2001

This is to certify that the

dissertation entitled

Semiparametric Estimation For Current Status Data
With Flexible Covariate Effects

presented by

Wenliang Lu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Major professor
Hira L. Koul

Date December 4, 2000

SEMIPARAMETRIC ESTIMATION FOR CURRENT STATUS DATA WITH FLEXIBLE COVARIATE EFFECTS

By

Wenliang Lu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2000

ABSTRACT

SEMIPARAMETRIC ESTIMATION FOR CURRENT STATUS DATA WITH FLEXIBLE COVARIATE EFFECTS

By

Wenliang Lu

This thesis studies a semiparametric hazard model with parametric baseline hazard rate and nonparametric covariate dependency based on current status data. Two estimators are proposed. One is the generalized profile maximum likelihood estimator (GPMLE) and the other is the sieve maximum likelihood estimator (SMLE). The GPMLE is obtained by maximizing the profile likelihood function where the nonparametric covariate part is estimated using kernel and least square methods. Under some regular conditions, the thesis establishes the square root consistency and asymptotic normality of this estimator. The SMLE of the parameter is obtained by maximizing the log-likelihood function with respect to both the finite dimensional and the infinite dimensional nuisance parameters while the infinite dimensional nuisance parameter is constrained to a subset of the parameter space which increases with the increase in the sample size. This estimator is shown to be consistent and asymptotic normal. Moreover, its asymptotic variance achieves the semiparametric lower bound.

ACKNOWLEDGMENTS

I would like to thank my advisor Professor Hira L. Koul for his guidance and many helpful discussions on the subject of this thesis. He was always available when I had doubts or questions. His general thinking of a statistical problem and ways to solve the problem will help my future research and working. I would also like to thank all the other committee members, Professor Joseph Gardiner, James Hannan and Habib Salehi, for serving my guidance committee, especially Professor James Hannan who was also my academic advisor. Finally I would like to thank the department of Statistics and Probability for offering me graduate assistantships so that I could come to the states and finish my thesis and my graduate study at Michigan State University.

TABLE OF CONTENTS

LIST OF TABLES

Introduction

0.1 Overview

Current status data arise in some clinical setting when the survival time of interest can only be determined to lie below or above a random examination time. In the settings such as destructive testing, animal experiments in which the occurrence of a survival time is only observable upon sacrifice, and epidemiologic studies in which obtaining more than one examination is not cost effective, current status data are commonly encountered.

The nonparametric estimation of the survival time distribution and some smooth functionals thereof have been discussed for current status data by a number of authors, including Groeneboom and Wellner (1992, §2.3), Huang and Wellner (1995), Geskus and Groeneboom (1996) and Geskus and Groeneboom (1997).

Semiparametric models based on current status data have also been studied in the literature. Klein and Spady (1993), Rabinowitz, Tsiatis and Aragon (1995), Li and Zhang (1998), and Murphy, Van der Vaart and Wellner (1999) considered the linear regression model based on current status data. Klein and Spady used the profile maximum likelihood method to derive the estimator of the regression parameters which were shown to achieve the semiparametric lower bound. In Rabinowitz, Tsiatis and Aragon's paper, a class of score statistics that may be used for estimation and confidence procedures is proposed. Li and Zhang minimized a class of U-statistics of order 3 to obtain estimators of the parameters. Murphy, Van der Vaart and Wellner considered the penalized maximized likelihood estimator of the regression parameter which was shown to be efficient. Koul and Schick (1999) studied the estimation and hypothesis testing of the ratio of scale parameters in the two-sample setting, using a

U-statistic of order 2.

Cox's regression model has been also studied based on the current status data. Finkelstein (1986), Diamond and McDonald (1991), and Shiboski and Jewell (1992) developed several methods to fit the model. Huang (1996) showed that, profiled over the cumulative baseline hazard function, the profile maximum likelihood estimator for the regression parameter is asymptotically normal with $n^{\frac{1}{2}}$-convergence rate.

Among the other semiparametric models for the current status data, additive hazards regression model was studied by Lin, Oaks and Ying (1998) and the proportional odds regression model was studied by Rossini and Tsiatis (1996). Under certain conditions on the examination time, Lin, Oaks and Ying found that one can make inferences about the regression parameters of the additive hazards model by using the familiar asymptotic theory and software for the proportional hazards model with right censoring data. Rossini and Tsiatis's approach in the proportional odds regression model is based on approximating the infinite-dimensional nuisance parameter, the baseline log-odds of failure, with a step function, and carrying out a maximum likelihood procedure. The resulting finite dimensional parameter estimates for the regression parameters are shown to be asymptotically normal and semiparametrically efficient.

Although these models, especially the Cox's regression model, are popular and widely used in practice, in many applications the shape of the baseline hazard is thought to be well understood but the covariate effect is rarely specified precisely. For example, in insurance problems the Gompertz-Makeham hazard has a long tradition of successful application, [Jordan (1975), page 21]. Meshalkin and Kagan (1972) claimed that the logarithm of the baseline hazard is approximately linear for a number of chronic diseases. As an alternative to Cox's regression model, Nielsen, Linton and Bickel (1998) studied a model where the baseline hazard rate belongs to a parametric class of hazard functions but the covariate part is of unknown functional form. They obtained an estimator of the the underlying parameter by profile maximum likelihood method when the data is randomly right censored.

This dissertation discusses the estimation of the underlying parameter in this model (Nielsen, Linton and Bickel, 1998) for current status data. Two estimators are proposed. The first one is obtained by maximizing a profile likelihood where the infinite dimensional nuisance parameter is estimated nonparametrically. This is called the *generalized profile maximum likelihood estimator.* A set of sufficient conditions are provided for consistency and asymptotic normality.

The second estimator, called *sieve maximum likelihood estimator*, is obtained by maximizing the log-likelihood function with respect to both the finite dimensional and the infinite dimensional nuisance parameters while the infinite dimensional nuisance parameter is constrained to a subset of the parameter space which increases with the increase in the sample size. It is shown to be consistent, asymptotically normal, with its asymptotic variance achieving the semiparametric lower bound.

Simulations are conducted to study the behavior of these estimators for small and moderate sample sizes. The generalized profile maximum likelihood estimator seems to have a slightly lower bias and variance than the sieve maximum likelihood estimator. Since the latter achieves the lower bound, as the sample size increase, it should behave better than the generalized profile maximum likelihood estimator for large samples.

0.2 The model

Let X, T, Z be a random vector, where X represents the survival time, T the monitoring variable and Z the covariate which could be a vector. Let (X_1, T_1, Z_1), $\cdots$, (X_n, T_n, Z_n) be i.i.d copies of X, T, Z.

Assume that, conditioned on Z, X and T are conditionally independent. The conditional distribution of X, given Z, is assumed to depend on some parameter and the covariate. In Cox's regression model, the cumulative hazard rate function of X given Z has the form

$$\Lambda_0(x)e^{\beta' Z},$$

where the first part Λ_0, with unspecified form, is called the baseline cumulative hazard function, and β is a vector of parameters. Nielsen, Linton, and Bickel (1998) proposed an alternative model with the first part depending only on some parameter θ_0 and the second part with unspecified form. More specifically, the cumulative hazard rate function is of the form

$$\Lambda(x, \theta_0)g(Z),$$

where $\Lambda(x, \theta_0)$ is a known function with unknown parameter θ_0, but g is an unknown function. Here θ_0 belongs to Θ, a subset of $\mathcal{R}^d$ for some $d \geq 1$. They discussed the estimation of θ_0 and g under right censoring.

In this dissertation we discuss the estimation of θ_0 and $g(z)$ based on current status data or interval censoring Case I data, where one observes $(T_i, \delta_i, Z_i), i = 1, 2, \dots, n$, with $\delta_i = I_{(X_i \leq T_i)}$. It is assumed in the following sections that θ_0 is a scalar. For θ_0 as a vector, similar results can be obtained. Because of the curse of dimensionality, Z is assumed to be a scalar also.

Let $F(x, Z, \theta_0)$ be the conditional distribution function of X, given Z. Assume that the cumulative hazard rate function is continuous. Then

$$F(x, Z, \theta_0) = 1 - \exp(-\Lambda(x, \theta_0)g(Z)).$$

We also assume that the distribution of (T, Z) does not depend on θ_0 or g, and that if $\Lambda(t, z, \theta_1)g_1(z) = \Lambda(t, z, \theta_0)g(z)$ for all (t, z) in the support of (T, Z), then $\theta_1 = \theta_0$ and $g_1(z) = g(z)$ for all z. The latter is the identifiability condition.

Chapter 1

Generalized Profile Maximum Likelihood Estimation

1.1 Definition of estimators of θ_0 and g

In this dissertation we first use a semiparametric profile likelihood method to define the estimator of the parameter. Both Klein and Spady (1993) and Nielsen, Linton, and Bickel (1998) used generalized profile likelihood methods to estimate the finite dimensional parameter while the infinitely dimensional nuisance parameter was estimated by the kernel method. The ensuing discussion in this section will be a bit informal. The precise conditions under which all definitions are valid are stated in the next section.

In this chapter, Θ is assumed to be a compact subset of $\mathcal{R}^1$, and is rewritten as $\mathcal{N}_0$.

One notes that, given $(T_i, Z_i), i = 1, 2, \ldots, n$, the (conditional) log-likelihood for θ and g based on $(T_i, \delta_i, Z_i), i = 1, 2, \ldots, n$ is

$$\sum_{i=1}^{n}[\delta_i log(1 - exp(-\Lambda(T_i, \theta)g(Z_i))) - (1 - \delta_i)\Lambda(T_i, \theta)g(Z_i)].$$

The idea of generalized profile likelihood methods is as follows:

(1) For a fixed θ, obtain the estimates, $\hat{g}_\theta(Z_i)$, of $g(Z_i)$, $i = 1, \cdots, n$, by using some

method such as the kernel method.

(2) The generalized profile likelihood for θ arising when $g(Z_i)$ is replaced by $\hat{g}_\theta(Z_i)$ is

$$\sum_{i=1}^{n}[\delta_i log(1-exp(-\Lambda(T_i,\theta)\hat{g}_\theta(Z_i)))-(1-\delta_i)\Lambda(T_i,\theta)\hat{g}_\theta(Z_i)].$$

Maximize it with respect to θ to obtain the estimate $\hat{\theta}$ of θ.

(3) If we want to estimate $g(z)$, we treat $\hat{\theta}$ as the real parameter and use some method as in step (1) or some other method to estimate it.

When $\theta=\theta_0$, $\hat{g}_{\theta_0}(Z_i)$ should approach $g(Z_i)$ for all fixed Z_i as the sample size n tends to infinite. Moreover, the convergence must be faster than some particular rate. This is hard to achieve for all Z_i, $i=1,\cdots,n$, because of the edge effects in the kernel estimation. Hence we use the following modified likelihood for θ and g:

$$l_n(\theta,g)=\sum_{i=1}^{n}w_1(T_i,b)w_2(Z_i,b)[\delta_i log(1-exp(-\Lambda(T_i,\theta)g(Z_i)))-(1-\delta_i)\Lambda(T_i,\theta)g(Z_i)]$$

where $w_2(Z_i,b)=1$ if Z_i is at least b far away from the boundary and 0 otherwise, $w_1(T_j,b)=1$ if T_j is at least b far away from the boundary and 0 otherwise. More precisely, for example, if the support of Z is an interval $[z_1^*,z_2^*]$, then $w_2(Z_i,b)=1$ if Z_i is in the interval $[z_1^*+b,z_2^*-b]$ and 0 otherwise, where b depends on n and $b\to 0$ as $n\to\infty$. Therefore, the modified likelihood is almost the same as the real likelihood for n large enough.

In this dissertation, the support of a random variable (or possibly a random vector) with a density with respect to Lebesgue measure means the closure of the set of all points at which the density is positive.

To estimate g for any fixed θ, our approach uses two dimensional kernel method to estimate

$$F(T_j,Z_i,\theta_0),\quad i,j=1,\cdots,n,$$

and then combines these estimates for each fixed i to obtain $\hat{g}_\theta(Z_i)$, $i=1,\cdots,n$. The least square method is used in the latter step.

Let K be a kernel and b the bandwidth. Define

$$\hat{F}(T_j, Z_i) = \frac{\sum_{l\neq j,i} \delta_l K_b(T_l - T_j) K_b(Z_l - Z_i)}{\sum_{l\neq j,i} K_b(T_l - T_j) K_b(Z_l - Z_i)}, \quad 1 \leq i, j \leq n, \tag{1.1.1}$$

where

$$K_b(t) = \frac{1}{b} K(\frac{t}{b}).$$

Note that $\hat{F}(\cdot, \cdot)$ depends on j, i, but we don't make it explicit until it is necessary.

Under certain conditions on K, F and the density of (T, Z), and if $b \to 0$ and $nb^2 \to \infty$, then, conditioned on T_j, Z_i, in probability,

$$\begin{aligned}
\hat{F}(T_j, Z_i) \to & \lim_{b\to 0} \frac{E_{j,i}[\delta K_b(T - T_j) K_b(Z - Z_i)]}{E_{j,i}[K_b(T - T_j) K_b(Z - Z_i)]} \\
= & \frac{F(T_j, Z_i, \theta_0) h(T_j, Z_i)}{h(T_j, Z_i)} \\
= & F(T_j, Z_i, \theta_0),
\end{aligned}$$

where $h(t, z)$ is the joint density of (T, Z) and $E_{j,i}$ denote conditional expectation, given T_j, Z_i. Therefore $\hat{F}(T_j, Z_i)$ can be used to estimate $F(T_j, Z_i, \theta_0)$.

Now if θ is the real parameter, then $-log(1 - F(T_j, Z_i, \theta)) = \Lambda(T_j, \theta) g(Z_i)$ and $-log(1 - \hat{F}(T_j, Z_i))$ should be close to $-log(1 - F(T_j, Z_i, \theta))$ for all j and i when the sample size is large enough. For fixed Z_i, we shall estimate $g(Z_i)$ such that $\Lambda(T_j, \theta) g(Z_i)$ is close to $-log(1 - \hat{F}(T_j, Z_i))$, $j = 1, \cdots, n$.

Let

$$\hat{g}_\theta(Z_i) = -\frac{\sum_{j\neq i} w_1(T_j, b) \Lambda(T_j, \theta) log(1 - \hat{F}(T_j, Z_i))}{\sum_{j\neq i} w_1(T_j, b) \Lambda^2(T_j, \theta)}, \tag{1.1.2}$$

a least square estimator of $g(Z_i)$, attaining

$$\min_{g(Z_i)} \sum_{j\neq i} w_1(T_j, b)[log(1 - \hat{F}(T_j, Z_i)) + \Lambda(T_j, \theta) g(Z_i)]^2.$$

The counterpart of $\hat{g}_\theta(z)$ in limit is

$$g_\theta(z) = -\frac{E\left[\Lambda(T, \theta) log(1 - F(T, Z, \theta_0))\right]}{E\Lambda^2(T, \theta)} = \frac{E[\Lambda(T, \theta)\Lambda(T, \theta_0)]}{E\Lambda^2(T, \theta)} g(z), \tag{1.1.3}$$

where E means the expectation w.r.t. the real parameter θ_0 and g. Note that, by (1.1.3), $g_{\theta_0} = g$. Let

$$F(t,z,\theta) = 1 - e^{-\Lambda(t,\theta)g_\theta(z)}, \quad \bar{F} = 1 - F \tag{1.1.4}$$

and

$$\hat{F}(t,z,\theta) = 1 - e^{-\Lambda(t,\theta)\hat{g}_\theta(z)}. \tag{1.1.5}$$

The modified profile log-likelihood that arises when g is replaced by $\hat{g}_\theta$ is

$$l_{n1}(\theta)$$
$$= \sum_{i=1}^{n} w_1(T_i,b)w_2(Z_i,b)[\delta_i log(1 - exp(-\Lambda(T_i,\theta)\hat{g}_\theta(Z_i))) - (1-\delta_i)\Lambda(T_i,\theta)\hat{g}_\theta(Z_i)]. \tag{1.1.6}$$

The estimator, $\hat{\theta}$, of θ_0 is the maximizer of the above likelihood over $\theta \in \mathcal{N}_0$.

Finally, the estimator of $g(z)$ is defined as

$$\tilde{g}(z) = -\frac{\sum_{i=1}^{n} w_1(T_j,b)\Lambda(T_i,\hat{\theta})log(1-\hat{F}(T_i,z))}{\sum_{i=1}^{n} w_1(T_j,b)\Lambda^2(T_i,\hat{\theta})}.$$

1.2 Asymptotic properties of the estimators

1.2.1 Consistency

In this section, we state the consistency of the generalized profile maximum likelihood estimator $\hat{\theta}$. Before doing this, we give various assumptions which will be used to prove the consistency and asymptotic normality of $\hat{\theta}$.

We list the following assumptions.

(A1) The respective supports $\mathcal{Z}$ and $\mathcal{T}$ of Z and T are closed intervals of $\mathcal{R}^1$. $\Lambda(t,\theta)$, $g(z)$ and $h(t,z)$ are positive and continuous on their domains of definition $\mathcal{T} \times \mathcal{N}_0$, $\mathcal{Z}$ and $\mathcal{T} \times \mathcal{Z}$. Moreover $\Lambda(t,\theta)$ is continuous in θ uniformly for t. The first and second derivatives of $\Lambda(t,\theta)$ w.r.t. θ, $\dot\Lambda(t,\theta)$ and $\ddot\Lambda(t,\theta)$, exist, and $\dot\Lambda(t,\theta)$, $\ddot\Lambda(t,\theta)$ are continuous in θ uniformly for t, and continuous in t for any fixed θ.

(A2) The function $g(z)$ and $h(t,z)$ are four times differentiable on their domains of definition with continuous 4th (partial) derivatives. Assume $\Lambda(t,\theta_0)$ is four times differentiable in t with continuous 4th (partial) derivatives.

(A3) The kernel function K is an r-th order kernel supported on $[-1,1]$, symmetric about zero and Lipschitz continuous on its support. (r-th order kernel means K satisfies: $\int K(t)dt = 1$, $\int t^s K(t)dt = 0$ for $s = 2,\cdots,r-1$ and $\int |t|^r|K(t)|dt < \infty$.)

(A3′) The kernel function K is Lipschitz continuous, supported on $[-1,1]$, and satisfies: $\int K(t)dt = 1$.

(A4) $b = O(n^{-a})$ with $\frac{1}{8} < a < \frac{1}{4}$.

(A5) θ_0 is an interior point of $\mathcal{N}_0$, which is a compact subset of $\mathcal{R}^1$.

(A6)

$$E\left[(\dot\Lambda(T,\theta_0)g(Z) + \Lambda(T,\theta_0)\dot g_{\theta_0}(Z))^2\frac{\bar F(T,Z,\theta_0)}{F(T,Z,\theta_0)}\right] > 0,$$

where $\dot g_\theta(z)$ is the (partial) derivative of $g_\theta(z)$ with respect to θ.

Assumption (A1) or similar assumptions have been seen in the literature, see, for example, Huang (1996), Klein and Spady (1993), Nielsen, Linton and Bickel (1998). Assumption (A2) is a smooth condition on the model, which is used mainly for the asymptotic normality. Assumptions (A3) and (A3′) are made for the kernel. One notes that (A3) implies (A3′). Assumption (A4) is the bandwidth condition in kernel estimation, which is crucial to the asymptotic normality. For the consistency of the estimator, this bandwidth condition can be weakened. Assumptions (A1), (A3′) and (A5) are imposed for consistency of the estimator. To prove the asymptotic normality, we use assumptions (A1) -(A6).

Next we state the theorem on the consistency of the estimator. The proof will be given in Section 1.4.2 following the general preliminary Lemmas 1-5 on kernel estimations in Section 1.4.1. Before the proof of the theorem in Section 1.4.2, we give first Lemmas 6-9 on the uniform consistency of $\hat{F}(T_j, Z_i)$ for $F(T_j, Z_i)$ and of $\hat{g}_\theta(Z_i)$ and its derivatives for $g_\theta(Z_i)$ and its derivatives, $1 \leq i, j \leq n$.

Theorem 1 *Suppose that (A1), (A3′) and (A5) hold, $b = O(n^{-a})$ with $0 < a < \frac{1}{4}$. Then the generalized profile likelihood estimator, $\hat{\theta}$, which is obtained by maximizing $l_{n1}(\theta)$, converges in probability to the real parameter θ_0.*

1.2.2 Asymptotic normality

In this section, we state the theorem on the asymptotic distribution of the estimator and the proof will be given after the proof of Theorem 1 in Section 1.4.2.

Theorem 2 *(Asymptotic distribution of $\hat{\theta}$) Suppose (A1)-(A6) hold with $r = 4$ for (A3). Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow \mathcal{N}(0, \sigma^2),$$

where

$$\sigma^2 = \frac{E\{[D_1(T, Z, \theta_0) - \Delta(T, Z, \theta_0)]^2 R(T, Z, \theta_0)\}}{[E(D_1^2(T, Z, \theta_0) R(T, Z, \theta_0))]^2},$$

$$\Delta(T, Z, \theta_0) = \frac{\Lambda(T, \theta_0) h_1(T)}{c_0 h(T, Z) R(T, Z, \theta_0)} \int \Lambda(t, \theta_0) D_1(t, Z, \theta_0) R(t, Z, \theta_0) h(t, Z) dt,$$

$$c_0 = E\Lambda^2(T, \theta_0), \quad D_1(t, z, \theta_0) = \dot{\Lambda}(t, \theta_0) g(z) + \Lambda(t, \theta_0) \dot{g}_{\theta_0}(z),$$

and

$$h_1(t) = \int_Z h(t, z) dz, \quad R(t, z, \theta_0) = \frac{\bar{F}(t, z, \theta_0)}{F(t, z, \theta_0)}.$$

1.3 Simulation

Before we prove the stated asymptotic properties of the estimator, let's take a look at its behavior for small and moderate samples.

