Summer 2010

The partially monotone tensor spline estimation of joint distribution function with bivariate current status data

Yuan Wu
University of Iowa

THE PARTIALLY MONOTONE TENSOR SPLINE ESTIMATION OF JOINT DISTRIBUTION FUNCTION WITH BIVARIATE CURRENT STATUS DATA

by

Yuan Wu

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

July 2010

Thesis Supervisor: Professor Ying Zhang

ABSTRACT

The analysis of joint distribution function with bivariate event time data is a challenging problem both theoretically and numerically. This thesis develops a tensor spline-based nonparametric maximum likelihood estimation method to estimate the joint distribution function with bivariate current status data.

Tensor I-splines are developed to replace the traditional tensor B-splines in approximating joint distribution function in order to simplify the restricted maximum likelihood estimation problem in computing. The generalized gradient projection algorithm is used to compute the restricted optimization problem. We show that the proposed tensor spline-based nonparametric estimator is consistent and that the rate of convergence can be as good as $n^{1/4}$. Simulation studies with moderate sample sizes show that the finite-sample performance of the proposed estimator is generally satisfactory.

Abstract Approved: ________________________________
Thesis Supervisor

____
Title and Department

____
Date

THE PARTIALLY MONOTONE TENSOR SPLINE ESTIMATION OF JOINT DISTRIBUTION FUNCTION WITH BIVARIATE CURRENT STATUS DATA

by

Yuan Wu

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Applied Mathematical and Computational Sciences
in the Graduate College of
The University of Iowa

July 2010

Thesis Supervisor: Professor Ying Zhang

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Yuan Wu

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Applied Mathematical and Computational Sciences at the July 2010 graduation.

Thesis Committee: ______________________________
Ying Zhang, Thesis Supervisor

____
Weimin Han

____
Jian Huang

____
Michael Jones

____
Lihe Wang

ACKNOWLEDGEMENTS

I am grateful to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

I deeply owe my gratitude to my advisor, Dr. Ying Zhang for his time and energy he put in developing my research skills and advising my thesis work, for his encouragement and support in setting my career goal, and for his patience in correcting my English writing. His wide knowledge and logical way of thinking have been of great value for me. Without his bright insights and personal guidance, I would never have been able to finish my dissertation.

I am really grateful to my academic advisor in the Department of Mathematics, Dr. Lihe Wang for his caring during these years at the University of Iowa. I would never forget he supported me financially and in all other possible ways when I had a difficult time due to my spoken English.

I would like to express my warm and sincere thanks to faculties and staffs in the Department of Mathematics, especially to Dr. Weimin Han and Dr. Yi Li. They both helped me in so many ways during my study at the University of Iowa.

I am grateful to Dr. Jian Huang and Dr. Michael Jones for their detailed review and valuable comments on this work.

I owe my thanks to my wife Xiaoli, who brought our newborn son, Aaron. Without her love and encouragement, it would have been impossible for me to finish this work. My special gratitude is due to my parents, my parents-in-law.

ABSTRACT

The analysis of joint distribution function with bivariate event time data is a challenging problem both theoretically and numerically. This thesis develops a tensor spline-based nonparametric maximum likelihood estimation method to estimate the joint distribution function with bivariate current status data.

Tensor I-splines are developed to replace the traditional tensor B-splines in approximating joint distribution function in order to simplify the restricted maximum likelihood estimation problem in computing. The generalized gradient projection algorithm is used to compute the restricted optimization problem. We show that the proposed tensor spline-based nonparametric estimator is consistent and that the rate of convergence can be as good as $n^{1/4}$. Simulation studies with moderate sample sizes show that the finite-sample performance of the proposed estimator is generally satisfactory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.1 Background of Bivariate Current Status Data

In some survival analysis applications, observation of the random event time T is restricted to the knowledge of whether or not T exceeds a random monitoring time C. This type of data is known as current status data, and sometimes referred to as interval censored data case I (Groenoboom and Wellner (1992)). Current status data arise naturally in many applications, see for example, in animal tumorigenicity experiments (Hoel and Walburg (1972), Finkelstein and Wolfe (1985), and Finkelstein (1986)). In these examples, for each experimental animal, T is the time from exposure to a potential carcinogen until occurrence of the tumor, and C is the time, on the same scale, of sacrifice. Upon sacrifice, the presence or absence of the tumor can be determined providing current status information on T. Another example of current status data arises in studies of the distribution of the age at weaning (Diamond et al. (1986), Diamond and McDonald (1991), and Grummer-Strawn (1993)). Here T represents the age of a child at weaning and C the age at observation. It also arises in studies of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) (Shiboski and Jewell (1992), and Jewell et al. (1994)).

The univariate current status data have been thoroughly studied in recent years. Groenoboom and Wellner (1992) studied the asymptotic properties of the nonparametric maximum likelihood estimator of the distribution function with current status data. Huang (1996) considered Cox's proportional hazards model with current status data and showed that the MLE of the regression parameter is asymptotically normal with $\sqrt{n}$ convergence rate, even through the MLE of the baseline cumulative hazard function only converges at $n^{1/3}$ rate. For a review of regression models for interval censored data, see Huang and Wellner (1997).

Bivariate event time data occur in many applications. For example, in an Australian twin study (Duffy et al. (1990)) the researchers were interested in times to a certain event such as disease or disease-related symptoms in both twins. As in univariate case, both failure times can be censored.

Some work has addressed the estimation of the joint distribution function of the correlated event times with bivariate right censored data. Analogous to the well-known Kaplan-Meier estimator for survival function, the bivariate Kaplan-Meier estimation on plane was proposed by Dabrowska (1988). Kooperberg (1998) discussed a tensor-spline estimation of the logarithm of joint density function with bivariate right censored data without studying the asymptotic properties of the estimator. For bivariate interval censored data case 2, a nonparametric two-stage method was proposed to estimate the joint distribution function in the literature. In this method, first the non-zero mass intersection rectangles are found, and then EM algorithm is applied to find the MLE of the joint distribution (Betensky and Finkelstein (1999) and Yu, et al. (2000)). In the second stage the EM algorithm could be replaced by the more efficient Iterative Convex Minarant method proposed by Groenoboom and Wellner (1992). But this nonparametric estimation is not uniquely defined (Yu, Wong and He (2000)) because the non-zero mass rectangles can not be uniquely determined. Moreover, the asymptotic properties of this type of nonparametric estimators are difficulty to study.

