0020	.0968
MGV	.0005	.0145	.0138	.930	.0005	.0144	.0047	.0934	.0895	.942	.0026	.0885
QIF	.0004	.0142	.0133	.928	.0001	.0141	.0053	.0925	.0864	.932	.0059	.0938
WLS	.0006	.0142	.0136	.934	.0005	.0140	.0051	.0927	.0897	.942	.0041	.0900
Assuming Exchangeable correlation structure(Misspecification)												
QL	.0001	.0166	.0160	.937	-.0002	.0165	.0041	.1096	.1020	.938	.0012	.1131
MGV	.0001	.0166	.0160	.936	-.0003	.0166	.0041	.1061	.0992	.940	.0009	.1084
QIF	.0003	.0168	.0160	.930	-.0001	.0167	.0053	.0954	.0915	.940	.0033	.0930
WLS	.0004	.0171	.0164	.934	.0002	.0171	.0052	.0946	.0912	.941	.0022	.0902

Table 2: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.8$ based on 1000 simulations under model (A1) with moderate correlation coefficients $(0.85, 0.6)$.

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0009	.0242	.0236	.942	.0007	.0233	.0040	.0672	.0666	.946	.0024	.0653
Assuming ARMA correlation structure(True)												
QL	.0008	.0182	.0174	.935	.0014	.0181	.0028	.0618	.0601	.938	.0030	.0612
MGV	.0007	.0187	.0179	.933	.0006	.0183	.0031	.0618	.0604	.942	.0030	.0608
QIF	.0007	.0195	.0182	.928	.0006	.0191	.0036	.0629	.0606	.941	.0036	.0621
WLS	.0008	.0194	.0184	.933	.0002	.0192	.0033	.0622	.0609	.941	.0019	.0598
Assuming Mixed correlation structure												
QL	.0008	.0182	.0174	.935	.0014	.0182	.0028	.0619	.0601	.937	.0029	.0613
MGV	.0005	.0192	.0183	.942	.0000	.0187	.0027	.0659	.0630	.945	.0026	.0660
QIF	.0007	.0196	.0181	.924	.0002	.0192	.0036	.0633	.0600	.944	.0029	.0614
WLS	.0008	.0193	.0184	.933	.0004	.0193	.0033	.0622	.0609	.941	.0021	.0594
Assuming Exchangeable correlation structure(Misspecification)												
QL	.0003	.0220	.0210	.937	-.0003	.0225	.0033	.0665	.0634	.941	.0019	.0644
MGV	.0001	.0221	.0212	.939	-.0008	.0222	.0024	.0937	.0895	.949	.0003	.0958
QIF	.0004	.0221	.0209	.931	-.0006	.0225	.0035	.0647	.0618	.943	.0013	.0634
WLS	.0005	.0222	.0212	.937	-.0002	.0224	.0035	.0645	.0623	.943	.0009	.0630

Table 3: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.8$ based on 1000 simulations under model (A1) with weak correlation coefficients $(0.85, 0.3)$.

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0008	.0241	.0236	.953	.0003	.0236	.0031	.0518	.0513	.947	.0017	.0514
Assuming ARMA correlation structure(True)												
QL	.0010	.0214	.0204	.935	.0007	.0216	.0024	.0484	.0476	.941	.0025	.0468
MGV	.0008	.0216	.0208	.932	.0011	.0216	.0025	.0490	.0480	.946	.0025	.0484
QIF	.0009	.0226	.0210	.928	.0006	.0225	.0030	.0497	.0479	.940	.0011	.0488
WLS	.0009	.0220	.0209	.933	.0003	.0219	.0026	.0487	.0478	.944	.0014	.0471
Assuming Mixed correlation structure												
QL	.0010	.0214	.0204	.936	.0007	.0215	.0024	.0484	.0476	.941	.0020	.0464
MGV	.0007	.0218	.0208	.937	.0005	.0207	.0021	.0530	.0513	.946	.0033	.0534
QIF	.0008	.0227	.0209	.926	.0006	.0228	.0030	.0500	.0475	.939	.0019	.0490
WLS	.0009	.0220	.0209	.934	.0005	.0219	.0026	.0488	.0478	.943	.0014	.0472
Assuming Exchangeable correlation structure(Misspecification)												
QL	.0006	.0231	.0221	.933	-.0005	.0231	.0027	.0502	.0484	.941	.0005	.0490
MGV	.0003	.0235	.0226	.939	-.0004	.0240	.0019	.0790	.0769	.953	.0004	.0811
QIF	.0006	.0233	.0220	.929	-.0007	.0235	.0027	.0501	.0480	.942	.0016	.0490
WLS	.0006	.0231	.0221	.935	-.0004	.0231	.0027	.0500	.0484	.940	.0004	.0491

Table 4: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 1.2$ based on 1000 simulations under model (A2) with strong correlation $\rho = 0.8$.

	$\beta_1 = 0.5$						$\beta_2 = 1$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0023	.1535	.1544	.949	.0021	.1583	.0048	.1691	.1629	.934	.0041	.1656
Assuming Exchangeable correlation structure(True)												
QL	.0069	.0853	.0855	.956	.0040	.0861	-.0145	.1625	.1423	.910	-.0208	.1624
MGV	.0044	.1113	.1119	.950	.0049	.1115	.0022	.1588	.1493	.928	.0012	.1538
QIF	.0066	.1069	.1037	.950	.0042	.1054	.0021	.1584	.1473	.922	.0004	.1540
WLS	.0053	.0942	.0945	.956	.0053	.0952	-.0009	.1563	.1449	.926	-.0041	.1495
Assuming Mixed correlation structure												
QL	.0068	.0855	.0853	.953	.0055	.0851	-.0134	.1622	.1422	.911	-.0195	.1617
MGV	.0031	.1246	.1249	.949	.0027	.1238	.0033	.1616	.1532	.929	.0001	.1577
QIF	.0060	.1044	.0976	.930	.0056	.1044	-.0003	.1573	.1421	.914	-.0021	.1456
WLS	.0052	.0942	.0944	.954	.0055	.0945	-.0008	.1563	.1448	.927	-.0034	.1484
Assuming ARMA correlation structure(Misspecification)												
QL	.0064	.0884	.0878	.949	.0061	.0889	-.0084	.1605	.1427	.915	-.0139	.1573
MGV	.0047	.0974	.0973	.948	.0029	.0972	.0008	.1574	.1466	.933	-.0037	.1532
QIF	.0048	.1064	.1019	.935	.0040	.1006	-.0001	.1576	.1434	.917	-.0028	.1448
WLS	.0051	.0945	.0948	.953	.0047	.0929	-.0004	.1564	.1451	.929	-.0026	.1499

Table 5: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 1.2$ based on 1000 simulations under model (A2) with moderate correlation $\rho = 0.5$.

	$\beta_1 = 0.5$						$\beta_2 = 1$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0060	.1952	.1960	.949	.0020	.1963	.0062	.1863	.1787	.940	.0059	.1876
Assuming Exchangeable correlation structure(True)												
QL	.0118	.1602	.1591	.954	.0083	.1608	-.0001	.1773	.1646	.935	-.0025	.1766
MGV	.0083	.1723	.1731	.952	.0068	.1705	.0048	.1795	.1695	.933	.0068	.1795
QIF	.0114	.1676	.1637	.951	.0054	.1643	.0045	.1787	.1659	.928	.0028	.1770
WLS	.0092	.1659	.1654	.956	.0059	.1645	.0037	.1777	.1666	.927	.0028	.1772
Assuming Mixed correlation structure												
QL	.0116	.1602	.1588	.951	.0082	.1607	.0001	.1773	.1646	.933	-.0023	.1769
MGV	.0042	.1837	.1777	.949	.0070	.1767	.0055	.1806	.1710	.936	.0080	.1816
QIF	.0106	.1692	.1611	.946	.0065	.1707	.0039	.1793	.1632	.921	.0018	.1759
WLS	.0092	.1658	.1652	.954	.0063	.1638	.0037	.1777	.1665	.927	.0036	.1777
Assuming ARMA correlation structure(Misspecification)												
QL	.0112	.1611	.1597	.953	.0075	.1591	.0006	.1774	.1647	.932	.0001	.1753
MGV	.0113	.1723	.1667	.946	.0095	.1715	.0020	.1785	.1665	.929	.0037	.1762
QIF	.0091	.1740	.1684	.939	.0043	.1722	.0033	.1815	.1658	.917	.0028	.1732
WLS	.0090	.1661	.1655	.952	.0059	.1647	.0039	.1778	.1667	.927	.0041	.1773

Table 6: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 1.2$ based on 1000 simulations under model (A2) with weak correlation $\rho = 0.2$.

	$\beta_1 = 0.5$						$\beta_2 = 1$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0133	.3109	.3105	.946	.0073	.3124	.0090	.2416	.2312	.947	.0156	.2523
Assuming Exchangeable correlation structure(True)												
QL	.0185	.3005	.2973	.952	.0122	.3088	.0077	.2381	.2246	.939	.0155	.2452
MGV	.0163	.3009	.2992	.953	.0098	.3058	.0081	.2380	.2254	.941	.0167	.2501
QIF	.0183	.3033	.2955	.948	.0105	.3010	.0082	.2406	.2230	.932	.0185	.2517
WLS	.0159	.3023	.2999	.948	.0084	.3062	.0083	.2389	.2256	.939	.0154	.2521
Assuming Mixed correlation structure												
QL	.0183	.3003	.2969	.954	.0120	.3053	.0077	.2382	.2245	.938	.0172	.2469
MGV	.0158	.4308	.3414	.931	.0071	.3248	.0087	.2537	.2380	.938	.0130	.2537
QIF	.0165	.3054	.2925	.946	.0076	.3090	.0085	.2422	.2212	.929	.0125	.2505
WLS	.0157	.3023	.2996	.948	.0077	.3047	.0083	.2389	.2254	.938	.0154	.2526
Assuming ARMA correlation structure(Misspecification)												
QL	.0181	.3008	.2974	.957	.0091	.3065	.0078	.2381	.2246	.938	.0157	.2480
MGV	.0408	.5037	.3951	.934	.0186	.3768	.0125	.2641	.2476	.942	.0111	.2549
QIF	.0153	.3090	.2998	.942	.0093	.3149	.0078	.2436	.2248	.928	.0149	.2470
WLS	.0156	.3026	.2999	.949	.0074	.3054	.0083	.2389	.2256	.938	.0158	.2528

Table 7: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.7$ based on 1000 simulations under Random intercept model (B1).

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0005	.0250	.0242	.933	.0000	.0162	.0008	.0749	.0723	.946	-.0004	.0507
Assuming ARMA correlation structure												
QL	-.0002	.0177	.0175	.939	-.0005	.0117	.0020	.0712	.0656	.927	.0004	.0466
MGV	-.0001	.0184	.0180	.938	.0002	.0121	.0016	.0707	.0660	.936	-.0014	.0465
QIF	-.0002	.0189	.0183	.931	.0003	.0121	.0018	.0717	.0662	.927	.0008	.0477
WLS	.0000	.0191	.0186	.944	.0001	.0126	.0012	.0708	.0665	.939	-.0017	.0467
Assuming Mixed correlation structure												
QL	-.0002	.0177	.0175	.940	-.0006	.0118	.0019	.0713	.0655	.925	-.0004	.0471
MGV	-.0001	.0188	.0183	.940	-.0004	.0127	.0012	.0729	.0679	.929	.0008	.0495
QIF	-.0002	.0190	.0182	.932	.0003	.0121	.0017	.0717	.0654	.923	.0004	.0469
WLS	.0000	.0191	.0186	.943	.0001	.0126	.0012	.0708	.0665	.940	-.0019	.0470
Assuming Exchangeable correlation structure												
QL	.0000	.0220	.0211	.943	.0004	.0143	.0003	.0737	.0693	.933	-.0014	.0499
MGV	-.0001	.0221	.0213	.943	.0001	.0143	.0000	.0944	.0914	.938	.0009	.0627
QIF	.0001	.0222	.0210	.938	-.0001	.0148	.0007	.0717	.0673	.935	-.0005	.0488
WLS	.0001	.0223	.0214	.948	.0001	.0149	.0005	.0717	.0679	.941	-.0007	.0482

Table 8: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.7$ based on 1000 simulations under Random slope model (B2).

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	-.0002	.0085	.0084	.947	-.0006	.0057	-.0002	.0254	.0259	.955	.0005	.0175
Assuming ARMA correlation structure												
QL	.0001	.0060	.0059	.942	-.0002	.0041	-.0007	.0244	.0238	.943	-.0002	.0166
MGV	.0000	.0063	.0061	.937	-.0001	.0044	-.0008	.0242	.0239	.948	-.0003	.0169
QIF	.0001	.0064	.0062	.941	-.0002	.0043	-.0005	.0245	.0240	.945	-.0001	.0169
WLS	.0000	.0064	.0063	.935	-.0001	.0045	-.0008	.0242	.0241	.948	-.0002	.0171
Assuming Mixed correlation structure												
QL	.0001	.0061	.0059	.941	-.0002	.0041	-.0007	.0245	.0237	.943	-.0003	.0167
MGV	-.0001	.0067	.0065	.933	-.0002	.0045	-.0009	.0245	.0242	.947	-.0006	.0170
QIF	.0001	.0064	.0062	.932	-.0002	.0043	-.0007	.0246	.0238	.944	-.0004	.0172
WLS	.0000	.0065	.0063	.937	-.0001	.0045	-.0008	.0242	.0241	.947	-.0006	.0169
Assuming Exchangeable correlation structure												
QL	-.0001	.0075	.0073	.937	-.0003	.0051	-.0013	.0260	.0251	.945	-.0023	.0173
MGV	-.0001	.0075	.0074	.940	-.0002	.0050	-.0010	.0250	.0246	.944	-.0010	.0168
QIF	-.0001	.0075	.0073	.940	-.0002	.0051	-.0010	.0250	.0244	.947	-.0009	.0169
WLS	-.0001	.0075	.0074	.939	-.0002	.0051	-.0010	.0249	.0246	.945	-.0008	.0169

Table 9: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.7$ based on 1000 simulations under Unimodal mixnormal model (B3).

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0010	.0251	.0241	.937	.0006	.0167	.0016	.1097	.1049	.942	-.0005	.0741
Assuming ARMA correlation structure												
QL	-.0001	.0128	.0127	.941	-.0001	.0085	.0045	.1064	.0875	.896	.0041	.0710
MGV	.0001	.0146	.0142	.939	.0000	.0096	.0021	.1027	.0956	.943	-.0009	.0689
QIF	-.0001	.0139	.0137	.938	.0000	.0091	.0045	.1029	.0911	.912	.0044	.0654
WLS	.0001	.0141	.0137	.941	.0000	.0094	.0023	.1014	.0932	.939	-.0003	.0677
Assuming Mixed correlation structure												
QL	-.0001	.0128	.0127	.941	-.0001	.0086	.0043	.1066	.0875	.897	.0021	.0716
MGV	.0000	.0143	.0139	.943	.0000	.0096	.0019	.1014	.0929	.930	-.0013	.0687
QIF	-.0001	.0138	.0133	.934	.0001	.0092	.0037	.1021	.0897	.916	.0057	.0654
WLS	.0001	.0140	.0137	.942	-.0001	.0094	.0022	.1015	.0932	.939	-.0011	.0680
Assuming Exchangeable correlation structure												
QL	.0000	.0167	.0160	.937	.0003	.0111	.0007	.1139	.1053	.930	.0023	.0754
MGV	.0000	.0167	.0161	.936	.0003	.0111	.0010	.1096	.1010	.931	.0011	.0743
QIF	.0000	.0168	.0161	.945	.0001	.0109	.0014	.1027	.0952	.934	.0007	.0693
WLS	.0001	.0172	.0165	.936	.0001	.0116	.0013	.1019	.0946	.935	-.0029	.0694

Table 10: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.7$ based on 1000 simulations under Bimodal mixnormal model (B4).

	$\beta_1 = 1$						$\beta_2 = 2$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0002	.0154	.0157	.956	.0000	.0161	-.0019	.0675	.0681	.940	.0011	.0663
Assuming ARMA correlation structure												
QL	-.0003	.0093	.0089	.941	-.0008	.0091	-.0006	.0721	.0611	.903	-.0006	.0747
MGV	-.0002	.0100	.0099	.941	-.0006	.0097	-.0017	.0668	.0652	.932	-.0002	.0666
QIF	-.0003	.0102	.0096	.933	-.0009	.0101	-.0013	.0679	.0631	.925	-.0002	.0704
WLS	-.0002	.0098	.0096	.942	-.0008	.0094	-.0016	.0668	.0643	.931	-.0024	.0662
Assuming Mixed correlation structure												
QL	-.0003	.0093	.0089	.941	-.0008	.0091	-.0005	.0723	.0610	.903	-.0012	.0742
MGV	-.0001	.0107	.0107	.949	-.0004	.0103	-.0017	.0671	.0658	.934	-.0008	.0650
QIF	-.0003	.0100	.0094	.936	-.0009	.0103	-.0012	.0689	.0624	.918	.0001	.0686
WLS	-.0002	.0098	.0096	.943	-.0008	.0094	-.0016	.0669	.0643	.929	-.0022	.0663
Assuming Exchangeable correlation structure												
QL	-.0002	.0115	.0112	.943	-.0003	.0112	-.0009	.0712	.0640	.919	.0017	.0704
MGV	.0000	.0122	.0122	.950	-.0002	.0121	-.0017	.0677	.0665	.935	.0004	.0644
QIF	-.0002	.0115	.0112	.942	-.0004	.0115	-.0015	.0689	.0658	.927	.0000	.0661
WLS	-.0001	.0116	.0115	.945	-.0004	.0115	-.0015	.0679	.0658	.932	-.0006	.0648

Table 11: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 1.2$ based on 1000 simulations under Bernoulli I model (D1).

	$\beta_1 = 0.01$						$\beta_2 = 0.01$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	.0002	.0524	.0525	.954	.0007	.0548	.0014	.0516	.0525	.962	.0007	.0512
Assuming Exchangeable correlation structure(True)												
QL	.0000	.0396	.0395	.947	.0001	.0393	-.0013	.0396	.0395	.949	-.0013	.0383
MGV	.0000	.0396	.0395	.947	-.0003	.0391	-.0013	.0396	.0395	.948	-.0016	.0383
QIF	.0003	.0404	.0397	.944	-.0003	.0401	-.0012	.0404	.0397	.943	-.0011	.0405
WLS	.0001	.0408	.0407	.950	.0011	.0406	-.0005	.0406	.0407	.956	.0008	.0407
Assuming Mixed correlation structure												
QL	.0000	.0396	.0394	.947	.0004	.0392	-.0012	.0396	.0394	.949	-.0013	.0382
MGV	.0003	.0434	.0432	.946	.0004	.0435	.0000	.0428	.0432	.951	.0003	.0414
QIF	.0001	.0406	.0393	.941	-.0008	.0412	-.0014	.0406	.0393	.936	-.0017	.0407
WLS	.0001	.0408	.0406	.948	.0011	.0407	-.0005	.0406	.0406	.956	.0006	.0404
Assuming ARMA correlation structure(Misspecification)												
QL	-.0002	.0406	.0404	.945	-.0005	.0389	-.0012	.0407	.0405	.948	-.0002	.0407
MGV	-.0004	.0429	.0428	.937	-.0004	.0417	-.0005	.0429	.0429	.950	-.0001	.0424
QIF	-.0002	.0445	.0434	.932	-.0007	.0418	-.0008	.0440	.0435	.942	-.0010	.0442
WLS	.0000	.0411	.0410	.944	.0006	.0397	-.0005	.0410	.0410	.955	.0006	.0400

Table 12: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 1.2$ based on 1000 simulations under Poisson model (D2).

	$\beta_1 = 0.01$						$\beta_2 = 0.01$					
Method	Bias	SEE	ESE	CP	Median	MAD	Bias	SEE	ESE	CP	Median	MAD
Working Independence												
WI	-.0008	.0271	.0268	.944	.0001	.0271	-.0005	.0271	.0269	.943	.0004	.0265
Assuming Exchangeable correlation structure(True)												
QL	-.0006	.0199	.0197	.951	-.0002	.0206	-.0007	.0191	.0197	.955	-.0009	.0192
MGV	-.0007	.0219	.0216	.947	-.0003	.0227	-.0006	.0215	.0217	.950	-.0003	.0217
QIF	-.0005	.0201	.0197	.950	-.0004	.0214	-.0006	.0193	.0197	.953	-.0003	.0197
WLS	-.0006	.0206	.0203	.952	.0000	.0217	-.0007	.0199	.0203	.954	-.0007	.0202
Assuming Mixed correlation structure												
QL	-.0006	.0199	.0197	.948	.0000	.0206	-.0007	.0192	.0197	.956	-.0010	.0193
MGV	-.0006	.0222	.0218	.956	-.0003	.0225	-.0008	.0219	.0219	.945	-.0011	.0213
QIF	-.0006	.0202	.0195	.949	-.0006	.0215	-.0007	.0194	.0195	.954	-.0005	.0195
WLS	-.0006	.0206	.0202	.954	-.0001	.0218	-.0007	.0200	.0203	.955	-.0007	.0202
Assuming ARMA correlation structure(Misspecification)												
QL	-.0005	.0206	.0202	.954	.0002	.0209	-.0008	.0199	.0203	.955	-.0006	.0198
MGV	-.0008	.0243	.0241	.943	-.0009	.0238	-.0011	.0245	.0240	.949	-.0009	.0247
QIF	-.0005	.0220	.0216	.952	-.0001	.0226	-.0010	.0224	.0217	.940	-.0009	.0223
WLS	-.0006	.0208	.0205	.954	-.0006	.0213	-.0007	.0202	.0205	.954	-.0009	.0205

Table 13: Summary of Bias, SEE, ESE, CP, Median and MAD for β_1 and β_2 with $n = 200, h = 0.66$ based on 1000 simulations under Bernoulli II model (D3).

$\beta = 0.4$						
Method	Bias	SEE	ESE	CP	Median	MAD
Working Independence						
WI	-.0031	.1462	.1449	.951	.0047	.1421
Assuming Exchangeable correlation structure(Perturbation)						
WLS	-.0049	.1369	.1373	.950	.0000	.1341
QL	-.0038	.1398	.1395	.954	.0031	.1359
MGV	-.0036	.1376	.1372	.949	.0027	.1348
QIF	-.0041	.1393	.1389	.954	.0028	.1366
Assuming Mixed correlation structure						
WLS	-.0049	.1371	.1371	.949	.0001	.1342
QL	-.0044	.1399	.1410	.950	-.0031	.1388
MGV	-.0026	.1381	.1367	.953	.0027	.1350
QIF	-.0041	.1393	.1388	.953	.0032	.1370
Assuming ARMA correlation structure(Misspecification)						
WLS	-.0047	.1371	.1372	.950	.0001	.1344
QL	-.0097	.2418	.2258	.948	-.0079	.2275
MGV	-.0035	.1401	.1390	.952	-.0019	.1408
QIF	-.0040	.1393	.1389	.953	.0025	.1352

Table 14: Point estimates of β based on HIV-1 RNA data under model (2.28) by different approaches.

β (NNRTI treatment)			
Method	Estimate	SD	p-value
WI	0.6265	0.0888	< 0.0001
QL	0.0376	0.0671	0.575
MGV	0.1492	0.0662	0.024
QIF	0.2239	0.0657	0.001
WLS	0.3489	0.0670	< 0.0001

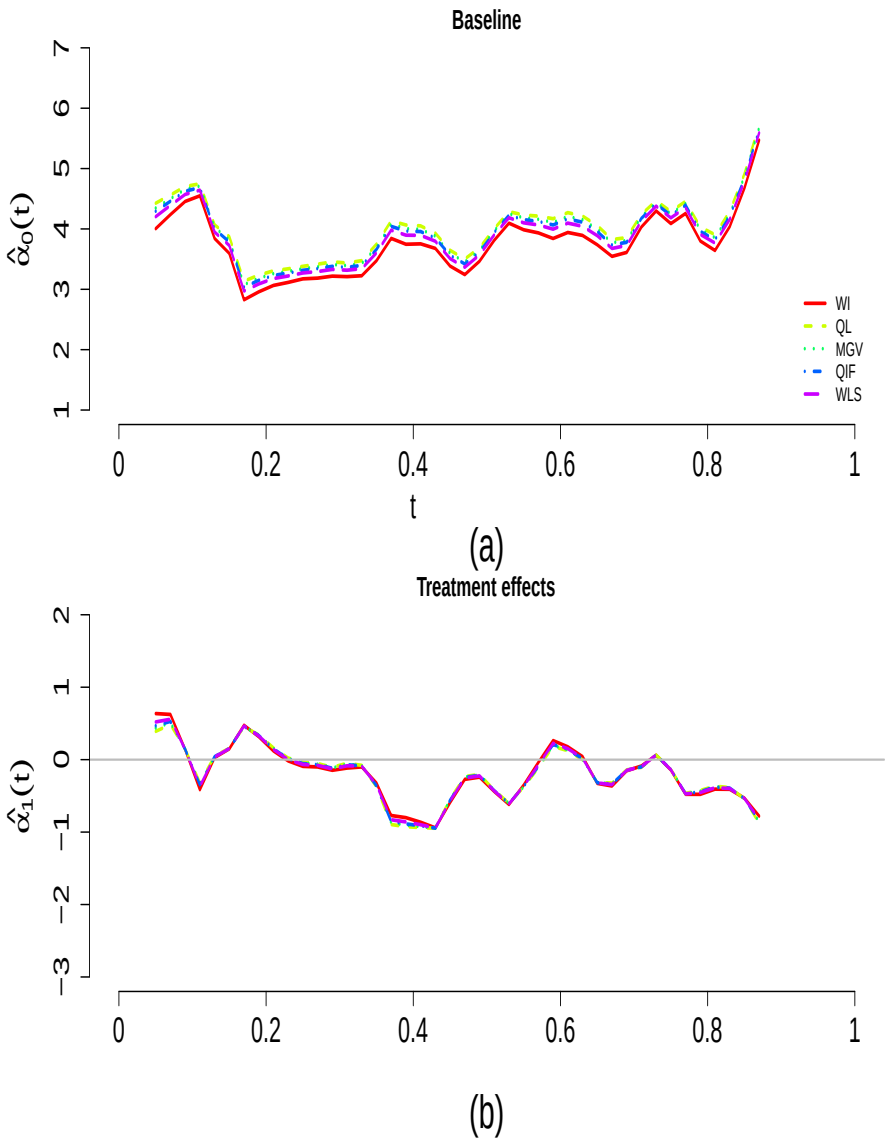

Figure 1: Estimates of baseline and varying-coefficient functions based on HIV-1 RNA data under model (2.28) by different approaches using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by different approaches; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by different approaches.