Assume that the conditional distribution of X, given Z, is a Weibull distribution with distribution function

$$1 - e^{-x^{\theta_0} g(Z)}.$$

where $g(z) = z$. Also assume that T and Z are uniformly distributed on $[1, 2]$ and $[0.2, 1.2]$ respectively.

For each fixed sample size (n=30, 60, 100, 200 respectively) and appropriate b's, 100 samples are generated with the real parameter $\theta_0 = 1.5$ and 100 replications of the estimate of θ_0 based on the generalized profile maximum likelihood estimator (GPMLE) are obtained. The means and standard deviations are shown in the following table.

Table 1. Simulation results for GPMLE

n	b	mean	s.d.
30	0.0400	1.3847	1.3299
	0.0420	1.4915	1.4075
	0.0450	1.7596	1.3905
60	0.0308	1.4720	0.9801
	0.0310	1.4824	0.9947
	0.0312	1.4908	1.0043
100	0.0238	1.4535	0.7720
	0.0240	1.4876	0.7702
	0.0242	1.5075	0.7943
200	0.0166	1.4560	0.4795
	0.0168	1.4990	0.4902
	0.0170	1.5421	0.5103

The kernel function used in the simulation is $K(x) = 9/8 - 15/8x^2, -1 \leq x \leq 1; 0,$ otherwise.

From the table we can see that the mean is around the true value for all the sample sizes but the standard deviation decreases with the increase in the sample size. The choice of b is crucial to the reduction of the bias of the estimator.

1.4 Proof of the consistency and asymptotic normality

1.4.1 Lemmas preliminary to the proof

To prove the consistency and asymptotic normality of the generalized profile maximum likelihood estimator, the uniform consistency of $\hat{F}(T_j, Z_i)$ for $F(T_j, Z_i)$ over all $1 \leq i, j \leq n$, and of $\hat{g}_\theta(Z_i)$ for $g_\theta(Z_i)$ over all $1 \leq i \leq n$ and $\theta \in \mathcal{N}_0$ is proved first. Since $\hat{g}_\theta(Z_i)$ is a function of $\hat{F}(T_j, Z_i)$ which, in view of (1.1.1), is a ratio of two sums (or means) of independent random variables, we first discuss some uniform convergence results of the sums (or means) of independent random variables in a general setting.

Lemma 1 *Let $Y_1, \cdots, Y_n$ be i.i.d. d-dimensional random vectors. Let $\mathcal{D}$ be a compact subset of $\mathcal{R}^d$, and for each $t \in \mathcal{D}$, let $W_n(t, \cdot), n \geq 1$, be a sequence of measurable functions on $\mathcal{R}^d$. Let*

$$\xi_n(t) = \frac{1}{n}\sum_{i=1}^{n} W_n(t, Y_i), \quad t \in \mathcal{D}. \tag{1.4.1}$$

Let $0 < h_n = O(n^{-a_0})$ with $a_0 > 0$ and assume that for some $0 \leq s, r < \infty$, and finite real number C_0,

$$h_n^r |W_n(t, y)| \leq C_0, \quad h_n^s |W_n(t_1, y) - W_n(t_2, y)| \leq C_0 \sum_{j=1}^{d} |t_{1j} - t_{2j}|, \tag{1.4.2}$$

uniformly for $y \in \mathcal{R}^d$ and for all t, t_1, t_2 in $\mathcal{D}$. Assume also that

$$E(W_n(t, Y_i)) = 0, \quad t \in \mathcal{D}.$$

Then, for all $a > 0$,

$$\sup_{t\in\mathcal{D}} |\xi_n(t)| = o_p(n^{-\frac{1-a}{2}} h_n^{-r}). \tag{1.4.3}$$

Proof. Let

$$\Delta_n = \frac{\varepsilon_n h_n^s}{2C_0 d},$$

where $0 < \varepsilon_n \to 0$, to be chosen later. By (1.4.1) and the second part of (1.4.2), for all $t_1, t_2 \in \mathcal{D}$ with $t_i = (t_{i1}, \cdots, t_{id})$, $i = 1, 2$,

$$|\xi_n(t_1) - \xi_n(t_2)| \leq C_0 h_n^{-s} \sum_{j=1}^{d} |t_{1j} - t_{2j}|.$$

If $|t_{1j} - t_{2j}| < \Delta_n$, then this inequality and the definition of Δ_n lead to

$$|\xi_n(t_1) - \xi_n(t_2)| \leq C_0 h_n^{-s} d\Delta_n = \frac{\varepsilon_n}{2}.$$

Since $\mathcal{D}$ is a compact space, it is contained in a hypercube. Without loss of generality, let it be contained in a unit cube. Let $N_n = 1/\Delta_n$ if $1/\Delta_n$ is an integer, and $([1/\Delta_n]+1)^d$ otherwise, where $[x]$ means the integer part of x. Divide the unit cube into small cubes $\mathcal{C}_{in}$, $i = 1, \cdots, N_n$, each with length less than or equal to Δ_n. Cover $\mathcal{D}$ with sets $\mathcal{D} \cap \mathcal{C}_{in}$, $i = 1, \cdots, N_n$. Discard empty sets and let $\mathcal{D}_{in}$, $i = 1, \cdots, M_n$, be the remaining sets. Then $t_1, t_2 \in \mathcal{D}_{in}$ implies that $|t_{1j} - t_{2j}| < \Delta_n$, $j = 1, \cdots, d$. Note also that

$$M_n \leq \left(\frac{1}{\Delta_n} + 1\right)^d.$$

For $i = 1, \cdots, M_n$, let t_i be a point in $\mathcal{D}_{in}$. Then, by triangle inequality,

$$\sup_{t\in\mathcal{D}} |\xi_n(t)| \leq \sup_{i=1,\cdots,M_n} \left[|\xi_n(t_i)| + \sup_{t\in\mathcal{D}_{in}} |\xi_n(t) - \xi_n(t_i)|\right] \leq \sup_{i=1,\cdots,M_n} |\xi_n(t_i)| + \frac{\varepsilon_n}{2}.$$

It follows that

$$P\left(\sup_{t\in\mathcal{D}} |\xi_n(t)| > \varepsilon_n\right) \leq P\left(\sup_{i=1,\cdots,M_n} |\xi_n(t_i)| > \frac{\varepsilon_n}{2}\right) \leq \sum_{i=1}^{M_n} P\left(|\xi_n(t_i)| > \frac{\varepsilon_n}{2}\right). \tag{1.4.4}$$

Notice that, by (1.4.1) and the first part of (1.4.2), $nh_n^r\xi_n(t)$ is a sum of independent and bounded random variables. Recall Bernstein's inequality (for example, from Shorack and Wellner(1986), page 855): for independent random variables $\xi_1, \cdots, \xi_n$ with bounded ranges $[-M, M]$ and zero means,

$$P(|\xi_1 + \cdots + \xi_n| > x) \leq 2exp\left(-\frac{1}{2}\frac{x^2}{v + Mx/3}\right), \tag{1.4.5}$$

for $v \geq var(\xi_1 + \cdots + \xi_n)$.

Apply the above inequality with $\xi_i = h_n^r W_n(t, Y_i)$, $x = nh_n^r\varepsilon_n/2$ and $v = nC_0^2$ to obtain

$$P\left(|\xi_n(t_i)| > \frac{\varepsilon_n}{2}\right) \leq 2exp\left(-\frac{1}{2}\frac{n^2h_n^{2r}\varepsilon_n^2/4}{nC_0^2 + C_0nh_n^r\varepsilon_n/6}\right).$$

Since $h_n^r\varepsilon_n \to 0$ as $n \to \infty$, the second term in the denominator of the fraction will be less than the first term for large enough n, and hence the above is less than

$$2exp(-Cnh_n^{2r}\varepsilon_n^2),$$

for some $0 < C < \infty$, not depending on n, h_n and ε_n.

It now readily follows from (1.4.4), the upper bound for M_n and the definition of Δ_n that

$$P\left(\sup_{t\in\mathcal{D}} |\xi_n(t)| > \varepsilon_n\right) \leq 2\left(\frac{2C_0d}{h_n^s\varepsilon_n} + 1\right)^d exp(-Cnh_n^{2r}\varepsilon_n^2), \tag{1.4.6}$$

which is $o(1)$ if $\varepsilon_n = \varepsilon n^{-\frac{1-a}{2}}h_n^{-r}$ for all $\varepsilon > 0$ and $a > 0$. The lemma is proved.

Next we are going to use Lemma 1 to show the uniform convergences in probability of the means of independent random variables which have the same forms as those in the definition of $\hat{F}(T_j, Z_i)$, $1 \leq i, j \leq n$. Moreover, their mean square convergence is also established, which is crucial to the proof of the asymptotic normality of the generalized profile maximum likelihood estimator.

Let $\tilde{U} = (U_1, U_2, \cdots, U_d)$ be a random vector in $\mathcal{R}^d$ and γ be a random variable taking values 0 or 1, and $\tilde{U}_i = (U_{i1}, U_{i2}, \cdots, U_{id}), \gamma_i$, $i = 1, \cdots, n$, be i.i.d. copies of $\tilde{U}$, γ, respectively.

Let g be a function on $\mathcal{R}^d$ and K be a function on $\mathcal{R}^1$. Let $K_b(t) = K(t/b)/b$, $t \in \mathcal{R}^1$, b depends on n, $b \to 0$ as $n \to \infty$. Also let $\tilde{v} = (v_1, v_2, \cdots, v_d)$ be a vector in $\mathcal{R}^d$. If $\tilde{u}, \tilde{v} \in \mathcal{R}^d$ and $x, y \in \mathcal{R}^1$, then $x\tilde{u} + y\tilde{v} = (xu_1 + yv_1, \cdots, xu_d + yv_d)$. Let also $d\tilde{u} = du_1 \cdots du_d$ in the integration.

Define

$$T_n(\tilde{v}) = \frac{1}{n} \sum_{i=1}^{n} g(\tilde{U}_i) K_b(U_{i1} - v_1) \cdots K_b(U_{id} - v_d).$$

The following two lemmas establish the convergence of $T_n(\tilde{v})$. Lemma 2 establishes the convergence rate of $T_n(\tilde{v})$ to its mean, in probability and in mean square, uniformly in $\tilde{v}$. Lemma 3 studies the rate behavior of the asymptotic bias of $T_n(\tilde{v})$.

Lemma 2 *Assume $\tilde{U}$ has a bounded (joint) density $f(\tilde{u})$ with support $\mathcal{D}_f = [s_1^*, t_1^*] \times \cdots \times [s_d^*, t_d^*]$, where $s_i^*, t_i^* \in \mathcal{R}^1$, $i = 1, \cdots, d$. Also assume that $K(\cdot)$ is a bounded and Lipschitz continuous function with*

$$\int_{-\infty}^{\infty} K^2(t)dt < \infty,$$

and $g(\tilde{u})$ is bounded. Then,

$$nb^d \sup_{\tilde{v} \in \mathcal{D}_f} E|T_n(\tilde{v}) - ET_n(\tilde{v})|^2 = O(1) \tag{1.4.7}$$

and for all $a > 0$,

$$\sqrt{n^{1-a} b^{2d}} \sup_{\tilde{v} \in \mathcal{D}_f} |T_n(\tilde{v}) - ET_n(\tilde{v})| = o_p(1). \tag{1.4.8}$$

Proof. Using the fact that $Var(Y) \leq EY^2$, for any random variable Y, and the change of variable formula, we obtain

$$\begin{aligned}
Var(T_n(\tilde{v})) &= \frac{1}{n} Var(g(\tilde{U}) K_b(U_1 - v_1) \cdots K_b(U_d - v_d)) \\
&\leq \frac{1}{n} E\{g(\tilde{U}) K_b(U_1 - v_1) \cdots K_b(U_d - v_d)\}^2 \\
&= \frac{1}{n} \int \cdots \int g^2(\tilde{u}) \frac{1}{b^{2d}} K^2(\frac{u_1 - v_1}{b}) \cdots K^2(\frac{u_d - v_d}{b}) f(\tilde{u}) d\tilde{u} \\
&= \frac{1}{nb^d} \int \cdots \int g^2(\tilde{v} + b\tilde{t}) K^2(t_1) \cdots K^2(t_d) f(\tilde{v} + b\tilde{t}) d\tilde{t}.
\end{aligned}$$

Therefore, by the boundedness of f and g, and the square integrability of K,

$$\sup_{\tilde{v} \in \mathcal{D}_f} nb^d Var(T_n(\tilde{v})) = O(1).$$

Hence (1.4.7) is proved.

Apply Lemma 1 with $t = \tilde{v}$, $\mathcal{D} = \mathcal{D}_f$, $\xi_n(t) = T_n(\tilde{v}) - ET_n(\tilde{v})$, $h_n = b$, $r = d$ and $s = d + 1$ to obtain (1.4.8).

Lemma 3 *Assume the conditions of Lemma 2 hold.*

(1) If f and g are also Lipschitz continuous and K has support $[-1, 1]$ and satisfies

$$\int K(t) dt = 1, \quad \int |K(t)| dt < \infty.$$

Then,

$$\sup_{\tilde{v} \in \mathcal{D}_f^0} |ET_n(\tilde{v}) - g(\tilde{v}) f(\tilde{v})| = O(b),$$

where $\mathcal{D}_f^0 = [s_1^* + b, t_1^* - b] \times \cdots [s_d^* + b, t_d^* - b]$

(2) Suppose f and g have up to rth bounded and continuous (partial) derivatives, and K is an rth order kernel supported on $[-1, 1]$, and symmetric around zero. Then

$$\sup_{\tilde{v} \in \mathcal{D}_f^0} |ET_n(\tilde{v}) - g(\tilde{v}) f(\tilde{v})| = O(b^r).$$

Proof. We only prove the second assertion since the first one can be proved in a similar but simpler way. Change of variables and Taylor expansion yields

$$\begin{aligned}
E[T_n(\tilde{v})] &= E\left(g(\tilde{U})K_b(U_1 - v_1)\cdots K_b(U_d - v_d)\right) \\
&= \int\cdots\int g(\tilde{u})K(\frac{u_1 - v_1}{b})\cdots K(\frac{u_d - v_d}{b})f(\tilde{u})d\tilde{u} \\
&= \int_0^1\cdots\int_0^1 g(\tilde{v} + b\tilde{t})K(t_1)\cdots K(t_d)f(\tilde{v} + b\tilde{t})d\tilde{t} \\
&= \int_0^1\cdots\int_0^1 g(\tilde{v})f(\tilde{v})d\tilde{t} \\
&\qquad + \frac{1}{r!}\int_0^1\cdots\int_0^1 \left[\sum_{j=1}^{d}\frac{\partial^r}{\partial {v_j}^r}\left(g(\tilde{v}^*)f(\tilde{v}^*)\right)b^r t_j^r\right]\prod_{j=1}^{d}K(t_j)dt_j \\
&= g(\tilde{v})f(\tilde{v}) + O(b^r),
\end{aligned}$$

uniformly in $\tilde{v} \in \mathcal{D}_f^0$, where $\tilde{v}^* = ({v_1}^*, \cdots, {v_d}^*)$ and ${v_j}^*$ is between $v_j - b$ and $v_j + b$. In the last two steps, the assumption $\int_{-1}^{1} t^s K(t)dt = 0, s = 1, \cdots, r-1$ and $\int_{-1}^{1} K(t)dt = 1$ were used.

The following two lemmas discuss the convergences of two other forms of means of independent random variables based on kernels. They will be used to prove the theorems in the following section. In Lemma 4, it is already centered; and in Lemma 5 there is some kind of centering.

Lemma 4 *Assume that the conditions of Lemma 2 hold. Assume also that $g(\tilde{v})$ is the conditional expectation of γ given $\tilde{U} = \tilde{v}$. Let*

$$S_n(\tilde{v}) = \frac{1}{n}\sum_{i=1}^{n}[\gamma_i - g(\tilde{U}_i)]\prod_{j=1}^{d}K_b(U_{ij} - v_j).$$

Then,

$$nb^d \sup_{\tilde{v}\in\mathcal{D}_f} E|S_n(\tilde{v})|^2 = O(1),$$

and for all $a > 0$,

$$\sqrt{n^{1-a}b^{2d}} \sup_{\tilde{v}\in\mathcal{D}_f} |S_n(\tilde{v})| = o_p(1).$$

Proof. Note that

$$E[S_n(\tilde{v})|\tilde{U}_i, i = 1, \cdots, n)] = 0.$$

Hence

$$\begin{aligned} Var\,(S_n(\tilde{v})) &= E\left[Var\left(S_n(\tilde{v})|\tilde{U}_i, i = 1, \cdots, n\right)\right] \\ &= E\left[\frac{1}{n}\sum_{i=1}^{n} g(\tilde{U}_i)(1 - g(\tilde{U}_i))\prod_{j=1}^{d} K_b^2(U_{ij} - v_j)\right]. \end{aligned}$$

The rest of the proof is exactly the same as that of Lemma 2.

Lemma 5 *Assume the conditions of Lemma 2 hold and that g is Lipschitz continuous. Let*

$$T_n'(\tilde{v}) = \frac{1}{n}\sum_{i=1}^{n}[g(\tilde{U}_i) - g(\tilde{v})]K_b(U_{i1} - v_1)\cdots K_b(U_{id} - v_d).$$

For the variance part of $T_n'(\tilde{v})$, we have

$$nb^{d-2}\sup_{\tilde{v}\in\mathcal{D}_f} E|T_n'(\tilde{v}) - ET_n'(\tilde{v})|^2 = O(1) \tag{1.4.9}$$

and for all $a > 0$,

$$\sqrt{n^{1-a}b^{2d}}\sup_{\tilde{v}\in\mathcal{D}_f}|T_n'(\tilde{v}) - ET_n'(\tilde{v})| = o_p(1). \tag{1.4.10}$$

For the bias part, we have the following.

(1) If f is also Lipschitz continuous and K satisfies: $\int K(t)dt = 1$, $\int |K(t)|dt < \infty$, and has support $[0, 1]$, then

$$\sup_{\tilde{v}\in\mathcal{D}_f}|ET_n'(\tilde{v})| = O(b). \tag{1.4.11}$$

(2) If f and g have up to rth bounded and continuous (partial) derivatives, and that K is a rth order kernel supported on $[-1, 1]$, symmetric around zero, then

$$\sup_{\tilde{v}\in\mathcal{D}_f^0}|ET_n'(\tilde{v})| = O(b^r). \tag{1.4.12}$$

Proof. Since we have the difference term $g(\tilde{U}_i) - g(\tilde{v})$ in $\tilde{T}_n(\tilde{v})$, we should expect a better convergence rate than that of $T_n(\tilde{v})$. The proof is similar to that of Lemmas 2 and 3.

For any $\tilde{v} \in \mathcal{D}_f$,

$$
\begin{aligned}
Var(T_n'(\tilde{v})) &= \frac{1}{n} Var\left[(g(\tilde{U}) - g(\tilde{v}))\prod_{j=1}^{d} K_d(U_j - v_j)\right] \\
&\leq \frac{1}{n} E\left[(g(\tilde{U}) - g(\tilde{v}))^2 \prod_{j=1}^{d} K_b^2(U_j - v_j)\right] \\
&= \frac{1}{n}\int \cdots \int [g(\tilde{u}) - g(\tilde{v})]^2 \frac{1}{b^{2d}} K^2\left(\frac{u_1 - v_1}{b}\right)\cdots K^2\left(\frac{u_d - v_d}{b}\right) f(\tilde{u}) d\tilde{u} \\
&= \frac{1}{nb^d}\int \cdots \int [g(\tilde{v} + b\tilde{t}) - g(\tilde{v})]^2 K^2(t_1)\cdots K^2(t_d) f(\tilde{v} + b\tilde{t}) d\tilde{t} \\
&\leq \frac{C}{nb^{d-2}},
\end{aligned}
$$

for some finite real number C, not depending on $\tilde{v}$. Thus (1.4.9) is proved.

Apply Lemma 1 with $t = \tilde{v}$, $\mathcal{D} = \mathcal{D}_f$, $\xi(t) = T_n'(\tilde{v}) - E(T_n'(\tilde{v}))$, $h_n = b$, $r = d$ and $s = d + 1$ to obtain (1.4.10). Assertion (1.4.11) and (1.4.12) can be proved in the same way as in the proof of Lemma 3.

