ESTIMATION OF TREATMENT DOSE-EFFECT BY ADJUSTING FOR DEPENDENT CENSORING USING HIGH-DIMENSIONAL AUXILIARY INFORMATION

Donglin Zeng *

Abstract

In right-censored data, one goal is to obtain an estimator of treatment dose-effect, which is represented by some parameter in a marginal model of lifetime given treatment dose variable. When dependent censoring is explained by both the dose variable and many other auxiliary covariates (high-dimensional auxiliary information), an intuitive approach to estimate the dose effect is to first estimate the conditional distribution of lifetime given the whole covariates using a semiparametric model then average out the auxiliary information. However, this intuitive approach is problematic in practice since the semiparametric model can be easily misspecified. In this article, a novel way is proposed to enable us to condense the high-dimensional auxiliary information through the utilization of two working models for the distribution of lifetime given all the covariates and the distribution of censoring time given all the covariates. The estimator of the treatment dose-effect is then obtained by maximizing a pseudo-likelihood function over a sieve space. Such an estimator is shown to be consistent and asymptotically normal when either of the two working models is correct; additionally, its asymptotic variance is the same as the generalized Cramér-Rao bound when both working models are correct.

KEY WORDS: semiparametric inference, dimension reduction, B-spline sieves, double robustness;

*Donglin Zeng is Assistant Professor with Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599.
1 Introduction.

Right-censored observations with high-dimensional auxiliary information are common in many clinical trials of analyzing patients’ lifetimes. Such observations contain the observed event times (either lifetime or censoring time), the censoring status indicating whether the patients were censored or not, the treatment doses the patients received and many other auxiliary covariates. In notations, if denote $T$ as the lifetime, $C$ as the censoring time, $V$ as the treatment dose variable and denote $X$ as the auxiliary covariates, then right censored observations from $n$ patients are

$$(Y_i, R_i, V_i, X_i), i = 1, ..., n,$$

where $Y_i = T_i \wedge C_i, R_i = I_{T_i \leq C_i}$. $X$ is called the auxiliary information: $X$ contains some variables which clinician are not primarily interested in; but these variables are informative since they either predict the lifetime or explain the censoring mechanism; for example, $X$ may include the feeling of the patients about participation in the trials, the social support of the patients, their accessibility to the trials and the genetic information of the patients etc. When a large amount of such information of $X$ is collected, it is often safe to assume, the dependence between $T$ and $C$ can be fully explained by $X$ and $V$; that is, conditional on $X$ and $V$, $T$ and $C$ are assumed to be independent.

One main goal of many clinical trials is to estimate the effect of the treatment variable on the lifetime. Among many definitions of the treatment effect, one definition is attractive due to its causal explanation. In this causal definition, the treatment effect is defined as the coefficient of the treatment variable in a Cox proportional hazard model of the lifetime given the treatment variable; i.e., we suppose

$$h_{T|V}(t|v) = \lambda(t)e^{(\alpha_0, \alpha_1)(1,v)'}, \int_0^\tau \lambda(t)dt = 1,$$

where $h_{T|V}(t|v)$ is the hazard rate function of the lifetime given the treatment, $\lambda(t)$ is the baseline hazard rate function and $\tau$ is the ending time of the trials. The coefficient $\alpha_1$ in this model has the following interpretation: suppose that we randomly select a group of patients from the whole population and randomly assign treatments to these patients, then one unit increase of the treatment dose causes the hazard rate increase ($\alpha_1 \geq 0$) or decrease ($\alpha_1 < 0$) by a factor of $e^{\alpha_1}$.

The purpose of this paper is to obtain an estimator of $\alpha = (\alpha_0, \alpha_1)$ using right-censored observations. As described above, $T$ and $C$ are assumed to be independent given $(X, V)$ and $X$ is high-dimensional. Moreover, $(X, V)$ are supposed to be time-independent covariates and the dose variable $V$ is a continuous variable.

Many literature have contributed to estimating the treatment effect using right-censored data.
However, most of them only consider the situations that $X$ has a low dimension and the dose effect is defined as the coefficient of $V$ in a semiparametric model of $T$ given $(X, V)$. Such models may be a Cox proportional hazard model (Cox (1972)), a proportional odds model (Pettitt (1982) and Bennett (1983)), etc. Estimation approaches such as using partial likelihood method (Cox (1972)), profile likelihood method (Murphy et al (1998)) can be used to derive the estimators of the coefficients. However, when $X$ is high-dimensional, this semiparametric model may be easily misspecified. For example, when using the Cox proportional hazard model for $T$ given $(X, V)$, we may hesitate about including the high-order interactions among $(X, V)$ (if the dimension for $X$ is 20, the two-way interactions have 200 terms). Consequently, once the model of $T$ given $(X, V)$ is misspecified, the estimators of the parameters in the marginal model of $T$ given $V$ are inconsistent. In addition, the treatment effect defined in this way does not possess the causal explanation as the one we defined above.

In this paper, we propose one novel way to condense the high-dimensional covariates $(X, V)$ by utilizing two working models for the distribution of $T$ given $(X, V)$ and the distribution of $C$ given $(X, V)$; then an estimator of the treatment dose effect $\alpha = (\alpha^0, \alpha^1)$ is obtained by maximizing a pseudo-likelihood function of a reduced data, which consists of the observed event times, the censoring status, and the condensed covariates. In the maximization, the nuisance parameters for $\alpha$ are profiled out over a sieve space consisting of B-splines. We demonstrate that the estimator for $\alpha$ has the following properties: if either of the two working models is correct, the estimator is consistent and asymptotically normal; if both working models are correct, the estimator’s asymptotic variance attains the generalized Cramér-Rao bound. The first property is called the double robustness by Robins et al (2000).

The paper is organized as follows. Section 2 illustrates the procedure of estimating $\alpha$ using our approach. Since in the calculation of the estimator of $\alpha$, the nuisance parameters are profiled over a sieve space consisting of B-splines, Section 3 details the construction of such a sieve space. In Section 4, we list all the assumptions which are sufficient in proving the consistency and asymptotic normality of the estimator for the treatment dose-effect; and the proofs are given in both Section 5 and Section 6. In Section 7, a heuristic method is proposed to estimate the asymptotic variance but without rigorous justification. Final discussion is given in Section 8. The appendix at the end of this paper, whose contents are extracted from Shumaker (1981), summarizes the properties of B-splines.

Notations. The following notations are used in this paper.

1. $P_n, P$ are the empirical measure and the probability measure respectively, i.e., if $X_1, \ldots, X_n$ are
i.i.d observations, then for any \( g(X) \),
\[
P_n g(X) = \frac{1}{n} \sum_{i=1}^{n} g(X_i), \quad P g(X) = E g(X).
\]
\( \sqrt{n}(P_n - P) \) denotes the empirical process generated from \( X_1, \ldots, X_n \).

2. Denote \( U(\gamma) = \gamma(X, V)' \). \( f_{U(\gamma)}(\cdot|y, v) \) denotes the conditional density of \( U(\gamma) \) given \( (T = y, V = v) \).

2 Estimation Procedure

The whole procedure for estimating \( \alpha \) is the following. First, we make two working models for the distribution of \( T \) and \( C \) given \( (X, V) \):

(Working Model 1): we tentatively assume that \( T \) is independent of \( X \) given \( V \) so \( f_{T|X,V}(y|x, v) = f_{T|V}(y|v) \);

(Working Model 2): we tentatively assume that the model of \( C \) given \( (X, V) \) is a Cox proportional hazard model, i.e., \( f_{C|X,V}(y|x, v) = h_c(y)e^{\gamma(x,v)'e^{-e^{\gamma(x,v)'H_c(y)}}} \) for some constant \( \gamma \) and unknown baseline hazard rate function \( h_c(.) \).

In here, \( f_{T|X,V}(\cdot|x, v) \) denotes the conditional density of \( T \) given \( (X = x, V = v) \) and \( f_{C|X,V}(\cdot|x, v) \) denotes the conditional density of \( C \) given \( (X = x, V = v) \). So basically, Working Model 1 implies that \( T \) depends on \( (X, V) \) only via \( V \) and Working Model 2 implies that \( C \) depends on \( (X, V) \) via a linear combination of \( (X, V) \). Recall the notation \( U(\gamma) = \gamma(X, V)' \) for fixed \( \gamma \). Then if either of the two working models is right, then for any \( (t, s) \),
\[
P(T < t, C < s|U(\gamma), V) = E[P(T < t|X, V)P(C < s|X, V)|U(\gamma), V]
= P(T < t|U(\gamma), V)P(C < s|U(\gamma), V).
\]

We conclude that if either working model is correct, \( T \) and \( C \) are independent given \( (U(\gamma), V) \); in other words, the dependence between \( T \) and \( C \) can be fully explained by the two-dimensional condensed information \( (U(\gamma), V) \).

Second, supposing that \( \gamma \) is a known constant, we replace the observed statistics \( (X, V) \) by \( (U(\gamma), V) \) and obtain a reduced data \( (Y_i, R_i, (U_i, V_i)), i = 1, \ldots, n \), where \( U_i = \gamma(X_i, V_i)' \). When either working model is correct, the resulting conditional independence between \( T \) and \( C \) given \( (U(\gamma), V) \) gives the observed likelihood function of the reduced data as
\[
\prod_{i=1}^{n} \left( \int_{Y_i}^{\infty} f_{T|U(\gamma), V}(s|U_i, V_i)ds \right)^{R_i} \left( \int_{Y_i}^{\infty} f_{T|U(\gamma), V}(s|U_i, V_i)ds \right)^{1-R_i}
\]
\[ [f_{C|U(\gamma)}, V(Y_i|U_i, V_i)]^{1-R_i} \left[ \int_{Y_i}^\infty f_{C|U(\gamma), V}(s|U_i, V_i)ds \right]^{R_i} f_{U(\gamma), V}(U_i, V_i) \].

Furthermore,
\[ f_{T|U(\gamma), V}(Y|U, V) = f_{T|V}(Y|V) f_{U(\gamma)}(U|Y, V)/f_{U(\gamma)}(U|V) \]
where \( f_{U(\gamma)}(\cdot|V) \) is the conditional density of \( U(\gamma) \) given \( V \) and \( f_{T|V}(\cdot|V) \) is equal \( e^{\alpha(Y)} e^{-e^{\alpha(Y)}} \Lambda(Y) \).

Hence, the above likelihood can be rewritten as
\[
\prod_{i=1}^n \left\{ e^{-e^{\alpha(Y_i)} \Lambda(Y_i)} e^{\alpha(Y_i)} \lambda(Y_i) f_{U(\gamma)}(U_i|Y_i, V_i) \right\}^{R_i} \left[ \int_{Y_i}^\infty e^{-e^{\alpha(Y_i)} \Lambda(s)} e^{\alpha(Y_i)} \lambda(s) f_{U(\gamma)}(U_i|s, V_i)ds \right]^{1-R_i} f_{C|U(\gamma), V}(Y_i|U_i, V_i)^{1-R_i} \left[ \int_{Y_i}^\infty f_{C|U(\gamma), V}(s|U_i, V_i)ds \right]^{R_i} f_V(V_i) \right\}.
\]

One convenience of the above expression is that we have managed to absorb the constraint imposed by the form of the conditional density of \( T \) given \( V \).

Third, we maximize the following function
\[
\prod_{i=1}^n \left\{ \left[ e^{-e^{\alpha(Y_i)} \Lambda(Y_i)} e^{\alpha(Y_i)} \lambda(Y_i) f_{U(\gamma)}(Y_i|X_i, V_i) \right]^{R_i} \left[ \int_{Y_i}^\infty e^{-e^{\alpha(Y_i)} \Lambda(s)} e^{\alpha(Y_i)} \lambda(s) f_{U(\gamma)}(Y_i|s, V_i)ds \right]^{1-R_i} \right\}.
\]

over the parameters \( (\alpha, \lambda(y), f_{U(\gamma)}(u|y, v)) \) for some fixed \( \gamma \).

However, two issues need to be solved to realize this maximization.

1. What value does \( \gamma \) take in the maximization? In fact, \( \gamma \) can be estimated using the working model for \( C \) given \( (X, V) \). One way to obtain the estimator of \( \gamma \) is to maximize the pseudo censoring partial-likelihood for the observations \( (Y_i, R_i, (X_i, V_i)), i = 1, \ldots, n \),
\[
\sum_{i=1}^n (1 - R_i) \log h_c(Y_i) + \gamma(X_i, V_i) - H_c(Y_i) e^{\gamma(X_i, V_i)};
\]
equivalently, \( \gamma \) is estimated by performing the Cox regression. We denote this estimator as \( \hat{\gamma}_n \) and substitute it into the above maximization.

2. What function space is used for \( (\lambda(y), f_{U(\gamma)}(u|y, v)) \) so that the maximization can proceed? In the next section, we will propose a sieve space consisting of B-splines for \( f_{U(\gamma)}(u|y, v) \) and \( \lambda(y) \). By using sieve estimators, the parameter space on which the maximization is realized is a compact set of a finite dimensional real space. Therefore, the estimator of \( \alpha \) exists.

In summary, we want to maximize
\[
\prod_{i=1}^n \left\{ e^{-e^{\alpha(Y_i)} \Lambda(Y_i)} e^{\alpha(Y_i)} \lambda(Y_i) f_{U(\gamma)}(Y_i|X_i, V_i) \right\}^{R_i}
\]
\[
\int_{Y_i}^{\infty} e^{-e^{\alpha(1,V_i)'\Lambda(s)}e^{\alpha(1,V_i)'}\lambda(s)f_{U(\gamma)}(\tilde{U}_i|s, V_i)ds}^{1-R_n}
\]

over the parameters \((\alpha, \lambda(y), f_{U(\gamma)}(u|y, v))\), where \(\tilde{U}_i = \hat{\gamma}_n(X_i, V_i)'\) for \(i = 1, ..., n\) and the parameters \((\alpha, \lambda(y), f_{U(\gamma)}(u|y, v))\) belong to a sieve space given in the next section. Computationally, many constrained optimization algorithms such as conjugate gradient method, steepest decent method combining with the use of either penalty or barrier function, can be applied to calculate the estimator for \(\alpha\).

3 A Sieve Space for Parameters

As assumed in Section 4, we suppose that \(\alpha \in [-M, M]^2\) and that the support for \((Y, X, V)\) is a cube \([0, 1] \times [e, f]^D \times [0, 1]\), where \(M\) is a known constant and \(D = \text{dim}(X)\). Define

\[
b(\gamma) = \max_{x,v}\{\gamma(x, v) : (x, v) \in [0, 1] \times [e, f]^D \times [0, 1]\}
\]

and

\[
a(\gamma) = \min_{x,v}\{\gamma(x, v) : (x, v) \in [0, 1] \times [e, f]^D \times [0, 1]\}.
\]

By appropriate shifting and rescaling of \(X\), we also suppose that \(0 \leq a(\hat{\gamma}_n) < b(\hat{\gamma}_n) \leq 1\).

Before giving the sieve space for the parameters, we reparameterize \((f_{U(\gamma)}(u|y, v), \lambda(y))\) by introducing

\[
f_{U(\gamma)}(u|y, v) = \frac{e^{\eta(u, y, v)}}{\int_{a(\hat{\gamma}_n)}^{b(\hat{\gamma}_n)} e^{\eta(u, y, v)}du}, \lambda(y) = e^{\xi(y)}
\]

where \((\eta(u, y, v), \xi(y))\) satisfy that

\[
\eta(0, y, v) = 0, \int_0^1 e^{\xi(y)}dy = 1.
\]

After the reparameterization, the new parameters are \((\alpha, \xi(y), \eta(u, y, v))\). A sieve space thus is defined for these new parameters. To do that, we obtain an extended partition with equal length \(1/K_n\) for the interval \([0, 1]\):

\[
\Delta_e = \{s_{-m} = ... = s_{-1} = 0 = s_0 < s_1 < ... < s_{K_n} = 1 = ... = s_{K_n+m}\}
\]

where \(m\) and \(K_n\) are two integers to be chosen later. Let \(\{N_j^m(s)\}\) be a normalized B-Spline basis associated with \(\Delta_e, j = 1, ..., K_n + m\). Then the sieve space for the parameters \((\alpha, \xi(y), \eta(u, y, v))\) is defined as

\[
\mathcal{S}_n(m, K_n, M_n) = \{ (\alpha, \xi(y), \eta(u, y, v)) : \alpha \in [-M, M]^2; \}
\]
\[ \eta(u, y, v) = \sum_{i_1, i_2, i_3 = 1}^{m + K_n} \eta_{i_1, i_2, i_3} N_{i_1}^m(u) N_{i_2}^m(y) N_{i_3}^m(v); \]

\[ \xi(y) = \sum_{i = 1}^{m + K_n} \xi_i N_i^m(y); \]

\[ \sum_{i_1, i_2, i_3 = 1}^{m + K_n} |\eta_{i_1, i_2, i_3}| \leq M_n, \quad \sum_{i = 1}^{m + K_n} |\xi_i| \leq M_n; \]

\[ \eta(0, y, v) = 0, \quad \int_0^1 e^{\xi(y)} dy = 1 \}, \]

where \( M_n \) is another constant to be chosen later.

