



UW Biostatistics Working Paper Series

2-22-2011

BATE Curve in Assessment of Clinical Utility of Predictive Biomarkers

Xiao-Hua Zhou

University of Washington, azhou@u.washington.edu

Yunbei Ma

University of Washington, yunbeim@u.washington.edu

Suggested Citation

Zhou, Xiao-Hua and Ma, Yunbei, "BATE Curve in Assessment of Clinical Utility of Predictive Biomarkers" (February 2011). *UW Biostatistics Working Paper Series*. Working Paper 376.
<http://biostats.bepress.com/uwbiostat/paper376>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

BATE Curve in Assessment of Clinical Utility of Predictive Biomarkers

Xiao-Hua Zhou

Department of Biostatistics, University of Washington

Yunbei Ma

Department of Biostatistics, University of Washington

February 20, 2011

Abstract

In this paper, for time-to-event data, we propose a new statistical framework for casual inference in evaluating clinical utility of predictive biomarkers and in selecting an optimal treatment for a particular patient. This new casual framework is based on a new concept, called Biomarker Adjusted Treatment Effect (BATE) curve, which can be used to represent the clinical utility of a predictive biomarker and select an optimal treatment for one particular patient. We then propose semi-parametric methods for estimating the BATE curves of biomarkers and establish asymptotic results of the proposed estimators for the BATE curves. We also conduct extensive simulation studies to evaluate finite-sample properties of the proposed estimation methods. Finally, we illustrate the application of the proposed method in a real-world data set.

KEYWORDS: Predictive biomarker; cutoff points; interaction; BATE curve; time-to-event outcome; varying-coefficient.

1. INTRODUCTION

Due to complexity of cancer, current staging and risk-stratification methods in oncology, while helpful, often fail to adequately predict malignancy aggressiveness and/or response to a specific treatment. The rapid advance of molecular genetic technology and accompanying proliferation of

molecular diagnostics companies have set the stage for a new era in personalized medicine. This development allows the tumor-node-metastasis (TNM) staging system to incorporate additional biomarkers, such as gene expression data, which could provide more precise information for risk stratification and treatment selection. By identifying patients who are at high risk and who are more likely to benefit from a given treatment, we hope to be able to provide the most effective treatment to those who are most in need.

A predictive marker is a biomarker that predicts the differential efficacy (benefit) of a particular therapy based on the value of a biomarker (e.g., only patients expressing the biomarker will respond to the specific treatment or will respond to a greater degree than those without the biomarker) (Sargent 2005). To apply these exciting results to maximize patient benefit, we need to develop a systematic statistical methodology to assess the clinical utility of promising biomarkers for predicting patients' responses to particular treatments.

Most of the current statistical methods for assessing the clinical utility of a predictive biomarker are based on a comparison of estimated survival curves between a treatment and control group, stratified by the biomarker values. Such an approach has two main limitations. First, if the biomarker yields a continuous-scale, such the approach requires dichotomization, which is artificial and may lose important information. Second, such the approach does not adequately quantify the clinical utility of predictive biomarkers. Specifically, many biomarkers that may have a significant p-value but do not have true clinical utility. Statistical significance does not imply clinical relevance.

Other two methods for assessing clinical utility of predictive biomarkers have been also proposed in the literature (Freidlin and Simon, 2005; Jiang, 2007). Freidlin and Simon (2005) proposed a design that combines prospective development of a gene expression-based classifier to select sensitive patients with a properly powered test for overall effect, by assuming the biomarker is binary. Jiang *et al.* (2007) extended Freidlin's design to allow a continuous-scale biomarker and proposed a parametric model to select a cut point for a pre-specified biomarker of the sensitive subpopulation.

Both these methods have limitations. The first drawback is the strong assumption that the effect of the interaction between the biomarker and the treatment group on patient outcomes is a step-function with only one jump of the biomarker value. However, when the biomarker is continuous, it's most likely that the interaction continuously varies with the value of the biomarker. Another

main drawback of Jiang’s method is that the proposed model with a unknown cut point is not identifiable. One additional limitation of all above mentioned existing methods is that they are not based on a causal framework.

In this paper, for a randomized clinical trial with full compliance, we propose a new statistical framework for casual inference in assessing the effectiveness of predictive biomarkers and in selecting sensitive patients to one particular treatment. Our new statistical framework overcomes the limitations of the existing methods for evaluating predictive biomarkers in selecting optimal treatments for individual patients. Specifically, in this paper, we introduce a new concept, called the Biomarker Adjusted Treatment Effect (BATE) curve, which is a graphical plot of the treatment effect as a function of the biomarker value and can be used to select an optimal treatment for an individual patient. Compared to the existing methods, the BATE curve can visually display the treatment effect on the patient’s outcome as a function of the biomarker value.

This article is organized as follows. In Sections 2 and 3, we illustrate the mathematic definitions and applications of our proposed BATE curves to select cutoff points, compare biomarkers, choose the optimal treatment for individual patients for the situation with multiple treatments, and test the usefulness of a predictive biomarker, In Section 4, we propose an estimation method for BATE curves and derive asymptotic properties of the estimated BATE curves. In Section 5, we derive the asymptotic distribution of the maximum deviations of the estimated BATE curves from the true BATE curves, and use it to construct the test statistics for the null hypothesis that the BATE curve is a constant and to construct simultaneous confidence bands for the BATE curve. Simulation studies and a real data example are reported in Sections 6 and 7, respectively. We give detailed proofs of the asymptotic results in Appendix C.

2. THE BATE CURVE FOR A TWO-ARM RANDOMIZED CLINICAL TRIALS WITH FULL COMPLIANCE

As discussed above, it is reasonable to assume that the treatment effect is a continuous curve of the biomarker value. This curve, which is named as the Biomarker Adjusted Treatment Effect (BATE) curve, is also flexible enough to indicate both local and global association between treatment and the biomarker value. In this section, for the randomized clinical trial with full compliance and with

one experimental and one control arms, we define the BATE curve mathematically for time-to-event outcome based on a casual framework, and demonstrate its detailed applications, such as the selection of sensitive patients and the identification of cutoff points. We then extend the BATE curve to the randomized clinical trial with full compliance and with multiple treatment arms in Section 3.

We first give some necessary notation and definitions.

2.1 Notation

Suppose that n independent subjects are randomized over experimental treatment or control. We let $T_i(1)$ be the event time of a subject that would be observed if the i th subject had received the experimental arm and $T_i(0)$ be the event time that would be observed if the i th subject had received the control arm. For one single biomarker V , we further define two potential conditional hazard rate functions as follows:

$$\begin{aligned}\lambda^{(1)}(t|v) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_i(1) \leq t + \Delta t | T_i(1) > t, V_i = v)}{\Delta t}, \\ \lambda^{(0)}(t|v) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_i(0) \leq t + \Delta t | T_i(0) > t, V_i = v)}{\Delta t}.\end{aligned}$$

Let T_i denote the event time, and we let C_i denote the censoring time and let $X_i = \min(T_i, C_i)$ denote the observed time for the i th individual, $i = 1, \dots, n$. Let Δ_i be an indicator which equals 1 if X_i is an event time and 0 otherwise. Let V_i and Z_i denote the biomarker value and treatment group indicator for the i th individual, respectively, where $Z_i = 1$ if individual i is assigned to the experimental group and $Z_i = 0$ otherwise, $i = 1, \dots, n$. We assume that the censoring times are independent of the failure times conditional on the covariates and that the observation period is $[0, \tau]$, where τ is a constant denoting the time for end of the study.

2.2 The BATE Curve

To evaluate the treatment effect adjusted for one single continuous biomarker V , we assume that the multiplicative interaction is of interest, we define the biomarker-adjusted treatment effect (BATE) curve at time t as follows:

$$\beta(v; t) = \log \left\{ \frac{\lambda^{(1)}(t|v)}{\lambda^{(0)}(t|v)} \right\}. \quad (1)$$

Here $\beta(v; t)$ is an unknown smooth function of v for any fixed $t \in (0, \tau)$.

In a randomized clinical trial with full compliance, since $\lambda^{(1)}(t|v) = \lambda(t|Z = 1, V = v)$ and $\lambda^{(0)}(t|v) = \lambda(t|Z = 0, V = v)$, we can express the BATE curve as

$$\beta(v; t) = \log \left\{ \frac{\lambda(t|Z = 1, V = v)}{\lambda(t|Z = 0, V = v)} \right\}. \quad (2)$$

In this paper, we focus on a special case that $\beta(v; t) \equiv \beta(v)$. This assumption holds under many well known models, such as Cox's model and varying-coefficient Cox's model. To get our idea across, hence we assume the observed data follow the following varying-coefficient proportional hazard regression model:

$$\lambda(t|Z_i, V_i) = \lambda_0(t) \exp\{\beta(V_i)Z_i + g(V_i)\}, \quad (3)$$

where $\beta(\cdot)$, an unknown smooth function, is the BATE curve, and $g(\cdot)$ is an unknown smooth function. Here $\lambda(t|Z_i, V_i)$ is the conditional hazard function of the observed event time T , given Z_i and V_i . Next we give a detailed discussion on the use of the BATE curve $\beta(v)$.

First, the BATE curve $\beta(\cdot)$ can be used graphically to identify a subset of patients who will not benefit or even may be harmed by the new treatment, and define the cutoff points based on different clinical significant levels in order to identify sensitive patients for the new treatment. For example, a monotonically increasing BATE curve implies that patients with lower biomarker levels are more likely benefited from the new therapy, see Figure 1(A). In Figure 1, we let $\alpha = -\log(8) \approx -2.08$, which means that the risk of the event for patients in the control group is eight times as large as the risk in the experimental group. Without loss of generality, we also assume that the range of the biomarker is $[0, 1]$. In this situation, the constant c_α such that $\beta(c_\alpha) = \alpha$, will be the standard cutoff point, and patients with the biomarker levels falling into $[0, c_\alpha]$ should be the sensitive patients. Conversely, a monotonically decreasing BATE curve implies that patients with higher biomarker levels are more likely benefited from the new therapy, see Figure 1(B). We choose the cutoff point, c_α , such that $\beta(c_\alpha) = \alpha$, and patients with the biomarker levels in $[c_\alpha, 1]$ should be the sensitive patients. For a non-monotonically BATE curve, such as the one in Figure 1(C), patients with medium biomarker levels are more likely benefited from the new therapy. Note that there are two cutoff points in this case. They are c_α^1 and c_α^2 such that $\beta(c_\alpha^1) = \beta(c_\alpha^2) = \alpha$. Hence patients with biomarker levels in $[c_\alpha^1, c_\alpha^2]$ should be sensitive patients.

[Figure 1 about here.]

Secondly, researchers often want to know if a biomarker is good enough as a predictive marker to distinguish sensitive patients from nonsensitive patients for the new treatment. This amounts to a hypothesis testing that $\beta(\cdot) \equiv C$. In this regard, if $\beta(\cdot) \equiv C$, then the corresponding biomarker can not predict the differential responses of individual patients to the new treatment. In this paper, we test the null hypothesis of $\beta(\cdot) \equiv C$ based on the asymptotic distributions of the normalized maximum deviations of the estimated BATE curves from the true BATE curves. We introduce two methods for solving this problem. One is by constructing a test statistics, and the other one is by constructing simultaneous confidence bands.

Thirdly, researchers are interested in comparing the clinical utility of different biomarkers based on their capacities for predicting responses to a treatment. However, the definition (2) may not be appropriate for such a comparison due to potentially different measurements of different biomarkers. To compare different biomarkers, we may give a slightly modified definition of the BATE curve at time t based on the quantile of the biomarker:

$$\psi(p, t) = \log \left\{ \frac{\lambda(t|Z = 1, V = \pi_p)}{\lambda(t|Z = 0, V = \pi_p)} \right\}, \quad (4)$$

where π_p is the p th quantile of the biomarker V , that is $\pi_p = \inf\{y : F_v(y) \leq p\}$. Here, $F_v(\cdot)$ is the cumulative distribution function of V . Similarly, in the special case that the BATE curve is independent with time, we have $\psi(p, t) \equiv \psi(p)$ for any $t \in (0, \tau)$.

Since $\psi(p)$ is defined on the same scale between 0 and 1, we can compare the BATE curves of two biomarkers to identify the more powerful one; that is we can distinguish “weak” and “strong” predictive biomarkers. Take Figure 2 as an example. A weak predictive biomarker is the biomarker whose BATE curve is near a constant (see the dashed line in Figure 2).