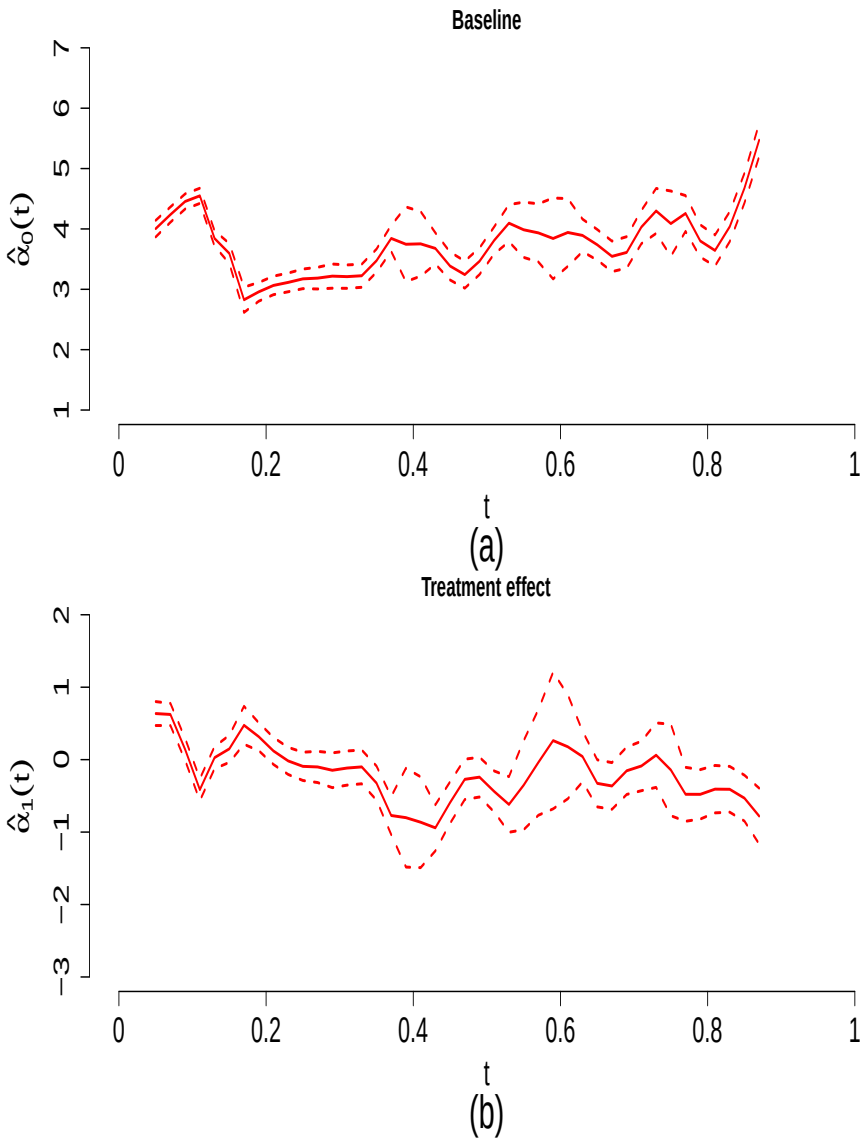

Figure 2: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on HIV-1 RNA data under model (2.28) by WI approach using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by WI approach; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by WI approach.

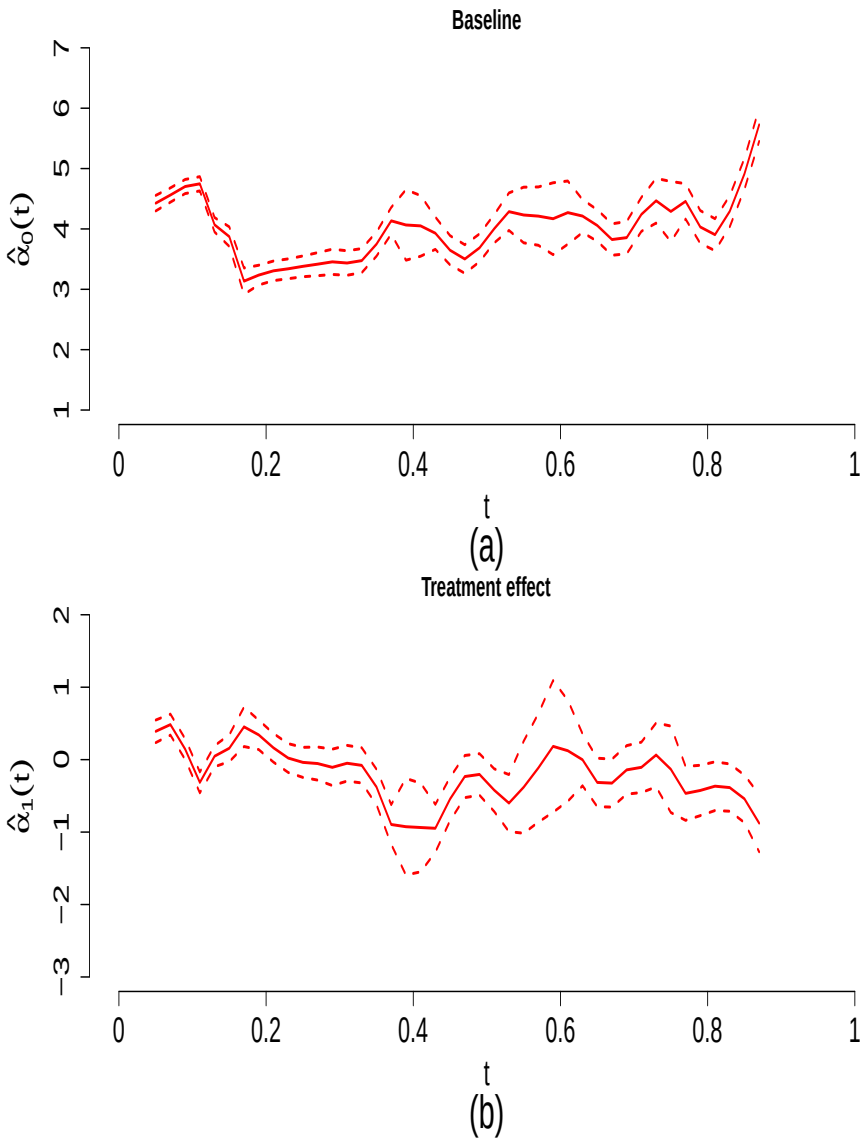

Figure 3: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on HIV-1 RNA data under model (2.28) by QL approach using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by QL approach; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by QL approach.

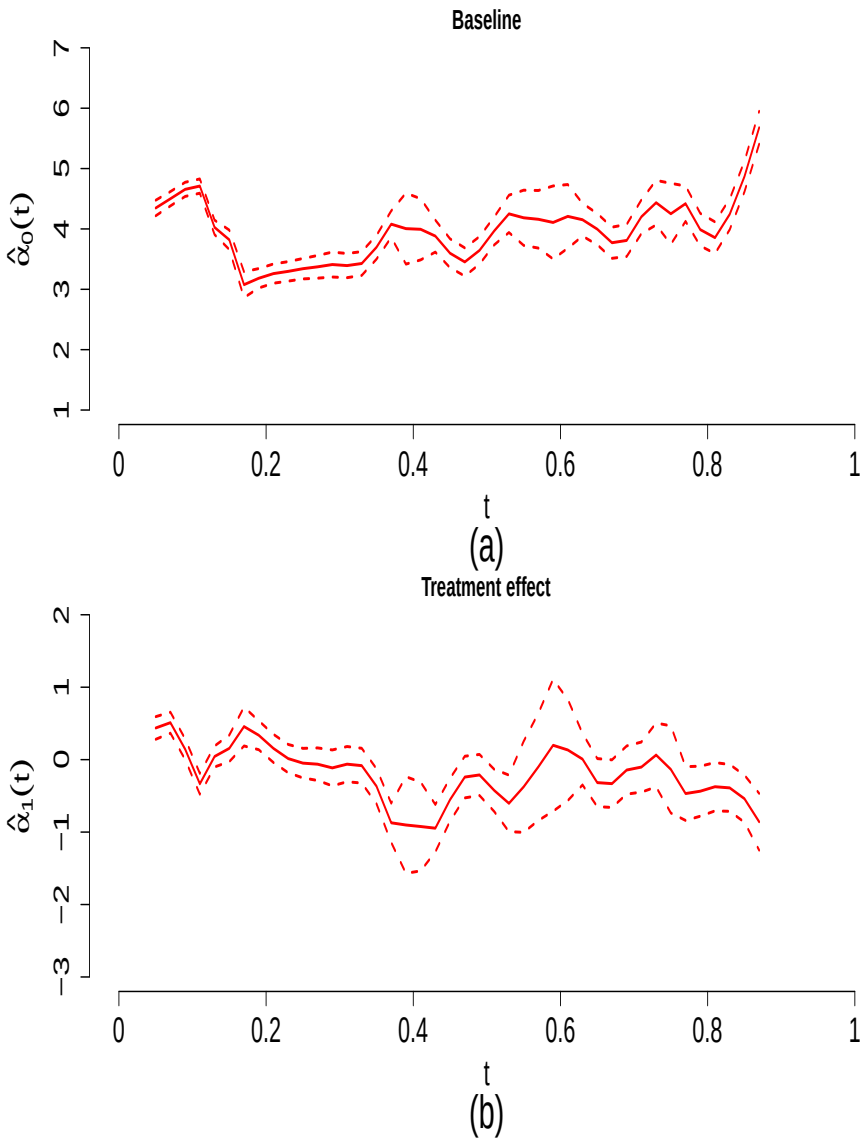

Figure 4: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on HIV-1 RNA data under model (2.28) by MGV approach using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by MGV approach; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by MGV approach.

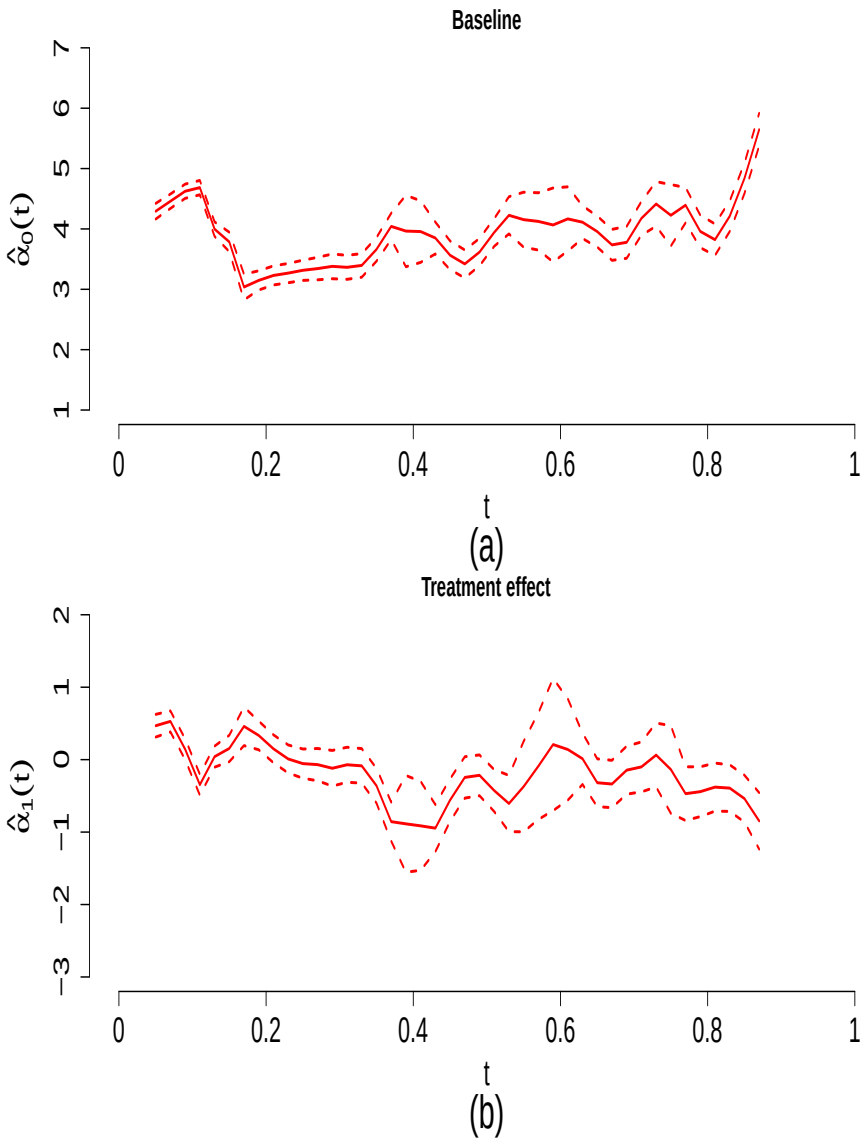

Figure 5: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on HIV-1 RNA data under model (2.28) by QIF approach using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by QIF approach; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by QIF approach.

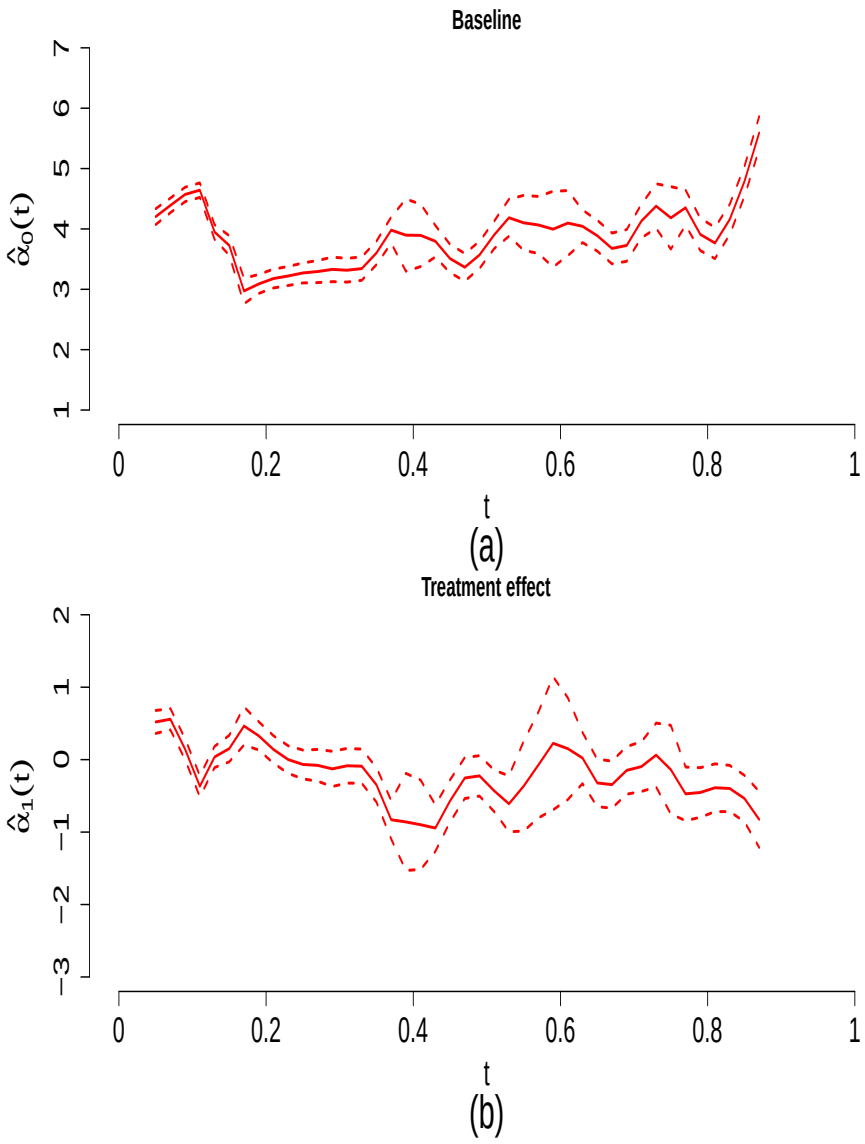

Figure 6: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on HIV-1 RNA data under model (2.28) by WLS approach using $h = 0.06$. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ by WLS approach; (b) is the estimated vaccine effects $\hat{\alpha}_1(t)$ by WLS approach.

CHAPTER 3: IMPROVING ESTIMATION OF GENERALIZED SEMI-PARAMETRIC VARYING-COEFFICIENT MODELS USING COVARIANCE FUNCTION

Chapter 3 is organized as follows. In Section 3.1, we introduce the generalized semi-parametric varying-coefficient model with treatment switching effects, and estimation procedure that can be applied when considering the correlation within subjects. In particular, estimation of parametric parameters and nonparametric functions are implemented by profile-kernel method, and iterated with the estimation of covariance function of responses, which can be fulfilled by various methods in literature, such as GEE, QL, MGV and QIF methods. The asymptotic properties for the nonparametric and parametric estimators of the above methods are developed in Section 3.2. The finite sample performance of the proposed estimators with different methods is examined in simulations in Section 3.3.

3.1 Estimation Procedures

3.1.1 Model Description

Suppose the data consist of n clusters with the ith (i=1,. . . , n) cluster having J_i observations. In particular, in longitudinal study, a cluster represents an individual. Suppose there is a random sample of n subjects and τ is the end of follow-up. Let $X_i(t)$ and $U_i(t)$ be possibly time-dependent covariates for the ith subject. Suppose that observations of the response process $Y_i(t)$ for subject i are taken at the sampling

time points $0 \leq T_{i1} < T_{i2} < \cdots < T_{iJ_i} \leq \tau$. The sampling times can be irregular and dependent on covariates. In addition, some subjects may drop out of the study early. Let $N_i(t) = \sum_{j=1}^{J_i} I(T_{ij} \leq t)$ be the number of observations taken from the ith subject by time t, where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time, whichever comes first. The responses for subject i can only be observed at time points before C_i. Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the counting process of sampling times.

Assume that $\{Y_i(\cdot), X_i(\cdot), U_i(\cdot), N_i(\cdot), i = 1, \cdots, n\}$ are independent identically distributed (iid) random processes. The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t) \,|X_i(t), U_i(t), C_i \geq t\} = E\{dN_i^*(t)|X_i(t), U_i(t)\}$ and $E\{Y_i(t)|X_i(t), U_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t),\, U_i(t)\}$. Assume that $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $U_i(t)$ and $C_i \geq t$. The censoring time C_i is allowed to depend on $X_i(\cdot)$ and $U_i(\cdot)$.

Let $X_i(t) = (X_{1i}^T(t), X_{2i}^T(t), X_{3i}^T(t))^T$ consist of three parts of dimensions p_1, p_2 and p_3, respectively, over the time interval $[0, \tau]$. Let $U_i(t)$ be the scalar covariate process with support $\mathcal{U}$. To characterize the treatment switching effects of $X_{3i}(t)$ with respect to $U_i(t)$, we propose the generalized semiparametric varying-coefficient model

$$\mu_i(t) = E\{Y_i(t)|X_i(t), U_i(t)\} = g^{-1}\{\alpha^T(t)X_{1i}(t) + \beta^T X_{2i}(t) + \gamma^T(U_i(t);\theta)X_{3i}(t)\}, \tag{3.1}$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known monotonic and differentiable link function, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters and $\gamma(u) = \gamma(u;\theta)$ is a p_3-dimensional vector of para-

metric functions specified up to a finite number of unknown parameters θ. Setting the first component of $X_{1i}(t)$ as 1 gives a nonparametric baseline function. $\gamma(u)$ is the effect of $X_{3i}(t)$ at the covariate level $U_i(t) = u$. Both discrete and continuous longitudinal responses can be modelled with appropriately chosen link functions. For example, the identity link function can be used for continuous response variables while the logit link function can be used for binary responses and logarithm link for Poisson process.

For the motivating example the ACTG 244 study, t is the time since initiation of antiretroviral therapy (ART). It is of interest to know how biomarkers such as viral load and CD4 counts respond to the new treatments. It is natural to assume that the effects of the new treatments depend on the time duration $U_i(t) = t - S_i$ since the switching, where S_i is the time of treatment switching. Letting $X_{3i}(t) = I(t > S_i)$ in (3.1), $\gamma(u)$ represents the change in the conditional mean response at time u after treatment switching adjusting for other covariates $X_{1i}(t)$ and $X_{2i}(t)$. On the other hand, if we let $X_{3i}(t) = X^o_{3i}(t)I(t > S_i)$ where $X^o_{3i}(t)$ are the indicators for the new treatments after switching, then $\gamma(u)$ are the effects of new treatments starting from treatment switching.

3.1.2 Estimation of Regression Coefficients

In this section, we apply different methods to model (3.1) and compare the performance of these methods. The proposed approach utilizes profile-kernel method to estimate the parametric, and nonparametric coefficients, meanwhile uses quasi-likelihood (QL), Minimum Generalized Variance (MGV) or Quadratic Inference Function (QIF)

method to estimate the variance and correlation coefficients within subjects.

We use a profile weighted least square approach to estimate the model coefficients: first, given parametric part we approximate the non-parametric part by local linear smoothing and ignore the correlation within subjects in the estimation function; second, given non-parametric part, we incorporate the correlation matrix as a weight in a weighted least square function to estimate the parametric part. By iteration of the above two steps until convergence we achieve the estimation of model coefficients.

At each t_0, let $\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2)$ be the first order Taylor expansion of $\alpha(\cdot)$ for $t \in \mathcal{N}_{t_0}$, a neighborhood of t_0, where $\dot{\alpha}(t_0)$ is the derivative of $\alpha(t)$ at $t = t_0$. Denote $\alpha^*(t_0) = (\alpha^T(t_0), \dot{\alpha}^T(t_0))^T$, $X^*_{1i}(t, t - t_0) = X_{1i}(t) \otimes (1, t - t_0)^T$, where $\otimes$ is the Kronecker product. In the following, we denote $Y_i(T_{ij}) = Y_{ij}$, $\mu_i(T_{ij}) = \mu_{ij}$, $\mathbf{X}_i(T_{ij}) = \mathbf{X}_{ij}$ and $U_i(T_{ij}) = U_{ij}$ for simplicity. Let $\zeta = (\beta^T, \theta^T)^T$, $X^*_{2ij} = (X^T_{2ij}, X^T_{3ij})^T$, and $\eta(U_{ij}, \zeta) = (\beta^T, \gamma^T(U_{ij}, \theta))^T$. For $t \in \mathcal{N}_{t_0}$, model (3.1) can be approximated by

$$\mu_{ij}(t_0, \alpha^*(t_0), \beta, \theta | X_{ij}, U_{ij}) = g^{-1}\{\alpha^{*T}(t_0)X^*_{1ij}(t_0) + \beta^T X_{2ij} + \gamma^T(U_{ij}, \theta)X_{3ij}\},$$

which can be simplified further as

$$\mu_{ij}(t_0, \alpha^*(t_0), \zeta | X_{ij}, U_{ij}) = g^{-1}\{\alpha^{*T}(t_0)X^*_{1ij}(t_0) + \eta^T(U_{ij}, \zeta)X^*_{2ij}\}, \tag{3.2}$$

At each t_0 and for fixed ζ, we propose the following local linear estimating function for $\alpha^*(t_0)$:

$$U_\alpha(\alpha^*; \zeta, t_0) = \sum_{i=1}^{n} X^*_{1i}(t_0)^T \Delta_i(t_0) K^{1/2}_{ih}(t_0) V^{-1}_{1i}(t_0) K^{1/2}_{ih}(t_0) \left[Y_i - \mu^*_i\right], \tag{3.3}$$

where $Y_i = (Y_{i1}, \cdots, Y_{iJ_i})^T$, $X^*_{1i} = (X^*_{1i1}, \cdots, X^*_{1iJ_i})^T$, $\mu^*_i = (\mu^*_{i1}, \cdots, \mu^*_{iJ_i})^T$, $\mu^*_{ij} = \mu_{ij}(t_0, \alpha^*(t_0), \zeta | X_{ij}, U_{ij})$, and $X^*_{1i}(t_0)$ denotes a $J_i \times 2p_1$ matrix with each row vector being $X^*_{1ij}(t_0) = X_{1ij} \otimes (1, T_{ij} - t_0)^T$, $\Delta_i = diag\{\dot{\mu}_{ij}\}$, $\dot{\mu}(\cdot)$ is the first derivative of $\mu(\cdot)$, $V^{-1}_{1i}(t_0)$ is a nonnegative weight process, $K(\cdot)$ is a kernel function, $h = h_n > 0$ is a bandwidth parameter, $K_h(\cdot) = K(\cdot/h)/h$, and $K_{ih}(t_0) = diag\{K_h(T_{ij} - t_0)\}$. The solution to the equation $U_\alpha(\alpha^*; \zeta, t_0) = 0$ is denoted by $\tilde{\alpha}^*(t_0, \zeta)$. Let $\tilde{\alpha}(t_0, \zeta)$ be the first p_1 components of $\tilde{\alpha}^*(t_0, \zeta)$.

Second, given estimated $\alpha(t)$, the profile weighted least-squares estimator $\hat{\zeta}$ is obtained by minimizing the following profile least-squares function with respect to ζ:

$$\ell_\zeta(\zeta) = \frac{1}{n} \sum_{i=1}^{n} [Y_i - \hat{\mu}_i(\zeta)]^T V^{-1}_{2i} [Y_i - \hat{\mu}_i(\zeta)], \tag{3.4}$$

where V^{-1}_{2i} is the working covariance matrix, and μ_i and μ_{ij} are the same as those defined in (2.4) except that they are evaluated at $\hat{\mu}_{ij}(\zeta) = g^{-1}\{\tilde{\alpha}^T(T_{ij}, \zeta) X_{1ij} + \eta^T(U_{ij}, \zeta) X^*_{2ij}\}$. The profile estimator for $\alpha(t_0)$ is obtained by $\hat{\alpha}(t_0) = \tilde{\alpha}(t_0, \hat{\zeta})$ through substitution.

The Newton-Raphson iterative method can be used to find the estimator $\hat{\zeta}$ that minimizes (3.4). Taking the derivative of $\ell_\zeta(\zeta)$ with respect to ζ leads to score function for ζ.

$$U_\zeta(\zeta) = \sum_{i=1}^{n} \{ \frac{\partial \tilde{\alpha}(T_i, \zeta)}{\partial \zeta} X_{1i} + \frac{\partial \eta(U_i, \zeta)}{\partial \zeta} X^*_{2i} \}^T \Delta_i V^{-1}_{2i} [Y_i - \hat{\mu}_i(\zeta)] \tag{3.5}$$

where

$$\frac{\partial \eta(U_i, \zeta)}{\partial \zeta} X^*_{2i} = (\frac{\partial \eta^T(U_{i1}, \zeta)}{\partial \zeta} X^*_{2i1}, \cdots, \frac{\partial \eta^T(U_{iJ_i}, \zeta)}{\partial \zeta} X^*_{2iJ_i})^T,$$

$$\partial \eta^T(U_{ij}, \zeta)/\partial \zeta = \text{diag}\{I_{p_2}, \partial \gamma^T(U_{ij}, \theta)/\partial \theta\}$$

and

$$\frac{\partial\tilde{\alpha}(T_i,\zeta)}{\partial\zeta}X_{1i} = (\frac{\partial\tilde{\alpha}^T(T_{i1},\zeta)}{\partial\zeta}X_{1i1},\cdots,\frac{\partial\tilde{\alpha}^T(T_{iJ_i},\zeta)}{\partial\zeta}X_{1iJ_i})^T.$$

Here $\frac{\partial\tilde{\alpha}(t,\zeta)}{\partial\zeta}$ be the first p_1 components of $\frac{\partial\alpha^*(t,\zeta)}{\partial\zeta}$ which can be expressed in terms of the partial derivatives of $U_\alpha(\alpha^*;\zeta,t)$ at $\alpha^* = \tilde{\alpha}^*(t,\zeta)$. Specifically, since $U_\alpha(\tilde{\alpha}^*(t,\zeta);\zeta,t) \equiv \mathbf{0}_{2p_1}$, it follows that $\tilde{\alpha}^*(t,\zeta)$ satisfies

$$\left\{\frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\alpha^*}\frac{\partial\alpha^*(t,\zeta)}{\partial\zeta} + \frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\zeta}\right\}\Bigg|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)} = \mathbf{0}_{2p_1\times(p_2+p_3)}.$$

Therefore,

$$\frac{\partial\alpha^*(t,\zeta)}{\partial\zeta} = -\left\{\frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\alpha^*}\right\}^{-1}\frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\zeta}\Bigg|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}, \tag{3.6}$$

where

$$\frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\alpha^*} = -\sum_{i=1}^{n} X_{1i}^*(t)^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\Delta_i(t)X_{1i}^*(t), \tag{3.7}$$

and $X_{1i}^*(t)$ denotes a $J_i \times 2p_1$ matrix with each row vector being $X_{1i}^*(T_{ij},T_{ij}-t) = X_{1i}(T_{ij})\otimes(1,T_{ij}-t)^T$ and

$$\frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial\zeta} = -\sum_{i=1}^{n} X_{1i}^*(t)^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\Delta_i(t)\{\frac{\partial\eta(U_i,\zeta)}{\partial\zeta}X_{2i}^*\}. \tag{3.8}$$

When estimating $\alpha(t)$, we let $R_{1i}(t)$ in $V_{1i}(t)$ be an identity matrix to ignore correlation in estimating equation (2.4), since it has been shown in literature that a more efficient estimation of nonparametric part is achieved by ignoring the correlation within subjects (Lin and Carroll, 2000, 2001b; Fan et al., 2007). However, the most efficient estimation for ζ is obtained by letting V_{2i} in equation (3.5) be the true covariance matrix of Y, which is proved in Theorem 3.2.