Finally, we maximize the function

\[
P_n \ln \left\{ \left[ e^{-e^{\alpha(1,V)' \int_0^Y e^{\alpha(1,V)' e^{\xi(Y)}} - e^{\alpha(1,V)' e^{\xi(Y)}}}} \right]^{R-1} \left[ \int_0^\infty e^{-e^{\alpha(1,V)' \int_0^s e^{\alpha(1,V)' e^{\xi(s)}} - e^{\alpha(1,V)' e^{\xi(s)}}}} ds \right]^{1-R} \right\}
\]

on the sieve space \( S_n(m, K_n, M_n) \). As indicated in the next section, one possible choice of \( (m, K_n, M_n) \) is

\[
(k + 2, \tilde{M} n^\beta, \tilde{M} \ln \ln n)
\]

for some constant \( \tilde{M} \), where \( k \) is a known integer and \( \beta \) satisfies that \( \frac{1}{2k} < \beta < \frac{3}{4k + 9} \).

It will be shown \( |a(\hat{\gamma}_n) - b(\hat{\gamma}_n)| \) has a positive limit in probability. So the probability of the denominator in the above expression being 0 is zero asymptotically. Hence, we recognize that, in probability, none of the terms in the above object function is infinite. The arguments of the maximum exist since we are maximizing over a compact set in a finite dimensional space. However, the solution itself may not be unique. We simply select any one of these maximizers and denote it as \( (\hat{\alpha}_n, \hat{\xi}_n(y), \hat{\eta}_n(u, y, v)) \) so we obtain the respective estimators

\[
\hat{\alpha}_n = \hat{\alpha}_n, \hat{\lambda}_n(y) = e^{\hat{\xi}_n(y)}, \hat{f}_{U|Y}(u|y, v) = \frac{e^{\hat{\eta}_n(u, y, v)}}{\int_{a(\hat{\gamma}_n)}^{b(\hat{\gamma}_n)} e^{\hat{\eta}_n(u, y, v)} du}.
\]

### 4 Assumptions

We list all the assumptions in this section. But before that, we state the asymptotic normality of \( \hat{\gamma}_n \), which was proved in Zeng (2001): there exists a constant \( \gamma^* \) such that

\[
\sqrt{n}(\hat{\gamma}_n - \gamma^*) = \sqrt{n}(P_n - P) S_{\gamma^*}(\gamma^*; Y, R, X, V) + o_P(1),
\]

where \( S_{\gamma^*}(\gamma^*; Y, R, X, V) \) is the influence function for \( \hat{\gamma}_n \).
In addition to the assumption that $T$ and $C$ are independent given $(X, V)$, all the other assumptions are in the following.

(A1.) There exists a known constant $\tau$ and an unknown constant $\tau'$ such that

$$P(C \geq \tau | X, V) = P(C = \tau | X, V) > \tau' > 0, P(T > \tau | X, V) > \tau' > 0, \text{a.s.}$$

(A2.) $\alpha$ is in $[-M, M]^2$ and $\gamma^*$ is in $[-M, M]^{D+1}$ where $M$ is a known constant; moreover, at least one of the first $D$ components of $\gamma^*$ is nonzero;

(A3.) The joint densities of $(Y, X, V, R = 1)$ and $(Y, X, V, R = 0)$ have positive lower bounds on $[0, \tau] \times [e, f]^{D} \times [0, 1]$;

(A4.) $P\{(X, V) \text{ is in a hyperplane } \} = 0$ (non-collinearity).

(A5.) For a known integer $k \geq 11$, the conditional density of $X$ given $(T = y, V)$, denoted by $f_{X|T,V}(x|y,v)$, and the true baseline hazard rate for $T$, i.e., $\lambda_0(y)$, satisfy that,

$$\ln f_{X|T,V}(x|y,v) \in W^{k+4, 2}(R^D), \ln \lambda_0(y) \in W^{k+4, 2}(R)$$

after appropriate extension to the whole space ($W^{k+4, 2}(R^d)$ is a Sobolev space consisting of the functions with $(k + 4)$-th derivatives in $L^2(R^d)$). The same assumption holds for the conditional density $\ln f_{C|X,V}$, the conditional density of $C$ given $X$ and $V$.

(A6.) $\Lambda_0(\tau) = 1$.

(A7.) $(M_n, K_n)$ satisfy that

$$\frac{K_n^{4k/3+3}(\ln K_n)^2}{n} \to 0, \quad e^{13M_n} K_n^{4k/3+3}(\ln K_n)^2 \to 0.$$ 

(A8.) $\frac{\sqrt{n}}{K_n^2} \to 0$.

Remark 4.1. The $\tau$ in Assumption (A.1) is equal to the ending time of the trials and in the following context we assume $\tau = 1$. Without the loss of the generality, we also assume $0 < a(\gamma^*) < b(\gamma^*) < 1$ after appropriate shifting and re-scaling of $X$. Therefore, one obvious fact is that $|a(\hat{\gamma}_n) - b(\hat{\gamma}_n)| \to |a(\gamma^*) - b(\gamma^*)| > 0$ in probability and $0 \leq a(\hat{\gamma}_n) < b(\hat{\gamma}_n) \leq 1$.

Remark 4.2. In Assumption (A.2), the condition on $\gamma^*$ seem to be artificial; however, since $\gamma^*$ solves the equation

$$P[(1 - R)(X, V)] = \mathbb{P}\{(1 - R) \frac{\mathbb{P}[|Y| \geq y(X, V)e^{\gamma(X, V)}]}{\mathbb{P}[|Y| \geq y e^{\gamma(X, V)}]} | y = Y\},$$

such condition can be expressed in terms of the joint density of $(T, C, X, V)$. For simplicity, we skip the details while impose the assumptions directly on $\gamma^*$. 

7
Remark 4.3. For simplicity, $X$ is assumed to be continuous; however, when $X$ includes categorical variables, the variable $U(\gamma)$ may have two or more disconnected supports. Thus, instead of using $[a(\hat{\gamma}_n), b(\hat{\gamma}_n)]$ as the integration domain with respect to $U(\gamma)$ in the maximization, we need to use some finite disconnected intervals. This change will not affect our following proofs.

Remark 4.4. From Assumption (A.2), without the loss of the generality, suppose the first component of $\gamma^* = (\gamma^{*}_1, \ldots, \gamma^{*}_D, \gamma^{*}_{D+1})$ is not zero. Then in terms of $f_{X|T,V}(x|y,v)$, the conditional density of $U(\gamma^*) = \gamma^*(X,V)'$ given $(T,V)$, denoted by $f_{U(\gamma^*)}(u|y,v)$, can be expressed by

$$
\frac{1}{\gamma^{*}_1} \int_0^1 \cdots \int_0^1 f_{X|T,V}\left(\frac{u - \sum_{i=2}^{D} \gamma^{*}_i x_i - \gamma^{*}_{D+1} v}{\gamma^{*}_1}, x_2, x_3, \ldots, x_D|y,v\right) dx_2 \ldots dx_D.
$$

Hence, Assumptions (4-5) also imply that $f_{U(\gamma^*)}(u|y,v)$ is bounded from below and its $(k + 4)$th derivatives are $L^2$ integrable. Moreover, by the embedding theorem in the Sobolev space, this also implies that each of $\ln f_{U(\gamma^*)|T,V}, \ln f_{C|U(\gamma^*)}, \ln \lambda_0(t)$ is in $W^{k,\infty}$ space; i.e, their $k$th derivatives are bounded essentially.

Remark 4.5. Assumption (A.7) and Assumption (A.8) give the size of the sieve space in terms of the number of knots in the partition $(K_n)$ and the bounds of the sieve functions $(M_n)$. When $k \geq 11$, such $K_n$ satisfying both Assumption (A.7) and Assumption (A.8) exists since $K_n^{4k/3+3}(\ln K_n)^{8/3} < K_n^{2k}$ for $K_n$ is large enough. For example, we can choose $K_n = n^\beta$, $\frac{1}{2k} < \beta < \frac{3}{4k+9}$. Additionally, the choice of $M_n$ can be of order $O(\ln \ln n)$.

Although all these assumptions guarantee the validity of the following arguments, they are not minimal.

5 Consistency

We prove the consistency of the estimator $\hat{\alpha}_n$ in this section. In addition, the convergence rates of $\hat{\alpha}_n$ and the estimators for the nuisance parameters are also obtained.

Theorem 5.1. (Consistency of $\hat{\alpha}_n$) Under Assumptions (1-7), suppose that either of the two working models is true: $T$ is independent of $X$ given $V$; or, the distribution of $C$ given $(X,V)$ follows a Cox proportional hazard model. Then $\hat{\alpha}_n$ is a consistent estimator of the true coefficients $\alpha_0$.

Proof of Theorem 5.1

The whole proof can be divided into three steps: first, we construct some functions in the sieve space which approximates the true parameters; then by using empirical process theory, we obtain one key inequality; finally, this inequality is used to obtain the consistency.
Recall the simplification we have assumed to construct the sieve space: \( \tau = 1, 0 < a(\gamma^*) < b(\gamma^*) < 1 \); thus, in probability 0 ≤ \( a(\tilde{\gamma}_n) < b(\tilde{\gamma}_n) \) ≤ 1.

Step 1. We construct some functions in \( S_n(m, K_n, M_n) \) to approximate the true parameters.

The true conditional density of \( U(\gamma^*) \) given \( (T, V) \), \( f_{U(\gamma^*)}(u|y, v) \) is positive on its support \([a(\gamma^*), b(\gamma^*)] \times [0, 1]^2\). So we extend the function \( \ln f_{U(\gamma^*)}(u|y, v) \) over \([0, 1]^3\) while the extended function after extension still belongs to \( W^{k, \infty}([0, 1]^3) \) and moreover, it is positive on \([a(\gamma^*)/2, 1] \times [0, 1]^2\) but it is zero on \([0, a(\gamma^*)/4] \times [0, 1]^2\). This extension is possible (cf. Adams (1975)). From the properties of spline functions (Appendix section), we can define a linear operator \( Q(.) \) mapping \( L^\infty([0, 1]^3) \) to the sieve space, i.e., \( Q(.) \) satisfies that for any \( g(u, y, v) \in L^\infty([0, 1]^3) \),

\[
Q(g(u, y, v)) = \sum_{i_1, i_2, i_3 = 1}^{m+K_n} \Lambda_{i_1, i_2, i_3}(g)N_{i_1}^m(u)N_{i_2}^m(y)N_{i_3}^m(v),
\]

where \( \Lambda_{i_1, i_2, i_3} \) are the linear functionals in \( L^\infty([0, 1]^3) \).

Let

\[
\eta_n(u, y, v) = Q^*(\ln f_{U(\gamma^*)}(u|y, v)),
\]

where \( Q^*(g) \) contains the remaining terms in \( Q(g) \) after deleting those terms in \( \{\lambda_{i_1, i_2, i_3}(g)N_{i_1}(u)N_{i_2}(y)N_{i_3}(v) : N_{i_1}(0) \neq 0\} \). We also define

\[
\xi_n(y) = Q(\ln \lambda_0(y)) - \ln \int_0^1 e^{Q(\ln \lambda_0(y))} dy.
\]

Hence, \( \eta_n(0, y, v) = 0 \) and \( \int_0^1 e^{\xi(y)} dy = 1 \).

From Theorem A.4 in the appendix, we have that \( \forall 1 \leq p \leq \infty \),

\[
\left| \Lambda_{i_1, i_2, i_3}(\ln f_{U(\gamma^*)}(u|y, v)) \right| \leq (2m + 1)^{3(m-1)}(K_n)^{3/p} \ln f_{U(\gamma^*)}(u|y, v) \| L_p([s_{i_1}, s_{i_1+1}) \times [s_{i_2}, s_{i_2+1}) \times [s_{i_3}, s_{i_3+1})],
\]

where \( \{s_1, ..., s_{K_n+m}\} \) is the partition of \([0, 1]\). Sum over all the indexes then

\[
\sum_{i_1, i_2, i_3 = 1}^{m+K_n} \left| \Lambda_{i_1, i_2, i_3}(\ln f_{U(\gamma^*)}(u|y, v)) \right| \leq C(m)K_n^{3/p} \ln f_{U(\gamma^*)}(u|y, v) \| L_p([0, 1]^3).
\]

Let \( p \) tend to \( \infty \) then

\[
\sum_{i_1, i_2, i_3 = 1}^{m+K_n} \left| \Lambda_{i_1, i_2, i_3}(\ln f_{U(\gamma^*)}(u|y, v)) \right| \leq C(m) < \infty.
\]

Therefore, we have verified that \( \eta_n(u, y, v) \) satisfies the conditions in the class \( S_n \). Similarly, we can obtain

\[
\sum_{i=1}^{m+K_n} \left| \Lambda_i(\ln \lambda_0(y)) \right| \leq C(m) < \infty,
\]

\[
where \( \tilde{\lambda}_i(\ln \lambda_0(y)) \) is the coefficient of \( N_i^m(y) \) in \( Q(\ln \lambda_0(y)) \). In conclusion, \((\alpha_0, \xi_n(y), \eta_n(u, y, v))\) is in the sieve space \( S_n(m, K_n, M_n) \).

Furthermore, according to Theorem A.4 in the appendix (let \( p = \infty \)), since \( k \leq m \), we have

\[
\|Q(\ln f_U(\gamma^*)(u|y, v)) - \ln f_U(\gamma^*)(u|y, v)\|_{L^\infty([0,1]^3)} \leq \frac{C(m)}{K_n^k} \| \ln f_U(\gamma^*)(u|y, v)\|_{W^{k, \infty}([0,1]^3)}.
\]

If \( N_i(0) \neq 0 \) then \( N_i(u) \) has a support in \([0, m/K_n]\) on which \( f_U(\gamma^*)(u|y, v) \) is also zero. Thus we conclude that the corresponding splines coefficients \( \tilde{\lambda}_{i_1, i_2, i_3} \) are zeros; that is, \( Q^*(\ln f_U(\gamma^*)(u|y, v)) = Q(\ln f_U(\gamma^*)(u|y, v)) \). So

\[
\|\eta_n(u|y, v) - \ln f_U(\gamma^*)(u|y, v)\|_{L^\infty([0,1]^3)} \leq \frac{C(m)}{K_n^k} \| \ln f_U(\gamma^*)(u|y, v)\|_{W^{k, \infty}([0,1]^3)}.
\]

Similarly by Theorem A.4,

\[
\|Q(\ln \lambda_0(y)) - \ln \lambda_0(y)\|_{L^\infty([0,1]^3)} \leq \frac{C(m)}{K_n^k} \| \ln \lambda_0(y)\|_{W^{k, \infty}([0,1])}.
\]

Therefore,

\[
\int_{a(\gamma^*)}^{b(\gamma^*)} e^{Q(\ln f_U(\gamma^*)(u|y,v))} du = 1 + O\left(\frac{1}{K_n^k}\right)
\]

and

\[
\int_0^1 e^{Q(\ln \lambda_0(y))} dy = 1 + O\left(\frac{1}{K_n^k}\right).
\]

Hence, if we define \( f_n(u|y, v) = \frac{e^{\eta_n(u,y,v)}}{\int_{a(\gamma_n)}^{b(\gamma_n)} e^{\eta_n(u,y,v)} du} \) and define \( \lambda_n(y) = e^{\xi_n(y)} \), then

\[
\|f_n(u|y, v) - f_U(\gamma^*)(u|y, v)\|_{L^\infty([0,1]^3)} \leq \frac{e^{\eta_n(u,y,v)}}{\int_{a(\gamma_n)}^{b(\gamma_n)} e^{\eta_n(u,y,v)} du} - f_U(\gamma^*)(u|y, v)\|_{L^\infty([0,1]^3)} \leq O_p(|a(\gamma_n) - a(\gamma^*)|) + O_p(|b(\gamma_n) - b(\gamma^*)|) + O\left(\frac{1}{K_n^k}\right)
\]

and similarly,

\[
\|\lambda_n(y) - \lambda_0(y)\|_{L^\infty([0,1])} \leq O\left(\frac{1}{K_n^k}\right).
\]

In other words, we construct the functions \((\alpha_0, \xi_n(y), \eta_n(u, y, v))\) in the sieve space such that the corresponding \((\alpha_0, \lambda_n(y), f_n(u|y, v))\) generated from these functions approximates the true parameters in \( L^\infty([0,1]^3) \) norm.
Step 2. We obtain a key inequality from the empirical process theory.