This thesis focuses on bivariate current status data. Let (T_1, T_2) be the two event times of interest and (C_1, C_2) the two corresponding random monitoring times. In this setting, bivariate current status data consist of

$$(C_1, C_2, \Delta_1 = I(T_1 \leq C_1), \Delta_2 = I(T_1 \leq C_1)),$$

where $I(\cdot)$ is the indicator function. This data structure arises in the studies of two diseases in same patients or some common disease for two correlated subjects. For example, Wang and Ding (2000) studied whether or not the onsets of hypertension and diabetes are correlated for people in two towns in Taiwan. By assuming a bivariate copula model Wang and Ding (2000) proposed a two-stage estimation of the association parameter of two event times, in which the joint distribution of the failure time variables is assumed to follow a bivariate copula model (Nelsen (2006)). First the nonparametric estimates of the marginal distributions are obtained, and then the association parameter is estimated by the maximum pseudo-likelihood method. This two-stage method facilitates an easy estimator of the joint distribution function through copula model, and is the only available method in the literature to estimate the joint distribution function with bivariate current status data. But if the copula model is not correctly specified, this estimator could be seriously biased. Jewell et al. (2005) studied the relationship between the time to HIV infection to the partner and the time to diagnosis of AIDS for the index case by estimating smooth functionals of marginal distribution functions. In both examples, the bivariate event times have the same monitoring time, that is $C_1 = C_2 = C$. Hence the joint distribution function can be only studied on the diagonal, that is, only $F(c, c)$ is identifiable. However, the common censoring assumption may not be true. For example, suppose a study is conducted to explore the times to first use of marijuana between siblings. Although the interview may be conducted at same times for both siblings, the ages at interview (monitoring times) of the siblings are different and these give arise to the bivariate current status data with possibly different C_1 and C_2. Ding and Wang (2004) proposed a nonparametric procedure for testing marginal dependence in the general scenario when two censoring times for two events could be different. But their goal was not the estimation of joint distribution function. This thesis is concerned about the estimation of joint distribution function in the general scenario when C_1 and C_2 may or may not be equal.

1.2 Thesis Objective and Proposed Method

In this thesis we propose to estimate the joint distribution function with current status data, using the tensor spline-based sieve maximum likelihood method. Conventionally, the estimate can be defined using the method similar to Betensky and Finkelstein (1999), and Yu et al. (2000). For these methods, the non-zero masses are determined for the intersection rectangles made by the collection of monitoring times. In addition to the uniqueness problem, as sample size increases, this problem is a high dimensional problem which is hard to deal with. To overcome these difficulties, we proposed a partially monotone tensor spline-based sieve estimation method for the the general bivariate current status data described in Section 1.1. Under some regularity conditions, we show the tensor spline estimator is consistent and we derive the convergence rate of this estimator. Our simulation studies indicate that the proposed tensor spline estimation method performs very well and better than a three-stage pseudo likelihood method extended from Wang and Ding' method (2000).

The rest of thesis is organized as follows.

In Chapter 2, Some technical backgrounds are introduced. First, we introduce the B-splines and the I-splines. We study the equivalency between the B-spines and the I-splines. The partially monotone tensor B-splines are used in studying the asymptotic properties of the proposed estimator. The partially monotone tensor I-splines are used in computing the restricted maximum likelihood estimate, since the constraints of the problem using the tensor I-splines is much simpler than that using the tensor B-splines. Second we introduce some basic concepts and results on empirical process theory, which will be heavily used in consistency proof and the derivation of convergence rate. Finally, we derive some useful technical results on the B-splines and the tensor B-splines.

In Chapter 3, the tensor spline-based sieve nonparametric maximum likelihood estimation method of joint distribution function with bivariate current status data is proposed. First, we derive the likelihood of bivariate current status data. Second we derive the likelihood with the B-splines, in which the partially monotone tensor B-spline function is used for the joint distribution function and the monotone B-spline functions are used for the two marginal distribution functions and we represent the spline-based sieve nonparametric maximum likelihood estimation problem as a constrained optimization problem with respect to the coefficients of the B-splines. Finally we similarly derive the likelihood with the I-splines and represent the problem in terms of the I-splines in order to have a set of constraints that are easily dealt with in computation.

In Chapter 4, the asymptotic properties of the proposed estimator are studied using the technical results on the B-splines. Some regularity conditions are provided for the theoretical development regarding the joint distribution function, the censoring time distribution functions and the B-splines. First, we show that the proposed estimator is consistent. Second, we derive the convergence rate of the proposed estimator, which can be as good as $n^{1/4}$.

In Chapter 5, the numerical studies are carried out. First, the generalized gradient projection method is introduced to compute the estimate. Then, extensive simulation studies are conducted to examine the finite sample performance of the proposed method. We also compare the proposed method with the Wang and Ding's method (2000).

In Chapter 6, we describe the further applications of the proposed method are discussed.

CHAPTER 2
TECHNICAL BACKGROUNDS

In this chapter, we describe three versions of spline functions and their relationships. We also summarize some results on modern empirical process theory that will be heavily used in studying the asymptotic properties of the spline-based nonparametric estimator of the joint distribution function with bivariate current status data.

2.1 Splines

In this section, we introduce three types of splines, the B-splines, the M-splines, and the I-splines. We describe the constructions of these splines and their relationships.

2.1.1 The B-splines

The normalized B-splines or simply the B-splines for brevity in this thesis, can be evaluated by the de Boor algorithm (de Boor, 2001) as follows. For a partition or a knot sequence, that is, a nondecreasing sequence $\{u_i\}$, the B-splines of order 1 with this knot sequence are the characteristic functions given by

$$N_i^1(s) = \begin{cases} 1, \ u_i \le s < u_{i+1}, \\ 0, \ \text{otherwise}. \end{cases} \tag{2.1}$$

If $u_i = u_{i+1}$, $N_i^1(s) = 0$.

The key characteristics of these functions are

(i) $N_i^1(s)$ for $i = 1, 2, \cdots$ are right continuous;

(ii) $\sum_i N_i^1(s) = 1$.

From the B-splines of order 1 (2.1) the B-splines of higher order can be obtained recursively by

$$N_i^k(s) = \omega_i^k(s)N_i^{k-1}(s) + (1 - \omega_{i+1}^k(s))N_{i+1}^{k-1}(s),$$

with

$$\omega_i^k(s) = \begin{cases} \frac{s-u_i}{u_{i+k-1}-u_i}, \ u_{i+k-1} \neq u_i, \\ \\ 0, \ \text{otherwise.} \end{cases}$$

If $u_{i+k-1} \neq u_i$ and $u_{i+k} \neq u_{i+1}$, we have

$$N_i^k(s) = \frac{s - u_i}{u_{i+k-1} - u_i}N_i^{k-1}(s) + \frac{u_{i+k} - s}{u_{i+k} - u_{i+1}}N_{i+1}^{k-1}(s). \tag{2.2}$$

There are some important properties for the B-splines which will be used throughout the rest of thesis and are summarized below.