When link function is the identity link, $\tilde{\alpha}^*(t_0, \zeta)$ and $\hat{\zeta}$ can be solved explicitly as the root of the estimating function (3.3) and (3.5). However, under a general link function, they need to be solved using iterative algorithm such as Newton-Raphson method. The estimation procedure iteratively updates estimates of the nonparametric component $\tilde{\alpha}^*(t, \zeta)$ and the parametric component $\hat{\zeta}$ until convergence. We denote the first p_1 component of the convergent $\tilde{\alpha}^*(t, \zeta)$ as $\hat{\alpha}(t)$.

3.1.3 Estimation of Covariance Function

The following section concentrates on estimation of V_{1i} and V_{2i}. To ensure positive definite of the working covariance matrix V_{1i} and V_{2i} , we decompose it as the sandwich form:

$$V_{1i}(t) = A_i(t) R_{1i}(t, \rho) A_i(t) \tag{3.9}$$

where $R_{1i}(t, \rho) = \mathbf{I}$. and similarly,

$$V_{2i} = A_i R_{2i}(\rho) A_i \tag{3.10}$$

where $A_i(t) = diag\{\sigma(t)\}$ and $R_{2i}(\rho)$ is the working correlation matrix of Y_i. Given $\hat{\alpha}(t)$ and $\hat{\zeta}$ obtained in Section (3.1.2), we can get estimated residual $\hat{r}_{ij} = Y_{ij} - \hat{\mu}_{ij}$, where $\hat{\mu}_{ij} = g^{-1}\{\hat{\alpha}(T_{ij}) X_{1ij} + \eta^T(U_{ij}, \hat{\zeta}) X^*_{2ij}\}$. The marginal variance of Y can be estimated either by methods of moments in GEE, or via univariate kernel smoothing in Fan et al. (2007). Especially when variance of Y depend on time and is nonstationary, we use the following formula in Fan et al. (2007):

$$\widehat{var}(Y_{ij}) = \hat{\sigma}^2(t) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{J_i} \hat{r}_{ij}^2 K_{h1}(t - T_{ij})}{\sum_{i=1}^{n} \sum_{j=1}^{J_i} K_{h1}(t - T_{ij})} \tag{3.11}$$

To estimate the correlation matrix used in above weighted least square function, there are quasi-likelihood(QL) approach, minimum generalized variance (MGV) approach both proposed by Fan et al. (2007) and weighted least square(WLS) approach newly proposed here. QL, MGV and WLS approaches share the same framework which is derived from Generalized Estimating Equation(GEE) framework, and their only difference is the objective functions in their optimization procedure, see equation (2.19) for QL, equation (2.20) for MGV and equation (2.21) for WLS. Quadratic inference function method (QIF) proposed by Qu et al. (2000) doesn't involve estimating the correlation parameter but incorporate correlation structure in estimating equations to improve efficiency, see equation (2.23).

Specifically, the estimations of $\alpha(t)$, ζ, $\sigma^2(t)$ and ρ can be accomplished through the following iterated algorithm:

Computational algorithm

1. Let $\hat{\alpha}(t)^{\{0\}}$ and $\hat{\zeta}^{\{0\}}$ be initial values which can obtained assuming working independence simultaneously in score function (3.3) and (3.5);

2. Given $\hat{\alpha}(t)^{\{0\}}$ and $\hat{\zeta}^{\{0\}}$, use formula (3.11) or methods of moments to estimate $\hat{A}_i$, and use QL, or MGV, or WLS approach to estimate ρ in $R_{2i}(\rho)$, hence $\hat{\Sigma}_i^{\{0\}} = \hat{A}_i \hat{R}_{2i}(\hat{\rho}) \hat{A}_i$.

3. Let the correlation matrix $R_{1i}(t)$ in score function of $\alpha(t)$ equation (3.3) be identity matrix, while the weight matrix in score function of ζ equation (3.5) be $V_{2i}^{-1} = \{\hat{\Sigma}_i^{\{0\}}\}^{-1}$, by profile-kernel procedure in Section 3.1.2, we get updated $\hat{\alpha}(t)^{\{1\}}$ and $\hat{\zeta}^{\{1\}}$.

4. Repeating step 2 and 3, the estimators $\hat{\alpha}^{\{m\}}(t)$, $\hat{\zeta}^{\{m\}}$ and $\hat{\Sigma}_i^{\{m\}}$ are updated at each iteration until convergence. $\hat{\zeta}$, $\hat{\alpha}(t)$ and $\hat{\Sigma}_i$ are $\hat{\zeta}^{\{m\}}$, $\hat{\alpha}^{\{m\}}(t)$ and $\hat{\Sigma}_i^{\{m\}}$ respectively, at convergence.

In our empirical experience, m will be no more than 3 and we will not lose any statistical efficiency provided the initial estimator is good enough. The initial estimators are obtained by assuming working independence which are consistent, thus after only a few iterations the convergence is achieved.

3.2 Asymptotic Properties

In what follows, let $J_i = J < \infty$ (i.e.,assuming finite cluster size) and T_i be a continuous observational-level covariate (e.g., a time-varying covariate in longitudinal studies). We allow $T_i = (T_{i1}, \cdots, T_{iJ})$ as well as X_i and U_i to be correlated unless stated otherwise, let $f_j(\cdot)$ denote the marginal density of T_{ij}. We further assume that the $(Y_i, X_i, U_i, T_i)(i = 1, \cdots, n)$ are iid observations with a continuous density function, and both $V_{1i}(\mu_i, \rho) = V_1(\mu_i, \rho)$ and $V_{2i}(\mu_i, \rho) = V_2(\mu_i, \rho)$ are invertible. Let ν_1^{jk} and ν_2^{jk} denote the (j,k)th element of $V_1^{-1}(\cdot)$ and $V_2^{-1}(\cdot)$ respectively. Let $C_K(r) = \int z^r K(z)dz$ and $\gamma_K(r) = \int z^r K^2(z)dz$. We further assume that $nh \to \infty$ as $n \to \infty$ and $h \to 0$.

We first rewrite the profile estimating equations for ζ in (3.5) as

$$U_\zeta(\zeta) = \sum_{i=1}^{n} \{\frac{\partial \tilde{\alpha}(T_i, \zeta)}{\partial \zeta} X_{1i} + \frac{\partial \eta(U_i, \zeta)}{\partial \zeta} X_{2i}^*\}^T \Delta_i V_{2i}^{-1} \left[Y_i - \hat{\mu}_i(\zeta)\right]] \tag{3.12}$$

Let ζ_0 and $\alpha_0(t)$ be the true values of ζ and $\alpha(t)$ under model (3.2), respectively. Let $\mu_{ij} = g^{-1}\{\alpha_0^T(T_{ij})X_{1ij} + \eta^T(U_{ij}, \zeta_0)X_{2ij}^*\}$, $\dot{\mu}_{ij}$ is the first derivative of $\mu_{ij}(\cdot)$, $\Delta_i =$

$diag\{\dot{\mu}_{ij}\}$ and $\epsilon_{ij} = Y_{ij} - \mu_{ij}$. Appendix A shows that, asymptotically,

$$\frac{\partial \alpha^*(t,\zeta)}{\partial \zeta} = -e_{11}^{-1}(t)e_{12}(t), \tag{3.13}$$

where, suppressing the index i denoting $\mu_j = g^{-1}\{\alpha_0^T(T_j)X_{1j} + \eta^T(U_j,\zeta_0)X_{2j}^*\}$ $(j = 1,\cdots,J)$, define

$$e_{11}(t) = \sum_{j=1}^{J} E\big[\Delta_{jj}^2(t)\nu_1^{jj}(t)X_{1j}X_{1j}^T|\mathbf{T}_j = \mathbf{t}\big]f_j(\mathbf{t})$$

and

$$e_{12}(t) = \sum_{j=1}^{J} E\big[\Delta_{jj}^2(t)\nu_1^{jj}(t)X_{1j}\{\frac{\partial \eta^T(U_j,\zeta)}{\partial \zeta}X_{2j}^*\}^T|\mathbf{T}_j = \mathbf{t}\big]f_j(\mathbf{t}),$$

where $\Delta_{jj}(t) = \dot{\mu}_j(\alpha(t),\zeta|X_j,U_j)$. Denote

$$B_{ij} = -e_{12}^T(T_{ij})e_{11}^{-1}(T_{ij})X_{1ij} + \frac{\partial \eta^T(U_{ij},\zeta_0)}{\partial \zeta}X_{2ij}^*$$

and

$$B_i = (B_{i1},\cdots,B_{iJ_i})^T.$$

Let $\hat{\mu}_{ij} = g^{-1}\{\hat{\alpha}^T(T_{ij})X_{1ij} + \eta^T(U_{ij},\hat{\zeta})X_{2ij}^*\}$, $\hat{\Delta}_i = diag\{\dot{\hat{\mu}}_{ij}\}$ and $\hat{\epsilon}_{ij} = Y_{ij} - \hat{\mu}_{ij}$. Let

$$\hat{E}_{11}(t) = n^{-1}\sum_{i=1}^{n}\{X_{1i}^T\hat{\Delta}_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\hat{\Delta}_i(t)X_{1i}\}$$

and

$$\hat{E}_{12}(t) = n^{-1}\sum_{i=1}^{n}\big[X_{1i}^T\hat{\Delta}_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\hat{\Delta}_i(t)\{\frac{\partial \eta(U_i,\hat{\zeta})}{\partial \zeta}X_{2i}^*\}\big].$$

Denote

$$\hat{B}_{ij} = -\hat{E}_{12}^T(T_{ij})\hat{E}_{11}^{-1}(T_{ij})X_{1ij} + \frac{\partial \eta^T(U_{ij},\hat{\zeta})}{\partial \zeta}X_{2ij}^*$$

and

$$\hat{B}_i = (\hat{B}_{i1}, \cdots, \hat{B}_{iJ_i})^T.$$

The following theorems characterize the asymptotic properties of the proposed estimator $\hat{\zeta}$ and $\hat{\alpha}(t)$. Conditions and proofs are given in the Appendix.

Theorem 3.1. *Under Condition I in the Appendix, the estimator $\hat{\zeta}$ is consistent for ζ_0, and $\sqrt{n}(\hat{\zeta} - \zeta_0)$ converges in distribution to a mean zero Gaussian random vector with covariance matrix $P^{-1}DP^{-1}$, i.e.,*

$$\sqrt{n}(\hat{\zeta} - \zeta_0) \xrightarrow{\mathcal{D}} (0, P^{-1}DP^{-1}), \tag{3.14}$$

where suppressing the subscript i in each term inside the expectations,

$$P = E\left[B^T \Delta V_2^{-1} \Delta B\right],$$

and

$$D = E\left[B^T \Delta V_2^{-1} \Sigma V_2^{-1} \Delta B\right],$$

where $\Sigma = cov(Y|X, U, T)$.

The matrix P can be consistently estimated by

$$\hat{P} = n^{-1} \sum_{i=1}^{n} \left[\hat{B}_i^T \hat{\Delta}_i \hat{V}_{2i}^{-1} \hat{\Delta}_i \hat{B}_i\right]$$

and D can be consistently estimated by

$$\hat{D} = n^{-1} \sum_{i=1}^{n} \left[\hat{B}_i^T \hat{\Delta}_i \hat{V}_{2i}^{-1} (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T \hat{V}_{2i}^{-1} \hat{\Delta}_i \hat{B}_i\right].$$

Theorem 3.2. *When V_{2i} is taken to be the conditional variance-covariance matrix*

of Y_i given X_i and U_i for $i = 1, \ldots, n$, then $P = D$. In this case

$$\sqrt{n}(\hat{\zeta} - \zeta_0) \xrightarrow{\mathcal{D}} \left(0, D_0^{-1}\right),$$

where $D_0 = E\left\{B^T \Delta \Sigma^{-1} \Delta B\right\}$ and

$$P^{-1} D P^{-1} - D_0^{-1} \geq 0, \tag{3.15}$$

for any matrix V_{2i}, where $A \geq 0$ means that the matrix A is nonnegative definite.

Theorem 3.3. *Under Condition I, we have that $\hat{\alpha}(t)$ converges to $\alpha_0(t)$ uniformly in $t \in [t_1, t_2]$, and*

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2} C_K(2) h^2 \ddot{\alpha}_0(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\alpha(t)\right), \tag{3.16}$$

where $C_K(2) = \int_{-1}^{1} t^2 K(t)\, dt$, $\gamma_K(0) = \int_{-1}^{1} K^2(t)\, dt$, $\sigma_{jj} = var(Y_j | X_j, U_j, T_j)$, σ_{jj} is the (j,j)th element of $\Sigma = cov(Y_i | X_i, U_i)$ and $\ddot{\alpha}_0(t)$ is the second derivative of the true $\alpha_0(t)$ with respect to t,

$$\Sigma_\alpha(t) = e_{11}^{-1}(t) \Sigma_e(t) e_{11}^{-1}(t)$$

and

$$\Sigma_e(t) = \gamma_K(0) \sum_{j=1}^{J} E\{\Delta_{jj}^2(t) (\nu_1^{jj}(t))^2 \sigma_{jj} X_{1j} X_{1j}^T | \mathbf{T}_j = \mathbf{t}\} f_j(t).$$

Thus

$$\Sigma_\alpha(t) \approx \frac{\gamma_K(0) \sum_{j=1}^{J} E\{\Delta_{jj}^2(t) (\nu_1^{jj}(t))^2 \sigma_{jj} X_{1j} X_{1j}^T | \mathbf{T}_j = \mathbf{t}\} f_j(t)}{\left\{ \sum_{j=1}^{J} E[\Delta_{jj}^2(t) \nu_1^{jj}(t) X_{1j} X_{1j}^T | \mathbf{T}_j = \mathbf{t}] f_j(t) \right\}^{-2}}$$

The variance-covariance matrix $\Sigma_\alpha(t)$ can be estimated consistently replacing $e_{11}(t)$

by $\hat{E}_{11}(t)$ and $\Sigma_e(t)$ by

$$\hat{\Sigma}_e(t) = n^{-1}h\sum_{i=1}^{n}\{X_{1i}^T\hat{\Delta}_i K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)(Y_i-\hat{\mu}_i)(Y_i-\hat{\mu}_i)^T K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\hat{\Delta}_i X_{1i}\}.$$

It follows that $\Sigma_\alpha(t)$ is minimized when assuming working independence $R_1 = \mathbf{I}$ and is

$$\Sigma_\alpha(t) \approx \left\{\gamma_K(0)\sum_{j=1}^{J} E[\Delta_{jj}^2(t)\sigma_{jj}^{-1}X_{1j}X_{1j}^T|\mathbf{T}_j = \mathbf{t}]f_j(t)\right\}^{-1}$$

3.3 Simulation Studies

For correlated continuous responses in generalized varying-coefficient model, we use a similar setting in Qi et al. (2016)'s simulation studies with identity link.

$$E\{Y_i(t)|\mathbf{X}_i, S_i\} = \alpha_0(t) + \alpha_1(t)X_{1i}(t) + \beta X_{2i} + \gamma(t - S_i, \theta)X_{3i}I(t > S_i),$$

for $0 \le t \le \tau$ with $\tau = 3.5$, where $\alpha_0(t) = 0.2\sqrt{t}$, $\alpha_1(t) = 0.1\sin(t)$, $\gamma(u, \theta) = \theta_1 exp(-\theta_2 u)$ and $\zeta = (\beta, \theta_1, \theta_2) = (0.1, 1.0, 0.5)$, $X_{1i}(t) = (t/3 + N(0, 1.5^2))/6$ and X_{3i} is a uniform random variable on $[-1, 1]$, X_{2i} is a Bernoulli random variable with success probability of 0.5 and S_i is a uniform random variable on $[0, 1]$. S_i, X_{2i} and X_{3i} are all subject-level covariates which doesn't change within subject i. The observation time follows a Poisson process with the proportional mean rate model $h(t|X_i, S_i) = 1.5\exp(0.7X_{2i})$. The censoring time C_i is generated from a uniform distribution on $[1.5, 8]$. There are approximately six observations per subject on $[0, \tau]$ and about 30% of subjects are censored before $\tau = 3.5$.

We consider two different model settings for the error part:

(C1) ARMA model. The error part $\epsilon_i(t)$ is a Gaussian process with mean 0,

variance changing with time $\sigma_i^2(t) = 0.5exp(t/12)$ and the correlation structure is ARMA(1,1), i.e., $corr(\epsilon_i(s), \epsilon_i(t)) = \gamma\rho^{|t-s|}$ for $s \neq t$. We let $\theta = (\gamma, \rho) = (0.85, 0.9)$, $\theta = (\gamma, \rho) = (0.85, 0.6)$, $\theta = (\gamma, \rho) = (0.85, 0.3)$ to consider strong, moderate and weak correlation respectively.

(C2) Exchange Model. The error $\epsilon_i(t) = Y_i(t) - E\{Y_i(t)|\mathbf{X}_i, S_i\}$ has a normal distribution with mean ϕ_i and variance ν^2, and ϕ_i is $N(0, 1)$. By this setting, it can be shown that the correlation structure of error within subjects is exchangeable correlation structure with correlation coefficient $\rho = 0.8$ if $\nu = 0.5$, $\rho = 0.5$ if $\nu = 1$ and $\rho = 0.2$ if $\nu = 2$. We choose above three different values of ν to consider strong, moderate and weak correlations.

In the following, we present simulation results of the proposed methods and select the bandwidth by the K-fold cross-validation bandwidth selection method. The bandwidth selected is $h = 1.4$ for ARMA model and $h = 2.9$ for Exchange Model when sample size $n = 400$. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ is used. We take $t_1 = h/2$ and $t_2 = \tau - h/2$ in the estimating functions for $\alpha(t)$ to avoid larger variations on the boundaries.

The performances of the estimators for ζ are measured through the Bias, the sample standard error of the estimators (SEE), the sample mean of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). Boxplots are drawn to show the median, quantiles and outliers. Each entry of the table is calculated based on 1000 repetitions for sample size $n = 400$. Each table we have four blocks, and each block is for one of the following scenarios: (1) working independence case; (2) assuming Exchangeable correlation structure; (3) assuming ARMA correlation

structure; (4) assuming a mixed correlation structure. QL, MGV, WLS and QIF approaches are compared when assuming an exchangeable, ARMA or mixed correlation structure.

Table 15, Table 16 and Table 17 show the results for strongly, moderately and weakly correlated observations within subjects for ARMA model respectively. Table 19, Table 20 and Table 21 show the results for strongly, moderately and weakly correlated observations within subjects for Exchange model respectively. Since QL, MGV and WLS methods involve estimating the correlation coefficients, we list the estimation results when assuming true correlation and a mixed correlation for each model. Table 18 is for ARMA model and Table 22 is for Exchange model.

We get several conclusions from these tables:

1. All the estimates are unbiased under each scenario by each method. There are a good agreement of SEE and ESE for θ_1 and θ_2, while ESE tends to be smaller than SEE for β. Since θ_1 and θ_2 are the coefficients of time-variant covariate $X_{3i}I(t > S_i)$, while β is the coefficient of time-invariant covariate X_2, we can say that the standard deviation formula performs well for the coefficient of time-variant covariate but tends to underestimate that of time-invariant covariate. CP value are near the 95% nominal level for most cases.

2. When estimating the coefficients of time-variant covariate θ_1 and θ_2, QL, MGV QIF and WLS approaches all perform well and reduce the SEE comparing with working independence(WI) case, while for β, the coefficient of time-invariant covariate, SEE doesn't change much comparing with WI case. This phenomenon

is also shown in (Lin and Carroll, 2001b; Wang et al., 2005).

3. When estimating θ_1 and θ_2, QL approach achieve the smallest SEE for each scenario which means QL is most efficient estimator. From results of correlation coefficients estimates in Table 18 and Table 22, QL approach almost target the correlation coefficient either assuming the true correlation structure or assuming a mixed one, while MGV and WLS tend to underestimate the true correlation coefficients.

4. The stronger the correlation is, more efficiency will be obtained when estimating the time-variant covariates. The reduction of SEE by QL, MGV, QIF and WLS methods compared to WI in Table 15 where correlation coefficient is $\theta = (\gamma, \rho) = (0.85, 0.9)$ will be more than those in Table 16 and Table 17 where $\theta = (\gamma, \rho) = (0.85, 0.6)$ and $\theta = (\gamma, \rho) = (0.85, 0.3)$ respectively. Similar results can be drawn from Table 19, Table 20 and Table 21.

5. When the correlation structure is misspecified, the efficiency achieved is not as much as the case of the correlation is true or assuming a mixed correlation structure and there are extreme outliers. The performance of assuming true correlation structure and assuming a mixed correlation structure are close. Thus assuming a mixed correlation structure when we don't know the true correlation structure performs as well as when we know the correlation structure in advance.

Table 15: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 1.4$ based on 1000 simulations under ARMA model with strong correlation coefficients $(0.85, 0.9)$.

	$\theta = (0.85, 0.9)$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	-.0009	.0766	.0754	.950	-.0022	.0953	.0923	.944	.0164	.1146	.1118	.945
Assuming ARMA correlation structure(True)												
QL	-.0016	.0708	.0662	.926	-.0003	.0485	.0475	.954	.0073	.0677	.0662	.944
MGV	-.0016	.0697	.0674	.944	-.0008	.0540	.0520	.941	.0092	.0794	.0786	.945
QIF	-.0018	.0702	.0673	.938	-.0010	.0638	.0597	.937	.0094	.0873	.0845	.945
WLS	-.0018	.0698	.0674	.943	-.0009	.0568	.0545	.944	.0088	.0760	.0756	.944
Assuming Mixed correlation structure												
QL	-.0017	.0708	.0662	.926	-.0004	.0484	.0475	.954	.0071	.0676	.0661	.945
MGV	-.0018	.0695	.0670	.941	-.0008	.0538	.0519	.938	.0078	.0727	.0724	.942
QIF	-.0024	.0704	.0666	.938	-.0008	.0616	.0576	.935	.0094	.0818	.0795	.939
WLS	-.0018	.0698	.0674	.943	-.0010	.0570	.0547	.945	.0087	.0760	.0755	.944
Assuming Exchangeable correlation structure(Misspecification)												
QL	-.0020	.0720	.0675	.932	-.0005	.0554	.0545	.945	.0055	.0730	.0735	.955
MGV	-.0020	.0698	.0670	.939	-.0010	.0574	.0560	.942	.0063	.0714	.0721	.947
QIF	-.0024	.0730	.0689	.930	-.0009	.0696	.0669	.937	.0104	.0821	.0834	.952
WLS	-.0020	.0701	.0676	.940	-.0014	.0614	.0596	.940	.0078	.0743	.0748	.937

Table 16: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 1.4$ based on 1000 simulations under ARMA model with moderate correlation coefficients $(0.85, 0.6)$.

	$\theta = (0.85, 0.6)$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	-.0015	.0660	.0643	.943	-.0022	.1033	.1015	.946	.0117	.1088	.1141	.963
Assuming ARMA correlation structure(True)												
QL	-.0024	.0599	.0569	.936	.0013	.0641	.0648	.949	.0104	.0888	.0888	.949
MGV	-.0022	.0598	.0572	.938	.0012	.0686	.0681	.939	.0099	.0884	.0902	.953
QIF	-.0020	.0606	.0577	.935	.0005	.0756	.0734	.942	.0096	.0957	.0950	.946
WLS	-.0022	.0603	.0579	.936	.0004	.0732	.0723	.938	.0101	.0906	.0931	.950
Assuming Mixed correlation structure												
QL	-.0025	.0600	.0569	.936	.0012	.0641	.0648	.949	.0103	.0888	.0887	.950
MGV	-.0024	.0605	.0576	.937	.0011	.0717	.0714	.941	.0091	.0901	.0928	.958
QIF	-.0026	.0605	.0570	.933	.0010	.0753	.0723	.931	.0098	.0935	.0925	.946
WLS	-.0022	.0604	.0579	.936	.0003	.0735	.0726	.940	.0100	.0905	.0930	.950
Assuming Exchangeable correlation structure(Misspecification)												
QL	-.0027	.0627	.0592	.936	.0009	.0823	.0828	.952	.0084	.1036	.1069	.963
MGV	-.0026	.0626	.0591	.934	.0009	.0823	.0828	.951	.0084	.1029	.1063	.965
QIF	-.0025	.0633	.0595	.927	.0003	.0879	.0858	.942	.0099	.0957	.0996	.959
WLS	-.0022	.0622	.0592	.932	-.0007	.0860	.0854	.950	.0087	.0957	.1007	.964

Table 17: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 1.4$ based on 1000 simulations under ARMA model with weak correlation coefficients $(0.85, 0.3)$.