Some notations: suppose \( f_1(u, y, v) \) be a function on \([0, 1]^3\) and \( f_2(y) \) be a function on \([0, 1]\). We denote \( G(r, f_1(u, y, v), f_2(y), \alpha) \) as a function of \((y, v, x)\) given by

\[
\int_0^y e^{-e^{\alpha(1,v)' L_f}} ds e^{\alpha(1,v)' f_2(y)f_1(u, y, v)} r
\]

Then, since \((\hat{\alpha}_n, \hat{\lambda}_n(y), \hat{f}_U(\gamma)(u|y, v))\) maximizes the pseudo-likelihood function over \(S_n\), we have

\[
\begin{align*}
(P_n - P) \ln G(R, f_n(\hat{U}|Y, V), \hat{\lambda}_n(Y), \hat{\alpha}_n) \geq P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0) + P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0) - P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0),
\end{align*}
\]

where \( U = \hat{\gamma}_n(X, V)' \). As a result,

\[
\begin{align*}
(P_n - P) \ln G(R, f_n(\hat{U}|Y, V), \hat{\lambda}_n(Y), \hat{\alpha}_n) \geq P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0) + P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0) - P \ln G(R, f_n(\hat{U}|Y, V), \lambda_n(Y), \alpha_0),
\end{align*}
\]

Consider a class of functions \( \mathcal{L}_n \):

\[
\begin{align*}
\{ \ln G(r, \tilde{f}_n(u|y, v), \tilde{\lambda}_n(y), \alpha) : (\tilde{\eta}(u, y, v), \tilde{\xi}(y), (\tilde{\eta}(u, y, v), \tilde{\xi}(y), \alpha) \in S_n(m, K_n, M_n) \}
\end{align*}
\]

\[
\begin{align*}
\{ \ln \left\{ \frac{\tilde{f}_n(u|y, v)e^{-\alpha(1,v)' \tilde{\lambda}_n(y)}}{f_n(u|y, v)e^{-\alpha(1,v)' \lambda_n(y)}} \right\} & \geq \ln \left\{ \frac{\tilde{f}_n(u|y, v)e^{-\alpha(1,v)' \tilde{\lambda}_n(y)}}{f_n(u|y, v)e^{-\alpha(1,v)' \lambda_n(y)}} \right\} \\
& + \ln \left\{ \frac{\int_y^\infty \tilde{f}_n(u|s, v)e^{-\alpha(1,v)' \tilde{\lambda}_n(s)}}{f_n(u|s, v)e^{-\alpha(1,v)' \lambda_n(s)}} ds \right\} \\
& = \ln \left\{ \frac{\tilde{f}_n(u|y, v)e^{-\alpha(1,v)' \tilde{\lambda}_n(y)}}{f_n(u|y, v)e^{-\alpha(1,v)' \lambda_n(y)}} \right\}.
\end{align*}
\]

Since \( \|N_i^m(.)\|_{L^\infty[0,1]} = 1 \), for any \( \tilde{f}_n(u|y, v), \tilde{\lambda}_n(y) \) given in \( \mathcal{L}_n \),

\[
\|\tilde{f}_n(U|Y, V)\|_{L^\infty([0,1]^3)} \leq O_p(e^{2M_n}); \|\tilde{\lambda}_n(Y)\|_{L^\infty([0,1])} \leq O(e^{M_n}),
\]

11
and by Assumption (A.1) and Assumption (A.4),

\[
\int_Y^\infty f_n(U|s,V) e^{-\alpha_0(1,V)\Lambda_n(s)} e^{\alpha_0(1,V)\lambda_n(s)} ds \\
\geq \int_1^\infty f_n(U|s,V) e^{-\alpha_0(1,V)\Lambda_n(s)} e^{\alpha_0(1,V)\lambda_n(s)} ds \geq O_p(1) > 0, \forall Y \leq 1.
\]

Hence, the class \( \mathcal{L}_n \) has an upper bound \( O_p(M_n) \). Moreover, this class can be regarded as the class of functions indexed by \( \{\tilde{\eta}_{i_1,i_2,i_3}^{m+K_n}\}_{i_1,i_2,i_3=1}^{m+K_n} \), which are the respective coefficients of \( \tilde{\eta}(u,v) \), \( \tilde{\xi}(y) \) in \( S_n(m,K_n,M_n) \). Careful checking indicates that the function in \( \mathcal{L}_n \) is Lipschitz with respect to \( \{\tilde{\eta}_{i_1,i_2,i_3}^{m+K_n}\}_{i_1,i_2,i_3=1}^{m+K_n} \), \( \{\tilde{\xi}_i\}_{i=1}^{m+K_n} \) and the Lipschitz constant is bounded by \( O_p(e^{6M_n}) \). In addition, since \( |\tilde{\eta}_{i_1,i_2,i_3}, \tilde{\xi}_i| \leq M_n \), all these coefficients lie in a hyperspace of a real space \( R^{(m+K_n)^3+m+K_n} \). Therefore, for any \( \epsilon > 0 \), if we partition this hyperspace into subspaces with scale length \( \epsilon \), the total number of subspaces is at most \( O((\frac{M_n}{\epsilon})(m+K_n)^3+m+K_n) \). According to the Lipschitz property of the functions in \( \mathcal{L}_n \), the \( L^\infty([0,1]^3) \) distance between any two functions of \( \mathcal{L} \) with respective indexes in the same subspace is no more than \( O_p(e^{6M_n})e((m+K_n)^3 + m + K_n) \). Consequently, we obtain

\[
N_\epsilon(O_p(e^{6M_n})e((m+K_n)^3 + m + K_n), \mathcal{L}_n, L_2(P)) \leq O((\frac{M_n}{\epsilon})^{(m+K_n)^3+m+K_n}).
\]

According to Theorem 19.35, van der Vaart (1998), in probability we have

\[
\sqrt{n} E_p \| \mathbf{P}_n - \mathbf{P} \|_{\mathcal{L}_n} \leq O_p(1) \int_0^{O(M_n)} \sqrt{\ln \left( \frac{2M_n e^{6M_n}(m+K_n)^3}{\epsilon} \right)} d\epsilon \\
\leq O_p(1)(m+K_n)^{3/2} \ln M_n \left( e^{6M_n}(m+K_n)^3 + M_n^2 \right) \\
\leq O_p(1)(m+K_n)^{3/2} \ln(m+K_n) M_n^2 \\
\leq O_p(1) K_n^{3/2} (\ln K_n) M_n^2.
\]

Therefore,

\[
(\mathbf{P}_n - \mathbf{P})[\ln \frac{G(R, f_U(\gamma)) \hat{U}[Y,V], \hat{\lambda}(Y), \hat{\alpha}_0)}{G(R, f_n(U|Y,V), \lambda_n(Y), \alpha_0)}] \leq O_p(\frac{K_n^{3/2} (\ln K_n) M_n^2}{\sqrt{n}}).
\]

So the inequality (5.1) implies

\[
O_p(\frac{K_n^{3/2} (\ln K_n) M_n^2}{\sqrt{n}}) \\
\geq \mathbf{P}[\ln \frac{G(R, f_n(U|Y,V), \lambda_n(Y), \alpha_0)}{G(R, f_U(\gamma) \hat{U}[Y,V], \lambda_0(Y), \alpha_0)} + \mathbf{P}[\ln \frac{G(R, f_U(\gamma) \hat{U}[Y,V], \lambda_0(Y), \alpha_0)}{G(R, f_U(\gamma) U[Y,V], \lambda_0(Y), \alpha_0)}] \\
=: (I) + (II).
\]

Furthermore, as the result of the Lipschitz property of

\[
G(R, f_U(\gamma) (U^*|Y,V), \lambda_0(Y), \alpha_0)
\]

12
at \((f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(Y), \alpha_0)\), we have that

\[
(I) = \mathbf{P}[\ln \frac{G(R, f_n(\hat{U}|Y,V), \lambda_n(Y), \alpha_0)}{G(R, f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(y), \alpha_0)}] \\
\geq -O_p(1)\|f_n(\hat{U}|Y,V) - f_{U(\gamma^*)}(U|Y,V)\|_{L^\infty([0,1]^2)} \\
- O_p(1)(\|\lambda_n(Y) - \lambda_0(Y)\|_{L^\infty([0,1])} + |\hat{\gamma}_n - \gamma^*|) \\
\geq -O_p(1)(\frac{1}{K_n^t} + \frac{1}{\sqrt{n}}).
\]

On the other hand, when either of two working models is correct, \(T\) is independent of \(C\) conditionally on \((U(\gamma^*), V)\). Hence the joint density of \((T \wedge C, R, (U(\gamma^*), V))\) is exactly

\[
G(R, f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(Y), \alpha_0)F_{C|U(\gamma^*)}(Y|U^*, V)^{1-R}F_{C|U(\gamma^*)}(Y|U^*, V)^R.
\]

In addition, from Theorem A.1 in the appendix, it holds that

\[
\left| \frac{d\hat{f}_{U(\gamma)}(\hat{\gamma}_n(X,V)|Y,V)}{d\gamma} \right| \\
\leq O(e^{2M_n} \sum_{i_1,i_2,i_3} |\eta_{i_1,i_2,i_3}| \|\frac{dN_{i_1}^m(\gamma,x,v)}{d\gamma}\| N_{i_2}^m(v) N_{i_3}^m(y)|\gamma=\hat{\gamma}_n \right| \\
\leq O(e^{2M_n} M_n K_n) K_n.
\]

So if define \(\hat{f}_{U(\gamma)}(U^*|Y,V) = e^{\hat{\eta}_n(U^*,Y,V)} \int_{\hat{a}(\gamma^*)}^{\hat{b}(\gamma^*)} e^{\hat{\eta}_n(s,Y,V)} ds\), we obtain

\[
(II) = \mathbf{P}[\ln \frac{G(R, f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(Y), \alpha_0)}{G(R, f_{U(\gamma)}(U^*|Y,V), \hat{\lambda}_n(Y), \hat{\alpha}_n)}] \\
\geq -O_p(e^{2M_n} M_n K_n |\hat{\gamma}_n - \gamma^*|) + \mathbf{P}[\ln \frac{G(R, f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(Y), \alpha_0)}{G(R, f_{U(\gamma)}(U^*|Y,V), \hat{\lambda}_n(Y), \hat{\alpha}_n)}] \\
\geq -O_p(e^{2M_n} M_n K_n |\hat{\gamma}_n - \gamma^*| + |a(\hat{\gamma}_n) - a(\gamma^*)| + |b(\hat{\gamma}_n) - b(\gamma^*)|) \\
+ \mathbf{P}[\ln \frac{G(R, f_{U(\gamma)}(U^*|Y,V), \lambda_0(Y), \alpha_0)}{G(R, f_{U(\gamma)}(U^*|Y,V), \hat{\lambda}_n(Y), \hat{\alpha}_n)}] \\
\geq -O_p(e^{2M_n} M_n K_n \frac{1}{\sqrt{n}}) + O(e^{-3M_n}) \mathbf{P}[G(R, f_{U(\gamma^*)}(U^*|Y,V), \lambda_0(Y), \alpha_0) \\
- G(R, \hat{f}_{U(\gamma)}(U^*|Y,V), \hat{\lambda}_n(Y), \hat{\alpha}_n)]^2 \\
\geq -O_p(e^{2M_n} M_n K_n \frac{1}{\sqrt{n}}) + O(e^{-3M_n}) \times \\
\int_{(U^*,Y,V)} [G(1, f_{U(\gamma^*)}, \lambda_0(Y), \alpha_0) - G(1, \hat{f}_{U(\gamma)}, \hat{\lambda}_n(Y), \hat{\alpha}_n)]^2 dU^* dY dV.
\]

Finally, we obtain the key inequality

\[
\int_{[a(\gamma), b(\gamma)] \times [0,1]^2} [G(1, f_{U(\gamma^*)}, \lambda_0(Y), \alpha_0) - G(1, \hat{f}_{U(\gamma)}, \hat{\lambda}_n(Y), \hat{\alpha}_n)]^2 dU^* dY dV \\
\leq O_p(1)(\frac{e^{5M_n} M_n K_n}{\sqrt{n}} + \frac{e^{2M_n} M_n K_n}{K_n^t} + \frac{e^{2M_n} M_n^2 K_n^{3/2}}{K_n^t} \ln K_n).
\]
Step 3. Obtain the $L^2$ convergence of the estimators.

Suppose we select $K_n$ and $M_n$ such that they satisfy Assumption (A.7). Then it holds that

$$
\frac{e^{3M_n}}{K_n^k} \frac{e^{6M_n} K_n^{-3/2} \ln K_n}{\sqrt{n}} \to 0.
$$

Notice that

$$
\int_{a(\gamma^*)}^{b(\gamma^*)} G(1, \tilde{f}_U(\gamma^*)(U^*|Y, V), \lambda_0(Y), \alpha_0)dU^* = e^{-e^{a_0(1, v)}\tilde{\Lambda}_0(Y)}e^{\alpha_0(1, v)'}\lambda_0(Y)
$$

and

$$
\int_{a(\gamma^*)}^{b(\gamma^*)} G(1, \tilde{f}_U(\gamma)(U^*|Y, V), \hat{\lambda}_n(Y), \hat{\alpha}_n)dU^* = e^{-e^{\hat{\alpha}_n(1, v)'}\hat{\Lambda}_n(Y)}e^{\hat{\alpha}_n(1, v)'}\hat{\lambda}_n(Y).
$$

From the inequality (5.2), we obtain

$$
\int_{0}^{1} \int_{0}^{1} \left[ e^{-e^{\alpha_n(1, v)'}\hat{\Lambda}_n(y)}e^{\alpha_n(1, v)'}\hat{\lambda}_n(y) - e^{-e^{\alpha_0(1, v)'}\Lambda_0(y)}e^{\alpha_0(1, v)'}\lambda_0(y) \right]^2 dy dv 
\leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{-3/2} \ln K_n}{\sqrt{n}} \right).
$$

So it holds that for any $s \in [0, 1],$

$$
\int_{0}^{1} \left( e^{-e^{\alpha_n(1, v)'}\hat{\Lambda}_n(s)} - e^{-e^{\alpha_0(1, v)'}\Lambda_0(s)} \right)^2 dv 
\leq \int_{0}^{1} \int_{0}^{s} e^{-e^{\alpha_n(1, v)'}\hat{\Lambda}_n(y)}e^{\alpha_n(1, v)'}\hat{\lambda}_n(y)dy - \int_{0}^{s} e^{-e^{\alpha_0(1, v)'}\Lambda_0(y)}e^{\alpha_0(1, v)'}\lambda_0(y)dy \right]^2 dv 
\leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{-3/2} \ln K_n}{\sqrt{n}} \right).
$$

Since $\hat{\Lambda}_n(1) = 1,$ by choosing subsequence, we can assume that $\hat{\Lambda}_n(y)$ converges to a $\Lambda^*(y)$ uniformly on $[0, 1]$ and also suppose $\hat{\alpha}_n \to \alpha^*.$ Following the above inequality, it simply results in $\alpha^* = \alpha_0$ and $\Lambda^*(y) = \Lambda_0(y), \forall 0 < y < 1.$ Additionally, let $s = 1$ then

$$
\int_{0}^{1} \left( e^{-e^{\alpha_n(1, v)'}\hat{\Lambda}_n(y)} - e^{-e^{\alpha_0(1, v)'}\Lambda_0(y)} \right)^2 dv \leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{-3/2} \ln K_n}{\sqrt{n}} \right).
$$

Hence,$|\hat{\alpha}_n - \alpha_0|^2 \leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{-3/2} \ln K_n}{\sqrt{n}} \right).$ Moreover, $\|\hat{\Lambda}(y) - \Lambda_0(y)\|_{L^\infty[0, 1]} \to 0.$ This completes the proof. □

Furthermore, by using the properties of Sobolev space, we can obtain the consistency of the nuisance parameters in a Sobolev-norm.