(B1) (Theorem 4.18, Schumaker (1981)) The B-splines have support on several knot intervals. Specifically, for the B-splines of order l, if $u_j \leq s < u_{j+1}$, then

$$N_i^l(s) \begin{cases} \neq 0, \ j - l + 1 \leq i \leq j, \\ \\ = 0, \ \text{else.} \end{cases}$$

(B2) (Theorem 4.20, Schumaker (1981)) We already mentioned that the B-splines of order 1 form a partition of unity. Actually, this property is true for the B-splines of any order. Specifically, for the B-splines of order l $N_i^l(s)$'s we have

$$\sum_i N_i^l(s) = 1.$$

(B3) (Theorem 5.9, Schumaker (1981)) The derivatives of the B-splines can be calculated by:

$$\frac{\partial N_i^l(s)}{\partial s} = \frac{l-1}{u_{i+l-1}-u_i} N_i^{l-1}(s) - \frac{l-1}{u_{i+l}-u_{i+1}} N_{i+1}^{l-1}(s).$$

2.1.2 The M-splines

In numerical analysis the M-splines are non-negative spline functions. In this thesis the M-splines are used to construct the I-splines. Curry and Schoenberg (1966) proposed the construction of the M-splines as follows. Suppose a knot sequence is given by $\{u_i\}$, the M-splines of order 1 for this knot sequence are defined as

$$M_i^1(s) = \begin{cases} \frac{1}{u_{i+1}-u_i}, \ u_i \leq s < u_{i+1}, \\ 0, \ \text{otherwise.} \end{cases} \tag{2.3}$$

From the M-splines of order 1 (2.3) the M-splines of higher order can be obtained recursively by

$$M_i^l(s) = \frac{l[(s-u_i)M_i^{l-1}(s) + (u_{i+l}-s)M_{i+1}^{l-1}(s)]}{(l-1)(u_{i+l}-u_i)}. \tag{2.4}$$

Lemma 2.1. *Suppose the M-splines given by (2.4) and the B-splines given by (2.2) are associated with the same knot sequence, then they are closely related by*

$$M_i^l(s) = \frac{l}{u_{i+l}-u_i} N_i^l(s). \tag{2.5}$$

Proof. (i) Note that $M_i^1(s) = \frac{1}{u_{i+1}-u_i} N_i^1(s)$, so the relationship is true for $l = 1$.

(ii) Suppose the relation holds for $l = k-1$, then

$$M_i^k(s) = \frac{k[(s-u_i)M_i^{k-1}(s) + (u_{i+k}-s)M_{i+1}^{k-1}(s)]}{(k-1)(u_{i+k}-u_i)}$$

$$= \frac{k[(s-u_i)\frac{k-1}{u_{i+k-1}-u_i}N_i^{k-1}(s) + (u_{i+k}-s)\frac{k-1}{u_{i+k}-u_{i+1}}N_{i+1}^{k-1}(s)]}{(k-1)(u_{i+k}-u_i)}$$
$$= \frac{k}{u_{i+k}-u_i}[\frac{s-u_i}{u_{i+k-1}-u_i}N_i^{k-1}(s) + \frac{u_{i+k}-s}{u_{i+k}-u_{i+1}}N_{i+1}^{k-1}(s)]$$
$$= \frac{k}{u_{i+k}-u_i}N_i^k(s),$$

which implies (2.5) holds for $l = k$.

By induction (2.5) holds for any positive integer l, which completes the proof. □

2.1.3 The I-splines

Ramsay (1988) proposed the I-splines, which are monotone functions constrained between 0 and 1. The I-splines can be used as spline basis functions for regression analysis and data transformation when monotonicity is desired. The I-splines are constructed through the M-splines as follows.

Suppose the M-splines M_i^l's have the knot sequence $\{u_i\}_1^{p+l+1}$ satisfying

$$u_1 = \cdots = u_{l+1} < u_{l+2} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l+1}, \tag{2.6}$$

where the first and the last $l+1$ knots are equal because the M-splines are of order l.

The I-splines are defined as

$$I_i^l(s) = \begin{cases} 1, & i = 1, \\ \int_L^s M_i^l(t)dt, & 1 < i \leq p, \end{cases} \tag{2.7}$$

with $L \leq s \leq U$, where L and U are the left and the right end point of the knot sequence $\{u_i\}_1^{p+l+1}$, respectively.

Since $M_i^l(s)$'s given by (2.4) are piecewise polynomials of degree $l-1$, then $I_i^l(s)$'s given by (2.7) are piecewise polynomials of degree l. De Boor (2001) showed $\int_L^U M_i^l(t) =$

1 for positive integers i and l, which, along with the fact that $M_i^l(t)$ are nonnegative, implies that the I-splines in (2.7) are monotone non-decreasing function constrained between 0 and 1.

Suppose the knot sequence $\{u_i\}_1^{p+l+1}$ of the M-splines M_i^{l+1}'s satisfies (2.6), a more convenient expression of the I-splines was given by Ramsay (1988) as the following:

$$I_i^l(s) = \begin{cases} 0, \ i > j, \\ \sum_{m=i}^{j} (u_{m+l+1} - u_m) M_m^{l+1}(s)/(l+1), \ j-l+1 \leq i \leq j, \\ 1, \ i < j-l+1, \end{cases} \tag{2.8}$$

for $u_j \leq s < u_{j+1}$ and $1 \leq i \leq p$.

The following Lemma 2.2 and Lemma 2.3 together indicate that the I-splines given by (2.7) are equivalent to the I-splines given by (2.8).

Lemma 2.2. *Suppose M_i^{l+1}'s in (2.8) and N_i^{l+1}'s are associated with the same knot sequence, then the I-splines given by (2.8) can be expressed by*

$$I_i^l(s) = \sum_{m=i}^{p} N_m^{l+1}(s), \tag{2.9}$$

for $1 \leq i \leq p$.

Proof. By Lemma 2.1, (2.8) can be further expressed in terms of the B-splines:

$$I_i^l(s) = \begin{cases} 0, \ i > j, \\ \sum_{m=i}^{j} N_m^{l+1}(s), \ j-l+1 \leq i \leq j, \\ 1, \ i < j-l+1, \end{cases} \tag{2.10}$$

for $u_j \leq s < u_{j+1}$ and $1 \leq i \leq p$.