1.4.2 Proof of Theorem 1 and 2

Before giving proofs of Theorem 1 and 2, we shall use the general results of the previous section to obtain some preliminaries for their proofs. To begin with, we shall first establish the uniform convergence of $\hat{F}(T_j, Z_i)$ to $F(T_j, Z_i)$ over all $1 \leq i, j \leq n$, $\hat{g}_\theta(Z_i)$ to $g_\theta(Z_i)$, $\dot{\hat{g}}_\theta(Z_i)$ to $\dot{g}_\theta(Z_i)$ and $\ddot{\hat{g}}_\theta(Z_i)$ to $\ddot{g}_\theta(Z_i)$ over all $1 \leq i \leq n$ and $\theta \in \mathcal{N}_0$. The expected square differences between $\hat{F}(T_j, Z_i)$ and $F(T_j, Z_i)$, between $\hat{g}_\theta(Z_i)$ and $g_\theta(Z_i)$, and between $\dot{\hat{g}}_\theta(Z_i)$ and $\dot{g}_\theta(Z_i)$ are established as well.

By assumption (A1), let $\mathcal{Z} = [z_1^*, z_2^*]$ and $\mathcal{T} = [t_1^*, t_2^*]$, two finite real intervals. Let $\mathcal{Z}^0 = [z_1^* + b, z_2^* - b]$ and $\mathcal{T}^0 = [t_1^* + b, t_2^* - b]$. Then the support of h, $\mathcal{D}_h = [t_1^*, t_2^*] \times [z_1^*, z_2^*]$. Also, let $\mathcal{D}_h^0 = \mathcal{T}^0 \times \mathcal{Z}^0$.

Recall the definition of $F(T_j, Z_i)$ from (1.1.1). Write

$$\hat{F}(T_j, Z_i) - F(T_j, Z_i, \theta_0) = \frac{V_n^{(j,i)}(T_j, Z_i) + B_n^{(j,i)}(T_j, Z_i)}{B_{n0}^{(j,i)}(T_j, Z_i)}, \tag{1.4.13}$$

where

$$V_n^{(j,i)}(t, z) = \frac{1}{n} \sum_{l \neq i,j} [\delta_l - F(T_l, Z_l, \theta_0)] K_b(T_l - t) K_b(Z_l - z), \tag{1.4.14}$$

$$B_n^{(j,i)}(t, z) = \frac{1}{n} \sum_{l \neq i,j} [F(T_l, Z_l, \theta_0) - F(t, z, \theta_0)] K_b(T_l - t) K_b(Z_l - z) \tag{1.4.15}$$

and

$$B_{n0}^{(j,i)}(t, z) = \frac{1}{n} \sum_{l \neq i,j} K_b(T_l - t) K_b(Z_l - z).$$

We first show that $V_n^{(j,i)}(T_j, Z_i)$ and $B_n^{(j,i)}(T_j, Z_i)$ converge to 0 in probability, uniformly over $1 \leq i, j \leq n$, and the conditional expectation of the squares of them, given T_j and Z_i, converge to 0 at certain rate, uniformly over $1 \leq i, j \leq n$. The same convergence results of $B_{n0}^{(j,i)}(T_j, Z_i)$ to $h(T_j, Z_i)$ are obtained as well. The previous lemmas are used to obtain these convergence results. More specifically, we have the following lemma.

In the following, $\sup_{j,i}$ stands for $\sup_{1 \leq j,i \leq n}$ and $\sup_{(T_j, Z_i) \in \mathcal{D}_h}$ stands for $\sup_{(i,j: 1 \leq j,i \leq n, (T_j, Z_i) \in \mathcal{D}_h)}$.

Lemma 6 *(1) Assume that the conditions (A1) and (A3′) hold, and $b = O(n^{-a})$ with $0 < a < \frac{1}{4}$. Then,*

$$\sup_{j,i} |V_n^{(j,i)}(T_j, Z_i)| = o_p(1), \tag{1.4.16}$$

$$\sup_{j,i} E_{j,i} |V_n^{(j,i)}(T_j, Z_i)|^2 = o_p(n^{-\frac{1}{2}}), \tag{1.4.17}$$

$$\sup_{(T_j, Z_i) \in \mathcal{D}_h} |B_n^{(j,i)}(T_j, Z_i)| = o_p(1), \tag{1.4.18}$$

$$\sup_{(T_j,Z_i)\in\mathcal{D}_h^0} |B_{n0}^{(j,i)}(T_j,Z_i) - h(T_j,Z_i)| = o_p(1), \tag{1.4.19}$$

where $E_{j,i}$ stands for conditional expectation, given T_j, Z_i.

(2) Assume (A1)-(A3) hold with $r = 4$, and $b = O(n^{-a})$ with $\frac{1}{16} < a < \frac{1}{4}$. Then

$$\sup_{(T_j,Z_i)\in\mathcal{D}_h^0} E_{j,i}|B_n^{(j,i)}(T_j,Z_i)|^2 = o_p(n^{-\frac{1}{2}}) \tag{1.4.20}$$

$$\sup_{(T_j,Z_i)\in\mathcal{D}_h^0} E_{j,i}|B_{n0}^{(j,i)}(T_j,Z_i) - h(T_j,Z_i)|^2 = o_p(n^{-\frac{1}{2}}). \tag{1.4.21}$$

Proof. Define

$$V_n(t,z) = \frac{1}{n}\sum_{l=1}^{n}[\delta_l - F(T_l,Z_l,\theta_0)]K_b(T_l - t)K_b(Z_l - z).$$

Apply Lemma 4 with $S_n(\tilde{v}) = \tilde{V}_n(t,z)$, $d = 2$, $\gamma_i = \delta_i$ and $g = F(t,z,\theta_0)$ to obtain

$$\sup_{(t,z)\in\mathcal{D}_h} E|V_n(t,z)|^2 = O(\frac{1}{nb^2}) = o(n^{-\frac{1}{2}}), \tag{1.4.22}$$

and

$$\sup_{(t,z)\in\mathcal{D}_h} |V_n(t,z)| = o_p(1). \tag{1.4.23}$$

Since K is bounded, by the definition of $V_n^{(j,i)}(t,z)$ and $V_n(t,z)$, we obtain

$$\sup_{1\le i,j\le n}\sup_{(t,z)\in\mathcal{D}_h} |V^{(j,i)}(t,z) - \tilde{V}(t,z)| \le \frac{C}{nb^2} = o(n^{-\frac{1}{2}}), \tag{1.4.24}$$

for some constant $0 < C < \infty$. It follows from (1.4.23), (1.4.24) and the triangle inequality that

$$\sup_{1\le j,i\le n}\sup_{(t,z)\in\mathcal{D}_h} |V_n^{(j,i)}(t,z)| = o_p(1).$$

Hence (1.4.16) is obtained.

Similarly, by (1.4.22) and (1.4.24), apply the inequality $(x-y)^2 \le 2(x^2+y^2)$, $x, y \in \mathcal{R}^1$ to obtain

$$\sup_{1\le j,i\le n}\sup_{(t,z)\in\mathcal{D}_h} E|V_n^{(j,i)}(t,z)|^2 = O(\frac{1}{nb^2}) = o(n^{-\frac{1}{2}}).$$

Hence (1.4.17) is obtained.

Define

$$B_n(t,z) = \frac{1}{n}\sum_{l=1}^{n}[F(T_l, Z_l, \theta_0) - F(t, z, \theta_0)]K_b(T_l - t)K_b(Z_l - z)$$

and

$$B_{n0}(t,z) = \frac{1}{n}\sum_{l=1}^{n} K_b(T_l - t)K_b(Z_l - z).$$

Apply (1.4.10) and (1.4.11) of Lemma 5 with $T'_n(\tilde{v}) = B_n(t,z)$, $g(\tilde{v}) = F(t,z,\theta_0)$, $\tilde{U}_i = (T_i, Z_i)$, $d = 2$ and $a = a_0$ to obtain that, for each $a_0 > 0$,

$$\sup_{(t,z)\in\mathcal{D}_h} |B_n(t,z)| = o_p(n^{-1+a_0}b^{-4}) + O(b),$$

which is $o_p(1)$ as a_0 is chosen to be small enough. This is because of the assumption on the convergence rate of b to 0. Similar argument as above leads to (1.4.18).

Similarly, apply (1.4.8) of Lemma 2 and part (1) of Lemma 3 with $T_n(\tilde{v}) = B_{n0}(t,z)$, $g(\tilde{v}) = 1$, $d = 2$ and $\tilde{U}_i = (T_i, Z_i)$ to obtain that

$$\sup_{(t,z)\in\mathcal{D}_h^0} |B_{n0}(t,z) - h(t,z)| = o_p(1), \tag{1.4.25}$$

and (1.4.19) follows from the same discussion as the above.

Use the identity $E(Y^2) = var(Y) + (EY)^2$ for any random variable Y, and then apply (1.4.9) and (1.4.12) of Lemma 5 with $T'_n(\tilde{v}) = B_n(t,z)$, $r = 4$ and $d = 2$ to obtain that

$$\sup_{(t,z)\in\mathcal{D}_h^0} E|B_n(t,z)|^2 = O(n^{-1}) + O(b^8),$$

which is $o(n^{-\frac{1}{2}})$ since $b = O(n^{-a})$ with $a > \frac{1}{16}$. Thus (1.4.20) follows from the same discussion as above.

Similarly, apply (1.4.7) of Lemma 2 and part (2) of Lemma 3 with $T_n(\tilde{v}) = B_{n0}(t,z)$, $r = 4$ and $d = 2$ to obtain

$$\sup_{(t,z)\in\mathcal{D}_h^0} E|B_{n0}(t,z) - h(t,z)|^2 = o_p(n^{-\frac{1}{2}}).$$

(1.4.21) follows from the same discussion as above. The lemma is proved.

Since, by assumption (A1), $h(t,z)$ is bounded away from 0 and ∞, and $F(t,z,\theta_0)$ bounded away from 0 and 1, their estimators will also have these properties with probability approaching 1 as the sample size tends to infinity. We then discuss the convergence of these estimators to their limits only on the set on which these properties are satisfies.

There exist real numbers $0 < a_1 \le a_2 < \infty$ such that $a_1 < \inf_{(t,z)\in\mathcal{D}_h} h(t,z)$ and $a_2 > \sup_{(t,z)\in\mathcal{D}_h} h(t,z)$, and $0 < d_1 \le d_2 < 1$ such that $d_1 < \inf_{(t,z)\in\mathcal{D}_h} F(t,z)$ and $d_2 > \sup_{(t,z)\in\mathcal{D}_h} F(t,z)$. Particularly, choose

$$a_1 = \inf_{(t,z)\in\mathcal{D}_h} h(t,z) - \epsilon, \quad a_2 = \sup_{(t,z)\in\mathcal{D}_h} h(t,z) + \epsilon,$$

and

$$d_1 = \inf_{(t,z)\in\mathcal{D}_h} F(t,z,\theta_0) - \epsilon, \quad d_2 = \sup_{(t,z)\in\mathcal{D}_h} F(t,z,\theta_0) + \epsilon,$$

for some $\epsilon > 0$.

Write $\hat{F}^{(j,i)}(T_j, Z_i)$ for $\hat{F}(T_j, Z_i)$ as the latter depends on (j,i), and let $\hat{F}^{(j,i)}(t,z)$ be obtained from (1.1.1) with T_j, Z_i replaced by t, z respectively. Let

$$A_{n1} = \{a_1 \le \min_{\substack{(t,z)\in\mathcal{D}_h^0\\ 1\le i,j\le n}} B_{n0}^{(j,i)}(t,z) \le \max_{\substack{(t,z)\in\mathcal{D}_h^0\\ 1\le i,j\le n}} B_{n0}^{(j,i)}(t,z) \le a_2\},$$

$$A_{n2} = \{d_1 \le \min_{\substack{(t,z)\in\mathcal{D}_h^0\\ 1\le i,j\le n}} \hat{F}^{(j,i)}(t,z) \le \max_{\substack{(t,z)\in\mathcal{D}_h^0\\ 1\le i,j\le n}} \hat{F}^{(j,i)}(t,z) \le d_2\}.$$

In the definition of $\hat{g}(Z_i)$, see (1.1.2), the summation is taken over these j such that $T_j \in \mathcal{T}^0$, i.e. $w_1(T_j, b) = 1$, $j = 1, \cdots, n$. As we discuss the convergence rate of $\hat{g}(Z_i)$ to $g(Z_i)$, we want to exclude the case when all the T_j fall into the edge area, more specifically, $\sum_{j=1}^n w_1(T_j, b) = 0$. Therefore, define

$$A_{n3} = \Big\{\sum_{j=1}^n w_1(T_j) > 0\Big\}.$$

It is easy to see that, the probability of the complement of A_{n3}, $P(A_{n3}^c) = O(b^n)$, by the assumption (A1).

Let

$$A_n = A_{n1} \bigcap A_{n2} \bigcap A_{n3}.$$

The probability of A_n is expected to go to 1 as n tends to ∞. This is proved later. Next the main results used to prove the consistency and asymptotic normality of the generalized profile maximum likelihood estimator are established in the following two lemmas.

Lemma 7 *(1) Assume condition (A1) and (A3′) hold, and* $b = O(n^{-a})$ *with* $0 < a < \frac{1}{4}$. *Then*

$$\sup_{(T_j, Z_i) \in \mathcal{D}_h^0} |\hat{F}(T_j, Z_i) - F(T_j, Z_i, \theta_0)| = o_p(1). \tag{1.4.26}$$

(2) Assume condition (A1),(A2) and (A3) hold, and $b = O(n^{-a})$ *with* $\frac{1}{16} < a < \frac{1}{4}$. *Then*

$$\sup_{(T_j, Z_i) \in \mathcal{D}_h^0} E_{j,i} |\hat{F}(T_j, Z_i) - F(T_j, Z_i, \theta_0)|^2 I_{A_n} = o_p(n^{-\frac{1}{2}}). \tag{1.4.27}$$

Proof. Note that $h(t, z)$ is bounded away from 0. Thus (1.4.26) follows from (1.4.13), (1.4.16), (1.4.18), (1.4.19), and (1.4.27) follows from (1.4.13), (1.4.17), (1.4.20), (1.4.19). The lemma is proved.

Recall that $\dot{g}_\theta(z)$ is the first (partial) derivative of $g_\theta(z)$ with respect to θ. Let $\ddot{g}_\theta(z)$ be the second (partial) derivative of $g_\theta(z)$ with respect to θ. Similarly define $\dot{\hat{g}}_\theta(z)$ and $\ddot{\hat{g}}_\theta(z)$.

Lemma 8 *(1) If (A1)-(A3) hold, and* $b = O(n^{-a})$ *with* $\frac{1}{16} < a < \frac{1}{4}$, *then*

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} E_i |\hat{g}_\theta(Z_i) - g_\theta(Z_i)|^2 I_{A_n} = o_p(n^{-\frac{1}{2}}), \tag{1.4.28}$$

and

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} E_i |\dot{\hat{g}}_\theta(Z_i) - \dot{g}_\theta(Z_i)|^2 I_{A_n} = o_p(n^{-\frac{1}{2}}). \tag{1.4.29}$$

where E_i stands for the conditional expectation given Z_i.

(2) If condition (A1), (A3′) hold, and $b = O(n^{-a})$ with $0 < a < \frac{1}{4}$, then

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} |\hat{g}_\theta(Z_i) - g_\theta(Z_i)| = o_p(1), \tag{1.4.30}$$

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} |\dot{\hat{g}}_\theta(Z_i) - \dot{g}_\theta(Z_i)| = o_p(1),$$

and

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} |\ddot{\hat{g}}_\theta(Z_i) - \ddot{g}_\theta(Z_i)| = o_p(1).$$

Proof. We prove only (1.4.30) and (1.4.28). The proof of the remaining results will be similar.

In view if (1.1.2) and (1.1.3), $\hat{g}_\theta(Z_i) - g_\theta(Z_i)$ can be decomposed into $R_{n1,\theta}(Z_i) + R_{n2,\theta}(Z_i)$, where

$$R_{n1,\theta}(Z_i) = -\frac{\sum_{j\neq i} w_1(T_j, b)\Lambda(T_j, \theta)[log(1 - \hat{F}(T_j, Z_i)) - log(1 - F(T_j, Z_i, \theta_0))]}{\sum_{j\neq i} w_1(T_j, b)\Lambda^2(T_j, \theta)},$$

and

$$R_{n2,\theta}(Z_i) = \left[\frac{\sum_{j\neq i} w_1(T_j, b)\Lambda(T_j, \theta)\Lambda(T_j, \theta_0)}{\sum_{j\neq i} w_1(T_j, b)\Lambda^2(T_j, \theta)} - \frac{E\Lambda(T, \theta)\Lambda(T, \theta_0)}{E\Lambda^2(T, \theta)}\right] g(Z_i).$$

It is enough to show that

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} E_i |R_{nk,\theta}(Z_i)|^2 I_{A_n} = o_p(n^{-\frac{1}{2}}), \quad k = 1, 2, \tag{1.4.31}$$

under the conditions of part (1), and

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} |R_{nk,\theta}(Z_i)| = o_p(1), \quad k = 1, 2, \tag{1.4.32}$$

under the conditions of part (2). By the mean value theorem,

$$|log(x) - log(y)| \leq \frac{|x-y|}{x \wedge y}, \tag{1.4.33}$$

for all positive x, y. Apply this with $x = 1 - \hat{F}(T_i, Z_i)$ and $y = 1 - F(T_i, Z_i, \theta_0)$ to obtain

$$|log(1 - \hat{F}(T_i, Z_i)) - log(1 - F(T_i, Z_i, \theta_0))| \leq \frac{|\hat{F}(T_i, Z_i) - F(T_i, Z_i, \theta_0)|}{(1 - \hat{F}(T_i, Z_i)) \wedge (1 - F(T_i, Z_i, \theta_0))}. \tag{1.4.34}$$

By the definition of $R_{n1,\theta}(Z_i)$, and the boundedness of $\Lambda(t, \theta)$ away from 0 and ∞, we obtain that, on A_{n3},

$$\sup_{\substack{Z_i \in \mathcal{Z}^0 \\ \theta \in \mathcal{N}_0}} \sup |R_{n1,\theta}(Z_i)| \leq C \sup_{(T_j, Z_i) \in \mathcal{D}_h} |log(1 - \hat{F}(T_i, Z_i)) - log(1 - F(T_i, Z_i, \theta_0))|,$$

for some constant $0 < C < \infty$. This, (1.4.34), (1.4.26), (1.4.27) and the boundedness of $F(t, z, \theta_0)$ away from 0 and 1 imply (1.4.31) and (1.4.32) with $k = 1$.

The Lipschitz continuity of $\Lambda(t, \theta)$ with respect to θ uniformly in t, and the uniform SLLN imply (1.4.31) and (1.4.32) with $k = 2$. (They can also be proved by applying (1).)

Notice that $\hat{F}(T_j, Z_i)$ does not depend on θ and $\hat{g}_\theta(Z_i)$ depends on θ only through $\Lambda(T_j, \theta)$. By the assumption on $\Lambda(T_j, \theta)$, similarly we can prove the remaining assertions. The lemma is proved.

We shall show that the probability of A_n approaches 1 as $n \to \infty$.

Lemma 9 *Assume that (A1) and (A3′) hold. Then*

$$\lim_{n \to \infty} P(A_n) = 1.$$

Proof. It suffices to show that

$$\lim_{n \to \infty} P(A_{nk}) = 1 \quad \text{or equivalently} \quad \lim_{n \to \infty} P(A_{nk}^c) = 0, \quad k = 1, 2, 3.$$

We have seen that $\lim_{n\to\infty} P(A_{n3}) = 0$ by its definition. We first prove the above assertion with $k = 1$. By the definition of A_{n1}, its compliment equal to

$$\left(\sup_{\substack{(t,z)\in\mathcal{D}_h^0 \\ 1\le i,j\le n}} |B_{n0}^{(j,i)}(t,z) - h(t,z)| > \epsilon \right).$$

One also notes that

$$B_{n0}(t,z) - B_{n0}^{(j,i)}(t,z) = \left[K_b(T_j - t)K_b(Z_j - z) + K_b(T_i - t)K_b(Z_i - z)\right]/n,$$

the absolute value of which is less than $C/(nb^2)$ for all $1 \le i,j \le n$, $(t,z) \in \mathcal{D}_h$, and for some finite constant C, since K is bounded. Hence we have

$$\begin{aligned} &P\left(\sup_{\substack{(t,z)\in\mathcal{D}_h^0 \\ 1\le i,j\le n}} |B_{n0}^{(j,i)}(t,z) - h(t,z)| > \epsilon \right) \\ &\le P\left(\sup_{(t,z)\in\mathcal{D}_h^0} |B_{n0}(t,z) - h(t,z)| > \epsilon - C/(nb^2) \right) \end{aligned}$$

which is $o(1)$ in view of (1.4.25) and that $nb^2 \to \infty$ as $n \to \infty$. We thus obtain

$$\lim_{n\to\infty} P\left(A_{n1}\right) = 1.$$

Let

$$\tilde{F}(t,z) = \frac{\sum_{l=1}^n \delta_l K_b(T_l - t)K_b(Z_l - z)}{\sum_{l=1}^n K_b(T_l - t)K_b(Z_l - z)}.$$

Similarly, one can obtain

$$P\left(A_{n2}^c\right) \le P\left(\sup_{(t,z)\in\mathcal{D}_h^0} |\tilde{F}(t,z) - F(t,z,\theta_0)| > \epsilon - \sup_{j,i,(t,z)\in\mathcal{D}_h^0} |\tilde{F}(t,z) - \hat{F}^{(j,i)}(t,z)| \right),$$

which is $o(1)$ if

$$\sup_{j,i,(t,z)\in\mathcal{D}_h^0} |\tilde{F}(t,z) - \hat{F}^{(j,i)}(t,z)| = o_p(1).$$

This is easy to show and omitted here. The lemma is proved.

