1
2006

This is to certify that the
dissertation entitled

SOME INFERENCE PROBLEMS FOR INTERVAL
CENSORED DATA

presented by

TINGTING YI

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Statistics

Major Professor's Signature

7/22/2005

Date

MSU is an Affirmative Action/Equal Opportunity Institution

SOME INFERENCE PROBLEMS FOR INTERVAL CENSORED DATA

By

Tingting Yi

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

2005

ABSTRACT

SOME INFERENCE PROBLEMS FOR INTERVAL CENSORED DATA

By

Tingting Yi

This thesis consists of two parts. The first part studies asymptotically efficient estimation of the baseline hazard parameters in a modified Cox model where covariate effects are nonparametric when data are interval censored. These estimators are obtained by maximizing the log-likelihood function with respect to both the finite dimensional and infinite dimensional nuisance parameters using method of sieves. The sequence of these estimators is shown to be consistent, asymptotically normal, with the asymptotic variance achieving the semiparametric lower bound.

The second part of the thesis pertains to constructing tests for fitting a class of parametric regression models to the regression function of the log of the event occurrence time variable when the data are interval censored case I and when the error distribution is known. These tests are based on certain martingale innovations of a marked empirical process. They are asymptotically distribution free in the sense that their asymptotic null distributions depend on neither the null model nor the covariate and the inspection time distributions.

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Hira L. Koul for his guidance and many helpful discussions on the subjects of this thesis. He was always available when I had doubts and questions. His general thinking of statistical problems and ways to solve the problems will help my future research and working. I would also like to thank all the other committee members, Professor Habib Salehi, Lijian Yang and Yijun Zuo, for serving my guidance committee. Finally I would like to thank the department of Statistics and Probability for offering me graduate assistantship so that I could come to the states and complete my graduate studies at the Michigan State University.

Contents

List of Tables

Chapter 1

Introduction

In an interval censoring set up one only knows that an event time X lies in a random interval. In the case I interval censoring model, the event time X is known to be either to the left of the observation time T or to its right. This type of data is also known as the current status data. In the case II interval censoring set up, there are two observation times T and U with $0 \leq T \leq U$, and one knows that either $X \leq T$, or $T < X \leq U$ or $X > U$.

Interval censored data occur frequently in clinical trials and longitudinal studies. For example, in a long-term follow up study, the subjects are given yearly screening to detect cancer. Cancer onset can only be known to occur between screenings. Hoel and Walburg (1972), Finkelstein and Wolfe (1985), Finkelstein (1986), Diamond et al. (1986), Diamond and McDonald (1991), Keiding (1991), among others, contain several examples of interval censoring data sets from clinical, tumorigenicity and demographic studies. The recent review article by Jewell and van der Laan (2004) contains some additional applications to health related studies.

The first part of the thesis studies asymptotically efficient estimation of the baseline hazard parameters in a modified Cox model where covariate effects are nonparametric when data are interval censored.

The Cox's regression model is widely used in survival analysis. In this model often the baseline hazard is assumed to be nonparametric while the covariate effects are modelled parametrically. 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 effect of covariates is nonparametric. They obtained an asymptotically efficient estimator of the underlying parameter by profile maximum likelihood method when the data are randomly right censored.

The estimator of the baseline hazard parameter, called *sieve maximum likelihood estimator*, is obtained by maximizing the log-likelihood function with respect to both the finite dimensional and 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. The sequence of these estimators is shown to be consistent, asymptotically normal, with the asymptotic variance achieving the semiparametric lower bound. This work thus generalizes the work of Lu (2000) from the current status data case to the general interval censoring case.

The nonparametric and semiparametric models for interval censored data have been studied in the literature. The monograph of Groeneboom and Wellner (1992) provides some basic results about the information bounds and nonparametric maximum likelihood estimators of a distribution function with current status and case II interval censored data. Huang and Wellner (1995) and Huang (1996) study NPMLE of the linear functionals and underlying parameters in Cox's proportional hazards model with interval censored data. Klein and Spady (1993) use the profile maximum likelihood method to derive estimators of the regression parameters that achieve the semiparametric lower bound. Li and Zhang (1998) derive an asymptotical efficient M-estimator of the regression parameters in a linear regression model with interval censored data. Rossini and Tsiatis (1996) use the sieve method to obtain asymptotically efficient estimators of the regression parameters in a semiparametric proportional odds regression model with current status data.

The second part of the thesis pertains to constructing the lack-of-fit tests of a parametric regression model when the response variables is subject to interval censoring case I. The proposed tests are based on certain martingale innovations of a marked empirical process developed by Stute (1997) and Stute, Thies and Zhu (1998). These two papers study the marked empirical process and its innovation martingale transformation for a general regression model. Our work extends the methodology to the interval censored case I data while the inspection time and covariate distributions are unknown but the error distribution is known. The tests are shown to be asymptotically distribution free in the sense that their asymptotic null distributions depend on neither the null model nor the covariate and the inspection time distributions.

Several other papers deal with model checking problems under censorship. Nikabadze and Stute (1997) use the Kaplan-Meier process to test the null hypothesis that the unknown distribution function of the true survival time is from a parametric family of distributions when the data are right censored. The Kapan-Meier process in their paper has also been transformed to be asymptotically distribution free. Stute (2000) constructs the tests based on the empirical process of the regressors marked by the residuals for model checking in right censored regression. The process in this paper is attached with the Kaplan-Meier weight and the weak limit of this process is a Gaussian process with a covariate function depending on the null model. Rabinowitz, Tsiatis and Aragon (1995) propose a class of score statistics that may be used for estimation and confidence procedures of an acceleration failure time model for interval censored data.

Simulations are conducted for both parts of the thesis. To study the behavior of the sieve MLE, a finite sample simulation shows a very desirable behavior in terms of bias and standard error. For the second part of the thesis, the simulation studies assessing some finite sample level and power behavior of the proposed tests for small and moderate sample sizes are also given.

Chapter 2

Sieve Estimation

2.1 The model

Let (X, T, U, Z) be a random vector, where X represents the survival or event time, T and U are the monitoring variables and Z the covariate which could be a vector. Assume that, conditional on Z, T and U are independent of X, with a joint continuous distribution function H such that $T \leq U$ with probability one.

In Cox's regression model, the conditional cumulative hazard rate function of X, given Z, has the form

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

where Λ_0, with unspecified form, is called the baseline cumulative hazard function, and β is a vector of regression parameters. Nielsen, Linton, and Bickel (1998) proposed an alternative model with Λ_0 depending only on some parameter θ_0 and the covariate effect is of an unspecified form. More specifically, the conditional cumulative hazard rate function of X, given Z, is of the form

$$\Lambda(x, \theta_0)g(Z), \tag{2.1.1}$$

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

Lu (2000) discusses the estimation of θ_0 and g based on the interval censoring case I data. In this paper we discuss the estimation of θ_0 and g based on interval censoring case II data, where one observes independent random vectors $(T_i, U_i, \delta_i, \gamma_i, Z_i)$, with $\delta_i = I_{(X_i \leq T_i)}$, $\gamma_i = I_{(T_i < X_i \leq U_i)}$, $i = 1, 2, \ldots, n$.

A consequence of (2.1.1) is that the conditional distribution of X, given Z, now depends on θ_0. Let $F(x, Z, \theta_0)$ denote this conditional distribution function. Assuming that the baseline cumulative hazard rate function $\Lambda(x, \theta_0)$ is continuous in x, we obtain

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

For the convenience and the transparency of the exposition we shall assume in the following sections that θ_0 is a scalar. For the case when θ_0 is a higher dimensional vector, similar results can be obtained. Because of the curse of dimensionality, we shall also assume that Z is a scalar variable.

2.2 Estimation

The goal is to estimate θ efficiently, with $\alpha(z) = \log g(z)$ as an infinite dimensional nuisance parameter. The conditional log-likelihood of θ and α given Z_i based on

$(T_i, U_i, \delta_i, \gamma_i, Z_i), i = 1, 2, \ldots, n$ is

$$
\begin{aligned}
Ln(\theta, \alpha) &= \frac{1}{n} \sum_{i=1}^{n} \Big\{ \delta_i \log F(T_i, Z_i, \theta, \alpha) \\
&\qquad + \gamma_i \log \left[F(U_i, Z_i, \theta, \alpha) - F(T_i, Z_i, \theta, \alpha) \right] \\
&\qquad + (1 - \delta_i - \gamma_i) \log \big(1 - F(U_i, Z_i, \theta, \alpha)\big) \Big\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \Big\{ \delta_i \log (1 - e^{-\Lambda(T_i, \theta) e^{\alpha(Z_i)}}) \\
&\qquad + \gamma_i \log \left[e^{-\Lambda(T_i, \theta) e^{\alpha(Z_i)}} - e^{-\Lambda(U_i, \theta) e^{\alpha(Z_i)}} \right] \\
&\qquad - (1 - \delta_i - \gamma_i) \Lambda(U_i, \theta) e^{\alpha(Z_i)} \Big\}. \qquad (2.2.1)
\end{aligned}
$$

Here

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

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 $\delta_i = 0$ and $\gamma_i = 0$. Hence a meaningful 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 sequence of step function with known jump points and maximize the log-likelihood as α varies over these 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

$$
\alpha_n(z) = \sum_{j=1}^{k} \alpha_{nj} I_j(z), \qquad (2.2.2)
$$

where $I_j(z)$ is the indicator function for the jth interval, defined by $I_j(z) = 1$ if $z_{j-1} < z \le z_j$ 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 (2.2.2) 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 (2.2.2) for α in (2.2.1). Since k is an increasing integer-valued function of n, written as $k(n)$, α_n will tend to α. The next section show that when $k(n) = O(n^r)$ with $0 < r < 1/2$, $(\hat{\theta}, \hat{\alpha}_n)$ is consistent.

Now define a step function, α_{0n}, of the form (2.2.2) as an approximation to α_0. Precisely,

$$\alpha_{0n}(z) = \sum_{j=1}^{k(n)} \alpha_0(z_j) I_j(z). \tag{2.2.3}$$

Let

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

$$\beta_0 = (\theta_0, \alpha_0), \qquad \hat{\beta}_n = (\hat{\theta}, \hat{\alpha}_n),$$

and

$$\begin{aligned} F_{\beta_0}(t,z) &= F(t,z,\theta_0,\alpha_0), \qquad & F_{\beta_n}(t,z) &= F(t,z,\theta,\alpha_n), \\ F_{\beta_{0n}}(t,z) &= F(t,z,\theta_0,\alpha_{0n}), \qquad & F_{\hat{\beta}_n}(t,z) &= F(t,z,\hat{\theta},\hat{\alpha}_n). \end{aligned} \tag{2.2.4}$$

We shall be assuming $\Lambda(t,\theta)$ to be twice differentiable with respect to θ, for all t. Let $\dot{\Lambda}(t,\theta)$, $\ddot{\Lambda}(t,\theta)$ denote the first and second derivatives of $\Lambda(t,\theta)$ with respect to θ.

The first and second partial derivatives of the log-likelihood are used to generate the estimates and their variance. In view of (2.2.1), the first derivative with respect to θ is

$$
\begin{aligned}
&S_{n,\,0}(\theta,\alpha_n)\\
&= \frac{\partial L_n(\theta,\alpha_n)}{\partial\theta}\\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i - F_{\beta_n}(T_i,Z_i)}{F_{\beta_n}(T_i,Z_i)}\dot\Lambda(T_i,\theta)e^{\alpha_n(Z_i)}\\
&\quad + \left[\delta_i - F_{\beta_n}(T_i,Z_i) + \frac{\gamma_i - [F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)]}{F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)}\bar F_{\beta_n}(T_i,Z_i)\right]\\
&\quad \times [\dot\Lambda(U_i,\theta) - \dot\Lambda(T_i,\theta)]e^{\alpha_n(Z_i)}\Big\}
\end{aligned}
\tag{2.2.5}
$$

and that with respect to α_{nj} is

$$
\begin{aligned}
&S_{n,\,j}(\theta,\alpha_n)\\
&= \frac{\partial L_n(\theta,\alpha_n)}{\partial\alpha_{nj}}\\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i - F_{\beta_n}(T_i,Z_i)}{F_{\beta_n}(T_i,Z_i)}\Lambda(T_i,\theta)e^{\alpha_n(Z_i)}I_j(Z_i)\\
&\quad + \left[\delta_i - F_{\beta_n}(T_i,Z_i) + \frac{\gamma_i - [F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)]}{F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)}\bar F_{\beta_n}(T_i,Z_i)\right]\\
&\quad \times [\Lambda(U_i,\theta) - \Lambda(T_i,\theta)]e^{\alpha_n(Z_i)}I_j(Z_i)\Big\} \qquad j = 1,2,\cdots,k.
\end{aligned}
\tag{2.2.6}
$$

The score vector is defined as

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

The estimates $(\hat\theta,\hat\alpha_n)$ are defined to be a solution to the score equation

$$
\tilde S_n(\theta,\alpha_n) = 0. \tag{2.2.7}
$$

The derivative of $\tilde{S}_n$ with respect to (θ, α_n) is called the Hessian matrix and related to the observed information. This is defined as

$$\tilde{H}_n(\theta, \alpha_n) = \frac{\partial \tilde{S}_n(\theta, \alpha_n)}{\partial\theta\partial\alpha_n}.$$