[Figure 2 about here.]

3. THE BATE CURVES OF A MULTIPLE-ARM RANDOMIZED CLINICAL TRIAL WITH FULL COMPLIANCE

In Section 2 when there is one treatment arm, we proposed to use the BATE curve to select sensitive patients for the treatment. When multiple treatments are involved in a randomized clinical trial,

we further extend the use of the BATE curve to select the optimal treatment for an individual patient. Next, we give detailed discussions on this extension .

3.1 Notation and Definitions

We first give some necessary notations. Let $\mathbf{Z} = (Z_1, \dots, Z_K)^T$ be a vector of treatment group indicators, where $Z_k = 1$ for the k th treatment group and $Z_k = 0$ otherwise, $k = 1, \dots, K$, $K > 1$. For $i = 1, \dots, n$, we let $T_i(k)$ be the event time of an individual that would be observed if the individual had received the k th treatment arm, where $k = 1, \dots, K$, and $T_i(0)$ be the event time that would be observed if the individual had received the control arm. We also define the potential conditional hazard rate functions as

$$\begin{aligned}\lambda^{(k)}(t|v) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_i(k) \leq t + \Delta t | T_i(k) > t, V_i = v)}{\Delta t}, \quad k = 1, \dots, K, \\ \lambda^{(0)}(t|v) &= \lim_{\Delta t \rightarrow 0} \frac{P(t < T_i(0) \leq t + \Delta t | T_i(0) > t, V_i = v)}{\Delta t}.\end{aligned}$$

Hence the BATE curve, at time t , of the k th treatment versus the control , based on the multiplicative interaction, is defined as follows:

$$\log \left\{ \frac{\lambda^{(k)}(t|v)}{\lambda^{(0)}(t|v)} \right\} = \beta_k(v, t). \quad (5)$$

Similarly to the discuss in Section 2, in a randomized clinical trial with full compliance, we can express the BATE curve, at time t , of the k th treatment as

$$\log \left\{ \frac{\lambda(t|V = v, \mathbf{Z} = \mathbf{e}_k)}{\lambda(t|V = v, \mathbf{Z} = \mathbf{0}_K)} \right\} = \beta_k(v, t), \quad (6)$$

where $\lambda(t|V = v, \mathbf{Z} = \mathbf{e}_k)$ is the conditional hazard function of the observed survival time T , given $V = v$ and $\mathbf{Z} = \mathbf{e}_k$.

We can also define the modified definition of the BATE curve at time t based on the quantile of the biomarker as

$$\log \left\{ \frac{\lambda(t|V = \pi_p, \mathbf{Z} = \mathbf{e}_k)}{\lambda(t|V = \pi_p, \mathbf{Z} = \mathbf{0}_K)} \right\} = \psi_k(p, t). \quad (7)$$

Since we are focus on the special case that the BATE curve for each treatment arm is independent with time, we have $\beta_k(v, t) \equiv \beta_k(v)$ and $\psi_k(v, t) \equiv \psi_k(v)$, for each $k = 1, \dots, K$ and for any

$t \in (0, \tau)$. Consequently, we assume the following varying-coefficient proportional hazard regression model for event time:

$$\lambda(t|Z_i, V_i) = \lambda_0(t) \exp\{\boldsymbol{\beta}^T(V_i)Z_i + g(V_i)\}, \quad (8)$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \dots, \beta_K(\cdot))^T$ denotes the vector of the BATE curves. Apparently, the model (3) is a special case of the model (8). This model has been studied by Fan, *et al.* (2006) and Cai, *et al.* (2007). They proposed the estimates of coefficient functions by local partial likelihood and established their pointwise asymptotic normalities. Their results, however, can only be used to construct pointwise confidence intervals but cannot be used to construct simultaneous confidence bands.

Since $K > 1$, we may also want to compare different treatments in order to find the optimal treatment for individual patients. This amounts to the problem of comparing the estimated BATE curves. We take Figure 3 ($K = 3$) as an example, which is also under the assumption that the range of the biomarker is $[0, 1]$. From Figure 3, we see that $\beta_3(v) < \beta_2(v) < \beta_1(v) < 0$ when $v \in [0, a]$, that $\beta_2(v) < \beta_3(v) < \beta_1(v) < 0$ when $v \in (a, b]$, and that $\beta_1(v) < \beta_2(v) < \beta_3(v) < 0$ when $v \in (b, 1]$. Hence, we would assign patients with the biomarker values in $[0, a]$, $(a, b]$ and $(b, 1]$ to receive Treatment 3, Treatment 2 and Treatment 1, respectively.

[Figure 3 about here.]

4. ESTIMATION AND ASYMPTOTIC PROPERTIES

Since the model (3) is a special case of the model (8), in this section we focus on estimation for the model (8).

4.1 Estimation of BATE Curves and Cutoff Points

We are using a partial likelihood function with local polynomial (linear) fitting to estimate $\boldsymbol{\beta}(v)$ and $g(v)$.

For any given point v_0 , by Taylor series expansions, we obtain that

$$\begin{aligned} \boldsymbol{\beta}(v) &\approx \boldsymbol{\beta}(v_0) + \boldsymbol{\beta}'(v_0)(v - v_0) \equiv \boldsymbol{\delta} + \boldsymbol{\gamma}(v - v_0), \text{ and} \\ g(v) &\approx g(v_0) + g'(v_0)(v - v_0) \equiv a + b(v - v_0). \end{aligned} \quad (9)$$

Let $\mathcal{R}(t)$ denote the set of the individuals at risk prior to time t and assume the observations are independent. Using (9) for the data around v_0 and utilizing a kernel function, we obtain the following logarithm of the local partial-likelihood:

$$\begin{aligned} & \ell_n(\delta, \gamma, b; v_0) \\ &= \frac{1}{n} \sum_{i=1}^n K_h(V_i - v_0) \Delta_i \left\{ \delta^T Z_i + \gamma^T Z_i (V_i - v_0) + b(V_i - v_0) \right. \\ & \quad \left. - \log \left(\sum_{j \in \mathcal{R}(X_i)} K_h(V_j - v_0) \exp \{ \delta^T Z_j + \gamma^T Z_j (V_j - v_0) + b(V_j - v_0) \} \right) \right\}, \end{aligned} \tag{10}$$

where $K_h(\cdot) = K(\cdot/h)/h$ is a symmetric kernel function, and h is a bandwidth. Suppose that (10) is maximized at $(\hat{\delta}, \hat{\gamma}, \hat{b})$. Then $\hat{\beta} = \hat{\delta}$ is a local linear estimator of the coefficient function $\beta(\cdot)$ at the point v_0 . An estimator of $g'(\cdot)$ at v_0 is simply the local slope $\hat{b}(v_0)$, i.e., $\hat{g}'(v_0) = \hat{g}$. The estimated curve $\hat{g}(\cdot)$ can be obtained by integrating $\hat{g}'(v_0)$, using the Trapezoidal rule (Hastie and Tibshirani, 1990). For the purpose of ensuring the identifiability of $g(\cdot)$, we set $g(0) = 0$ without loss of generality.

Next, we consider the estimation of the modified BATE curve $\psi(p) = (\psi_1(p), \dots, \psi_K(p))^T$. Note that $\psi(p) = \beta(\pi_p)$. Hence, based on $\hat{\beta}(\cdot)$, it remains to estimate the quantile π_p of the continuous biomarker V . The estimator $\hat{\pi}_p$ is given by $\inf\{y : \hat{F}_v(y) \geq p\}$, where $\hat{F}_v(\cdot)$ denotes the empirical estimator of $F_v(\cdot)$. As a result, the estimator of $\psi(p)$ has the following form

$$\hat{\psi}(p) = \hat{\beta}(\hat{\pi}_p) = (\hat{\beta}_1(\hat{\pi}_p), \dots, \hat{\beta}_K(\hat{\pi}_p))^T.$$

Since $\hat{\pi}_p$ is a unique sample quantile, according to Koenker and Bassett (1978), if $F_v(\cdot)$ is continuous and has a continuous and positive density, $f_v(\cdot)$, at π_p , then $\sqrt{n}(\hat{\pi}_p - \pi_p)$ converges to a mean zero normal distribution, with the variance $\omega^2 = p(1-p)/f_v^2(\pi_p)$.

Given a clinical significant level α , since $\beta(\cdot)$ is a vector of smooth functions, and since $F_v(\cdot)$ is continuous and has a continuous and positive density, we can directly select the tolerant cutoff points $c_{\alpha,i}$ such that $\beta_i(c_{\alpha,i}) = \alpha$, for the i th treatment arm.

Note that, for $i = 1, \dots, K$, the tolerant cutoff points $c_{\alpha,i}$ may not be unique due to the potential non-monotonicity of $\beta_i(\cdot)$. Hence we select the cutoff points as follows:

$$c_{\alpha,i} = \{c : \hat{\beta}_i(c) = \alpha\}, \text{ and } p_{\alpha,i} = \{p : \hat{\psi}_i(p) = \alpha\},$$

$i = 1, \dots, K$. Obviously, for $i = 1, \dots, K$, we can also get $p_{\alpha,i}$ by letting $p_{\alpha,i} = \hat{F}_p(c_{\alpha,i})$.

4.2 Asymptotic Properties

To obtain asymptotic properties of the maximum partial likelihood estimator $\widehat{\beta}_1(v)$, we need a few notations on the counting process. Let $N_i(t) = I(T_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$. Define $\mathcal{F}_{t,i}$ to be the failure, censoring and covariate information up to time t , and let $M_i(t) = N_i(t) - Y_i(t)\lambda(t|Z_i, V_i)$ be the $\mathcal{F}_{t,i}$ -martingale. Let $\mu_i = \int v^i K(v)dv$, and $\nu_i = \int v^i K^2(v)dv$, $i = 1, 2, \dots$. Let $\beta_0(\cdot)$ and $g_0(\cdot)$ be the true functions of $\beta(\cdot)$ and $g(\cdot)$, respectively. Denote

$$\rho(u, \mathbf{z}, v_0) = P(X \geq u | \mathbf{z}, v_0) \exp\{\beta_0^T(v_0)\mathbf{z} + g_0(v_0)\}.$$

For $k = 0, 1, 2$ and $l = 0, 1, 2$, we define

$$a_{kl}(u, v_0) = f(v_0)\mu_l E\{\rho(u, \mathbf{Z}, v_0)\mathbf{Z}^{\otimes k} | V = v_0\},$$

where $f(\cdot)$ is the density of V and $\mathbf{Z}^{\otimes k} = 1, \mathbf{Z}$ and $\mathbf{Z}\mathbf{Z}^T$ for $k = 0, 1$ and 2 , respectively.

PROPOSITION 1: *Under Conditions in the Appendix, we can obtain the following results:*

(1). As $n \rightarrow \infty$,

$$\sup_{v \in \mathcal{V}} |(\widehat{\beta}(v) - \beta_0(v))| \xrightarrow{\mathcal{P}} 0,$$

(2). For each $v_0 \in \mathcal{V}$,

$$\sqrt{nh}(\widehat{\beta}(v_0) - \beta_0(v_0) - \frac{\mu_2}{2}h^2\beta''(v_0)) \xrightarrow{\mathcal{L}} N(0, \nu_0\Sigma^{-1}(v_0)).$$

Here \mathcal{V} is the compact support of the biomarker V , and $\beta''(v_0)$ is the second derivative of $\beta(v)$ at v_0 and

$$\begin{aligned} \Sigma(v_0) = & \int_0^\tau \{a_{20}(u, v_0) \\ & - a_{10}(u, v_0)a_{10}^T(u, v_0)a_{00}^{-1}(u, v_0)\lambda_0(u)\} du. \end{aligned}$$

Proof of Proposition 1 can be found in Fan, *et al.* (2006), and is omitted here for saving some space.

Let $\psi_0(\cdot)$ denote the true function of $\psi(\cdot)$. Next, we state the pointwise asymptotic distribution of $\widehat{\psi}(p)$,

Theorem 1 Under Conditions in the Appendix, if $f_v(\cdot)$ is continuous and positive at π_p , we can obtain the following asymptotic normality result:

$$\sqrt{nh}(\widehat{\psi}(p) - \psi_0(p) - \frac{\mu_2}{2} h^2 \boldsymbol{\beta}''(\pi_p)) \xrightarrow{\mathcal{L}} N(0, \nu_0 \Sigma^{-1}(\pi_p)).$$

The proof of Theorem 1 can be directly obtained by Proposition 1 and the fact that $\widehat{\pi}_p$ is \sqrt{n} -consistent.