**Theorem 5.2.** *(Consistency of Nuisance Parameters)* Under Assumptions (1-7), suppose that either of the two working models is true: $T$ is independent of $X$ given $V;$ or, the distribution of $C$ given $(X, V)$ follows a Cox proportional hazard model. Then in probability,

$$
\|\hat{\Lambda}(y) - \Lambda_0(y)\|_{W^1, \infty([0, 1])}, \|\tilde{f}_U(\gamma)(u|y, v) - f_U(\gamma^*)(u|y, v)\|_{W^1, \infty([a(\gamma^*), b(\gamma^*)] \times [0, 1]^2)} \to 0.
$$
Proof of Theorem 5.2

After repeating application of (5.2) and (5.3), we can further obtain
\[ \| \tilde{\lambda}_n(y) - \lambda_0(y) \|^2_{L^2([0,1])} \leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{3/2} \ln K_n}{\sqrt{n}} \right) \]
and
\[ \| \tilde{f}_U(\gamma)(u|y, v) - f_U(\gamma^*)(u|y, v) \|^2_{L^2([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} \leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{3/2} \ln K_n}{\sqrt{n}} \right). \]
Thus
\[ \| \tilde{f}_U(\gamma)(u|y, v) - f_U(\gamma^*)(u|y, v) \|^2_{L^2([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} \leq O_p(1) \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{3/2} \ln K_n}{\sqrt{n}} \right). \]

Now we can apply the following Sobolev interpolation inequality (cf. Adams (1975)) to further obtain the convergence rates for \( \tilde{f}_U(u|y, v; \gamma_n) \), \( \tilde{\lambda}_n(y) \) in other Sobolev norm.

(Sobolev Interpolation Theorem) If a function \( H(x) \) belongs \( W^k,2(R^d) \) \( (k > d/2) \) then for any \( w < k' - d/2 \),
\[ \| \nabla^w H(x) \|_{L^\infty} \leq C \| H(x) \|_{W^{k',2}} \| H(x) \|_{L^2}^{1-\tau_1}, \]
where \( \tau_1 = \frac{w+d/2}{k'} \).

Use Theorem A.2 in the appendix for B-spline functions and Assumption (A.5).
\[ \| \nabla_k \hat{\eta}_n(u, y, v) \|_{L^\infty([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} \leq C K_n^k \sum_{i_1, i_2, i_3=1} \| \hat{\eta}_{k_1, k_2, k_3} \|_{L^\infty} \leq O(M_n K_n^k), \]
where \( k_1 + k_2 + k_3 = k \). Hence
\[ \| \nabla_k \hat{f}_U(\gamma)(u|y, v) \|_{L^\infty([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} \leq C e^{(k+1)M_n M_n K_n^k}. \]

The Sobolev inequality tells that for \( (w_1, w_2, w_3) \) such that \( w_1 + w_2 + w_3 < k - 3/2 \),
\[ \| \nabla^w (\hat{f}_U(\gamma)(u|y, v) - f_U(\gamma^*)(u|y, v)) \|_{L^\infty([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} \leq C e^{(k+2)M_n \tau_1} K_n^{k \tau_1} \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{3/2} \ln K_n}{\sqrt{n}} \right)^{(1-\tau_1)/2}, \]
where \( \tau_1 = \frac{w+3/2}{k} \). Similarly,
\[ \| \nabla^w (\hat{\lambda}_n(y) - \lambda_0(y)) \|_{L^\infty([0,1])} \leq C e^{2M_n \tau_2} K_n^{\tau_2} \left( \frac{e^{3M_n}}{K_n^k} + \frac{e^{6M_n} K_n^{3/2} \ln K_n}{\sqrt{n}} \right)^{(1-\tau_2)/2}, \]
where \( \tau_2 = \frac{w+1/2}{k} \).

Therefore, if \( w \) satisfies that \( w < k - 3/2 \) and suppose \( K_n \) is chosen such that \( K_n \rightarrow \infty \) and

\[
K_n^{k+3(1-\tau_1)/4} (\ln K_n)^{1-\tau_2/n} \rightarrow \infty, \quad K_n^{k+3(1-\tau_2)/4} (\ln K_n)^{1-\tau_2/n} \rightarrow 0,
\]

then we can select \( M_n \) such that

\[
\| \hat{f}_{U(\gamma)}(U|Y,V) - f_{U(\gamma^*)}(U|Y,V) \|_{W^{1,\infty}([a(\gamma^*),b(\gamma^*)] \times [0,1]^2)} = o_p(1),
\]

\[
\| \hat{\lambda}_n(Y) - \lambda_0(Y) \|_{W^{1,\infty}([0,1])} = o_p(1).
\]

In particular, let \( w = 1 \). Then under Assumption (A.5) and Assumption (A.7), \( \tau_1, \tau_2 < 1/4 \). Therefore, both \( K_n^{k+3(1-\tau_1)/4} (\ln K_n)^{1-\tau_2/n} \) and \( K_n^{k+3(1-\tau_2)/4} (\ln K_n)^{1-\tau_2/n} \) are smaller than \( \frac{K_n^{k+4+9/16}(\ln K_n)^{1/2}}{n^{3/16}} \rightarrow 0 \). So we obtain

\[
\| \hat{\lambda}_n(Y) - \lambda_0(Y) \|_{W^{1,\infty}([0,1])}, \| \hat{f}_{U(\gamma)}(u|y,v) - f_{U(\gamma^*)}(u|y,v) \|_{W^{1,\infty}([a(\gamma^*),b(\gamma^*)] \times [0,1]^2)} \rightarrow 0.
\]

\( \square \)

The result in Theorem 5.2 can help to obtain a useful converge rate of the estimators in \( L^2 \)-norm, which is given in Theorem 5.3.

**Theorem 5.3.** (*Convergence Rate*) Under Assumptions (1-7), suppose that either of the two working models is true: \( T \) is independent of \( X \) given \( V \); or, the distribution of \( C \) given \( (X,V) \) follows a Cox proportional hazards model. Then

\[
| \hat{\alpha}_n - \alpha_0 |^2 \leq O_p(1)(\frac{1}{K_n^2}) + o_p(\frac{1}{\sqrt{n}}),
\]

\[
\| \hat{\lambda}_n(Y) - \lambda_0(Y) \|_{L^2([0,1])}^2 \leq O_p(1)(\frac{1}{K_n^2}) + o_p(\frac{1}{\sqrt{n}}),
\]

and

\[
\| \hat{f}_{U(\gamma)}(u|y,v) - f_{U(\gamma^*)}(u|y,v) \|_{L^2([a(\gamma^*),b(\gamma^*)] \times [0,1]^2)}^2 \leq O_p(1)(\frac{1}{K_n^2}) + o_p(\frac{1}{\sqrt{n}}).
\]

**Proof of Theorem 5.3**

Using the results from Theorem 5.1 and Theorem 5.2, redo the proof of Theorem 5.1. Since the class \( \mathcal{L}_n \) now is bounded by a constant and all the functions are bounded above from infinity and below from zero in probability and moreover, the class can be indexed by \( (\hat{f}_n(u|y,v), \hat{\lambda}_n(y), \alpha) \) which lies in a compact set in \( W^{1,\infty}([a(\gamma^*),b(\gamma^*)] \times [0,1]^2) \times W^{1,\infty}([0,1]) \times [-M,M]^2 \), the integration of entropy of the class \( \mathcal{L}_n \) is finite. Thus, we can apply the Donsker theorem to \( (P_n - P)\mathcal{L}_n \). Notice that
the functions in $\mathcal{L}_n$ converge to zero uniformly. So the left hand side of the inequality (5.1) is bounded by $o_p(1/\sqrt{n})$. It is obtained that

$$o_p\left(\frac{1}{\sqrt{n}}\right) \geq P \ln \frac{G(R, f_{Un}(U|Y,V), \lambda_n(Y), \alpha_0)}{G(R, f_{U0}(U|Y,V), \lambda_0(Y), \alpha_0)} + P \ln \frac{G(R, f_{U0}(U|Y,V), \lambda_0(Y), \alpha_0)}{G(R, f_{U0}(U|Y,V), \lambda_n(Y), \alpha_0)}.$$

We perform Taylor expansion on the right hand side. Since the first order in the expansion vanishes, we obtain that

$$\int_{[a(\gamma^*), b(\gamma^*)] \times [0,1]^2} [G(1, f_{U0(\gamma)}, \lambda_0(y), \alpha_0) - G(1, f_{U0(\gamma)}, \lambda_0(y), \alpha_0)]^2 dudvdv$$

$$\leq o_p\left(\frac{1}{\sqrt{n}}\right) + O_p(1)(|\gamma_n - \gamma^*|^2 + |a(\gamma_n) - a(\gamma^*)|^2 + |b(\gamma_n) - b(\gamma^*)|^2)$$

$$+ O_p(1)(||f_n(u|y,v) - f_{U0}(u|y,v)||_{L^2([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)} + ||\lambda_n(y) - \lambda_0(y)||_{L^2([0,1])}^2)$$

After repeating the arguments of Step 3 in the proof of Theorem 5.1, we obtain that each of

$$|\hat{\alpha}_n - \alpha_0|^2, ||\hat{f}_{U0}(u|y,v) - f_{U0}(u|y,v)||_{L^2([a(\gamma^*), b(\gamma^*)] \times [0,1]^2)}^2, ||\hat{\lambda}_n(y) - \lambda_0(y)||_{L^2([0,1])}^2$$

is bounded by $O_p(1)(\frac{1}{K_n}) + o_p(\frac{1}{\sqrt{n}})$. □

6 Asymptotic Normality

We start to show the asymptotic normality of $\hat{\alpha}_n$.

**Theorem 6.1.** (Asymptotic Normality of $\hat{\alpha}_n$) Under Assumptions (1-8), when either of the two working models is correct, $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ is asymptotically normal. Furthermore, when both working models are correct, the asymptotic variance of $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ is the same as the generalized Cramér-Rao bound.

**Proof of Theorem 6.1:**

The proof is direct if we can write $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ in the form of an empirical process. However, to do that, we need derive what we call the "pseudo least favorable direction" (a definition of the least favorable direction is elaborated in Murphy and van der Vaart (1996, 1998)). In detail, the whole proof can be divided into the following steps: first, we give the definition of the "pseudo least favorable direction" for the parameter $\alpha$; second, we prove the existence of the "pseudo least favorable direction"; third, this direction is verified to be in a smooth Sobolev space; fourth, we find the projection of this least favorable direction on the sieve space; fifth, the empirical process for $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$ is derived by linearization; finally, we complete the proof of Theorem 6.1.
The following consists the details of each step.

Step 1. We give the definition of the "pseudo least favorable direction" when $\gamma^*$ is known.

Suppose $\gamma$ is already known to equal $\gamma^*$. Then the whole parameters contain $(\alpha, \lambda(y), f_{U,\gamma^*}(u|y,v))$ of which the nuisance parameters are $(f_{U,\gamma^*}(u|y,v), \lambda(y))$ and they are denoted as $\psi$. The "pseudo least favorable direction" in this context is defined as a tangent function $h(u,y,v) = (h_1(u,y,v), h_2(y))$ for $\psi$ at the true parameter $(\psi_0, \alpha_0)$ in the parameter space. In other words, we can find a sub-model $(\psi_*, \alpha_*)$ such that

$$ (\psi_*, \alpha_*)|_{\epsilon=0} = (\psi_0, \alpha_0) $$

and

$$ \frac{d}{d\epsilon} \psi_0|_{\epsilon=0} = h(u,y,v). $$

In addition, if define

$$ l(\psi, \alpha; \gamma^*) = r \ln f_{U,\gamma^*,T|V}(u,y|v) + (1-r) \ln \int_0^\infty f_{U,\gamma^*,T|V}(u,s|v)ds $$

where $f_{U,\gamma^*,T|V}(u,y|v)$ is the conditional density of $(U(\gamma^*), T)$ given $V$, then the "pseudo least favorable direction" $h(u,y,v) = (h_1(u,y,v), h_2(y))$ satisfies that

$$ (l^*_\psi(\psi_0, \alpha_0; \gamma^*) l_\psi(\psi_0, \alpha_0; \gamma^*))[h(u,y,v)] = l^*_\psi(\psi_0, \alpha_0; \gamma^*) l_\alpha(\psi_0, \alpha_0; \gamma^*), a.s. $$

where $l_\psi(\psi_0, \alpha_0; \gamma^*)[h(u,y,v)]$ is the derivative of $l$ with respective $\psi$ along the direction $h(U,Y,V)$ and $l^*_\psi(\psi_0, \alpha_0; \gamma^*)$ is the adjoint operator of $l_\psi(\psi_0, \alpha_0; \gamma^*)$ in the Hilbert space $L^2(P)$; $l_\alpha(\psi_0, \alpha_0; \gamma^*)$ is the derivative of $l$ respective to $\alpha$; $P$ is the probability measure generated by the joint density of $(U^* = \gamma^*(X,V), Y = T \wedge C, V, R)$ (it is equivalent to the Lebesgue measure on $[a(\gamma^*), b(\gamma^*)] \times [0,1]^2$).

Our definition of the "pseudo least favorable direction" imitates the definition of the least favorable direction (Murphy and van der Vaart(1998)) except that the function $l(\psi, \alpha, \gamma^*)$ is not a full likelihood function for $\alpha$ since $\gamma^*$ is unknown .

Step 2. We derive the existence of the pseudo least favorable direction.

In our problem, notice that the tangent space for $\psi$ is a subspace of

$$ H = \{h(u,y,v) = (h_1(u,y,v), h_2(y)) : \int_{a(\gamma^*)}^{b(\gamma^*)} h_1(u,y,v)du = 0, $$

$$ \int_0^1 h_2(y)dy = 0, h(u,y,v) \in L^2(P) \}.$$ 

$H$ is a Hilbert space with norm defined as

$$ < h, h >^1_H = \sqrt{\|h_1(U,Y,V)\|_{L^2(P)}^2 + \|h_2(V)\|_{L^2(P)}^2}. $$
We will show that there exists a unique \( h \in H \) such that at \( \psi_0, \alpha_0, \gamma = \gamma^* \),

\[
(l^*_\psi(\psi, \alpha; \gamma)l_\psi(\psi, \alpha; \gamma))[h(u, y, v)] = l^*_\psi(\psi, \alpha; \gamma)l_\alpha(\psi, \alpha; \gamma), \text{ a.s.}
\]

We put it into a lemma.

**Lemma 6.1.** Under Assumptions (1-6), there exists a unique \( h \in H \) such that

\[
(l^*_\psi(\psi, \alpha_0; \gamma^*)l_\psi(\psi, \alpha_0; \gamma^*))[h(u, y, v)] = l^*_\psi(\psi, \alpha_0; \gamma^*)l_\alpha(\psi, \alpha_0; \gamma^*), \text{ a.s.}
\]

**Proof of Lemma 6.1**

After some calculation,

\[
l_\psi(\psi, \alpha_0; \gamma^*)[h] = \int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)} + \frac{h_1(u, y, v)}{f_U(\gamma^*)(u|y,v)},
\]

\[
+ (1 - r)\int_0^\infty f_U(\gamma^*), T, V(u, y|v)\left(-e^{2(1, v')}\int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)} + \frac{h_1(u, y', v)}{f_U(\gamma^*)(u|y',v)}\right)dy'.
\]

We denote \(<,>_H\) as inner product in Hilbert space \( H \) and use \( O^+(1) \) in the following as some positive constant. Then

\[
<l^*_\psi(\psi, \alpha_0; \gamma^*)l_\psi(\psi, \alpha_0; \gamma^*)[h(u, y, v)], h(u, y, v)>_H
\]

\[
= \|l_\psi(\psi, \alpha_0; \gamma^*)[h(u, y, v)]\|^2_{L^2(P)}
\]

\[
\geq \|R(-e^{2(1, v')}\int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)} + \frac{h_1(u, y, v)}{f_U(\gamma^*)(u|y,v)})\|^2_{L^2(P)}
\]

\[
\geq O^+(1)\int_{(u,y,v)\in[a(\gamma^*), b(\gamma^*)] \times [0,1]^2} (-e^{2(1, v')}\int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)} + \frac{h_1(u, y, v)}{f_U(\gamma^*)(u|y,v)})^2dy'dydv
\]

\[
\geq O^+(1)\int_{(y,v)\in[0,1]^2} dydv \int_{a(\gamma^*)}^{b(\gamma^*)} h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)}
\]

\[
+ \frac{h_1(u, y, v)}{f_U(\gamma^*)(u|y,v)} dy'dydv
\]

\[
\geq O^+(1)\int_{(y,v)\in[0,1]^2} (-e^{2(1, v')}\int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)})^2dydv
\]

\[
+ \int_{(u,y,v)\in[a(\gamma^*), b(\gamma^*)] \times [0,1]^2} h_1(u, y, v)^2dy'dydv
\]

(since \( \int_{a(\gamma^*)}^{b(\gamma^*)} h_1(u, y, v)du = 0 \))

\[
\geq O^+(1)\int_{(y,v)\in[0,1]^2} \lambda_0(y)(-e^{2(1, v')}\int_0^y h_2(s)ds + \frac{h_2(y)}{\lambda_0(y)})^2dydv
\]

\[
+ \int_{(u,y,v)\in[a(\gamma^*), b(\gamma^*)] \times [0,1]^2} h_1(u, y, v)^2dy'dydv
\]

\[
\geq O^+(1)\int_{(y,v)\in[0,1]^2} h_2(y)^2dydv - 2\int_{(y,v)\in[0,1]^2} e^{2(1, v')} h_2(y)\int_0^y h_2(s)dsdydv
\]

19
Second, from the consistency proof, we know \( \hat{\eta} \) where \( \eta \) is invertible in \( H \) is one to one mapping from \( H \) to \( H \); thus

\[
\mathcal{O}(1) < h, h > H.
\]

So by using the Lax-Milgram theorem, we conclude that the operator \( l^*_\psi(\psi_0, \alpha_0; \gamma^*)l_\psi(\psi_0, \alpha_0; \gamma^*) \) is one to one mapping from \( H \) to \( H \); thus

\[
l^*_\psi(\psi_0, \alpha_0; \gamma^*)l_\psi(\psi_0, \alpha_0; \gamma^*)
\]

is invertible in \( H \). Lemma 6.1 is proved. \( \square \)

Step 3. The smoothness of the least favorable direction is obtained. In fact, we obtain

\[
\text{Lemma 6.2.}
\]

\[
h(u, y, v) \in W^{k, \infty}([a(\gamma^*), b(\gamma^*)] \times [0, 1]^2).
\]

The proof of this lemma is given in the appendix.