	$\theta = (0.85, 0.3)$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	-.0009	.0548	.0537	.948	-.0008	.1036	.1030	.951	.0115	.1119	.1108	.943
Assuming ARMA correlation structure(True)												
QL	-.0018	.0497	.0480	.945	-.0018	.0771	.0764	.948	.0079	.0914	.0918	.949
MGV	-.0019	.0499	.0483	.947	-.0006	.0806	.0797	.948	.0098	.0950	.0950	.944
QIF	-.0015	.0510	.0493	.948	-.0010	.0854	.0828	.946	.0090	.0999	.0966	.941
WLS	-.0015	.0505	.0487	.947	-.0007	.0832	.0825	.947	.0092	.0955	.0951	.946
Assuming Mixed correlation structure												
QL	-.0019	.0497	.0480	.943	-.0019	.0771	.0764	.948	.0079	.0915	.0917	.950
MGV	-.0017	.0505	.0489	.945	-.0003	.0858	.0849	.939	.0098	.0978	.0979	.946
QIF	-.0021	.0511	.0485	.945	-.0008	.0850	.0819	.940	.0091	.0987	.0950	.945
WLS	-.0016	.0505	.0487	.947	-.0007	.0834	.0828	.947	.0092	.0956	.0952	.945
Assuming Exchangeable correlation structure(Misspecification)												
QL	-.0020	.0521	.0504	.940	-.0001	.0952	.0943	.937	.0113	.1100	.1109	.955
MGV	-.0021	.0530	.0513	.942	.0003	.0970	.0957	.938	.0131	.1176	.1184	.955
QIF	-.0017	.0532	.0504	.927	.0008	.0949	.0935	.938	.0115	.1044	.1033	.946
WLS	-.0015	.0523	.0506	.938	-.0002	.0956	.0951	.944	.0106	.1052	.1057	.950

Table 18: Estimation of correlation coefficients on 1000 simulations for ARMA model by different approaches.

Method	Assuming Mixed Structure	Assuming ARMA Structure(true)
	$\theta = (0.85, 0.9)$	
QL	$0.750 \times 0.882^{-\lvert s-t\rvert} + 0.091$	$0.847 \times 0.899^{-\lvert s-t\rvert}$
MGV	$0.303 \times 0.607^{-\lvert s-t\rvert} + 0.250$	$0.607 \times 0.607^{-\lvert s-t\rvert}$
WLS	$0.298 \times 0.637^{-\lvert s-t\rvert} + 0.150$	$0.439 \times 0.787^{-\lvert s-t\rvert}$
	$\theta = (0.85, 0.6)$	
QL	$0.834 \times 0.577^{-\lvert s-t\rvert} + 0.005$	$0.847 \times 0.596^{-\lvert s-t\rvert}$
MGV	$0.303 \times 0.607^{-\lvert s-t\rvert} + 0.250$	$0.607 \times 0.607^{-\lvert s-t\rvert}$
WLS	$0.442 \times 0.360^{-\lvert s-t\rvert} + 0.039$	$0.471 \times 0.430^{-\lvert s-t\rvert}$
	$\theta = (0.85, 0.3)$	
QL	$0.839 \times 0.282^{-\lvert s-t\rvert} + 0.002$	$0.848 \times 0.297^{-\lvert s-t\rvert}$
MGV	$0.303 \times 0.607^{-\lvert s-t\rvert} + 0.050$	$0.607 \times 0.607^{-\lvert s-t\rvert}$
WLS	$0.482 \times 0.152^{-\lvert s-t\rvert} + 0.012$	$0.489 \times 0.178^{-\lvert s-t\rvert}$

Table 19: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 2.9$ based on 1000 simulations under Exchange model with strong correlation $\rho = 0.8$.

	$\rho = 0.8$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	.0047	.1170	.1138	.943	.0012	.1239	.1329	.964	.0219	.1562	.1549	.940
Assuming Exchangeable correlation structure(True)												
QL	.0018	.1054	.1016	.935	-.0010	.0598	.0600	.957	.0065	.0656	.0637	.947
MGV	.0035	.1084	.1063	.939	.0011	.0913	.0979	.968	.0140	.1129	.1132	.950
QIF	.0021	.1074	.1046	.936	.0003	.0889	.0890	.952	.0118	.1061	.1042	.941
WLS	.0027	.1044	.1027	.937	.0003	.0685	.0717	.964	.0081	.0799	.0795	.946
Assuming Mixed correlation structure												
QL	.0018	.1054	.1016	.936	-.0010	.0598	.0600	.955	.0064	.0658	.0637	.945
MGV	.0029	.1058	.1041	.939	-.0007	.0780	.0815	.964	.0107	.1036	.1056	.952
QIF	.0022	.1063	.1023	.936	-.0011	.0829	.0813	.948	.0108	.1012	.0978	.943
WLS	.0027	.1044	.1027	.937	.0003	.0685	.0717	.965	.0081	.0805	.0800	.945
Assuming ARMA correlation structure(Misspecification)												
QL	.0021	.1050	.1020	.935	-.0014	.0650	.0649	.954	.0070	.0747	.0725	.941
MGV	.0026	.1050	.1034	.943	-.0006	.0722	.0736	.956	.0112	.1054	.1054	.949
QIF	.0030	.1076	.1038	.937	-.0014	.0888	.0879	.952	.0122	.1140	.1122	.940
WLS	.0027	.1045	.1028	.939	.0002	.0691	.0720	.969	.0087	.0862	.0862	.948

Table 20: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 2.9$ based on 1000 simulations under Exchange model with moderate correlation $\rho = 0.5$.

	$\rho = 0.5$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	.0034	.1189	.1184	.947	-.0091	.1670	.1584	.942	.0307	.1830	.1823	.946
Assuming Exchangeable correlation structure(True)												
QL	.0016	.1081	.1086	.954	-.0068	.1184	.1138	.945	.0183	.1243	.1216	.952
MGV	.0023	.1110	.1112	.951	-.0059	.1338	.1288	.944	.0208	.1422	.1422	.949
QIF	.0016	.1112	.1101	.948	-.0079	.1309	.1250	.937	.0239	.1449	.1402	.938
WLS	.0021	.1092	.1097	.955	-.0072	.1266	.1213	.944	.0212	.1344	.1330	.954
Assuming Mixed correlation structure												
QL	.0016	.1081	.1086	.955	-.0068	.1185	.1137	.944	.0184	.1246	.1216	.953
MGV	.0022	.1086	.1098	.957	-.0079	.1266	.1214	.947	.0211	.1400	.1405	.954
QIF	.0023	.1105	.1090	.948	-.0103	.1297	.1226	.939	.0223	.1410	.1370	.942
WLS	.0021	.1092	.1097	.955	-.0072	.1267	.1213	.944	.0213	.1350	.1333	.953
Assuming ARMA correlation structure(Misspecification)												
QL	.0018	.1078	.1088	.957	-.0069	.1207	.1159	.942	.0200	.1294	.1253	.951
MGV	.0022	.1086	.1101	.957	-.0083	.1329	.1275	.948	.0229	.1509	.1472	.947
QIF	.0041	.1139	.1127	.947	-.0121	.1421	.1359	.940	.0235	.1646	.1573	.951
WLS	.0021	.1091	.1098	.953	-.0072	.1266	.1216	.946	.0218	.1375	.1359	.953

Table 21: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for $n = 400, h = 2.9$ based on 1000 simulations under Exchange model with weak correlation $\rho = 0.2$.

	$\rho = 0.2$											
	$\beta = 0.1$				$\theta_1 = 1$				$\theta_2 = 0.5$			
Method	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
Working Independence												
WI	.0018	.1426	.1350	.936	-.0047	.2482	.2315	.936	.0306	.2598	.2496	.940
Assuming Exchangeable correlation structure(True)												
QL	.0004	.1359	.1289	.937	.0013	.2213	.2108	.945	.0276	.2477	.2214	.953
MGV	.0005	.1361	.1290	.936	.0020	.2242	.2118	.941	.0322	.2690	.2250	.951
QIF	.0012	.1375	.1286	.934	-.0005	.2290	.2117	.937	.0273	.2442	.2237	.944
WLS	.0012	.1370	.1298	.936	-.0011	.2279	.2149	.938	.0265	.2405	.2266	.945
Assuming Mixed correlation structure												
QL	.0004	.1359	.1288	.936	.0017	.2213	.2107	.944	.0280	.2482	.2215	.952
MGV	-.0001	.1361	.1296	.938	.0056	.2234	.2158	.948	.0366	.2693	.2303	.952
QIF	.0012	.1379	.1279	.932	-.0014	.2302	.2101	.932	.0266	.2464	.2218	.940
WLS	.0012	.1370	.1298	.935	-.0009	.2279	.2148	.938	.0268	.2409	.2268	.944
Assuming ARMA correlation structure(Misspecification)												
QL	.0004	.1357	.1289	.936	.0022	.2211	.2112	.947	.0288	.2467	.2225	.946
MGV	-.0012	.1390	.1330	.939	.0087	.2459	.2428	.947	.0570	.3450	.2573	.963
QIF	.0008	.1411	.1326	.931	-.0017	.2396	.2230	.936	.0308	.2568	.2404	.940
WLS	.0011	.1369	.1299	.936	-.0007	.2277	.2150	.939	.0272	.2416	.2277	.943

Table 22: Estimation of correlation coefficients on 1000 simulations for Exchange model by different approaches.

Method	Assuming Mixed Structure	Assuming Exchangeable Structure(true)
	$\rho = 0.8$	
QL	$0.012 \times 0.417^{-\lvert s-t \rvert} + 0.791$	0.796
MGV	$0.303 \times 0.607^{-\lvert s-t \rvert} + 0.050$	0.100
WLS	$0.020 \times 0.299^{-\lvert s-t \rvert} + 0.354$	0.362
	$\rho = 0.5$	
QL	$0.023 \times 0.348^{-\lvert s-t \rvert} + 0.484$	0.496
MGV	$0.288 \times 0.605^{-\lvert s-t \rvert} + 0.050$	0.100
WLS	$0.014 \times 0.361^{-\lvert s-t \rvert} + 0.176$	0.192
	$\rho = 0.2$	
QL	$0.014 \times 0.107^{-\lvert s-t \rvert} + 0.180$	0.196
MGV	$0.290 \times 0.606^{-\lvert s-t \rvert} + 0.051$	0.195
WLS	$0.007 \times 0.061^{-\lvert s-t \rvert} + 0.075$	0.079

CHAPTER 4: REAL DATA APPLICATIONS

4.1 DATA EXAMPLE 1: Application to ACTG 244 Trial

AIDS Clinical Trials Group(ACTG) protocol 244, a randomized double-blinded trial, aims to evaluate the effects of randomizing HIV infected patients to combined drug therapy based on development of drug resistant mutations. HIV infected patients who received a Zidovudine (ZDV) monotherapy were enrolled in this study. Zidovudine is demonstrated as an effective drug to control the clinical progress of disease in HIV infected patients (Montaner et al., 1998). However after taking the drug, some patients may develop Zidovudine resistance (ZDV^R), which can be detected by monitoring the ZDV^R mutation T215Y/F in HIV reverse transcriptase. Patients' CD4 cell counts and $\log_{10}$plasma HIV RNA were measured and T215Y/F mutation are monitored from patients' plasma bimonthly since study entry. The visit dates varied across individuals. Upon detection of the mutation in some patient's plasma, he or she was randomized to continue ZDV, or add another drug didanosine (ddI) or add ddI plus nevirapine (NVP). Patients' demographic information were collected at the study entry.

289 HIV infected patients were enrolled in this study, among which 284 were dispensed Zidovudine (ZDV). 57 of them developed T215Y/F mutation during the study and 234 of them were not detected of the mutation. Among the 57 who developed

mutation, 8 were off the study before the randomization, among the rest 49 patients, 17, 15 and 17 of them are randomized to ZDV, $ZDV+ddI$ and $ZDV+ddI+NVP$ respectively. Among those 234 patients who didn't develop the mutation during the trial, 97 were off treatment before the interim interview, while for the rest 137 patients, 69 and 68 were randomized to $ZDV+ddI$ and $ZDV+ddI+NVP$ respectively after the interim interview.

To see whether switching from ZDV monotherapy to combined therapy $ZDV+ddI$ or $ZDV+ddI+NVP$ based on ZDV^{R} mutation T215Y/F will alter the deterioration of the disease, we apply a varying coefficient model to study the treatment switching effects. We choose the CD4 cell counts as our endpoint, which is an independent predictor of AIDS/death (Mocroft et al., 2003).

4.1.1 Analysis of the effects of switching treatments *after* drug-resistant mutation was detected

First, we examine the effects of switching treatments following detection of the T215Y/F mutation. After preliminary exploration of the data, we propose the following model for each subject i:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \gamma_1(U_{1i}(t), \theta_{\mathbf{1}}) T_{A1i}(t) \\ &+ \gamma_2(U_{1i}(t), \theta_{\mathbf{2}}) T_{A2i}(t) + \gamma_3(U_{1i}(t), \theta_{\mathbf{3}}) T_{A3i}(t) + \epsilon_i(t) \end{aligned} \tag{4.1}$$

for $t \in [0, \tau]$, where $\tau = 2.5$ years. Let $Y_i(t)$ be the square root of CD4 cell counts at t years since study entry for subject i, Z_{1i} be sex of subject i (1 if Female; 0 if Male), Z_{2i} be age in years at study entry of subject i, Z_{3i} and Z_{4i} be dummy variables

coding race ($Z_{3i} = 1$ if white and 0 otherwise, $Z_{4i} = 1$ if black and 0 otherwise). Let $U_{1i}(t) = t - S_{1i}$ be the time elapsed from the treatment randomization after T215Y/F mutation was detected, where S_{1i} is the treatment switching time, i.e., the time from study entry to the first randomization based on detection of the T215Y/F mutation. We set $S_{1i} = 3$ years which is longer than the study duration for the 234 subjects who did not experience mutation during the study. The length of the range of the observed values for $U_{1i}(t)$, $t \in [0, 2.5]$, is 2.25. Let $T_{A1i}(t) = 1$ if $t > S_{1i}$ and randomized to ZDV and 0 otherwise, $T_{A2i}(t) = 1$ if $t > S_{1i}$ and randomized to ZDV+ddI and 0 otherwise, and $T_{A3i}(t) = 1$ if $t > S_{1i}$ and randomized to ZDV+ddI+NVP and 0 otherwise; note that all the three treatment indicators are zero prior to detection of the mutation. All $n = 284$ enrolled subjects dispensed ZDV monotherapy were studied in this analysis. The eight subjects who were off treatment prior to the first randomization as well as the 90 subjects who were off treatment prior to the interim review were censored at the time of drop-off; the 137 subjects who did not develop the mutation and were randomized at the interim review were censored at the time of the second randomization.

We assume that $\gamma_k(u, \theta_\mathbf{k})$, $k = 1, 2, 3$, are the second order polynomial functions. Let $\gamma_1(u, \theta_\mathbf{1}) = \theta_{10} + \theta_{11}u + \theta_{12}u^2$, $\gamma_2(u, \theta_\mathbf{2}) = \theta_{20} + \theta_{21}u + \theta_{22}u^2$ and $\gamma_3(u, \theta_\mathbf{3}) = \theta_{30} + \theta_{31}u + \theta_{32}u^2$, where $\theta_\mathbf{1} = (\theta_{10}, \theta_{11}, \theta_{12})$, $\theta_\mathbf{2} = (\theta_{20}, \theta_{21}, \theta_{22})$ and $\theta_\mathbf{3} = (\theta_{30}, \theta_{31}, \theta_{32})$. The 3-fold cross-validation method for bandwidth selection yields $h = 0.41$ while the bandwidth formula $h = C\hat{\sigma}_T n^{-1/3}$ yields $h = 0.36$ for $C = 4$ $h = 0.45$ for $C = 5$. The results of analysis are hardly affected by the bandwidth choice between 0.36 and 0.47.

The parameter estimates of β_1, β_2, β_3 and β_4 by different approaches are presented in Table 23. Since we don't know the true correlation structure, we assumed a mixed correlation structure when using WLS, QL and MGV methods. As the number of parameters in above model is totally 13, which is too large to achieve convergence when solving the quadratic inference functions in QIF approach, we didn't report the result of QIF approach here. The estimates of varying-coefficient functions $\alpha(t)$, $\gamma_1(u, \theta_\mathbf{1})$, $\gamma_2(u, \theta_\mathbf{2})$ and $\gamma_3(u, \theta_\mathbf{3})$ are presented in Figure 7. We also list the confidence intervals for each method in Figure 8, Figure 9, Figure 10 and Figure 11.

From Table 23 under model (4.1) where switching treatment after mutation was detected(shown in the left column), we can see that none of the estimated parameters are significant when using WI method, but all of them become significant at level 0.01 when using methods considering within subject correlation. Females tend to have less CD4 counts, and younger people have slightly less CD4 counts. People of black or white race tend to have less CD4 counts than other race of people.

From Figure 7, Figure 8, Figure 9, Figure 10 and Figure 11, baseline level of CD4 counts by all the methods show similar downward trend and switching to the combination therapies show a less decreasing trend than the ZDV monotherapy. We notice that when switching to ZDV+ddI+NVP, CD4 counts increase until 1.5 years after switching by all the methods except WI approach. Thus switching to the combination therapies from ZDV monotherapy have benefits for HIV patients even after they developed the drug-resistant mutation.

Since we have demonstrated in the simulation study that QL approach always target the true correlation coefficients and have a better performance than MGV and

WLS approaches. We only list the result from QL approach. When assuming a mixed correlation structure by QL approach, the correlation coefficient $\rho(s,t)$ between time s and t is $\rho(s,t) = 0.4987 \times 0.5828^{-|s-t|} + 0.2656$ for $s \neq t$ and $\rho(s,t) = 1$ for $s = t$. As the distance of time points s and t increases, the correlation of responses at these time points within the same subject decreases.

4.1.2 Analysis of the effects of switching treatments *before* drug-resistant mutation was detected

Next, we examine the effects of switching treatments *before* drug-resistant mutation was detected. In September 1996, the Data Safety Monitoring Board reviewed the data in this study independently. After this interim review, all the 234 subjects who were not detected mutation were offered randomization to the ZDV+ddI or ZDV+ddI+NVP arms with six months of additional follow-up. This section we focus on this group of people. We exclude the subjects who developed the T215/F mutation here as the time to develop T215/F mutation likely introduces dependent censoring. 90 out of the 234 patients were off the treatment prior to the interim review, whose censoring time is the time of dropping-off. Similarly as above analysis, we used a varying coefficient model:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} \\ &+ \gamma_2(U_{2i}(t), \boldsymbol{\theta_2}) T_{B2i}(t) + \gamma_3(U_{2i}(t), \boldsymbol{\theta_3}) T_{B3i}(t) + \epsilon_i(t), \end{aligned} \tag{4.2}$$

for $t \in [0, 2.5]$. $Y_i(t)$, Z_{1i}, Z_{2i}, Z_{3i} and Z_{4i} are defined the same as in Model (4.1). $U_{2i}(t) = t - S_{2i}$, where S_{2i} is the second randomization time after interim review

for subject i. The range of observed values of $U_{2i}(t)$, $t \in [0, 2.5]$, is $[0, 0.70]$. Let $T_{B2i}(t) = 1$ if $t > S_{2i}$ and randomized to ZDV+ddI and 0 otherwise. Let $T_{B3i}(t) = 1$ if $t > S_{2i}$ and randomized to ZDV+ddI+NVP and 0 otherwise. $T_{B2i}(t) = 0$ and $T_{B3i}(t) = 0$ indicates a subject is on ZDV monotherapy at time t after the interim review. Also we use the second order polynomial functions for $\gamma_2(u, \theta_\mathbf{2})$ and $\gamma_3(u, \theta_\mathbf{3})$.

From Table 23 under model (4.2) where switching treatment before mutation was detected(shown in the right column), we get similar results of β_1, β_2 and β_3 as those under model (4.1) where switching treatment after mutation was detected. None of the estimated parameters are significant when using WI method, but all of them except β_1 become significant at level 0.01 when using QL and MGV methods. All the estimated coefficients of parameters have the same sign no matter before or after mutation, that is, females tend to have lower CD4 level, and younger people have a slightly lower CD4 level. People of black or white race tend to have lower CD4 level than other race of people.

Figure 12 lists the results of estimated baseline level $\alpha_0(t)$, estimated $\gamma_2(u, \theta_\mathbf{2})$ and $\gamma_3(u, \theta_\mathbf{3})$ by all methods in the same plot. We also list the confidence intervals of these results by WI, QL, MGV and WLS approaches in Figure 13, Figure 14, Figure 15 and Figure 16 respectively. We can see that baseline level of CD4 counts of all patients show similar downward trend until it bumps up a little at 1.5 years since study entry by all the methods. The estimated switching-treatment effects are above the horizontal zero line by all methods, suggesting that combination therapy improve CD4 counts for patients who have not yet developed the T215/F drug resistance mutation.

From the estimation of covariance function by QL method assuming a mixed correlation structure, the correlation coefficient $\rho(s,t)$ between time s and t is $\rho(s,t) = 0.3440 \times 0.3907^{-|s-t|} + 0.3884$ for $s \neq t$ and $\rho(s,t) = 1$ for $s = t$. As the distance of time points s and t increases, the correlation of responses at these time points within the same subject decreases.

4.2 DATA EXAMPLE 2: STEP Study with MITT Cases

In HIV vaccine efficacy trials, as soon as the patients were diagnosed with Ab+, they may start antiretroviral treatments and their longitudinal biomarkers, e.g., viral loads and CD4 counts in their blood samples are monitored regularly. However instead of assessing the effects from the time point that Ab+ were diagnosed for these patients, we may be more interest in studying the vaccine effects from the time when they actually became HIV infected. Hence two time scales are involved, one is the time from diagnosis of Ab+, the other is the time from actual HIV acquisition. The varying-coefficient model proposed above can be used to solve such two-time-scale problems and help us to understand the treatment effects or vaccine effects on patients' disease progress. Using advanced PCR test can approximate well the actual HIV acquisition time for patients shown Ab+ (Piatak Jr et al., 1993).

The MRKAd5 HIV-1 gag/pol/nef vaccine, which elicits T cell immunity, aims to control the replication of Human immunodeficiency virus (HIV) among the participants who got HIV-infected after vaccination (Gray et al., 2011). STEP study (cf. Buchbinder et al. (2008); McElrath et al. (2008)), a multi-center, double-blinded, randomized phase II clinical trial, was to determine whether the MRKAd5 HIV-1

gag/pol/nef vaccine is capable to fulfill this goal and how the effects evolve with time.

This study started in December 2004. 3000 HIV-1 negative participants with high risk of HIV-infection aged from 18 to 45 were enrolled in this study. They are from 34 sites selected in North America, the Caribbean, South America, and Australia and were randomized to receive vaccine or placebo in a ratio 1:1, stratified by sex, study site and adenovirus type 5 (Ad5) antibody titer at baseline. Some of the participants were fully adherent to vaccinations while others were not.

MITT cases are the modified intention-to-treat subjects who became HIV infected during this trial. The modified intention-to-treat refers to all randomized subjects, excluding the few that were found to be HIV infected at study entry. Our analysis focuses on all the 174 MITT cases as of September 22, 2009. Since there are only 15 females that are less than 10% of the sample, we exclude these female patients to avoid the sex effect. Thus we have totally 159 HIV-infected males in our analysis. For each participant, we have the records of their first positive diagnosis date and estimated infection date. The first positive diagnosis date was the date their first positive Elisa was confirmed by Western Blot or RNA, and the estimated infection date was determined by the dates of their first positive PCR test.

After the first positive diagnosis for each of 159 HIV-infected patients, he or she would be scheduled 18 post-infection visits, that is, weeks 0, 1, 2, 8, 12, 26, and every 26 weeks thereafter through week 338. However, the actual visiting dates and time may vary due to each individual. We define a patient is right censored if he or she started the antiretroviral therapy(ART) or was censored traditionally such as went

off the study or reached the end of study. The right censoring time is the time from the first positive diagnosis date to right censoring. We have the measurements of HIV virus load and CD4 cell counts for each patient before the right censoring time.

Among the 791 visits from all these 159 males, 156 were missing in CD4 cell counts and 5 were missing in HIV virus load and there are no missing data in CD4 and virus load simultaneously. We used a linear regression method to impute these missing values. Firstly we built a linear regression model based on complete data of $\log_{10}$(viral load) and square root of CD4 count for each time point. Secondly, we used the built model to predict the missing viral load value from corresponding CD4 cell counts at one time point or predict the missing CD4 count from virus load value. However, there are no complete data at three time points and only one complete data at two other time points where fail to conduct the linear regression. At another time point, the predicted value of virus load was far beyond the normal range. Thus we delete the observations at these six time points and conduct our data analysis based on the rest 785 observations for 159 subjects. 97 of them were randomized to vaccine and 62 were in the placebo group. 122 are in the sites of North America or Australia, and 37 reside in other sites in Caribbean or South America.

In order to see how vaccine affects HIV virus load and CD4 counts over time since the actual HIV acquisition date, we built two varying coefficient models as model (4.3) and model (4.4) for HIV virus load and CD4 counts respectively. There are two time scales in these models. $T_i(t)$ is the time from the first positive diagnosis date to the jth visit for ith subject. The time elapsed from estimated infection date is denoted by $U_i(t) = T_i(t) + O_i$, where O_i is the gap between estimated infection date and the

first positive diagnosis date. Time is measured in years. The right censoring rate of $T_i(t)$ is 69.81%. We choose $\tau = 2.5$ since there are few data after time point 2.5. The range for observed $U_i(t)$, $t \in [0, 2.5]$, is $[0, 3.0]$. Let $Y_{1i}(t)$ and $Y_{2i}(t)$ be the common logarithm of HIV virus load and the square root of CD4 counts for ith subject at time t respectively. We choose the same covariates for both models. Let $X_{1i}(t)$ be the natural logarithm of Ad5 (adenovirus type 5 antibody titer at baseline), $X_{2i}(t)$ be the site indicator (1 if North America or Australia; 0 otherwise), $X_{3i}(t)$ be the pre-protocol indicator (1 if the subject was fully adherent to vaccinations; 0 otherwise) and $X_{4i}(t)$ be the treatment indicator (1 if the subject received vaccine; 0 if receiving placebo).

After preliminary exploration of the data, X_1, X_2 and X_3 show no evidence of varying coefficients. We propose the following models: Virus Load model

$$Y_{1i}(t) = \alpha(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + \gamma(U_i(t))X_{4i}(t) + \epsilon_i(t), \tag{4.3}$$

and CD4 model

$$Y_{2i}(t) = \alpha(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + \gamma(U_i(t))X_{4i}(t) + \epsilon_i(t). \tag{4.4}$$

where

$$\gamma(u) = \theta_0 + \theta_1 u + \theta_2 u^2.$$

By the empirical bandwidth formula a possible reasonable choice of the bandwidth for this data set is 0.35 for both model (4.3) and model (4.4). When we ignore the correlation among observations within subjects using Working Independence(WI) approach, the estimates of β_1, β_2 and β_3 are shown in Table 24 for Virus Load model

(4.3) and in Table 25 for CD4 model (4.4). We can see that none of them are significant at 5% level except β_2 in CD4 model (4.4). Thus there are no significant effects of baseline Ad5 titer, study sites or the pre-protocol on the HIV viral load level, and only study sites have significant effect on the CD4 counts by WI approach.

However when we incorporate the correlation within subjects using QL, MGV, QIF and WLS approaches by assuming a mixed correlation structure(since we don't know the true correlation structure), some of the estimates of parameters β_1, β_2 and β_3 become significant as shown in Table (4.3) and Table (4.4). In Virus Load model $\hat{\beta}_1$ becomes significantly negative by QL and MGV methods at level 0.05, while $\hat{\beta}_2$ is significantly negative when using QL, MGV and QIF methods at level 0.05, which means a patient with higher Ad5 antibody titer level at study entry will have a lower virus load and one lives in North America or Australia will have a lower virus load than those reside in Caribbean or South America. In CD4 model, $\hat{\beta}_1$ is significantly negative by QL, MGV and WLS methods at level 0.05, $\hat{\beta}_2$ is significantly negative for all methods and $\hat{\beta}_3$ is significantly positive by QL and QIF methods at level 0.05. This means that a person with a higher Ad5 antibody level at study entry tends to have a lower CD4 level, which is counterintuitive. People in North America or Australia tend to have a higher CD4 counts and those adherent to vaccination will have a higher CD4 level, which is consistent with the results from Virus Load model.