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

$$\begin{aligned}
h_{00}(\theta, \alpha_n) &= \frac{\partial S_{n,\,0}(\theta, \alpha_n)}{\partial\theta} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\delta_i - F_{\beta_n}(T_i, Z_i)}{F_{\beta_n}(T_i, Z_i)}\ddot{\Lambda}(T_i, \theta)e^{\alpha_n(Z_i)} \\
&\quad + \left[\delta_i - F_{\beta_n}(T_i, Z_i) + \frac{\gamma_i - [F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)]}{F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)}\bar{F}_{\beta_n}(T_i, Z_i)\right] \\
&\quad \times[\ddot{\Lambda}(U_i, \theta) - \ddot{\Lambda}(T_i, \theta)]e^{\alpha_n(Z_i)}\Bigg\} \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\delta_i\bar{F}_{\beta_n}(T_i, Z_i)}{F^2_{\beta_n}(T_i, Z_i)}\dot{\Lambda}^2(T_i, \theta)e^{2\alpha_n(Z_i)} \\
&\quad +\frac{\gamma_i\bar{F}_{\beta_n}(T_i, Z_i)\bar{F}_{\beta_n}(U_i, Z_i)}{[F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)]^2}[\dot{\Lambda}(U_i, \theta) - \dot{\Lambda}(T_i, \theta)]^2e^{2\alpha_n(Z_i)}\Bigg\},
\end{aligned} \tag{2.2.8}$$

$$\begin{aligned}
h_{0j}(\theta, \alpha_n) &= \frac{\partial S_{n,\,0}(\theta, \alpha_n)}{\partial\alpha_j} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\delta_i - F_{\beta_n}(T_i, Z_i)}{F_{\beta_n}(T_i, Z_i)}\dot{\Lambda}(T_i, \theta)e^{\alpha_n(Z_i)}I_j(Z_i) \\
&\quad + \left[\delta_i - F_{\beta_n}(T_i, Z_i) + \frac{\gamma_i - [F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)]}{F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)}\bar{F}_{\beta_n}(T_i, Z_i)\right] \\
&\quad \times[\dot{\Lambda}(U_i, \theta) - \dot{\Lambda}(T_i, \theta)]e^{\alpha_n(Z_i)}I_j(Z_i)\Bigg\} \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\Bigg\{\frac{\delta_i\bar{F}_{\beta_n}(T_i, Z_i)}{F^2_{\beta_n}(T_i, Z_i)}\dot{\Lambda}(T_i, \theta)\Lambda(T_i, \theta)e^{2\alpha_n(Z_i)}I_j(Z_i) \\
&\quad +\frac{\gamma_i\bar{F}_{\beta_n}(T_i, Z_i)\bar{F}_{\beta_n}(U_i, Z_i)}{[F_{\beta_n}(U_i, Z_i) - F_{\beta_n}(T_i, Z_i)]^2} \\
&\quad \times[\dot{\Lambda}(U_i, \theta) - \dot{\Lambda}(T_i, \theta)][\Lambda(U_i, \theta) - \Lambda(T_i, \theta)]e^{2\alpha_n(Z_i)}I_j(Z_i)\Bigg\},
\end{aligned} \tag{2.2.9}$$

$$
\begin{aligned}
h_{jj}(\theta,\alpha_n) &= \frac{\partial S_{n,\,j}(\theta,\alpha_n)}{\partial \alpha_j} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i - F_{\beta_n}(T_i,Z_i)}{F_{\beta_n}(T_i,Z_i)}\Lambda(T_i,\theta)e^{\alpha_n(Z_i)}I_j(Z_i) \\
&\quad + \left[\delta_i - F_{\beta_n}(T_i) + \frac{\gamma_i - [F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)]}{F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)}\bar{F}_{\beta_n}(T_i,Z_i)\right] \\
&\quad \times[\Lambda(U_i,\theta) - \Lambda(T_i,\theta)]e^{\alpha_n(Z_i)}I_j(Z_i)\Big\} \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i\bar{F}_{\beta_n}(T_i,Z_i)}{F^2_{\beta_n}(T_i,Z_i)}\Lambda^2(T_i,\theta)e^{2\alpha_n(Z_i)}I_j(Z_i) \\
&\quad +\frac{\gamma_i\bar{F}_{\beta_n}(T_i,Z_i)\bar{F}_{\beta_n}(U_i,Z_i)}{[F_{\beta_n}(U_i,Z_i) - F_{\beta_n}(T_i,Z_i)]^2}[\Lambda(U_i,\theta) - \Lambda(T_i,\theta)]^2e^{2\alpha_n(Z_i)}I_j(Z_i)\Big\} \qquad (2.2.10) \\
&\qquad\qquad j = 1,\cdots,k, \\
h_{ij}(\theta,\alpha_n) &= 0, \qquad\qquad i \neq j = 1,2,\cdots,k.
\end{aligned}
$$

In the above expressions the expectation is taken with respect to the true parameters (θ_0, α_0).

2.3 Consistency and Asymptotic Normality

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

(3) The first and second derivative of $\Lambda(x,\theta)$ with respect to θ, $\dot{\Lambda}(x,\theta)$ and $\ddot{\Lambda}(x,\theta)$

exist, are bounded for $x \in \mathcal{T}$ or $x \in \mathcal{U}$ and $\theta \in \mathcal{N}_1$, and continuous in θ, for any fixed x.

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

For any function $b(z)$ defined on $\mathcal{Z}$, 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.

Theorem (2.3.1) below states the existence of at least one consistent (in sup-norm) estimator, $\hat{\theta}$, which is a solution of the score equation.

First, let's define

$$
\begin{aligned}
D_{00}(T,U,Z,\theta_0,\alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T,Z)}{F_{\beta_{0n}}(T,Z)}\dot{\Lambda}^2(T,\theta_0) + \frac{\bar{F}_{\beta_{0n}}(T,Z)\bar{F}_{\beta_{0n}}(U,Z)}{F_{\beta_{0n}}(U,Z) - F_{\beta_{0n}}(T,Z)} \\
&\qquad \times[\dot{\Lambda}(U,\theta_0) - \dot{\Lambda}(T,\theta_0)]^2\Big]e^{2\alpha_{0n}(Z)}, \\
D_{01}(T,U,Z,\theta_0,\alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T,Z)}{F_{\beta_{0n}}(T,Z)}\dot{\Lambda}(T,\theta_0)\Lambda(T,\theta_0) + \frac{\bar{F}_{\beta_{0n}}(T,Z)\bar{F}_{\beta_{0n}}(U,Z)}{F_{\beta_{0n}}(U,Z) - F_{\beta_{0n}}(T,Z)} \\
&\qquad \times[\dot{\Lambda}(U,\theta_0) - \dot{\Lambda}(T,\theta_0)][\Lambda(U,\theta_0) - \Lambda(T,\theta_0)]\Big]e^{2\alpha_{0n}(Z)}, \\
D_{11}(T,U,Z,\theta_0,\alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T,Z)}{F_{\beta_{0n}}(T,Z)}\Lambda^2(T,\theta_0) + \frac{\bar{F}_{\beta_{0n}}(T,Z)\bar{F}_{\beta_{0n}}(U,Z)}{F_{\beta_{0n}}(U,Z) - F_{\beta_{0n}}(T,Z)} \\
&\qquad \times[\Lambda(U,\theta_0) - \Lambda(T,\theta_0)]^2\Big]e^{2\alpha_{0n}(Z)}, \qquad (2.3.1)
\end{aligned}
$$

where $F_{\beta_{0n}}(T,Z)$ is defined in (2.2.4).

Theorem 2.3.1 *Assume that the Condition A holds, and the number of intervals is increasing at a rate $k(n) = n^r$, with $0 < r < 1/2$. Assume that*

$$(2.3.2) \qquad P(F(U,Z,\theta,\alpha) - F(T,Z,\theta,\alpha) > c) = 1, \qquad \forall\ \theta \in \Theta, \alpha,$$

for some $0 < c < \infty$. 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, there exists a $0 < C < \infty$, not depending on n, such that

$$(2.3.3) \quad P(I_j(Z) = 1) = o(1), \quad kP(I_j(Z) = 1) > C, \quad j = 1, 2, \cdots, k$$

and

$$(2.3.4)\; E[D_{00}(T,U,Z,\theta_0,\alpha_{0n})] - \sum_{j=1}^{k} \frac{E[D_{01}(T,U,Z,\theta_0,\alpha_{0n})I_j(Z)]^2}{E[D_{11}(T,U,Z,\theta_0,\alpha_{0n})I_j(Z)]} > C.$$

Then, there is at least one consistent (in sup-norm) solution to (2.2.7), i.e. there exists one $(\hat{\theta}, \hat{\alpha}_n)$ such that

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

Theorem 2.3.2 *Assume that the conditions in Theorem 2.3.1 hold. Assume also $k(n) = n^r$, with $1/4 < r < 1/2$, and*

$$(2.3.5) \qquad E[D_{00}(T,U,Z,\theta_0,\alpha_0)] - \frac{(E[D_{01}(T,U,Z,\theta_0,\alpha_0)])^2}{E[D_{11}(T,U,Z,\theta_0,\alpha_0)]} > 0.$$

Then the estimator $(\hat{\theta}, \hat{\alpha}_n)$ in Theorem (2.3.1) has the following convergence rate

$$|\hat{\theta} - \theta_0| = o_P(n^{-1/4}), \qquad ||\hat{\alpha}_n - \alpha_0||_\infty = o_P(n^{-1/4}).$$

Theorem 2.3.3 *Assume that the conditions of Theorem 2.3.2 hold, and σ^2 defined below is finite. Assume also 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

$$(2.3.6) \quad \sigma^2 = \left[E(D_{00}(T,U,Z,\theta_0,\alpha_0)) - E\left(\frac{(E(D_{01}(T,U,Z,\theta_0,\alpha_0)|Z))^2}{E(D_{11}(T,U,Z,\theta_0,\alpha_0)|Z)}\right)\right]^{-1}.$$

2.4 Information bound for θ_0

The true model has two parameters: θ, a finite dimensional, and α, an infinite dimensional functional parameter. The semi-parametric 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 (Bickel et al. 1993). It is shown in this section that the above asymptotic variance σ^2 achieves this bound. Projection method is used to find the efficient score for the semi-parametric model and hence the variance bound.

The conditional log-likelihood of θ and α. given Z based on $(T, U, \delta, \gamma, Z)$ is given by

$$
\begin{aligned}
&\delta \log\left(1 - e^{-\Lambda(T,\theta)e^{\alpha(Z)}}\right) + \gamma \log\left[e^{-\Lambda(T,\theta)e^{\alpha(Z)}} - e^{-\Lambda(U,\theta)e^{\alpha(Z)}}\right] \\
&-(1-\delta-\gamma)\Lambda(U,\theta)e^{\alpha(Z)}. \qquad (2.4.1)
\end{aligned}
$$

Consider a general parametric sub-model with $\alpha = \alpha_\tau$, specified by τ (a real number) where $\frac{\partial}{\partial\tau}\alpha_\tau(z)|_{\tau=0} = a(z)$ for some function $a(z)$ with $Ea^2(Z) < \infty$. Take derivatives of (2.4.1) with respect to θ and τ at $(\alpha = \alpha_0, \tau = 0)$ to obtain the scores

$$
\begin{aligned}
&S_0(T, U, Z, \delta, \gamma, \theta_0, \alpha_0) \\
&= \frac{\delta - F_{\beta_0}(T,Z)}{F_{\beta_0}(T,Z)}\dot{\Lambda}(T,\theta_0)e^{\alpha_0(Z)} \\
&\quad + \left[\delta - F_{\beta_0}(T,Z) + \frac{\gamma - [F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)]}{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)}\bar{F}_{\beta_0}(T,Z)\right] \\
&\qquad \times[\dot{\Lambda}(U,\theta_0) - \dot{\Lambda}(T,\theta_0)]e^{\alpha_0(Z)}, \qquad (2.4.2)
\end{aligned}
$$

$$
\begin{aligned}
&S_a(T, U, Z, \delta, \gamma, \theta_0, \alpha_0) \\
&= \frac{\delta - F_{\beta_0}(T,Z)}{F_{\beta_0}(T,Z)}\Lambda(T,\theta_0)e^{\alpha_0(Z)}a(Z) \\
&\quad + \left[\delta - F_{\beta_0}(T,Z) + \frac{\gamma - [F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)]}{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)}\bar{F}_{\beta_0}(T)\right] \\
&\qquad \times\Lambda(U,\theta_0) - \Lambda(T,\theta_0)]e^{\alpha_0(Z)}a(Z). \qquad (2.4.3)
\end{aligned}
$$

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

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

Note that

$$\begin{aligned}
E(\delta|T,U,Z) &= F_{\beta_0}(T,Z), \\
E(\gamma|T,U,Z) &= F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z), \\
Var(\delta|T,U,Z) &= F_{\beta_0}(T,Z)\bar{F}_{\beta_0}(T,Z), \\
Var(\gamma|T,U,Z) &= \left[F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)\right] \times \left[1 - (F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z))\right], \\
E(\gamma\delta|T,U,Z) &= 0. \qquad (2.4.5)
\end{aligned}$$

Substituting (2.4.2), (2.4.3) for S_0 and S_a in (2.4.4), taking conditional expectation given (T,U,Z) first, and then taking expectation with respect to (T,U,Z), we obtain

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

where D_{01} and D_{11} were defined in (2.3.1). Taking conditional expectation given Z first, and then expectation with respect to Z, we obtain

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

It can be verified that

$$a^*(Z) = \frac{E(D_{01}(T,U,Z,\theta_0,\alpha_0)|Z)}{E(D_{11}(T,U,Z,\theta_0,\alpha_0)|Z)} \tag{2.4.7}$$

solves (2.4.6) and hence also solves (2.4.4).

Therefore, the efficient score is given by

$$
\begin{aligned}
& S_0(T,U,Z,\delta,\gamma,\theta_0,\alpha_0) - S_{a^*}(T,U,Z,\delta,\gamma,\theta_0,\alpha_0) \\
= \ & \frac{\delta - F_{\beta_0}(T,Z)}{F_{\beta_0}(T,Z)} e^{\alpha_0(Z)} \left(\dot{\Lambda}(T,\theta_0) - \Lambda(T,\theta_0) \frac{E(D_{01}(T,U,Z,\theta_0,\alpha_0)|Z)}{E(D_{11}(T,U,Z,\theta_0,\alpha_0)|Z)} \right) \\
& + \left[\delta - F_{\beta_0}(T,Z) + \frac{\gamma - [F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)]}{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)} \bar{F}_0(T,Z) \right] e^{\alpha_0(Z)} \\
& \times \left([\dot{\Lambda}(U,\theta_0) - \dot{\Lambda}(T,\theta_0)] - [\Lambda(U,\theta_0) - \Lambda(T,\theta_0)] \frac{E(D_{01}(T,U,Z,\theta_0,\alpha_0)|Z)}{E(D_{11}(T,U,Z,\theta_0,\alpha_0)|Z)} \right).
\end{aligned}
$$

The semiparametric information bound is equal to the variance of the efficient score:

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

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,U,Z) first, and then expectation with respect to (T,U,Z) to obtain

$$
\begin{aligned}
& E[S_0(T,U,Z,\delta,\gamma,\theta_0,\alpha_0) - S_{a^*}(T,U,Z,\delta,\gamma,\theta_0,\alpha_0)]^2 = E[S_0^2 + S_{a^*}^2 - 2S_0 S_{a^*}] \\
= \ & E\left[D_{00}(T,U,Z,\theta_0,\alpha_0) - \frac{(E(D_{01}(T,U,Z,\theta_0,\alpha_0)|Z))^2}{E(D_{11}(T,U,Z,\theta_0,\alpha_0)|Z)} \right].
\end{aligned}
$$

In view of (2.3.6), it follows that σ^2, the asymptotic variance of $\hat{\theta}$ achieves the asymptotic variance bound.