By the B-splines property (B1) in Section 2.1.1. for the B-spline of order $l+1$, if $u_j \leq s < u_{j+1}$,

$$N_m^{l+1}(s) \begin{cases} \neq 0, \ j-(l+1)+1 \leq m \leq j, \\ \\ = 0, \ \text{else}. \end{cases}$$

So the expression of the I-splines given by (2.10) can be rewritten as

$$I_i^l(s) = \sum_{m=i}^{p} N_m^{l+1}(s).$$

□

Lemma 2.3. *Suppose M_i^{l+1}'s in (2.8) and N_i^{l+1}'s in (2.9) are associated with the same knot sequence, the I-splines given by (2.9) are equivalent to the I-splines given by (2.7).*

Proof. (i) We show $\sum_{m=i}^{p} N_m^{l+1}(s) = \int_L^s M_i^l(t)dt$ for $i = 2, \ldots, p$. We shall demonstrate this by the following two steps:

(a) Prove $\frac{\partial(\sum_{m=i}^{p} N_m^{l+1}(s))}{\partial s} = \frac{\partial(\int_L^s M_i^l(t)dt)}{\partial s}$ as follows.

$$\begin{aligned} \frac{\partial(\sum_{m=i}^{p} N_m^{l+1}(s))}{\partial s} =& \sum_{m=i}^{p} \{ \frac{l}{u_{m+l}-u_m} N_m^l(s) - \frac{l}{u_{m+l+1}-u_{m+1}} N_{m+1}^l(s) \} \\ =& \frac{l}{u_{i+l}-u_i} N_i^l(s) + \sum_{m=i}^{p-1} \frac{l-l}{u_{m+l+1}-u_{m+1}} N_{m+1}^l(s) \\ =& \frac{l}{u_{i+l}-u_i} N_i^l(s) = M_i^l(s) \\ =& \frac{\partial(\int_L^s M_i^l(t)dt)}{\partial s} \end{aligned}$$

(b)

$$\sum_{m=i}^{p} N_m^{l+1}(L) = 0 = \int_L^L M_i^l(t)dt.$$

(a) and (b) imply for $i = 2, \ldots, p$, $\sum_{m=i}^{p} N_m^{l+1}(s)$ and $\int_L^s M_i^l(t)dt$ have same derivative and they both are zero at the left end point of the knot sequence. Therefore $\sum_{m=i}^{p} N_m^{l+1}(s) = \int_L^s M_i^l(t)dt$, for $i = 2, \ldots, p$.

(ii) For $i = 1$, it is trivial because of property (B2) in Section 2.1.1.

□

Remark 2.1. *(2.9) provides a much easier way to compute the I-splines than using definition (2.7) due to the available softwares for the B-splines. In Appendix the explicit form of the I-spline with order 3 is given, along with the steps to to construct the I-splines through the B-splines in statistics package R.*

2.1.4 Monotone Spline Functions

In the literature, $\{N_i^l : i = 1, 2, \cdots, l = 1, 2, \cdots\}$ are referred to as the B-spline basis functions. In this thesis, we denote the B-spline functions as the linear combinations of the B-spline basis functions. Suppose we have the lth-order B-spline basis functions $N_i^l(s)$ with knot sequence $\{u_i\}_1^{p+l}$ satisfying $u_1 = \cdots = u_l < u_{l+1} \cdots < u_p < u_{p+1} = \cdots = u_{p+l}$. A B-spline function of order l is given by

$$f(s) = \sum_{i=1}^{p} \beta_i N_i^l(s). \tag{2.11}$$

It is obvious by (B1) that

$$f(s) = \sum_{i=j-l+1}^{j} \beta_i N_i^l(s)$$

for $u_j \leq s < u_{j+1}$.

According to (B3) and (2.11), the derivative of $f(s)$ is given by

$$\begin{aligned}\frac{\partial f}{\partial s} &= \sum_{i=1}^{p} \beta_i \frac{\partial N_i^l(s)}{\partial s} \\ &= \sum_{i=1}^{p} \beta_i \{ \frac{l-1}{u_{i+l-1} - u_i} N_i^{l-1}(s) - \frac{l-1}{u_{i+l} - u_{i+1}} N_{i+1}^{l-1}(s) \} \\ &= \sum_{i=1}^{p-1} \frac{(l-1)(\beta_{i+1} - \beta_i)}{u_{i+l} - u_{i+1}} N_{i+1}^{l-1}(s),\end{aligned} \tag{2.12}$$

the last equality due to the fact that $N_1^{l-1}(u) \equiv 0$ and $N_{p+1}^{l-1}(u) \equiv 0$.

Lemma 2.4. *If* $0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p$*, the B-spline function* $f(s)$ *given by (2.11) is a nonnegative and nondecreasing function.*

Proof. It is obvious that $f(s)$ is nonnegative. By (2.12), the derivative of $f(s)$ is nonnegative as well. □

Similarly the I-spline functions in the thesis are denoted by linear combinations of the I-spline basis functions $\{I_i^l : i = 1, 2, \cdots, l = 1, 2, \cdots\}$. Specifically, a I-spline function of $(l-1)$th-order is given by

$$f(s) = \sum_{i=1}^{p} \gamma_i I_i^{l-1}(s). \tag{2.13}$$

Lemma 2.5. *If* $\gamma_i \geq 0$ *for* $i = 1, \cdots, p$*,* $f(s)$ *given by (2.13) is nonnegative and nondecreasing.*

Proof. Because the I-spline basis functions are nonnegative and nondecreasing, $\gamma_i \geq 0$ for $i = 1, \cdots, p$ are sufficient to guarantee $f(s)$ for being nonnegative and nondecreasing. □

Lemma 2.6. $f(s) = \sum_{i=1}^{p} \beta_i N_i^l(s)$ *with* $0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p$ *and* $f(s) = \sum_{i=1}^{p} \gamma_i I_i^{l-1}(s)$ *with* $\gamma_i \geq 0$ *for* $i = 1, \cdots, p$ *are equivalent to each other.*

Proof. For a monotone spline function expressed in the I-splines by (2.13), substitute $I_i^{l-1}(s) = \sum_{j=i}^{p} N_j^l(s)$, it follows that

$$f(s) = \sum_{i=1}^{p} \{\sum_{j=1}^{i} \gamma_j\} N_i^l(s).$$

By $\gamma_i \geq 0,\ i = 1, \cdots, p$, we can rewrite the spline function (2.13) as

$$f(s) = \sum_{i=1}^{p} \beta_i N_i^l(s),$$

where $\beta_i = \sum_{j=1}^{i} \gamma_j,\ i = 1, \ldots, p$. Then $0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p$. □

Remark 2.2. *Lemma 2.6 implies that a nonnegative and nondecreasing spline function can be equivalently expressed by either the B-splines or the I-splines.*

2.1.5 Partially Monotone Tensor Spline Functions

Suppose we have the lth-order B-splines $N_i^{(1),l}(s)$ with knot sequence $\{u_i\}_1^{p+l}$ satisfying $u_1 = \cdots = u_l < u_{l+1} \cdots < u_p < u_{p+1} = \cdots = u_{p+l}$ and the lth-order B-splines $N_j^{(2),l}(t)$ with knot sequence $\{v_i\}_1^{q+l}$ satisfying $v_1 = \cdots = v_l < v_{l+1} \cdots < v_q < v_{q+1} = \cdots = v_{q+l}$. The tensor B-spline functions are given by

$$f(s,t) = \sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t). \tag{2.14}$$