The estimated results of $\alpha(t)$ and $\gamma(u)$ by all approaches are illustrated in Figure (17) and Figure (23) for Virus Load model and CD4 model respectively. We also plot $\hat{\alpha}(t)$ and $\hat{\gamma}(u)$ together with their 95% pointwise confidence intervals by each approach in Figure 18, Figure 19, Figure 20, Figure 21 and Figure 22 for WI, QL,

MGV, QIF and WLS approach respectively for Virus Load model (4.3). Similarly Figure 24, Figure 25, Figure 26, Figure 27 and Figure 28 are for WI, QL, MGV, QIF and WLS respectively under CD4 model (4.4).

From Figure (17), Figure 18, Figure 19, Figure 20, Figure 21 and Figure 22, except QL method, baseline function $\hat{\alpha}_0(t)$ doesn't show much trend and $\hat{\gamma}(u)$ is below the zero horizontal line for all other methods under Virus Load model, which implies that controlling the effects from other covariates in the model, patients' virus load didn't change much with time while the vaccine have a negative effect on virus load since actual HIV infection.

From Figure (23), Figure 24, Figure 25, Figure 26, Figure 27 and Figure 28, baseline function $\hat{\alpha}(t)$ shows a downward trend for all methods but bounced up a little at the second year since first positive diagnosis, and $\hat{\gamma}(u)$ goes upward from negative before the second year to be postive after the second year since actual infection for CD4 model, which indicates excluding the effects from other covariates in the model, patients' CD4 level was decreasing with time and the vaccine effect changes over time since actual infection and improves the CD4 counts at a later time.

When using QL method by assuming a mixed correlation structure, the estimated covariance function under Virus Load model is $\rho(s,t) = 0.8498 \times 0.3701^{-|s-t|} + 0.0567$ for $s \neq t$ and $\rho(s,t) = 1$ for $s = t$, while the estimated covariance function under CD4 model is $\rho(s,t) = 0.5481 \times 0.8189^{-|s-t|}$ for $s \neq t$ and $\rho(s,t) = 1$ for $s = t$. As the distance of time points s and t increases, the correlation of responses at these time points within the same subject decreases. The correlation structure is close to an ARMA(1,1) process.

Table 23: Point estimates of β_1, β_2, β_3 and β_4 on the ACTG 244 data under model (4.1) and model (4.2), where switching treatments *after* and *before* drug-resistant mutation was detected respectively by different approaches.

	β_1 (Sex:1 if female)					
	after mutation			*before* mutation		
Method	Estimate	SD	p-value	Estimate	SD	p-value
WI	-1.3155	0.7216	0.0683	-0.6719	0.6488	0.3004
QL	-2.4615	0.3989	< 0.0001	-1.1806	0.5142	0.0217
MGV	-1.6663	0.4147	0.0001	-0.9666	0.5352	0.0709
WLS	-1.4344	0.4694	0.0022	-0.8623	0.5531	0.1190
	β_2 (Age)					
	after mutation			*before* mutation		
Method	Estimate	SD	p-value	Estimate	SD	p-value
WI	0.0619	0.0284	0.0292	0.0203	0.0251	0.4186
QL	0.0885	0.0117	< 0.0001	0.0434	0.0145	0.0027
MGV	0.0765	0.0123	< 0.0001	0.0430	0.0149	0.0038
WLS	0.0556	0.0147	0.0002	0.0313	0.0170	0.0645
	β_3 (Race:1 if White)					
	after mutation			*before* mutation		
Method	Estimate	SD	p-value	Estimate	SD	p-value
WI	-0.7575	0.7746	0.3281	-1.0888	0.7227	0.1319
QL	-4.2366	0.3309	< 0.0001	-2.8251	0.4132	< 0.0001
MGV	-2.5806	0.3245	< 0.0001	-2.0906	0.4084	< 0.0001
WLS	-1.3100	0.3814	0.0006	-1.2749	0.4521	0.0048
	β_4 (Race:1 if Black)					
	after mutation			*before* mutation		
Method	Estimate	SD	p-value	Estimate	SD	p-value
WI	-0.8163	0.8565	0.3405	-1.7619	0.7781	0.0235
QL	-1.9153	0.4111	< 0.0001	-2.6880	0.4994	< 0.0001
MGV	-1.7389	0.4015	< 0.0001	-2.5629	0.4881	< 0.0001
WLS	-1.0980	0.4548	0.0158	-1.8945	0.5208	0.0003

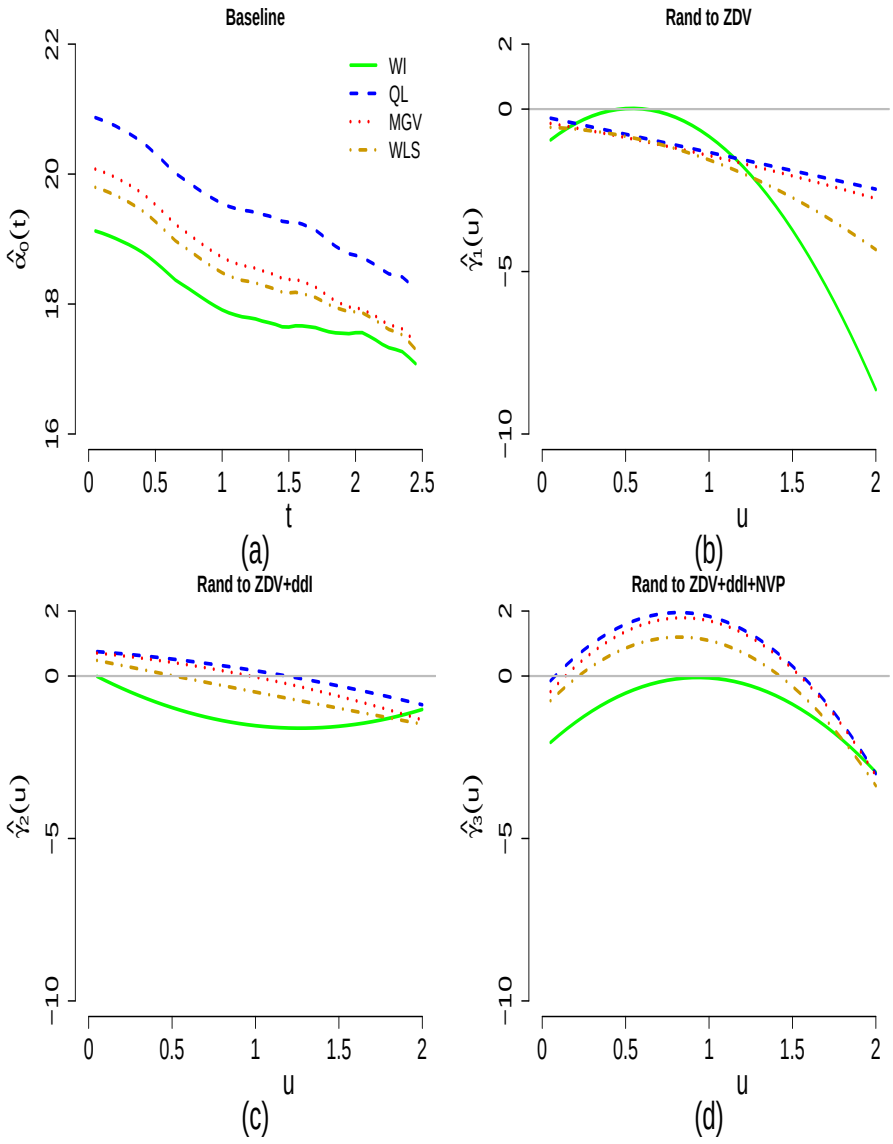

Figure 7: Estimates of baseline and treatment effects *after* drug-resistant mutation was detected based on the ACTG 244 data under model (4.1) by different approaches. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by different approaches; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively by different approaches.

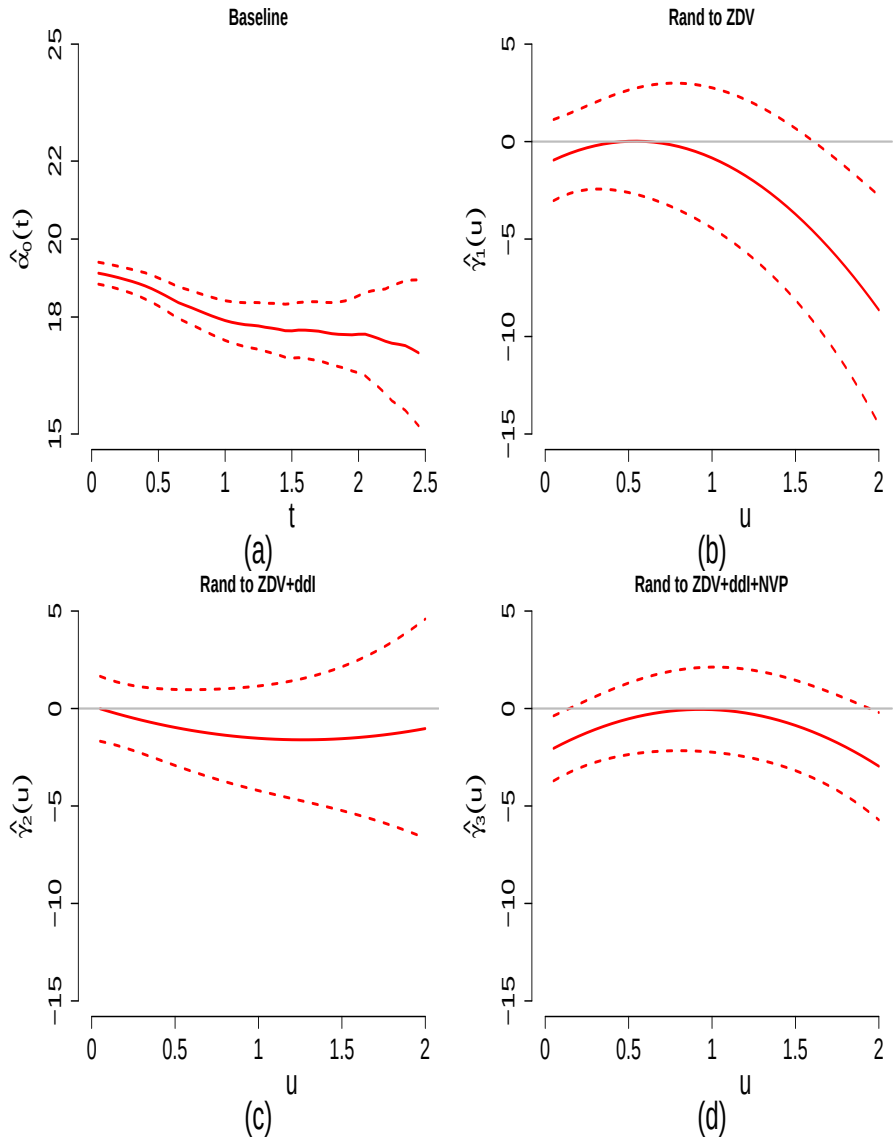

Figure 8: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *after* drug-resistant mutation was detected based on the ACTG 244 data under model (4.1) by WI approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by WI approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively by WI approach.

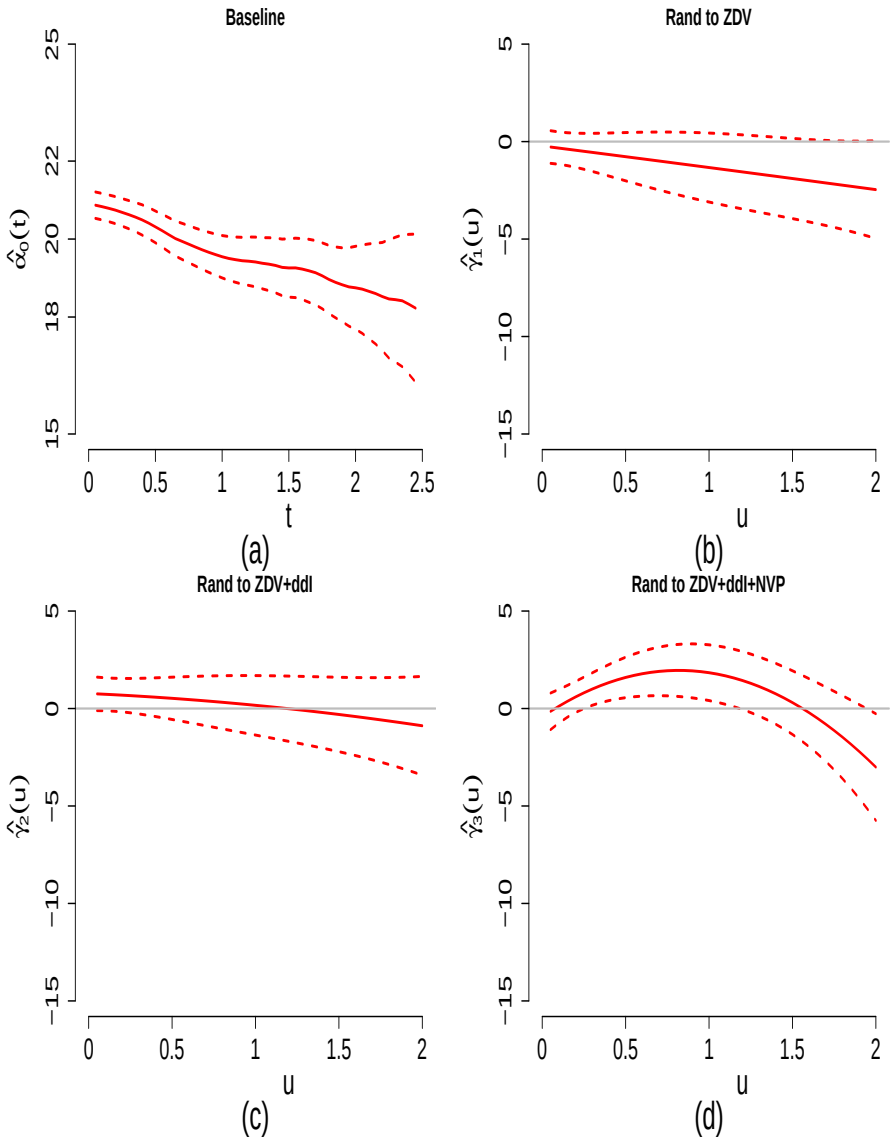

Figure 9: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *after* drug-resistant mutation was detected based on the ACTG 244 data under model (4.1) by QL approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by QL approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively by QL approach.

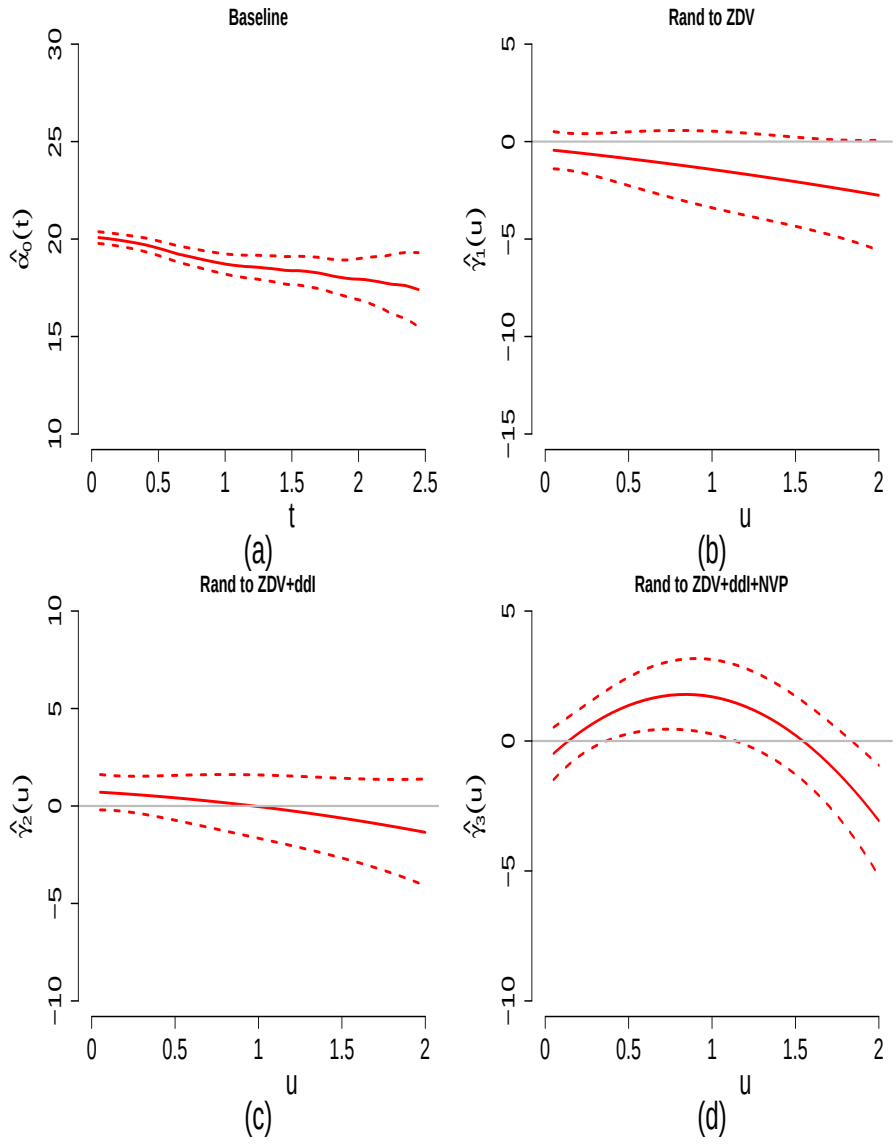

Figure 10: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *after* drug-resistant mutation was detected based on the ACTG 244 data under model (4.1) by MGV approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by MGV approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively by MGV approach.

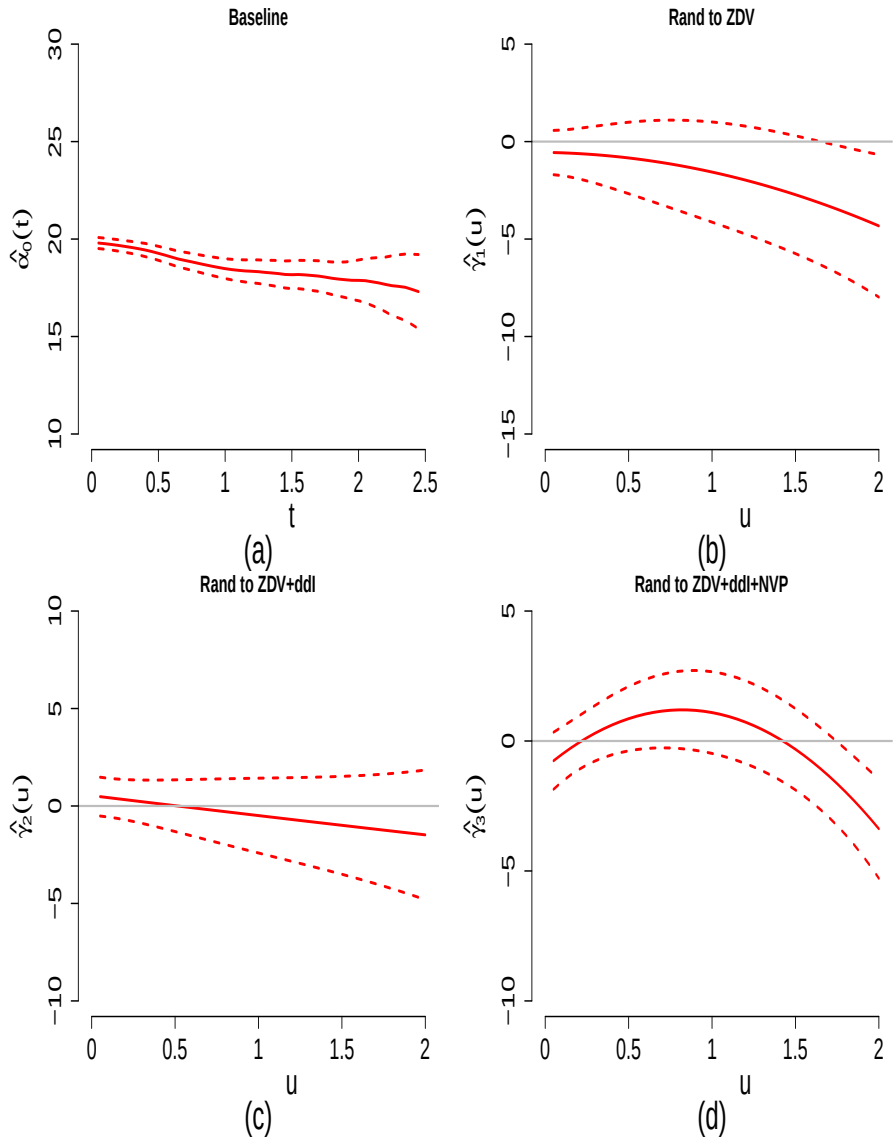

Figure 11: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *after* drug-resistant mutation was detected based on the ACTG 244 data under model (4.1) by WLS approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by WLS approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 1, 2, 3$, respectively by WLS approach.

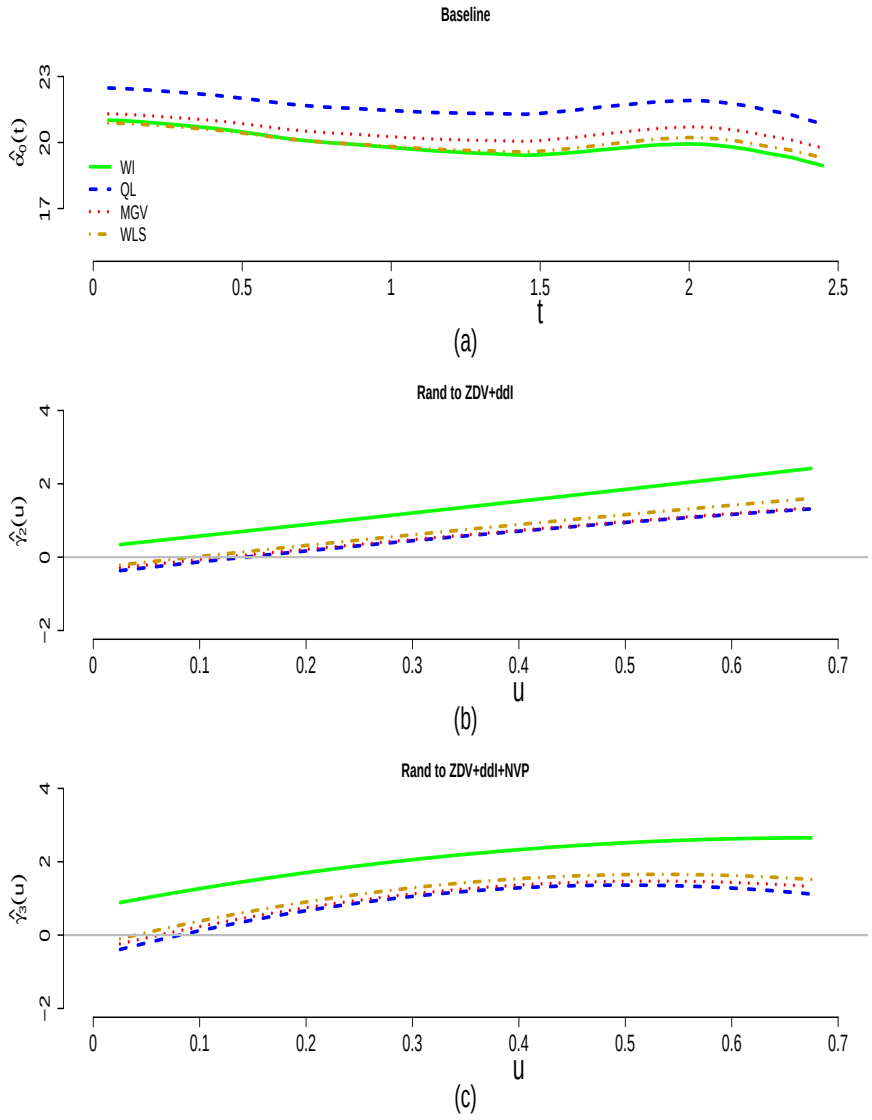

Figure 12: Estimates of baseline and treatment effects *before* drug-resistant mutation was detected based on the ACTG 244 data under model (4.2) by different approaches. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by different approaches; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 2, 3$, respectively by different approaches.

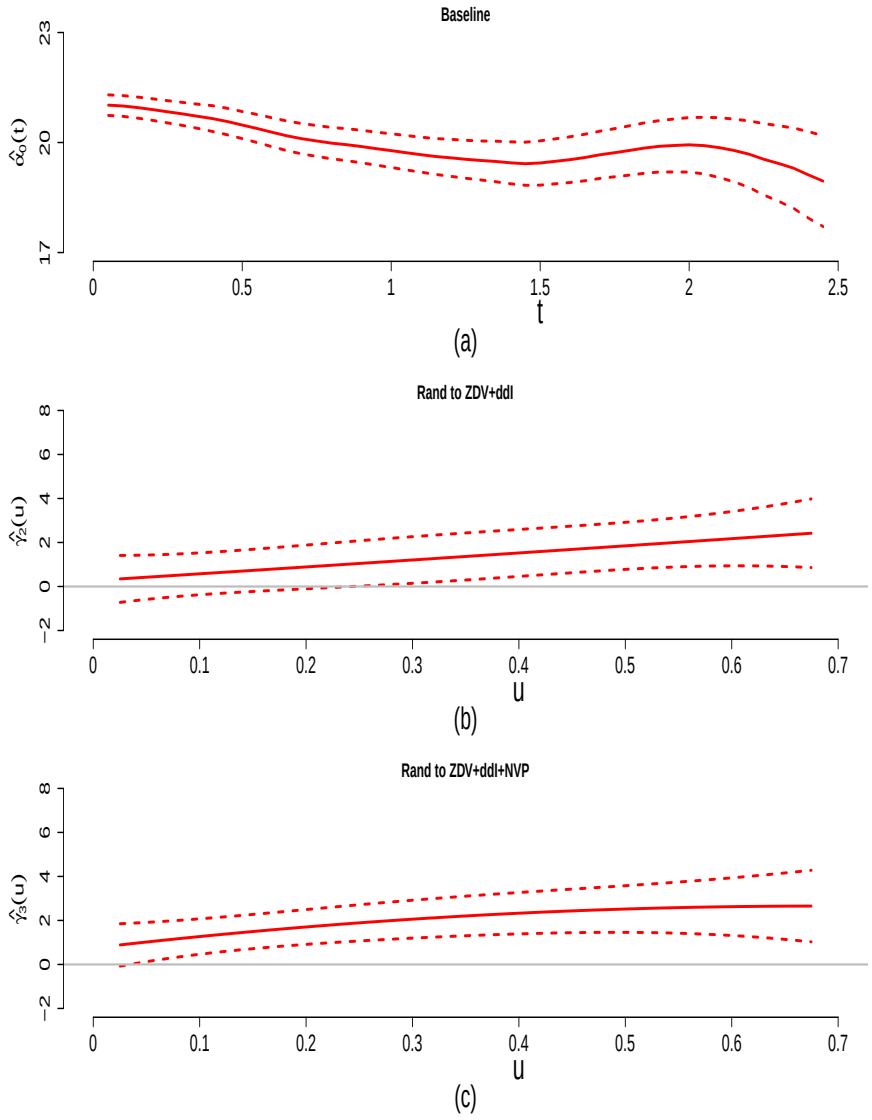

Figure 13: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *before* drug-resistant mutation was detected based on the ACTG 244 data under model (4.2) by WI approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by WI approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 2, 3$, respectively by WI approach.