5. HYPOTHESIS TESTING BASED ON ASYMPTOTIC DISTRIBUTIONS OF THE NORMALIZED MAXIMUM DEVIATIONS

As discussed in Section 2, the usefulness of a predictive biomarker can be tested, based on a null hypothesis that the BATE curve is a constant. In this section we propose two methods to test this null hypothesis. The basic idea is using the asymptotic distribution of the normalized maximum deviations of the estimated BATE curves from the true BATE curves.

5.1 Asymptotic Distributions of the Normalized Maximum Deviations

For any function $b(v)$ and any matrix $B(v) = (b_{ij}(v))_K$, we define $\|b\|_\infty = \sup_{v \in [0,1]} |b(v)|$ and $\|B\|_\infty = \left(\sum_{i=1}^K \sum_{j=1}^K \|b_{ij}\|_\infty \right)^{1/2}$. We now give our main results below.

Theorem 2 Assume that Conditions in the Appendix hold, if $[0, 1] \subseteq \mathcal{V}$, and $h = n^{-d}$, $1/5 \leq d < 1 - 2/s$, $s > 2$, then we have for $k = 1, \dots, K$,

$$\begin{aligned} P\{ \{2 \log(1/h)\}^{1/2} \left(\left\| \frac{\widehat{\beta}_k - \beta_k - \text{bias}(\widehat{\beta}_k)}{\sigma_k} \right\|_\infty - d_n \right) < x \} \\ \rightarrow \exp\{-2 \exp(-x)\}, \end{aligned} \tag{11}$$

where $\sigma_k^2(v)$ is the k th diagonal element of $\nu_0 \Sigma(v)$, $\text{bias}(\widehat{\beta}_k(v)) = e_{k,K}^T \frac{\mu_2}{2} h^2 \boldsymbol{\beta}''(v)$, $e_{k,K}$ is a K -dimensional vector with the k th element 1 and else 0. Here

$$d_n = \{2 \log(1/h)\}^{1/2} + \frac{1}{\{2 \log(1/h)\}^{1/2}} \log \left\{ \frac{\int (K'(v))^2 dv}{4\nu_0 \pi} \right\}.$$

If the supremum in Theorem 2 is taken on an interval $[a, b]$ instead of $[0, 1]$, then by replacing $\log(1/h)$ in (11) and d_n with $\log((b-a)/h)$, Theorem 2 continues to hold. The proof of Theorem 2 is given in Appendix C.

5.2 Hypothesis Testing of the BATE Curves

Researchers often want to know if a biomarker is good enough as a predictive marker to distinguish sensitive patients from nonsensitive patients for one treatment of interest, or if one particular treatment is noneffective for all patients. This amounts to testing whether its BATE curve is equivalent to a non-zero constant or zero. Without loss of generality, suppose that we are interested in the k th ($1 \leq k \leq K$) treatment, and we consider the hypothesis problem over $[0, 1]$. Hence we state the null and alternative hypothesis as follows:

$$H_0 : \beta_k(v) = C \text{ v.s } H_1 : \beta_k(v) \neq C \quad (12)$$

or

$$H_0^* : \psi_k(p) = C \text{ v.s } H_1^* : \psi_k(p) \neq C. \quad (13)$$

Under the null hypothesis H_0 , the model (8) is a semiparametric varying coefficient Cox's hazard ratio model. We are estimating C using a profile method.

Note that $\sigma_k(v)$ and $bias(\hat{\beta}_k)$ in (11) are unknown, we can not directly use Theorem 2 to construct test statistics. Hence it's required to first estimate the bias and variance of $\hat{\beta}_k(v)$, $k = 1, \dots, K$. From the expressions for the asymptotic bias, $bias(\hat{\beta}(v))$, and variance, $\Sigma(v)$, of $\hat{\beta}(v)$ in Appendix A-C, we can obtain their consistent estimators $\widehat{bias}(\hat{\beta}(v))$ and $\widehat{\Sigma}(v)$, which are also given in Appendix A-C before the proof of Theorem 3.

Theorem 3 *Under the conditions of Theorem 2, if the third derivative of $\beta(\cdot)$ is continuous on $[0, 1]$, then we have for $k = 1, \dots, K$,*

$$P\{\{2 \log(1/h)\}^{1/2} \left(\left\| \frac{\hat{\beta}_k - \beta_k - \widehat{bias}(\hat{\beta}_k)}{\widehat{\sigma}_k} \right\|_\infty - d_n \right) < x\} \\ \rightarrow \exp\{-2 \exp(-x)\} \quad , \quad (14)$$

where $\widehat{\sigma}_k^2(v)$ is the k th diagonal element of $\nu_0 \widehat{\Sigma}(v)$ and $\widehat{bias}(\hat{\beta}_k(v))$ is the k th element of $\widehat{bias}(\hat{\beta}(v))$.

Next, we establish the the asymptotic distribution of the maximum deviation of $\hat{\psi}(p)$.

Theorem 4 *Under the conditions of Theorem 2, if the third derivative of $\beta(\cdot)$ is continuous on $[0, 1]$ and $f_v(\cdot)$ is continuous and positive, then we have for $k = 1, \dots, K$,*

$$P\{\{2 \log(1/h)\}^{1/2} \left(\left\| \frac{\hat{\psi}_k(p) - \psi_k(p) - \widehat{bias}(\hat{\psi}_k(p))}{\widehat{\sigma}_k} \right\|_\infty - d_n \right) < x\} \\ \rightarrow \exp\{-2 \exp(-x)\} \quad , \quad (15)$$

where $\widehat{bias}(\widehat{\psi}_k(p))$ is same as $\widehat{bias}(\widehat{\beta}(\pi_p))$.

The proofs of Theorem 3 and Theorem 4 are given in Appendix C.

According to Theorem 3 and Theorem 4, we can test the null hypothesis, defined by (12) (or (13)), by calculating the following test statistics

$$\{2 \log(1/h)\}^{1/2} (\| \frac{\widehat{\beta}_k - \widehat{C} - \widehat{bias}(\widehat{\beta}_k)}{\widehat{\sigma}_1} \|_\infty - d_n),$$

or

$$\{2 \log(1/h)\}^{1/2} (\| \frac{\widehat{\psi}_k - \widehat{C} - \widehat{bias}(\widehat{\psi}_k)}{\widehat{\sigma}_1} \|_\infty - d_n)$$

and rejecting H_0 (or H_0^*) when test statistic exceeds the asymptotic critical value $v_{\alpha^*} = -\log(-0.5 \log(1 - \alpha^*))$ for the nominal level α^* .

An alternative method for the hypothesis testing problem (12) (or (13)) is to see whether \widehat{C} falls inside the simultaneous confidence band of $\beta_k(v)$ (or $\psi_k(p)$), which can also be constructed based on the asymptotic distribution of the maximum deviations. It is followed from Theorem 3 and Theorem 4 that for $k = 1, \dots, K$, the $(1 - \alpha^*)\%$ confidence bands of $\widehat{\beta}_k(v)$ and $\widehat{\psi}_k(p)$ on $[0, 1]$ is

$$(\widehat{\beta}_k(v) - \widehat{bias}(\widehat{\beta}_k(v)) \pm q_{n,k}(v)), \tag{16}$$

and

$$(\widehat{\psi}_k(p) - \widehat{bias}(\widehat{\psi}_k(p)) \pm q_{n,k}(\pi_p)), \tag{17}$$

respectively, where $q_{n,k}(v) = (d_n + [\log 2 - \log\{-\log(1 - \alpha^*)\}](2 \log(1/h))^{-1/2})\widehat{\sigma}_k(v)$.

Remark 1 Note that, by using the MSE optimal bandwidth h_{opt} , whose converge rate is $O_p(n^{-1/5})$, we obtain a asymptotically biased estimator, $\widehat{\beta}_k(v)$, of $\beta_k(v)$. Although we can propose a consistent estimate of the asymptotic bias of $\widehat{\beta}_k(v)$, it is unstable. According to Theorem 2, we can choose a bandwidth which converges to 0 at a slightly faster rate than the optimal bandwidth rate $O_p(n^{-1/5})$ to obtain an asymptotically unbiased estimate. In the simulation studies, we ignore the asymptotic biases based on the two-stage method of Fan and Zhang (2000): We first select the optimal bandwidth h_{opt} for estimating the coefficient functions, and then use $h_{opt}/2$ for calculating test statistics or constructing confidence bands.

Remark 2 *It's well known that based on the asymptotically unbiased estimator by choosing a bandwidth which converges to 0 at a slightly faster rate than the optimal bandwidth rate, the test statistics' approximations are not very accurate (Hall 1993). Tian, Zucker and Wei (2005) argued that by employe an alternative strategy based on the Lin, Fleming and Wei (1994) stochastic perturbation technique, we can obtain a more accurate approximation to the distribution of the maximum deviations of $\widehat{\beta}_k(\cdot)$ from $\beta_k(\cdot)$. However, our simulations studied in the next section show that by using the two-stage method, we can also obtain accurate approximations.*

6. SIMULATION STUDIES

In this section, we carried out two simulation studies to examine the finite-sample properties of the proposed methods. The first simulation study assessed the performance of the proposed estimators for the BATE curves, and the second simulation study evaluated the type I error rate of the proposed test for the null hypothesis that some BATE curves are a constant. In both simulation studies, the sample size was set to be 500, and the number of the simulation was chosen to be 500.

In the first simulation study, we generated the failure time data from the following varying-coefficient hazard regression model: $\lambda(t|V, Z_1, Z_2) = \lambda_0(t) \exp\{\beta_1(V)Z_1 + \beta_2(V)Z_2 + g(V)\}$, where $\beta_1(v) = 3v(1.5 - v)$, $\beta_2(v) = 2 \sin(3v)$, $g(v) = v^2$ and $\lambda_0(t) = 3t^2/5$. We then generated V from a normal distribution with mean 1 and variance 1/9. Covariates Z_1 and Z_2 were generated form multiple Bernoulli random variables $(z_1, z_2, z_1 + z_2)$, and each component took value 0 or 1 with probability of being 1 as 0.3, 0.35 and 0.65, respectively. We generated the censoring time C from an exponential distribution with mean cc , where cc was chosen to yield a approximately censoring rate (CR) of 20% and 40%, respectively. For each censoring rate, we also calculated the standard deviations (SDs), the average standard errors (SEs), and the 95% confidence interval coverage rates (CPs) of $\widehat{\beta}_1(v)$, $\widehat{\beta}_2(v)$ and $\widehat{\psi}_1(p)$, $\widehat{\psi}_2(p)$. Simulation results are represented in Table 1 and Figure 4-5. The results show that the proposed estimators perform well.

[Figure 4 about here.]

[Figure 5 about here.]

[Table 1 about here.]

In the second simulation study, we focus on the performance of the proposed hypothesis test for the first null hypothesis that the functional coefficient $\beta_1(\cdot)$ is a constant. That is we consider the null hypotheses: $H_0 : \beta_1(\cdot) \equiv C$, which implies that the causal effect of the first treatment versus control does not depend on the value of the biomarker. We generated failure times from the hazard regression model $\lambda(t|V, Z_1, Z_2) = 0.6t^2 \exp\{\beta_1(V)Z_1 + \beta_2(V)Z_2 + V^2\}$. To evaluate the proposed method, we carried out several different simulation studies. We let $\beta_1(v) = -1$ and $\beta_2(v) = 2 \sin(3v), -\exp(v), 0$, respectively. For each case, we calculated the type I error rates based on 500 simulations and 500 sample size, for the significant level $\alpha^* = 0.05, 0.01$, respectively. Results are reported in Table 2. For each case, we chose cc to yield an approximately censoring rate of 20%, and generated covariates Z_1, Z_2 and V in the same way as in the first simulation study.

[Table 2 about here.]

In the both simulation studies, we adapted the method of Fan and Huang (2005) to select the optimal bandwidth \hat{h}_{opt} , which was found to be round 0.113 in our setting, for estimating the coefficient functions, and then as discussed in Remark 1, we used $\hat{h}_{opt}/2$ for calculating test statistics or constructing confidence bands (Fan and Zhang 2000).