Step 4. The projection of \( h(u, y, v) \) on the tangent space of the sieve space is constructed.

Since from Lemma 6.2, \( h(u, y, v) \) lies in \( W^{k, \infty}([a(\gamma^*), b(\gamma^*)] \times [0, 1]^2) \), we can find an approximate direction \( h_n(u, y, v) \) which is one tangent direction for the nuisance parameters at \( \hat{\psi}_n = (\hat{f}_U(\gamma)(u|y, v), \hat{\lambda}_n(y)) \) in the sieve space. The procedure is as follows:

At first, recall that the nuisance parameters for the sieve space have the forms

\[
\frac{e^{\eta(u, y, v)}}{\int_{a(\gamma_n)}^{b(\gamma_n)} e^{\eta(u, y, v)} du}, e^{\xi(y)}
\]

where \( \eta(u, y, v), \xi(y) \) have the expression specified in \( S_n(m, K_n, M_n) \) and moreover, \( \eta(0, y, v) = 0, \int_0^1 e^{\xi(y)} dy = 1 \). Second, from the consistency proof, we know \( (\hat{\eta}_n(u, y, v), \hat{\xi}_n(y)) \) are the interior points in \( S_n(m, K_n, M_n) \) under \( L^2 \)-norm. Hence, by simple computation, the tangent vectors \( h_n(u, y, v) \) for the nuisance parameters at \( \hat{\psi}_n = (\hat{f}_U(\gamma)(u|y, v), \hat{\lambda}_n(y)) \) have the form as

\[
(\hat{f}_U(\gamma)(u|y, v))e_1(u, y, v) = \hat{f}_U(\gamma)(u|y, v) \frac{\int_{a(\gamma_n)}^{b(\gamma_n)} e^{\hat{\eta}_n(u, y, v)} \xi_1(u, y, v) du}{\int_{a(\gamma_n)}^{b(\gamma_n)} e^{\hat{\eta}_n(u, y, v)} du}, \hat{\lambda}_n(y) e_2(u, y, v)
\]

where

\[
\xi_1(u, y, v) = \sum_{i_1, i_2, i_3} c_{i_1i_2i_3} N_{i_1}^m(u) N_{i_2}^m(y) N_{i_3}^m(u, y, v), \xi_2(y) = \sum_{i=1}^{m+K_n} c_i N_{i}^m(y)
\]
and
\[ \xi_1(0, y, v) = 0, \int_0^1 \hat{\lambda}_n(y) \xi_2(y) dy = 0. \]

One good approximation to the "pseudo least favorable direction" \( h(u, y, v) = (h_1(u, y, v), h_2(y)) \) is to choose \( h_n(u, y, v) = (h_1^n(u, y, v), h_2^n(y)) \) whose corresponding \( (\xi_1(u, y, v), \xi_2(y)) \) satisfy:

\[ \xi_1(u, y, v) = Q^*(h_1(u, y, v) / f_{U(\gamma^*)}(u|y, v)) \]

for some extendedly defined \( h_1(u, y, v) \) such that \( h_1(u, y, v) = 0 \) if \( u < a(\gamma^*)/2 \); and

\[ \xi_2(y) = Q(h_2(y)/\lambda_0(y)) - \int_0^1 \hat{\lambda}_n(y) Q(h_2(y)/\lambda_0(y)) dy. \]

Here the operators \( Q \) and \( Q^* \) were defined in Section 3.

From the conclusion in Lemma 6.2, we can easily show that

\[
\|h_n(u, y, v) - h(u, y, v)\|_{L^2(P)} \\
\leq \|\hat{f}_{U(\gamma)}(u|y, v) - f_{U(\gamma^*)}(u|y, v)\|_{L^2(P)} + \|\hat{\lambda}_n(y) - \lambda_0(y)\|_{L^2(P)} \\
+ \|Q^*(h_1(u, y, v) / f_{U(\gamma^*)}(u|y, v)) - h_1(u, y, v) / f_{U(\gamma^*)}(u|y, v)\|_{L^2(P)} \\
+ \|Q(h_2(y)/\lambda_0(y)) - h_2(y)/\lambda_0\|_{L^2(P)} \\
+ O_p(\|\hat{\gamma}_n - \gamma^*\| + |a(\hat{\gamma}_n) - a(\gamma^*)| + |b(\hat{\gamma}_n) - b(\gamma^*)|) \\
\leq O(\frac{1}{\sqrt{K^k}}) + O_p(\frac{1}{\sqrt{n}}).
\]

Step 5. Derivation of the empirical process for \( \sqrt{n}(\hat{\alpha}_n - \alpha_0) \).

After we obtain the approximate direction \( h_n(u, y, v) \), we can start to prove Theorem 6.1. First, when either working model is correct, \( l(\psi_0, \alpha_0; \gamma^*) \) is exactly the log-likelihood of the observation \( (T \wedge C = y_i, R = r_i, V = v_i, U(\gamma^*) = \gamma^* (x_i, v_i') \) for \( (\psi, \alpha) \). Hence, the fact that the expectation of the score function is 0 and the fact that the expectation of the square score is the same as the fisher information is also true.

Since \( (\hat{\psi}_n, \hat{\alpha}_n) \) maximizes the log-likelihood among the sieve space,

\[
P_n l_{\psi}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)[h_n] + P_n l_{\alpha}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n) = 0.
\]

Therefore,

\[
(P_n - P) l_{\psi}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)[h_n] + (P_n - P) l_{\alpha}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)
\]

\[
= -P \{ l_{\psi}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)[h_n] + l_{\alpha}(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n) \}
\]
 Recall the result in Theorem 5.3 and the fact \( \|h\|_\infty \) on the right hand side then we obtain

\[
(P_n - P)l_\psi(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)[h_n] + (P_n - P)l_\alpha(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n) = -P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[\hat{\psi}_n - \psi_0, h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)[\hat{\alpha}_n - \alpha_0])
\]

Further approximate the \( n \) by \( h \) on the right hand side then we obtain

\[
(P_n - P)l_\psi(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n)[h_n] + (P_n - P)l_\alpha(\hat{\psi}_n, \hat{\alpha}_n; \hat{\gamma}_n) = -P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[\hat{\psi}_n - \psi_0, h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)[\hat{\alpha}_n - \alpha_0])
\]

However, \( h(u, y, v) \) satisfies that

\[
l_\psi^*(\psi_0, \alpha_0; \gamma^*)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] = l_\psi^*(\psi_0, \alpha_0; \gamma^*)l_\alpha(\psi_0, \alpha_0; \gamma^*), a.s.
\]

so

\[
P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[\hat{\psi}_n - \psi_0, h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)[\hat{\alpha}_n - \alpha_0]) = -<l_\psi^*(\psi_0, \alpha_0; \gamma^*)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] - l_\alpha(\psi_0, \alpha_0; \gamma^*), \hat{\psi}_n - \psi_0 >_H = 0.
\]

We can apply the central limit theorem for the Donsker class to the left hand side by noticing that this class is indexed by both the bounded functions \((\hat{\psi}_n, h_n) \in W^{1, \infty} \) (Lemma 6.2) and \( \hat{\alpha}_n \in [-M, M]^2, \hat{\gamma}_n \in [-M, M]^{D+1} \). Finally, we have

\[
-P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)\} \sqrt{n}(\hat{\alpha}_n - \alpha_0) = \sqrt{n}(P_n - P)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + \sqrt{n}(P_n - P)l_\alpha(\psi_0, \alpha_0; \gamma^*)
\]

Recall the result in Theorem 5.3 and the fact \( \|h - h_n\|_2^2 \) then we obtain

\[
-P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)\} \sqrt{n}(\hat{\alpha}_n - \alpha_0) = \sqrt{n}(P_n - P)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + \sqrt{n}(P_n - P)l_\alpha(\psi_0, \alpha_0; \gamma^*)
\]

Finally, we have

\[
-P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)\} \sqrt{n}(\hat{\alpha}_n - \alpha_0) = \sqrt{n}(P_n - P)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + \sqrt{n}(P_n - P)l_\alpha(\psi_0, \alpha_0; \gamma^*)
\]
\[
\sqrt{n}(P_n - P)l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + \sqrt{n}(P_n - P)l_\alpha(\psi_0, \alpha_0; \gamma^*)
\]
\[+ P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)\}\sqrt{n}(\hat{\gamma}_n - \gamma^*)
\]
\[+ O_p(\sqrt{n}K_n) + o_p(1).
\]

Step 6. Completion of the proof.

First, we need the following result.

**Lemma 6.3.**

\[-P\{l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)\} > 0.
\]

**Proof of Lemma 6.3**

It is sufficient to show the matrix

\[
M = -\begin{bmatrix}
P_l\psi(\psi_0, \alpha_0; \gamma^*)[h, h] & P_l\alpha(\psi_0, \alpha_0; \gamma^*)[h] \\
P_l\psi(\psi_0, \alpha_0; \gamma^*)[h] & P_l\alpha(\psi_0, \alpha_0; \gamma^*)
\end{bmatrix}
\]

is positive definite since if so,

\[-(P_l\alpha(\psi_0, \alpha_0; \gamma^*) - P_l\psi(\psi_0, \alpha_0; \gamma^*)[h]'[P_l\psi(\psi_0, \alpha_0; \gamma^*)[h, h]]^{-1}P_l\alpha(\psi_0, \alpha_0; \gamma^*)[h])\]

is positive. However, but

\[P_l\psi(\psi_0, \alpha_0; \gamma^*)[h, h] + P_l\alpha(\psi_0, \alpha_0; \gamma^*)[h] = 0;\]

thus,

\[-(P_l\alpha + P_l\psi[h]) > 0,
\]

which is the result stated in Lemma 6.3.

First for any \(\beta = (\beta_1, \beta_2),\)

\[\beta' M \beta = P[l_\psi(\psi_0, \alpha_0; \gamma^*)[(\beta_1 h, \beta_2)]]^2 \geq 0.
\]

Second, if \(\beta' M \beta = 0\) then \(l_\psi(\psi_0, \alpha_0; \gamma^*)[(\beta_1 h, \beta_2)] = 0,\ a.s.\) This implies that

\[0 = \beta_1 (h_1(u, y, v) - f_U(\gamma^*)(u|y, v)h_2(y)/\lambda_0(y)
\]
\[+ f_U(\gamma^*)(u|y, v)e^{\alpha_0(1, v)'}\int_0^y h_2(s)ds + \beta_2 (e^{-\alpha_0(1, v)'(1, v)'\Lambda_0(y) + (1, v)'})
\]

So \(\beta = 0,\ i.e.,\ M\) is positive definite. \(\square\)

Therefore \(\sqrt{n}(\hat{\alpha}_n - \alpha_0)\) is asymptotically normal. Moreover, its asymptotic variance can be expressed in terms of the variance of the influence function

\[-\{P[l_\psi(\psi_0, \alpha_0; \gamma^*)[h] + l_\alpha(\psi_0, \alpha_0; \gamma^*)]\}^{-1}
\]

23
This implies that when both working models are right, $\hat{\alpha}$ function, and therefore it equals 0 whatever $\gamma$ takes. So $P[l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)]$ is always an expectation of a score function, and therefore it equals 0 whatever $\gamma$ is, i.e.,

$$P[l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)] = 0.$$ 

This implies that when both working models are right, $\hat{\alpha}_n$ has an influence function which is

$$-\{P[l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)]\}^{-1}\{l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)\}.$$

However, $l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)$ is simply the score function along the direction $(h, 1)$ so it must be on the space spanned by all the scores. Then $\sqrt{n}(\hat{\alpha}_n - \alpha_0)$’s asymptotic variance is the same as the generalized Cramér-Rao bound. The proof is completed. □

## 7 Heuristic Variance Estimation

We propose some heuristic approaches for estimating the asymptotic variance of $\hat{\alpha}_n$. No rigorous justification is made here. One approach for estimating the asymptotic variance of $\hat{\alpha}_n$ is by bootstrapping. Though, computationally, this approach can be conveniently programmed and realized, we are not very confident that such bootstrapped estimator can converge to the asymptotic variance of $\hat{\alpha}_n$ in a large sample. An alternative approach we will propose is the direct estimation of the influence function of $\hat{\alpha}_n$. Recall from the last section that, $\hat{\alpha}_n$ has an influence function

$$-\{P[l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)]\}^{-1}\{l_\psi(\psi_0, \alpha_0; \gamma)[h] + l_\alpha(\psi_0, \alpha_0; \gamma)\},$$

where $\psi = (f_U(\gamma)(u|y, v), \lambda_0(y))$ and

$$l(\psi_0, \alpha_0; \gamma) = r \ln f_U(\gamma)(u|y, v)e^{-e^{\alpha_0(1,v)'}\lambda_0(y)}e^{\alpha_0(1,v)'}\lambda_0(y)$$

$$+ (1 - r) \ln \int_y^{\infty} f_U(\gamma)(u|s, v)e^{-e^{\alpha_0(1,v)'}\Lambda_0(s)}e^{\alpha_0(1,v)'}\Lambda_0(s)ds.$$ 

Moreover, $h(u, y, v) = (h_1(u, y, v), h_2(y))$ is the solution to the equation

$$l_\psi^*l_\psi[h] = l_\alpha^*l_\alpha.$$
and $S_{\gamma^*}(\gamma^*; Y, R, X, V)$ is the influence function of $\hat{\gamma}_n$. Since $S_{\gamma^*}(\gamma^*; Y, R, X, V)$ has an explicit expression and can be estimated consistently (cf. Zeng (2001)), the remaining key issues are how to estimate or calculate

$$A_1 = -\{P[l_{\psi\alpha}(\psi_0, \alpha_0; \gamma^*)[h] + l_{\alpha\alpha}(\psi_0, \alpha_0; \gamma^*)]\}^{-1},$$

$$A_2 = P[l_{\psi\gamma}(\psi_0, \alpha_0; \gamma^*)[h] + l_{\alpha\gamma}(\psi_0, \alpha_0; \gamma^*)],$$

and

$$A_3(Y, R, X, V) = l_{\psi}(\psi_0, \alpha_0; \gamma^*)[h] + l_{\alpha}(\psi_0, \alpha_0; \gamma^*).$$

In the below, we in turn describe their heuristic estimation.

$A_1$ in fact is the generalized Cramér-Rao efficiency bound for $\alpha$ in the likelihood function of $(Y, R, U(\gamma^*), V)$, i.e., $l(\psi, \alpha; \gamma^*)$, assuming $\gamma^*$ is known. Therefore, if define a profiled likelihood for $\alpha$ by

$$p_n(\alpha; \gamma^*) = \max_{\psi \in S_n} P_n l(\psi, \alpha; \gamma^*),$$

then according to Murphy and van der Vaart (2000), $A_1$ can be approximated by

$$-p_n(\hat{\alpha}_n + \epsilon_n, \hat{\gamma}_n) - 2p_n(\hat{\alpha}_n, \hat{\gamma}_n) + p_n(\hat{\alpha}_n - \epsilon_n, \hat{\gamma}_n)$$

for some $\epsilon_n = O(1/\sqrt{n})$.