Proof of Theorem 1 (Consistency) It is enough to show that $l_{n1}(\theta)/n$ converges in probability, uniformly in $\mathcal{N}_0$, to a nonrandom function that has unique maximizer at θ_0.

We are going to prove later that

$$\sup_{\theta\in\mathcal{N}_0} |l_{n1}(\theta)/n - \bar{l}_{n1}(\theta)| = o_p(1), \tag{1.4.35}$$

where

$$\bar{l}_{n1}(\theta)$$
$$= \frac{1}{n}\sum_{i=1}^{n} w_1(T_i,b)w_2(Z_i,b)[\delta_i log(1-exp(-\Lambda(T_i,\theta)g_\theta(Z_i))) - (1-\delta_i)\Lambda(T_i,\theta)g_\theta(Z_i)].$$

By a uniform law of large numbers, which holds under our conditions, and the fact that

$$P(w_1(T,b)=0) = O(b) \quad \text{and} \quad P(w_2(Z,b)=0) = O(b),$$

we obtain

$$\sup_{\theta\in\mathcal{N}_0} |\bar{l}_{n1}(\theta) - l(\theta)| = o_p(1), \tag{1.4.36}$$

where

$$\begin{aligned} l(\theta) &= E[\delta log(1-exp(-\Lambda(T,\theta)g_\theta(Z))) - (1-\delta)\Lambda(T,\theta)g_\theta(Z)] \\ &= \iint [(1-e^{-\Lambda(t,\theta_0)g(z)})log(1-e^{-\Lambda(t,\theta)g_\theta(z)}) - \Lambda(t,\theta)g_\theta(z)e^{-\Lambda(t,\theta_0)g(z)}]h(t,z)dtdz. \end{aligned}$$

This can also be obtained by apply Lemma 1 with $t=\theta$, $\xi(t) = \bar{l}_{ni}(\theta)$, $\mathcal{D} = \mathcal{N}_0$, $r=s=0$.

Next we prove that $l(\theta)$ has a unique maximizer at θ_0. One notes that the function

$$f(y) = (1-e^{-x})log(1-e^{-y}) - ye^{-x}$$

attains its maximum at $y=x$ for any $x>0$ and $y>0$, because

$$f'(y) = \frac{e^{-y}-e^{-x}}{1-e^{-y}},$$

which is positive for $y < x$, equals 0 for $y = x$ and negative for $y > x$. Apply this with $x = \Lambda(t,\theta_0)g(z)$ and $y = \Lambda(t,\theta)g_\theta(z)$ to obtain that $l(\theta) \le l(\theta_0)$, and $l(\theta_1) = l(\theta_0)$ iff $\Lambda(t,\theta_1)g_{\theta_1}(z) = \Lambda(t,\theta_0)g(z)$, $(t,z) \in \mathcal{D}_h$. This and the identifiability condition imply that $l(\theta) < l(\theta_0)$ for any $\theta \ne \theta_0$. Therefore $l(\theta)$ is uniquely maximized at θ_0. This, (1.4.35) and (1.4.36) prove the theorem.

Now we establish (1.4.35). Write w_{ii} for $w_1(T_i,b)w_2(Z_i,b)$. It is enough to prove that

$$\sup_{\theta\in\mathcal{N}_0} |\frac{1}{n}\sum_{i=1}^{n} w_{ii}\delta_i \log(\hat{F}(T_i,Z_i,\theta)) - \frac{1}{n}\sum_{i=}^{n} w_{ii}\delta_i \log(F(T_i,Z_i,\theta))| = o_p(1) \qquad (1.4.37)$$

and

$$\sup_{\theta\in\mathcal{N}_0} |\frac{1}{n}\sum_{i=1}^{n} w_{ii}(1-\delta_i)\Lambda(T_i,\theta)\hat{g}_\theta(Z_i) - \frac{1}{n}\sum_{i=}^{n} w_{ii}(1-\delta_i)\Lambda(T_i,\theta)g_\theta(Z_i)| = o_p(1). \qquad (1.4.38)$$

Apply (1.4.33) with $x = \hat{F}(T_i,Z_i,\theta)$ and $y = F(T_i,Z_i,\theta)$ to obtain

$$|log(\hat{F}(T_i,Z_i,\theta)) - log(F(T_i,Z_i,\theta))| \le \frac{|\hat{F}(T_i,Z_i,\theta) - F(T_i,Z_i,\theta)|}{\hat{F}(T_i,Z_i,\theta) \wedge F(T_i,Z_i,\theta)} \qquad (1.4.39)$$

By the mean value theorem,

$$|e^{-x} - e^{-y} \le |x-y|, \qquad (1.4.40)$$

for all positive x, y. Apply this with $x = \Lambda(T_i,\theta)\hat{g}_\theta(Z_i)$ and $y = \Lambda(T_i,\theta)g_\theta(Z_i)$ and recall the definition of $\hat{F}(T_i,Z_i,\theta)$ and $F(T_i,Z_i,\theta)$ (see (1.1.5) and (1.1.4)) to obtain that the right hand side of (1.4.39) is no more than

$$\frac{\Lambda(T_i,\theta)|\hat{g}(Z_i) - g(Z_i)|}{\hat{F}(T_i,Z_i,\theta) \wedge F(T_i,Z_i,\theta)}.$$

Therefore the left hand side of (1.4.37) is no more than

$$\frac{\sup_{\theta\in\mathcal{N}_0,(T_i,Z_i)\in\mathcal{D}_h^0} \Lambda(T_i,\theta)|\hat{g}_\theta(Z_i) - g_\theta(Z_i)|}{\inf_{\theta\in\mathcal{N}_0,(T_i,Z_i)\in\mathcal{D}_h^0} \hat{F}(T_i,Z_i,\theta) \wedge F(T_i,Z_i,\theta)},$$

which is $o_p(1)$ because of the boundedness of Λ, (1.4.26), (1.4.30) and the boundedness of $F(t, z, \theta_0)$ away from 0. This proves (1.4.37), and (1.4.38) can be proved in a similar way. Hence (1.4.35) is proved.

Proof of Theorem 2 (Asymptotic normality) We first prove the following.

$$\sup_{\substack{(T_j, Z_i) \in \mathcal{D}_h^0 \\ \theta \in \mathcal{N}_0}} |\hat{F}(T_j, Z_i, \theta) - F(T_j, Z_i, \theta)| = o_p(1), \tag{1.4.41}$$

where $\hat{F}(t, z, \theta)$ is defined in (1.1.5) and $F(t, z, \theta)$ is defined in (1.1.4). Apply (1.4.40) with $x = \Lambda(t, \theta)\hat{g}_\theta(z)$ and $y = \Lambda(t, \theta)g_\theta(z)$ to obtain

$$|\hat{F}(t, z, \theta) - F(t, z, \theta)| \leq \Lambda(t, \theta)|\hat{g}_\theta(z) - g_\theta(z)|.$$

This, the boundedness of Λ and (1.4.30) imply (1.4.41).

By the definition of $g_\theta(z)$, see (1.1.3), and the assumption on Λ and g (see Assumption (A1)), $g_\theta(z)$, as a function of θ and z on $\mathcal{N}_0 \times \mathcal{Z}$, is bounded away from 0 and ∞. It follows from the definition of $F(t, z, \theta)$ that, as a function of t, z and θ, it is boundedness from 0 and ∞.

Let

$$\hat{D}_1(T_i, Z_i, \theta) = \dot{\Lambda}(T_i, \theta)\hat{g}_\theta(Z_i) + \Lambda(T_i, \theta)\dot{\hat{g}}_\theta(Z_i), \tag{1.4.42}$$

and

$$D_1(T_i, Z_i, \theta) = \dot{\Lambda}(T_i, \theta)g_\theta(Z_i) + \Lambda(T_i, \theta)\dot{g}_\theta(Z_i).$$

It follows from part (2) of Lemma 8 that

$$\sup_{(T_i, Z_i) \in \mathcal{D}_h^0} |\hat{D}_1(T_i, Z_i, \theta) - D_1(T_i, Z_i, \theta)| = o_p(1). \tag{1.4.43}$$

Now we begin to prove the theorem. The derivative, with respect to θ, of the modified profile log-likelihood, $l_{n1}(\theta)$, defined in (1.1.6), is given by

$$\frac{\partial}{\partial \theta} l_{n1}(\theta) = \sum_{i=1}^{n} w_{ii} \left[\frac{\delta_i}{\hat{F}(T_i, Z_i, \theta)} - 1 \right] \hat{D}_1(T_i, Z_i, \theta).$$

Let

$$S_n(\theta) = \frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}l_{n1}(\theta).$$

Then, by the mean value theorem,

$$0 = S_n(\hat{\theta}) = S_n(\theta_0) + \sqrt{n}(\hat{\theta} - \theta_0)\dot{S}_n(\theta^*), \tag{1.4.44}$$

where θ^* is between θ_0 and $\hat{\theta}$, and

$$\begin{aligned}\dot{S}_n(\theta) = &-\frac{1}{n}\sum_{i=1}^{n} w_{ii}\frac{\delta_i[1-\hat{F}(T_i,Z_i,\theta)]}{\hat{F}^2(T_i,Z_i,\theta)}\hat{D}_1^2(T_i,Z_i,\theta)\\ &+\frac{1}{n}\sum_{i=1}^{n} w_{ii}\left[\frac{\delta_i}{\hat{F}(T_i,Z_i,\theta)} - 1\right]\\ &\qquad\times[\ddot{\Lambda}(T_i,\theta)\hat{g}_\theta(Z_i) + 2\dot{\Lambda}(T_i,\theta)\dot{\hat{g}}_\theta(Z_i) + \Lambda(T_i,\theta)\ddot{\hat{g}}_\theta(Z_i)].\end{aligned}$$

We are going to show that $\dot{S}_n(\theta^*)$ converges in probability to a positive number. To do this, let $\dot{S}_n^*(\theta)$ is obtained from $\dot{S}_n(\theta)$ with $\hat{F}(T_i,Z_i,\theta)$ replaced by $F(T_i,Z_i,\theta)$, $\hat{D}_1(T_i,Z_i,\theta)$ replaced by $D_1(T_i,Z_i,\theta)$, $\hat{g}_\theta(Z_i)$ replaced by $g_\theta(Z_i)$, $\dot{\hat{g}}_\theta(Z_i)$ replaced by $\dot{g}_\theta(Z_i)$ and $\ddot{\hat{g}}_\theta(Z_i)$ replaced by $\ddot{g}_\theta(Z_i)$. In view of (1.4.41), (1.4.43), part (2) of Lemma 8, boundedness of $F(t,z,\theta)$ away from 0, and the boundedness of Λ, $\dot{\Lambda}$ and $\ddot{\Lambda}$, we obtain

$$\sup_{\theta\in\mathcal{N}_0} |\dot{S}_n(\theta) - \dot{S}_n^*(\theta)| = o_p(1). \tag{1.4.45}$$

One also notes that, under assumption (A1), $\dot{S}_n^*(\theta)$ is Lipschitz continuous in θ on $\mathcal{N}_0$. This, (1.4.45) and the triangle inequality imply

$$|\dot{S}_n(\theta^*) - \dot{S}_n^*(\theta_0)| = o_p(1). \tag{1.4.46}$$

Since $\dot{S}_n^*(\theta_0)$ is the mean of bounded random variables, it follows from the SLLN (Strong Law of Large Numbers) that $\dot{S}_n^*(\theta_0)$ converges with probability 1 to

$$-E\left[\frac{\bar{F}(T,Z,\theta_0)}{F(T,Z,\theta_0)}D_1^2(T,Z,\theta_0)\right] := -d(\theta_0).$$

This and (1.4.46) imply that $\dot{S}_n(\theta^*)$ converges to $-d(\theta_0)$ in probability. Hence it follows from this and (1.4.44) that

$$\sqrt{n}(\hat{\theta} - \theta_0) = d^{-1}(\theta_0) S_n(\theta_0)[1 + o_p(1)]. \tag{1.4.47}$$

Next we are going to find the limiting distribution of $S_n(\theta_0)$. Write $\hat{g}(z)$ for $\hat{g}_{\theta_0}(z)$. Write

$$S_n(\theta_0) = E_n + Q_n, \tag{1.4.48}$$

where

$$E_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ii} \left[\delta_i - F(T_i, Z_i, \theta_0)\right] \frac{\hat{D}_1(T_i, Z_i, \theta_0)}{\hat{F}(T_i, Z_i, \theta_0)}, \tag{1.4.49}$$

and

$$Q_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ii} \left[F(T_i, Z_i, \theta_0) - \hat{F}(T_i, Z_i, \theta_0)\right] \frac{\hat{D}_1(T_i, Z_i, \theta_0)}{\hat{F}(T_i, Z_i, \theta_0)}. \tag{1.4.50}$$

Both E_n and Q_n has contributions to the limiting distribution of $S_n(\theta_0)$.

First we deal with E_n. Write w_{j1} for $w_1(T_j, b)$. By the definition of E_n, it can be rewritten as the following

$$E_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ii}[\delta_i - F(T_i, Z_i, \theta_0)] \frac{D_1(T_i, Z_i, \theta_0)}{F(T_i, Z_i, \theta_0)} + R_n, \tag{1.4.51}$$

where

$$R_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ii}[\delta_i - F(T_i, Z_i, \theta_0)] \tilde{D}(T_i, Z_i)$$

and

$$\tilde{D}(T_i, Z_i) = \frac{\hat{D}_1(T_i, Z_i, \theta_0)}{\hat{F}(T_i, Z_i, \theta_0)} - \frac{D_1(T_i, Z_i, \theta_0)}{F(T_i, Z_i, \theta_0)}. \tag{1.4.52}$$

In view of (1.4.41) and (1.4.43), R_n is expected to go to 0 in probability. To prove this, we show the expectation of R_n^2 converges to 0 as n tends to infinity. Note that

$$\begin{aligned} R_n^2 = &\frac{1}{n}\sum_{i=1}^{n} w_{ii}[\delta_i - F(T_i, Z_i, \theta_0)]^2\tilde{D}^2(T_i, Z_i) \\ &+ \frac{1}{\sqrt{n}}\sum\sum_{i_1\neq i_2} w_{i_1i_1}[\delta_{i_1} - F(T_{i_1}, Z_{i_1}, \theta_0)]\tilde{D}(T_{i_1}, Z_{i_1}) \\ &\qquad \times w_{i_2i_2}[\delta_{i_2} - F(T_{i_2}, Z_{i_2}, \theta_0)]\tilde{D}(T_{i_2}, Z_{i_2}). \end{aligned}$$

That the first term on the right hand side of the above expression is $o_p(1)$ follows from

$$\sup_{(T_i,Z_i)\in\mathcal{D}_h^0} |\tilde{D}(T_i, Z_i)| = o_p(1),$$

which in turn follows from (1.4.41), (1.4.43) and the fact that $F(t, z, \theta)$ is bounded away from 0 and 1.

To prove that the second term of the expression of R_n^2 goes to 0 in probability, define the following

$$\hat{F}^{(i_2)}(T_j, Z_{i_1}) = \frac{\sum_{k\neq i_1,i_2,j} \delta_k K_b(T_k - T_j)K_b(Z_k - Z_{i_1})}{\sum_{k\neq i_1,j} K_b(T_k - T_j)K_b(Z_k - Z_{i_1})}, \quad 1 \leq j, i_1, i_2 \leq n, \quad i_1 \neq i_2.$$

For $1 \leq i_1, i_2 \leq n$ and $i_1 \neq i_2$, let $\tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})$ be obtained from $\tilde{D}(T_{i_1}, Z_{i_1})$ with $\hat{F}(T_j, Z_{i_1})$, $1 \leq j \leq n$, replaced by $\hat{F}^{(i_2)}(T_j, Z_{i_1})$. For any $1 \leq i_1 \leq n$, by the definitions of $\tilde{D}(T_{i_1}, Z_{i_1})$ (see (1.4.52)) and $\hat{D}_1(T_{i_1}, Z_{i_1}, \theta)$ (see (1.4.42)), $\tilde{D}(T_{i_1}, Z_{i_1})$ depends on $\hat{F}(T_j, Z_{i_1})$, $1 \leq j \leq n$, through $\hat{g}_{\theta_0}(Z_{i_1})$. See (1.1.2) for the dependence of $\hat{g}_{\theta_0}(Z_{i_1})$ on $\hat{F}(T_j, Z_{i_1})$, $1 \leq j \leq n$.

One can see that, for $1 \leq j, i_1, i_2 \leq n$ and $i_1 \neq i_2$,

$$\hat{F}(T_j, Z_{i_1}) - \hat{F}^{(i_2)}(T_j, Z_{i_1}) = \begin{cases} \frac{\delta_{i_2} K_b(T_{i_2}-T_j)K_b(Z_{i_2}-Z_{i_1})}{\sum_{k\neq i_1,j} K_b(T_k-T_j)K_b(Z_k-Z_{i_1})} & \text{if } i_2 \neq j, \\ 0 & \text{if } i_2 = j. \end{cases} \tag{1.4.53}$$

In order to study this difference, denote $W(t, z) = \delta K_b(T - t)K_b(Z - z)/n$. Since $nb^2|W(t, z)| < C_0$ for some finite number C_0, by Bernstein inequality (see (1.4.5)),

$$P\left(nb^2|W(t, z)| > nb^2\varepsilon_n\right) \leq 2exp\left(-\frac{1}{2}\frac{n^2b^4\varepsilon_n^2}{C_0^2 + C_0nb^2\varepsilon_n/3}\right) \leq 2exp\left(-Cnb^2\varepsilon_n\right). \tag{1.4.54}$$

The last inequality holds for some finite and positive number C if $b^2\varepsilon_n = O(n^{-a})$ for some $0 < a < 1$. In view of the proof of Lemma 1, instead of using $\xi_n(t)$, ε_n, h_n, d, r and s in the proof of Lemma 1, here using $nW(t,z)$, $n\varepsilon_n$, b, 2, 2 and 3, then (1.4.6) there, with the exponential part replaced by that of (1.4.54), leads to

$$P\left(\sup_{(t,z)\in\mathcal{D}_h} |W(t,z)| > \varepsilon_n\right) \leq 2\left(\frac{4C_0}{nb\varepsilon_n}+1\right)^2 exp\left(-Cnb^2\varepsilon_n\right),$$

which is $o(1)$ if ε_n is chosen to be $n^{-(1-a_0)}b^{-2}$ for all $0 < a_0 < 1$. For these values of a_0, $\varepsilon_n b^2 = O(n^{-(1-a_0)})$, so that (1.4.54) holds.