By (B3) the partial derivatives of the tensor B-spline function can be expressed by

$$\frac{\partial f(s,t)}{\partial s} = \sum_{i=1}^{p-1} \sum_{j=1}^{q} \frac{(l-1)(\alpha_{i+1,j} - \alpha_{i,j})}{u_{i+l} - u_{i+1}} N_{i+1}^{(1),l-1}(s) N_j^{(2),l}(t), \tag{2.15}$$

and

$$\frac{\partial f(s,t)}{\partial t} = \sum_{i=1}^{p} \sum_{j=1}^{q-1} \frac{(l-1)(\alpha_{i,j+1} - \alpha_{i,j})}{v_{j+l} - v_{j+1}} N_i^{(1),l}(s) N_{j+1}^{(2),l-1}(t). \tag{2.16}$$

Lemma 2.7. *If* $0 \leq \alpha_{1,j} \leq \alpha_{2,j} \leq \cdots \leq \alpha_{p,j}$ *for* $j = 1, 2, \cdots, q$ *and* $0 \leq \alpha_{i,1} \leq \alpha_{i,2} \leq \cdots \leq \alpha_{i,q}$ *for* $i = 1, 2, \cdots, p$*, the tensor B-spline function* $f(s,t)$ *given by (2.14) is nonnegative and nondecreasing in both* s *and* t *directions.*

Proof. It is obvious that $f(s,t)$ is nonnegative. By (2.15) and (2.16), both partial derivatives of $f(s,t)$ are nonnegative. □

Similarly the tensor I-spline functions are given by

$$f(s,t) = \sum_{i=1}^{p} \sum_{i=1}^{q} \eta_{i,j} I_i^{(1),l-1}(s) I_j^{(2),l-1}(t). \tag{2.17}$$

Lemma 2.8. *If* $\eta_{i,j} \geq 0$ *for* $i = 1, \cdots, p$ *and* $j = 1, \cdots, q$*,* $f(s,t)$ *given by (2.17) is nonnegative and nondecreasing in both* s *and* t *directions.*

Proof. Because the I-spline basis functions are nonnegative and nondecreasing, $\eta_{i,j} \geq 0$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$ are sufficient to guarantee $f(s,t)$ for being nonnegative and nondecreasing in both s and t directions. □

Let

$$S_I = \{f(s,t) = \sum_{i=1}^{p} \sum_{j=1}^{q} \eta_{i,j} I_i^{(1),l-1}(s) I_j^{(2),l-1}(t) : \eta_{i,j} \geq 0 \text{ for } i = 1, \ldots, p \text{ and } j = 1, \ldots, q\},$$

and

$$S_B = \{f(s,t) = \sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t) : \alpha'_{i,j}s \text{ satisfy } (2.18)\},$$

$$\begin{aligned} &0 \leq \alpha_{1,j} \leq \alpha_{2,j} \leq \cdots \leq \alpha_{p,j}, \text{ for } j = 1, 2, \cdots, q, \\ &0 \leq \alpha_{i,1} \leq \alpha_{i,2} \leq \cdots \leq \alpha_{i,q}, \text{ for } i = 1, 2, \cdots, p, \\ &(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0, \\ &\text{for } i = 1, 2, \cdots, p-1 \text{ and } j = 1, 2, \cdots, q-1. \end{aligned} \tag{2.18}$$

Lemma 2.9. *Set S_B is equivalent to set S_I.*

Proof. Substitute $I_i^{(1),l-1} = \sum_{x=i}^{p} N_x^{(1),l}$ and $I_j^{(2),l-1} = \sum_{y=j}^{q} N_y^{(2),l}$ in (2.17), then

$$f(s,t) = \sum_{i=1}^{p}\sum_{j=1}^{q}\{\sum_{x=1}^{i}\sum_{y=1}^{j}\eta_{x,y}\}N_i^{(1),l}(s)N_j^{(2),l}(t). \tag{2.19}$$

Let $\alpha_{i,j} = \sum_{x=1}^{i}\sum_{y=1}^{j}\eta_{x,y}$, then (2.19) can be written as

$$f(s,t) = \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_{i,j}N_i^{(1),l}(s)N_j^{(2),l}(t).$$

Then the following results (i), (ii), (iii) and (iv) can be easily obtained.

(i)

$$\alpha_{1,1} = \eta_{1,1} \geq 0.$$

(ii) For $i = 1, \ldots, p$ and $j = 1, \ldots, q-1$,

$$\begin{aligned}\alpha_{i,j+1} - \alpha_{i,j} &= \sum_{x=1}^{i}\sum_{y=1}^{j+1}\eta_{x,y} - \sum_{x=1}^{i}\sum_{y=1}^{j}\eta_{x,y}\\ &= \sum_{x=1}^{i}\eta_{x,j+1} \geq 0.\end{aligned}$$

(iii) For $i = 1, \ldots, p-1$ and $j = 1, \ldots, q$, we have

$$\begin{aligned}\alpha_{i+1,j} - \alpha_{i,j} &= \sum_{x=1}^{i+1}\sum_{y=1}^{j}\eta_{x,y} - \sum_{x=1}^{i}\sum_{y=1}^{j}\eta_{x,y}\\ &= \sum_{y=1}^{j}\eta_{i+1,y} \geq 0.\end{aligned}$$

(iv) For $i = 1, \ldots, p-1$ and $j = 1, \ldots, q-1$,

$$\begin{aligned}(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) =& (\sum_{x=1}^{i+1}\sum_{y=1}^{j+1}\eta_{x,y} - \sum_{x=1}^{i+1}\sum_{y=1}^{j}\eta_{x,y}) \\ & - (\sum_{x=1}^{i}\sum_{y=1}^{j+1}\eta_{x,y} - \sum_{x=1}^{i}\sum_{y=1}^{j}\eta_{x,y}) \\ =& \sum_{x=1}^{i+1}\eta_{x,j+1} - \sum_{x=1}^{i}\eta_{x,j+1} \\ =& \eta_{i+1,j+1} \geq 0.\end{aligned} \tag{2.20}$$

By (i), (ii), (iii) and (iv), S_I and S_B are equivalent. □

Remark 2.3. *Lemma 2.9 implies the partially monotone tensor I-spline functions are not equivalent to the partially monotone tensor B-spline functions, since there must be an extra condition (2.20) for the partially monotone tensor B-spline function to make them equivalent. However this condition is very necessary for estimating the joint distribution function, because it corresponds to the fact that the mass of the joint distribution on any rectangle region of its domain is nonnegative.*

2.2 Results on Empirical Process Theory

Empirical process theory is a very powerful tool in studying the asymptotic properties of nonparametric or semiparametric estimator in Statistics. In this section we introduce some basic concepts and results which will be used in our theoretical development.