We define a class of estimators for $\alpha$, which are indexed by $\gamma$ and denoted as $\hat{\alpha}(\gamma)$. $\hat{\alpha}(\gamma)$ is obtained using the same maximization procedure as in obtaining $\hat{\alpha}_n$, except that $\gamma$ is regarded as constant. Obviously, $\hat{\alpha}_n = \hat{\alpha}(\hat{\gamma}_n)$. Then the term $A_2$, whose similar expression was observed and estimated in Zeng (2001), can be considered as the factor of the change in $\hat{\alpha}(\gamma)$ caused by the change of $\gamma$ around $\gamma^*$, i.e.,

$$A_2 \approx \nabla_\gamma \hat{\alpha}(\gamma)|_{\gamma = \gamma^*}.$$ 

Using the similar arguments given in Zeng (2001), we can show that

$$\frac{\hat{\alpha}(\hat{\gamma}_n + \epsilon_n e) - \hat{\alpha}_n}{\epsilon_n}$$

approximates $P[l_{\psi\gamma}(\psi_0, \alpha_0; \gamma^*)[h] + l_{\alpha\gamma}(\psi_0, \alpha_0; \gamma^*)]e$ for any unit vector $e$, where $\epsilon_n$ satisfies that $\sqrt{n}\epsilon_n \to \infty$. Hence, $A_2$ can be estimated by the above numerical differentiation.

We have to compute the function $h(u, y, v) = (h_1(u, y, v), h_2(y))$ in estimating the function $A_3(Y, R, X, V)$. However, as already seen in the proof of Lemma 6.2, we can obtain a numerical solution for $h(u, y, v)$ by solving a second-order differential equation with smooth boundary conditions. This step needs much effort of computation.

The final estimator for the asymptotic variance of $\hat{\alpha}_n$ is

$$\hat{A}_1^{-1}P_n[\hat{A}_3(Y, R, X, V) + \hat{A}_2 \hat{S}_{\gamma^*}(Y, R, X, V)]^2 \hat{A}_1^{-1}.$$
8 Discussion

For right-censored data, when the dependence between lifetime and censoring time is explained by high-dimensional covariates, we utilize two working models to condense this high-dimensional information thus derive the estimator of the treatment effect by maximizing some pseudo likelihood function. We have shown that the estimator is consistent and asymptotically normal when either working model is correct. Computationally, the estimation procedure can be implemented through constrained optimization algorithms and its variance can be estimated using either bootstrap or numerical solutions.

For simplicity, the working model for $T$ given $(X,V)$ given in Section 2 is assumed to be the same as $T$ given $V$. This may seem very restrictive. However, in practice, any semiparametric model can be adopted as the working model for $T$ given $(X,V)$. For example, suppose that we use a semiparametric model for $T$ given $(X,V)$ as follows:

$$f_{T|X,V}(y|x,v) = g(y, \beta(x,v)^t),$$

then the condensed information will include

$$(\beta(X,V)^t, \gamma(X,V)^t, V);$$

hence, the estimator of $\alpha$ can be derived by maximizing

$$P_n \ln \{ [e^{\alpha(1,V)^t}\lambda(Y)e^{-\Lambda(Y)}e^{\alpha(1,V)^t}]^{f_{U_1,U_2}(\hat{U_1},\hat{U_2}|Y,V)} \}
\int_{Y}^{\infty} e^{\alpha(1,V)^t}\lambda(s)e^{-\Lambda(s)}e^{\alpha(1,V)^t} f_{U_1,U_2}(\hat{U_1},\hat{U_2}|s,V)ds \}^{1-R}$$

over a sieve space of the parameters $(\alpha, f_{U_1,U_2}(u_1,u_2|y,v), \lambda(y))$, where

$$U_1 = \beta(X,V)^t, U_2 = \gamma(X,V)^t, \hat{U_1} = \hat{\beta}_n(X,V)^t,$$

and $\hat{U_2} = \hat{\gamma}_n(X,V)^t$ for some estimators $\hat{\beta}_n, \hat{\gamma}_n$. The slight difference from the previous context is that B-splines in the sieve space are constructed on a four-dimensional space. Consequently, with some change in the technical assumptions, one of the following two conclusions is expected to be true: if the semiparametric model for $T$ given $(X,V)$ does not satisfy the constraint that $h_{T|V}(y|v) = \lambda(y)e^{\alpha(1,v)^t}$, the consistency of $\hat{\alpha}_n$ holds only if the working model for $C$ given $(X,V)$ is specified correctly since the working model for $T$ given $(X,V)$ is always misspecified; on the contrary, if the working model for $T$ given $(X,V)$ satisfies the constraint, the double robustness of $\hat{\alpha}_n$ holds as before. Though we expect the latter consequence, to specify a working model for $T$ given $(X,V)$ satisfying the constraint is not easy except that we assume $T$ depends on $(X,V)$ only via $V$. 

26
In Zeng (2001), a whole section explains why such a pseudo-likelihood approach can provide a double robust estimator. The essential idea is that the likelihood function for the reduced data, which contain the event times, the censoring indicators, and the condensed low-dimensional information from the working models, approximates a partial likelihood function when either working model is correct. More details can be seen in Zeng (2001) and we do not want to repeat here.

9 Acknowledgments.

This work is part of my Ph.D. Dissertation advised by Susan Murphy at the University of Michigan. I owe many thanks to her for numerous discussions and helpful comments.

APPENDIX

A Basic Properties of Spline Functions

Spline functions, especially the well-known cubic splines, are important functions used in numerical computation and statistical estimation of probability densities. A general definition for spline functions is given in the following (Shumaker(1981)):

Let \([a,b]\) be a finite closed interval and let \(\Delta = \{x_i\}_{i=1}^k\) be a partition of \([a,b]\), i.e.,

\[
a = x_0 < x_1 < \ldots < x_k < x_{k+1} = b.
\]

Define \(I_i = [x_i, x_{i+1})\), \(i = 0, 1, \ldots, k - 1\); \(I_k = [x_k, x_{k+1}]\). In addition, let \(m\) be a positive integer and \(\mathcal{M} = (m_1, \ldots, m_k)\) be a vector of integers with \(1 \leq m_i \leq m, i = 1, 2, \ldots, k\). Suppose \(\mathcal{P}_m\) to be a class consisting of all polynomials with degrees at most \(m\). Then a polynomial spline space of order \(m\) with knots \(x_1, \ldots, x_k\) of multiplicity \(m_1, \ldots, m_k\) is defined as

\[
S(\mathcal{P}_m; \mathcal{M}; \Delta) = \{ s : \text{there exist polynomials } s_0, \ldots, s_k \\
\text{in } \mathcal{P}_m \text{ such that } s(x) = s_i(x) \text{ for } x \in I_i, i = 0, 1, \ldots, k \\
\text{and } D^j s_{i-1}(x_i) = D^j s_i(x_i) \text{ for } j = 0, 1, \ldots, m - 1 - m_i, i = 1, \ldots, k \}
\]

Indeed this space consists of piecewise polynomials with certain smoothness joining end points of each subinterval. Clearly, the higher \(m_i\) are, the less smooth the function is at the knots. Usually, we let \(m_1 = m_2 = \ldots = m_k = 1\).

It turns out that \(S(\mathcal{P}_m; \mathcal{M}; \Delta)\) is a linear space of dimension \(m + K\) where \(K = \sum_{i=1}^k m_i\) (Shumaker(1981), Theorem 4.4). A typical basis for this linear space consists of so called B-spline functions, which are constructed based on an extended partition \(\tilde{\Delta} = \{y_i\}_{i=1}^{2m+K}\) of \([a, b]\):

\[
y_1 \leq \ldots \leq y_m \leq a; b \leq y_{m+K+1} \leq \ldots \leq y_{2m+K}
\]
where each $x_i$ has multiplicity of $m_i$ for $i = 1, \ldots, k$ (the $\tilde{\Delta}$ is called an extended partition associated with $S(\mathcal{P}_m; \mathcal{M}; \Delta)$). Then, B-spline functions associated with this partition are defined as

$$N_i^m(x) = (-1)^m(y_{i+m} - y_i)[y_i, \ldots, y_{i+m}]f_x I_{a \leq x \leq b},$$

where $f_x(y) = (x - y)_{+}^{m-1}$ and $[y_i, \ldots, y_{i+m}]f$ is a functional of $f$, defined as

$$[y_i, \ldots, y_{i+m}]f = \frac{\det \begin{pmatrix} 1 & \ldots & 1 \\ y_i & \ldots & y_{i+m} \\ \vdots & \ldots & \vdots \\ y_{i+m-2} & \ldots & y_{i+m-2} \\ f(y_i) & \ldots & f(y_{i+m}) \end{pmatrix}}{\det \begin{pmatrix} 1 & \ldots & 1 \\ y_i & \ldots & y_{i+m} \\ \vdots & \ldots & \vdots \\ y_{i+m-2} & \ldots & y_{i+m-2} \\ y_i & \ldots & y_i \end{pmatrix}}.$$

As shown in Theorem 4.9 of Shumaker (1981), $\{N_i^m(x), i = 1, \ldots, m+K\}$ form a basis for $S(\mathcal{P}_m; \mathcal{M}; \Delta)$ and moreover, they satisfy

$$1 \geq N_i^m(x) > 0 \text{ iff } x \in (y_i, y_{i+m}); \sum_{i=1}^{m+K} N_i^m(x) = 1, x \in [a, b].$$

Because of this property, $\{N_i^m(x)\}$ are sometimes called normalized version of B-splines.

If $y_i < y_{i+m}$, then $N_i^m(x)$ is called the $m$th order B-spline associated with the knots $y_i, \ldots, y_{i+m}$. Furthermore, suppose

$$y_i \leq \ldots \leq y_{i+m} = \tau_1, \ldots, \tau_1, \ldots, \tau_d, \ldots, \tau_d$$

where each $\tau_i$ has a multiplicity of $l_i$, then (Shumaker (1981), Theorem 4.14)

$$N_i^m(x) = \sum_{j=1}^{d} \sum_{k=1}^{l_j} \alpha_{jk}(x - \tau_j)_{+}^{m-k}$$

with $\alpha_{jl} \neq 0$. Moreover,

$$D_kN_i^m(\tau_j) = D_kN_i^m(\tau_j), k = 0, 1, \ldots, m - l_j - 1, j = 1, 2, \ldots, d.$$  

Notice that the lower the multiplicity are, the more smoothness $N_i^m(x)$ has.

Computationally, these B-splines evaluated at $x$ can be obtained from an iterative equation (Shumaker (1981), Theorem 4.15): define

$$Q_i^m(x) = \begin{cases} \frac{N_i^m(x)}{(y_{i+m} - y_i)}, & y_i < y_{i+m} \\ 0, & O.W. \end{cases}$$
then
\[ Q_i^m(x) = \frac{(x - y_i)Q_{i+1}^{m-1}(x) + (y_{i+m} - x)Q_{i}^{m-1}(x)}{y_{i+m} - y_i}, \quad m \geq 2. \]

In a very special case when the knots are equally spaced, i.e., \( y_{i+1} - y_i = h \) for some constant \( h \) (so multiplicity are all equal to 1), normalized B-splines turn out to have simple forms. Let
\[ N_i^m(x) = m! \sum_{i=0}^{m} (-1)^i \binom{m}{i} (x - i)^{m-1} \]
be a function defined for \( x \in [0, m] \). Then
\[ N_i^m(x) = N_i^m(\frac{x - y_i}{h}), y_i \leq x \leq y_i + mh \]
i.e., the functions originating from the same function by shifting and scaling, are exactly the same B-spline basis for \( S(P_m; M; \Delta) \) as defined before. Moreover,
\[ \|N_i^m(x)\|_{L^1[0,m]} = \|N_i^m(x)\|_{L^\infty[0,m]} = 1. \]

One property regarding the bound of the derivatives of B-splines is in the following (Shumaker (1981), Theorem 4.22):

**Theorem A.1.** Let \( N_i^m(x) \) be the normalized B-spline over the knots \( y_i \leq ... \leq y_{i+m} \). Suppose \( x \in [y_l, y_{l+1}) \). Define
\[ \Delta_{i,l,j} = \max \{ (y_{v+j} - y_i) : y_i \leq y_v \leq y_l < y_{l+1} \leq y_{v+j} \leq y_{l+m} \} \]
Then for any \( \sigma > 0, \Delta_{i,l,m-\sigma+1} > 0, \)
\[ |D_+^\sigma N_i^m(x)| \leq \frac{2^\sigma (m-1)!}{(m-\sigma-1)! \Delta_{i,l,m-\sigma} \Delta_{i,l,m-\sigma}} \leq \frac{2^\sigma (m-1)!}{(m-\sigma-1)!(y_{l+1} - y_l)}\sigma. \]

As mentioned above, the linear space \( S(P_m; M; \Delta) \) has its normalized B-spline basis \( \{N_i^m(x)\}_{i=1}^{m+K} \). Hence, we can define a dual basis \( \{\lambda_j\}_{j=1}^{m+K} \) in its dual space such that \( \lambda_j(N_i^m) = \delta_{ij} \) where \( \delta_{ij} = 1 \) if \( i = j \) or 0 if \( i \neq j \). By the Hahn-Banach Theorem, the functional \( \lambda_j(.) \) can be extended to on \( L^p[y_j, y_{j+m}] \) for any \( 1 \leq p \leq \infty \) (\( p = \infty \) means that the norm is essential supreme norm). It is shown that there exists such a set of dual basis satisfying
\[ |\lambda_j(f)| \leq (2m + 1)g^{m-1}(y_{j+m} - y_j)^{-1/p} \|f\|_{L^p[y_j, y_{j+m}]}, \forall f \in L^p[y_j, y_{j+m}]. \]

After fixing the dual basis, we can introduce a linear operator \( Q \) on \( L^p[a, b] \) for \( 1 \leq p \leq \infty \) such that:
\[ Q(f) = \sum_{i=1}^{m+K} \lambda_i(f) N_i^m(x). \]
We know of one dimensional dual basis function: $N_{i,j}^{m_i}(x_i) = \prod_{k=1}^{d} N_{i_k}^{m_k}(x_k), 1 \leq i_k \leq m_k + K_k$.

$N_{i_1,i_2,...,i_d}(x_1,...,x_d)$ is the B-spline associated with the product of the knots which generate each $N_{i_k}^{m_k}(x_k)$. Similarly, this tensor-product space has a dual basis which is simply the composite product of one dimensional dual basis function:

$$\lambda_{i_1,i_2,...,i_d}(f(x_1,...,x_d)) = \lambda_{1,i_1}(...(\lambda_{d,i_d}(f(x_1,...,x_d)))...).$$
Define

\[ Q(f) = \sum_{i_1=1}^{m_1+K_1} \cdots \sum_{i_d=1}^{m_d+K_d} (\lambda_{i_1,i_2,\ldots,i_d}(f)) N_{i_1,i_2,\ldots,i_d}(x_1,\ldots,x_d), \]

then a property similar to Theorem A.2 (from Shumaker(1981) Theorem 12.7) exists for multivariate settings.

**Theorem A.3.** For \( f \in W^{r,p}(H) \) and \( r_j \leq m_j, j = 1, 2, \ldots, d, \)

\[ \| f - Q(f) \|_{L^p(H)} \leq C \sum_{i=1}^d \bar{\Delta}_{i_j}^r \| D_{x_i}^r f \|_{L^p(H)}, 1 \leq p \leq \infty, \]

where \( C \) is a constant depending only on \( (m_i)_{i=1}^d, (r_i)_{i=1}^d, d. \)

In particular, if we choose equally spaced partition and suppose \( k_i = k, m_i = m, M_i = \{ 1, 1, \ldots, 1 \} \) as well as \( r_i = r, \) then Theorem A.3 becomes

**Theorem A.4.** Without loss of generality, suppose \( a_i = 0, b_i = 1. \) With the equally spaced partition and the above selection of \( M, k_i, r_i, \forall 1 \leq p \leq \infty, \)

\[ |\lambda_{i_1,\ldots,i_d}(f)| \leq (2m + 1)^d 9^d (m-1)^{d/p} \| f \|_{L^p(H_{i_1,\ldots,i_d})}, \]

\[ \| f - Q(f) \|_{L^p(H)} \leq C \| f \|_{W^{r,p}/k^r}, \]

where \( H_{i_1,\ldots,i_d} \) is define as follows: suppose that \( y_1 = \ldots y_m = 0 < y_{m+1} < \ldots < y_{m+k} = 1 = y_{m+k+1} = \ldots = y_{i,k+2m} \) is an extended partition of \([0,1]\) then \( H_{i_1,\ldots,i_d} \) is the tensor-product of \([y_{ij}, y_{ij+1}].\)

All the proofs are given in Shumaker(1981).