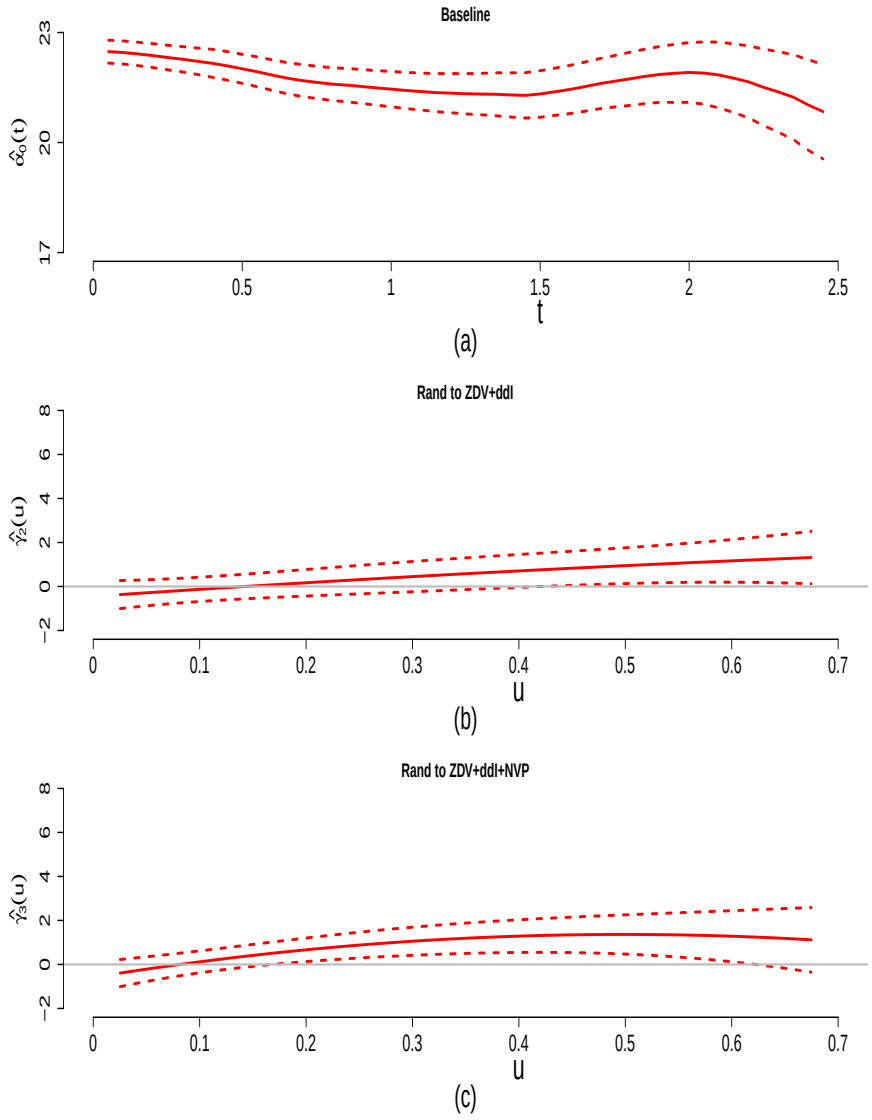

Figure 14: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *before* drug-resistant mutation was detected based on the ACTG 244 data under model (4.2) by QL approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by QL approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 2, 3$, respectively by QL approach.

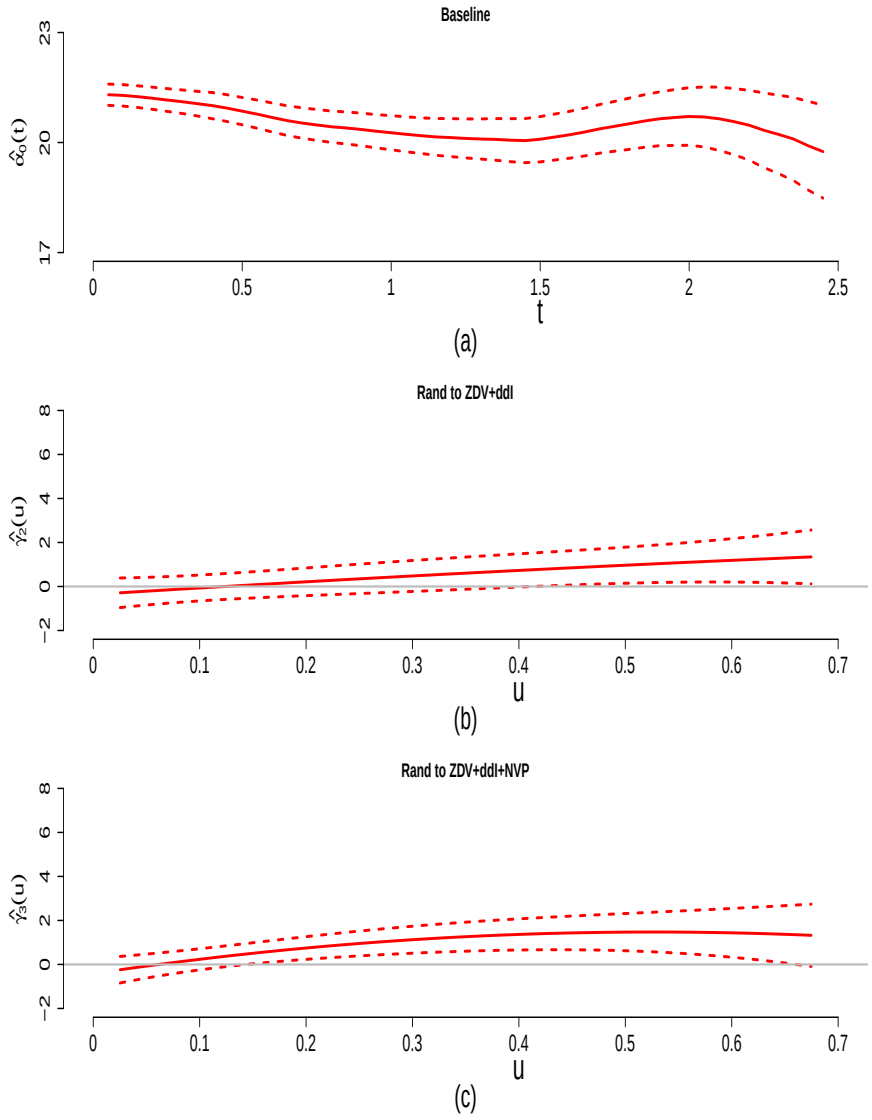

Figure 15: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *before* drug-resistant mutation was detected based on the ACTG 244 data under model (4.2) by MGV approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by MGV approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 2, 3$, respectively by MGV approach.

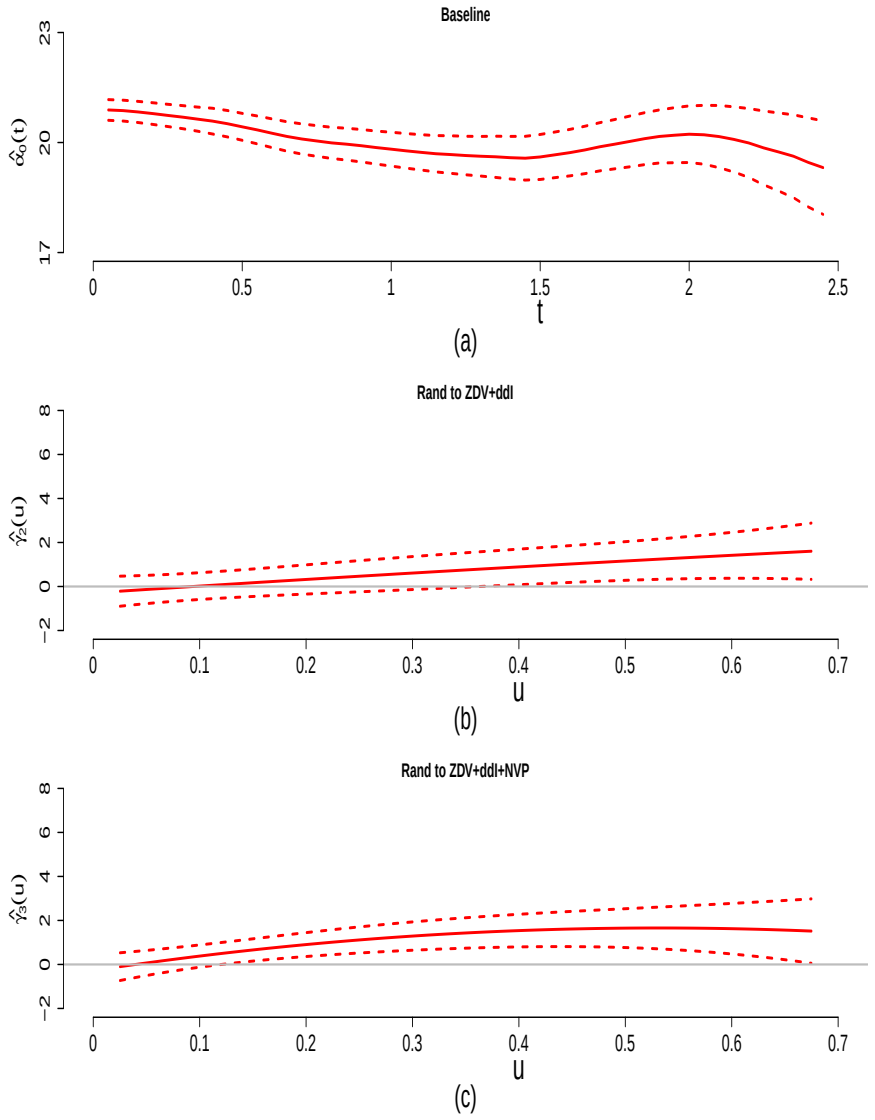

Figure 16: Estimates of baseline and treatment effects with 95% pointwise confidence intervals *before* drug-resistant mutation was detected based on the ACTG 244 data under model (4.2) by WLS approach. (a) is the estimated baseline $\hat{\alpha}_0(t)$ using $h = 0.41$ by WLS approach; (b) and (c) are the estimates of $\gamma_k(u)$, $k = 2, 3$, respectively by WLS approach.

Table 24: Point estimates of β_1, β_2 and β_3 based on STEP data under Virus Load model (4.3) by different approaches.

β_1 (Ad5 antibody level)			
Method	Estimate	SD	p-value
WI	0.0105	0.0436	0.8090
QL	-0.0701	0.0264	0.0078
MGV	-0.0668	0.0302	0.0267
QIF	-0.0239	0.0302	0.4299
WLS	-0.0270	0.0315	0.3903
β_2 (1 if in North America or Australia)			
Method	Estimate	SD	p-value
WI	-0.1405	0.1645	0.3929
QL	-0.5863	0.1026	< 0.0001
MGV	-0.4201	0.1109	0.0002
QIF	-0.2569	0.1062	0.0156
WLS	-0.2288	0.1201	0.0568
β_3 (1 if adherent to vaccination)			
Method	Estimate	SD	p-value
WI	-0.0888	0.1938	0.6469
QL	-0.1811	0.1144	0.1135
MGV	-0.2931	0.1388	0.0347
QIF	-0.1351	0.1299	0.2985
WLS	-0.1841	0.1338	0.1687

Table 25: Point estimates of β_1, β_2 and β_3 based on STEP data under CD4 model (4.4) by different approaches.

β_1 (Ad5 antibody level)			
Method	Estimate	SD	p-value
WI	-0.2179	0.1873	0.2446
QL	-0.4004	0.1219	0.0010
MGV	-0.3344	0.1274	0.0087
QIF	-0.0715	0.1270	0.5731
WLS	-0.2940	0.1396	0.0352
β_2 (1 if in North America or Australia)			
Method	Estimate	SD	p-value
WI	3.3128	0.7816	< 0.0001
QL	5.5436	0.5270	< 0.0001
MGV	4.6546	0.5457	< 0.0001
QIF	2.2328	0.4887	< 0.0001
WLS	3.9258	0.5914	< 0.0001
β_3 (1 if adherent to vaccination)			
Method	Estimate	SD	p-value
WI	-0.0682	0.8579	0.9367
QL	1.4325	0.5401	0.0080
MGV	0.9835	0.5545	0.0761
QIF	1.4971	0.4927	0.0024
WLS	0.4815	0.5943	0.4178

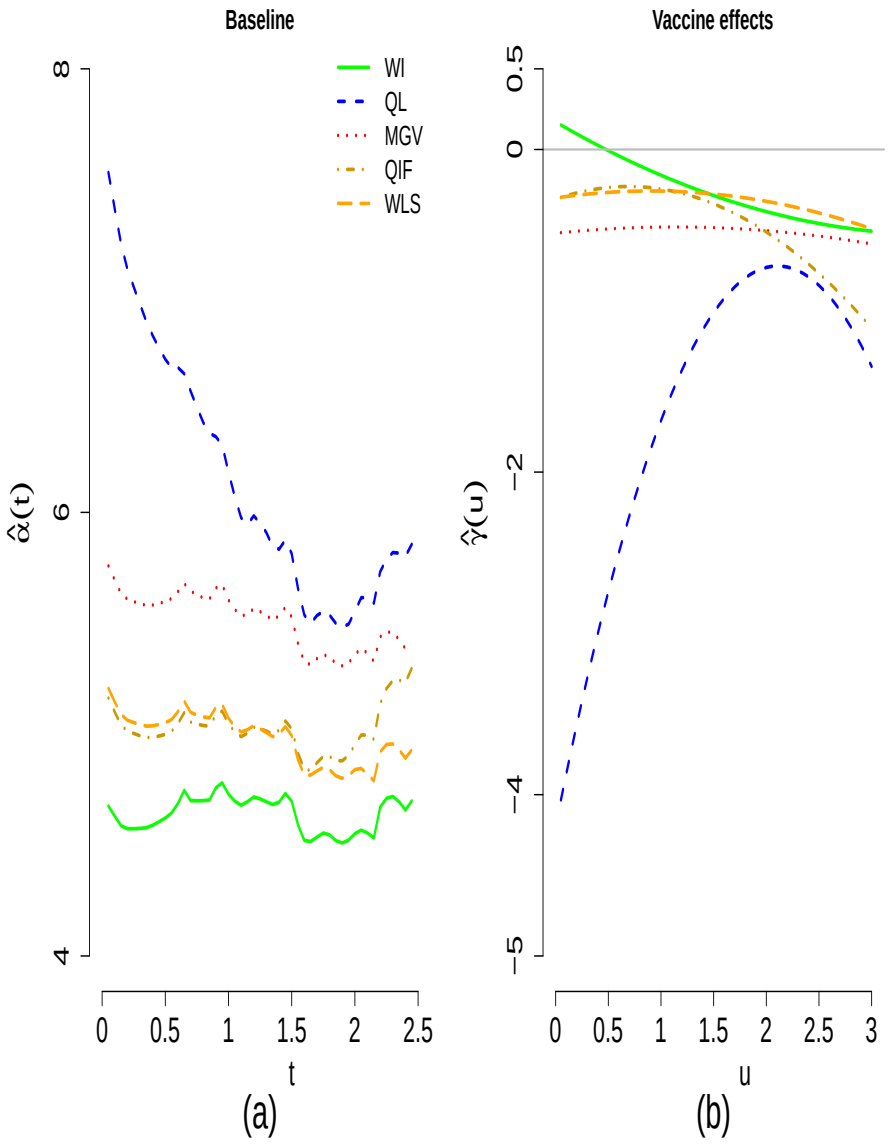

Figure 17: Estimates of baseline and varying-coefficient functions based on STEP data under Virus Load model (4.3) by different approaches. (a) is the estimated baseline function $\hat{\alpha}(t)$ by different approaches using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by different approaches.

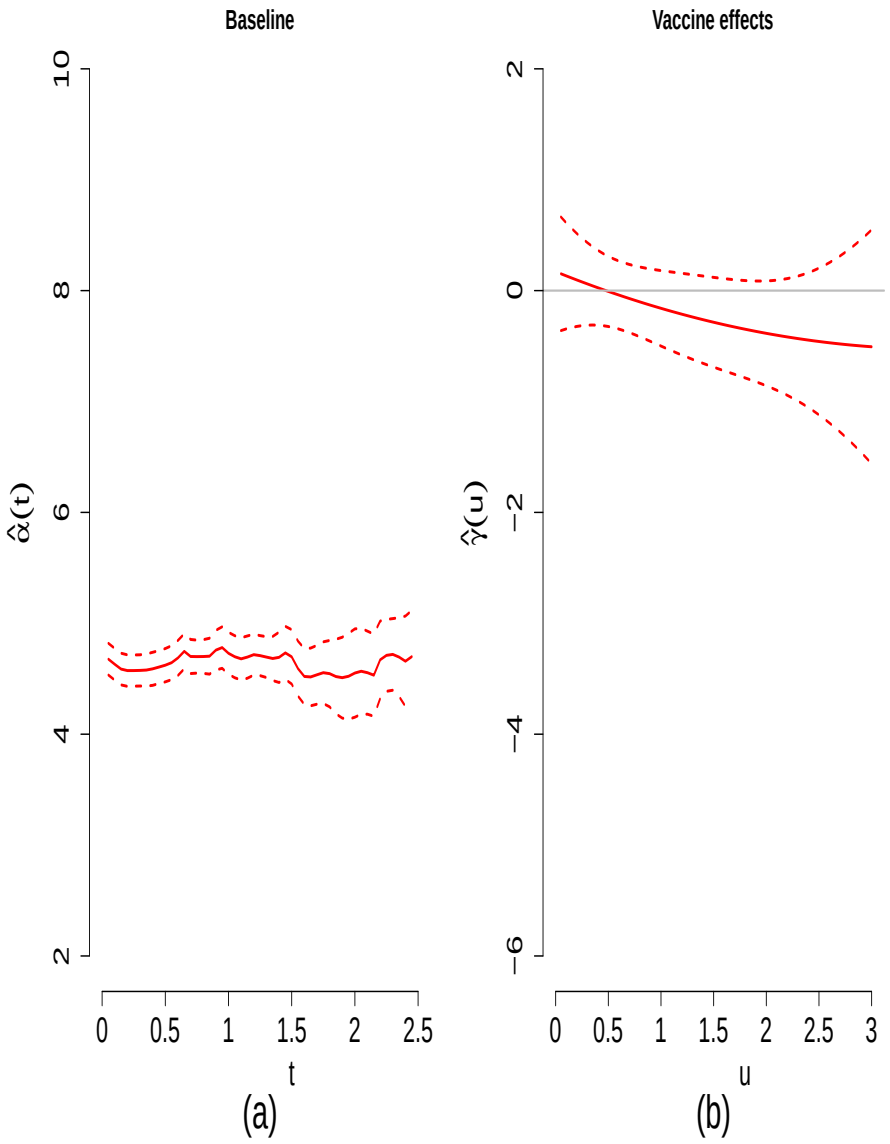

Figure 18: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under Virus Load model (4.3) by WI approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by WI approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by WI approach.

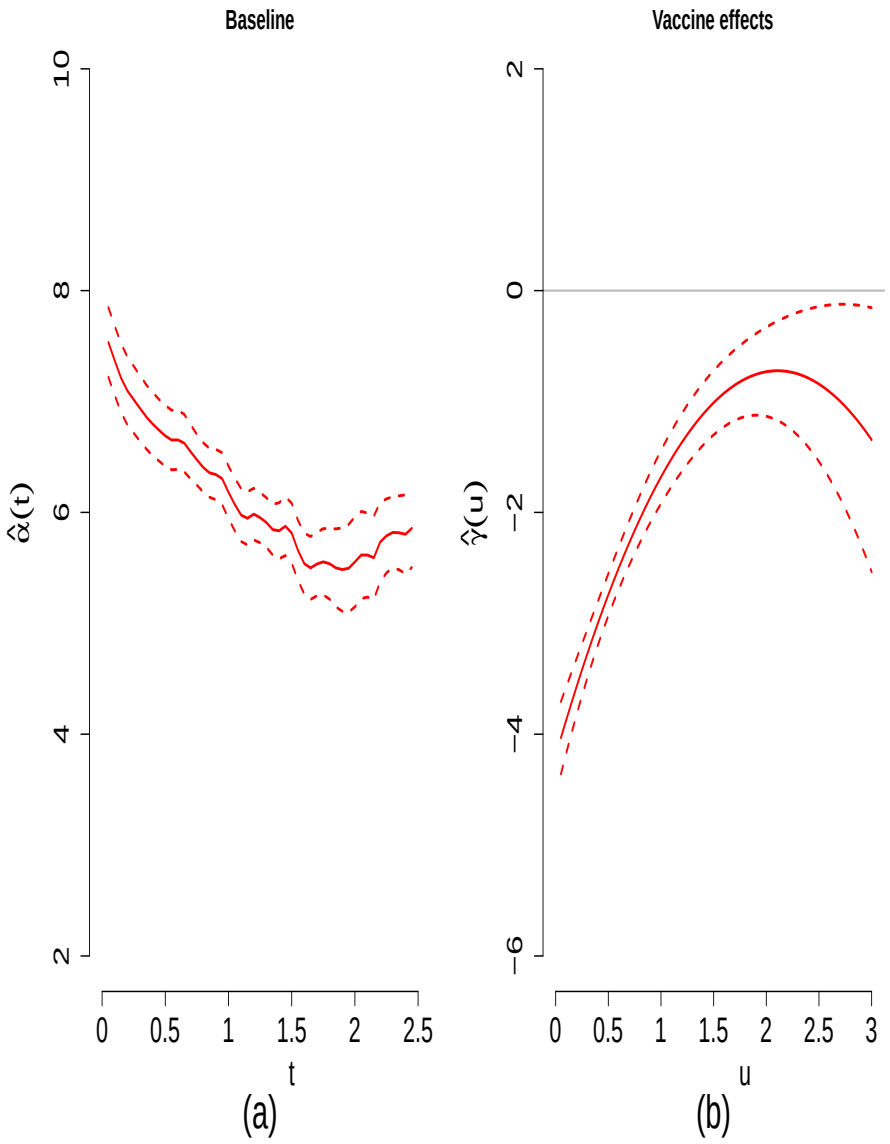

Figure 19: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under Virus Load model (4.3) by QL approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by QL approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by QL approach.

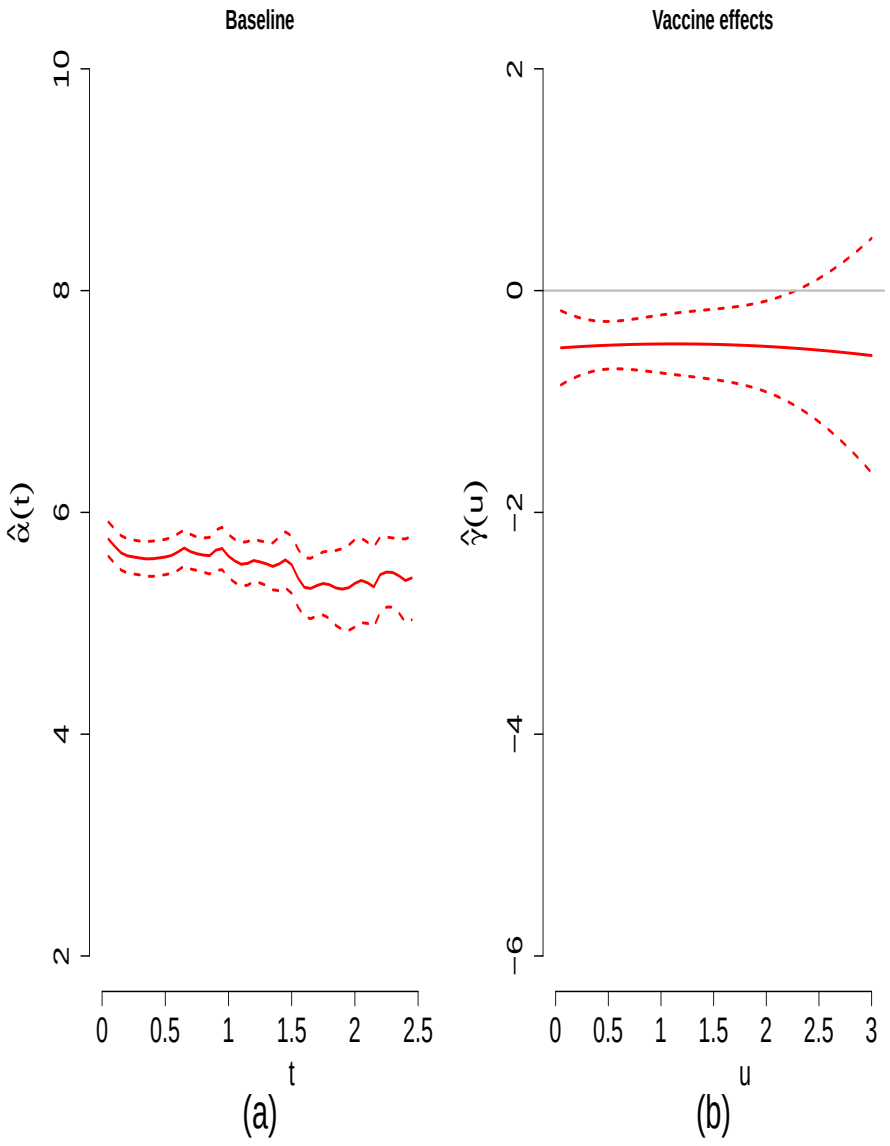

Figure 20: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under Virus Load model (4.3) by MGV approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by MGV approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by MGV approach.