Definition 2.1. *(Covering number, van der Vaart and Wellner (1996), P. 83) The covering number $N(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$. The entropy with covering $H(\epsilon, \mathcal{F}, \|\cdot\|)$ is the logarithm of the covering number.*

Definition 2.2. *(Bracketing number, van der Vaart and Wellner (1996), P. 83) Given two functions* l *and* u*, the bracket* $[l, u]$ *is the set of all functions* f *with* $l \leq f \leq u$*. An* ϵ*-bracket is a bracket* $[l, u]$ *with* $\|u - l\| < \epsilon$*. The bracketing number* $N_{[\,]}(\epsilon, \mathcal{F}, \|\cdot\|)$ *is the minimum number of* ϵ*-brackets needed to cover* $\mathcal{F}$*. The entropy with bracketing* $H_{[\,]}(\epsilon, \mathcal{F}, \|\cdot\|)$ *is the logarithm of the bracketing number.*

Definition 2.3. *(VC-index and VC-class of sets, van der Vaart and Wellner (1996), P. 134-135) Let* $\mathcal{C}$ *a collection of subsets of a set* $\mathcal{X}$*. An arbitrary set of* n *points* $\{x_1, \ldots, x_n\}$ *possesses* 2^n *subsets. Say that* C *picks out a certain subset from* $\{x_1, \ldots, x_n\}$ *if this can be formed as a set of the form* $C \cap \{x_1, \ldots, x_n\}$ *for a* C *in* $\mathcal{C}$*. The collection* $\mathcal{C}$ *is said to shatter* $\{x_1, \ldots, x_n\}$ *if each of its* 2^n *subsets can be picked out in this manner. The VC-index* $V(\mathcal{C})$ *of the class* $\mathcal{C}$ *is the smallest* n *for which no set of size* n *is shattered by* $\mathcal{C}$*. A collection of measurable sets* $\mathcal{C}$ *is called a VC-class of sets if its VC-index is finite.*

Definition 2.4. *(Symmetric convex hull, van der Vaart and Wellner (1996), P. 142) The symmetric convex hull* $sconv(\mathcal{F})$ *of a class of functions is defined as the set of functions* $\sum_{i=1}^{m} \alpha_i f_i$*, with* $\sum_{i=1}^{m} |\alpha_i| \leq 1$ *and each* f_i *contained in* $\mathcal{F}$*.*

Definition 2.5. *(VC-class of functions, van der Vaart and Wellner (1996), P. 141) The subgraph of a function* $f : \mathcal{X} \mapsto \mathbb{R}$ *is the subset of* $\mathcal{X} \times \mathbb{R}$ *given by* $\{(x, t) : t < f(x)\}$*. A collection* $\mathcal{F}$ *of measurable functions on a sample space is called a VC-class of functions, if the collection of all subgraphs of the functions in* $\mathcal{F}$ *forms a VC-classes of sets in* $\mathcal{X} \times \mathbb{R}$*.*

Definition 2.6. *(Envelop function, van der Vaart and Wellner (1996), P. 84) An envelop function of a class* $\mathcal{F}$ *is any function* $x \mapsto F(x)$ *such that* $|f(x)| \leq F(x)$*, for every* x *and*

$f \in \mathcal{F}$.

In the following lemmas, the $L_r(Q)$ norm associated with probability measure Q is defined by

$$\|f\|_{L_r(Q)} = (Q|f|^r)^{1/r} = (\int |f|^r dQ)^{1/r},$$

for $r > 0$, the empirical measure $\mathbb{P}_n$ of a sample of random elements $x_1, \ldots, x_n$ is defined by

$$\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(x_i),$$

and accordingly $\mathbb{G}_n$ is denoted as $\sqrt{n}(\mathbb{P}_n - P)$.

Lemma 2.10. *(Theorem 5.7 in van der Vaart (1998)) Let M_n be random functions and let M be a fixed function of θ such that for every $\epsilon > 0$*

$$\sup_{\theta\in\Theta} |M_n(\theta) - M(\theta)| \to_P 0,$$

$$\sup_{\theta: d(\theta,\theta_0)\geq\epsilon} M(\theta) < M(\theta_0).$$

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_P(1)$ converges in probability to θ_0.

Lemma 2.11. *(Theorem 2.6.7 in van der Vaart and Wellner (1996)) For a VC-class of functions with measurable envelop function F and $r \geq 1$, one has for any probability measure Q with $\|F\|_{L_r(Q)} > 0$,*

$$N(\epsilon\|F\|_{L_r(Q)}, \mathcal{F}, L_r(Q)) \leq KV(\mathcal{F})(16e)^{V(\mathcal{F})}(\frac{1}{\epsilon})^{r(V(\mathcal{F})-1)},$$

for a universal constant K and $0 < \epsilon < 1$.

Lemma 2.12. *(Theorem 2.6.9 in van der Vaart and Wellner (1996)) Let Q be a probability measure, and let $\mathcal{F}$ be a class of measurable functions with measurable square integrable envelop F such that $QF^2 < \infty$ and $N(\epsilon\|\mathcal{F}\|_{L_2(Q)}, \mathcal{F}, L_2(Q)) \leq C(\frac{1}{\epsilon})^V$ for $0 < \epsilon < 1$. Then there exists a constant K that depends on C and V only such that*

$$N(\epsilon\|\mathcal{F}\|_{L_2(Q)}, \overline{sconv}(\mathcal{F}), L_2(Q)) \leq K(\frac{1}{\epsilon})^{2V/(V+2)}.$$

Lemma 2.13. *(Theorem 2.5.2 in van der Vaart and Wellner (1996)) Let $\mathcal{F}$ be a class of measurable functions that satisfies $\int_0^\infty \sup_Q \sqrt{\log N(\epsilon\|\mathcal{F}\|_{L_2(Q)}, \mathcal{F}, L_2(Q))} d\epsilon < \infty$, where the envelop function F of $\mathcal{F}$ is square integrable and the supremum is taken over all finitely discrete probability measures Q with $\|F\|^2_{L_2(Q)} > 0$. Let the classes $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \|f - g\|_{L_2(P)} < \delta\}$ and $\mathcal{F}_\delta^2$ be P-measurable for every $\delta > 0$. If $PF^2 < \infty$, then $\mathcal{F}$ is P-Donsker .*

Lemma 2.14. *(Example 2.10.7 in van der Vaart and Wellner (1996)) If $\mathcal{F}$ and $\mathcal{G}$ are P-Donsker Classes with $\sup_{f\in\mathcal{F}\cup\mathcal{G}} |Pf| < \infty$, then the pairwise infima $\mathcal{F} \wedge \mathcal{G}$, the pairwise suprema $\mathcal{F} \vee \mathcal{G}$, and pairwise sums $\mathcal{F} + \mathcal{G}$ are P-Donsker classes.*

Lemma 2.15. *(Corollary 2.3.12 in van der Vaart and Wellner (1996)) Let $\mathcal{F}$ be a class of measurable functions and the semi-norm ρ_P on $\mathcal{F}$ be defined as $\rho_P(f) = \{P(f - Pf)^2\}^{1/2}$. Then $\mathcal{F}$ being P-Donsker class implies that, if $f_n \to f$ as $n \to \infty$ in semi-norm ρ_P for all f_n and f in $\mathcal{F}$, then $(\mathbb{P}_n - P)(f_n - f) = o_P(n^{-1/2})$.*