**B Proof of Lemma 6.2**

Proof

All the following parameters are true parameters. For convenience, we omit the subscript “0” and simplify the notations of

\[ l(\psi_0, \alpha_0; \gamma^*), l_\psi(\psi_0, \alpha_0; \gamma^*), l_\alpha(\psi_0, \alpha_0; \gamma^*) \]

as \( l, l_\psi, l_\alpha. \) We will show that, under Assumptions (1-8), the function \( h(u, y, v) = (h_1(u, y, v), h_2(y)) \) of the "pseudo least favorable direction", which satisfies

\[ l_\psi^* l_\alpha [h] = l_\psi^* l_{\alpha}, a.s. \]

is in the Sobolev space \( W^{k,\infty}([a(\gamma^*), b(\gamma^*)] \times [0,1]^2) \times W^{k,\infty}[a(\gamma^*), b(\gamma^*)]. \) In fact, we will even show \( h(u, y, v) \) belongs to a Sobolev space with higher smoothness, \( W^{k+2,2}([a(\gamma^*), b(\gamma^*)] \times [0,1]^2) \times \)
\( W^{k+2.2}[a(\gamma^*), b(\gamma^*)] \) (without confusion, either of \( W^{k+2.2}([a(\gamma^*), b(\gamma^*)] \times [0,1]^2) \) and \( W^{k+2.2}([a(\gamma^*), b(\gamma^*)]) \) is simplified as \( W^{k+2.2} \) in the following context).

The proof consists of four steps.

**Step 1.** We show \( h(u, y, v) \) satisfies an integration equation and give the explicit form of the equation.

**Step 2.** By using the equation \( h(u, y, v) \) satisfies, we argue that \( h_2(y) \) belongs to \( W^{k+4.2} \).

**Step 3.** By transformation, we obtain another function \( \tilde{H}_1(u, y, v) \), which satisfies a second-order differential equation for fixed \((u, v)\) with boundary conditions. To prove \( h_1(u, y, v) \in W^{k+2.2} \) is equivalent to proving \( \frac{\partial^2}{\partial y^2} \tilde{H}_1(u, y, v) \in W^{k+2.2} \).

**Step 4.** \( \tilde{H}_1(u, y, v) \) can also be considered as a fixed point of an differential operator. So we show that the operator has a fixed point satisfying the condition of Step 3.

**Notation.** \( C^* \) in the following context denotes positive constant which may vary from place to place.

**Step 1.** First, recall that the equation

\[
\ell^*_{\psi}[h] = \ell^*_{\psi}l_{\alpha}, \text{a.s.}
\]

is derived based on a Hilbert space with \( L^2(P) \)-norm

\[
H = \{(h_1(u, y, v), h_2(y)) : h_1(u, y, v), h_2(y) \in L^2(P), \int_u h_1(u, y, v)du = 0, \int_y h_2(y)dy = 0\}.
\]

Second, \( \ell_{\psi} \) is written as the composition of two linear operators \( l_A(A_{\psi}) \) so the above equation can be re-expressed as \( A^*_{\psi}l_{\alpha}A_{\psi}[h] = A^*_{\psi}l_{\alpha} \). Notice that the log-likelihood \( l(\psi, \alpha; \gamma^*) \) can be written in two steps: let

\[
A(\psi, \alpha; \gamma^*) = e^{-e^{\psi(1,v)'A(y)e^{\psi(1,v)'A(y)f_{U(\gamma^*)}(\gamma^*(x,v)|y,v)}}}
\]

\[
l(A(\psi, \alpha; \gamma^*)) = r \ln A(\psi, \alpha; \gamma^*) + (1 - r) \int_y^\infty A(\psi, \alpha; \gamma^*)ds.
\]

Hence, \( \ell_{\psi}[h] = l_A(A_{\psi}[h]) \) where \( A_{\psi}[h] \) is the derivative of \( A(\psi, \alpha; \gamma^*) \) with respect to \( \psi \) along the direction \( h \), i.e,

\[
A_{\psi}[h] = A(\psi, \alpha; \gamma^*)(\frac{h_1(u, y, v)}{f_{U(\gamma^*)}(u, y, v)} + \frac{h_2(y)}{\lambda(y)} - e^{\psi(1,v)'h_2(y)} \int_0^y h_2(s)ds).
\]
Third, we obtain the kernel space of \( A_\psi^* \). If a function \( g(u, y, v) \) satisfies

\[
0 = < A_\psi^*[g], h^* >_{L^2(P)} = < g, A_\psi[h^*] >_{L^2(P)}, \forall h^* \in H,
\]

then

\[
\int_{u,y,v} f_{U(\gamma^*),Y,V}(u, y, v)g(u, y, v)[f_{T|V}(y|v)h_1^*(u|y, v) + f_{T|V}(y|v)f_{U(\gamma^*)}(u|y, v)(\frac{h_2^*(y)}{\lambda(y)}) - e^{\alpha(1,v)'} \int_0^y h_2^*(s)ds]dudydv = 0,
\]

where \( f_{U(\gamma^*),Y,V} \) is the joint density of \((U(\gamma^*), Y, V)\). So

\[
f_{U(\gamma^*),Y,V}(u, y, v)g(u, y, v)f_{T|V}(y|v) = \xi(y, v)
\]

for some function \( \xi(y, v) \) and moreover, \( \xi(y, v) \) satisfies that

\[
\int_v \xi(y, v)dv/\lambda(y) - \int_y^\infty \int_v \xi(y, v)e^{\alpha(1,v)'}dvdy = c_0
\]

for some constant \( c_0 \).

Therefore, from \( A_\psi^* l_A^* l_A A_\psi[h] = A_\psi^* l_A^* l_\alpha \), we obtain that there exist some function \( \xi(y, v) \) and a constant \( c_0 \) such that

\[
l_A^*[l_A A_\psi[h] - l_\alpha] = \xi(y, v)/(f_{U(\gamma^*),Y,V}(u, y, v)f_{T|V}(y, v))
\]

and

\[
\int_v \xi(y, v)dv/\lambda(y) - \int_y^\infty \int_v \xi(y, v)e^{\alpha(1,v)'}dvdy = c_0.
\]

Finally, we write out the explicit expression of \( l_A^* l_A A_\psi \). For any \( g(u, y, v) \in L^2([a(\gamma^*), b(\gamma^*)] \times [0, 1]) \),

\[
l_A[g] = r \frac{g(u, y, v)}{f_{T|V}(y|v)f_{U(\gamma^*)}(u|y, v)} + (1 - r) \frac{\int_y^\infty g(u, s, v)ds}{\int_y^\infty f_{T|V}(s|v)f_{U(\gamma^*)}(u|s, v)ds}, a.s.
\]

Then by using the equation

\[
< l_A[g], \tilde{h}(r, u, y, v) >_{L^2(P)} = < g, l_A^*(\tilde{h}(r, u, y, v)) >_{L^2(P)}
\]

for any \( \tilde{h}(r, u, y, v) \in L^2(P) \), we have

\[
l_A^*[\tilde{h}(r, u, y, v)] = \left\{ \frac{\tilde{h}(1, u, y, v)f_V(v)(1 - F_C(y|u, v))}{f_{Y,U(\gamma^*)}(y|u, v)} \right\}
\]

\[
+ \int_0^y f_V(v)f_C(s|u, v)\tilde{h}(0, u, s, v)ds \frac{1}{f_{Y,U(\gamma^*)}(y|u, v)}
\]

33
where \( f_C(y|u, v) \) is the conditional density of \( C \) given \((U(\gamma^*), V)\) and \( f_{Y|U(\gamma^*), V}(u, y, v) \) is the joint density of \((Y = T \land C, U(\gamma^*), V)\). Let \( f_{U(\gamma^*), Y|V}(u, y|v) \) be the density of \((U(\gamma^*), Y)\) given \( V = v \). Then

\[
L^*_\alpha L_A[g(u, y, v)] = \frac{1 - F_C(y|u, v)}{f_{T|V}(y|v)f_{U(\gamma^*)}(u|y, v)f_{U(\gamma^*), Y|V}(u, y|v)} + \int_0^y \int_s^\infty g(u, s', v)ds' \int_s^\infty f_{T|V}(s'|v)f_{U(\gamma^*)}(u|s', v)ds' f_C(s|u, v)ds \frac{1}{f_{U(\gamma^*), Y|V}(u, y|v)}.
\]

Substituting the expression

\[
g(u, y, v) = A_\psi[h] = f_{T|V}(y|v)h_1(u, y, v) + f_{T|V}(y|v)f_{U(\gamma^*)}(u|y, v)(h_2(y)/\lambda(y) - e^{\alpha(1, v)'} \int_0^y h_2(s)ds)
\]

into the above expression. So we obtain

\[
[h_1(u, y, v) + (h_2(y)/\lambda(y) - e^{\alpha(1, v)'} \int_0^y h_2(s)ds) f_{U(\gamma^*)}(u|y, v)] + s_1(u, y, v) \int_0^y s_2(u, y', v) \int_{y'}^\infty s_3(y'', v)[h_1(u, y'', v) + (h_2(y'')/\lambda(y'')) - e^{\alpha(1, v)'} \int_0^{y''} h_2(s)ds f_{U(\gamma^*)}(u|y'', v)] dy'' dy'
\]

\[
= s_4(u, y, v)\xi(y, v) + s_4(u, y, v)L^*_\alpha L_A,
\]

where

\[
s_1(u, y, v) = f_{U(\gamma^*)|Y|V}(u, y, v)/(1 - F_C(y|u, v)),
\]

\[
s_2(u, y, v) = f_C(y|u, v)/\int_y f_{T|V}(s|v)f_{U(\gamma^*)|T,V}(u|s, v)ds,
\]

\[
s_3(y, v) = f_{T|V}(y|v),
\]

\[
s_4(u, y, v) = f_{U(\gamma^*)|Y|V}(u, y, v)/[(1 - F_C(y|u, v))f_{T|V}(y|v)f_V(v)],
\]

and each of \( \{s_i(u, y, v), i = 1, 2, 3, 4\} \) is positive function in \( W^{k+2} \) by our assumptions.

**Step 2.** Integrate (4) over \( u \) and solve for \( c(y, v) \) then we obtain

\[
\xi(y, v) = (h_2(y) - B(h_1(u, y, v), h_2(y)))/\int_u s_4(u, y, v)dy,
\]

where \( B \) is a linear operator mapping \((h_1(u, y, v), h_2(y))\) to a function of \((y, v)\). Substitute the expression for \( \xi(y, v) \) into the equation

\[
\int_v \xi(y, v)dv/\lambda(y) - \int_y^\infty \int_v \xi(y, v)e^{\alpha(1, v)'}dvdv = c_0.
\]

34
Then the equation (4) becomes

\[ s_5(y)h_2(y) - Ing(h_2(y), h_1(u, y, v)) = c_0, \]

where \( Ing \) is another linear operator and \( s_5 \) is a positive function in \( W^{k+4.2} \) space. Moreover, since \( B \) and \( Ing \) are integral operators with respect to \( y \), we can easily show that if \((h_1(u, y, v), h_2(y))\) has \( j \)th derivatives with respect to \( y \) and their \( j \)th derivative is in \( L^2(P) \), then both \( B(h_1(u, y, v), h_2(y)) \) and \( Ing(h_2(y), h_1(u, y, v)) \) have \((j+1)\)th derivative with respect to \( y \) and their \((j+1)\)th derivative with respect to \( y \) is in \( L^2(P) \).

Hence, we can prove \( h_2(y) \in W^{k+4.2} \) by the following iterative arguments: for any \( 0 \leq j < k+4 \), if \((h_1(u, y, v), h_2(y))\) has \( j \)th derivatives with respect to \( y \) and their \( j \)th derivative is in \( L^2(P) \), then both \( B(h_1(u, y, v), h_2(y)) \) and \( Ing(h_2(y), h_1(u, y, v)) \) have \((j+1)\)th derivative with respect to \( y \) and such derivatives are in \( L^2(P) \). Thus, from the equation (6), \( h_2(y) \) has a \( L^2 \)-integrable \((j+1)\)th derivative with respect to \( y \); then (5) gives that \( \xi(y, v) \) has \((j+1)\)th derivative with respect to \( y \) and its derivative belongs to \( L^2(P) \); further from the equation (4), we obtain \( h_1(u, y, v) \) also has \((j+1)\) derivative with respect \( y \) and its derivative belongs to \( L^2(P) \). As a result, \((h_1(u, y, v), h_2(y))\) has \((j+1)\)th derivatives with respect to \( y \) and the \((j+1)\)th derivative is in \( L^2(P) \).

**Step 3.** We define

\[ H_1(u, y, v) = \int_0^u s_2(u, y', v) \int_{y'}^\infty s_3(y'', v)h_1(u, y'', v)dy''dy'. \]

Then the equation (4) becomes

\[
-[(H_1(u, y, v))'/y]/s_2(u, y, v)'/s_3(y, v) + s_1(u, y, v)H_1(u, y, v)]/s_4(u, y, v) \\
= c(y, v) + s_6(u, y, v),
\]

where \( s_6(u, y, v) \) only depends on the underlying densities and \( h_2(y) \) and it also belongs to \( W^{k+4.2} \). Moreover,

\[ c(y, v) = \int_u s_1(u, y, v)H_1(u, y, v)du/\int_u s_4(u, y, v)du. \]

\( H_1(u, y, v) \) also satisfies the boundary conditions

\[ H_1(u, 0, v) = 0, \frac{\partial}{\partial y}H_1(u, \tau, v) = 0. \]

In other words, \( H_1(u, y, v) \) solves a second-order differential equation

\[
-a(u, y, v)\frac{\partial^2}{\partial y^2}H_1(u, y, v) + b(u, y, v)\frac{\partial}{\partial y}H_1(u, y, v) + d(u, y, v)H_1(u, y, v) \\
= c(y, v) + e(u, y, v),
\]

\[ H_1(u, 0, v) = 0, \frac{\partial}{\partial y}H_1(u, \tau, v) = 0, \]
where the functions $a, b, d, e, q$ are known to be in $W^{k+4,2}$ and in addition, $a, d, q > 0$.

Clearly, if we can show that $\frac{\partial^2}{\partial y^2} H_1(u, y, v)$ has bounded $k$-th derivatives, then $h_1(u, y, v) = -((H_1(u, y, v))'_{y} / s_2(u, y, v))'_{y} / s_3(y, v)$ will belong to $W^{k+2,2}$.

We further define another function

$$\tilde{H}_1(u, y, v) = H_1(u, y, v) / [e^{\delta(y-\tau)^2} g(u, y, v)],$$

where $\delta$ is a small constant to be chosen later and $g(u, y, v)$ satisfies the following equation

$$-a(u, y, v) \frac{\partial^2}{\partial y^2} g(u, y, v) + b(u, y, v) \frac{\partial}{\partial y} g(u, y, v) + d(u, y, v) g(u, y, v) = 0, g(u, 0, v) = -1, \frac{\partial}{\partial y} g(u, \tau, v) = -1. \tag{8}$$

Since $a(u, y, v) > 0$, the elliptic equation theory says that the solution to the above equation exists and $g(u, y, v)$ has $L^2$-integrable $(k + 4)$th derivatives. In addition, since $d(u, y, v) > 0$, by the maximum principle (Section 6.4, Evans (1994)), we know that $g(u, y, v)$ cannot attain its non-negative maximum for some $y_0 \in (0, \tau)$ because otherwise, $g(u, y, v)$ must be a constant. So $g(u, y, v)$ can only attain its possible non-negative maximum at $\tau$; however, this gives $\frac{\partial}{\partial y} g(u, \tau, v) \geq 0$ and we obtain the contradiction again. Therefore, $g(u, y, v)$ must be strictly smaller than zero so $\frac{1}{g(u, y, v)}$ has $L^2$-integrable $(k + 4)$th derivatives.

We substitute $H_1(u, y, v) = e^{\delta(y-\tau)^2} g(u, y, v) \tilde{H}_1(u, y, v)$ into the equation (7). After simplification using the equation (8), we have

$$-\frac{\partial^2}{\partial y^2} \tilde{H}_1(u, y, v) + \tilde{b}(u, y, v) \frac{\partial}{\partial y} \tilde{H}_1(u, y, v) + \delta \tilde{d}(u, y, v) \tilde{H}_1(u, y, v) \tag{9}$$

$$= \tilde{s}(u, y, v) e(y, v) + \tilde{e}(u, y, v),$$

where

$$\tilde{b}(u, y, v) = \frac{-4\delta(y - \tau) g(u, y, v) a(u, y, v) - 2 \frac{\partial}{\partial y} g(u, y, v) a(u, y, v) + b(u, y, v) g(u, y, v)}{a(u, y, v) g(u, y, v)},$$

$$\tilde{d}(u, y, v) = \frac{-4\delta(y - \tau)^2 \frac{\partial}{\partial y} g(u, y, v) - 4\delta(y - \tau)^2 g(u, y, v) - 2 g(u, y, v)}{g(u, y, v)},$$

$$\tilde{s}(u, y, v) = \frac{e^{-\delta(y-\tau)^2}}{a(u, y, v) g(u, y, v)},$$

$$\tilde{e}(u, y, v) = \frac{e(u, y, v) e^{-\delta(y-\tau)^2}}{a(u, y, v) g(u, y, v)}.$$
Notice that $\bar{b}, \bar{d}, \bar{s}, \bar{e}$ all belong to $W^{k+4.2}$ and their $W^{k+4.2}$ norms are bounded by a constant independent of $\delta$. Obviously, in order to show that $h_1(u, y, v)$ is in $W^{k+2.2}$, it is sufficient to show $\frac{\partial^2}{\partial y^2} \bar{H}_1(u, y, v)$ is in $W^{k+2.2}$.