2.5 An Extension

In this section, we will extend the above results to multiple interval censoring case. Let $(X, \mathbf{T}, Z)$ be a random vector, where X still represents the survival time, Z the covariate, and $\mathbf{T} = (T_1, T_2, \cdots, T_p)$, are the vector of monitoring times with $P(0 < T_1 < T_2 < \cdots < T_p < \infty) = 1$. As before, $\mathbf{T}$ and X are conditionally independent, given Z. Let $\mathbf{\Delta} = (\delta_1, \delta_2, \cdots, \delta_p)$ where $\delta_j = I(T_{j-1} < X \leq T_j)$ $j = 1, \cdots, p$. Set $T_0 = 0$, $T_{p+1} = \infty$.

Suppose we observe n i.i.d. copies $(\mathbf{T_i}, \mathbf{\Delta_i}, Z_i)$ $i = 1, \cdots, n$ of $(\mathbf{T}, \mathbf{\Delta}, Z)$, where $\mathbf{T_i} = (T_{i,1}, \cdots, T_{i,p})$, $\mathbf{\Delta_i} = (\delta_{i,1}, \cdots, \delta_{i,p})$ and $\delta_{i,j} = I(T_{i,j-1} < X_i \leq T_{i,j})$, $j = 1, \cdots, p$, and $T_{i,0} = 0$, $T_{i,p+1} = \infty$ for any $i = 1, \cdots, n$.

Then the conditional log-likelihood of (θ, α) based on $(\mathbf{T_i}, \mathbf{\Delta_i})$, given Z_i, $i = 1, \cdots, n$, is

$$
\begin{aligned}
Ln(\theta, \alpha) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p+1} \delta_{i,j} \log \Big(F(T_{i,j}, Z_i, \theta, \alpha) - F(T_{i,j-1}, Z_i, \theta, \alpha) \Big) \\
&= \frac{1}{n} \sum_{i=1}^{n} \Big\{ \delta_{i,1} \log (1 - e^{-\Lambda(T_{i,1}, \theta) e^{\alpha(Z_i)}}) - \delta_{i,p+1} \Lambda(T_{i,p}, \theta) e^{\alpha(Z_i)} \\
&\qquad + \sum_{j=2}^{p} \delta_{i,j} \log [e^{-\Lambda(T_{i,j-1}, \theta) e^{\alpha(Z_i)}} - e^{-\Lambda(T_{i,j}, \theta) e^{\alpha(Z_i)}}] \Big\}.
\end{aligned}
$$

Here $\delta_{i,p+1} = 1 - \sum_{j=1}^{p} \delta_{i,j}$, $i = 1, \cdots, n$.

The first and second derivatives of this log-likelihood with respect to (θ, α_n) are

$$
\begin{aligned}
& S_{n,0}(\theta, \alpha n) \\
&= \frac{\partial Ln(\theta, \alpha n)}{\partial \theta} \\
&= \frac{1}{n} \sum_{i=1}^{n} e^{\alpha n(Z_i)} \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta_n}(T_{i,1}, Z_i)}{F_{\beta_n}(T_{i,1}, Z_i)} \dot{\Lambda}(T_{i,1}, \theta) - \delta_{i,p+1} \dot{\Lambda}(T_{i,p}, \theta) \\
&\quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta_n}(T_{i,j}, Z_i) \dot{\Lambda}(T_{i,j}, \theta) - \bar{F}_{\beta_n}(T_{i,j-1}, Z_i) \dot{\Lambda}(T_{i,j-1}, \theta)}{F_{\beta_n}(T_{i,j}, Z_i) - F_{\beta_n}(T_{i,j-1}, Z_i)} \Big\},
\end{aligned}
$$

$$
\begin{aligned}
& S_{n,s}(\theta, \alpha n) \\
&= \frac{\partial Ln(\theta, \alpha n)}{\partial \alpha ns} \\
&= \frac{1}{n} \sum_{i=1}^{n} e^{\alpha n(Z_i)} I_s(Z_i) \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta_n}(T_{i,1}, Z_i)}{F_{\beta_n}(T_{i,1}, Z_i)} \Lambda(T_{i,1}, \theta) - \delta_{i,p+1} \Lambda(T_{i,p}, \theta) \\
&\quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta_n}(T_{i,j}, Z_i) \Lambda(T_{i,j}, \theta) - \bar{F}_{\beta_n}(T_{i,j-1}, Z_i) \Lambda(T_{i,j-1}, \theta)}{F_{\beta_n}(T_{i,j}, Z_i) - F_{\beta_n}(T_{i,j-1}, Z_i)} \Big\}, \\
&\qquad\qquad\qquad\qquad s = 1, \cdots, k.
\end{aligned}
$$

$$
\begin{aligned}
& h_{00}(\theta, \alpha n) \\
= \; & \frac{\partial S_{n,0}(\theta, \alpha n)}{\partial \theta} \\
= \; & \frac{1}{n} \sum_{i=1}^{n} e^{\alpha n(Z_i)} \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta n}(T_{i,1}, Z_i)}{F_{\beta n}(T_{i,1}, Z_i)} \ddot{\Lambda}(T_{i,1}, \theta) - \delta_{i,p+1} \ddot{\Lambda}(T_{i,p}, \theta) \\
& \quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta n}(T_{i,j}, Z_i) \ddot{\Lambda}(T_{i,j}, \theta) - \bar{F}_{\beta n}(T_{i,j-1}, Z_i) \ddot{\Lambda}(T_{i,j-1}, \theta)}{F_{\beta n}(T_{i,j}, Z_i) - F_{\beta n}(T_{i,j-1}, Z_i)} \Big\} \\
& - \frac{1}{n} \sum_{i=1}^{n} e^{2\alpha n(Z_i)} \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta n}(T_{i,1}, Z_i)}{F^2_{\beta n}(T_{i,1}, Z_i)} \dot{\Lambda}^2(T_{i,1}, \theta) \\
& \quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta n}(T_{i,j}, Z_i) \bar{F}_{\beta n}(T_{i,j-1}, Z_i)}{[F_{\beta n}(T_{i,j}, Z_i) - F_{\beta n}(T_{i,j-1}, Z_i)]^2} \\
& \qquad [\dot{\Lambda}(T_{i,j}, \theta) - \dot{\Lambda}(T_{i,j-1}, \theta)]^2 \Big\},
\end{aligned}
$$

$$
\begin{aligned}
& h_{0s}(\theta, \alpha n) \\
= \; & \frac{\partial S_{n,0}(\theta, \alpha n)}{\partial \alpha_{ns}} \\
= \; & \frac{1}{n} \sum_{i=1}^{n} e^{\alpha n(Z_i)} I_s(Z_i) \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta n}(T_{i,1}, Z_i)}{F_{\beta n}(T_{i,1}, Z_i)} \dot{\Lambda}(T_{i,1}, \theta) - \delta_{i,p+1} \dot{\Lambda}(T_{i,p}, \theta) \\
& \quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta n}(T_{i,j}, Z_i) \dot{\Lambda}(T_{i,j}, \theta) - \bar{F}_{\beta n}(T_{i,j-1}, Z_i) \dot{\Lambda}(T_{i,j-1}, \theta)}{F_{\beta n}(T_{i,j}, Z_i) - F_{\beta n}(T_{i,j-1}, Z_i)} \Big\} \\
& - \frac{1}{n} \sum_{i=1}^{n} e^{2\alpha n(Z_i)} I_s(Z_i) \Big\{ \delta_{i,1} \frac{\bar{F}_{\beta n}(T_{i,1}, Z_i)}{F^2_{\beta n}(T_{i,1}, Z_i)} \dot{\Lambda}(T_{i,1}, \theta) \Lambda(T_{i,1}, \theta) \\
& \quad + \sum_{j=2}^{j=p} \delta_{i,j} \frac{\bar{F}_{\beta n}(T_{i,j}, Z_i) \bar{F}_{\beta n}(T_{i,j-1}, Z_i)}{[F_{\beta n}(T_{i,j}, Z_i) - F_{\beta n}(T_{i,j-1}, Z_i)]^2} \\
& \quad \times [\dot{\Lambda}(T_{i,j}, \theta) - \dot{\Lambda}(T_{i,j-1}, \theta)][\Lambda(T_{i,j}, \theta) - \Lambda(T_{i,j-1}, \theta)] \Big\},
\end{aligned}
$$

$$
\begin{aligned}
&h_{ss}(\theta, \alpha_n) \\
&= \frac{\partial S_{n,0}(\theta, \alpha_n)}{\partial \theta} \\
&= \frac{1}{n}\sum_{i=1}^{n} e^{\alpha_n(Z_i)} I_s(Z_i)\Big\{\delta_{i,1}\frac{\bar{F}_{\beta_n}(T_{i,1}, Z_i)}{F_{\beta_n}(T_{i,1}, Z_i)}\Lambda(T_{i,1}, \theta) - \delta_{i,p+1}\Lambda(T_{i,p}, \theta) \\
&+ \sum_{j=2}^{j=p} \delta_{i,j}\frac{\bar{F}_{\beta_n}(T_{i,j}, Z_i)\Lambda(T_{i,j}, \theta) - \bar{F}_{\beta_n}(T_{i,j-1}, Z_i)\Lambda(T_{i,j-1}, \theta)}{F_{\beta_n}(T_{i,j}, Z_i) - F_{\beta_n}(T_{i,j-1}, Z_i)}\Big\} \\
&- \frac{1}{n}\sum_{i=1}^{n} e^{2\alpha_n(Z_i)} I_s(Z_i)\Big\{\delta_{i,1}\frac{\bar{F}_{\beta_n}(T_{i,1}, Z_i)}{F^2_{\beta_n}(T_{i,1}, Z_i)}\Lambda^2(T_{i,1}, \theta) \\
&+ \sum_{j=2}^{j=p} \delta_{i,j}\frac{\bar{F}_{\beta_n}(T_{i,j}, Z_i)\bar{F}_{\beta_n}(T_{i,j-1}, Z_i)}{[F_{\beta_n}(T_{i,j}, Z_i) - F_{\beta_n}(T_{i,j-1}, Z_i)]^2}[\Lambda(T_{i,j}, \theta) - \Lambda(T_{i,j-1}, \theta)]^2\Big\},
\end{aligned}
$$

where $s = 1, \cdots, k$, and $F_{\beta_n}(\mathbf{T}, Z)$ is defined in (2.2.4).

Under the assumptions similar to Condition A, we can prove analog of Theorem (2.3.1) to (2.3.3) for this multiple interval censoring case. Note that the bold $\mathbf{T}$ is the random vector while T_i is its ith coordinate.

First we need to define

$$
\begin{aligned}
D_{00}(\mathbf{T}, Z, \theta_0, \alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T_1, Z)}{F_{\beta_{0n}}(T_1, Z)}\dot{\Lambda}^2(T_1, \theta_0) + \sum_{j=2}^{p} \frac{\bar{F}_{\beta_{0n}}(T_j, Z)\bar{F}_{\beta_{0n}}(T_{j-1}, Z)}{F_{\beta_{0n}}(T_j, Z) - F_{\beta_{0n}}(T_{j-1}, Z)} \\
&\quad \times [\dot{\Lambda}(T_j, \theta_0) - \dot{\Lambda}(T_{j-1}, \theta_0)]^2\Big] e^{2\alpha_{0n}(Z)} \\
D_{01}(\mathbf{T}, Z, \theta_0, \alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T_1, Z)}{F_{\beta_{0n}}(T_1, Z)}\dot{\Lambda}(T, \theta_0)\Lambda(T, \theta_0) + \sum_{j=2}^{p} \frac{\bar{F}_{\beta_{0n}}(T_j, Z)\bar{F}_{\beta_{0n}}(T_{j-1}, Z)}{F_{\beta_{0n}}(T_j, Z) - F_{\beta_{0n}}(T_{j-1}, Z)} \\
&\quad \times [\dot{\Lambda}(T_j, \theta_0) - \dot{\Lambda}(T_{j-1}, \theta_0)][\Lambda(T_j, \theta_0) - \Lambda(T_{j-1}, \theta_0)]\Big] e^{2\alpha_{0n}(Z)} \\
D_{11}(\mathbf{T}, Z, \theta_0, \alpha_{0n}) &= \Big[\frac{\bar{F}_{\beta_{0n}}(T_1, Z)}{F_{\beta_{0n}}(T_1, Z)}\Lambda^2(T_1, \theta_0) + \sum_{j=2}^{p} \frac{\bar{F}_{\beta_{0n}}(T_j, Z)\bar{F}_{\beta_{0n}}(T_{j-1}, Z)}{F_{\beta_{0n}}(T-j, Z) - F_{\beta_{0n}}(T_{j-1}, Z)} \\
&\quad \times [\Lambda(T_j, \theta_0) - \Lambda(T_{j-1}, \theta_0)]^2\Big] e^{2\alpha_{0n}(Z)},
\end{aligned}
$$

where $F_{\beta_{0n}}(\mathbf{T}, Z)$ is defined in (2.2.4).