Lemma 2.16. *(Lemma 3.4.2 in van der Vaart and Wellner (1996)) Let $\mathcal{F}$ be the class of measurable functions such that $Pf^2 < \delta^2$ and $\|f\|_\infty \leq M$ for every f in $\mathcal{F}$. Then there* *exists $K > 0$, which is related to M, such that*

$$E_P \|\mathbb{G}_n\|_{\mathcal{F}} \leq K \tilde{J}_{[\,]}\{\delta, \mathcal{F}, L_2(P)\}[1 + \frac{\tilde{J}_{[\,]}\{\delta, \mathcal{F}, L_2(P)\}}{\delta^2 \sqrt{n}}],$$

where $\tilde{J}_{[\,]}\{\delta, \mathcal{F}, L_2(P)\} = \int_0^\delta \sqrt{1 + \log N_{[\,]}\{\epsilon, \mathcal{F}, L_2(P)\}} d\epsilon$ and $\|\mathbb{G}_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|$.

Lemma 2.17. *(Theorem 3.4.1 in van der Vaart and Wellner (1996)) For each n, let $\mathbb{M}_n$ and M_n be stochastic processes indexed by a set Θ Let $\theta_n \in \Theta$ and $0 \leq \delta_n < \eta$ be arbitrary. Suppose that, for every n and $\delta_n < \delta \leq \eta$*

$$\sup_{\delta/2 < d_n(\theta,\theta_n) \leq \delta, \theta\in\Theta_n} M_n(\theta) - M_n(\theta_n) \leq -\delta^2,$$

$$E \sup_{\delta/2 < d_n(\theta,\theta_n) \leq \delta, \theta\in\Theta_n} \sqrt{n}[(\mathbb{M}_n - M_n)(\theta) - (\mathbb{M}_n - M_n)(\theta_n)]^+ \leq C\phi(\delta),$$

for functions ϕ such that $\delta \mapsto \phi(\delta)/\delta^\alpha$ is decreasing on (δ_n, η), for some $\alpha < 2$. Let $r_n \leq C\delta_n^{-1}$ satisfy

$$r_n^2 \phi(\frac{1}{r_n}) \leq \sqrt{n}, \text{ for every } n.$$

If the sequence $\hat{\theta}_n$ takes its values in Θ_n and satisfies $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_n) - O_P(r_n^{-2})$ and $d_n(\hat{\theta}_n, \theta_n)$ converges to zero in probability, then $r_n d_n(\hat{\theta}_n, \theta_n) = O_P(1)$.

2.3 Some Useful Results on B-splines

Lemma 2.18. *(Jackson type Theorem, De boor (2001), P. 149) Suppose $g(x)$ is a function with the continuous derivative $\frac{d^w g(x)}{dx^w}$. Then there exists a B-spline function $Ag(x) = \sum_{i=1}^p \beta_i N_i^l(x)$ with order l of the B-spline basis functions satisfying $l \geq w + 2$ and have knot sequence $\{u_i\}_1^p$ with $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U_1$, such that*

$$\|g - Ag\|_\infty \leq c|T|^w \| \frac{d^w g}{dx^w} \|_\infty$$

for some constant $c > 0$ *depending on* l *only, where*

$$|T| = \max_{l \le i \le p}(u_{i+1} - u_i).$$

The result of Lemma 2.18 can be generalized to bivariate function and is given in Lemma 2.19.

Lemma 2.19. *Suppose* $g(x, y)$ *is a bivariate function with the continuous mixed derivatives of order* w, $\nabla^w_m g = \frac{\partial^w g(x,y)}{\partial x^m y^{w-m}}$ *for* $m = 1, 2, \ldots, w$. *Then there exists a bivariate tensor B-spline function*

$Ag(x, y) = \sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_{i,j} N_i^{(1),l}(x) N_j^{(2),l}(y)$ *with order* l *in both directions satisfying* $l \ge w + 2$ *and have knot sequences* $\{u_i\}_1^{p+l}$ *with* $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U_1$, $\{v_i\}_1^{q+l}$ *with* $L_2 = v_1 = \cdots = v_l < v_{l+1} < \cdots < v_q < v_{q+1} = \cdots = v_{q+l} = U_2$, *such that*

$$\|g - Ag\|_\infty \le c|T|^w(\|g\|_{w,\infty}),$$

for some constant $c > 0$ *depending on* l *only, where*

$$|T| = \max\{\max_{l \le i \le p}(u_{i+1} - u_i), \max_{l \le j \le q}(v_{j+1} - v_j)\},$$

and

$$\|g\|_{w,\infty} = \max_{0 \le m \le w} \|\frac{\partial^w g}{\partial x^m \partial y^{w-m}}\|_\infty.$$

Proof. We define $\omega(g; h) = \max\{|g(x_1, y_1) - g(x_2, y_2)| : |x_1 - x_2| \le h, |y_1 - y_2| \le h, x_1, x_2 \in [L_1, U_1], y_1, y_2 \in [L_2, U_2]\}$. Then $\omega(g; h)$ is a monotone and subadditivity function of h, that is, $\omega(g; h_1) \le \omega(g; h_1 + h_2) \le \omega(g; h_1) + \omega(g; h_2)$ for nonnegative h_1

and h_2. The monotonicity of $\omega(g;h)$ is obvious by the definition. The proof of subadditivity is as follows.

For any (x_1, y_1) and (x_2, y_2) with $|x_1 - x_2| \le h_1 + h_2$ and $|y_1 - y_2| \le h_1 + h_2$, we can find (x_3, y_3) such that $|x_1 - x_3| \le h_1$, $|y_1 - y_3| \le h_1$ and $|x_2 - x_3| \le h_2$, $|y_2 - y_3| \le h_2$.

Therefore, for any $|x_1 - x_2| \le h_1 + h_2$ and $|y_1 - y_2| \le h_1 + h_2$, we have

$$\begin{aligned} |g(x_1, y_1) - g(x_2, y_2)| \le & |g(x_1, y_1) - g(x_3, y_3)| + |g(x_3, y_3) - g(x_2, y_2)| \\ \le & \max_{\substack{|x_1 - x_3| \le h_1 \\ |y_1 - y_3| \le h_1}} |g(x_1, y_1) - g(x_3, y_3)| \\ & + \max_{\substack{|x_2 - x_3| \le h_2 \\ |y_2 - y_3| \le h_2}} |g(x_3, y_3) - g(x_2, y_2)| \\ = & \omega(g; h_1) + \omega(g; h_2). \end{aligned} \tag{2.21}$$

By (2.21), $\omega(g; h_1 + h_2) \le \omega(g; h_1) + \omega(g; h_2)$ for nonnegative h_1 and h_2, that is, subadditivity of $\omega(g;h)$ holds.