**Step 4.** Let $\mathcal{B}$ be the Banach space $\{\bar{H}_1(u, y, v) : \frac{\partial^2}{\partial y^2} \bar{H}_1(u, y, v) \in \mathcal{W}^{k+2.2}, \bar{H}_1(u, 0, v) = 0, \frac{\partial}{\partial y} \bar{H}_1(u, 0, v) = 0\}$ with norm $\|H_1(u, y, v)\|_B = \|\frac{\partial^2}{\partial y^2} \bar{H}_1(u, y, v)\|_{W^{k+2.2}} + \|\bar{H}_1(u, y, v)\|_{W^{k+2.2}}$. We consider the function $\bar{H}_1(u, y, v)$ as a fixed point of an operator $\mathcal{X}$ from $\mathcal{B}$ to $\mathcal{B}$. $\mathcal{X}$ is defined as follows: for a $w(u, y, v) \in \mathcal{B}$, $\mathcal{X}(w(u, y, v))$ is defined as a solution $\tilde{w}(u, y, v)$ to the following equation,

$$-\frac{\partial^2}{\partial y^2} \tilde{w}(u, y, v) + \bar{b}(u, y, v) \frac{\partial}{\partial y} \tilde{w}(u, y, v) = -\delta \tilde{d}(u, y, v)w(u, y, v) + \bar{s}(u, y, v)c(y, v) + \bar{e}(u, y, v)$$

with boundary conditions

$$\tilde{w}(u, 0, v) = 0, \quad g(u, \tau, v) \frac{\partial}{\partial y} \tilde{w}(u, \tau, v) + \tilde{w}(u, \tau, v) \frac{\partial}{\partial y} g(u, \tau, v) = 0.$$

Here,

$$c(y, v) = \frac{\int_u s_1(u, y, v) e^{\delta (y - \tau)^2} g(u, y, v) \tilde{w}(u, y, v) du}{\int_u s_1(u, y, v) du}.$$

Our eventual goal is to show $\bar{H}_1(u, y, v) \in \mathcal{B}$. This is true if we can prove

**(4A).** The operator $\mathcal{X}$ is well defined.

**(4B).** The operator $\mathcal{X}$ has a fixed point in $\mathcal{B}$.

**Proof of (4A).** $\mathcal{B}$ is well defined. From the above equation and the first two boundary conditions, we can easily solve for $\tilde{w}(u, y, v)$ in terms of $c(y, v)$:

$$\tilde{w}(u, y, v) = \int_0^y e^{\bar{B}(u,y',v)} \int_y^{\tau} e^{-\bar{B}(u,y'',v)} \left[ -\delta \tilde{d}(u, y'', v)w(u, y'', v) + \bar{s}(u, y'', v)c(y'', v) + \bar{e}(u, y'', v) \right] dy'' dy'$$

$$- \frac{\partial}{\partial y} g(u, \tau, v) \int_0^\tau e^{\bar{B}(u,y',v)} dy'$$

$$\times \left\{ \int_0^\tau e^{-\bar{B}(u,y'',v)} \left[ -\delta \tilde{d}(u, y'', v)w(u, y'', v) + \bar{s}(u, y'', v)c(y'', v) + \bar{e}(u, y'', v) \right] dy'' dy' \right\},$$

37
where \( \tilde{B}(u, y, v) = \int_0^y \tilde{b}(u, y', v)dy' \). Then by substituting this expression into the equality for \( c(y, v) \), we obtain an integral equation

\[
\begin{align*}
&c(y, v) + \int_0^y c(y', v)k_1(y', y, v)dy' + \int_y^\tau c(y', v)k_2(y', y, v)dy' \\
&= \mathcal{Q}(\delta \tilde{d}(u, y, v)w(u, y, v)),
\end{align*}
\]

where \( k_1, k_2 \) have bounded \((k + 2)\)-times derivatives with respect \((y', y, v)\), and \( \mathcal{Q} \) is a linear operator satisfying that

\[
\|\mathcal{Q}(\delta \tilde{d}(u, y, v)w(u, y, v))\|_B \leq C^*(\|w(u, y, v)\|_B + 1),
\]

for some constant \( C^* \). Therefore, that \( \mathcal{X} \) is well defined is equivalent to the above equation has a solution \( c(y, v) \).

If denote the above equation as \((I + K_v)[c(y, v)] = \mathcal{Q}(\delta \tilde{d}(u, y, v)w(u, y, v))\), then this is a Fredholm-type equation for any fixed \( v \) (the two integrations on the left hand side are compact operators from \( L^2[0, \tau] \) to itself). So it has a unique solution in \( L^2[0, \tau] \) if we can show that the corresponding homogeneous equation only has zero solution. To prove the latter, we reconstruct the original equation for \( H_1(u, y, v) \) but with minor difference. Suppose a \( \tilde{c}(y, v) \) satisfies

\[
\tilde{c}(y, v) + \int_0^y \tilde{c}(y', v)k_1(y', y, v)dy' + \int_y^\tau \tilde{c}(y', v)k_2(y', y, v)dy' = 0.
\]

We define

\[
\begin{align*}
\tilde{w}^*(u, y, v) &= \int_0^y e^{\tilde{B}(u, y', v)} \int_y^\tau e^{-\tilde{B}(u, y'', v)} \tilde{s}(u, y'', v) \tilde{c}(y'', v)dy''dy' \\
&\quad - \frac{\partial}{\partial y}g(u, \tau, v) \int_0^y e^{\tilde{B}(u, y', v)} dy' \\
&\quad - \frac{\partial}{\partial y}g(u, \tau, v) \int_0^\tau e^{B(u, \tau, v)} dy' \\
&\quad \times \{ \int_0^\tau e^{B(u, y', v)} \int_y^{\tau} e^{-B(u, y'', v)} \tilde{s}(u, y'', v) \tilde{c}(y'', v)dy''dy' \}.
\end{align*}
\]

Then \( \tilde{w}^*(u, y, v) \) satisfies

\[
\begin{align*}
-\frac{\partial^2}{\partial y^2} \tilde{w}^*(u, y, v) + \tilde{b}(u, y, v) \frac{\partial}{\partial y} \tilde{w}^*(u, y, v) &= \tilde{s}(u, y, v) \tilde{c}(y, v), \\
\tilde{w}^*(u, 0, v) &= 0, \quad g(u, \tau, v) \frac{\partial}{\partial y} \tilde{w}^*(u, \tau, v) + \tilde{w}^*(u, \tau, v) \frac{\partial}{\partial y} g(u, \tau, v) = 0, \\
\tilde{c}(y, v) &= \frac{\int_u s_1(u, y, v) e^{\delta(y-\tau)^2} g(u, y, v) \tilde{w}^*(u, y, v) du}{\int_u s_4(u, y, v) du}.
\end{align*}
\]

We further define \( H_1^*(u, y, v) = e^{\delta(y-\tau)^2} g(u, y, v) \tilde{w}^*(u, y, v) \). So it holds that

\[
-(H_1^*(u, y, v)'_y/s_2(u, y, v))'_y/s_3(y, v) + s_1(u, y, v)H_1^*(u, y, v)
\]

\[38\]
\[
\begin{align*}
= s_4(u, y, v) \int_u s_1(u, y, v) H_1^*(u, y, v) du / \int_u s_4(u, y, v) du + \delta \tilde{d}(u, y, v) s_4(u, y, v) H_1^*(u, y, v),
\end{align*}
\]

\[H_1^*(u, 0, y) = 0, \quad \frac{\partial}{\partial y} H_1^*(u, \tau, v) = 0.\]

Now let
\[
h_1^*(u, y, v)
= -(H_1^*(u, y, v)'/s_2(u, y, v))' / s_3(y, v) - \int \delta \tilde{d}(u, y, v) s_4(u, y, v) H_1^*(u, y, v) du.
\]

Thus, \(\int h_1^*(u, y, v) du = 0\) by integrating both sides of the equation over \(u\). Moreover,
\[
h_1^*(u, y, v) + s_1(u, y, v) \int_0^y s_2(u, y', v) \int_{y'}^\infty s_3(y'', v) h_1^*(u, y'', v) dy'' dy'
= s_4(u, y, v) \int_u s_1(u, y, v) H_1^*(u, y, v) du / \int_u s_4(u, y, v) du
+ \delta (d(u, y, v) s_4(u, y, v) H_1^*(u, y, v) - \int_u \tilde{d}(u, y, v) s_4(u, y, v) H_1^*(u, y, v) du)
- \delta s_1(u, y, v) \int_0^y s_2(u, y', v) \int_{y'}^\infty s_3(y'', v) \int_u \tilde{d}(u, y'', v) s_4(u, y'', v) H_1^*(u, y'', v) du dy'' dy'.
\]

Recall the derivation of \(\{s_i(u, y, v), i = 1, 2, 3, 4\}\) in the equation (4). We obtain
\[
< l_n A [f_{TV}(y \mid v)(H_1^*(u, y, v))], f_{TV}(y \mid v) h_1^*(u, y, v) >_{L^2(P)}
\leq \delta \|H_1^*(u, y, v)\|_{L^2(P)} \|h_1^*(u, y, v)\|_{L^2(P)}.
\]

On the other hand, since
\[
h_1^*(u, y, v) = -(H_1^*(u, y, v)'/s_2(u, y, v))' / s_3(y, v) - \int \delta \tilde{d}(u, y, v) s_4(u, y, v) H_1^*(u, y, v) du,
\]

it holds that
\[
\|H_1^*(u, y, v)\|_{L^2(P)} \leq C \|h_1^*(u, y, v)\|_{L^2(P)} + \delta C \|H_1^*(u, y, v)\|_{L^2(P)}.
\]

Moreover,
\[
< l_n A [f_{TV}(y \mid v)(H_1^*(u, y, v))], f_{TV}(y \mid v) h_1^*(u, y, v) >_{L^2(P)}
\geq C \|h_1^*(u, y, v)\|^2_{L^2(P)}.
\]

Combining the above two inequalities, we conclude that
\[
\|h_1^*(u, y, v)\|^2_{L^2(P)} \leq C \delta \|h_1^*(u, y, v)\|^2_{L^2(P)}.
\]

If we choose \(\delta\) small enough, \(h_1^*(u, y, v) = 0, a.s.\); that is,
\[
((H_1^*(u, y, v)' / s_2(u, y, v))' / s_3(y, v) = \delta \int \tilde{d}(u, y, v) s_4(u, y, v) H_1^*(u, y, v) du.
\]

39
Again, it deduces that \( \|H^*_1(u, y, v)\|_{L^2(P)}^2 \leq C \delta \|H^*_1(u, y, v)\|_{L^2(P)}^2 \). Therefore, \( H^*_1(u, y, v) = 0 \) when \( \delta \) is chosen small enough. Finally, \( \hat{c}(y, v) = 0 \). This implies there is a unique function \( \hat{w}(u, y, v) \) such that \( X[w(u, y, v)] = \hat{w}(u, y, v) \).

Moreover, it is clear that for each \( v \),
\[
\|\hat{c}(y, v)\|_{L^2[0, r]} \leq \|(I + K_v)^{-1}[Q(\delta \hat{d}(u, y, v)w(u, y, v))]\|_{L^2[0, r]}
\]
so the \( L^2 \)-norm of \( \hat{c}(y, v) \) as a function of \( (y, v) \) is bounded by a constant from the continuity of \( (I + K_v)^{-1} \) in \( v \in [0, 1] \). Additionally,
\[
\|\hat{c}(y, v)\|_{L^2} \leq \hat{C}_0 + C^*_0 \|w(u, y, v)\|_{L^2}.
\]

We show the solution \( \hat{c}(y, v) \) belongs to \( W^{k+2,2} \) by induction on \( m \leq k + 2 \). When \( m = 0 \), this has already been verified in the above. Suppose \( \hat{c}(y, v) \in W^{m-1,2} \). We differentiate the equation (12) \( m \) times in the generalized function space (i.e., the \( m \)-th derivative for a generalized function \( g(x) \) is defined as a \( L^1 \) function \( g_m(x) \) such that \( \int g(x)D^m\phi(x)dx = (-1)^m \int g_m(x)\phi(x)dx \) for any \( \phi(x) \in C_0^\infty \)). Therefore, we obtain that the generalized derivative \( D^m \hat{c}(y, v) \) satisfies the equation of \( f \)
\[
(I + K_v)[f] = D^mQ(\delta \hat{d}(u, y, v)w(u, y, v)) + \sum_{||\leq m-1} G(D^l\hat{c}(y, v)),
\]
where \( D^l \) is any \( l \)-th derivative with respect to \( y \) and \( v \) and \( G \) is a smooth functional of \( D^l\hat{c}(y, v) \). However, by similar arguments as the case \( m = 0 \), we know that the unique solution to this equation, which is equal to \( D^m \hat{c}(y, v) \) exists and belongs to \( L^2 \). Furthermore,
\[
\|D^m \hat{c}(y, v)\|_{L^2} \leq \hat{C}_m + C^*_m(\|\delta w(u, y, v)\|_{W^{k+2,2}} + \sum_{l \leq m-1} \|D^l \hat{c}(y, v)\|_{L^2})
\]
for some constants \( \hat{C}_m \) and \( C^*_m \). Hence, \( \hat{c}(y, v) \) belongs to \( W^{k+2,2} \). We sum the inequalities of (13) for \( m = 0, ..., k \). The
\[
\|\hat{c}(y, v)\|_{W^{k+2,2}} \leq C^* \|w(u, y, v)\|_{W^{k+2,2}} + C^* \|\hat{c}(y, v)\|_{W^{k-1,2}}.
\]
The interpolation theorem in the Sobolev space gives
\[
\|\hat{c}(y, v)\|_{W^{k-1,2}} \leq \epsilon \|\hat{c}(y, v)\|_{W^{k+2,2}} + C^*_\epsilon \|\hat{c}(y, v)\|_{L^2}
\]
\[
\leq \epsilon \|\hat{c}(y, v)\|_{W^{k+2,2}} + C^*_\epsilon (\hat{C}_0 + \delta C^*_0 \|w(u, y, v)\|_{L^2})
\]
for any small \( \epsilon \). Therefore, we finally obtain
\[
\|\hat{c}(y, v)\|_{W^{k+2,2}} \leq C^*(\delta \|w(u, y, v)\|_{W^{k+2,2}} + 1)
\]
for some constant $C^*$. With similar arguments by differentiating (12) $(k + 4)$ times with respect to $y$, it is easy to see

$$\|\tilde{c}(y, v)\|_B \leq C^*(\delta \|w(u, y, v)\|_B + 1).$$

Finally, the solution to the equation (10), which is given by (11), belongs to $B$, i.e., $\tilde{w}(u, y, v) \in B$. We have verified $\mathcal{X}$ is an linear operator from $B$ to $B$.

**proof of (4B).** Using the equation (11), we also obtain that

$$\|\mathcal{X}(w(u, y, v))\|_B = \|\tilde{w}(u, y, v)\|_B \leq C_5(1 + \|\tilde{d}(u, y, v)w(u, y, v)\|_B + \|\tilde{s}(u, y, v)\tilde{c}(y, v)\|_B) \leq C_6(1 + \delta \|w(u, y, v)\|_B),$$

for a constant $C_6$ and

$$\|\mathcal{X}(w_1(u, y, v) - w_2(u, y, v))\|_B \leq C_0\delta \|w_1(u, y, v) - w_2(u, y, v)\|_B,$$

So $\mathcal{X}$ a continuous operator from $B$ to $B$. Additionally, it maps a closed convex set $\{w(u, y, v) : \|w(u, y, v)\|_B \leq C_7\}$ to itself for a large $C_7$ (say, let $C_6(1 + \delta C_7) \leq C_7$). By the Schauder’s fixed point theorem, there exists a function $w(u, y, v)$ in $B$ satisfying $\mathcal{X}(w(u, y, v)) = w(u, y, v)$. Obviously, such $w(u, y, v)$ is the same as $\tilde{H}_1(u, y, v)$ due to the uniqueness of the solution.

At last, we have shown $\tilde{H}_1(u, y, v)$ is in $B$. That is, $\frac{\partial^2}{\partial y^2} \tilde{H}_1(u, y, v)$ thus $h_1(u, y, v)$ belongs to $W^{k+2,2}$. Lemma 6.2 holds. $\square$
REFERENCES