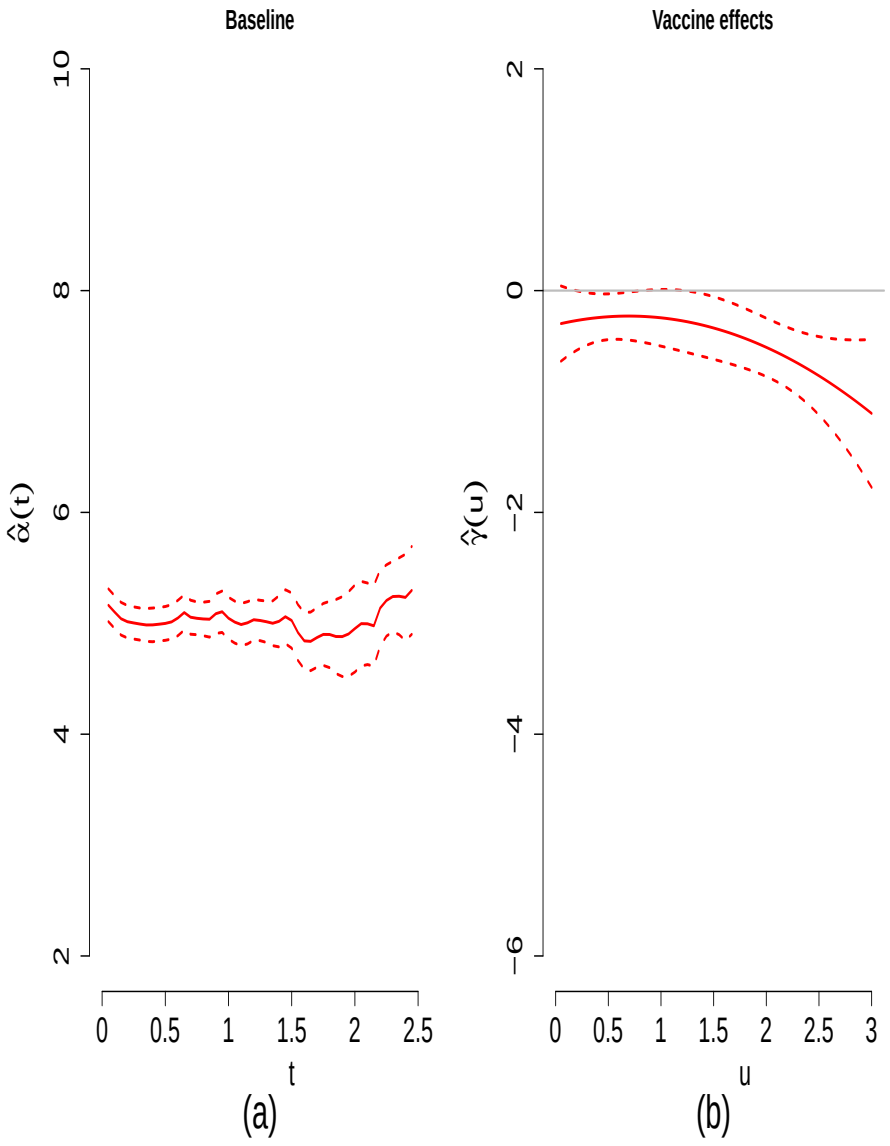

Figure 21: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under Virus Load model (4.3) by QIF approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by QIF approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by QIF approach.

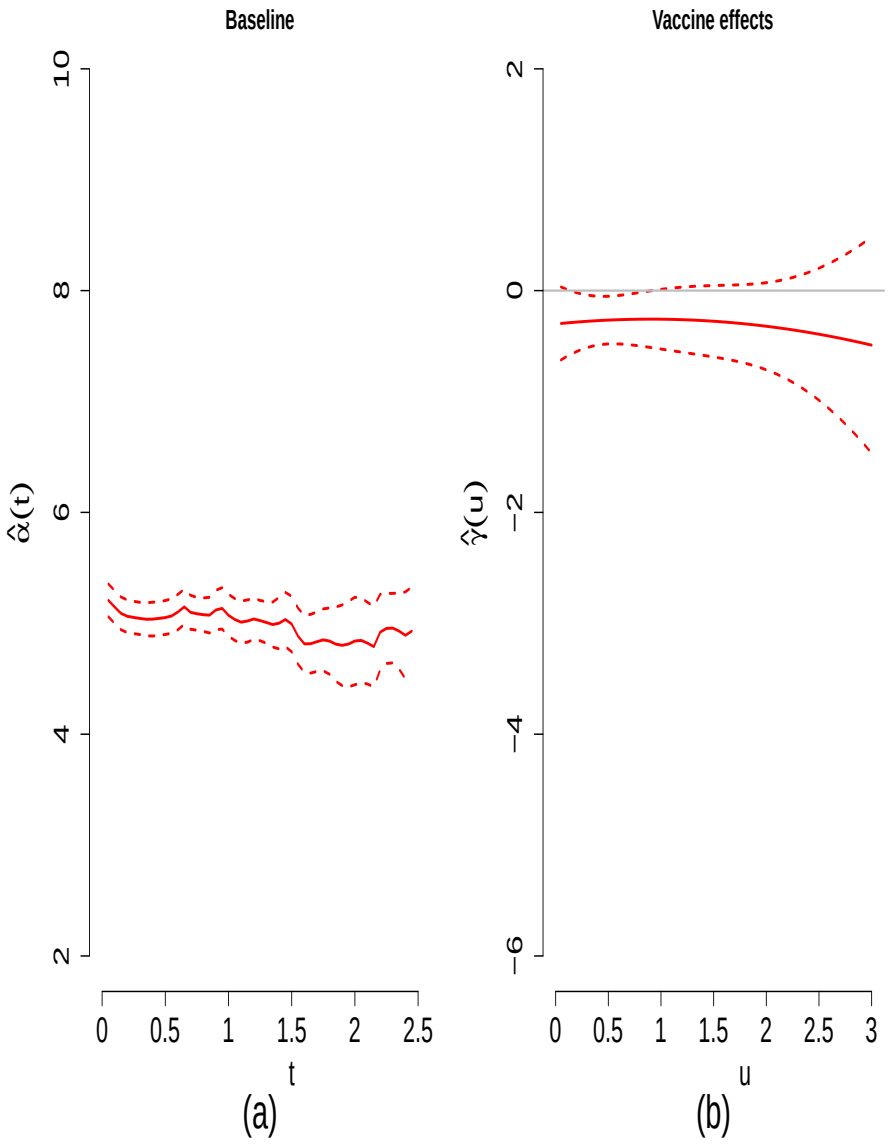

Figure 22: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under Virus Load model (4.3) by WLS approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by WLS approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by WLS approach.

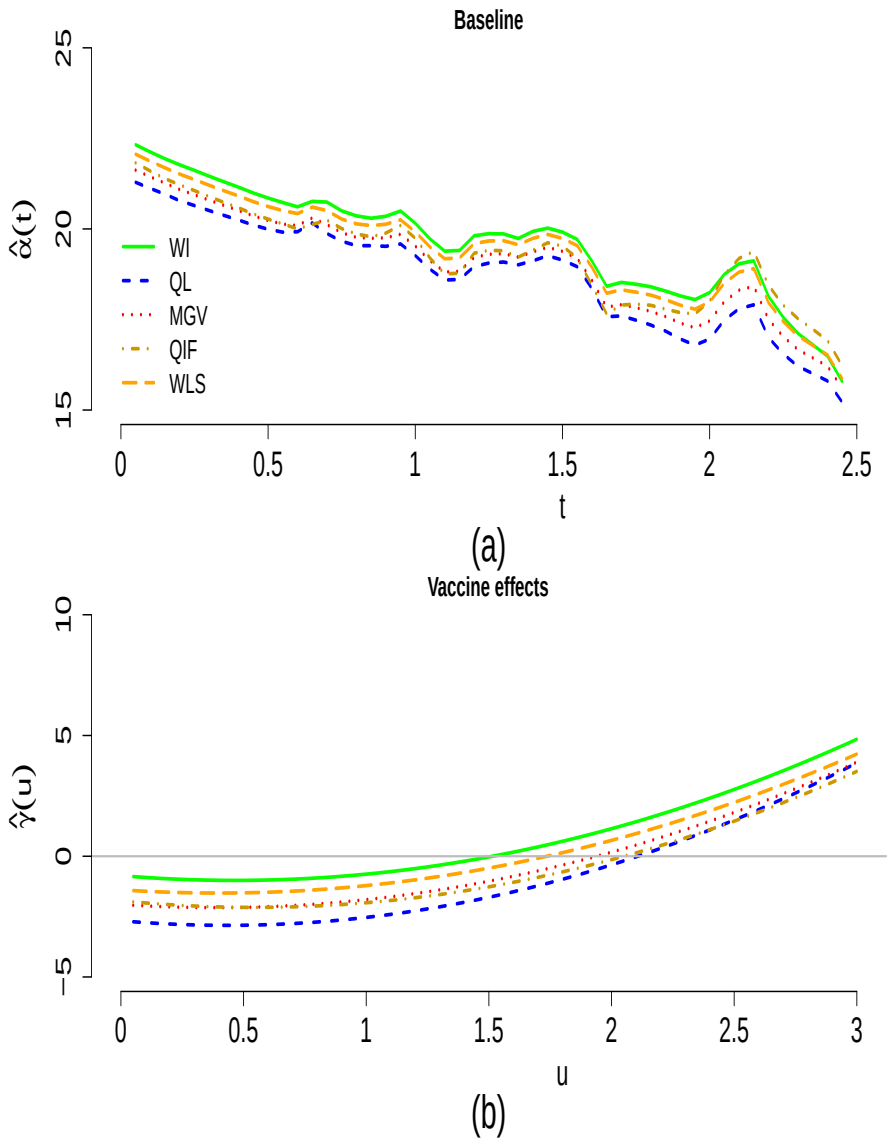

Figure 23: Estimates of baseline and varying-coefficient functions based on STEP data under CD4 model (4.4) by different approaches. (a) is the estimated baseline function $\hat{\alpha}(t)$ by different approaches using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by different approaches.

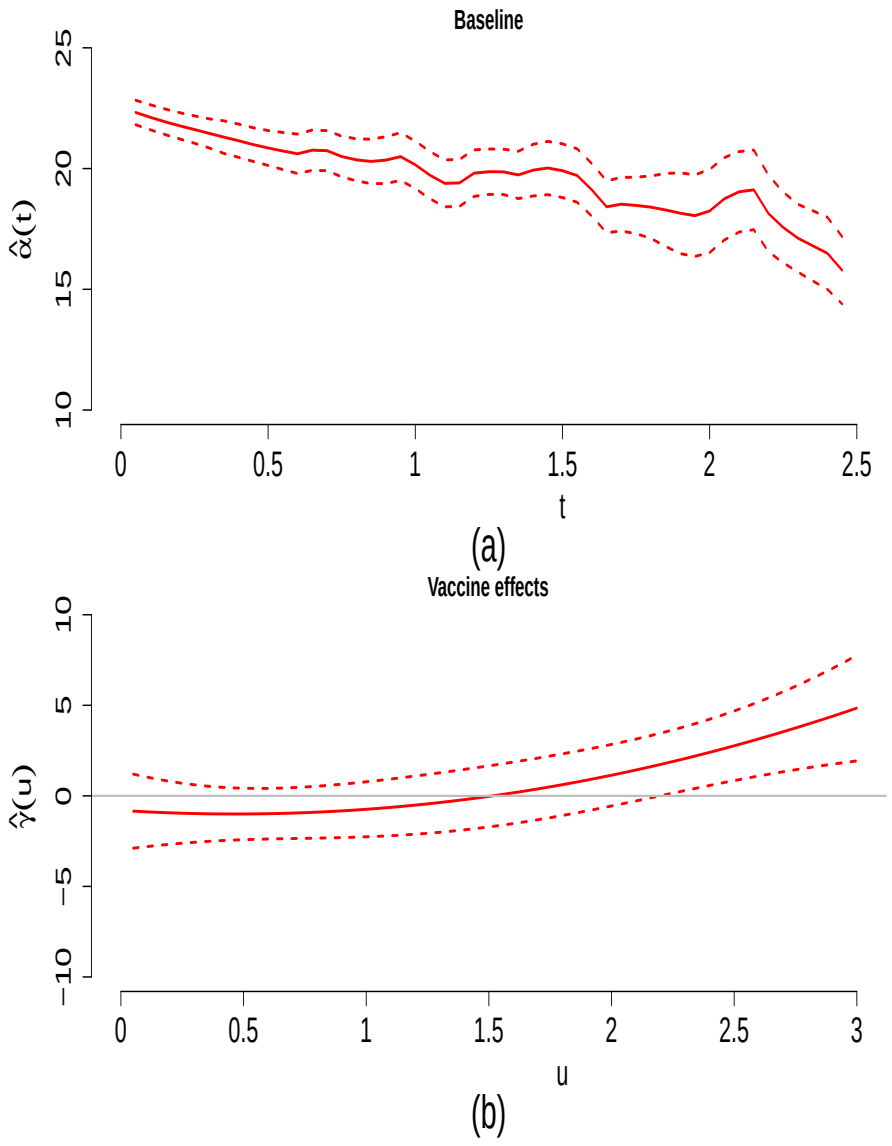

Figure 24: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under CD4 model (4.4) by WI approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by WI approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by WI approach.

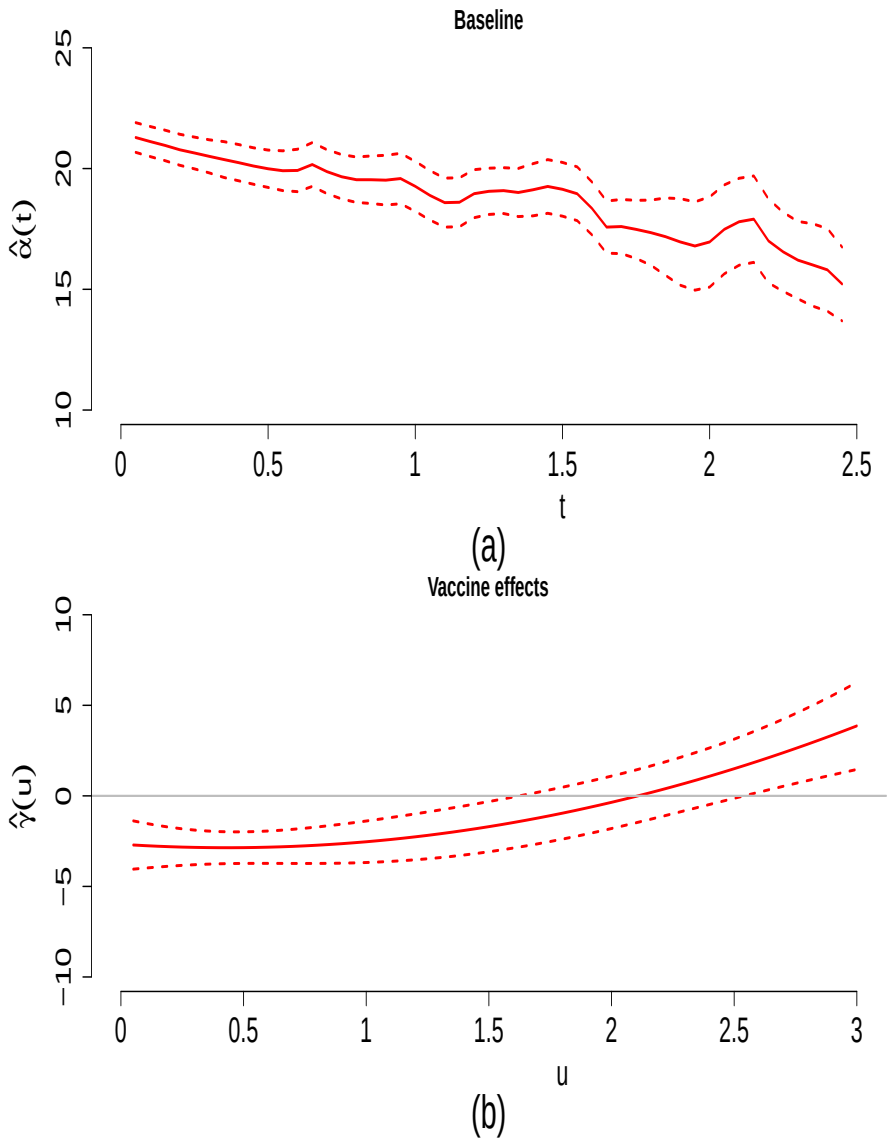

Figure 25: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under CD4 model (4.4) by QL approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by QL approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by QL approach.

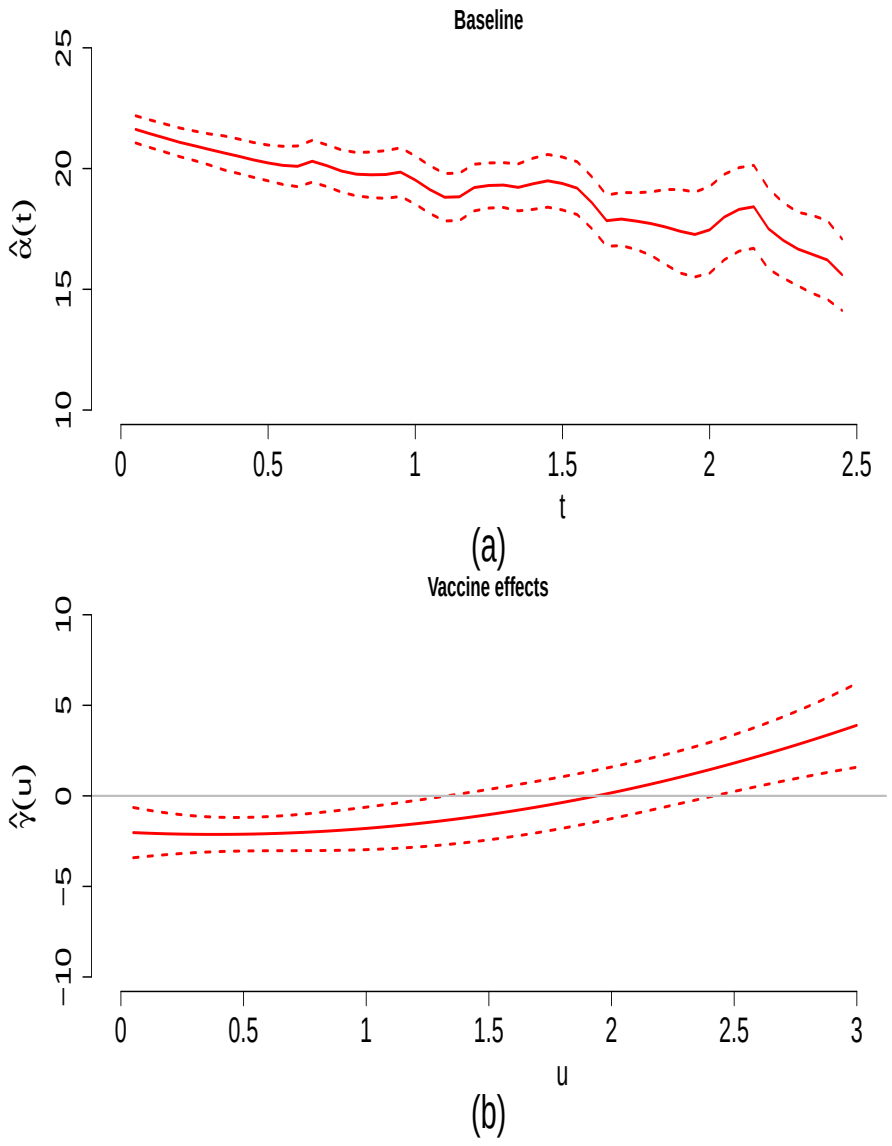

Figure 26: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under CD4 model (4.4) by MGV approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by MGV approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by MGV approach.

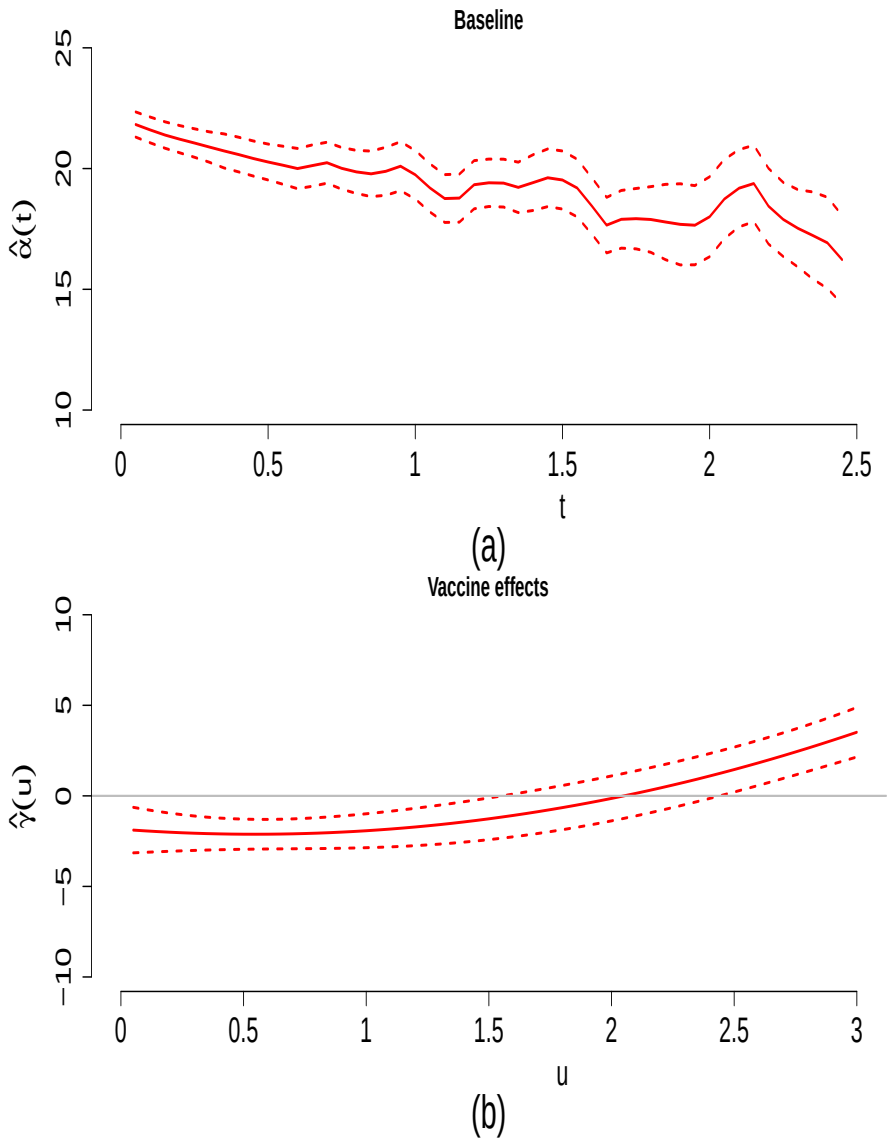

Figure 27: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under CD4 model (4.4) by QIF approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by QIF approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by QIF approach.

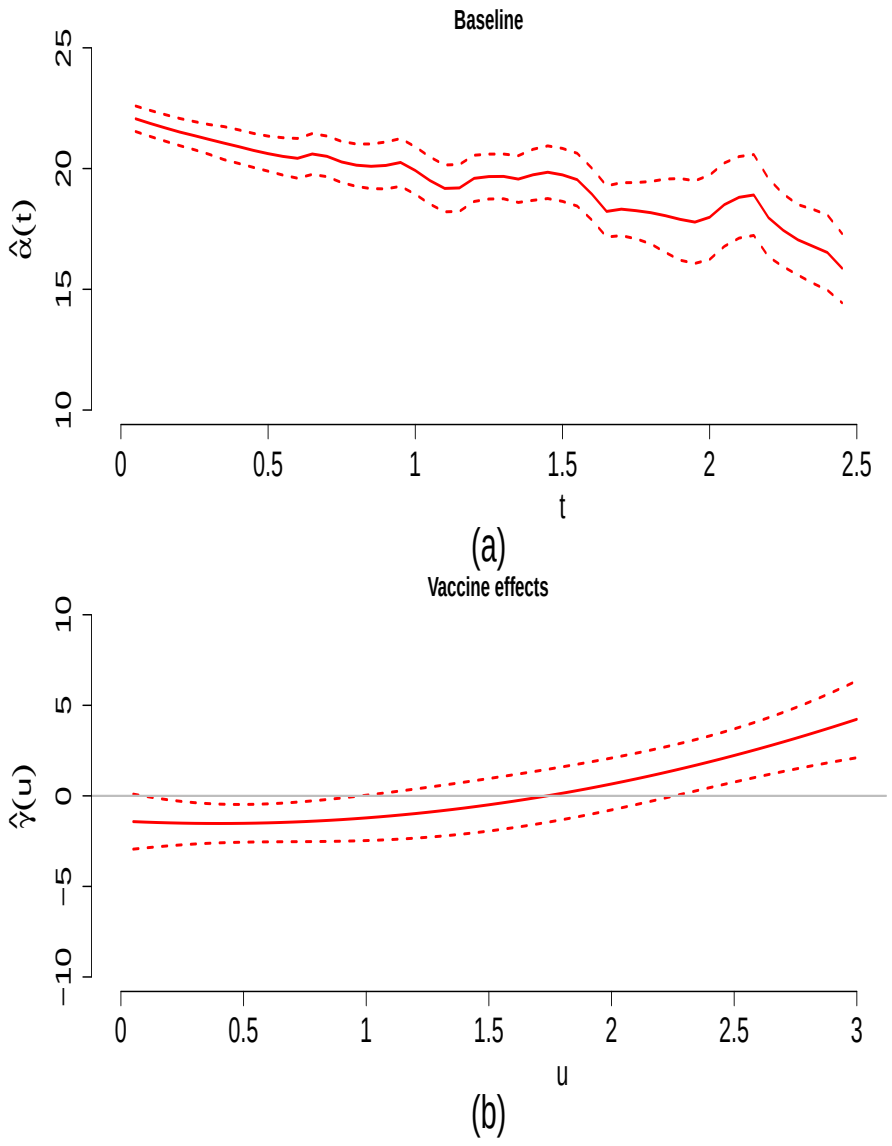

Figure 28: Estimates of baseline and varying-coefficient functions with 95% pointwise confidence intervals based on STEP data under CD4 model (4.4) by WLS approach. (a) is the estimated baseline function $\hat{\alpha}(t)$ by WLS approach using $h = 0.35$; (b) is the estimated vaccine effects $\hat{\gamma}(u)$ by WLS approach.

REFERENCES

Bai, Y., Zhu, Z., and Fung, W. K. (2008). Partial linear models for longitudinal data based on quadratic inference functions. *Scandinavian Journal of Statistics*, 35(1):104–118.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P. B., Lama, J. R., Marmor, M., del Rio, C., et al. (2008). Efficacy assessment of a cell-mediated immunity hiv-1 vaccine (the step study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet*, 372(9653):1881–1893.

Cai, Z., Fan, J., and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902.

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438):477–489.

Chen, K. and Jin, Z. (2006). Partial linear regression models for clustered data. *Journal of the American Statistical Association*, 101(473):195–204.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, pages 196–216.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, 102(478):632–641.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710–723.

Grabar, S., Le Moing, V., Goujard, C., Leport, C., Kazatchkine, M. D., Costagliola, D., and Weiss, L. (2000). Clinical outcome of patients with hiv-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy. *Annals of internal medicine*, 133(6):401–410.

Gray, G. E., Allen, M., Moodie, Z., Churchyard, G., Bekker, L.-G., Nchabeleng, M., Mlisana, K., Metch, B., de Bruyn, G., Latka, M. H., et al. (2011). Safety and efficacy of the hvtn 503/phambili study of a clade-b-based hiv-1 vaccine in south africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. *The Lancet infectious diseases*, 11(7):507–515.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.

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

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 757–796.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.

He, X., Fung, W. K., and Zhu, Z. (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, 100(472):1176–1184.

He, X., Zhu, Z.-Y., and Fung, W.-K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, 89(3):579–590.

Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98.