The consistency of the sieve MLE can be proved similarly as Theorem (2.3.1) when we change the assumption (2.3.2) to

$$P(F(T_j, Z, \theta, \alpha) - F(T_{j-1}, Z, \theta, \alpha) > c) = 1, \quad j = 1, \cdots, p+1, \quad \forall\ \theta \in \Theta, \alpha,$$

and the assumption (2.3.4) to

$$E[D_{00}(\mathbf{T}, Z, \theta_0, \alpha_{0n})] - \sum_{s=1}^{k} \frac{E[D_{01}(\mathbf{T}, Z, \theta_0, \alpha_{0n}) I_s(Z)]^2}{E[D_{11}(\mathbf{T}, Z, \theta_0, \alpha_{0n}) I_s(Z)]} > C.$$

The consistency rate of the sieve MLE is still $o_p(n^{-1/4})$. The proof is analogous to that in Theorem (2.3.2) except that we need to change the assumption (2.3.5) in Theorem (2.3.2) to be

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

The proof of the asymptotic normality is similar to that of Theorem (2.3.3). The asymptotic variance now is

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

and this asymptotic variance also achieves the information bound.

2.6 Simulation

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

Assume the conditional distribution function of X given Z is the Weibull distribution

$$1 - e^{-x^{\theta_0} e^{\alpha_0(z)}},$$

where $\alpha_0(z) = \log(z)$. Also assume that U and T are uniformly distributed on the upper triangle of $[1, 2] \times [1, 2]$ and Z is uniformly distributed on $[0.2, 1.2]$, and true parameter $\theta_0 = 2$.

For each fixed sample size (n=30, 60, 100, 200 respectively) and appropriate k's, k is the number of jumps in the step function and increases with the increase of the sample size n. 100 replications of the estimate of θ_0 based on the sieve maximum likelihood estimator are obtained. The means and standard deviations of these estimators thus computed are reported in the following table.

Table 2.1: *Simulation results for the Sieve MLE*

	n=30		n=60		n=100		n=200	
	mean	std	mean	std	mean	std	mean	std
k=1	1.9855	0.2673	1.8380	0.3417	2.0189	0.3150	1.9880	0.2565
k=2	2.0357	0.3969	1.9430	0.2608	1.8871	0.3106	1.8022	0.1886
k=3	2.0975	0.4351	2.0002	0.1892	1.8456	0.2399	1.7890	0.1522
k=4	2.0529	0.2704	1.8907	0.2638	1.8676	0.2185	1.7865	0.1633
k=5			2.0067	0.3790	1.8841	0.1979	1.7821	0.1715
k=6			1.9103	0.3303	1.9250	0.2121	1.9425	0.2278
k=7					1.9241	0.1781	1.9200	0.1571

From the above table we can see that the mean is around the true value for all the sample sizes and the standard deviations decreases with the increase of the sample size n and k.

2.7 Proofs

Before we go through the proofs, first we need some notation. 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 $m \times m$ matrix whose (i,j)th element is denoted by a_{ij}, then

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

Proof of Theorem (2.3.1)

Recall the definition of α_{0n} in (2.2.3). The Lipschitz continuity of α_0 implies that

$$\|\alpha_{0n} - \alpha_0\|_\infty = O(k(n)^{-1}). \tag{2.7.1}$$

Note that $\tilde{S}_n(\beta_n) = 0$ is equivalent to

$$S_n(\beta_n) := \begin{pmatrix} S_{n,\,0}(\beta_n) \\ kS_{n,\,1}(\beta_n) \\ \vdots \\ kS_{n,\,k}(\beta_n) \end{pmatrix} = 0. \tag{2.7.2}$$

The derivative of $S_n(\beta_n)$ with respect to β_n is

$$H_n(\beta_n) := \begin{pmatrix} h_{00}(\beta_n) & h_{01}(\beta_n) & \cdots & h_{0k}(\beta_n) \\ kh_{01}(\beta_n) & kh_{11}(\beta_n) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ kh_{0k}(\beta_n) & 0 & 0 & kh_{kk}(\beta_n) \end{pmatrix} \tag{2.7.3}$$

where h_{ij} is defined in (2.2.8), (2.2.9) and (2.2.10). The lower-right $k \times k$ sub-matrix is a diagonal matrix.

Let

$$\mu(\beta_n) = ES_n(\beta_n),$$

and

$$
\begin{aligned}
W_1(T,U,Z,\beta_n) &= \frac{F_{\beta_0}(T,Z) - F_{\beta_n}(T,Z)}{F_{\beta_n}(T,Z)} e^{\alpha_n(Z)},\\
W_2(T,U,Z,\beta_n) &= \Big[F_{\beta_0}(T,Z) - F_{\beta_n}(T,Z)\\
&\quad + \frac{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z) - (F_{\beta_n}(U,Z) - F_{\beta_n}(T,Z))}{F_{\beta_n}(U,Z) - F_{\beta_n}(T,Z)} \bar{F}_{\beta_n}(T,Z)\Big] e^{\alpha_n(Z)},\\
W_3(T,U,Z,\beta_n) &= \frac{F_{\beta_0}(T,Z)\bar{F}_{\beta_n}(T,Z)}{F^2_{\beta_n}(T,Z)} e^{2\alpha_n(Z)},\\
W_4(T,U,Z,\beta_n) &= \frac{[F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z)]\bar{F}_{\beta_n}(T,Z)\bar{F}_{\beta_n}(U,Z)}{[F_{\beta_n}(U,Z) - F_{\beta_n}(T,Z)]^2} e^{2\alpha_n(Z)},
\end{aligned}
$$

(2.7.4)

where $F_{\beta_0}(T,Z)$ and $F_{\beta_n}(T,Z)$ is defined in (2.2.4). Then by (2.7.2), (2.2.5), (2.2.6) and by (2.4.5) we obtain

$$
\mu(\beta_n) = \begin{pmatrix}
E(W_1(T,U,Z,\beta_n)\dot{\Lambda}(T,\theta) + W_2(T,U,Z,\beta_n)[\dot{\Lambda}(U,\theta) - \dot{\Lambda}(T,\theta)])\\
kE(W_1(T,U,Z,\beta_n)\Lambda(T,\theta)I_1(Z) + W_2(T,U,Z,\beta_n)[\Lambda(U,\theta) - \Lambda(T,\theta)]I_1(Z))\\
\vdots\\
kE(W_1(T,U,Z,\beta_n)\Lambda(T,\theta)I_k(Z) + W_2(T,U,Z,\beta_n)[\Lambda(U,\theta) - \Lambda(T,\theta)]I_k(Z))
\end{pmatrix}.
$$

(2.7.5)

By the 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 check that $\mu(\beta_0) = 0$ and $\|\mu(\beta_n)\|_\infty = o(1)$ if $\|\beta_n - \beta_{0n}\|_\infty = O(k^{-1})$ and $P(I_j(Z) = 1) = o(1)$ for $j = 1, \cdots, k$.

Let $\Sigma(\beta_n) = EH_n(\beta_n)$. By (2.7.3) and the definition of $h_{ij}(\beta_n), 0 \le i,j \le k$,

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

where

$$\begin{aligned}
b_{00}(\beta_n) &= E\Big(W_1(T,U,Z,\beta_n)\ddot{\Lambda}(T,\theta) + W_2(T,U,Z,\beta_n)[\ddot{\Lambda}(U,\theta) - \ddot{\Lambda}(T,\theta)] \\
&\quad - W_3(T,U,Z,\beta_n)\dot{\Lambda}^2(T,\theta) - W_4(T,U,Z,\beta_n)[\dot{\Lambda}(U,\theta) - \dot{\Lambda}(T,\theta)]^2\Big), \\
b_{0j}(\beta_n) &= E\Big(\Big[W_1(T,U,Z,\beta_n)\dot{\Lambda}(T,\theta) + W_2(T,U,Z,\beta_n)[\dot{\Lambda}(U,\theta) - \dot{\Lambda}(T,\theta)] \\
&\quad - W_3(T,U,Z,\beta_n) \times \dot{\Lambda}(T,\theta)\Lambda(T,\theta) - W_4(T,U,Z,\beta_n)[\dot{\Lambda}(U,\theta) - \dot{\Lambda}(T,\theta)] \\
&\quad \times [\Lambda(U,\theta) - \Lambda(T,\theta)]\Big] I_j(Z)\Big), \\
b_{jj}(\beta_n) &= E\Big(\Big[W_1(T,U,Z,\beta_n)\Lambda(T,\theta) + W_2(T,U,Z,\beta_n)[\Lambda(U,\theta) - \Lambda(T,\theta)] \\
&\quad - W_3(T,U,Z,\beta_n)\Lambda^2(T,\theta) - W_4(T,U,Z,\beta_n)[\Lambda(U,\theta) - \Lambda(T,\theta)]^2\Big] I_j(Z)\Big), \\
b_{ij}(\beta_n) &= 0 \qquad\qquad i \neq j = 1, \cdots, k,
\end{aligned}$$

with W_1 to W_4 as in (2.7.4).

The inverse of $\Sigma(\beta_n)$ is

(2.7.6)
$$\Sigma^{-1}(\beta_n) = \begin{pmatrix} q_{00} & k^{-1}q_{01} \\ q_{01}' & k^{-1}q_{11} \end{pmatrix},$$

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

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

Since $\mu(\beta_0) = 0$ by (2.7.5), and by Condition A , $\mu(\beta_n)$ is continuous in β_n, by (2.7.1),

$$\|\mu(\beta_{0n})\|_\infty = o(1). \tag{2.7.7}$$

Since $\Sigma(\beta_n)$ is continuous in β_n by Condition A, $\Sigma^{-1}(\beta_{0n})$ exists and $\|\Sigma^{-1}(\beta_{0n})\|_\infty < c$ for large n by (2.3.2), (2.3.3) and (2.3.4), it follows from (2.7.7) and the inverse function theorem (IFT) with sup-norm (Lemma 1 of Rossini and Tsiatis (1996), which is stated in the following lemma) that there exists $\tilde{\beta}_n = (\tilde{\theta}, \tilde{\alpha}_n)$, with α_n of the form (2.2.2), such that

$$\mu(\tilde{\beta}_n) = 0, \tag{2.7.8}$$

and

$$\|\tilde{\beta}_n - \beta_{0n}\|_\infty = o(1). \tag{2.7.9}$$

Next, suppose that there exists some finite constant c such that

$$\|S_n(\tilde{\beta}_n)\|_\infty = o_P(1) \tag{2.7.10}$$

$$P(\|H_n^{-1}(\tilde{\beta}_n)\|_\infty < c) \longrightarrow 1, \tag{2.7.11}$$

then by IFT again, with probability tending to 1, there exists solution $\hat{\beta}_n = (\hat{\theta}, \hat{\alpha}_n)$ of the equation $S_n(\beta_n) = 0$ such that

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

This, (2.7.9), (2.7.1), and the triangle inequality imply that

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

This completes the proof of the theorem as soon as we verify (2.7.10) and (2.7.11).

Next, we shall prove (2.7.10) and (2.7.11). To prove (2.7.10), fix a $\varepsilon > 0$, from the definition of $S_n(\beta_n)$ in (2.7.2), we observe that

(2.7.12)
$$P(\|S_n(\tilde{\beta}_n)\|_\infty > \varepsilon) \le P\left(|S_{n,\,0}(\tilde{\beta}_n)| > \frac{\varepsilon}{2}\right) + P\left(\sup_{1 \le j \le k} |kS_{n,\,j}(\tilde{\beta}_n)| > \frac{\varepsilon}{2}\right).$$

Rewrite

$$S_{n,\,0}(\tilde{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n} A_{i,\,n}$$

where $\{A_{i,\,n}\}_{1\le i\le n}$ is the ith summand in the r.h.s. of (2.2.5) with β_n replaced by $\tilde{\beta}_n$. By the Condition A and (2.3.2), there exists a constant C such that $\sup_{1 \le i \le n} |A_{i,\,n}| \le C$ for all n. And by (2.7.8) we observe that

$$EA_{i,\,n}(\tilde{\beta}_n) = ES_{n,\,0}(\tilde{\beta}_n) = 0 \quad \forall \; i = 1, \cdots, n.$$

Then Chebyschev's inequality implies that

$$P\left(|S_{n,\,0}(\tilde{\beta}_n)| > \frac{\varepsilon}{2}\right) < \frac{4C^2}{n\varepsilon^2}. \tag{2.7.13}$$

The second term in the upper bound of (2.7.12) converges to zero by Bernstein's inequality:

$$P\left(\sup_{1 \le j \le k} |kS_{n,\,j}(\tilde{\beta}_n)| > \frac{\varepsilon}{2}\right) \le k \sup_{1 \le j \le k} P\left(|kS_{n,\,j}(\tilde{\beta}_n)| > \frac{\varepsilon}{2}\right). \tag{2.7.14}$$

Here rewrite

$$S_{n,\,j}(\tilde{\beta}_n) = \frac{1}{n}\sum_{i=1}^{n} B_{i,n}$$

where $\{B_{i,\,n}\}_{1\le i\le n}$ is the ith summand in the r.h.s. of (2.2.6). Similarly, by the Condition A and (2.3.2), there exists a constant C such that $\sup_{1\,\le\, i\,\le\, n}|B_{i,\,n}|\le C$ for all n, we also have

$$EB_{i,n}^2=\sigma^2<C^2,\quad E|B_{i,n}|^p\le C^{p-2}p!EB_{i,n}^2\quad \forall\ i=1,\cdots,n,\forall\ p\ge 2.$$

By (2.7.8),

$$EB_{i,\,n}(\tilde{\beta n})=ES_{n,\,j}(\tilde{\beta n})=0\ \ \forall\ \ i=1,\cdots,n.$$

Apply Bernstein's inequality to obtain:

$$\begin{aligned}P\left(|kS_{n,\,j}(\tilde{\beta n})|>\frac{\varepsilon}{2}\right) &= P\left(|\sum_{i=1}^{n}B_{i,\,n}(\tilde{\beta n})|>\frac{n\varepsilon}{2k}\right)\\ &\le 2\exp\left(-\frac{n\varepsilon^2}{16k^2\sigma^2+4Ck\varepsilon}\right).\end{aligned} \tag{2.7.15}$$

This together with (2.7.14), we obtain

$$P\left(\sup_{1\,\le\, j\,\le\, k}|kS_{n,\,j}(\tilde{\beta n})|>\frac{\varepsilon}{2}\right)\le 2k\exp\left(-\frac{n\varepsilon^2}{16k^2\sigma^2+4Ck\varepsilon}\right). \tag{2.7.16}$$

Combine (2.7.12), (2.7.13) and (2.7.16) to obtain:

$$P(\|S_n(\tilde{\beta n})\|_\infty>\varepsilon)=O\left(\frac{1}{n}+k\exp(-\frac{n}{k^2})\right)$$

This proves (2.7.10) upon taking $k=O(n^\gamma)$ for some $0<\gamma<1/2$.