It follows that

$$\sup_{(t,z)\in\mathcal{D}_h} |W(t,z)| = o_p(n^{-(1-a_0)}b^{-2}).$$

Since $b = O(n^{-a})$ with $\frac{1}{8} < a < \frac{1}{4}$, the above rate is $o_p(n^{-(1-a_0-2a)})$ and is $o_p(n^{-\frac{1}{2}})$ if $0 < a_0 < \frac{1}{2} - 2a$. Therefore, we obtain

$$\sup_{(t,z)\in\mathcal{D}_h} |W(t,z)| = o_p(n^{-\frac{1}{2}}). \tag{1.4.55}$$

By (1.4.19) of Lemma 6, and that $\inf_{(t,z)\in\mathcal{D}_h} h(t,z) > 0$, it follows from (1.4.53) and (1.4.55) that

$$\sup_{\substack{1\leq j,i_1,i_2\leq n\\ i_1\neq i_2}} |\hat{F}(T_j,Z_{i_1}) - \hat{F}^{(i_2)}(T_j,Z_{i_1})| = o_p(n^{-\frac{1}{2}}).$$

By the definition of $\tilde{D}^{(i_2)}(T_{i_1},Z_{i_1})$ and $\tilde{D}(T_{i_1},Z_{i_1})$, $1 \leq i_1, i_2 \leq n, i_1 \neq i_2$, and assumption (A1), we can obtain that, with probability approaching 1,

$$\sup_{\substack{1\leq i_1,i_2\leq n\\ i_1\neq i_2}} |\tilde{D}^{(i_2)}(T_{i_1},Z_{i_1}) - \tilde{D}(T_{i_1},Z_{i_1})| \leq C \sup_{\substack{1\leq j,i_1,i_2\leq n\\ i_1\neq i_2}} |\hat{F}(T_j,Z_{i_1}) - \hat{F}^{(i_2)}(T_j,Z_{i_1})|,$$

for some finite and positive C. It follows from the above two displays that

$$\sup_{\substack{1\leq i_1,i_2\leq n\\ i_1\neq i_2}} |\tilde{D}^{(i_2)}(T_{i_1},Z_{i_1}) - \tilde{D}(T_{i_1},Z_{i_1})| = o_p(n^{-\frac{1}{2}}). \tag{1.4.56}$$

Next we show that, for the second part of R_n^2, $\tilde{D}(T_{i_1}, Z_{i_2})$ can be replaced with $\tilde{D}(T_{i_1}, Z_{i_1}) - \tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})$, $1 \le i_1, i_2 \le n, i_1 \ne i_2$, without changing the expectation of it. This is because of the following reason. For any $1 \le i_1 \le n$, given $(T_k, Z_k), 1 \le k \le n$, and $\delta_k, k \ne i$, the conditional expectation of $\delta_{i_1} - F(T_{i_1}, Z_{i_1}, \theta_0)$ is 0. For $1 \le i_1, i_2 \le n, i_1 \ne i_2$, by their definitions, neither $\tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})$ nor $\tilde{D}^{(i_1)}(T_{i_2}, Z_{i_2})$ depends on δ_{i_1} and δ_{i_2}, and $\tilde{D}(T_{i_1}, Z_{i_1})$ does not depend on δ_{i_1}. Given $(T_k, Z_k), 1 \le k \le n$, and $\delta_k, 1 \le k \le n, k \ne i_1$, the conditional expections of

$$[\delta_{i_1} - F(T_{i_1}, Z_{i_1})][\delta_{i_2} - F(T_{i_2}, Z_{i_2})]\tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})\tilde{D}^{(i_1)}(T_{i_2}, Z_{i_2})$$

and

$$[\delta_{i_1} - F(T_{i_1}, Z_{i_1})][\delta_{i_2} - F(T_{i_2}, Z_{i_2})][\tilde{D}(T_{i_1}, Z_{i_1}) - \tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})]\tilde{D}^{(i_1)}(T_{i_2}, Z_{i_2})$$

are zero. Thus their expectations are 0 too. Therefore the expectation of the second part of R_n^2 is equal to

$$E\Bigg(\frac{1}{n}\sum_{i_1 \ne i_2}\sum w_{i_1 i_1}[\delta_{i_1} - F(T_{i_1}, Z_{i_1}, \theta_0)][\tilde{D}(T_{i_1}, Z_{i_1}) - \tilde{D}^{(i_2)}(T_{i_1}, Z_{i_1})]$$
$$\times\, w_{i_2 i_2}[\delta_{i_2} - F(T_{i_2}, Z_{i_2}, \theta_0)][\tilde{D}(T_{i_2}, Z_{i_2}) - \tilde{D}^{(i_1)}(T_{i_2}, Z_{i_2})]\Bigg)$$

which is $o_p(1)$ in view of (1.4.56) and the boundedness of $\delta_i - F(T_i, Z_i)$. Therefore,

$$R_n = o_p(1),$$

and hence, by (1.4.51),

$$E_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}[\delta_i - F(T_i, Z_i, \theta_0)]\frac{D_1(T_i, Z_i, \theta_0)}{F(T_i, Z_i, \theta_0)} + o_p(1).$$

Since $P(w_{ii} = 0) = O(b) = o(1)$, $D_1(T_i, Z_i, \theta_0)$ is bounded, $F(T_i, Z_i, \theta_0)$ is bounded away from 0, and the conditional expectation of $\delta_i - F(T_i, Z_i, \theta_0)$ given (T_i, Z_i) is 0, it is easy to see that

$$E_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\delta_i - F(T_i, Z_i, \theta_0)]\frac{D_1(T_i, Z_i, \theta_0)}{F(T_i, Z_i, \theta_0)} + o_p(1). \tag{1.4.57}$$

Now we deal with Q_n. Recall Q_n from (1.4.50). Under the condition of the boundedness of $\Lambda(t,\theta_0)$, $\dot{\Lambda}(t,\theta_0)$ and $F(t,z,\theta_0)$ away from 0 and ∞, by the uniform boundedness of $\hat{g}(Z_i)$ and hence $\hat{F}(T_i,Z_i,\theta_0)$ on A_n (defined before), applying (1.4.40) with $x=\Lambda(T_i,\theta_0)\hat{g}(Z_i)$ and $y=\Lambda(T_i,\theta_0)g(Z_i)$, we can see that

$$\begin{aligned}
&|Q_n - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}[F(T_i,Z_i,\theta_0)-\hat{F}(T_i,Z_i,\theta_0)]\frac{D_1(T_i,Z_i,\theta_0)}{F(T_i,Z_i,\theta_0)}|I_{A_n}\\
&\le \frac{C}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}|\hat{g}(Z_i)-g(Z_i)|[|\hat{g}(Z_i)-g(Z_i)|+|\dot{\hat{g}}(Z_i)-\dot{g}(Z_i)|]I_{A_n}\\
&\le \frac{C}{\sqrt{n}}\sum_{i=1}^{n} I_{(Z_i\in Z^0)}[|\hat{g}(Z_i)-g(Z_i)|^2+|\hat{g}(Z_i)-g(Z_i)||\dot{\hat{g}}(Z_i)-\dot{g}(Z_i)|]I_{A_n}.
\end{aligned}$$

Here C is a positive and finite number. Taking expectation first conditioned on Z_i for each sub-term, we obtain that, by (1.4.28), the expectation of the first term in the last display is $o_p(1)$. It follows from Cauchy-Schwartz inequality, (1.4.28) and (1.4.29) that the expectation of the second term is also $o_p(1)$. Hence we obtain that, on A_n,

$$Q_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}[F(T_i,Z_i,\theta_0)-\hat{F}(T_i,Z_i,\theta_0)]\frac{D_1(T_i,Z_i,\theta_0)}{F(T_i,Z_i,\theta_0)} + o_p(1). \qquad (1.4.58)$$

By Taylor expansion of $1-e^{-ax}$ with respect to x at some point x_0, applying this with $x=\Lambda(T_i,\theta_0)\hat{g}(Z_i)$ and $x_0=\Lambda(T_i,\theta_0)g(Z_i)$, noticing the boundedness of Λ, we obtain

$$\begin{aligned}
&|[F(T_i,Z_i,\theta_0)-\hat{F}(T_i,Z_i,\theta_0)]+\Lambda(T_i,\theta_0)\bar{F}(T_i,Z_i,\theta_0)[\hat{g}(Z_i)-g(Z_i)]|\\
&\le C[\hat{g}(Z_i)-g(Z_i)]^2,
\end{aligned}$$

for some finite and positive number C.

This, the boundedness of $\xi(T_i,Z_i)$ (defined below) and (1.4.28) imply that, on A_n,

$$Q_n = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i,Z_i)[\hat{g}(Z_i)-g(Z_i)] + o_p(1) \qquad (1.4.59)$$

where

$$\xi(t,z) = \Lambda(t,\theta_0)D_1(t,z,\theta_0)\frac{\bar{F}(t,z,\theta_0)}{F(t,z,\theta_0)}, \quad (t,z)\in\mathcal{D}_h. \qquad (1.4.60)$$

Let

$$C_{n0} = \frac{1}{n}\sum_{j\neq i}^{n} w_{j1}\Lambda^2(T_j,\theta_0).$$

By the law of large number and that $P(w_{j1}=0)=0(1)$, C_{n0} converges in probability to $E\Lambda^2(T,\theta_0)$ which is c_0 according to the notation used before.

By Taylor expansion of $log(1-x)$ with respect to x at some point x_0, applying this with $x=\hat{F}(T_j,Z_i)$ and $x_0=F(T_j,Z_i,\theta_0)$, noticing (1.4.27) of Lemma 7, we obtain that, on A_n,

$$\begin{aligned}\hat{g}(Z_i)-g(Z_i) &= -\frac{1}{C_{n0}}\frac{1}{n}\sum_{j\neq i} w_{j1}\Lambda(T_j,\theta_0)[log(1-\hat{F}(T_j,Z_i))-log(1-F(T_j,Z_i,\theta_0))] \\ &= \frac{1}{C_{n0}}\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j,\theta_0)[\hat{F}(T_j,Z_i)-F(T_j,Z_i,\theta_0)]}{\bar{F}(T_j,Z_i,\theta_0)}+o_p(n^{-\frac{1}{2}}).\end{aligned}$$

This and (1.4.13) imply that, on A_n,

$$\hat{g}(Z_i)-g(Z_i)=R_{n1}(Z_i)+R_{n2}(Z_i)+o_p(n^{-\frac{1}{2}}), \tag{1.4.61}$$

where

$$R_{n1}(Z_i)=\frac{1}{C_{n0}}\frac{1}{n}\sum_{j\neq i} w_{j1}\Lambda(T_j,\theta_0)\frac{V_n^{(j,i)}(T_j,Z_i)}{\bar{F}(T_j,Z_i,\theta_0)B_{n0}^{(j,i)}(T_j,Z_i)}, \tag{1.4.62}$$

and

$$R_{n2}(Z_i)=\frac{1}{C_{n0}}\frac{1}{n}\sum_{j=1} w_{j1}\Lambda(T_j,\theta_0)\frac{B_n^{(j,i)}(T_j,Z_i)}{\bar{F}(T_j,Z_i,\theta_0)B_{n0}^{(j,i)}(T_j,Z_i)}. \tag{1.4.63}$$

Substitute (1.4.61) into (1.4.59) to obtain that, on A_n,

$$Q_n=Q_{n1}+Q_{n2}+o_p(1), \tag{1.4.64}$$

where

$$Q_{n1}=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i,Z_i)R_{n1}(Z_i)$$

and

$$Q_{n2} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)R_{n2}(Z_i). \tag{1.4.65}$$

Let $C_{n1} = Q_{n1}C_{n0}$. Then

$$C_{n1} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j, \theta_0)V_n(T_j, Z_i)}{\bar{F}(T_j, Z_i)B_{n0}(T_j, Z_i)}.$$

By Taylor expansion of $\frac{1}{x}$ with respect to x at some point x_0, applying this with $x = B_{n0}^{(j,i)}(T_j, Z_i)$ and $x_0 = h(T_j, Z_i)$, we obtain

$$\begin{aligned} C_{n1} = & -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j, \theta_0)}{\bar{F}(T_j, Z_i)}V_n^{(j,i)}(T_j, Z_i) \\ & \times \left[\frac{1}{h(T_j, Z_i)} - \frac{1}{[h^*(T_j, Z_i)]^2}(B_{n0}^{(j,i)}(T_j, Z_i) - h(T_j, Z_i))\right], \end{aligned}$$

where $h^*(T_j, Z_i)$ is between $h(T_j, Z_i)$ and $B_{n0}^{(j,i)}(T_j, Z_i)$. By (1.4.17) and (1.4.21) of Lemma 6, and the boundedness of h, ξ, Λ and F, using Cauchy-Schwartz inequality, we obtain that

$$C_{n1} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j, \theta_0)}{\bar{F}(T_j, Z_i)h(T_j, Z_i)}V_n(T_j, Z_i) + o_p(1).$$

Let

$$\zeta_n(T_j, Z_l) = \frac{1}{n}\sum_{i\neq j,l} w_{ii}\frac{\xi(T_i, Z_i)}{\bar{F}(T_j, Z_i)h(T_j, Z_i)}K_b(Z_i - Z_l)$$

and

$$\zeta(t, z) = \frac{\int_{\mathcal{T}} \xi(s, z)h(s, z)ds}{\bar{F}(t, z)h(t, z)}, \quad (t, z) \in \mathcal{D}_h. \tag{1.4.66}$$

where $\xi(t, z)$ is defined (1.4.60). Then, by the definition of $V_n^{(j,i)}(T_j, Z_i)$, $1 \leq i, j \leq n$, (See (1.4.14)), and change of summations,

$$C_{n1} = -\frac{1}{\sqrt{n}}\sum_{l=1}^{n}[\delta_l - F(T_l, Z_l)]\frac{1}{n}\sum_{j\neq l} w_{j1}\Lambda(T_j, \theta_0)K_b(T_j - T_l)\zeta_n(T_j, Z_l) + o_p(1).$$

Let $w_{ll}(2b) = 1$ if $T_l \in [t_1^* + 2b, t_2^* - 2b]$ and $Z_l \in [z_1^* + 2b, z_2^* - 2b]$, and 0 otherwise, $1 \le l \le n$. Now we write the main part of C_{n1} as the sum of two parts according to whether $w_{ll}(2b) = 1$ or 0. The reason for doing this is because of the edge effect of the kernel estimation. Write

$$C_{n1} = \tilde{C}_{n1} + C_{n1}^0 + o_p(1), \tag{1.4.67}$$

where

$$\tilde{C}_{n1} = -\frac{1}{\sqrt{n}} \sum_{l=1}^{n} w_{ll}(2b)[\delta_l - F(T_l, Z_l)] \frac{1}{n} \sum_{j \neq l} w_{j1} \Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta_n(T_j, Z_l),$$

and

$$C_{n1}^0 = -\frac{1}{\sqrt{n}} \sum_{l=1}^{n} (1 - w_{ll}(2b))[\delta_l - F(T_l, Z_l)] \frac{1}{n} \sum_{j \neq l} w_{j1} \Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta_n(T_j, Z_l)$$

Since conditioned on $(T_i, Z_i), i = 1, \cdots, n$, $\delta_l - F(T_l, Z_l, \theta_0)$ and $\delta_k - F(T_k, Z_k, \theta_0)$ for $l \neq k$ are independent with mean zero and variances $F(T_l, Z_l, \theta_0)[1 - F(T_l, Z_l, \theta_0)]$ and $F(T_k, Z_k, \theta_0)[1 - F(T_k, Z_k, \theta_0)]$ respectively, by taking conditional expectation first, we can see that

$$\begin{aligned} E(C_{n1}^0)^2 = \frac{1}{n} \sum_{l=1}^{n} E\Big[&(1 - w_{ll}(2b)) F(T_l, Z_l, \theta_0)(1 - F(T_l, Z_l, \theta_0)) \\ &\times \Big(\frac{1}{n} \sum_{j \neq l} w_{j1} \Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta_n(T_j, Z_l)\Big)^2\Big] \end{aligned}$$

which is $o_p(1)$ because

$$\begin{aligned} &\frac{1}{n} \sum_{j \neq l} E|\Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta_n(T_j, Z_l)| \\ &\le \frac{1}{n} \sum_{j \neq l} \frac{1}{n} \sum_{i \neq j,l} E\Big| \frac{\Lambda(T_j, \theta_0) \xi(T_i, Z_i)}{\bar{F}(T_j, Z_i, \theta_0) h(T_j, Z_i)} K_b(Z_i - Z_l) K_b(T_j - T_l) \Big| = O(1) \end{aligned}$$

and $E(1 - w_{ll}(2b)) = o(1)$. Here we use the boundedness of $\frac{\Lambda(t, \theta_0) \xi(t, z)}{\bar{F}(t, z, \theta_0) h(t, z)}$ and that $sup_{(T_l, Z_l) \in \mathcal{D}_h^0} E_l |K_b(Z_i - Z_l) K_b(T_j - T_l)| < \infty$. Therefore, we obtain

$$C_{n1}^0 = o_p(1). \tag{1.4.68}$$

Write

$$\tilde{C}_{n1} = \tilde{C}_{n11} + \tilde{C}_{n12}, \tag{1.4.69}$$

where

$$\tilde{C}_{n11} = -\frac{1}{\sqrt{n}}\sum_{l=1}^{n} w_{ll}(2b)[\delta_l - F(T_l, Z_l, \theta_0)]\frac{1}{n}\sum_{j\neq l} w_{j1}\Lambda(T_j, \theta_0)K_b(T_j - T_l)\zeta(T_j, Z_l)$$

and

$$\tilde{C}_{n12} = -\frac{1}{\sqrt{n}}\sum_{l=1}^{n} w_{ll}(2b)[\delta_l - F(T_l, Z_l, \theta_0)]$$
$$\times \frac{1}{n}\sum_{j\neq l} w_{j1}\Lambda(T_j, \theta_0)K_b(T_j - T_l)[\zeta_n(T_j, Z_l) - \zeta(T_j, Z_l)].$$

Note that $\zeta(t, z)$ is four times differentiable under the assumptions. Apply (1.4.7) of Lemma 2 and part (1) of Lemma 3 to $\zeta_n(T_j, Z_l)$ with $d = 1$ and $r = 4$ to obtain

$$\sup_{Z_l\in[z_1^*+2b, z_2^*-2b], T_j\in\mathcal{T}^0} E_{j,l}|\zeta_n(T_j, Z_l) - \zeta(T_j, Z_l)|^2 = O_p(\frac{1}{nb}) + O_p(b^8). \tag{1.4.70}$$

Because of the conditional independence of $\delta_l - F(T_l, Z_l, \theta_0)$ and $\delta_k - F(T_k, Z_k, \theta_0)$ for $l \neq k$ with mean zero and variance $F(T_l, Z_l)[1-F(T_l, Z_l)]$ and $F(T_k, Z_k)[1-F(T_k, Z_k)]$ respectively when $(T_i, Z_i), i = 1, \cdots, n$ are given, as before, we have

$$E(\tilde{C}_{n12})^2 = \frac{1}{n}\sum_{l=1}^{n} E\Big\{w_{ll}(2b)F(T_l, Z_l, \theta_0)[1 - F(T_l, Z_l, \theta_0)]$$
$$\times \Big[\frac{1}{n}\sum_{j\neq l} w_{j1}\Lambda(T_j, \theta_0)K_b(T_j - T_l)(\zeta_n(T_j, Z_l) - \zeta_n(T_j, Z_l))\Big]^2\Big\}$$
$$\leq \frac{1}{n}\sum_{l=1}^{n} E\Big[w_{ll}(2b)F(T_l, Z_l, \theta_0)(1 - F(T_l, Z_l, \theta_0))$$
$$\times \frac{1}{n}\sum_{j\neq l} w_{j1}\Lambda^2(T_j, \theta_0)K_b^2(T_j - T_l)\frac{1}{n}\sum_{j\neq l}(\zeta_n(T_j, Z_l) - \zeta_n(T_j, Z_l))^2\Big].$$

The last inequality follows from Cauchy-Schwartz inequality. Since K is bounded, $|K_b^2| \leq C/b^2$ for some finite number C. This, the boundedness of $F(T_l, Z_l, \theta_0)$ and $\Lambda(T_l, \theta_0)$, and (1.4.70), imply that

$$E(\tilde{C}_{n12})^2 = O_p(\frac{1}{nb^3}) + O_p(b^6),$$

which is $o_p(1)$ as $b = O(n^{-a})$ with $\frac{1}{8} < a < \frac{1}{4}$ (See assumption (A4)). Therefore, $\tilde{C}_{n12} = o_p(1)$ and it follows from (1.4.69) that

$$\tilde{C}_{n1} = -\frac{1}{\sqrt{n}} \sum_{l=1}^{n} w_{ll}(2b)[\delta_l - F(T_l, Z_l)] \frac{1}{n} \sum_{j \neq l} w_{j1} \Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta(T_j, Z_l)$$
$$+ o_p(1).$$

Let

$$\eta_n(T_l, Z_l) = \frac{1}{n} \sum_{j \neq l} w_{j1} \Lambda(T_j, \theta_0) K_b(T_j - T_l) \zeta(T_j, Z_l), \quad 1 \leq l \leq n,$$

and

$$\eta(t, z) = \Lambda(t, \theta_0) \zeta(t, z) h_1(t), \quad (t, z) \in \mathcal{D}_h. \tag{1.4.71}$$

where $\zeta(t, z)$ is defined in (1.4.66), and $h_1(t)$ is the marginal distribution of T as defined before. Similarly, we can obtain

$$\tilde{C}_{n1} = -\frac{1}{\sqrt{n}} \sum_{l=1}^{n} w_{ll}(2b)[\delta_l - F(T_l, Z_l)] \eta(T_l, Z_l) + o_p(1).$$

Since $P(w_{ll}(2b) = 0) = O(b)$, $\eta(t, z)$ is bounded, and the conditional expectation of δ_l, given (T_l, Z_l), is 0, it is easy to see that

$$\tilde{C}_{n1} = -\frac{1}{\sqrt{n}} \sum_{l=1}^{n} [\delta_l - F(T_l, Z_l)] \eta(T_l, Z_l) + o_p(1). \tag{1.4.72}$$

Since $Q_{n1} = C_{n1}/C_{n0}$ and $C_{n0} - c_0 = o_p(1)$, it follows from (1.4.67), (1.4.68) and (1.4.72) that

$$Q_{n1} = -\frac{1}{c_0} \frac{1}{\sqrt{n}} \sum_{l=1}^{n} [\delta_l - F(T_l, Z_l)] \eta(T_l, Z_l) + o_p(1). \tag{1.4.73}$$

Now we deal with Q_{n2}. Let $C_{n2} = Q_{n2} C_{n0}$. If follows from (1.4.65) and (1.4.63) that

$$C_{n2} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_{ii} \xi(T_i, Z_i) \frac{1}{n} \sum_{j \neq i} w_{j1} \frac{\Lambda(T_j, \theta_0)}{\bar{F}(T_j, Z_i, \theta_0) B_{n0}^{(j,i)}(T_j, Z_i)} B_n^{(j,i)}(T_j, Z_i).$$

As $b = O(n^{-a})$ with $\frac{1}{16} < a < \frac{1}{4}$ (see assumption (A1)), (1.4.20) and (1.4.21) of Lemma 6 and the same arguments as we were dealing with C_{n1} lead to

$$C_{n2} = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j, \theta_0)}{\bar{F}(T_j, Z_i, \theta_0)h(T_j, Z_i)}B_n^{(j,i)}(T_j, Z_i) + o_p(1).$$