From Table 2, we see that by using the two-stage method for calculating test statistics or constructing confidence bands, type I error rates are very close to their corresponding significant levels α^* .

7. EXAMPLE

We illustrate the application of the proposed method in a real-world clinical study on the role of *c-myc* in selecting the optimal treatment for patients with colon cancer (Augenlicht (1997)). In this clinical trial, patients with colon cancer can be treated by surgery alone or surgery plus chemotherapy. Surgery alone is less invasive and less expensive than surgery plus chemotherapy. It is desirable to identify the patients who may benefit more from surgery plus chemotherapy based on their biomarkers. Based on a study conducted by the Eastern Cooperative Oncology Group (ECOG), Augenlicht (1997) suggested that the *c-myc* oncogene may be a predictive biomarker for patients with colon cancer. Using a subset of the cases from this clinical trial, Li and Ryan (2006)

found that there might be an interaction between the *c-myc* oncogene expression level and the two treatments on overall survival and disease progression free survival.

In this section, we apply our method to this data set to assess the role of *c-myc* oncogene in predicting response to treatment and distinguish sensitive patients from nonsensitive patients based on their *c-myc* levels. Disease progression free survivals of a total of 124 patients randomized to receive surgery alone or surgery plus chemotherapy or other treatments were measured. We let V_i be the *c-myc* oncogene expression level of the i th patient, Z_{i1} and Z_{i2} be the indicators of receiving other treatments and receiving surgery plus chemotherapy, respectively. We consider the following model:

$$\lambda(t|Z_i, V_i) = \lambda_0(t) \exp\{\beta_1(V_i)Z_{i1} + \beta_2(V_i)Z_{i2} + g(V_i)\}. \quad (18)$$

We fit the model (18) for the progression free survival. The 95% confidence limits and 95% confidence band are calculated by a sandwich method. The Gaussian kernel is employed, and the bandwidth is chosen to be 25% of the interval length. Results are presented in Figure 6 and Figure 7.

[Figure 6 about here.]

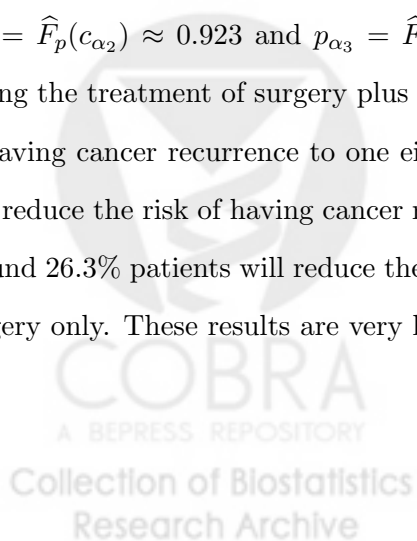
[Figure 7 about here.]

First we consider the hypothesis problems of whether the BATE curves or modified BATE curves are constants. We perform the following four hypothesis tests. $H_{01} : \beta_1(v) \equiv 0$; $H_{02} : \beta_1(v) \equiv C_1$; $H_{03} : \beta_1(v) \equiv 0$; and $H_{04} : \beta_1(v) \equiv C_2$, where C_1 and C_2 are unspecified constants. We obtain the p -value of each null hypothesis as follows: $p_1 = 0.056$, $p_2 = 0.002$, $p_3 = 0.052$ and $p_4 = 0.005$, respectively. The results show (1) that the effect of the treatment arm of surgery only versus the other treatment arm does not depend on the biomarker *c - myc* value and (2) that the treatment effect of the treatment arm of surgery plus chemotherapy versus the surgery varies with patients' *c - myc* levels.

As we discussed in Section 5, we can also test the null hypothesis problems by constructing confidence bands of the estimators of $\beta_1(v)$ and $\beta_2(v)$, and the resulting confidence bands are displayed in Figure 6 and Figure 7. From Figure 6, we can see that the constant curve 0 falls between

the upper and lower confidence bands of the BATE or modified BATE curve of the treatment arm of other treatments. This result is same as the conclusion (1) as above, which also implies that the treatment effect of the other treatment versus the surgery only does not depend patient's $c - myc$ levels. From Figure 7, we can see that, for the treatment arm of surgery plus chemotherapy, we reject the null hypothesis that the treatment effect of the surgery plus chemotherapy versus surgery only is the same, regarding less patient's $c - myc$ values. This conclusion implies that the biomarker $c - myc$ is capable of predicting the treatment effect of surgery plus chemotherapy, selecting cutoff points and distinguishing sensitive patients from nonsensitive patients.

Next, for the treatment arm of surgery plus chemotherapy, we consider three different clinical significant levels: $\alpha_1 = -\log(8) \approx -2.08$, $\alpha_2 = -\log(4) \approx -1.39$ and $\alpha_3 = -\log(2) \approx -0.69$, which imply that the risks of having cancer recurrence for patients in control group (surgery only) are eight times, four times and twice, respectively, as large as the risk in the treatment group of surgery plus chemotherapy. The cutoff points for the chosen clinical significant levels are shown in Figure 7. From Figure 7(A), we see that a patient with $c - myc$ level no less than the cutoff point $c_{\alpha_1} \approx 3.254$ should be treated by surgery plus chemotherapy, if this patient is willing to take at most one eighth risk of having cancer recurrence, as large as being treated by surgery only. Similarly, a patient with $c - myc$ level no less than the cutoff point $c_{\alpha_2} \approx 2.689$ ($c_{\alpha_3} \approx 1.716$) should be treated by surgery plus chemotherapy, if one-fourth (a half) risk of having cancer recurrence can be acceptable. Based on the quantile of $c - myc$, we also get the cutoff points for each clinical significant levels, see Figure 7(B). From Figure 7(B), we can see that the cutoff points are $p_{\alpha_1} = \widehat{F}_p(c_{\alpha_1}) \approx 0.955$, $p_{\alpha_2} = \widehat{F}_p(c_{\alpha_2}) \approx 0.923$ and $p_{\alpha_3} = \widehat{F}_p(c_{\alpha_3}) \approx 0.737$, respectively. Then we can conclude that, if taking the treatment of surgery plus chemotherapy, only around 4.5% patients will reduce the risk of having cancer recurrence to one eighth of the risk if taking surgery only, around 7.7% patients will reduce the risk of having cancer recurrence to one fourth of the risk if taking surgery only, and around 26.3% patients will reduce the risk of having cancer recurrence to a half of the risk if taking surgery only. These results are very helpful for making medical policies.



8. DISCUSSION

In this paper, we have introduced a new concept, the BATE curve, to represent the predictive ability of a biomarker in selecting patients who respond better to one particular treatment over another treatment, called sensible patients, when patient's outcome is time-to-event. Compared to the existing methods, the BATE curve not only visually displays the treatment effect on the patient's outcome as a function of the biomarker value but also allows one to compare the relative performance of different predictive biomarkers that may have different scales. We have also proposed semi-parametric time-varying coefficient regression methods for estimating the BATE curves and their confidence bands.

The proposed estimation methods have the following technical advantages over the existing methods. (1) By assuming the interaction between a continuous-scale biomarker and the treatment group is a continuous function of the biomarker, the proposed varying coefficient models are flexible enough to indicate both local and global association between treatment and the biomarker value. (2) We propose a new semi-parametric test for the null hypothesis that the entire BATE curve of a biomarker is constant; that is, the biomarker does not have predictive power in selecting sensible patients to one particular treatment.

Although this paper has focused on multiplicative interaction to represent the treatment effect when the patient's outcome is time-to-event, the proposed methods can be easily extended to the additive interaction between the treatment groups and biomarkers. In fact, we have already developed the BATE curve and associated inference procedures for the additive interactions. Since in the survival analysis, the multiplicative interaction effects, which describes derivations from multiplicative joint effects, are commonly used, due to the limited space in this paper, we have only reported the results for multiplicative interaction effects. The results on the BATE curves, based on additive interactions, will be reported in a technical report.

ACKNOWLEDGEMENT

This work was partially supported a grant from U.S. Department of Veterans Affairs, Veterans Affairs Health Administration, HSR&D grant ECI-03-206. This paper presents the findings and conclusions of the authors. It does not necessarily represent those of VA HSR&D Service. The

authors are also grateful to Dr. Yi Li for the data analyzed in the paper.

REFERENCES

- Augenlicht, L. (1997). Low-level c-myc amplification in human colonic carcinoma cell lines and tumors: a frequent, p53 independent, mutation associated with improved clinical outcome in a randomized multiinstitutional trial. *Cancer Res.* **57**. 1769–1775.
- Bonetti, M. and Gelber, R.D. (2000). A graphical method to assess treatment-covariate interactions using the Cox model on subsets of the data. *Statist. Med.* **19**. 2595–2609.
- Cai, J., Fan, J., Jiang, J. and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *J. Am. Statist. Assoc.* **102**. 538–551.
- Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. B.* **57**. 371–394.
- Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11**. 1031–1057.
- Fan, J., Lin, H. Z. and Zhou, Y. (2006). Local partial likelihood estimation for lifetime data. *Ann. Statist.* **34**. 290–325.
- Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27**. 715–731.
- Freidlin, B. and Simon, R. (2005). Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clinical Cancer Research* **11**. 7872–7878.
- Greenland, S. (1993). Basic problems in interaction assessment. *Environ. Health Persp.* **101**. 59–66.
- Hall, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *J. Roy. Statist. Soc. B.* **55**. 291–304.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman & Hall.

Jiang, W., Freidlin, B. and Simon, R. (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J. Natl. Cancer. Inst.* **99**. 1036–1043.

Koenker, R. and Bassett, G. W. (1978). Regression quantiles. *Econometrica* **46**. 33–50.

Li, Y. and Ryan, R. (2006). Inference on survival data with covariate measurement error - an imputation approach. *Scand. J. Statist.* **33**. 169–190.

Lin, D., Fleming, T. and Wei, L.J. (1994). Confidence bands for survival curves under the proportional hazards models. *Biometrika* **89**. 1354–1365.

Royston, P. and Sauerbrei, W. (2004). A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statist. Med.* **23**. 2509–2525.

Sargent, D. J. *et al.*, (2005). Clinical trial designs for predictive marker validation in cancer treatment trials. *J. Clin. Oncol.* **23**. 2020–2027.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6**. 177–184.

Tian, L., Zucker, D. and Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Am. Statist. Assoc.* **100**. 172–183.

Vansteelandt, S., VanderWeele, T., Tchetgen, E. J. and Robins, J. M. (2008). Multiply robust inference for statistical interaction. *J. Am. Statist. Assoc.* **103**. 1693–1704.

APPENDIX A: ASSUMPTIONS

We list the following assumptions in the Appendix for our results.

1. For $s > 2$, $E|Z_j|^{2s} < \infty$, $j = 1, \dots, K$.
2. The biomarker V has a compact support \mathcal{V} , in which $f_v(v)$ is continuous, and $\inf_{v \in \mathcal{V}_\varepsilon} f_v(v) > 0$ for some $\varepsilon > 0$, where $\mathcal{V}_\varepsilon = \{v : \inf_{v_0 \in \mathcal{V}} |v - v_0| \leq \varepsilon\}$.

3. Let $\Omega(v) = \text{diag}(\Sigma(v), \Gamma(v))$, where

$$\Gamma(v) = \begin{pmatrix} \int_0^\tau a_{22}(u, v) d\Lambda_0(u) & \int_0^\tau a_{12}(u, v) d\Lambda_0(u) \\ \int_0^\tau a_{12}^T(u, v) d\Lambda_0(u) & \int_0^\tau a_{22}(u, v) d\Lambda_0(u) \end{pmatrix}.$$

Then $\Omega(v)$ is non-singular for any $v \in \mathcal{V}_\varepsilon$, $\Sigma(v)$ is positive definite for any $v \in \mathcal{V}_\varepsilon$, and the elements in $\Omega(v)$ are continuous on the compact support \mathcal{V} .

4. $E(Z_j^{2s}|V = v)$ is bounded for $v \in \mathcal{V}$, $j = 1, \dots, K$. $\int \lambda_0(t) dt < \infty$, and $\|f_v(v)\|_{\mathcal{V}} < \infty$, where $\|\cdot\|_{\mathcal{V}}$ is the sup-norm of a function on \mathcal{V} .

5. Functions $\beta(\cdot)$ and $g(\cdot)$ have continuous second derivatives on the compact support \mathcal{V} .

6. The kernel function $K(\cdot)$ is a bounded, symmetric density function and uniformly continuous.

Furthermore, $K(x) \rightarrow 0$ as $x \rightarrow \infty$, $\int |K(x)| dx < \infty$ and $\int K^2(x) dx < \infty$.