Now we shall prove (2.7.11). Since

$$H_n(\tilde{\beta n}):=\begin{pmatrix} h_{00}(\tilde{\beta n}) & h_{01}(\tilde{\beta n}) & \cdots & h_{0k}(\tilde{\beta n})\\ kh_{01}(\tilde{\beta n}) & kh_{11}(\tilde{\beta n}) & 0 & 0\\ \vdots & 0 & \ddots & 0\\ kh_{0k}(\tilde{\beta n}) & 0 & 0 & kh_{kk}(\tilde{\beta n})\end{pmatrix},$$

the inverse of $H_n(\tilde{\beta}_n)$ is

$$H_n^{-1}(\tilde{\beta}_n) = \begin{pmatrix} a_{00}(\tilde{\beta}_n) & k^{-1}a_{01}(\tilde{\beta}_n) \\ a_{01}'(\tilde{\beta}_n) & k^{-1}a_{11}(\tilde{\beta}_n) \end{pmatrix}, \tag{2.7.17}$$

where

$$a_{00}(\tilde{\beta}_n) = \left(h_{00}(\tilde{\beta}_n) - \sum_{j=1}^{k} \frac{h_{0j}^2}{h_{jj}}(\tilde{\beta}_n) \right)^{-1},$$

$a_{01}(\tilde{\beta}_n)$ is a row vector with its jth element

$$-a_{00}(\tilde{\beta}_n)\frac{h_{0j}}{h_{jj}}(\tilde{\beta}_n), \quad j = 1, 2, \cdots, k,$$

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

$$I_{(i=j)}h_{jj}^{-1}(\tilde{\beta}_n) + a_{00}(\tilde{\beta}_n)\frac{h_{0i}h_{0j}}{h_{ii}h_{jj}}(\tilde{\beta}_n), \quad j = 1, 2, \cdots, k.$$

By the definitions of $h_{ij}(\beta_n)$ and $b_{ij}(\beta_n), 0 \le i, j \le k$, and the law of large numbers for triangle arrays, we obtain

$$|h_{00}(\tilde{\beta}_n) - b_{00}(\tilde{\beta}_n)| \xrightarrow{\mathcal{P}} 0 \qquad |h_{0j}(\tilde{\beta}_n) - b_{0j}(\tilde{\beta}_n)| \xrightarrow{\mathcal{P}} 0 \qquad |h_{jj}(\tilde{\beta}_n) - b_{jj}(\tilde{\beta}_n)| \xrightarrow{\mathcal{P}} 0$$

$0 \le j \le k$. Together with the definitions of a_{00}, a_{01}, a_{11} and q_{00}, q_{01}, q_{11}, we get

$$\begin{aligned} &|a_{00}(\tilde{\beta}_n) - q_{00}(\tilde{\beta}_n)| \xrightarrow{\mathcal{P}} 0 \\ &||a_{01}(\tilde{\beta}_n) - q_{01}(\tilde{\beta}_n)|| \xrightarrow{\mathcal{P}} 0 \\ &||a_{11}(\tilde{\beta}_n) - q_{11}(\tilde{\beta}_n)|| \xrightarrow{\mathcal{P}} 0. \end{aligned} \tag{2.7.18}$$

By the facts that $\Sigma^{-1}(\beta_n)$ is continuous, $\|\Sigma^{-1}(\beta_{0n})\|_\infty < c$, and $\|\tilde{\beta}_n - \beta_{0n}\|_\infty = o(1)$, we obtain $\|\Sigma^{-1}(\tilde{\beta}_n)\|_\infty < c$, this together with (2.7.6) , (2.7.17) and (2.7.18), we proved $\|H_n^{-1}(\tilde{\beta}_n)\|_\infty < c$ with probability approaching 1.

Lemma 2.7.1 *(Inverse Function Theorem with Sup-norm). Let $A(x)$ 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 $H(x) = \partial A(x)$ (derivatives of the elements of H with respect to the elements of x). If there exists constant C and δ^* such that*

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

and

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

then for $d < \delta^*/(4C)$ and all y such that $||y - A(x_0)||_\infty < d$, there exists a unique inverse value x in the δ^* neighborhood of x_0 such that $A(x) = y$ and $||x - x_0|| < 4Cd$.(Rossini and Tsiatis 1996)

Proof of Theorem (2.3.2)

We are going to use some general results on the convergence rate of sieve estimations. The following lemma is a part of Theorem 1 of Shen and Wong (1994). To state the lemma, we introduce some general notation. Let $Y_1, \cdots, Y_n$ be a sequence of independent random variables (or possible 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 $\ell : \Theta \times \mathcal{Y} \to \mathcal{R}$ be a suitably chosen function. We are interested in the properties of an estimator $\hat{\beta}_n$ over a subset Θ_n of Θ by maximizing the empirical criterion $C_n(\beta) = \frac{1}{n}\sum_{i=1}^n \ell(\beta, Y_i)$, that is, $C_n(\hat{\beta}_n) = max_{\beta \in \Theta_n} C_n(\beta)$. 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 additional assumptions are needed for the lemma.

C0. ℓ is bounded.

C1. For a given β_0, $\exists$ constants $A_1 > 0$ and $a > 0$, such that for all small $\epsilon > 0$,

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

C2. For a given β_0, $\exists$ constants $A_2 > 0$ and $b > 0$, such that for all small $\epsilon > 0$,

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

C3. Let $\mathcal{Q}_n = \{\ell(\beta, \cdot) - \ell(\pi_n\beta_0, \cdot) : \beta \in \Theta_n\}$. For some constant $r_0 < 1/2$ and $A_3 > 0$,

$$H(\epsilon, \mathcal{Q}_n) \leq A_3 n^{2r_0} log(\frac{1}{\epsilon}) \qquad \text{for all small } \epsilon > 0.$$

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

Lemma 2.7.2 *Suppose Assumptions C0 to C3 hold. Then*

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

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

$$\tau = \begin{cases} \dfrac{1-2r_0}{2a} - \dfrac{\log\log n}{2a\log n}, & \text{if } b \geq a, \\ \dfrac{1-2r_0}{4a-2b}, & \text{if } b < a. \end{cases}$$

From the proof of Theorem 1 of Shen and Wong (1994), it is noted that the global 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 (2.7.2) to our case, let $Y = (T, U, Z, \delta, \gamma)$, $\beta = (\theta, \alpha)$, $\pi_n\beta = (\theta, \alpha_n)$ where α_n is of the form (2.2.2) 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 β_n are chosen such that, with probability approaching 1, the MLE $(\hat{\theta}, \hat{\alpha_n})$ is in Θ_n. Define the metric as follows

$$\rho(\beta, \beta_0) = |\theta - \theta_0| + ||\alpha - \alpha_0||_\infty, \tag{2.7.19}$$

and let

$$\begin{aligned} \ell(\beta, Y) &= \Big\{\delta \log(1 - e^{-\Lambda(T,\theta)e^{\alpha(Z)}}) + \gamma \log[e^{-\Lambda(T,\theta)e^{\alpha(Z)}} - e^{-\Lambda(U,\theta)e^{\alpha(Z)}}] \\ &\quad -(1 - \delta - \gamma)\Lambda(U,\theta)e^{\alpha(Z)}\Big\}. \end{aligned}$$

We shall now verify the conditions C0 - C3 of Lemma (2.7.2) for this ℓ. Note that in the proof below we always denote C as some finite and positive number. Under our assumptions, C0 holds. Note that

$$\begin{aligned} &E\ell(\beta, Y) \\ &= E\Big\{(1 - e^{-\Lambda(T,\theta_0)e^{\alpha_0(Z)}})\log(1 - e^{-\Lambda(T,\theta)e^{\alpha(Z)}}) \\ &\quad +[e^{-\Lambda(T,\theta_0)e^{\alpha_0(Z)}} - e^{-\Lambda(U,\theta_0)e^{\alpha_0(Z)}}]\log[e^{-\Lambda(T,\theta)e^{\alpha(Z)}} - e^{-\Lambda(U,\theta)e^{\alpha(Z)}}] \\ &\quad -e^{-\Lambda(U,\theta_0)e^{\alpha_0(Z)}}\Lambda(U,\theta)e^{\alpha(Z)}\Big\}. \end{aligned}$$

The Taylor expansion of $\ell(\beta, Y)$ with respect to θ and α around (θ_0, α_0), and the fact that the expectation of the first derivative of $\ell(\beta, Y)$ w.r.t β vanishes at β_0 and the matrix of the second derivatives is negative definite by (2.3.5), we obtain

$$E(\ell(\beta_0, Y) - \ell(\beta, Y)) \geq c\rho^2(\beta, \beta_0), \tag{2.7.20}$$

for some finite and positive number c. Hence the condition C1 is satisfied with $a = 1$. Note that the Condition A implies that for all y

$$|\ell(\beta_0, y) - \ell(\beta, y)| \leq C(|\theta - \theta_0| + ||\alpha - \alpha_0||_\infty). \tag{2.7.21}$$

Hence

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

Thus the condition C2 holds with $b = 1$.

By (2.7.21), we also have

$$H(\epsilon, \mathcal{Q}_n) \leq H(\epsilon/C, \Theta_n), \tag{2.7.22}$$

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 $||\alpha_n|| \leq C_0$, $(\theta, \alpha_n) \in \Theta_n$, and α_n is of the form (2.2.2). For any $\eta > 0$, divide the interval $[0, C_0]$ into small intervals, with length at most $\eta/2$, such that the number of intervals is less than or equal to $2C_0/\eta + 1$. Then it is easy to see that

$$H(\eta, \Theta_n) \leq \log \left(\left(\frac{2C_0}{\eta} + 1 \right) \left(\frac{2C_0}{\eta} + 1 \right)^{k(n)} \right) \leq Ck(n) \log \left(\frac{1}{\eta} \right), \tag{2.7.23}$$

as η is small enough. Hence, by (2.7.22) and (2.7.23), for all small $\epsilon > 0$,

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

Therefore C3 is satisfied with $\gamma_0 = \frac{\gamma}{2}$.

Thus lemma (2.7.2) is applicable to this ℓ, which in turn yields that

$$\rho(\hat{\beta}_n, \beta_0) = O_P(\max(n^{-\tau}, \rho(\pi_n \beta_0, \beta_0), K^{1/2}(\pi_n \beta_0, \beta_0))), \tag{2.7.24}$$

where

$$\tau = \frac{1-\gamma}{2} - \frac{\log \log n}{2 \log n}.$$

Note that, for large n, $1/4 < \gamma < 1/2$ implies that $1/4 < \tau < 3/8$. Since $\beta_0 = (\theta_0, \alpha_0)$, $\pi_n \beta_0 = (\theta_0, \alpha_{0n})$, where α_{0n} is of the form (2.2.2), by (2.7.19) and (4) of the Condition A, we obtain that

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

Thus

$$\rho(\pi_n\beta_0, \beta_0) \leq Cn^{-\gamma}. \tag{2.7.25}$$

The same argument as that leading to (2.7.20) gives that,

$$K(\pi_n\beta_0, \beta_0) = E(\ell(\beta_0, Y) - \ell(\pi_n\beta_0, Y)) \leq C||\alpha_{0n} - \alpha_0||^2 = Cn^{-2\gamma}, \tag{2.7.26}$$

which is of order between $o(n^{-1/2})$ and $o(n^{-1})$ for $1/4 < \gamma < 1/2$. It follows then from (2.7.24), (2.7.25) and (2.7.26) that for $1/4 < \gamma < 1/2$,

$$\rho(\hat{\beta}_n, \beta_0) = o_P(n^{-1/4}),$$

thereby complete proving Theorem (2.3.2).

Proof of Theorem (2.3.3)

Recall the definition of $S_{n,0}(\theta, \alpha)$ from (2.2.5). Furthermore let a be a measurable function on $\mathcal{Z}$ with $Ea^2(\mathcal{Z}) < \infty$,

$$\begin{aligned} S_n(\theta,\alpha)[a] \ = \ & \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i - F_\beta(T_i,Z_i)}{F_\beta(T_i,Z_i)}\Lambda(T_i,\theta)e^{\alpha(Z_i)}a(Z_i) \\ & + \left[\delta_i - F_\beta(T_i,Z_i) + \frac{\gamma_i - [F_\beta(U_i,Z_i) - F_\beta(T_i,Z_i)]}{F_\beta(U_i,Z_i) - F_\beta(T_i,Z_i)}\bar{F}_\beta(T_i,Z_i)\right] \\ & \times [\Lambda(U_i,\theta) - \Lambda(T_i,\theta)]e^{\alpha(Z_i)}a(Z_i)\Big\} \end{aligned}$$

where $F_\beta(t,z)$ is defined in (2.2.4).