Huang, J. Z., Zhang, L., and Zhou, L. (2007). Efficient estimation in marginal partially linear models for longitudinal/clustered data using splines. *Scandinavian Journal of Statistics*, 34(3):451–477.

Larder, B. A., Kellam, P., and Kemp, S. D. (1991). Zidovudine resistance predicted by direct detection of mutations in dna from hiv-infected lymphocytes. *Aids*, 5(2):137–144.

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

Lin, D. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American statistical Association*, 95(450):520–534.

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

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

Macke, J. H., Berens, P., Ecker, A. S., Tolias, A. S., and Bethge, M. (2009). Generating spike trains with specified correlation coefficients. *Neural computation*, 21(2):397–423.

McElrath, M. J., De Rosa, S. C., Moodie, Z., Dubey, S., Kierstead, L., Janes, H., Defawe, O. D., Carter, D. K., Hural, J., Akondy, R., et al. (2008). Hiv-1 vaccine-induced immunity in the test-of-concept step study: a case–cohort analysis. *The Lancet*, 372(9653):1894–1905.

Mocroft, A., Ledergerber, B., Katlama, C., Kirk, O., Reiss, P, d., Monforte, A. d., Knysz, B., Dietrich, M., Phillips, A., Lundgren, J. D., et al. (2003). Decline in the aids and death rates in the eurosida study: an observational study. *The Lancet*, 362(9377):22–29.

Montaner, J. S., Reiss, P., Cooper, D., Vella, S., Harris, M., Conway, B., Wainberg, M. A., Smith, D., Robinson, P., Hall, D., et al. (1998). A randomized, double-blind trial comparing combinations of nevirapine, didanosine, and zidovudine for hiv-infected patients: the incas trial. *Jama*, 279(12):930–937.

Pepe, M. S. and Couper, D. (1997). Modeling partly conditional means with longitudinal data. *Journal of the American Statistical Association*, 92(439):991–998.

Piatak Jr, M., Saag, M., Yang, L., Clark, S., Kappes, J., Luk, K., Hahn, B., Shaw, G., and Lifson, J. (1993). High levels of hiv-1 in plasma during all stages of infection determined by competitive pcr. *SCIENCE-NEW YORK THEN WASHINGTON-*, 259:1749–1749.

Qi, L., Sun, Y., and Gilbert, P. B. (2016). Generalized semiparametric varying-coefficient model for longitudinal data with applications to adaptive treatment randomizations. *Biometrics*.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62(2):379–391.

Qu, A., Lindsay, B. G., and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika*, 87(4):823–836.

Qu, A. and Song, P. X.-K. (2004). Assessing robustness of generalised estimating equations and quadratic inference functions. *Biometrika*, 91(2):447–459.

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

Sun, Y., Gilbert, P. B., and McKeague, I. W. (2009). Proportional hazards models with continuous marks. *Annals of statistics*, 37(1):394.

Sun, Y., Sun, L., and Zhou, J. (2013). Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime data analysis*, 19(3):317–349.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, 32(1):21–47.

Tian, L., Zucker, D., and Wei, L. (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association*, 100(469):172–183.

Van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3. Cambridge university press.

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

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

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics*, pages 689–699.

APPENDIX A: PROOFS OF THE LEMMAS

Condition I.

(I.1) The processes $Y_i(t)$, $X_i(t)$ and $U_i(t)$, $0 \leq t \leq \tau$, are bounded and their total variations are bounded by a constant;

(I.2) The kernel function $K(\cdot)$ is symmetric with compact support on $[-1, 1]$ and bounded variation; bandwidth $h \to 0$; $nh^2 \to \infty$ and nh^5 is bounded.

(I.3) The link function $g(\cdot)$ is monotone and its inverse function $g^{-1}(\cdot)$ is twice differentiable;

(I.4) $\alpha_0(t)$, $e_{11}(t)$ and $e_{12}(t)$ are twice differentiable; $e_{11}^{-1}(t)$ is bounded over $0 \leq t \leq \tau$; the matrices P and D are positive definite;

(I.5) The weight process $V_1^{-1}(t, x) \xrightarrow{\mathcal{P}} v_1^{-1}(t, x)$ uniformly in the range of (t, x); $v_1^{-1}(t, x)$ is differentiable with uniformly bounded partial derivatives;

(I.6) The equation

$$hE\{X_1^T \Delta K_h^{1/2}(t) V_1^{-1}(t) K_h^{1/2}(t)(Y - \mu)(Y - \mu)^T K_h^{1/2}(t) V_1^{-1}(t) K_h^{1/2}(t) \Delta X_1\}$$

exists and is finite.

Lemmas

Let

$$u_\alpha(\alpha, \zeta, t) = E\{X_1^T \Delta(t) K_h^{1/2}(t) V_1^{-1}(t) K_h^{1/2}(t) \left[\mu(\alpha_0(t), \zeta_0 | X, U) - \mu(\alpha(t), \zeta | X, U)\right]\}$$

by suppressing index i. Define $\alpha_\zeta(t)$ as the unique root such that $u_\alpha(\alpha_\zeta, \zeta, t) = 0$ for $\zeta \in \mathcal{N}_{\zeta_0}$ and $\alpha^*_\zeta(t) = (\alpha^T_\zeta(t), \mathbf{0}^T_{p_1})^T$ where $\mathbf{0}_{p_1}$ is a $p_1 \times 1$ vector of zeros. Let

$$e_{\zeta,11}(t) = \sum_{j=1}^{J} E\big[\Delta^2_{\zeta,jj}(t)\nu_1^{jj}(t)X_{1j}X_{1j}^T|\mathbf{T}_j = \mathbf{t}\big]f_j(\mathbf{t})$$

and

$$e_{\zeta,12}(t) = \sum_{j=1}^{J} E\big[\Delta^2_{\zeta,jj}(t)\nu_1^{jj}(t)X_{1j}\{\frac{\partial \eta^T(U_j,\zeta)}{\partial \zeta}X^*_{2j}\}^T|\mathbf{T}_j = \mathbf{t}\big]f_j(\mathbf{t}),$$

where $\Delta_{\zeta,jj}(t) = \dot{\mu}_j(\alpha_\zeta(t), \zeta|X_j, U_j)$. When $\zeta = \zeta_0$, we have $\alpha_\zeta(t) = \alpha_0(t)$, $e_{\zeta,11}(t) = e_{11}(t)$ and $e_{\zeta,12}(t) = e_{12}(t)$.

The following lemmas are used for proving the main theorems. The proof of the lemmas make repeated applications of the Glivenko-Cantelli Theorem (Van der Vaart, 1998). A sufficient condition for applying the Glivenko-Cantelli Theorem can be checked by estimating the order of the bracketing number, similar to the proof of Lemma 2 of Sun et al. (2009). This sufficient condition holds under the conditions provided in Condition I. Let $H = \text{diag}\{I_{p_1}, hI_{p_1}\}$.

Lemma A.1. *Under Condition I, as $n \to \infty$, $H\tilde{\alpha}^*(t,\zeta) \xrightarrow{\mathcal{P}} \alpha^*_\zeta(t)$,*

$$H\partial\tilde{\alpha}^*(t,\zeta)/\partial\zeta \xrightarrow{\mathcal{P}} (-(e_{\zeta,11}(t)^{-1}e_{\zeta,12}(t))^T, \mathbf{0}^T_{p_1\times(p_2+p_3)})^T, \tag{A.1}$$

and $H\partial^2\tilde{\alpha}^*(t,\zeta)/\partial\zeta^2$ converges in probability to a deterministic function of (t,ζ) of bounded variation, uniformly in $t \in [t_1, t_2] \subset (0,\tau)$ and $\zeta \in \mathcal{N}_{\zeta_0}$ at the rate $n^{-1/2+\nu}$ for $\nu > 0$.

Proof of Lemma A.1

The first result of this lemma follows from Lemma 1 of Sun et al. (2013) directly.

We only prove the second and the third results.

By (3.6),

$$H\frac{\partial\tilde{\alpha}^*(t,\zeta)}{\partial\zeta} = -\left\{n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\alpha^*}\right\}^{-1} n^{-1}H^{-1}\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\zeta}\bigg|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}.$$

Note that

$$\begin{aligned}
&n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\alpha^*}\\
&= -n^{-1}H^{-2}\sum_{i=1}^{n} X_{1i}^*(t)^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\Delta_i(t)X_{1i}^*(t)\\
&= -H^{-2}\sum_{j=1}^{J} E\{\Delta_{jj}^2(t)\nu_1^{jj}(t)X_{1j}^*(t)X_{1j}^*(t)^T|\mathbf{T}_j=\mathbf{t}\}f_j(\mathbf{t}) + O_p(\frac{1}{\sqrt{n}h})
\end{aligned}$$

uniformly in t by Glivenko-Cantelli Theorem.

Since $H\tilde{\alpha}^*(t,\zeta)\xrightarrow{\mathcal{P}}\alpha_\zeta^*(t)$, we have

$$\begin{aligned}
&n^{-1}H^{-2}\left.\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\alpha^*}\right|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}\\
&= -\begin{pmatrix} 1 & 0\\ 0 & C_K(2)\end{pmatrix}\otimes\sum_{j=1}^{J} E\{\Delta_{jj}^2(t)\nu_1^{jj}(t)X_{1j}X_{1j}^T|\mathbf{T}_j=\mathbf{t}\}f_j(\mathbf{t})\\
&\qquad + O(h^2) + O_p(\frac{1}{\sqrt{n}h})\\
&\xrightarrow{\mathcal{P}} -\begin{pmatrix} 1 & 0\\ 0 & C_K(2)\end{pmatrix}\otimes\sum_{j=1}^{J} E\big[\Delta_{jj}^2(t)\nu_1^{jj}(t)X_{1j}X_{1j}^T|\mathbf{T}_j=\mathbf{t}\big]f_j(\mathbf{t}) \qquad \text{(A.2)}
\end{aligned}$$

uniformly in t and $\zeta \in \mathcal{N}_{\zeta_0}$. Similarly,

$$
\begin{aligned}
& n^{-1}H^{-1} \left. \frac{\partial U_\alpha(\alpha^*, \zeta, t_0)}{\partial \zeta} \right|_{\alpha^* = \tilde{\alpha}^*(t,\zeta)} \\
& \xrightarrow{\mathcal{P}} \begin{pmatrix} -\sum_{j=1}^{J} E\big[\Delta_{jj}^2(t)\nu_1^{jj}(t) X_{1j}\{\frac{\partial \eta^T(U_j,\zeta)}{\partial \zeta} X_{2j}^*\}^T | \mathbf{T}_j = \mathbf{t}\big] f_j(\mathbf{t}) \\ 0 \end{pmatrix}
\end{aligned} \tag{A.3}
$$

uniformly in t and $\zeta \in \mathcal{N}(\zeta_0)$. Therefore, (A.1) holds uniformly in t and $\zeta \in \mathcal{N}(\zeta_0)$.

By a similar argument, the third statement holds. □

Lemma A.2. *Under Condition I,*

$$
\sqrt{nh}\{\tilde{\alpha}(t,\zeta_0) - \alpha_0(t) - \frac{1}{2}C_K(2)h^2\ddot{\alpha}_0^T(t)\} = e_{11}^{-1}(t)(nh)^{1/2}n^{-1}U_\alpha(\alpha_0(t),\zeta_0) + o_p(1), \tag{A.4}
$$

uniformly in $t \in [t_1, t_2] \subset (0, \tau)$, *where*

$$
U_\alpha(\alpha_0(t),\zeta_0) = \sum_{i=1}^{n} X_{1i}^T \Delta_i(t) K_{ih}^{1/2}(t) V_{1i}^{-1}(t) K_{ih}^{1/2}(t) \left[Y_i - \mu_i\right]
$$

Further, $(nh)^{1/2}n^{-1}U_\alpha(\alpha_0(t),\zeta_0) = O_p(1)$ *uniformly in* $t \in [t_1, t_2] \subset (0, \tau)$.

Proof of Lemma A.2

Applying the first order Taylor expansion to $U_\alpha(\tilde{\alpha}^*(t,\zeta_0),\zeta_0)$, we have

$$
\sqrt{nh}H(\tilde{\alpha}^*(t,\zeta_0) - \alpha_0^*(t)) = -\left\{ n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha_0^*(t),\zeta_0)}{\partial \alpha^*} \right\}^{-1} \sqrt{\frac{h}{n}} H^{-1} U_\alpha(\tilde{\alpha}^*(t,\zeta_0),\zeta_0)
$$

The first p_1 components of the above equation is

$$
\sqrt{nh}(\tilde{\alpha}(t,\zeta_0) - \alpha_0(t)) = e_{11}^{-1}(t)(h/n)^{1/2} U_\alpha(\tilde{\alpha}(t,\zeta_0),\zeta_0)\{1 + o_p(1)\} \tag{A.5}
$$

By the local linear approximation for $\alpha_0(t)$ around t_0,

$$\begin{aligned}
&\mu_i(t) - g^{-1}\{\alpha^{*T}(t_0)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\} \\
&= g^{-1}\{\alpha_0^T(t)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\} - g^{-1}\{\alpha^{*T}(t_0)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\} \\
&= \dot{\mu}_i(t)\{\frac{1}{2}\ddot{\alpha}_0^T(t_0)X_{1i}(t)(t-t_0)^2 + O((t-t_0)^3)\}, \qquad \text{(A.6)}
\end{aligned}$$

Denote $\mu_i = (\mu_{i1}, \cdots, \mu_{iJ_i})^T$ and $\mu_i^* = (\mu_{i1}^*, \cdots, \mu_{iJ_i}^*)^T$ where

$$\mu_{ij} = g^{-1}\{\alpha_0^T(T_{ij})X_{1ij} + \eta^T(U_{ij},\zeta_0)X^*_{2ij}\}$$

and

$$\mu^*_{ij} = g^{-1}\{\alpha^{*T}(T_{ij})X_{1ij} + \eta^T(U_{ij},\zeta_0)X^*_{2ij}\}$$

It follows that

$$\begin{aligned}
&(h/n)^{1/2}U_\alpha(\tilde{\alpha}(t,\zeta_0),\zeta_0) \\
&= (h/n)^{1/2}\sum_{i=1}^n X_{1i}^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t) \\
&\qquad\qquad \times [Y_i - \mu_i + \mu_i - \mu_i^*] \\
&= (h/n)^{1/2}\sum_{i=1}^n X_{1i}^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\,[Y_i - \mu_i] + \frac{1}{2}\sqrt{nh}C_K(2)h^2\ddot{\alpha}_0^T(t)e_{11}(t)
\end{aligned}$$

Hence

$$\begin{aligned}
&\sqrt{nh}(\tilde{\alpha}(t,\zeta_0) - \alpha_0(t) - \frac{1}{2}C_K(2)h^2\ddot{\alpha}_0^T(t)) \\
&= e_{11}^{-1}(t)\sqrt{\frac{h}{n}}\sum_{i=1}^n X_{1i}^T\Delta_i(t)K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\,[Y_i - \mu_i] \\
&\qquad\qquad + o_p((nh)^{-1} + \sqrt{nh^5})
\end{aligned}$$

Follow Appendix A of Tian et al. (2005), the right hand side of above equation is

$O_p(1)$ uniformly in $t \in [t_1, t_2]$. □

APPENDIX B: PROOFS OF THE THEOREMS IN CHAPTER 3

Proof of Theorem 3.1.

We first consider the proof for the consistency of $\hat{\zeta}$. By Glivenko-Cantelli theorem and Lemma A.1, we have

$$n^{-1}U_\zeta(\zeta) \xrightarrow{\mathcal{P}} E\{B^T \Delta V_2^{-1}\,[Y - \hat{\mu}(\zeta)]\}, \tag{B.1}$$

uniformly for $\zeta \in \mathcal{N}_{\zeta_0}$, where

$$B_i = (B_{i1}, \cdots, B_{iJ_i})^T$$

and

$$B_{ij} = -e_{12}^T(T_{ij})(e_{11}(T_{ij}))^{-1}X_{1ij} + \frac{\partial \eta(U_{ij}, \zeta)}{\partial \zeta} X_{2ij}^*,$$

By suppressing the index i, we denote $B = (B_1, \cdots, B_J)^T$, similarly for Y, Δ, V_2 and μ.

The right side of equation (B.1) equals to

$$E\{B^T \Delta V_2^{-1}\,[\mu - \hat{\mu}(\zeta)]\}.$$

defined as $u(\zeta)$ by double expectation. Taking partial derivative of $U_\zeta(\zeta)$ with respect

to ζ and applying Lemma 1, we have

$$
\begin{aligned}
n^{-1}\frac{\partial U_\zeta(\zeta)}{\partial \zeta} = -n^{-1}\sum_{i=1}^{n}\{&\frac{\partial \tilde{\alpha}(T_i,\zeta)}{\partial \zeta}X_{1i} + \frac{\partial \eta(U_i,\zeta)}{\partial \zeta}X_{2i}^*\}^T \Delta_i V_{2i}^{-1}\Delta_i \\
&\{\frac{\partial \tilde{\alpha}(T_i,\zeta)}{\partial \zeta}X_{1i} + \frac{\partial \eta(U_i,\zeta)}{\partial \zeta}X_{2i}^*\} \\
&+ n^{-1}\sum_{i=1}^{n}\left\{\frac{\partial^2 \tilde{\alpha}(T_i,\zeta)}{\partial \zeta^2}X_{1i} + \frac{\partial^2 \eta(U_i,\zeta)}{\partial \zeta^2}X_{2i}^*\right\}^T \\
&\Delta_i V_{2i}^{-1}\left[\mu_i - \hat{\mu}_i(\zeta)\right], \qquad \text{(B.2)}
\end{aligned}
$$

where $\{\frac{\partial^2 \tilde{\alpha}(T_i,\zeta)}{\partial \zeta^2}X_{1i} + \frac{\partial^2 \eta(U_i,\zeta)}{\partial \zeta^2}X_{2i}^*\}$ is a $J_i \times (p_2+p_3)$ matrix with each element being $\{\frac{\partial^{2T} \tilde{\alpha}(T_{ij},\zeta)}{\partial \zeta^2}X_{1ij} + \frac{\partial^2 \eta(U_{ij},\zeta)}{\partial \zeta^2}X_{2ij}^*\}$. When $\zeta = \zeta_0$, the latter term goes to zero as n goes to infinity by Lemma A.1 and the Glivenko-Cantelli theorem. It follows that

$$
-n^{-1}\left.\frac{\partial U_\zeta(\zeta)}{\partial \zeta}\right|_{\zeta=\zeta_0} \xrightarrow{\mathcal{P}} E\left[B^T \Delta V_2^{-1}\Delta B\right] = P \qquad \text{(B.3)}
$$

uniformly in a neighborhood of ζ_0. Since $u(\zeta_0) = 0$ and P is positive definite, ζ_0 is the unique root of $u(\zeta)$. By Theorem 5.9 of Van der Vaart (1998), we have $\hat{\zeta} \xrightarrow{\mathcal{P}} \zeta_0$.

Now we show the asymptotic normality of $n^{-1/2}U_\zeta(\zeta_0)$.

$$
\begin{aligned}
&n^{-1/2}U_\zeta(\zeta_0) \\
&= n^{-1/2}\sum_{i=1}^{n}\{\frac{\partial \tilde{\alpha}(T_i,\zeta)}{\partial \zeta}X_{1i} + \frac{\partial \eta(U_i,\zeta)}{\partial \zeta}X_{2i}^*\}^T \Delta_i V_{2i}^{-1}\left[Y_i - \hat{\mu}_i(\tilde{\alpha}(T_i,\zeta_0),\zeta_0)\right] \\
&= n^{-1/2}\sum_{i=1}^{n}\{\frac{\partial \tilde{\alpha}(T_i,\zeta)}{\partial \zeta}X_{1i} + \frac{\partial \eta(U_i,\zeta)}{\partial \zeta}X_{2i}^*\}^T \Delta_i V_{2i}^{-1}\left[Y_i - \mu_i\right] \\
&\quad + n^{-1/2}\sum_{i=1}^{n}\{\frac{\partial \tilde{\alpha}(T_i,\zeta)}{\partial \zeta}X_{1i} + \frac{\partial \eta(U_i,\zeta)}{\partial \zeta}X_{2i}^*\}^T \Delta_i V_{2i}^{-1}\left[\mu_i - \hat{\mu}_i(\tilde{\alpha}(T_i,\zeta_0),\zeta_0)\right] \qquad \text{(B.4)}
\end{aligned}
$$

where

$$\mu_i - \hat{\mu}_i(\tilde{\alpha}(T_i, \zeta_0), \zeta_0)$$
$$= g^{-1}\{\alpha_0(T_i)X_{1i} + \eta(U_i, \zeta_0)X_{2i}^*\} - g^{-1}\{\tilde{\alpha}(T_i, \zeta_0)X_{1i} + \eta(U_i, \zeta_0)X_{2i}^*\}$$

the element of which is

$$g^{-1}\{\alpha_0^T(T_{ij})X_{1ij} + \eta^T(U_{ij}, \zeta_0)X_{2ij}^*\} - g^{-1}\{\tilde{\alpha}^T(T_{ij}, \zeta_0)X_{1ij} + \eta^T(U_{ij}, \zeta_0)X_{2ij}^*\}$$

The second term of (B.4) is negligible because by Taylor expansion, which is

$$- n^{-1/2}\sum_{i=1}^{n}\{\frac{\partial\tilde{\alpha}(T_i, \zeta)}{\partial\zeta}X_{1i} + \frac{\partial\eta(U_i, \zeta)}{\partial\zeta}X_{2i}^*\}^T\Delta_i V_{2i}^{-1}\Delta_i$$
$$\{\alpha_0(T_i)X_{1i} - \tilde{\alpha}(T_i, \zeta)X_{1i}\}$$
$$= o_p(1),$$

by Lemma 1 in Lin and Ying (2001), and the element of vector $\{\frac{\partial\tilde{\alpha}(T_i,\zeta_0)}{\partial\zeta}X_{1i} + \frac{\partial\eta(U_i,\zeta_0)}{\partial\zeta}X_{2i}^*\}$ is $\{\frac{\partial\tilde{\alpha}^T(T_{ij},\zeta_0)}{\partial\zeta}X_{1ij} + \frac{\partial\eta^T(U_{ij},\zeta_0)}{\partial\zeta}X_{2ij}^*\}$

Hence,

$$n^{-1/2}U_\zeta(\zeta_0)$$
$$= n^{-1/2}\sum_{i=1}^{n}\{\frac{\partial\tilde{\alpha}(T_i, \zeta)}{\partial\zeta}X_{1i} + \frac{\partial\eta(U_i, \zeta)}{\partial\zeta}X_{2i}^*\}^T\Delta_i V_{2i}^{-1}\left[Y_i - \mu_i\right] + o_p(1)$$
$$= n^{-1/2}\sum_{i=1}^{n}B_i^T\Delta_i V_{2i}^{-1}\left[Y_i - \mu_i\right] + o_p(1), \tag{B.5}$$

which converges in distribution to $N(0, D)$ by central limit theorem, where D is defined in (3.1).

It follows from (B.3) and (B.5) that $n^{1/2}(\hat{\zeta} - \zeta_0)\xrightarrow{\mathcal{D}}N(0, P^{-1}DP^{-1})$. □

Proof of Theorem 3.2.

Write $\Sigma = cov(Y|X_1, X_2)$. Define

$$S = \left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} B^T\Delta V_2^{-1}\Sigma^{\frac{1}{2}} - \left\{E(B^T\Delta\Sigma\Delta B)\right\}^{-1} B^T\Delta\Sigma^{-\frac{1}{2}}$$

Then

$$\begin{aligned} SS^T = &\left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} B^T\Delta V_2^{-1}\Sigma V_2^{-1}\Delta B \left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} \\ &- \left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} B^T\Delta V_2^{-1}\Delta B \left\{E(B^T\Delta\Sigma\Delta B)\right\}^{-1} \\ &- \left\{E(B^T\Delta\Sigma\Delta B)\right\}^{-1} B^T\Delta V_2^{-1}\Delta B \left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} \\ &- \left\{E(B^T\Delta\Sigma\Delta B)\right\}^{-1} B^T\Delta\Sigma\Delta B \left\{E(B^T\Delta\Sigma\Delta B)\right\}^{-1} \end{aligned}$$

Because SS^T is nonnegative definite, we have that

$$\begin{aligned} E(SS^T) = &\left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} E(B^T\Delta V_2^{-1}\Sigma V_2^{-1}\Delta B) \left\{E(B^T\Delta V_2^{-1}\Delta B)\right\}^{-1} \\ &- \left\{E(B^T\Delta\Sigma^{-1}\Delta B)\right\}^{-1} \end{aligned}$$

is nonnegative definite. Hence

$$P^{-1}DP^{-1} - D_0^{-1} \geq 0$$

The equality holds if and only if $S = 0$, which occurs when $V_2 = \Sigma$.

Proof of Theorem 3.3.

Since $\hat{\alpha}(t) = \tilde{\alpha}(t, \hat{\zeta})$, we have $\hat{\alpha}(t) \xrightarrow{\mathcal{P}} \alpha_0(t)$ uniformly in $t \in [t_1, t_2]$ by applying continuous mapping theorem and the uniform consistency results in Lemma A.1 and Theorem 3.1. Now we prove the asymptotic normality.

By Taylor expansion we have

$$\sqrt{nh}(\tilde{\alpha}(t,\hat{\zeta}) - \tilde{\alpha}(t,\zeta_0)) = -(nh)^{1/2}\frac{\partial\tilde{\alpha}(t,\tilde{\zeta})}{\partial\zeta}(\hat{\zeta} - \zeta_0),$$

where $\tilde{\zeta}$ is on the line segment between ζ_0 and $\hat{\zeta}$, which is $O_p(h^{1/2})$, by (A.1) and Theorem 3.1. Thus

$$\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}C_K(2)h^2\ddot{\alpha}_0^T(t)\} \\
&= \sqrt{nh}\{\tilde{\alpha}(t,\zeta_0) - \alpha_0(t) - \frac{1}{2}C_K(2)h^2\ddot{\alpha}_0^T(t)\} + \sqrt{nh}(\tilde{\alpha}(t,\hat{\zeta}) - \tilde{\alpha}(t,\zeta_0)) \\
&= e_{11}^{-1}(t)^{-1}(h/n)^{1/2}\sum_{i=1}^{n} X_{1i}^T\Delta_i K_{ih}^{1/2}(t)V_{1i}^{-1}(t)K_{ih}^{1/2}(t)\left[Y_i - \mu_i\right] + O_p(h^{1/2}) \\
&= e_{11}^{-1}(t)n^{-1/2}\sum_{i=1}^{n}\psi_i(t) + O_p(h^{1/2}),
\end{aligned}$$

for $t \in [t_1, t_2]$ by (A.4).

Note that $E(\psi_i(t)) = 0$. It follows that $n^{-1/2}\sum_{i=1}^{n}\psi_i(t) \xrightarrow{\mathcal{D}} N(0,\Sigma_e)$ by applying the Lindeberg-Feller central limit theorem. Consequently,

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}C_K(2)h^2\ddot{\alpha}_0^T(t)) \xrightarrow{\mathcal{D}} N\left(0, (e_{11}(t))^{-1}\Sigma_e(t)(e_{11}(t))^{-1}\right). \qquad \square$$