That is, $B_{n0}^{(j,i)}(T_j, Z_i)$ can be replaced by $h(T_j, Z_i)$ with a small difference $o_p(1)$. By the definition of $B_n^{(j,i)}(T_j, Z_i)$ (see (1.4.15)),

$$\begin{aligned}C_{n2} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_{ii}\xi(T_i, Z_i)\frac{1}{n}\sum_{j\neq i} w_{j1}\frac{\Lambda(T_j, \theta_0)}{\bar{F}(T_j, Z_i, \theta_0)h(T_j, Z_i)} \\ &\qquad \times \frac{1}{n}\sum_{l\neq i,j}[F(T_l, Z_l, \theta_0) - F(T_j, Z_i, \theta_0)]K_b(T_l - T_j)K_b(Z_l - Z_i) + o_p(1) \\ &:= C_{n20} + o_p(1), \quad \text{say}\end{aligned}$$

Note that

$$\begin{aligned}(C_{n20})^2 = \frac{1}{n^5}&\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}\sum_{l_1\neq i_1,j_1} w_{i_1i_1}w_{j_11}\frac{\xi(T_{i_1}, Z_{i_1})\Lambda(T_{j_1}, \theta_0)}{\bar{F}(T_{j_1}, Z_{i_1}, \theta_0)h(T_{j_1}, Z_{i_1})} \\ &\times [F(T_{l_1}, Z_{l_1}, \theta_0) - F(T_{j_1}, Z_{i_1}, \theta_0)]K_b(T_{l_1} - T_{j_1})K_b(Z_{l_1} - Z_{i_1}) \\ &\times \sum_{i_2=1}^{n}\sum_{j_2\neq i_2}\sum_{l_2\neq i_2,j_2} w_{i_2i_2}w_{j_21}\frac{\xi(T_{i_2}, Z_{i_2})\Lambda(T_{j_2}, \theta_0)}{\bar{F}(T_{j_2}, Z_{i_2}, \theta_0)h(T_{j_2}, Z_{i_2})} \\ &\times [F(T_{l_2}, Z_{l_2}, \theta_0) - F(T_{j_2}, Z_{i_2}, \theta_0)]K_b(T_{l_2} - T_{j_2})K_b(Z_{l_2} - Z_{i_2}).\end{aligned}$$

Since

$$\begin{aligned}&E\{[F(T, Z, \theta_0) - F(t, z, \theta_0)]K_b(T - t)K_b(Z - z)\} \\ &\quad = \int_0^1\int_0^1 [F(t + bu, z + bv, \theta_0) - F(t, z, \theta_0)]K(u)K(v)h(t + bu, z + bv)dudv \\ &\quad = O(b^4),\end{aligned}$$

uniformly in $(t, z) \in \mathcal{D}_h^0$, for the terms with $l_1 \neq l_2$, conditioned on $T_{j_1}, Z_{i_1}, T_{j_2}, Z_{i_2}$, its expectation is of order $O(b^8)$ uniformly in j_1, j_2, i_1, i_2 and hence the sum of the expectations of these terms is of order $O(nb^8)$ which is $o(1)$ if $nb^8 \to 0$. Note also that

$$E\left[w_{ii}|F(t,z,\theta_0) - F(T_j,Z_i,\theta_0)]K_b(t-T_j)K_b(z-Z_i)|\right]$$
$$\leq \int_0^1\int_0^1 |F(t+bu,z+bv,\theta_0) - F(t,z,\theta_0)||K(u)K(v)h(t+bu,z+bv|dudv$$
$$= O(b).$$

For those terms with $l_1 = l_2$ and $i_1 \neq i_2, j_1 \neq j_2$, conditioned on T_{l_1}, Z_{l_1}, its expectation is of order $O(b)$ uniformly in T_{l_1}, Z_{l_1}, hence the order of the sum of the expectations of these terms is also of order $O(b)$ since there are 5 summations. This order is also $o(1)$.

Similarly, the sum of the expectations of the other terms is of order $o(1)$. Therefore,

$$E(C_{n20})^2 = o(1)$$

and hence $C_{n20} = o_p(1)$. Therefore, C_{n2} and hence Q_{n2} is $o_p(1)$. This, (1.4.64) and (1.4.73) imply that

$$Q_n = -\frac{1}{c_0}\frac{1}{\sqrt{n}}\sum_{l=1}^{n}[\delta_l - F(T_l,Z_l,\theta_0)]\eta(T_l,Z_l) + o_p(1). \tag{1.4.74}$$

If follows from (1.4.48), (1.4.57) and (1.4.74) that, on A_n,

$$S_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{l=1}^{n}[\delta_l - F(T_l,Z_l,\theta_0)]\left[\frac{D_1(T_l,Z_l,\theta_0)}{F(T_l,Z_l,\theta_0)} - \frac{\eta(T_l,Z_l)}{c_0}\right] + o_p(1). \tag{1.4.75}$$

This and (1.4.47) imply that, on A_n,

$$\sqrt{n}(\hat{\theta}-\theta) = d^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{l=1}^{n}[\delta_l - F(T_l,Z_l,\theta_0)]\left[\frac{D_1(T_l,Z_l,\theta_0)}{F(T_l,Z_l,\theta_0)} - \frac{\eta(T_l,Z_l)}{c_0}\right] + o_p(1),$$

where $\eta(t,z)$ is defined in (1.4.71) and

$$d(\theta_0) = E\left[\frac{\bar{F}(T,Z,\theta_0)}{F(T,Z,\theta_0)}D_1^2(T,Z,\theta_0)\right].$$

Since $D_1(t,z,\theta_0)$ and $\eta(t,z)$ are bounded and $F(t,z,\theta_0)$ is bounded away from 0 on $\mathcal{D}_h$, the theorem follows from the central limit theorem and Lemma 9.

Chapter 2

Sieve Estimation

2.1 Estimation

The second approach uses the idea of sieve and is analogous to that of Rossini and Tsiatis (1996).

The goal of this chapter is to estimate θ efficiently, with $\alpha(z) = log(g(z))$ as an infinite dimensional nuisance parameter. The rescaled (conditional) log-likelihood of θ and α based on $(T_i, \delta_i, Z_i), i = 1, 2, \dots, n$ is

$$\begin{aligned} L_n(\theta, \alpha) &= \frac{1}{n} \sum_{i=1}^{n} \left[\delta_i log F(T_i, Z_i, \theta, \alpha) + (1 - \delta_i) log \bar{F}(T_i, Z_i, \theta, \alpha)\right] \\ &= \frac{1}{n} \sum_{i=1}^{n} \left[\delta_i log(1 - e^{-\Lambda(T_i, \theta) e^{\alpha(Z_i)}}) - (1 - \delta_i)\Lambda(T_i, \theta) e^{\alpha(Z_i)}\right] \end{aligned} \tag{2.1.1}$$

Here

$$F(t, z, \theta, \alpha) = 1 - e^{-\Lambda(t, \theta) e^{\alpha(z)}}, \quad \bar{F}(t, z, \theta, \alpha) = 1 - F(t, z, \theta, \alpha). \tag{2.1.2}$$

To maximize the log-likelihood over all possible θ and α, we should set $\alpha(Z_i)$ to be positive infinite if $\delta_i = 1$, and negative infinite if $\delta_i = 0$. Hence the maximum likelihood estimator over all possible functions α does not exist. The log-likelihood function is maximized as α varies over a small set of functions which depends on the sample size. More specifically, we approximate α by a step function with known jump points and maximize the log-likelihood as α varies over the step functions. As the number of steps increases along with the sample size, the bias from the approximation disappears. Assume that the covariate lies in a bounded interval. Without loss of generality, it will be taken to be an interval $[0, 1]$. To construct the step function, define a partition $0 = z_0 < z_1 < \cdots < z_k = 1$, where k depends on n and increases with the increase of n. The step function is then defined as

$$\alpha_n(z) = \sum_{j=1}^{k} \alpha_{nj} I_j(z), \tag{2.1.3}$$

where $I_j(z)$ is the indicator function for the jth interval, defined by $I_j(z) = 1$ if $z_{j-1} < z \leq z_j$ and zero otherwise. For the fixed partition, the step function is completely specified by the parameters $(\alpha_{n1}, \cdots, \alpha_{nk})$. Hence, from here on, α_n will denote either the function α_n given by (2.1.3) or, equivalently, the vector α, depending on the context.

The estimate $(\hat{\theta}, \hat{\alpha}_n)$ is obtained by maximizing the approximate likelihood formed by substituting (2.1.3) for α in (2.1.1). Since k is an increasing integer-valued function of n, written as $k(n)$, α_n will tend to α. The next two sections show that when $k(n) = O(n^{\gamma})$ with $\frac{1}{4} < \gamma < \frac{1}{2}$, $(\hat{\theta}, \hat{\alpha}_n)$ is consistent and $\hat{\theta}$ is also asymptotically normal.

The first and second partial derivatives of the approximate log-likelihood are used to generate the estimates and their variance. In view of (2.1.1), the first derivative with respect to θ is

$$S_{n,0}(\theta, \alpha_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{[\delta_i - F(T_i, Z_i, \theta, \alpha_n)]}{F(T_i, Z_i, \theta, \alpha_n)} \dot{\Lambda}(T_i, \theta) e^{\alpha_n(Z_i)}, \tag{2.1.4}$$

and that with respect to α_{nj} is

$$S_{n,j}(\theta, \alpha_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{[\delta_i - F(T_i, Z_i, \theta, \alpha_n)]}{F(T_i, Z_i, \theta, \alpha_n)} \Lambda(T_i, \theta) e^{\alpha_n(Z_i)} I_j(Z_i), \quad j = 1, \cdots, k, \tag{2.1.5}$$

where $\dot{\Lambda}(t,\theta)$ denote the derivative with respect to θ.

The score vector is defined as

$$S_n(\theta,\alpha_n) = \begin{pmatrix} S_{n,0}(\theta,\alpha_n) \\ S_{n,1}(\theta,\alpha_n) \\ \vdots \\ S_{n,k}(\theta,\alpha_n) \end{pmatrix}. \tag{2.1.6}$$

The estimates $(\hat{\theta},\hat{\alpha}_n)$ are defined to be a solution to the score equation

$$S_n(\theta,\alpha_n) = 0. \tag{2.1.7}$$

The derivative of S_n with respect to (θ,α_n) is called the Hessian matrix and related to the observed information. This is defined as

$$H_n(\theta,\alpha_n) = \frac{\partial}{\partial(\theta,\alpha_n)} S_n(\theta,\alpha_n), \tag{2.1.8}$$

which is the $(k+1)$ by $(k+1)$ matrix of partial derivatives with respect to θ and α_n of the elements of $S_n(\theta,\alpha_n)$. Let 0 denote the first element. Then the elements of H_n are defined by

$$\begin{aligned} h_{00}(\theta,\alpha_n) &= \frac{1}{n}\sum_{i=1}^{n} \frac{[\delta_i - F(T_i,Z_i,\theta,\alpha_n)]}{F(T_i,Z_i,\theta,\alpha_n)} \ddot{\Lambda}(T_i,\theta) e^{\alpha_n(Z_i)} \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i D_{00}(T_i,Z_i,\theta,\alpha_n)}{F(T_i,Z_i,\theta,\alpha_n)}, \end{aligned}$$

$$\begin{aligned} h_{0j}(\theta,\alpha_n) &= \frac{1}{n}\sum_{i=1}^{n} \frac{[\delta_i - F(T_i,Z_i,\theta,\alpha_n)]}{F(T_i,Z_i,\theta,\alpha_n)} \dot{\Lambda}(T_i,\theta) e^{\alpha_n(Z_i)} I_j(Z_i) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i D_{01}(T_i,Z_i,\theta,\alpha_n) I_j(Z_i)}{F(T_i,Z_i,\theta,\alpha_n)}, \quad j = 1,\cdots,k, \end{aligned}$$

$$h_{j0}(\theta,\alpha_n) = h_{0j}(\theta,\alpha_n), \quad j = 1,\cdots,k,$$

$$\begin{aligned} h_{jj}(\theta,\alpha_n) &= \frac{1}{n}\sum_{i=1}^{n} \frac{[\delta_i - F(T_i,Z_i,\theta,\alpha_n)]}{F(T_i,Z_i,\theta,\alpha_n)} \Lambda(T_i,\theta) e^{\alpha_n(Z_i)} I_j(Z_i) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i D_{11}(T_i,Z_i,\theta,\alpha_n) I_j(Z_i)}{F(T_i,Z_i,\theta,\alpha_n)}, \quad j = 1,\cdots,k, \end{aligned}$$

and

$$h_{ij}(\theta, \alpha_n) = 0, \quad i \neq j = 1, \cdots, k,$$

where

$$D_{00}(t, z, \theta, \alpha_n) = \frac{\bar{F}(t, z, \theta, \alpha_n)}{F(t, z, \theta, \alpha_n)} \dot{\Lambda}^2(t, \theta) e^{2\alpha_n(z)}, \tag{2.1.9}$$

$$D_{01}(t, z, \theta, \alpha_n) = \frac{\bar{F}(t, z, \theta, \alpha_n)}{F(t, z, \theta, \alpha_n)} \dot{\Lambda}(t, \theta) \Lambda(t, \theta) e^{2\alpha_n(z)}, \tag{2.1.10}$$

$$D_{11}(t, z, \theta, \alpha_n) = \frac{\bar{F}(t, z, \theta, \alpha_n)}{F(t, z, \theta, \alpha_n)} \Lambda^2(t, \theta) e^{2\alpha_n(z)}, \tag{2.1.11}$$

and $\ddot{\Lambda}(t, \theta)$ is the second derivative with respect to θ.

Expectation is taken with respect to the true parameters (θ_0, α_0).

2.2 Consistency

In order to have the consistency and asymptotic normality of the estimator, we use some assumptions. We call the following assumptions Condition A.

(1) The real parameter θ_0 is an interior point of Θ.

(2) Let $\mathcal{T}$ and $\mathcal{Z}$ be the supports of T and Z respectively, where $\mathcal{Z}$ is a closed interval of $\mathcal{R}^1$. $\Lambda(t, \theta)$ is bounded away from 0 and ∞ over $(t, \theta) \in \mathcal{T} \times \mathcal{N}_1$, where $\mathcal{N}_1 = \{\theta : |\theta - \theta_0| \leq \Delta\}$ for some $0 < \Delta < \infty$. The density of (T, Z), $h(t, z)$, is bounded on $\mathcal{T} \times \mathcal{Z}$, Lipschitz continuous in z uniformly for $t \in \mathcal{T}$.

(3) The first and second derivatives of $\Lambda(t, \theta)$ with respect to θ, $\dot{\Lambda}(t, \theta)$ and $\ddot{\Lambda}(t, \theta)$, exist, are bounded for $t \in \mathcal{D}_T$ and $\theta \in \mathcal{N}_1$, and continuous in θ for any fixed t;

(4) $\alpha_0(z)$ is Lipschitz continuous on $\mathcal{Z}$.

For any function $b(z)$ defined on the support of Z, let $||b||_\infty = \sup_{z \in \mathcal{Z}} |b(z)|$ and $||b|| = \sqrt{E(b(Z))^2}$ be sup-norm and L_2-norm respectively.

In the following, Theorem 3 states the existence of one consistent (in sup-norm) estimator, $\hat{\theta}$, which is a solution to the score equation. Theorem 4 establishes the convergence rate of the estimator (in L_2 norm), which will be used to prove the asymptotic normality of the estimator. The proof of them will be given later.

Theorem 3 *Assume that Condition A holds, and the number of intervals is increasing at a rate* $k(n) = n^{\gamma}$, *with* $0 < \gamma < 1$. *Assume also that for all* k *and* α_{0n} *with* $\|\alpha_{0n} - \alpha_0\|_\infty < \Delta_0$ *for some positive and finite number* Δ_0,

$$P(I_j(Z) = 1) = o(1), \quad kP(I_j(Z) = 1) > c, \quad j = 1, 2, \cdots, k, \tag{2.2.1}$$

and

$$E\left[D_{00}(T, Z, \theta_0, \alpha_{0n})\right] - \sum_{j=1}^{k} \frac{E\left(\frac{F(T,Z,\theta_0,\alpha_0)}{F(T,Z,\theta_0,\alpha_{0n})} D_{01}(T, Z, \theta_0, \alpha_{0n}) I_j(Z)\right)^2}{E\left(\frac{F(T,Z,\theta_0,\alpha_0)}{F(T,Z,\theta_0,\alpha_{0n})} D_{11}(T, Z, \theta_0, \alpha_{0n}) I_j(Z)\right)} > c, \tag{2.2.2}$$

for some $0 < c < \infty$, *not depending on* n. *Then there is at least one consistent (in sup-norm) solution to* (2.1.7), *i.e. there exists at least one* $(\hat{\theta}, \hat{\alpha}_n)$ *such that*

$$|\hat{\theta} - \theta_0| + \|\hat{\alpha}_n - \alpha_0\|_\infty = o_p(1).$$

The proof is given in Section 2.6.

Theorem 4 *Assume that the conditions in Theorem 3 holds. Assume also* $k(n) = n^{\gamma}$, *with* $\frac{1}{4} < \gamma < \frac{1}{2}$, *and*

$$E\left[D_{00}(T, Z, \theta_0, \alpha_0)\right] - \frac{E\left(D_{01}(T, Z, \theta_0, \alpha_0)\right)^2}{E\left(D_{11}(T, Z, \theta_0, \alpha_0)\right)} > 0. \tag{2.2.3}$$

Then the estimator $(\hat{\theta}, \hat{\alpha}_n)$ *in Theorem 3 has the following convergence rate*

$$|\hat{\theta} - \theta_0| = o_p(n^{-\frac{1}{4}}), \quad \|\hat{\alpha}_n - \alpha_0\| = o_p(n^{-\frac{1}{4}}).$$

The proof is given in Section 2.6.

2.3 Asymptotic normality of $\hat{\theta}$

In this section, the asymptotic normality of the estimator is stated and the proof will be given later.

Theorem 5 *Assume that the conditions in Theorem 4 hold, and σ^2 defined below is finite. Assume also that the third derivative of $\Lambda(t,\theta)$ with respect to θ exists for θ in a neighborhood of θ_0, and is continuous at θ_0. Then*

$$\sqrt{n}(\hat{\theta}-\theta_0) \to N(0,\sigma^2),$$

where the asymptotic variance is given by

$$\sigma^2 = \left[E(D_{00}(T,Z,\theta_0,\alpha_0)) - E\left(\frac{(E(D_{01}(T,Z,\theta_0,\alpha_0)|Z))^2}{E(D_{11}(T,Z,\theta_0,\alpha_0)|Z)}\right)\right]^{-1}. \tag{2.3.1}$$

The proof is given in Section 2.6.

2.4 Information bound for θ_0

The true model has two parameters: θ is finite dimensional, and α is an infinite-dimensional functional parameter. The semiparametric information bound for estimating θ is based on the maximum of the asymptotic variance bounds of regular estimators for θ obtained using parametric sub-models of α. It was shown in Section 2.3 that the estimator $\hat{\theta}$ is asymptotically normal with a certain asymptotic variance. It is shown in this section that this asymptotic variance achieves the bound. Projection methods are used to find the efficient score for the semiparametric model and hence the variance bound (Bickel et al. 1993).

The log-likelihood of θ and α based on (T,δ,Z) is given by

$$\delta log(1-e^{-\Lambda(T,\theta)e^{\alpha(Z)}}) - (1-\delta)\Lambda(T,\theta)e^{\alpha(Z)}. \tag{2.4.1}$$

Consider a general parametric submodel with $\alpha = \alpha_\gamma$, specified by γ (a real variable), where $\frac{\partial}{\partial\gamma}\alpha_\gamma(z)|_{\gamma=0} = a(z)$ for some function $a(z)$ with $Ea^2(Z) < \infty$. Take derivatives of (2.4.1) with respect to θ and γ at $(\theta = \theta_0, \gamma = 0)$ to obtain the scores

$$S_0(T, Z, \delta, \theta_0, \alpha_0) = [\delta - F(T, Z, \theta_0, \alpha_0)]\,\frac{\dot{\Lambda}(T, \theta_0)e^{\alpha_0(Z)}}{F(T, Z, \theta_0, \alpha_0)} \tag{2.4.2}$$

and

$$S_a(T, Z, \delta, \theta_0, \alpha_0) = [\delta - F(T, Z, \theta_0, \alpha_0)]\,\frac{\Lambda(T, \theta_0)e^{\alpha_0(Z)}a(Z)}{F(T, Z, \theta_0, \alpha_0)}. \tag{2.4.3}$$

To find the information bound, project S_0 to the linear span formed from all square integrable S_a. This projection is denoted by S_{a^*} and is computed by solving for all S_a,

$$E(S_0 S_a) = E(S_{a^*} S_a). \tag{2.4.4}$$

Note that the conditional expectation and variance of δ given (T, Z) is $F(T, Z, \theta_0, \alpha_0)$ and $F(T, Z, \theta_0, \alpha_0)\bar{F}(T, Z, \theta_0, \alpha_0)$ respectively. Substituting (2.4.2), (2.4.3) for S_0, S_a in the above expression, taking conditional expectation, given (T, Z) first, and then taking expectation with respect to (T, Z), we obtain

$$E(D_{01}(T, Z, \theta_0, \alpha_0)a(Z)) = E(D_{11}(T, Z, \theta_0, \alpha_0)a^*(Z)a(Z)),$$

where D_{01} and D_{11} were defined in (2.1.10) and (2.1.11) respectively. Take conditional expectation, given Z first, and then expectation with respect to Z to obtain

$$E[E(D_{01}(T, Z, \theta_0, \alpha_0)|Z)a(Z)] = E[E(D_{11}(T, Z, \theta_0, \alpha_0)|Z)a^*(Z)a(Z)]. \tag{2.4.5}$$

It is easy to see that

$$a^*(Z) = \frac{E(D_{01}(T, Z, \theta_0, \alpha_0)|Z)}{E(D_{11}(T, Z, \theta_0, \alpha_0)|Z)} \tag{2.4.6}$$

solve (2.4.5) and hence also solve (2.4.4).