7. The conditional probability $P(X \geq u|\mathbf{z}, v)$ is equi-continuous in the arguments (u, v) on $[0, \tau] \times \mathcal{V}_\varepsilon$.

8. Let

$$s_k^*(u, \zeta, v_0) = f(v_0) \int E \left[P(X \geq u|\mathbf{z}, v_0) \Phi(\zeta, y) \mathbf{R}(y)^{\otimes k} | v = v_0 \right] K(y) dy,$$

where $k = 0, 1, 2$, $\mathbf{R}(y) = (\mathbf{Z}^T, \mathbf{Z}^T y, y)^T$, and

$$\Phi(\zeta, y) = \exp\{\zeta^T \mathbf{R}(y) + \beta^T(v) \mathbf{Z}\}.$$

We suppose that for $k = 0, 1, 2$, $s_k^*(u, \zeta, v)$ is bounded away from 0 on the product space $[0, \tau] \times \mathbf{C}_1 \times \mathcal{V}_\varepsilon$, where $\mathbf{C}_1 \in \mathbf{R}^{2K+1}$; that is

$$\inf_{u \in [0, \tau]} \inf_{\zeta \in \mathbf{C}_1} \inf_{v \in \mathcal{V}_\varepsilon} s_k^*(u, \zeta, v) > 0,$$

and

$$\sup_{(\beta^T, g) \in \mathbf{C}_2} E|\mathbf{Z}|^2 \exp\{\beta^T \mathbf{Z} + g\} < \infty,$$

where $\mathbf{C}_2 \in \mathbf{R}^{K+1}$

9. We have $nh/\log n \rightarrow \infty$ as $n \rightarrow \infty$, and nh^5 is bounded.

APPENDIX B: TECHNICAL LEMMAS

Before we prove the theorems, we present several lemmas first as follows. Lemma 1-lemma 3 are needed for the proofs of lemma 4, Theorem 2, Theorem 3, and Theorem 4. Lemma 4 is needed for the proof of Theorem 2.

Given $t \in [0, \tau]$, let $(\xi_1(t), V_1), \dots, (\xi_n(t), V_n)$ be an i.i.d. random sample from $(\xi(t), V)$. We assume that V and the kernel function $K(\cdot)$ satisfy the conditions in A.1. We further assume that ξ satisfies the following three conditions: (1) for an $s > 2$, $\sup_{t \in [0, \tau]} E|\xi(t)|^s < \infty$; (2) the function $\sigma^2(v)$ is bounded away from zero for $v \in [0, 1]$ and has a bounded first derivative on $[0, 1]$, where $\sigma^2(v) = E\{\int_0^\tau \xi^2(t)\lambda(t)dt|V = v\}$; and (3) $\sup_x \sup_{t \in [0, \tau]} \int |y(t)|^s f(y(t), x)dy = c_x < \infty$, where $f(y(t), x)$ is the joint density of $(\xi(t), V)$. Let

$$m(v) = \frac{1}{n}\sigma^{-1}(v)f^{-1/2}(v)\sum_{i=1}^n\int_0^\tau\xi_i(t)K_h(V_i - v)dM_i(t).$$

For the process $m(v)$, we have the following lemma 1

Lemma 1 *Suppose Assumptions 6-7 hold. If $h = n^{-b}$, for some $0 < b < 1 - 2/s$, we have*

$$\begin{aligned} &P\{(-2\log h)^{1/2}\left(\sqrt{nh}\nu_0^{-1/2}\|m(v)\|_\infty - d_n\right) < x\} \\ &\rightarrow \exp\{-2\exp\{-x\}\}. \end{aligned} \tag{19}$$

Proof. Note that

$$\begin{aligned} m(v) &= \frac{1}{n}\sigma^{-1}(v)f^{-1/2}(v)\sum_{i=1}^n\int_0^\tau\xi_i(t)K_h(V_i - v)dM_i(t) \\ &= \frac{1}{n}\sigma^{-1}(v)f^{-1/2}(v)\sum_{i=1}^nK_h(V_i - v)\int_0^\tau\xi_i(t)dM_i(t), \end{aligned}$$

with $M_i^*(s, v) = \int_0^s \xi_i(t)dM_i(t)$ being a \mathcal{F}_s martingale for $s \in (0, \tau]$ and $i = 1, \dots, n$. Then $M_i^*(\tau, v)$, $i = 1, \dots, n$, is an i.i.d. random variable sequence, $E|M_i^*(\tau, v)|^s < \infty$, $E\{M_i^*(\tau, v)\} = 0$, and $Var\{M_i^*(\tau, v)\} = \sigma^2(v)$ for any $v \in [0, 1]$. Hence, according to Lemma 1 of Fan and Zhang (2000), (19) follows.

Similarly, we define

$$m^*(v) = \frac{1}{n}\sigma^{*-1}(v)f^{-1/2}(v)\sum_{i=1}^n\int_0^\tau\xi_i(t)K_h(V_i - v)d\Lambda_i(t),$$

where $\sigma^{*2}(v) = Var\{\int_0^\tau \xi(t)\lambda(t)dt|V = v\}$, and $\sigma^{*2}(v)$ is bounded away from zero for $v \in [0, 1]$ and has a bounded first derivative on $[0,1]$. Hence for the process $m^*(v)$, we have a similar lemma to lemma 1, as stated as lemma 2 below.

Lemma 2 *Suppose Assumptions 6-7 hold. If $h = n^{-b}$, for some $0 < b < 1 - 2/s$, we have*

$$\begin{aligned} & P\{(-2 \log h)^{1/2} \left(\sqrt{nh} \nu_0^{-1/2} \|m^*(v) - E\{m^*(v)\}\|_\infty - d_n \right) < x\} \\ & \rightarrow \exp\{-2 \exp\{-x\}\}. \end{aligned} \quad (20)$$

The proof of lemma 2 can be obtained directly by the Lemma 1 of Fan and Zhang (2000).

Lemma 3 *Let $\theta(v_0) = (\delta^T, h\gamma^T, hb)^T$, $\mathbf{Z}_i^* = (\mathbf{Z}_i^T, \mathbf{Z}_i^T(V_i - v_0)/h, (V_i - v_0)/h)^T$, where $i = 1, \dots, n$.*

Let $\theta_0(v_0) = (\beta_0^T(v_0), h\beta_0'^T(v_0), hg'(v_0))^T$ be the true value of $\theta(v_0)$, for any $v_0 \in \mathcal{V}$. Define

$$\begin{aligned} S_{nkl}(t, \theta_0(v_0)) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) \exp\{\theta_0^T(v_0) \mathbf{Z}_i^* + g_0(v_0)\} K_h(V_i - v_0) (Z_i^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^l, \\ S_{nkl}^*(t, v_0) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) \exp\{\beta_0^T(v_0) \mathbf{Z}_i + g_0(v_0)\} K_h(V_i - v_0) (Z_i^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^l, \end{aligned}$$

for $k = 0, 1, 2$, $l = 0, 1, 2$. Then if Assumptions 1-7 hold, we have

$$\sup_{t \in [0, \tau]} \|S_{nkl}^*(t, v_0) - a_{kl}(t, v_0)\|_\infty = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

and

$$\sup_{t \in [0, \tau]} \|S_{nkl}(t, \theta_0(v_0)) - a_{kl}(t, v_0)\|_\infty = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h^2 \right),$$

where $k = 0, 1, 2$, $l = 0, 1, 2$.

Proof. Note that for any $t \in [0, \tau]$, $k = 0, 1, 2$, $l = 0, 1, 2$ and $v_0 \in \mathcal{V}$, $a_{kl}(t, v_0) = E\{S_{nkl}^*(t, v_0)\}$, and $Cov\{\sqrt{n}S_{nkl}^*(t, v_0)\} = O_p(h)$. Hence according to Theorem 1 of Silverman (1978), we obtain the following result:

$$\left\{ \frac{1}{h} \log\left(\frac{1}{h}\right) \right\}^{-1/2} \sqrt{n} \sup_{t \in [0, \tau]} \sup_{v \in \mathcal{V}} |S_{nkl}^*(t, v_0) - a_{kl}(t, v_0)| = O_p(1),$$

which yields that

$$\sup_{t \in [0, \tau]} \|S_{nkl}^*(t, v_0) - a_{kl}(t, v_0)\|_\infty = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right),$$

$k = 0, 1, 2$, $l = 0, 1, 2$. Furthermore, since

$$\begin{aligned} & \sup_{t \in [0, \tau]} \|S_{nkl}(t, \theta_0(v_0)) - a_{kl}(t, v_0)\|_\infty \\ \leq & \sup_{t \in [0, \tau]} \|S_{nkl}(t, v_0) - S_{nkl}^*(t, v_0)\|_\infty + \sup_{t \in [0, \tau]} \|S_{nkl}^*(t, v_0) - a_{kl}(t, v_0)\|_\infty, \end{aligned}$$

we just need to prove that the first term on the right-hand side has the order of $O_p(h^2)$ to complete the proof. Notice that for $k = 0, 1, 2$, $l = 0, 1, 2$,

$$\begin{aligned} & \sup_{t \in [0, \tau]} \|S_{nkl}(t, \theta_0(v_0)) - S_{nkl}^*(t, v_0)\|_\infty \\ = & \sup_{t \in [0, \tau]} \left\| \frac{1}{n} \sum_{i=1}^n Y_i(t) K_h(V_i - v_0) (Z^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^l \right. \\ & \left. [\exp\{\theta_0^T(v_0) \mathbf{Z}_i^* + g_0(v_0)\} - \exp\{\beta_0^T(v_0) \mathbf{Z}_i + g_0(v_0)\}] \right\|_\infty \\ = & \sup_{t \in [0, \tau]} \left\| \frac{1}{n} \sum_{i=1}^n Y_i(t) K_h(V_i - v_0) (Z^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^l \right. \\ & \left. \exp\{\eta_i(v_0)\} \frac{1}{2} [\beta_0''^T(V_i^*) \mathbf{Z}_i + g_0''(V_i^*)] (V_i - v_0)^2 \right\|_\infty, \end{aligned}$$

where $\eta_i(v_0)$ is between $\beta_0^T(v_0) \mathbf{Z}_i + g_0(v_0)$ and $\theta_0^T(v_0) \mathbf{Z}_i^* + g_0(v_0)$, and V_i^* is between v_0 and V_i , $i = 1, \dots, n$. Hence it follows from Theorem 1 of Silverman (1978) that

$$\begin{aligned} & \sup_{t \in [0, \tau]} \left\| \frac{1}{n} \sum_{i=1}^n Y_i(t) K_h(V_i - v_0) (Z^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^l \right. \\ & \left. \exp\{\eta_i(v_0)\} \frac{1}{2} [\beta_0''^T(V_i^*) \mathbf{Z}_i + g_0''(V_i^*)] (V_i - v_0)^2 \right\|_\infty \\ = & \sup_{t \in [0, \tau]} \left\| \frac{h^2}{n} \sum_{i=1}^n Y_i(t) K_h(V_i - v_0) (Z^{\otimes k}) \left(\frac{V_i - v_0}{h}\right)^{l+2} \right. \\ & \left. \exp\{\eta_i(v_0)\} \frac{1}{2} [\beta_0''^T(V_i^*) \mathbf{Z}_i + g_0''(V_i^*)] \right\|_\infty \\ = & O_p(h^2). \end{aligned}$$

Thus lemma 3 follows.

Lemma 4 *If Assumptions 1-7 hold, then*

$$\begin{aligned} & \left\| -\frac{\partial^2 \ell_n(\theta^*(v_0); v_0)}{\partial \theta \partial \theta^T} \Big|_{\theta = \theta^*(v_0)} - \Omega(v_0) \right\|_\infty \\ = & O_p \left(\left(\frac{\log(1/h)}{nh} \right)^{1/2} + h^2 + \|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty \right), \end{aligned} \tag{21}$$

where $\theta^*(v_0)$ is between $\hat{\theta}(v_0)$ and $\theta_0(v_0)$, for any $v_0 \in \mathcal{V}$.