Denote the expectation of $S_{n,0}(\theta,\alpha)$ and $S_n(\theta.\alpha)[a]$ by $\mu_0(\theta,\alpha)$ and $\mu(\theta.\alpha)[a]$ respectively. By (2.4.5), we obtain

$$\begin{aligned} \mu_0(\theta,\alpha) = E\Big\{ & \frac{F_{\beta_0}(T,Z) - F_\beta(T,Z)}{F(T,Z)}\dot{\Lambda}(T,\theta)e^{\alpha(Z)} \\ & + \Big[F_{\beta_0}(T,Z) - F_\beta(T,Z) \\ & + \frac{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z) - [F_\beta(U,Z) - F_\beta(T,Z)]}{F_\beta(U,Z) - F_\beta(T,Z)}\bar{F}_\beta(T,Z)\Big] \\ & \times [\dot{\Lambda}(U,\theta) - \dot{\Lambda}(T,\theta)]e^{\alpha(Z)}\Big\} \end{aligned} \tag{2.7.27}$$

and

$$\mu(\theta,\alpha)[a] = E\Big\{\frac{F_{\beta_0}(T,Z) - F_\beta(T,Z)}{F_\beta(T,Z)}\Lambda(T,\theta)e^{\alpha(Z)}a(Z)$$

$$+ \Big[F_{\beta_0}(T,Z) - F_\beta(T,Z)$$

$$\frac{F_{\beta_0}(U,Z) - F_{\beta_0}(T,Z) - [F_\beta(U,Z) - F_\beta(T,Z)]}{F_\beta(U,Z) - F_\beta(T,Z)}\bar{F}(T,Z)\Big]$$

$$\times\, [\Lambda(U,\theta) - \Lambda(T,\theta)]e^{\alpha(Z)}a(Z)\Big\}, \tag{2.7.28}$$

where $F_{\beta_0}(t,z)$ and $F_\beta(t,z)$ are defined in (2.2.4).

The method used here is similar to that described in Huang (1996). From Lemma (2.7.3) below, we obtain the following stochastic equi-continuity results, for every $0 < C < \infty$, $\forall$ a with $Ea^2(Z) < \infty$,

$$\sup_{\substack{|\theta-\theta_0|\le Cn^{-1/4},\\ ||\alpha-\alpha_0||\le Cn^{-1/4}}} |S_{n,0}(\theta,\alpha) - \mu_0(\theta,\alpha) - (S_{n,0}(\theta_0,\alpha_0) - \mu_0(\theta_0,\alpha_0))| = o_p^*(n^{-1/2})$$

$$\sup_{\substack{|\theta-\theta_0|\le Cn^{-1/4},\\ ||\alpha-\alpha_0||\le Cn^{-1/4}}} |S_n(\theta,\alpha)[a] - \mu(\theta,\alpha)[a] - (S_n(\theta_0,\alpha_0)[a] - \mu(\theta_0,\alpha_0)[a])| = o_p^*(n^{-1/2}).$$

This and Theorem (2.3.2) yields

$$S_{n,0}(\hat\theta,\hat\alpha_n) - \mu_0(\hat\theta,\hat\alpha_n) - (S_{n,0}(\theta_0,\alpha_0) - \mu_0(\theta_0,\alpha_0)) = o_p(n^{-1/2}), \tag{2.7.29}$$

$$S_n(\hat\theta,\hat\alpha_n)[a^*] - \mu(\hat\theta,\hat\alpha_n)[a^*] - (S_n(\theta_0,\alpha_0)[a^*] - \mu(\theta_0,\alpha_0)[a^*]) = o_p(n^{-1/2}), \tag{2.7.30}$$

where a^* is defined in (2.4.7).

By the definition of $\hat{\theta}$ and $\hat{\alpha}_n$, $S_{n,0}(\hat{\theta}, \hat{\alpha}_n) = 0$. Also note that $\mu_0(\theta_0, \alpha_0) = 0$ by (2.7.27). It thus follows from (2.7.29) that

$$\mu_0(\hat{\theta}, \hat{\alpha}_n) = -S_{n,0}(\theta_0, \alpha_0) + o_P(n^{-1/2}). \tag{2.7.31}$$

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

$$S_n(\hat{\theta}, \hat{\alpha}_n)[a^*] = o_P(n^{-1/2}). \tag{2.7.32}$$

Together with $\mu(\theta_0, \alpha_0)[a^*] = 0$ by (2.7.28), we obtain from (2.7.30) that

$$\mu(\hat{\theta}, \hat{\alpha}_n)[a^*] = -S_n(\theta_0, \alpha_0)[a^*] + o_P(n^{-1/2}). \tag{2.7.33}$$

By the Condition A and that the third derivative of $\Lambda(t, \theta)$ with respect to θ exists and is continuous, the Taylor expansion of $\mu_0(\hat{\theta}, \hat{\alpha}_n)$ up to the second order with respect to θ and α_n, together with (2.7.31) and (2.7.27) yields that

$$-E[D_{00}(T, U, Z, \theta_0, \alpha_0)](\hat{\theta} - \theta_0) - E[D_{01}(T, U, Z, \theta_0, \alpha_0)](\hat{\alpha}_n(Z) - \alpha_0(Z))$$

$$= -S_{n,0}(\theta_0, \alpha_0) + O_P(|\hat{\theta} - \theta_0|^2 + ||\hat{\alpha}_n - \alpha_0||_\infty^2) + o_P(n^{-1/2}). \tag{2.7.34}$$

Similarly we can obtain from (2.7.33) and (2.7.28) that

$$-E[D_{01}(T, U, Z, \theta_0, \alpha_0)a^*(Z)](\hat{\theta} - \theta_0) - E[D_{11}(T, U, Z, \theta_0, \alpha_0)a^*(Z)](\hat{\alpha}_n(Z) - \alpha_0(Z))$$

$$= -S_n(\theta_0, \alpha_0)[a^*] + O_P(|\hat{\theta} - \theta_0|^2 + ||\hat{\alpha}_n - \alpha_0||_\infty^2) + o_P(n^{-1/2}). \tag{2.7.35}$$

By Theorem (2.3.2),

$$|\hat{\theta} - \theta_0|^2 + ||\hat{\alpha}_n - \alpha_0||_\infty^2 = o_P(n^{-1/2}).$$

Subtracting (2.7.34) from (2.7.35) and noticing the definition of a^*, we obtain

$$E[D_{00}(T, U, Z, \theta_0, \alpha_0) - D_{01}(T, U, Z, \theta_0, \alpha_0)a^*(Z)](\hat{\theta} - \theta_0)$$

$$= S_{n,0}(\theta_0, \alpha_0) - S_n(\theta_0, \alpha_0)[a^*] + o_P(n^{-1/2}).$$

The theorem follows from the central limit theorem and the calculation of the variance is straightforward.

Now we prove (2.7.32). Let

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

Condition A implies that

$$\|a_n^* - a^*\|_\infty = O(1/k(n)). \tag{2.7.36}$$

By the definition of $(\hat{\theta}, \hat{\alpha}_n)$, that is, it solves (2.7.2), we obtain

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

Thus to prove (2.7.32), it suffices to show

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

The left hand side of above equation is

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - F_{\hat{\beta}_n}(T_i, Z_i)}{F_{\hat{\beta}_n}(T_i, Z_i)} \Lambda(T_i, \hat{\theta}) e^{\hat{\alpha}_n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)] \\ &+\frac{1}{n}\sum_{i=1}^{n} \left[\delta_i - F_{\hat{\beta}_n}(T_i, Z_i) + \frac{\gamma_i - [F_{\hat{\beta}_n}(U_i, Z_i) - F_{\hat{\beta}_n}(T_i, Z_i)]}{F_{\hat{\beta}_n}(U_i, Z_i) - F_{\hat{\beta}_n}(T_i, Z_i)} \bar{F}_{\hat{\beta}_n}(T_i, Z_i)\right] \\ &\times \qquad \times [\Lambda(U_i, \hat{\theta}) - \Lambda(T_i, \hat{\theta})] e^{\hat{\alpha}_n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)] \\ &\stackrel{def}{=} \; I + II, \end{aligned}$$

where

$$I = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i - F_{\hat{\beta}_n}(T_i, Z_i)}{F_{\hat{\beta}_n}(T_i, Z_i)} \Lambda(T_i, \hat{\theta}) e^{\hat{\alpha}_n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)],$$

$$II = \frac{1}{n}\sum_{i=1}^{n}\left[\delta_i - F_{\hat\beta_n}(T_i,Z_i) + \frac{\gamma_i - [F_{\hat\beta_n}(U_i,Z_i) - F_{\hat\beta_n}(T_i,Z_i)]}{F_{\hat\beta_n}(U_i,Z_i) - F_{\hat\beta_n}(T_i,Z_i)}\bar F_{\hat\beta_n}(T_i,Z_i)\right]$$
$$\times[\Lambda(U_i,\hat\theta) - \Lambda(T_i,\hat\theta)]e^{\hat\alpha n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)].$$

Rewrite I as

$$\begin{aligned} I &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i - F_{\beta_0}(T_i,Z_i)}{F_{\hat\beta_n}(T_i,Z_i)}\Lambda(T_i,\hat\theta)e^{\hat\alpha n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)] \\ &+\frac{1}{n}\sum_{i=1}^{n}\frac{F_{\beta_0}(T_i,Z_i) - F_{\hat\beta_n}(T_i,Z_i)}{F_{\hat\beta_n}(T_i,Z_i)}\Lambda(T_i,\hat\theta)e^{\hat\alpha n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)]. \end{aligned}$$

By (2.7.36), Theorem (2.3.2) and the Lipschitz continuity of F with respect to θ and α by Condition A, the second term is $o_p(n^{-1/2})$. For the first term, by Theorem (2.3.2) and Condition A, we can write it as

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i - F_{\beta_0}(T_i,Z_i)}{F_{\beta_0}(T_i,Z_i)}\frac{F_{\beta_0}(T_i,Z_i)}{F_{\hat\beta_n}(T_i,Z_i)}\Lambda(T_i,\hat\theta)e^{\hat\alpha n(Z_i)}[a^*(Z_i) - a_n^*(Z_i)] \\ = &\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i - F_{\beta_0}(T_i,Z_i)}{F_{\beta_0}(T_i,Z_i)}\left(1 + Op(|\hat\theta - \theta_0| + ||\hat\alpha_n - \alpha_0||)\right)\left(\Lambda(T_i,\theta_0) + Op(|\hat\theta - \theta_0|)\right) \\ &\left(e^{\alpha_0(Z_i)} + Op(||\hat\alpha_n - \alpha_0||)\right)[a^*(Z_i) - a_n^*(Z_i)]. \end{aligned}$$

This is $o_p(n^{-1/2})$ by Central Limit Theorem, Theorem (2.3.2) and (2.7.36). A similar argument yields that II is also $o(n^{-1/2})$. This complete the proof of (2.7.32).

Lemma 2.7.3 *Under the condition of Theorem (2.3.3), for any positive and finite number C, and for function a on $\mathcal{Z}$ with $Ea^2(Z) < \infty$,*

$$\sup_{\substack{|\theta-\theta_0|\le Cn^{-1/4}\\ ||\alpha-\alpha_0||\le Cn^{-1/4}}} |\sqrt{n}(S_{n,0}(\theta,\alpha) - \mu_0(\theta,\alpha)) - \sqrt{n}(S_{n,0}(\theta_0,\alpha_0) - \mu_0(\theta_0,\alpha_0))|$$
$$o_p^*(1),$$
$$\sup_{\substack{|\theta-\theta_0|\le Cn^{-1/4}\\ ||\alpha-\alpha_0||\le Cn^{-1/4}}} |\sqrt{n}(S_n(\theta,\alpha)[a] - \mu_0(\theta,\alpha)[a]) - \sqrt{n}(S_n(\theta_0,\alpha_0)[a] - \mu_0(\theta_0,\alpha_0)[a])|$$
$$= o_p^*(1). \tag{2.7.37}$$

Proof We shall only prove the first part since the second one can be proved similarly. Note that $\sqrt{n}(S_{n,\,0}(\theta,\alpha)-\mu_0(\theta,\alpha))-\sqrt{n}(S_{n,\,0}(\theta_0,\alpha_0)-\mu_0(\theta_0,\alpha_0))$ is an empirical processes indexed by functions belong to the class

$$
\begin{aligned}
\mathcal{C} \;=\; & \Big\{f(\delta,\gamma,t,u,z,\theta,\alpha)=\frac{\delta-F_\beta(t,z)}{F_\beta(t,z)}\dot\Lambda(t,\theta)e^{\alpha(z)}-\frac{\delta-F_{\beta_0}(t,z)}{F_{\beta_0}(t,z)}\dot\Lambda(t,\theta_0)e^{\alpha_0(z)}\\
&+[\delta-F_\beta(t,z)+\frac{\gamma-[F_\beta(u,z)-F_\beta(t,z)]}{F_\beta(u,z)-F_\beta(t,z)}\bar F_\beta(t,z)]\times[\dot\Lambda(u,\theta)-\dot\Lambda(t,\theta)]e^{\alpha(z)}\\
&-[\delta-F_{\beta_0}(t,z)+\frac{\gamma-[F_{\beta_0}(u,z)-F_{\beta_0}(t,z)]}{F_{\beta_0}(u,z)-F_{\beta_0}(t,z)}\bar F_{\beta_0}(t,z)]\times[\dot\Lambda(u,\theta_0)-\dot\Lambda(t,\theta_0)]e^{\alpha_0(z)}:\\
&|\theta-\theta_0|\le Cn^{-1/4},||\alpha-\alpha_0||\le Cn^{-1/4}\Big\}, \qquad (2.7.38)
\end{aligned}
$$

that is, by the functional notation used in van der Vaart and Wellner (1996) for the empirical processes,

$$
\begin{aligned}
&\sqrt{n}(S_{n,\,0}(\theta,\alpha)-\mu_0(\theta,\alpha))-\sqrt{n}(S_{n,\,0}(\theta_0,\alpha_0)-\mu_0(\theta_0,\alpha_0))\\
&=\sqrt{n}(P_n-P)f(\delta,\gamma,t,u,z,\theta,\alpha), \qquad (2.7.39)
\end{aligned}
$$

where P_n is the empirical measure based on $(\delta_i,\gamma_i,T_i,U_i,Z_i), i=1,\cdots,n$ and P is the probability measure of (δ,γ,T,U,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

$$
|f(\delta,\gamma,t,u,z,\theta,\alpha)-f(\delta,\gamma,t,u,z,\theta_0,\alpha_0)|\le C_0(|\theta-\theta_0|+||\alpha-\alpha_0||_\infty), \qquad (2.7.40)
$$

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

$$
\mathcal{D}=\Big\{(\theta-\theta_0,\alpha-\alpha_0):\alpha \text{ is of form } (2.2.2),|\theta-\theta_0|\le Cn^{-1/4},||\alpha-\alpha_0||\le Cn^{-1/4}\Big\}
$$

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. When we prove Theorem (2.3.2), we already obtained

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

for some finite and positive number C_1. Hence

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

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

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

This and the fact that $k(n) = n^\gamma$, with $0 < \gamma < 1/2$, in turns imply that

$$J_{[\ ]}(Cn^{-1/4}, \mathcal{C}, L_2(P)) = o(1). \tag{2.7.41}$$

Note that $f(\delta, \gamma, t, u, z, \theta_0, \alpha_0) = 0$ by (2.7.38). This fact and (2.7.40) imply that, for any $f \in \mathcal{C}$,

$$P(f^2(\delta, \gamma, t, u, z, \theta, \alpha)) \leq C_4 n^{-1/2}, \tag{2.7.42}$$

for some finite and positive number C_4.