Therefore, the efficient score is given by

$$\begin{aligned}
&S_0(T, Z, \theta_0, \alpha_0) - S_{a^*}(T, Z, \theta_0, \alpha_0) \\
&\quad = \frac{(\delta - F(T, Z, \theta_0, \alpha_0))e^{\alpha_0(Z)}}{F(T, Z, \theta_0, \alpha_0)} \left(\dot{\Lambda}(T, \theta_0) - \Lambda(T, \theta_0) \frac{E(D_{01}(T, Z, \theta_0, \alpha_0)|Z)}{E(D_{11}(T, Z, \theta_0, \alpha_0)|Z)} \right).
\end{aligned}$$

The semiparametric information bound is equal to

$$E\left[S_0(T, Z, \theta_0, \alpha_0) - S_{a^*}(T, Z, \theta_0, \alpha_0)\right]^2$$

and the asymptotic variance bound is the inverse of the information bound. Take the conditional expectation of the square of the efficient score, given (T, Z) first, and then expectation with respect to (T, Z) to obtain

$$\begin{aligned}
&E\left[S_0(T, Z, \theta_0, \alpha_0) - S_{a^*}(T, Z, \theta_0, \alpha_0)\right]^2 \\
&\quad = E\left[\frac{\bar{F}(T, Z, \theta_0, \alpha_0)}{F(T, Z, \theta_0, \alpha_0)} e^{2\alpha_0(Z)} \left(\dot{\Lambda}(T, \theta_0) - \Lambda(T, \theta_0) \frac{E(D_{01}(T, Z, \theta_0, \alpha_0)|Z)}{E(D_{11}(T, Z, \theta_0, \alpha_0)|Z)} \right)^2 \right].
\end{aligned}$$

Expand the square term and take the conditional expectation given Z first to obtain that the right hand side of the previous display is equal to

$$E\left[D_{00}(T, Z, \theta_0, \alpha_0) - \frac{(E(D_{01}(T, Z, \theta_0, \alpha_0)|Z))^2}{E(D_{11}(T, Z, \theta_0, \alpha_0)|Z)} \right].$$

In view of (2.3.1), it follows that the asymptotic variance of $\hat{\theta}$ achieves the asymptotic variance bound.

2.5 Simulation

A simulation study is presented before we go to the proof of the stated asymptotic properties of the estimator.

As in Section 1.3, assume that the conditional distribution of X given Z is a Weibull distribution with distribution function

$$1 - e^{-x^{\theta_0} e^{\alpha_0(Z)}}.$$

where $\alpha_0(z) = log(z)$. Also assume that T and Z are uniformly distributed on $[1, 2]$ and $[0.2, 1.2]$ respectively.

For each fixed sample size (n=30, 60, 100, 200, 500, 1000 respectively) and appropriate k's, 100 samples are generated with the real parameter $\theta_0 = 1.5$ and 100 replications of the estimate of θ_0 based on the sieve maximum likelihood estimator (SMLE) are obtained. The means and standard deviations of these estimates are shown in the following table.

Table 2. Simulation results for the SMLE

n	k	mean	s.d.
30	1	1.6976	1.8180
	2	2.1060	2.0269
	4	2.4360	2.9598
60	4	1.7064	1.2145
	6	1.8189	1.2680
	8	1.9675	1.4248
100	4	1.5954	0.8047
	6	1.6427	0.8103
	8	1.6932	0.8502
200	6	1.5624	0.5154
	8	1.5838	0.5330
	10	1.6240	0.5365
500	6	1.5591	0.2946
	8	1.5671	0.2893
	15	1.6076	0.2964
1000	10	1.5432	0.2136
	15	1.5530	0.2125
	20	1.5651	0.2177

The above table shows that when the sample size is not large, the bias and variance are slightly larger than those of the generalized profile maximum likelihood estimator (see Table 1). However, they decrease with the increase of the sample size, and the variance will be eventually less than that of the generalized profile maximum likelihood estimator since it achieves the semiparametric lower bound. Unfortunately, a very large sample size is needed for this to happen. This can be seen when we compare the above table with the simulation results for the generalized profile maximum likelihood estimator in Section 1.3.

2.6 Proof of the theorems

2.6.1 Proof of Theorem 3

The definitions of sup-norms for a vector and a matrix are introduced first. If a is a vector with elements a_j, $1 \leq j \leq m$, then

$$\|a\|_\infty = \max_{1\leq j\leq m} |a_j|.$$

If A is an m by m matrix whose (i,j) element is denoted by a_{ij}, then

$$\|A\|_\infty = \max_{1\leq i\leq m} \left(\sum_{j=1}^{m} |a_{ij}| \right).$$

Now define a step function, α_{0n}, of form (2.1.3) as an approximation to α_0. Precisely,

$$\alpha_{0n}(z) = \sum_{j=1}^{k(n)} \alpha_0(z_j) I_j(z).$$

The Lipschitz continuity of α_0 implies that

$$\|\alpha_{0n} - \alpha_0\|_\infty = O(k(n)^{-1}). \tag{2.6.1}$$

Let

$$\beta_n = (\theta, \alpha_{n1}, \cdots, \alpha_{nk}), \quad \beta_{0n} = (\theta_0, \alpha_0(z_1), \cdots, \alpha_0(z_k)),$$

and

$$\beta_0 = (\theta_0, \alpha_0).$$

Note that $S_n(\beta_n) = 0$ is equivalent to

$$\tilde{S}_n(\beta_n) := \begin{pmatrix} S_{n,0}(\beta_n) \\ kS_{n,1}(\beta_n) \\ \vdots \\ kS_{n,k}(\beta_n) \end{pmatrix} = 0. \tag{2.6.2}$$

The derivative of $\tilde{S}_n(\beta_n)$ with respect to β_n is

$$\tilde{H}_n(\beta_n) := \begin{pmatrix} h_{00}(\beta_n) & h_{01}(\beta_n) & \cdots & h_{0k}(\beta_n) \\ kh_{01}(\beta_n) & kh_{11}(\beta_n) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ kh_{0k}(\beta_n) & 0 & 0 & kh_{kk}(\beta_n) \end{pmatrix}, \tag{2.6.3}$$

where h_{ij} is defined in Section 2.1. The low-right k by k sub-matrix is a diagonal matrix.

Let $\tilde{S}(\beta_n) = E\tilde{S}_n(\beta_n)$ (expectations for all the elements). Then, by (2.6.2), (2.1.4), (2.1.5) and the fact that the conditional expectation of δ_i given (T_i, Z_i) is $F(T_i, Z_i, \theta_0, \alpha_0)$, we obtain

$$\tilde{S}(\beta_n) = \begin{pmatrix} E(A(T, Z, \beta_0, \beta_n)\dot{\Lambda}(T, \theta)) \\ kE(A(T, Z, \beta_0, \beta_n)\Lambda(T, \theta)I_1(Z)) \\ \vdots \\ kE(A(T, Z, \beta_0, \beta_n)\Lambda(T, \theta)I_k(Z)) \end{pmatrix}, \tag{2.6.4}$$

where

$$A(T, Z, \beta_0, \beta_n) = e^{\alpha_n(Z)} \left[\frac{F(T, Z, \theta_0, \alpha_0)}{F(T, Z, \theta, \alpha_n)} - 1 \right].$$

By (2.1.2) and Assumption (2), (3) of Condition A, $F(t, z, \theta, \alpha)$ is Lipschitz in θ, α, uniformly for $(t, z) \in \mathcal{T} \times \mathcal{Z}$. It is easy to see that $\|\tilde{S}(\beta_n)\|_\infty = o(1)$ if $\|\beta_n - \beta_{0n}\|_\infty = O(k^{-1})$ and $P(I_j(Z) = 1) = o(1)$ for $j = 1, \cdots, k$.

Let $\tilde{H}(\beta_n) = E\tilde{H}_n(\beta_n)$. Similarly, by (2.6.3) and the definition of $h_{ij}(\beta_n)$ (see Section 2.1), $0 \le i, j \le k$,

$$\tilde{H}(\beta_n) = \begin{pmatrix} b_{00}(\beta_n) & b_{01}(\beta_n) & \cdots & b_{0k}(\beta_n) \\ kb_{01}(\beta_n) & kb_{11}(\beta_n) & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ kb_{0k}(\beta_n) & 0 & \cdots & kb_{kk}(\beta_n) \end{pmatrix},$$

where

$$b_{00}(\beta_n) = E\left[(R(T,Z,\beta_0,\beta_n)-1)\ddot{\Lambda}(T,\theta)e^{\alpha_n(Z)}\right] - E(R(T,Z,\beta_0,\beta_n)D_{00}(T,Z,\theta,\alpha_n)),$$

$$\begin{aligned} b_{0j}(\beta_n) =& E\left[(R(T,Z,\beta_0,\beta_n)-1)\dot{\Lambda}(T,\theta)e^{\alpha_n(Z)}I_j(Z)\right] \\ & - E(R(T,Z,\beta_0,\beta_n)D_{01}(T,Z,\theta,\alpha_n)I_j(Z)), \quad j=1,2,\cdots,k, \end{aligned}$$

$$\begin{aligned} b_{jj}(\beta_n) =& E\left[(R(T,Z,\beta_0,\beta_n)-1)\Lambda(T,\theta)e^{\alpha_n(Z)}I_j(Z)\right] \\ & - E(R(T,Z,\beta_0,\beta_n)D_{11}(T,Z,\theta,\alpha_n)I_j(Z)), \quad j=1,2,\cdots,k, \end{aligned}$$

and

$$R(T,Z,\beta_0,\beta_n) = \frac{F(T,Z,\theta_0,\alpha_0)}{F(T,Z,\theta,\alpha_n)}.$$

Notice that

$$\tilde{H}(\beta_n) = \frac{\partial}{\partial\beta_n}\tilde{S}(\beta_n).$$

The inverse of $\tilde{H}(\beta_n)$ is as follows

$$\tilde{H}^{-1}(\beta_n) = \begin{pmatrix} q_{00} & k^{-1}q_{01} \\ q_{01}' & k^{-1}q_{11} \end{pmatrix}, \tag{2.6.5}$$

where

$$q_{00} = \left(b_{00} - \sum_{j=1}^{k}\frac{b_{0j}^2}{b_{jj}}\right)^{-1},$$

q_{01} is a row vector with its jth element

$$-q_{00}\frac{b_{0j}}{b_{jj}}, \quad j=1,2,\cdots,k,$$

and q_{11} is a $k\times k$ matrix with its (i,j)th element

$$I_{(i=j)}b_{jj}^{-1} + q_{00}^2\frac{b_{0i}b_{0j}}{b_{ii}b_{jj}}, \quad j=1,2,\cdots,k.$$

Since $\tilde{S}(\beta_0)=0$ by (2.6.4), $\tilde{S}(\beta_n)$ in continuous in β_n by Condition A, by (2.6.1),

$$\|\tilde{S}(\beta_{0n})\|_\infty = o(1). \tag{2.6.6}$$

Since $\tilde{H}(\beta_n)$ is continuous in β_n by Condition A, and $\tilde{H}^{-1}(\beta_{0n})$ exists for large n by (2.2.1) and (2.2.2), by (2.6.6), it follows from the inverse function theorem with sup-norm (Lemma 1 of Rossini and Tsiatis (1996), which is stated in the following lemma. For the standard (L_2) formulation of the inverse function theorem, see Rudin (1964)) that there exists $\tilde{\beta}_n = (\tilde{\theta}, \tilde{\alpha}_n)$, with $\tilde{\alpha}_n$ of the form (2.1.3), such that

$$\tilde{S}(\tilde{\beta}_n) = 0$$

and

$$\|\tilde{\beta}_n - \beta_{0n}\|_\infty = o(1). \tag{2.6.7}$$

If $\|\tilde{S}_n(\tilde{\beta}_n)\|_\infty = o_p(1)$ and $\|\tilde{H}_n^{-1}(\tilde{\beta}_n)\|_\infty < c$ with probability approaching 1 for some finite constant c, then by the inverse function theorem again, with probability tending to 1, there exists solution $\hat{\beta}_n = (\hat{\theta}, \hat{\alpha}_n)$ to $\tilde{S}_n(\beta_n) = 0$ such that

$$\|\hat{\beta}_n - \tilde{\beta}_n\|_\infty = o_p(1).$$

This, (2.6.7), (2.6.1) and the triangle inequality imply that

$$\|\hat{\beta}_n - \beta_0\|_\infty = o_p(1).$$

$\|\tilde{S}_n(\tilde{\beta})\|_\infty = o_p(1)$ and $\|\tilde{H}_n^{-1}(\tilde{\beta}_n)\|_\infty < c$ with probability approaching 1 can be established in the same way as the proof of Theorem 1 in Rossini and Tsiatis (1996). The theorem is proved.

Lemma 10 *(Inverse Function Theorem with Sup-norm). Let $H(x)$ be a continuous differentiable mapping from $\mathcal{R}^m$ to $\mathcal{R}^m$ in a neighborhood of x_0. Define the Jacobian as the $m \times m$ matrix $A(x) = \partial H(x)$ (derivatives of the elements of H with respect to the elements of x). If there exists constants C and δ^* such that*

$$\|A^{-1}(x_0)\|_\infty < C$$

and

$$\sup_{\{x:\|x-x_0\|_\infty<\delta^*\}} \|A(x)-A(x_0)\|_\infty \leq (2C)^{-1},$$

then for $d < \delta^*/(4C)$ and all y such that $\|y - H(x_0)\|_\infty < d$, there exists a unique inverse value x in the δ^* neighborhood of x_0 such that $H(x) = y$ and $\|x - x_0\| < 4Cd$.

2.6.2 Proof of Theorem 4

We are going to use some general results on the convergence rate of sieve estimators. The following lemma is a part of Theorem 1 of Shen and Wong (1994). To state the lemma, we introduce some general notations.

Let $Y_1, \cdots, Y_n$ be a sequence of independent random variables (or possibly vectors) distributed according to a density $p_0(y)$ with respect to a σ-finite measure μ on a measurable space $(\mathcal{Y}, \mathcal{B})$ and let Θ be a parameter space of the parameter β. Let $l : \Theta \times \mathcal{Y} \to R$ be a suitably chosen function. We are interested in the properties of an estimate $\hat{\beta}_n$ over a subset Θ_n of Θ by maximizing the empirical criterion $C_n(\beta) = \frac{1}{n}\sum_{i=1}^n l(\beta, Y_i)$, that is, $C_n(\hat{\beta}) = max_{\beta\in\Theta_n} C_n(\beta)$. Here Θ_n is an approximation to Θ in the sense that for any $\beta \in \Theta$, there exists $\pi_n\beta \in \Theta_n$ such that for an appropriate pseudo-distance ρ, $\rho(\pi_n\beta, \beta) \to 0$ as $n \to \infty$. The following assumptions are needed for the lemma.

C0. l is bounded.

C1. For some constants $A_1 > 0$ and $a > 0$, and for all small $\epsilon > 0$,

$$\inf_{\rho(\beta,\beta_0)>\epsilon,\beta\in\Theta_n} E(l(\beta_0, Y) - l(\beta, Y)) \geq 2A_1\epsilon^{2a}.$$

C2. For some constants $A_2 > 0$ and $b > 0$, and for all small $\epsilon > 0$,

$$\inf_{\rho(\beta,\beta_0)\leq\epsilon,\beta\in\Theta_n} Var(l(\beta_0, Y) - l(\beta, Y)) \leq 2A_2\epsilon^{2b}.$$

C3. Let $\mathcal{F}_n = \{l(\beta, \cdot) - l(\pi_n\beta_0) : \beta \in \Theta_n\}$. For some constants $r_0 < \frac{1}{2}$ and $A_3 > 0$,

$$H(\epsilon, \mathcal{F}_n) \leq A_3 n^{2r_0} log\left(\frac{1}{\epsilon}\right) \quad \text{for all small} \quad \epsilon > 0,$$

where $H(\epsilon, \mathcal{F}_n)$ is the L_∞-metric entropy of the space $\mathcal{F}_n$, that is, $exp(H(\epsilon, \mathcal{F}_n))$ is the smallest number of ϵ-balls in the L_∞-metric needed to cover the space $\mathcal{F}_n$.

Lemma 11 *Suppose Assumptions C0 to C3 hold. Then*

$$\rho(\hat{\beta}, \beta_0) = O_p\left(max\left(n^{-\tau}, \rho(\pi_n\beta_0, \beta_0), K^{1/(2a)}(\pi_n\beta_0, \beta_0)\right)\right),$$

where $K(\pi_n\beta_0, \beta_0) = E(l(\beta_0, Y) - l(\pi_n\beta_0, Y))$ *and*

$$\tau = \begin{cases} \frac{1-2r_0}{2a} - \frac{loglogn}{2alogn}, & if \quad b \geq a, \\ \frac{1-2r_0}{4a-2b}, & if \quad b < a. \end{cases}$$

From the proof of Theorem 1 of Shen and Wong (1994), it is noted that the globe maximizer could be replaced by a local maximizer around the real parameter and the convergence rate is still true for the local maximizer. In this situation, the sieve Θ_n is a sequence of shrinking neighborhoods of the real parameter β_0. To apply the above Lemma to our case, let $Y = (T, Z, \delta)$, $\beta = (\theta, \alpha)$, $\pi_n\beta = (\theta, \alpha_n)$ where α_n is of form (2.1.3) with $\alpha_{nj} = \alpha(z_j)$. Also let

$$\Theta_n = \{(\theta, \alpha_n) : |\theta - \theta_0| \leq a_n, ||\alpha_n - \alpha_0||_\infty \leq b_n\},$$

where a_n and b_n are chosen such that, with probability approaching 1, $(\hat{\theta}, \hat{\alpha}_n)$ is the maximum point in Θ_n. Define the metric as follows

$$\rho(\beta, \beta_0) = |\theta - \theta_0| + ||\alpha - \alpha_0||, \tag{2.6.8}$$

and also define

$$l(\beta, Y) = \delta log\left(1 - e^{-\Lambda(T,\theta)e^{\alpha(Z)}}\right) - (1-\delta)\Lambda(T,\theta)e^{\alpha(Z)}.$$

Under our assumptions, C0 is true. Note that

$$\begin{aligned} &El(\beta, Y) \\ &= E\left[\left(1 - e^{-\Lambda(T,\theta_0)e^{\alpha_0(Z)}}\right) log\left(1 - e^{-\Lambda(T,\theta)e^{\alpha(Z)}}\right) - e^{-\Lambda(T,\theta_0)e^{\alpha_0(Z)}}\Lambda(T,\theta)e^{\alpha(Z)}\right]. \end{aligned}$$

Taking Taylor expansion of $l(\beta, Y)$ with respect to θ and α, noticing that the expectation of the first derivative vanishes at β_0 and the matrix of the second derivatives is negative definite by (2.2.3), we obtain

$$E(l(\beta_0, Y) - l(\beta, Y)) \geq c\rho^2(\beta, \beta_0) \tag{2.6.9}$$

for same finite and positive number c. Hence C1 is satisfied with $a = 1$. It is easy to see that, under Condition A,

$$Var(l(\beta_0, Y) - l(\beta, Y)) \leq E(l(\beta_0, Y) - l(\beta, Y))^2 \leq C\rho^2(\beta, \beta_0)$$

for some $0 < C < \infty$. Thus C2 holds with $b = 1$.

Since for all y,

$$|l(\beta, y) - l(\beta_0, y)| \leq C(|\theta - \theta_0| + \|\alpha - \alpha_0\|_\infty),$$

for some $0 < C < \infty$, not depending on y, it is easy to see that

$$H(\epsilon, \mathcal{F}_n) \leq H(\epsilon/C, \Theta_n),$$

where $H(\eta, \Theta_n)$ is the metric entropy of the space Θ_n with respect to the norm $|\theta - \theta_0| + \|\alpha - \alpha_0\|_\infty$. Since Θ_n is a sequence of shrinking neighborhoods of $\beta_0 = (\theta_0, \alpha_0)$, there exists a positive and finite number C_0 such that $|\theta| \leq C_0$ and $\|\alpha_n\|_\infty \leq C_0$, $(\theta, \alpha_n) \in \Theta_n$, and α_n of form (2.1.3). For any $\eta > 0$, divide the interval $[0, C_0]$ into small intervals, with length $\eta/2$ or less, such that the number of intervals is less than or equal to $\frac{2C_0}{\eta} + 1$. Then, it is easy to see that

$$H(\eta, \Theta_n) \leq log\left(\left(\frac{2C_0}{\eta} + 1\right)\left(\frac{2C_0}{\eta} + 1\right)^{k(n)}\right) \leq Ck(n)log\left(\frac{1}{\eta}\right),$$

for some positive and finite constant C, as η is small enough. Hence, for small $\epsilon > 0$,

$$H(\epsilon, \mathcal{F}_n) \leq Ck(n)log\left(\frac{1}{\eta}\right) = Cn^\gamma log\left(\frac{1}{\eta}\right),$$

for some positive and finite number C. Therefore C3 is satisfied with $r_0 = \frac{\gamma}{2}$.