Proof. Let

$$\begin{aligned} \Phi(t, \theta^*(v_0)) &= \left\{ \begin{pmatrix} S_{n20}(t, \theta^*(v_0)) & S_{n21}(t, \theta^*(v_0)) & S_{n11}(t, \theta^*(v_0)) \\ S_{n21}(t, \theta^*(v_0)) & S_{n22}(t, \theta^*(v_0)) & S_{n12}(t, \theta^*(v_0)) \\ S_{n11}(t, \theta^*(v_0)) & S_{n12}(t, \theta^*(v_0)) & S_{n02}(t, \theta^*(v_0)) \end{pmatrix} S_{n00}(t, \theta^*(v_0)) \right. \\ &\quad \left. - \begin{pmatrix} S_{n10}(t, \theta^*(v_0)) \\ S_{n11}(t, \theta^*(v_0)) \\ S_{n01}(t, \theta^*(v_0)) \end{pmatrix}^{\otimes 2} \right\} S_{n00}^{-2}(t, \theta^*(v_0)), \text{ and} \\ \varphi(t, v^*) &= \left\{ \begin{pmatrix} a_{20}(t, v^*) & a_{21}(t, v^*) & a_{11}(t, v^*) \\ a_{21}(t, v^*) & a_{22}(t, v^*) & a_{12}(t, v^*) \\ a_{11}(t, v^*) & a_{12}(t, v^*) & a_{02}(t, v^*) \end{pmatrix} a_{00}(t, v^*) \right. \\ &\quad \left. - \begin{pmatrix} a_{10}(t, v^*) \\ a_{11}(t, v^*) \\ a_{01}(t, v^*) \end{pmatrix}^{\otimes 2} \right\} a_{00}^{-2}(t, v^*), \end{aligned}$$

where $v^* \in \mathcal{V}$ such that $\theta_0(v^*) = \theta^*(v_0)$. Hence,

$$\begin{aligned} &\left\| -\frac{\partial^2 \ell_n(\theta^*(v_0); v_0)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta^*(v_0)} - \Omega(v_0) \right\|_{\infty} \\ &\leq \left\| \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \{\Phi(t, \theta^*(v_0)) - \varphi(t, v^*)\} dN_i(t) \right\|_{\infty} + \left\| \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \varphi(t, v_0) dM_i(t) \right\|_{\infty} \\ &\quad + \left\| \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \{\varphi(t, v_0) - \varphi(t, v^*)\} dN_i(t) \right\|_{\infty} \\ &\quad + \left\| \frac{1}{n} \sum_{i=1}^n \int_0^{\tau} \varphi(t, v_0) Y_i(t) \exp\{\beta_0^T(V_i) \mathbf{Z}_i + g_0(V_i)\} d\Lambda_0(t) - \Omega(v_0) \right\|_{\infty} \\ &\equiv I_{n1} + I_{n2} + I_{n3} + I_{n4}. \end{aligned} \tag{22}$$

According to Lemma 3, we have $I_{n1} = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} + h^2\right)$. Notice that it follows from lemma

1 and lemma 2 that

$$\left\| \frac{1}{n} \sum_{i=1}^n K_h(V_i - v) \int_0^{\tau} \xi_i(t) dM_i(t) \right\|_{\infty} = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right), \tag{23}$$

$$\left\| \frac{1}{n} \sum_{i=1}^n K_h(V_i - v) \int_0^{\tau} \xi_i(t) dN_i(t) - E\{m^*(v)\} \right\|_{\infty} = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right), \tag{24}$$

for any compact support \mathcal{V} of V . Since

$$I_{n2} = \left\| \frac{1}{n} \sum_{i=1}^n \int_0^\tau \frac{\partial \varphi(t, v)}{\partial \theta} \Big|_{\theta=\theta_0(v_0)} \{ \theta_0(v^*) - \theta_0(v_0) \} (1 + o_p(|\theta_0(v^*) - \theta_0(v_0)|)) dN_i(t) \right\|_\infty,$$

it follows from (24) that $I_{n2} = O_p(\|\theta_0(v^*) - \theta_0(v_0)\|_\infty) = O_p(\|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty)$. Similarly, since $K(\cdot)$ is symmetric, (23) yields that $I_{n3} = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right)$ and Lemma 3 yields that $I_{n4} = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2}\right)$. Thus Lemma 4 holds.

APPENDIX C: PROOF OF THE MAIN RESULTS

Proof of Theorem 2. For any $v_0 \in \mathcal{V}$,

$$-\frac{\partial \ell_n(\theta_0(v_0); v_0)}{\partial \theta} \Big|_{\theta=\theta_0(v_0)} = \frac{\partial^2 \ell_n(\theta^*(v_0); v_0)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta^*(v_0)} (\hat{\theta}(v_0) - \theta_0(v_0)),$$

where $\theta^*(v_0)$ is between $\hat{\theta}(v_0)$ and $\theta_0(v_0)$. Hence we have

$$\|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty = \left\| \left[-\frac{\partial^2 \ell_n(\theta^*(v_0); v_0)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta^*(v_0)} \right]^{-1} \frac{\partial \ell_n(\theta^*(v_0); v_0)}{\partial \theta} \Big|_{\theta=\theta_0(v_0)} \right\|_\infty. \quad (25)$$

Next, we focus on $\left\| \frac{\partial \ell_n(\theta^*(v_0); v_0)}{\partial \theta} \Big|_{\theta=\theta_0(v_0)} \right\|_\infty$. Let $\zeta_n(t, \theta_0(v_0)) = (S_{n10}^T(t, \theta_0(v_0)), S_{n11}^T(t, \theta_0(v_0)), S_{n01}(t, \theta_0(v_0)))^T$. Then

$$\begin{aligned} & \frac{\partial \ell_n(\theta^*(v_0); v_0)}{\partial \theta} \Big|_{\theta=\theta_0(v_0)} \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \left\{ \mathbf{Z}_i^* - \frac{\zeta_n(t, \theta_0(v_0))}{S_{n00}(t, \theta_0(v_0))} \right\} dN_i(t) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \left\{ \mathbf{Z}_i^* - \frac{\zeta_n(t, \theta_0(v_0))}{S_{n00}(t, \theta_0(v_0))} \right\} dM_i(t) \\ & \quad + \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \left\{ \mathbf{Z}_i^* - \frac{\zeta_n(t, \theta_0(v_0))}{S_{n00}(t, \theta_0(v_0))} \right\} Y_i(t) \exp\{\beta_0^T(V_i) \mathbf{Z}_i + g_0(V_i)\} d\Lambda_0(t) \\ &\equiv I_{n1}^* + I_{n2}^* \end{aligned}$$

By Lemma 1 and Lemma 3, we have

$$\begin{aligned} & \left\| I_{n1}^* - \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \left\{ \mathbf{Z}_i^* - \frac{(a_{10}^T(t, v_0), a_{11}^T(t, v_0), a_{01}(t, v_0))^T}{a_{00}(t, v_0)} \right\} dM_i(t) \right\|_\infty \\ & \leq O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} h^2 + \left\{\frac{\log(1/h)}{nh}\right\}\right), \end{aligned} \quad (26)$$

and

$$\|I_{n2}^* - \frac{1}{2}\mu_2 h^2 \Omega(v_0)(\beta_0''^T(v_0), 0_K^T, g_0''(v_0))^T\|_\infty \leq O_p(h^3). \quad (27)$$

Note that $(A + hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + O(h^2)$, thus it follows from Lemma 4 that

$$\begin{aligned} & \left\| \left[-\frac{\partial^2 \ell_n(\theta^*(v_0); v_0)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta^*(v_0)} \right]^{-1} - \Omega^{-1}(v_0) \right\|_\infty \\ &= O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h^2 + \|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty \right). \end{aligned} \quad (28)$$

As a result, by combining (25)-(28), we have

$$\begin{aligned} & \|\sqrt{nh}\{\hat{\theta}(v_0) - \theta_0(v_0) - \frac{1}{2}\mu_2 h^2(\beta_0''^T(v_0), 0_K^T, g_0''(v_0))^T\} \\ & \quad - \sqrt{\frac{h}{n}} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \{ \mathbf{Z}_i^* - \frac{(a_{10}^T(t, v_0), a_{11}^T(t, v_0), a_{01}(t, v_0))^T}{a_{00}(t, v_0)} \} dM_i(t)\|_\infty \\ & \leq O_p \left(\{\log(1/h)\}^{1/2} h^2 + \left\{ \frac{\log(1/h)}{\sqrt{nh}} \right\} \right) + O_p(\sqrt{nh^7}) \\ & \quad + O_p \left(\|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty \{\log(1/h)\}^{1/2} \right). \end{aligned}$$

It follows from (23) and $\|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty = o_p(1)$ that $\|\hat{\theta}(v_0) - \theta_0(v_0)\|_\infty = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h^2 \right)$, which yields that

$$\begin{aligned} & \|\sqrt{nh}\{\hat{\theta}(v_0) - \theta_0(v_0) - \frac{1}{2}\mu_2 h^2(\beta_0''^T(v_0), 0_K^T, g_0''(v_0))^T\} \\ & \quad - \sqrt{\frac{h}{n}} \sum_{i=1}^n \int_0^\tau K_h(V_i - v_0) \{ \mathbf{Z}_i^* - \frac{(a_{10}^T(t, v_0), a_{11}^T(t, v_0), a_{01}(t, v_0))^T}{a_{00}(t, v_0)} \} dM_i(t)\|_\infty \\ & \leq O_p \left(\frac{\log(1/h)}{nh} + h^2 \{\log(1/h)\}^{1/2} + \sqrt{nh^7} \right). \end{aligned}$$

Hence, Theorem 2 follows.

Next, we derive the consistent estimators of the asymptotic bias and covariance of $\hat{\beta}(\cdot)$. According to Fan *et al.* (2006), we can estimate the asymptotic bias and covariance by

$$\widehat{bias}(\hat{\beta}(v)) = A_n^{-1}(v)B_n(v), \quad \widehat{\Sigma}^{-1}(v) = (nh)^{-1}A_n^{-1}(v)\widehat{\Pi}_n(v)A_n^{-1}(v),$$

and

$$\widehat{\Pi}_n(v) = \frac{\nu_0^{-1}}{nh} \sum_{i=1}^n \int_0^\tau K_h^2(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} \right\}^{\otimes 2} dN_i(t),$$

respectively, where

$$\begin{aligned}
 A_n(v) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left[\frac{S_{n20}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} - \left\{ \frac{S_{n10}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} \right\}^{\otimes 2} \right] dN_i(t), \\
 B_n(v) &= \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} \right\} dN_i(t).
 \end{aligned}$$

Proof of Theorem 3. Note that $\|\hat{\theta}(v) - \theta_0(v)\|_\infty = O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} + h^2\right)$; hence it follows from lemma 3 that

$$\begin{aligned}
 &\|A_n(v) - \Sigma(v)\|_\infty \\
 \leq &\|A_n(v) - \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left[\frac{S_{n20}(t, \theta_0(v))}{S_{n00}(t, \theta_0(v))} - \left\{ \frac{S_{n10}(t, \theta_0(v))}{S_{n00}(t, \theta_0(v))} \right\}^{\otimes 2} \right] dN_i(t)\|_\infty \\
 &+ \left\| \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left[\frac{S_{n20}(t, \theta_0(v))}{S_{n00}(t, \theta_0(v))} - \left\{ \frac{S_{n10}(t, \theta_0(v))}{S_{n00}(t, \theta_0(v))} \right\}^{\otimes 2} \right] dN_i(t) - \Sigma(v) \right\|_\infty \\
 = &O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} + h^2\right).
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 &\|B_n(v) - \frac{1}{2}\mu_2 h^2 \Sigma(v) \beta''(v)\|_\infty \\
 \leq &\left\| \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} \right\} dM_i(t) \right\|_\infty \\
 &+ \left\| \frac{1}{n} \sum_{i=1}^n \int_0^\tau K_h(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \hat{\theta}(v))}{S_{n00}(t, \hat{\theta}(v))} \right\} d\Lambda_i(t) - \frac{1}{2}\mu_2 h^2 \Sigma(v) \beta''(v) \right\|_\infty.
 \end{aligned}$$

Similarly to (26) and (27), the first term on the right-hand side of the above equation is bounded by $O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} h^2 + \left\{\frac{\log(1/h)}{nh}\right\}\right)$, and the second term on the right-hand side of the above equation is bounded by $O_p\left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} h^2 + h^4\right)$. Hence we have

$$\|B_n(v) - \frac{1}{2}\mu_2 h^2 \Sigma(v) \beta''(v)\|_\infty \leq p \left(\left\{\frac{\log(1/h)}{nh}\right\}^{1/2} h^2 + \left\{\frac{\log(1/h)}{nh}\right\} + h^4 \right)$$

Finally, we consider $\hat{\Pi}_n(v)$. According to lemma 3, we have

$$\begin{aligned}
 &\|\hat{\Pi}_n(v) - \nu_0 \Sigma(v)\|_\infty \\
 \leq &\|\hat{\Pi}_n(v) - \frac{1}{nh} \sum_{i=1}^n \int_0^\tau K_h^2(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \theta(v))}{S_{n00}(t, \theta(v))} \right\}^{\otimes 2} dN_i(t)\|_\infty
 \end{aligned}$$

$$\begin{aligned}
& + \left\| \frac{1}{nh} \sum_{i=1}^n \int_0^\tau K_h^2(V_i - v) \left\{ \mathbf{Z}_i - \frac{S_{n10}(t, \theta(v))}{S_{n00}(t, \theta(v))} \right\}^{\otimes 2} dN_i(t) - \nu_0 \Sigma(v) \right\|_\infty \\
& = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} + h^2 \right).
\end{aligned}$$

Therefore, since $nh^5 = O(1)$, we have

$$\begin{aligned}
& \sqrt{nh} \|\widehat{bias}(\widehat{\beta}(v)) - \frac{1}{2} \mu_2 h^2 \beta''(v)\|_\infty = O_p \left(\{\log(1/h)\}^{1/2} h^2 \right), \\
& \text{and } \|\widehat{\Sigma}^{-1}(v) - \Sigma^{-1}(v)\|_\infty = O_p \left(\left\{ \frac{\log(1/h)}{nh} \right\}^{1/2} \right).
\end{aligned}$$

Hence, Theorem 3 follows.