Apply Lemma 3.4.2 of van der Vaart and Wellner (1996), which is stated in the following Lemma (2.7.4). Let $Y_i = (\delta_i, \gamma_i, T_i, U_i, Z_i), i = 1, \cdots, n, \mathcal{F} = \mathcal{C}$ and $\epsilon = Cn^{-1/4}$. By (2.7.42) and f is bounded, $f \in \mathcal{C}$, the conditions of the lemma hold. It follows from the lemma and (2.7.41) that

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

In view of (2.7.39), we obtain (2.7.37). The lemma is proved.

Lemma 2.7.4 *Let $Y_1, Y_2, \cdots, Y_n$ be i.i.d. random variables (or possible 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*

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

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

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

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

Chapter 3

Model Check

3.1 Introduction and Main Results

The purpose of this chapter is to develop tests of lack-of-fit of a regression model when the response variable is subject to interval censoring case 1. Now let Y^0 denote the times of the onset of an event and T^0 the time of inspection. Suppose, additionally, one is interested in assessing the effect of a covariate Z on the time of the onset of the event, for example one may wish to asses the effect of the age of a patient on the time of onset of a disease in the patient. One way to proceed is to use the classical regression analysis where one regresses $Y := \log Y^0$ on Z but one observes only (δ, T) with $T = \log T^0$. But then the question of which regression model to chose from a possible class of models becomes relevant.

More precisely, assume Y has finite expectation and let $\mu(z) := E(Y|Z = z)$ denote the regression function. Let $\mathcal{M} = \{m_\theta(z) : z \in \mathbb{R}, \theta \in \Theta\}$ be a given parametric family of functions, where Θ is a subset of the q-dimensional Euclidean space $\mathbb{R}^q$. This class of functions represents a possible class of regression models and the problem of interest is to test the hypothesis

$$H_0: \quad \mu(z) = m_{\theta_0}(z), \quad \text{for some } \theta_0 \in \Theta, \ \forall\, z \in \mathbb{R},$$

based on n i.i.d. observations $X_i = (\delta_i, T_i, Z_i)$, $1 \leq i \leq n$ on (δ, T, Z), where $\delta = I(Y \leq T)$. The alternative of interest is that H_0 is not true.

In the case Y_i's are fully observable tests for the lack-of-fit hypothesis H_0 have been based on the marked residual empirical process

$$(3.1.1) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - m_{\theta_n}(Z_i)}{s_{\theta_n}(Z_i)} I(Z_i \leq z), \quad z \in \mathbb{R},$$

where θ_n is $\sqrt{n}$-consistent estimator of θ_0 under the null hypothesis, and $s^2_{\theta}(z)$ is the conditional variance of $Y - m_{\theta}(Z)$, given $Z = z$, under H_0, cf., An and Cheng (1991), Stute (1997), and Stute, Thies and Zhu (1998), among others. The last paper showed that the tests based on its innovation martingale transforms are asymptotically distribution free.

Our focus here is to develop an analog of this transformation for the current status response data when the error distribution is known. Since the Y_i's are not observable, we need to replace them in (3.1.1) by $\hat{Y}_i$, a copy of

$$(3.1.2) \qquad \hat{Y} = E(Y|\delta, T, Z) = \delta E(Y|\delta, T, Z) + (1 - \delta) E(Y|\delta, T, Z).$$

To proceed further, let F denote the d.f. of the error $\varepsilon := Y - \mu(Z)$. Assume that

(3.1.3) F is continuous, $0 < F(y) < 1$, for all $y \in \mathbb{R}$, $E\varepsilon = 0$, $E\varepsilon^2 < \infty$.

ε is conditionally independent of T, given Z, and T is independent of Z.

Then, with $\bar{F} := 1 - F$, we obtain,

$$
\begin{aligned}
(3.1.4)\ E(Y|\delta = 1, T = t, Z = z) &= \frac{\int_{-\infty}^{t} y\, dF(y - \mu(z))}{F(t - \mu(z))} \\
&= \frac{\int_{-\infty}^{t - \mu(z)} y\, dF(y)}{F(t - \mu(z))} + \mu(z), \\
E(Y|\delta = 0, T = t, Z = z) &= \frac{\int_{t}^{\infty} y\, dF(y - \mu(z))}{1 - F(t - \mu(z))} \\
&= \frac{\int_{t - \mu(z)}^{\infty} y\, dF(y)}{\bar{F}(t - \mu(z))} + \mu(z), \qquad t, z \in \mathbb{R}.
\end{aligned}
$$

Let

$$
\begin{aligned}
R(\delta, t, z) &:= E(Y|\delta, T = t, Z = z) - \mu(z), \qquad \nu(z) := \int_{-\infty}^{z} y\, dF(y) \\
\sigma^2(z) &:= Var(R(\delta, T, Z)|Z = z), \qquad t, z \in \mathbb{R}.
\end{aligned}
$$

From (3.1.2), (3.1.4) and the fact $\nu(\infty) = 0$, we obtain

$$
\begin{aligned}
R(\delta, t, z) &= \delta \frac{\int_{-\infty}^{t - \mu(z)} y\, dF(y)}{F(t - \mu(z))} + (1 - \delta) \frac{\int_{t - \mu(z)}^{\infty} y\, dF(y)}{\bar{F}(t - \mu(z))} \\
&= \nu(t - \mu(z))[\frac{\delta}{F(t - \mu(z))} - \frac{1 - \delta}{\bar{F}(t - \mu(z))}] \\
&= \frac{\nu(t - \mu(z))[\delta - F(t - \mu(z))]}{F(t - \mu(z))\bar{F}(t - \mu(z))}.
\end{aligned}
$$

By the conditional independence of ε and T, given Z, $E\{R(\delta, T, Z)|T, Z\} = 0$ and

$$
\begin{aligned}
(3.1.5) \quad 0 < \sigma^2(z) &= E\Big\{\frac{\nu^2(T - \mu(z))}{F(T - \mu(z))\bar{F}(T - \mu(z))}\Big\} < \infty, \qquad \forall\, z \in \mathbb{R}, \\
E\sigma^2(Z) &< \infty, \qquad \text{by the assumption } E\varepsilon^2 < \infty.
\end{aligned}
$$

The entities $R(\delta_i, T_i, Z_i)/\sigma(Z_i)$ play the role of the standardized residuals in the current status data.

To test the simple hypothesis $\tilde{H} : \mu(z) = \mu_0(z)$, $z \in \mathbb{R}$, where μ_0 is a known function, the analogue of the process (3.1.1) suitable here would be

$$
V_n^0(z) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{R_0(\delta_i, T_i, Z_i)}{\sigma_0(Z_i)} I(Z_i \le z), \qquad z \in \mathbb{R},
$$

where R_0, σ_0 are the above R, σ functions with μ replaced by μ_0.

Let H and G denote the d.f.'s of T and Z, respectively, and B denote the standard Brownian motion on $[0, \infty)$. Using an argument similar to one used in Stute (1997), it can be verified that under (3.1.3),

$$V_n^0 \Longrightarrow B \circ G, \qquad D[-\infty, \infty], \text{ in uniform metric.}$$

Thus, for example the test that would reject $\tilde{H}$ whenever $K_n := \sup_{z \in \mathbb{R}} |V_n^0(z)| > b_\alpha$, where b_α is $100(1-\alpha)$th percentile of the distribution of $\sup_{0 \le t \le 1} |B(t)|$ would have the asymptotic size α.

To discuss the more interesting problem of testing H_0, we proceed as follows. For convenience, let P_θ denote the joint distribution of (δ, T, Z) when $\mu = m_\theta$, and E_θ and Var_θ denote the corresponding mean and variance operations. Let R_θ, σ_θ stand for R, σ when $\mu = m_\theta$ and θ_n denote a $\sqrt{n}$-consistent estimator of θ_0, under H_0, based on $(\delta_i, T_i, Z_i); 1 \le i \le n$. Tests of H_0 will be based on the process $\tilde{V}_n(z) := V_n(z, \theta_n)$, where

$$V_n(z, \theta) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{R_\theta(\delta_i, T_i, Z_i)}{\sigma_\theta(Z_i)} I(Z_i \le z), \qquad z \in \mathbb{R}, \theta \in \Theta.$$

To analyze asymptotic behavior of $\tilde{V}_n$, we need to make the following assumptions.

(3.1.6) $\quad \sqrt{n}\|\theta_n - \theta_0\| = O_p(1), \quad (P_{\theta_0})$

(3.1.7) $\quad$ The d.f. F has a continuous density f and m_θ is differentiable in a neighborhood of θ_0 with its $q \times 1$ vector of derivative $\dot{m}_{\theta_0}$, so that there exists a family of $q \times 1$ vectors of functions $g_{\theta_0}(t, z, \delta)$, $z \in \mathbb{R}$, $t \in \mathbb{R}$, $\theta \in \Theta^q$, which is the derivative of $\dfrac{R_\theta(t, z, \delta)}{\sigma_\theta(z)}$ with respect to θ at $\theta = \theta_0$, such that $\forall\ 0 < b < \infty$

$$\sup_{\substack{\sqrt{n}\|\theta - \theta_0\| \le b, \\ 1 \le i \le n}} \sqrt{n} \left| \frac{R_\theta(T_i, Z_i, \delta_i)}{\sigma_\theta(Z_i)} - \frac{R_{\theta_0}(T_i, Z_i, \delta_i)}{\sigma_{\theta_0}(Z_i)} - (\theta - \theta_0) g_{\theta_0}(T_i, Z_i, \delta_i) \right| = o_p(1),$$

$$(3.1.8) \qquad E_{\theta_0}\|g_{\theta_0}(T_i, Z_i, \delta_i)\|^2 < \infty.$$

Let $\dot{\nu}(x)$ denote the derivative of $\nu(x)$ with respect to $x \in \mathbb{R}$ and $\dot{R}_\theta$ denote the vector of the first derivatives of R_θ with respect to θ. Direct calculations show that with $x = t - m_\theta(z)$,

$$\dot{R}_\theta(\delta, t, z) = \left[\frac{-\dot{\nu}(x)[\delta - F(x)] + \nu(x)f(x)}{F(x)\bar{F}(x)} + \nu(x)[\delta - F(x)]\frac{f(x)(1 - 2F(x))}{(F(x)\bar{F}(x))^2}\right]\dot{m}_\theta(z).$$

Let

$$\begin{aligned} h_\theta(z) &:= \frac{1}{\sigma_\theta(z)}E_\theta \dot{R}_\theta(\delta, T, z), \\ \ell_\theta(z, H) &:= E\left(\frac{\nu(T - m_\theta(z))\, f(T - m_\theta(z))}{(F \cdot \bar{F})(T - m_\theta(z))}\right), \qquad z \in \mathbb{R},\, \theta \in \Theta \end{aligned}$$

Use a conditioning argument and the independence of ε and T, given Z, to obtain

$$E_{\theta_0}\dot{R}_{\theta_0}(\delta, T, z) = \ell_{\theta_0}(z, H)\dot{m}_{\theta_0}(z), \quad h_{\theta_0}(z) = \ell_{\theta_0}(z, H)\frac{\dot{m}_{\theta_0}(z)}{\sigma_{\theta_0}(z)}.$$

Next, let

$$D_\theta(z) := E_\theta\Big\{\frac{\dot{R}_\theta(\delta, T, Z)}{\sigma_\theta(Z)}I(Z \le z)\Big\}, \qquad z \in \mathbb{R}, \theta \in \Theta.$$

Note that the independence of T and Z enables one to write

$$(3.1.9) \qquad D_{\theta_0}(z) = \int_{-\infty}^{z} h_{\theta_0}(u)\, dG(u), \qquad z \in \mathbb{R}.$$

Arguing as in Stute (1997), which uses a standard Taylor expansion, a Glivenko-Cantelli type of an argument and a weak convergence argument, we obtain the following

Theorem 3.1.1. *Under the assumptions (3.1.3), (3.1.6) - (3.1.8), we obtain that uniformly in $z \in \mathbb{R}$, under P_{θ_0},*

$$\tilde{V}_n(z) = V_n(z, \theta_0) + n^{1/2}(\theta_n - \theta_0)' D_{\theta_0}(z) + o_P(1).$$

Moreover, $V_n(\cdot, \theta_0) \Longrightarrow B(G(\cdot))$, in $D[-\infty, \infty]$, with respect to the uniform metric.

Next, we develop an analog of the linear transformation of Stute, Thies and Zhu (1998). Let

$$\begin{aligned}A_{\theta_0}(z) &= \int_z^\infty h_{\theta_0}(u)h_{\theta_0}(u)'\,dG(u),\\ &= \int_z^\infty \ell^2_{\theta_0}(u,H)\frac{\dot m_{\theta_0}(u)\dot m_{\theta_0}(u)'}{\sigma^2_{\theta_0}(u)}\,dG(u), \qquad z\in\mathbb{R}.\end{aligned}$$

Note that this is a nonnegative definite $q\times q$-matrix. But we shall assume that $A_{\theta_0}(z)$ is nonsingular for all $z<\infty$ and define the linear functional transform

$$(Q\varphi)(z) = \varphi(z) - \int_{-\infty}^z h_{\theta_0}(z_1)'A^{-1}_{\theta_0}(z_1)\left[\int_{z_1}^\infty h_{\theta_0}(z_2)\varphi(dz_2)\right]dG(z_1), \quad z\in\mathbb{R}.$$

When we apply Q to Brownian motion $B\circ G$, the inner integral needs to be interpreted as a stochastic integral.