Apply the above lemma, we obtain

$$\rho(\beta, \beta_0) = O_p\left(max\left(n^{-\tau}, \rho(\pi_n\beta_0, \beta_0), K^{\frac{1}{2}}(\pi_n\beta_0, \beta_0)\right)\right), \tag{2.6.10}$$

where

$$\tau = \frac{1-\gamma}{2} - \frac{loglogn}{2logn}.$$

Note that, for large n, $\frac{1}{4} < \tau < \frac{3}{8}$ as $\frac{1}{4} < \gamma < \frac{1}{2}$. Since $\beta_0 = (\theta_0, \alpha_0)$, $\pi_n\beta_0 = (\theta_0, \alpha_{0n})$, where α_{0n} is of form (2.1.3), we obtain that, by (2.6.8) and (4) of Condition A,

$$\rho^2(\pi_n\beta_0, \beta_0) = ||\alpha_{0n} - \alpha_0||^2 \leq Ck(n)^{-2} = Cn^{-2\gamma},$$

for some positive and finite number C. Thus

$$\rho(\pi_n\beta_0, \beta_0) \leq Cn^{-\gamma}. \tag{2.6.11}$$

The same argument as that leading to (2.6.9) gives that, for some finite and positive number C,

$$K(\pi_n\beta_0, \beta_0) = E(l(\beta_0, Y) - l(\pi_n\beta_0, Y)) \leq C||\alpha_{0n} - \alpha_0||^2 = Cn^{-2\gamma}, \tag{2.6.12}$$

which is of order between $o(n^{-\frac{1}{2}})$ and $o(n^{-1})$ for $\frac{1}{4} < \gamma < \frac{1}{2}$. It follows from (2.6.10), (2.6.11) and (2.6.12) that, for $\frac{1}{4} < \gamma < \frac{1}{2}$,

$$\rho(\beta, \beta_0) = o_p(n^{-\frac{1}{4}}).$$

The theorem is proved.

2.6.3 Proof of Theorem 5

$S_{n,0}(\theta, \alpha)$ was defined in (2.1.4) and further denote, for a function a on $\mathcal{Z}$ with $E(a(Z))^2 < \infty$,

$$S_n(\theta, \alpha)[a] = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - F(T_i, Z_i, \theta, \alpha)}{F(T_i, Z_i, \theta, \alpha)} \Lambda(T_i, \theta) e^{\alpha(Z_i)} a(Z_i),$$

where $F(t, z, \theta, \alpha)$ is defined in (2.1.2).

Denote the expectation of $S_{n,0}$ and $S_n[a]$ by S_0 and $S[a]$ respectively. Since the conditional expectation of δ_i given (T_i, Z_i) is $F(T_i, Z_i, \theta_0, \alpha_0)$, we obtain

$$S_0(\theta, \alpha) = E\left[\frac{F(T, Z, \theta_0, \alpha_0) - F(T, Z, \theta, \alpha)}{F(T, Z, \theta, \alpha)}\dot{\Lambda}(T, \theta)e^{\alpha(Z)}\right] \tag{2.6.13}$$

and

$$S(\theta, \alpha)[a] = E\left[\frac{F(T, Z, \theta_0, \alpha_0) - F(T, Z, \theta, \alpha)}{F(T, Z, \theta, \alpha)}\Lambda(T, \theta)e^{\alpha(Z)}a(Z)\right]. \tag{2.6.14}$$

The method used here is similar to that described in Huang(1996). From Lemma 12 below, we obtain the following stochastic equi-continuity results

$$\begin{aligned}\sup_{|\theta-\theta_0|\leq Cn^{-\frac{1}{4}},\|\alpha-\alpha_0\|\leq Cn^{-\frac{1}{4}}} & |(S_{n,0}(\theta,\alpha) - S_0(\theta,\alpha)) - (S_{n,0}(\theta_0,\alpha_0) - S_0(\theta_0,\alpha_0))| \\ & = o_p^*(n^{-\frac{1}{2}})\end{aligned}$$

and

$$\begin{aligned}\sup_{|\theta-\theta_0|\leq Cn^{-\frac{1}{4}},\|\alpha-\alpha_0\|\leq Cn^{-\frac{1}{4}}} & |(S_n(\theta,\alpha)[a] - S(\theta,\alpha)[a]) - (S_n(\theta_0,\alpha_0)[a] - S(\theta_0,\alpha_0)[a])| \\ & = o_p^*(n^{-\frac{1}{2}}),\end{aligned}$$

for all a with $Ea^2(Z) < \infty$, and all positive and finite number C.

This and Theorem 4 results in

$$(S_{n,0}(\hat{\theta}, \hat{\alpha}_n) - S_0(\hat{\theta}, \hat{\alpha}_n)) - (S_{n,0}(\theta_0, \alpha_0) - S_0(\theta_0, \alpha_0)) = o_p(n^{-\frac{1}{2}}) \tag{2.6.15}$$

and

$$(S_n(\hat{\theta}, \hat{\alpha}_n)[a^*] - S(\hat{\theta}, \hat{\alpha}_n)[a^*]) - (S_n(\theta_0, \alpha_0)[a^*] - S(\theta_0, \alpha_0)[a^*]) = o_p(n^{-\frac{1}{2}}), \tag{2.6.16}$$

where a^* is defined in (2.4.6).

By the definition of $\hat{\theta}$ and $\hat{\alpha}_n$, $S_{n,0}(\hat{\theta}, \hat{\alpha}_n) = 0$. Also note that $S_0(\theta_0, \alpha_0) = 0$ by (2.6.13). It follows from (2.6.15) that

$$S_0(\hat{\theta}, \hat{\alpha}_n)) = -S_{n,0}(\theta_0, \alpha_0) + o_p(n^{-\frac{1}{2}}). \tag{2.6.17}$$

For the another part, we do not have $S_n(\hat{\theta}, \hat{\alpha}_n)[a^*] = 0$, but we will show that

$$S_n(\hat{\theta}, \hat{\alpha}_n)[a^*] = o_p(n^{-\frac{1}{2}}). \tag{2.6.18}$$

Together with $S(\theta_0, \alpha_0)[a^*] = 0$ by (2.6.14), we obtain from (2.6.16) that

$$S(\hat{\theta}, \hat{\alpha}_n))[a^*] = -S_n(\theta_0, \alpha_0)[a^*] + o_p(n^{-\frac{1}{2}}). \tag{2.6.19}$$

By Condition A and that the third derivative of $\Lambda(t, \theta)$ with respect to θ exists and is continuous, taking Taylor expansion of $S_0(\hat{\theta}, \hat{\alpha}_n))$ to the second order with respect to θ and α_n, we obtain from (2.6.17) and (2.6.13) that

$$\begin{aligned} &-E[D_{00}(T, Z, \theta_0, \alpha_0)](\hat{\theta} - \theta_0) - E\left[D_{01}(T, Z, \theta_0, \alpha_0)(\hat{\alpha}_n(Z) - \alpha_0(Z))\right] \\ &\quad = -S_{n,0}(\theta_0, \alpha_0) + O(|\hat{\theta} - \theta_0|^2 + \|\hat{\alpha}_n - \alpha_0\|^2) + o(n^{-\frac{1}{2}}). \end{aligned} \tag{2.6.20}$$

Similarly we can obtain from (2.6.19) and (2.6.14) that

$$\begin{aligned} &-E[D_{01}(T, Z, \theta_0, \alpha_0)a^*(Z)](\hat{\theta} - \theta_0) - E\left[D_{11}(T, Z, \theta_0, \alpha_0)a^*(Z)(\hat{\alpha}_n(Z) - \alpha_0(Z))\right] \\ &\quad = -S_n(\theta_0, \alpha_0)[a^*] + O(|\hat{\theta} - \theta_0|^2 + \|\hat{\alpha} - \alpha_0\|^2) + o(n^{-\frac{1}{2}}). \end{aligned} \tag{2.6.21}$$

By Theorem 4,

$$|\hat{\theta} - \theta_0|^2 + \|\hat{\alpha} - \alpha_0\|^2 = o_p(n^{-\frac{1}{2}}).$$

Subtracting (2.6.20) from (2.6.21) and noticing the definition of a^* (see (2.4.6)), we obtain

$$\begin{aligned} &E\left[D_{00}(T, Z, \theta_0, \alpha_0) - D_{01}(T, Z, \theta_0, \alpha_0)a^*(Z)\right](\hat{\theta} - \theta_0) \\ &\quad = S_{n,0}(\theta_0, \alpha_0) - S_n(\theta_0, \alpha_0)[a^*] + o_p(n^{-\frac{1}{2}}). \end{aligned}$$

The theorem follows from the central limit theorem and the calculation of the variance is straightforward. Now we prove (2.6.18). Let

$$a_n^*(z) = \sum_{j=1}^{k} a^*(z_j) I_j(z).$$

Condition A implies that

$$\|a_n^* - a^*\|_\infty = O(1/k(n)). \tag{2.6.22}$$

By the definition of $(\hat{\theta}, \hat{\alpha}_n)$, that is, it solves (2.1.7), we obtain

$$S_n(\hat{\theta}, \hat{\alpha}_n)[a_n^*] = 0.$$

We only need to show

$$S_n(\hat{\theta}, \hat{\alpha}_n)[a^*] - S_n(\hat{\theta}, \hat{\alpha}_n)[a_n^*] = o(n^{-\frac{1}{2}}).$$

Write the left hand side of the above display as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - F(T_i, Z_i, \theta_0, \alpha_0)}{F(T_i, Z_i, \hat{\theta}, \hat{\alpha}_n)} \Lambda(T_i, \hat{\theta}) e^{\hat{\alpha}_n(Z_i)}(a^*(Z_i) - a_n^*(Z_i)) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{F(T_i, Z_i, \theta_0, \alpha_0) - F(T_i, Z_i, \hat{\theta}, \hat{\alpha})}{F(T_i, Z_i, \hat{\theta}, \hat{\alpha}_n)} \Lambda(T_i, \hat{\theta}) e^{\hat{\alpha}_n(Z_i)}(a^*(Z_i) - a_n^*(Z_i)).
\end{aligned} \tag{2.6.23}$$

That the second term is $o(n^{-\frac{1}{2}})$ follows from (2.6.22), Theorem 4 and the Lipschitz continuity of F with respect to θ and α by Condition A. Similar proof as that of Lemma 12 below leads to that the first term is also $o(n^{-\frac{1}{2}})$.

Lemma 12 *For any positive and finite number C, and any function a on $\mathcal{Z}$ with $Ea^2(Z) < \infty$,*

$$\begin{aligned}
\sup_{|\theta-\theta_0| \le Cn^{-\frac{1}{4}}, \|\alpha-\alpha_0\| \le Cn^{-\frac{1}{4}}} &|\sqrt{n}(S_{n,0}(\theta, \alpha) - S_0(\theta, \alpha)) - \sqrt{n}(S_{n,0}(\theta_0, \alpha_0) - S_0(\theta_0, \alpha_0))| \\
&= o_p^*(1)
\end{aligned} \tag{2.6.24}$$

and

$$\sup_{\substack{|\theta-\theta_0|\le Cn^{-\frac{1}{4}}\\ ||\alpha-\alpha_0||\le Cn^{-\frac{1}{4}}}} |\sqrt{n}(S_n(\theta,\alpha)[a]-S(\theta,\alpha)[a])-\sqrt{n}(S_n(\theta_0,\alpha_0)[a]-S(\theta_0,\alpha_0)[a])|$$
$$= o_p^*(1). \quad (2.6.25)$$

Here $o_p^*(1)$ means tending to 0 in outer probability.

Proof. We prove the first part and the second one can be proved analogously. Note that $\sqrt{n}(S_{n,0}(\theta,\alpha)-S_0(\theta,\alpha))-\sqrt{n}(S_{n,0}(\theta_0,\alpha_0)-S_0(\theta_0,\alpha_0))$ is an empirical processes indexed by the set of functions

$$\mathcal{C}=\Big\{f(\delta,t,z,\theta,\alpha)=\frac{\delta-F(t,z,\theta,\alpha)}{F(t,z,\theta,\alpha)}\dot{\Lambda}(t,\theta)e^{\alpha(z)}-\frac{\delta-F(t,z,\theta_0,\alpha_0)}{F(t,z,\theta_0,\alpha_0)}\dot{\Lambda}(t,\theta_0)e^{\alpha_0(z)}:$$
$$|\theta-\theta_0|\le Cn^{-\frac{1}{4}}, ||\alpha-\alpha_0||\le Cn^{-\frac{1}{4}}\Big\}, \quad (2.6.26)$$

that is, by the functional notations used in the literature for empirical processes,

$$\sqrt{n}(S_{n,0}(\theta,\alpha)-S_0(\theta,\alpha))-\sqrt{n}(S_{n,0}(\theta_0,\alpha_0)-S_0(\theta_0,\alpha_0))$$
$$=\sqrt{n}(P_n-P)f(\delta,t,z,\theta,\alpha), \quad (2.6.27)$$

where P_n is the empirical measure based on $(\delta_i,T_i,Z_i), i=1,\cdots,n$ and P is the probability measure of (δ,T,Z) with respect to the real parameters (θ_0,α_0). Note that, under Condition A, functions in $\mathcal{C}$ are uniformly bounded for large n, and

$$|f(\delta,t,z,\theta,\alpha)-f(\delta,t,z,\theta_0,\alpha_0)|\le C_0(|\theta-\theta_0|+||\alpha-\alpha_0||_\infty), \quad (2.6.28)$$

for some finite and positive number C_0. Therefore, $\mathcal{C}$ is a set of functions which are Lipschitz in parameter $(\theta,\alpha)\in\mathcal{D}$, where

$$\mathcal{D}=\{(\theta-\theta_0,\alpha-\alpha_0):\alpha\quad\text{is of form (2.1.3)}, |\theta-\theta_0|\le Cn^{-\frac{1}{4}}, ||\alpha-\alpha_0||\le Cn^{-\frac{1}{4}}\}$$

and the norm in $L_\infty(\mathcal{D})$ is $||(\theta_1,\alpha_1)-(\theta_2,\alpha_2)||_\infty=|\theta_1-\theta_2|+||\alpha_1-\alpha_2||_\infty$. By Theorem 2.7.11 of Van der Vaart and Wellner (1996), the metric entropy of $\mathcal{C}$ with bracketing with respect to $L_2(P)$ norm

$$H_{[\,]}(\epsilon, \mathcal{C}, L_2(P)) \leq H(\epsilon/c, \mathcal{D}, L_\infty),$$

for some finite and positive number c. It is easy to see that

$$H(\epsilon, \mathcal{D}, L_\infty) \leq C_1 k(n) log\left(\frac{1}{\epsilon}\right),$$

for some finite and positive number C_1. Hence

$$H_{[\,]}(\epsilon, \mathcal{C}, L_2(P)) \leq C_2 k(n) log\left(\frac{1}{\epsilon}\right),$$

for some finite and positive number C_2. It is obtained that for any $\epsilon > 0$, there exists $0 < C_3 < \infty$, not depending on n, such that

$$J_{[\,]}(\epsilon, \mathcal{C}, L_2(P)) := \int_0^\epsilon \sqrt{1 + H_{[\,]}(t, \mathcal{C}, L_2(P))}dt \leq C_3 k(n)^{\frac{1}{2}} \epsilon^{1-\eta}, \quad \text{for any} \quad \eta > 0.$$

It follows from this that, as $k(n) = n^\gamma$ with $0 < \gamma < \frac{1}{2}$,

$$J_{[\,]}(Cn^{-\frac{1}{4}}, \mathcal{C}, L_2(P)) = o(1). \tag{2.6.29}$$

Note that $f(\delta, t, z, \theta_0, \alpha_0) = 0$ by (2.6.26). It follows from this and (2.6.28) that, for any $f \in \mathcal{C}$,

$$P(f(\delta, t, z, \theta, \alpha))^2 \leq C_4 n^{-\frac{1}{2}}, \tag{2.6.30}$$

for some finite and positive number C_4.

Apply Lemma 3.4.2 (page 324) of Van de vaart and Wellner (1996), which is stated in the following lemma. Let $Y_i = (\delta_i, T_i, Z_i)$, $i = 1, 2, \cdots, n$, $\mathcal{F} = \mathcal{C}$ and $\epsilon = Cn^{-\frac{1}{4}}$. By (2.6.30) and the boundedness of f, $f \in \mathcal{C}$, the conditions of the lemma hold. It follows from the lemma and (2.6.29) that

$$\sqrt{n} E^*(\sup_{f \in \mathcal{C}} |(P_n - P)f|) = o_p^*(1).$$

In view of (2.6.27), We obtain (2.6.24). The lemma is proved.

Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables (or possibly vectors) with distribution P and let P_n be the empirical measure of these random variables. Denote $G_n = \sqrt{n}(P_n - P)$ and $\|G_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |G_n f|$ for any measurable class of functions $\mathcal{F}$. Denote

$$J_{[\,]}(\epsilon, \mathcal{F}, L_2(P)) = \int_0^\epsilon \sqrt{1 + H_{[\,]}(t, \mathcal{F}, L_2(P))}dt.$$

Lemma 13 *Let $\mathcal{F}$ be a uniformly bounded class of measurable functions. Then*

$$E^*\|G_n\|_{\mathcal{F}} \leq CJ_{[\,]}(\epsilon, \mathcal{F}, L_2(P)) \left(1 + \frac{J_{[\,]}(\epsilon, \mathcal{F}, L_2(P))}{\epsilon^2 \sqrt{n}} M\right),$$

if every f in F satisfies $Pf^2 < \epsilon^2$ and $\|f\|_\infty \leq M$. Here E^* means outer expectation with respect to P.

Bibliography

[1] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. 1993. Efficient and adaptive estimation for semiparametric models. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD.

[2] DIAMOND, I. D. and MCDONALD, I.W. (1991). Analysis of current-status data. In *Demographic Applications of Event History Analysis*, Ed. J. Trussell, R. Hankinson and J. Tilton, pp. 231-52. Oxford: Oxford University Press.

[3] FINKELSTEIN, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 42, 845-54.

[4] GROENEBOON, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Basel: Birkhäuser Verlag.

[5] GESKUS, R. B. and GROENEBOON, P. (1996). Asymptotically optimal estimation of smooth functionals for interval censoring, part I. *Statist. Neerlandica* 50, 69-88.

[6] GESKUS, R. B. and GROENEBOON, P. (1997). Asymptotically optimal estimation of smooth functionals for interval censoring, part II. *Statist. Neerlandica* 51, 201-219.

[7] HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* 24,540-68.

[8] HUANG, J. and WELLNER, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data. *Statist. Neerlandica* 49, 153-63.

[9] JORDAN, C. W. (1975). *Life Contingencies.* The Society of Actuaries, Chicago.

[10] KLEIN, R. W. and SPADY, R. H. (1993). An efficient semiparametric estimator for binary response models. *Ecomometrica* 61, 387-421.

[11] KOUL, H. L. and SCHICK, A. (1999). Inference about the ratio of scale parameters in a two-sample setting with current status data. *Statist. and Probab. Letters* 45, 359-369.

[12] LI, G. and ZHANG, C.-H. (1998). Linear regression with interval censored data. *Ann. Atatist.* 26, 1306-1327.

[13] MESHALKIN, L. D. and KAGAN, A. R. (1972). Discussion of "Regression models and life tables," by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* 34, 213

[14] MURPHY, S. A., VAN DER VAART, A. W. and WELLNER J. A. Current status regression. *Mathematical Methods of Statistics* 8, 407-25.

[15] NIELSEN, J. P. and LINTON, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.* 23, 1735-1748.

[16] NIELSEN, J. P., LINTON, O. B. and BICKEL, P. J. (1998). On a semiparametric survival model with flexible covariate effect.

[17] RABINOWITZ, D., TSIATIS, A. and ARAGON, J. (1995). Regression with interval censored data. *Biometrika* 82, 501-13.

[18] ROSSINI, A. J. and TSIATIS, A. A. (1996). A semiparametric proportional odds regression model for the analysis of current status data. *J. Am. Statist. Assoc.* 91, 713-21.

[19] RUDIN, W. (1964). Principles of Mathematical Analysis, New York, McGraw-Hill.

[20] SEVERINI, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* 20, 1768–1802.

[21] SHEN, XIAOTONG and WONG, WING-HUNG. (1994). Convergence rate of sieve estimates. *Ann. Statist.* 22, 580–615.

[22] SHIBOSKI, S. C. and JEWELL, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Am. Statist. Assoc.* 87, 360-72.

[23] SHORACK, G. R. and WELLNER, J. A. (1986). Empirical Processes with Applications to Statistics. Wiley, New York.

[24] VAN DER VAART, A. W. AND WELLNER, J. A. (1996). Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York.