Proof of Theorem 4. Note that

$$\begin{aligned}
& \widehat{\psi}(p) - \psi_0(p) - \widehat{bias}(\widehat{\psi}(p)) \\
& = \widehat{\beta}(\widehat{\pi}_p) - \beta_0(\pi_p) - \widehat{bias}(\widehat{\beta}(\pi_p)) \\
& = \widehat{\beta}(\widehat{\pi}_p) - \widehat{\beta}(\pi_p) + \widehat{\beta}(\pi_p) - \beta_0(\pi_p) - \widehat{bias}(\widehat{\beta}(\pi_p)) \\
& = \widehat{\beta}'(\pi_p)(\widehat{\pi}_p - \pi_p) + o_p(|\widehat{\pi}_p - \pi_p|) + \{\widehat{\beta}(\pi_p) - \beta_0(\pi_p) - \widehat{bias}(\widehat{\beta}(\pi_p))\}.
\end{aligned}$$

Thus, it's sufficient to prove

$$\sup_{0 \leq p \leq 1} |(\log(1/h))^{1/2} \sqrt{nh} \{\widehat{\beta}'(\pi_p)(\widehat{\pi}_p - \pi_p) + o_p(|\widehat{\pi}_p - \pi_p|)\}| = o_p(1). \quad (29)$$

We first prove $\sup_{0 \leq p \leq 1} |\widehat{\pi}_p - \pi_p| = O_p(n^{-1/2})$. According to the definition of $\widehat{\pi}_p$, we have

$$\widehat{F}_v(\widehat{\pi}_p -) \leq p \leq \widehat{F}_v(\widehat{\pi}_p), \text{ and } \sup_{0 \leq p \leq 1} |\widehat{F}_v(\widehat{\pi}_p -) \leq p \leq \widehat{F}_v(\widehat{\pi}_p)| = o_p(n^{-1/2}).$$

Hence, it follows from $\sup_x |\widehat{F}_v(x) - F_v(x)| = o_p(n^{-1/2})$ that

$$\begin{aligned}
\sup_{0 \leq p \leq 1} |F_v(\widehat{\pi}_p) - F_v(\pi_p)| & \leq \sup_{0 \leq p \leq 1} |F_v(\widehat{\pi}_p) - \widehat{F}_v(\widehat{\pi}_p)| + \sup_{0 \leq p \leq 1} |\widehat{F}_v(\widehat{\pi}_p) - p| \\
& = O_p(n^{-1/2}) + o_p(n^{-1/2}) = O_p(n^{-1/2}).
\end{aligned}$$

Note that

$$F_v(\widehat{\pi}_p) - F_v(\pi_p) = f_v(\pi_p)(\widehat{\pi}_p - \pi_p) + o_p(|\widehat{\pi}_p - \pi_p|).$$

Thus, $\sup_{0 \leq p \leq 1} |\hat{\pi}_p - \pi_p| = O_p(n^{-1/2})$, which yields that

$$\begin{aligned} & \sup_{0 \leq p \leq 1} |(\log(1/h))^{1/2} \sqrt{nh} \{ \hat{\beta}'(\pi_p)(\hat{\pi}_p - \pi_p) + o_p(|\hat{\pi}_p - \pi_p|) \}| \\ &= O_p(\{h \log(1/h)\}^{1/2}) = o_p(1). \end{aligned}$$

Theorem 4 follows.



List of Figures

1	Different BATE curves with different sensitive subsets	32
2	Comparing between a "strong" predictive marker and a "weak" predictive marker. . .	33
3	The BATE curves for different treatment arms	34
4	Estimated results of the case with censoring rate $CR = 20\%$	35
5	Estimated results of the case with censoring rate $CR = 40\%$	36
6	Estimation of the BATE curves for the treatment arm of other treatments	37
7	Estimation of the BATE curves for the treatment arm of surgery plus chemotherapy	38



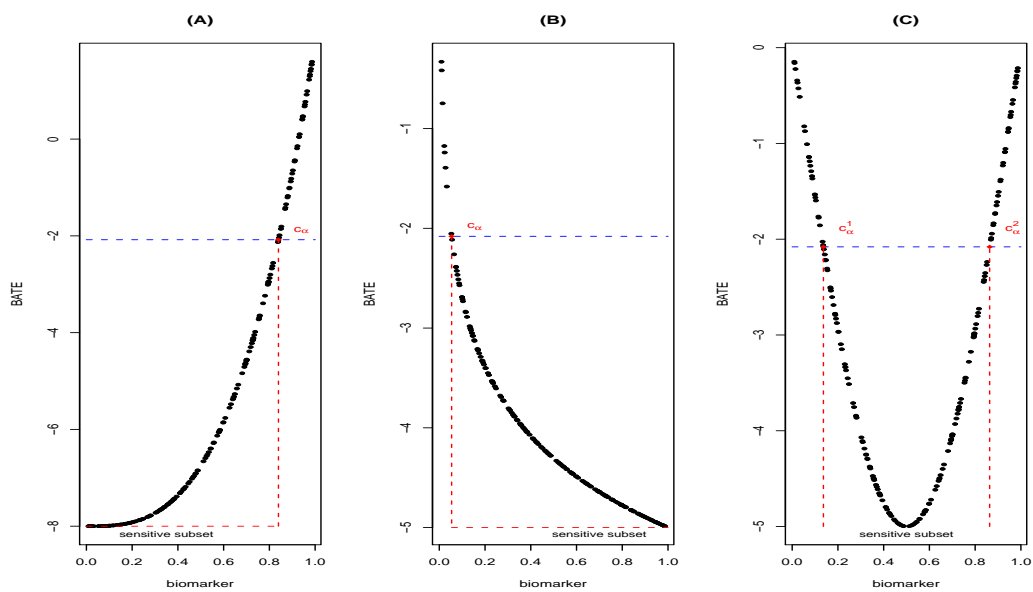


Figure 1: Different BATE curves with different sensitive subsets



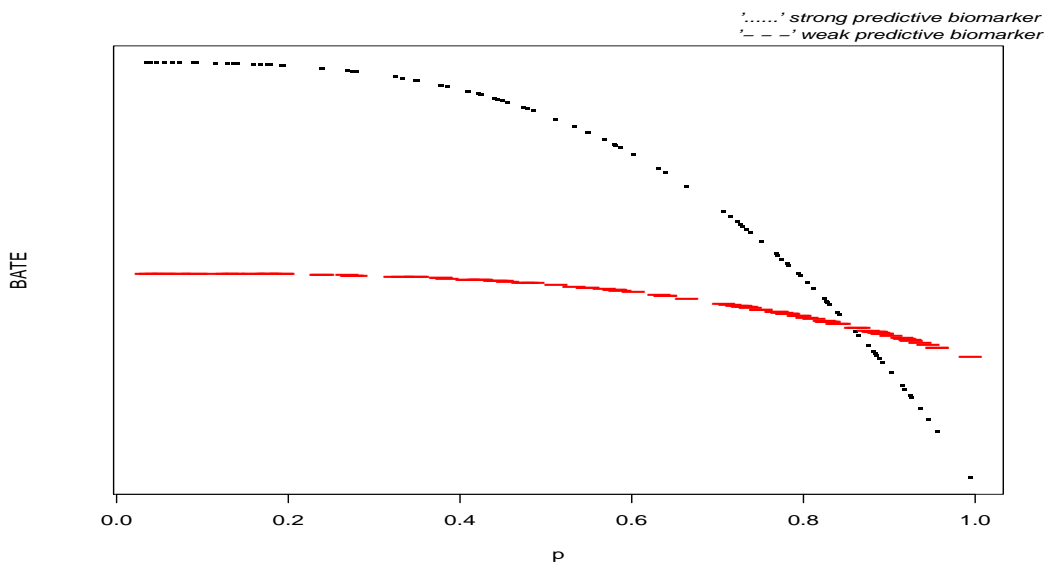


Figure 2: Comparing between a "strong" predictive marker and a "weak" predictive marker.



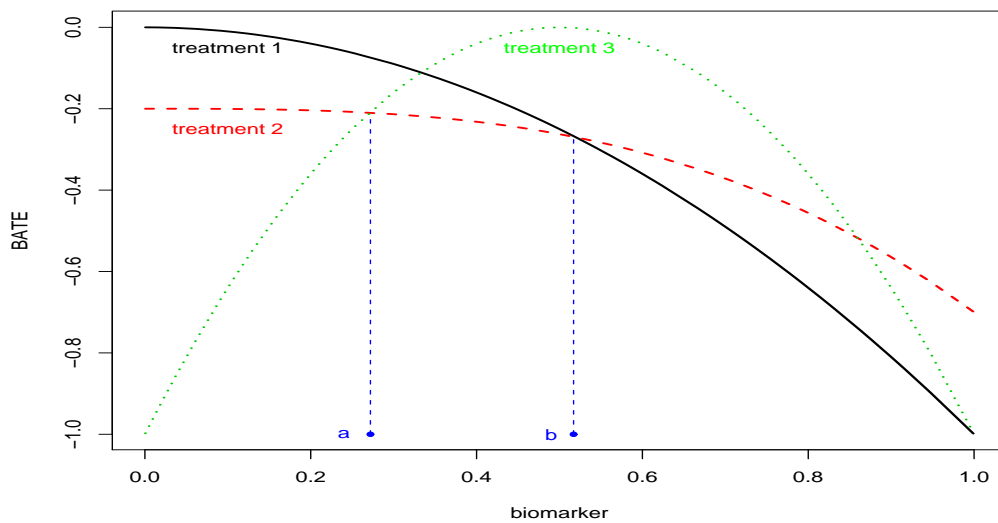


Figure 3: The BATE curves for different treatment arms: The solid, dashed and dotted curves are the BATE curves for Treatment 1,2 and 3, respectively.