Observe that (3.1.9) readily implies

$$Q(D'_{\theta_0}U)=0, \quad \text{for any random vector } U.$$

Arguing as in STZ, one can also verify that Q maps $B\circ G$ to $B\circ G$. Consequently, we have

$$Q(B\circ G + D'_0U) = Q(B\circ G) = B\circ G, \quad \text{for any random vector } U.$$

These observations together with Theorem 3.1.1. suggest that under H_0, $Q\tilde{V}_n$ would also converge weakly to $B\circ G$. But the transformation Q depends on the unknown parameters θ_0, H and G. Let h_n, A_n and σ_n denote the h_{θ_0}, A_{θ_0} and σ_{θ_0} after θ_0, H and G are replaced by θ_n, and the empirical d.f.'s H_n and G_n, respectively, in there. Define an estimate of Q to be

$$(Q_n\varphi)(z) = \varphi(z) - \int_{-\infty}^z h_n(z_1)'A_n^{-1}(z_1)\left[\int_{z_1}^\infty h_n(z_2)\varphi(dz_2)\right]dG_n(z_1).$$

To verify the weak convergence of $Q_n \tilde{V}_n$ we need the following additional smoothness condition on $h_\theta(z)$. For some $q \times q$ square matrix $\dot{h}_{\theta_0}(z)$ and a non-negative function $K_0(z)$, both measurable, the following holds:

$$E\|h_{\theta_0}(Z)\|^j K_0(z) < \infty, \qquad E\|\dot{h}_{\theta_0}(Z)\|\|h_{\theta_0}(Z)\|^j < \infty, \qquad j = 0, 1,$$

and $\forall \epsilon > 0$, there exists a $\delta > 0$ such that $\|\theta - \theta_0\| < \delta$ implies

$$\|h_\theta(z) - h_{\theta_0}(z) - \dot{h}_{\theta_0}(z)(\theta - \theta_0)\| \leq \epsilon K_0(z)\|\theta - \theta_0\|,$$

for almost all $z(G)$.

Using the methods of proof of STZ or Koul and Stute (1999), one can verify that under the above assumed conditions and under H_0, $Q_n \tilde{V}_n \Longrightarrow B \circ G$. Hence, under H_0, $\sup_{z \in \mathbb{R}} |Q_n \tilde{V}_n(z)| \Longrightarrow \sup_{0 \leq u \leq 1} |B(u)|$, $\int [Q_n \tilde{V}_n(z)]^2 dG_n(z) \Longrightarrow \int_0^1 B^2(u)du$, and the corresponding tests are asymptotically distribution free.

3.2 Estimation of θ

In order to apply the above results, it is important to have an estimator $\hat{\theta}$ of θ_0 under H_0 satisfying all the assumptions. Li and Zhang (1998) constructed an asymptotical efficient M-estimator of the regression coefficients in a linear regression model with interval censored data and when the error d.f. F is unknown. Since here F is assumed to be known and since their estimator is computational much more involved, we shall instead use the conditional least square estimator defined by

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} \left[\delta_i - F(T_i - m_\theta(Z_i))\right]^2.$$

Assume that F has a continuously differentiable density f and

$$\Sigma_{\theta_0} := E\Big(f^2(T - m_{\theta_0}(Z))\, \dot{m}_{\theta_0}(Z)\dot{m}_{\theta_0}(Z)'\Big)$$

is positive definite. In addition assume that $\dot{m}_\theta$ is continuously differentiable with the matrix of derivatives $\ddot{m}_{\theta_0}(z)$ satisfying $\|\ddot{m}_{\theta_0}(z)\| \le M_{\theta_0}(z)$, with $\int M_{\theta_0}(z)dG(z) < \infty$. Then using the classical Cramér type of argument one can verify that

$$
\begin{aligned}
& n^{1/2}(\hat{\theta} - \theta_0) \\
&= \Sigma_{\theta_0}^{-1} n^{-1/2} \sum_{i=1}^{n} [\delta_i - F(T_i - m_{\theta_0}(Z_i))] f(T_i - m_{\theta_0}(Z_i)) \dot{m}_{\theta_0}(Z_i) + o_P(1), \qquad (P_{\theta_0}).
\end{aligned}
$$

Consequently, under H_0,

$$
\begin{aligned}
& n^{1/2}(\hat{\theta} - \theta_0) \Longrightarrow \mathcal{N}_q(0, \Omega_0), \qquad \Omega_0 := \Sigma_{\theta_0}^{-1} M_0 \Sigma_{\theta_0}^{-1}, \\
& M_0 := E\Big\{(F\bar{F}f^2)(T - m_{\theta_0}(Z))\, \dot{m}_{\theta_0}(Z) \dot{m}_{\theta_0}(Z)'\Big\}.
\end{aligned}
$$

See, e.g., Liese and Vajda (2004) for a general method of proving asymptotic normality in nonlinear regression models.

3.3 A Simulation

Here we shall exhibit results of a finite sample simulation. For simplicity we took $\mathcal{M}$ to be simple linear regression model. Thus $q = 1$ and Let

$$
\mathcal{M} = \{m(\cdot, \theta) : m(z, \theta) = \theta z\}.
$$

In this case then several entities simplify as follows. Let $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ denote the ordered Z_i's and $T_{(i)}$'s, $\delta_{(i)}$'s denote the corresponding T_i's and δ_i's. Also, let $\ell_{nj}(z) \equiv \ell_{\theta_n}(Z_{(j)}, H_n)$, $R_{nj} := R_{\theta_n}(\delta_{(j)}, T_{(j)}, Z_{(j)})$, $\sigma_{nj} := \sigma_n(Z_{(j)})$, and $A_{nj} := A_n(Z_{(j)})$, where now

$$
A_n(z) := \frac{1}{n} \sum_{i=1}^{n} \frac{\ell_n^2(Z_{(i)})}{\sigma_n^2(Z_{(i)})} Z_{(i)}^2 I(Z_i \ge z).
$$

Then

$$
\begin{aligned}
& Q_n \tilde{V}_n(z) \\
= \ & \tilde{V}_n(z) - \frac{1}{n} \sum_{i=1}^{n} \frac{Z_{(i)} \ell_{ni}}{A_{ni} \sigma_{ni}} \frac{1}{n^{1/2}} \sum_{j=1}^{n} \frac{Z_{(j)} \ell_{nj} R_{nj}}{\sigma_{nj}^2} I(Z_{(j)} \wedge z \geq Z_{(i)}) \\
= \ & \frac{1}{n^{1/2}} \sum_{j=1}^{n} \left\{ I(Z_{(j)} \leq z) - \frac{1}{n} \sum_{i=1}^{n} \frac{Z_{(i)} Z_{(j)} \ell_{ni} \ell_{nj}}{A_{ni} \sigma_{ni} \sigma_{nj}} I(Z_{(j)} \wedge z \geq Z_{(i)}) \right\} \frac{R_{nj}}{\sigma_{nj}}
\end{aligned}
$$

To test the hypotheses $\tilde{H}$ and H_0, we consider the following two tests based on process $V_n^0(z)$ and $Q_n \tilde{V}_n(z)$ where

$$
K_n := \sup_{z \in \mathbb{R}} |V_n^0(z)| \qquad \hat{K}_n := \sup_{z \in \mathbb{R}} |Q_n \tilde{V}_n(z)|.
$$

We reject $\tilde{H}$ (H_0) whenever $K_n > b_\alpha$ ($\hat{K}_n > b_\alpha$), where b_α is the $100(1-\alpha)$ percentile of the distribution of $\sup_{0 \leq t \leq 1} |B(t)|$. Note that

$$
\begin{aligned}
K_n &= \frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{i=1}^{j} \frac{R_\theta(\delta_{(k)}, T_{(k)}, Z_{(k)})}{\sigma_\theta(Z_{(k)})} \right|, \\
\hat{K}_n &= \frac{1}{\sqrt{n}} \max_{1 \leq k \leq n} \left| \sum_{j=1}^{k} \left[1 - \frac{1}{n} \sum_{i=1}^{k \wedge j} \frac{Z_{(i)} Z_{(j)} \ell_{ni} \ell_{nj}}{A_{ni} \sigma_{ni} \sigma_{nj}} \right] \frac{R_{nj}}{\sigma_{nj}} \right|
\end{aligned}
$$

Next, we examine the finite sample performance of the test statistic K_n and $\hat{K}_n$ through some simulations. We generate the covariate Z_i's from the uniform distribution on the interval $[0, 1]$, and Y_i's according to

$$
Y_i = 3Z_i + aZ_i^2 + \varepsilon_i \qquad\qquad 1 \leq i \leq n,
$$

while ε_i's are simulated independently from the following distributions.

I : logistic(0,β): logistic distribution with location parameter 0 and scale parameter β.

II : Normal$(0, \sigma^2)$.

III : DE(0,β): double exponential distribution with location parameter 0 and scale parameter β.

We also generate the censoring time variable T_i's from the uniform distribution on the interval [0, 3]. Hence $\tilde{H} : \mu = 3Z$, $H_0 : m \in \mathcal{M}$ hold with $\theta_0 = 3$ if and only if $a = 0$.

We compute the empirical sizes and powers for different values of a and different error distributions. The results represent the Monte Carlo levels when $a = 0$ and the Monte Carlo powers when $a \neq 0$. The sample sizes used in the simulation are n=100, 200 and 400, each replicated 1000 times.

Table 3.1: *Empirical sizes and powers of K_n test, $\varepsilon \sim$ logistic* $(0, \beta)$

		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
a	β	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	1	0.088	0.095	0.043	0.043	0.022	0.021
0	2	0.092	0.098	0.046	0.048	0.021	0.022
	3	0.101	0.079	0.047	0.040	0.021	0.021
	1	0.262	0.488	0.162	0.348	0.108	0.252
1	2	0.134	0.237	0.090	0.150	0.054	0.104
	3	0.131	0.149	0.077	0.085	0.045	0.052
	1	0.641	0.922	0.527	0.869	0.406	0.784
2	2	0.327	0.568	0.221	0.455	0.146	0.341
	3	0.209	0.348	0.130	0.242	0.083	0.166

Table 3.2: *Empirical sizes and powers of K_n test, $\varepsilon \sim$ Normal* $(0, \sigma^2)$

		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
a	σ	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	1	0.105	0.100	0.057	0.045	0.034	0.018
0	2	0.089	0.091	0.043	0.040	0.017	0.019
	3	0.095	0.086	0.045	0.044	0.026	0.019
	1	0.416	0.742	0.313	0.630	0.233	0.053
1	2	0.208	0.385	0.135	0.270	0.089	0.187
	3	0.155	0.260	0.085	0.173	0.045	0.107
	1	0.890	0.987	0.808	0.976	0.722	0.964
2	2	0.581	0.870	0.472	0.782	0.363	0.686
	3	0.363	0.639	0.265	0.528	0.186	0.423

Table 3.3: *Empirical sizes and powers of $\hat{K}_n$ test, $\varepsilon \sim$ Normal* $(0, \sigma^2)$

		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
a	σ	$n = 200$	$n = 400$	$n = 200$	$n = 40$	$n = 200$	$n = 400$
	1	0.066	0.092	0.031	0.031	0.012	0.014
0	2	0.058	0.093	0.022	0.035	0.005	0.016
	3	0.083	0.156	0.035	0.075	0.019	0.037
	1	0.058	0.107	0.022	0.048	0.008	0.017
1	2	0.127	0.171	0.065	0.093	0.028	0.057
	3	0.119	0.229	0.052	0.123	0.026	0.077
	1	0.225	0.509	0.124	0.344	0.056	0.204
3	2	0.211	0.472	0.122	0.290	0.064	0.195
	3	0.195	0.344	0.113	0.219	0.056	0.124
	1	0.305	0.619	0.169	0.448	0.087	0.302
5	2	0.288	0.541	0.019	0.414	0.110	0.291
	3	0.243	0.396	0.157	0.279	0.091	0.192

Table 3.4: *Empirical sizes and powers of K_n test, $\varepsilon \sim DE\ (0, \beta)$*

		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
a	β	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
	1	0.088	0.089	0.040	0.045	0.025	0.020
0	2	0.078	0.099	0.041	0.045	0.022	0.020
	3	0.102	0.080	0.045	0.049	0.023	0.027
	1	0.350	0.574	0.231	0.433	0.147	0.340
1	2	0.205	0.342	0.135	0.246	0.075	0.172
	3	0.146	0.242	0.082	0.170	0.046	0.102
	1	0.747	0.968	0.644	0.939	0.537	0.904
2	2	0.508	0.792	0.388	0.691	0.264	0.591
	3	0.345	0.576	0.241	0.458	0.159	0.349

Table 3.5: *Empirical sizes and powers of $\hat{K}_n$ test, $\varepsilon \sim DE\ (0, \beta)$*

		$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
a	β	$n = 200$	$n = 400$	$n = 200$	$n = 40$	$n = 200$	$n = 400$
	1	0.102	0.083	0.058	0.038	0.030	0.019
0	2	0.083	0.089	0.047	0.041	0.025	0.025
	3	0.086	0.093	0.042	0.042	0.014	0.019
	1	0.106	0.169	0.051	0.089	0.017	0.050
1	2	0.080	0.135	0.040	0.067	0.019	0.040
	3	0.081	0.122	0.042	0.058	0.021	0.025
	1	0.177	0.477	0.091	0.301	0.042	0.171
3	2	0.191	0.355	0.098	0.233	0.050	0.144
	3	0.161	0.286	0.082	0.175	0.042	0.109
	1	0.217	0.534	0.090	0.343	0.042	0.217
5	2	0.274	0.548	0.165	0.401	0.082	0.263
	3	0.231	0.470	0.135	0.323	0.080	0.203

From the above tables, we can see the empirical sizes are close to the nominal level when sample size is large. Under the alternatives, the power decreases as β or σ increases, while it increases as a increases and sample size increases.

Packet has: 1 3pg. Bibliography