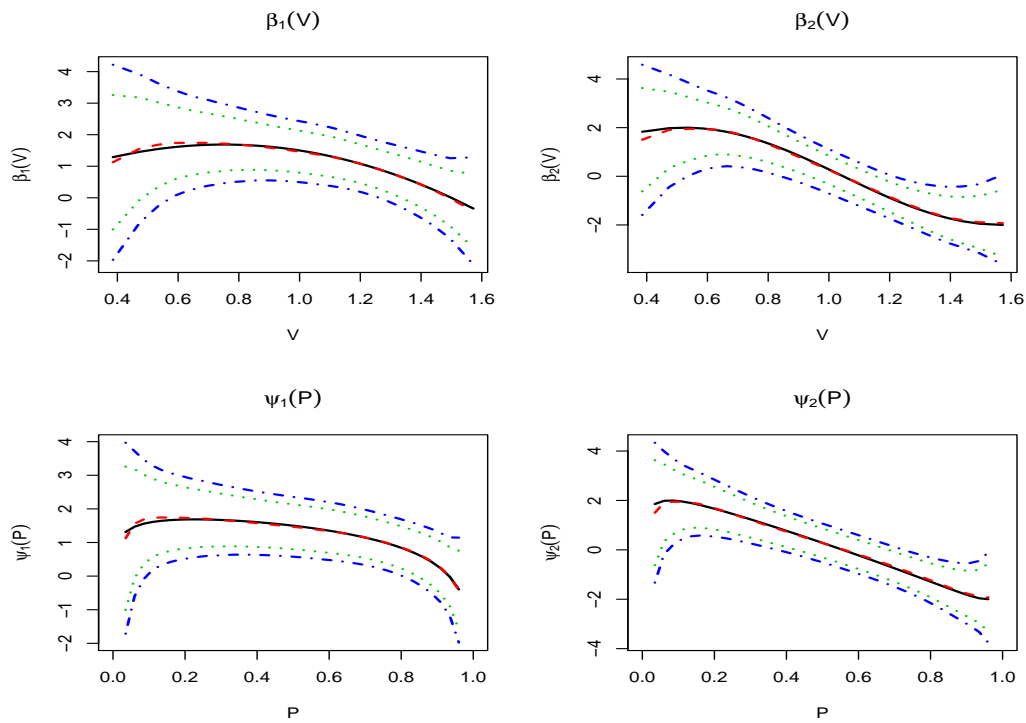


Figure 4: Estimated results of the case with censoring rate $CR = 20\%$. For each figure, solid and dashed lines are the true and estimated curves, respectively. Dotted and dash-dotted lines are 95% confidence limits and 95% confidence bands, respectively.

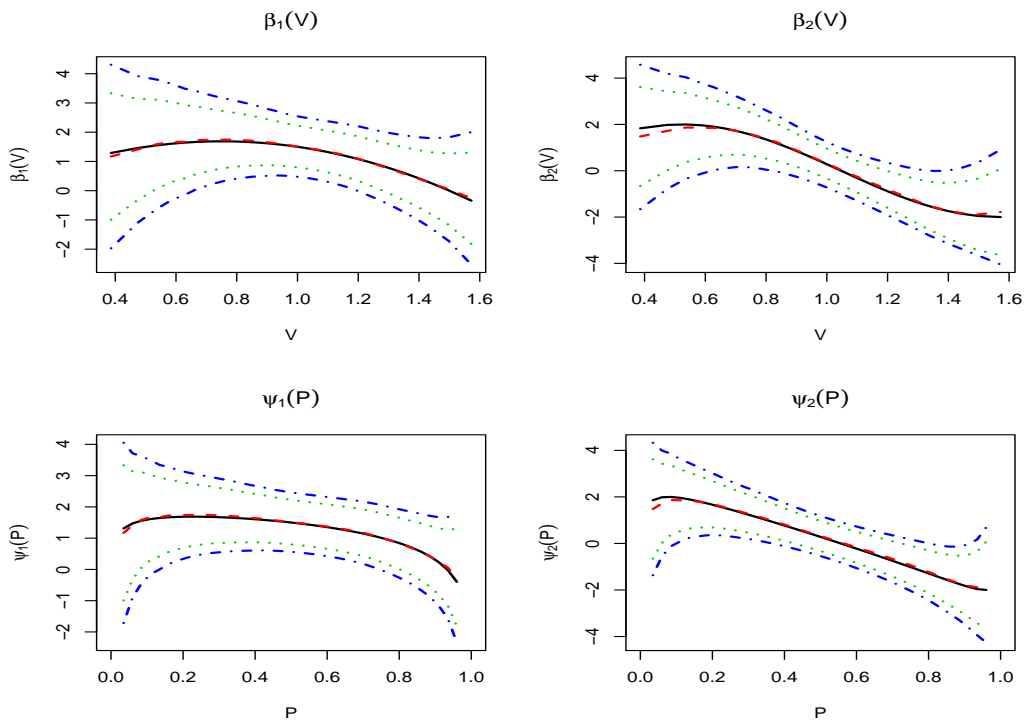


Figure 5: Estimated results of the case with censoring rate $CR = 40\%$. For each figure, solid and dashed lines are the true and estimated curves, respectively. Dotted and dash-dotted lines are 95% confidence limits and 95% confidence bands, respectively.

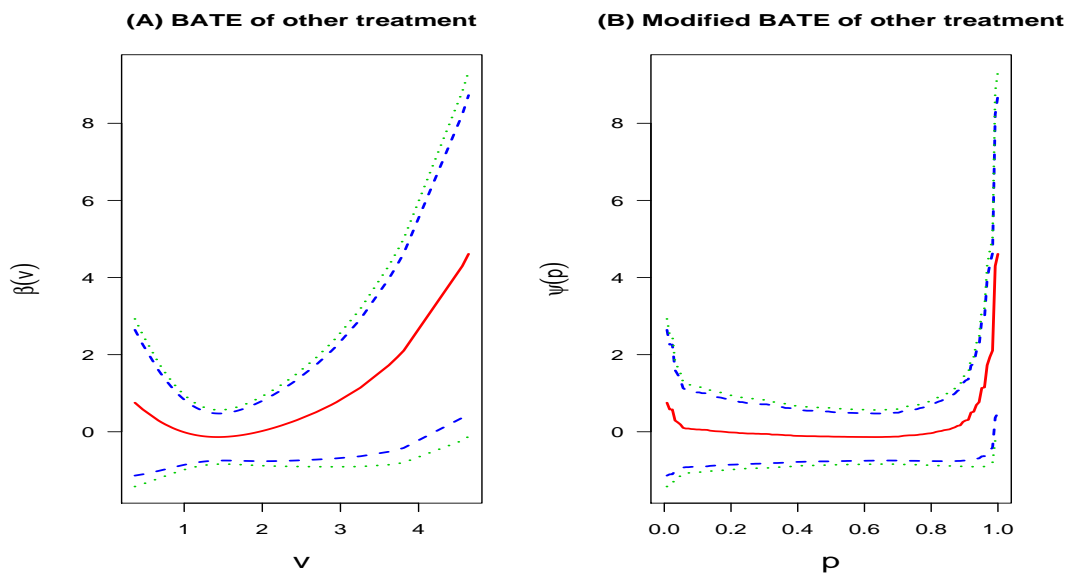


Figure 6: Estimation of the BATE curves: (A) is for the BATE curve versus $c - myc$, (B) is for the modified BATE curve versus p . Solid, dashed and dotted lines are estimated curves, 95% confidence limits and 95% confidence band, respectively.



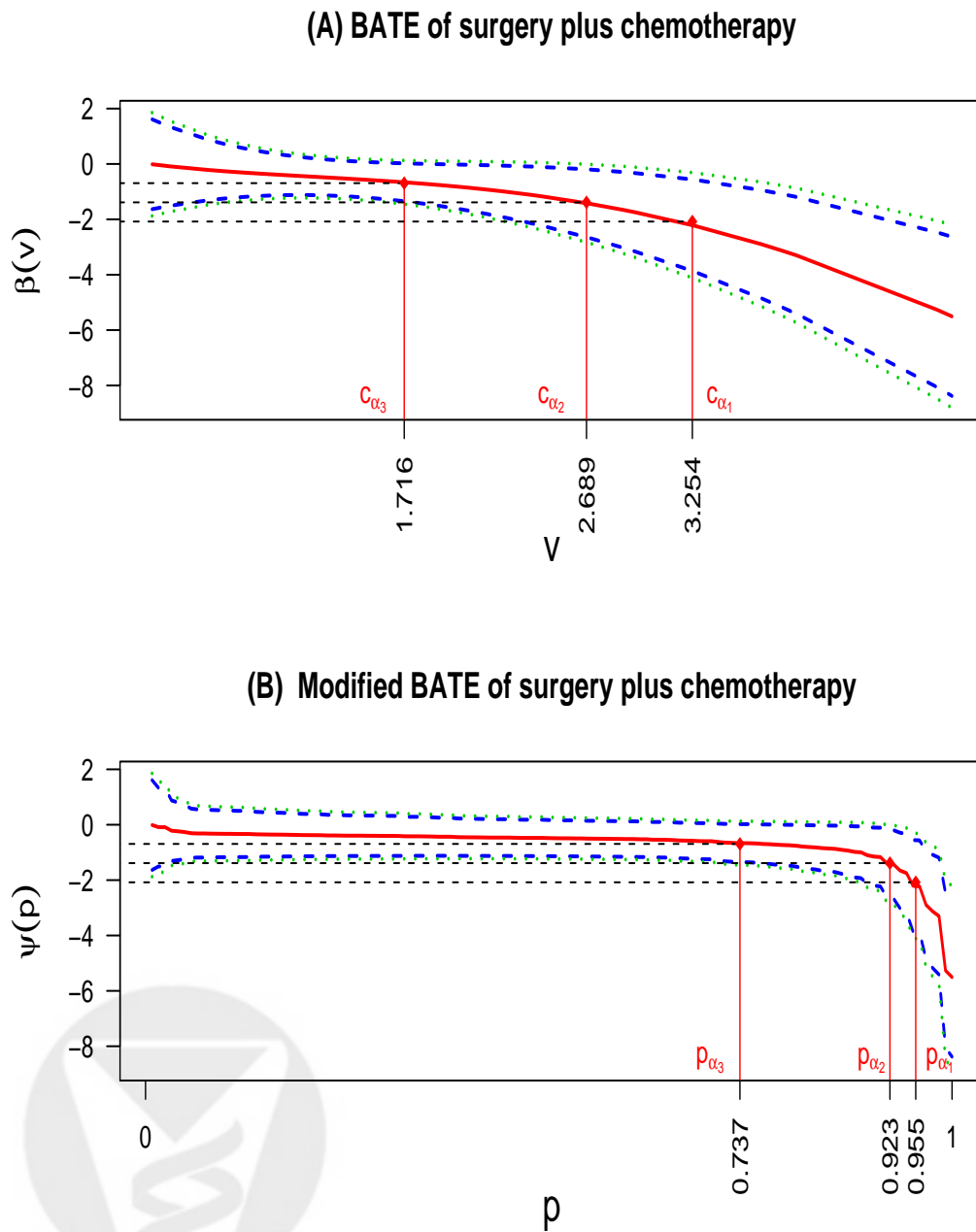


Figure 7: Estimation of the BATE curves: (A) is for the BATE curve versus $c - myc$, (B) is for the modified BATE curve versus p . Solid, dashed and dotted lines are estimated curves, 95% confidence limits and 95% confidence band, respectively.

List of Tables

1	Estimation results of the BATE curves	40
2	Simulation results of the hypothesis test	41



Table 1: Estimation results of the BATE curves

<i>CR</i>	<i>p</i>	$\psi_1(p)$				$\psi_2(p)$			
		Bias	SD	SE	CP	Bias	SD	SE	CP
20%	0.2	0.0374	0.3116	0.2995	0.94	-0.0165	0.3145	0.2894	0.93
	0.4	-0.0103	0.2416	0.2279	0.92	-0.0265	0.2321	0.2190	0.94
	0.6	-0.0343	0.2319	0.2228	0.94	0.0418	0.2350	0.2159	0.93
	0.8	-0.0339	0.2983	0.2855	0.91	0.1207	0.3264	0.2937	0.93
40%	0.2	0.0354	0.3348	0.3357	0.97	-0.0287	0.3175	0.3232	0.91
	0.4	-0.0143	0.2604	0.2545	0.90	-0.0422	0.2589	0.2439	0.94
	0.6	-0.0400	0.2572	0.2443	0.92	0.0286	0.2649	0.2384	0.92
	0.8	-0.0394	0.3303	0.3083	0.92	0.1268	0.3477	0.3216	0.95



Table 2: Percentage of rejecting null hypothesis given significant level $\alpha^* = 0.05$ and censoring rate 20% based on 500 simulations

	$\beta_1(v) \equiv -1, \alpha^* = 0.01$			$\beta_1(v) \equiv -1, \alpha^* = 0.05$		
$\beta_2(v)$	$2 \sin(3v)$	$-\exp(v)$	0.5	$2 \sin(3v)$	$-\exp(v)$	0.5
Type I errors	0.014	0.008	0.010	0.066	0.050	0.056