By choosing $\tau_1 < \tau_2 < \cdots < \tau_p$ in $[L_1, U_1]$ and $\xi_1 < \xi_2 < \cdots < \xi_q$ in $[L_2, U_2]$, we can construct a partially monotone tensor B-spline function Ag to approximate the smooth function g on $[L_1, U_1] \times [L_2, U_2]$ as follows.

$$Ag(x, y) = \sum_{i=1}^{p} \sum_{j=1}^{q} g(\tau_i, \xi_j) N_i^{(1),l}(x) N_j^{(2),l}(y) = \sum_{i=1}^{p} \{\sum_{j=1}^{q} N_j^{(2),l}(y) g(\tau_i, \xi_j)\} N_i^{(1),l}(x).$$

For $(\hat{x}, \hat{y})$ in $[u_{j_1}, u_{j_1+1}] \times [v_{j_2}, v_{j_2+1}] \in [L_1, U_1] \times [L_2, U_2]$,

$$Ag(\hat{x}, \hat{y}) = \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} g(\tau_i, \xi_j) N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}), \tag{2.22}$$

by property (B1) in Section 2.1.1. Also by property (B2) in Section 2.1.1, we have

$$\begin{aligned} g(\hat{x},\hat{y}) &= g(\hat{x},\hat{y}) \sum_{i=j_1+1-l}^{j_1} N_i^{(1),l}(\hat{x}) \\ &= g(\hat{x},\hat{y}) \sum_{i=j_1+1-l}^{j_1} \{ \sum_{j=j_2+1-l}^{j_2} N_j^{(2),l}(\hat{y}) \} N_i^{(1),l}(\hat{x}) \\ &= g(\hat{x},\hat{y}) \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}). \end{aligned} \tag{2.23}$$

By (2.22) and (2.23),

$$g(\hat{x},\hat{y}) - Ag(\hat{x},\hat{y}) = \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} \{ g(\hat{x},\hat{y}) - g(\tau_i,\xi_j) \} N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}).$$

Then we have

$$\begin{aligned} |g(\hat{x},\hat{y}) - Ag(\hat{x},\hat{y})| &\leq \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} |g(\hat{x},\hat{y}) - g(\tau_i,\xi_j)| N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}) \\ &\leq \max_{\substack{j_1+1-l \leq i \leq j_1 \\ j_2+1-l \leq j \leq j_2}} |g(\hat{x},\hat{y}) - g(\tau_i,\xi_j)|. \end{aligned}$$

Now choose the τ_i's and ξ_j appropriately. since the number of τ_i's is greater than the number of subintervals made from knot sequence $\{u_i\}_1^{p+l}$, in order to guarantee that $\tau_{i+1} - \tau_i > 0$ for $i = 1, \cdots, p-1$, we might choose

$$\tau_i = \begin{cases} u_1 + \frac{(i-1)(u_{l+1}-u_l)}{l}, \ i = 1, \ldots, l, \\ \\ u_i, \ i = l+1, \ldots, p. \end{cases} \tag{2.24}$$

Similarly, we might choose

$$\xi_j = \begin{cases} v_1 + \frac{(j-1)(v_{l+1}-v_l)}{l}, \ j = 1, \ldots, l, \\ \\ v_j, \ j = l+1, \ldots, q, \end{cases} \tag{2.25}$$

(2.24) and (2.25) imply $|\tau_i - u_i| \le |T|$ and $|\xi_j - v_j| \le |T|$ for $i = 1, \dots, p$ and $j = 1, \dots, q$. We also know $|u_i - \hat{x}| \le u_{j_1+1} - u_{j_1-l+1} \le l|T|$ for $j_1 - l < i \le j_1$ and $\hat{x} \in [u_{j_1}, u_{j_1+1}]$ and $|v_j - \hat{y}| \le v_{j_2+1} - v_{j_2-l+1} \le l|T|$ for $j_2 - l < j \le j_2$ and $\hat{y} \in [v_{j_2}, v_{j_2+1}]$.

So we have for $j_1 - l < i \le j_1$ and $\hat{x} \in [u_{j_1}, u_{j_1+1}]$

$$|\tau_i - \hat{x}| \le (l+1)|T|,$$

and for $j_2 - l < j \le j_2$ and $\hat{y} \in [v_{j_2}, v_{j_2+1}]$

$$|\xi_j - \hat{y}| \le (l+1)|T|.$$

Therefore we have

$$\begin{aligned}\max_{\substack{j_1+1-l \le i \le j_1 \\ j_2+1-l \le j \le j_2}} |g(\hat{x},\hat{y}) - g(\tau_i, \xi_j)| \le & \max\{|g(x_1,y_1) - g(x_2,y_2)| : \\ & |x_1 - x_2| \le (l+1)|T|, |y_1 - y_2| \le (l+1)|T|\} \\ = & \omega(g; (l+1)|T|) \\ = & (l+1)\omega(g; |T|),\end{aligned} \tag{2.26}$$

the last inequality is by subadditivity of $\omega(g; h)$.

By (2.26) we have

$$\|g - Ag\|_\infty = \sup_{\substack{L_1 \le x \le U_1 \\ L_2 \le y \le U_2}} |g(x,y) - Ag(x,y)| \le (l+1)\omega(g; |T|),$$

which means the distance between g and $\psi_{l,l}$

$$d(g, \psi_{l,l}) = \inf_{s \in \psi_{l,l}} \|g - s\| \le (l+1)\omega(g; |T|), \tag{2.27}$$

where $\psi_{l,l}$ denotes the set of all tensor B-splines with order l in both directions. Because the distance of function g from $\psi_{l,l}$ is the same as the distance of the function $g-s$ from $\psi_{l,l}$ for $s \in \psi_{l,l}$, by (2.27) we have

$$d(g,\psi_{l,l}) = d(g-s,\psi_{l,l}) \leq (l+1)\omega(g-s,|T|). \tag{2.28}$$

Furthermore, since g has bounded partial derivatives, we have

$$\begin{aligned}
\omega(g-s,|T|) &= \max_{\substack{|x_1-x_2|\leq|T| \\ |y_1-y_2|\leq|T|}} |(g-s)(x_1,y_1)-(g-s)(x_2,y_2)| \\
&\leq \max_{|y_1-y_2|\leq|T|} |(g-s)(x_1,y_1)-(g-s)(x_1,y_2)| \\
&\quad + \max_{|x_1-x_2|\leq|T|} |(g-s)(x_1,y_2)-(g-s)(x_1,y_2)| \\
&\leq \|\frac{\partial(g-s)}{\partial y}\|_\infty |T| + \|\frac{\partial(g-s)}{\partial x}\|_\infty |T|
\end{aligned}$$

Then by (2.28),

$$d(g,\psi_{l,l}) \leq (l+1)|T|(\|\frac{\partial(g-s)}{\partial y}\|_\infty + \|\frac{\partial(g-s)}{\partial x}\|_\infty). \tag{2.29}$$

Since we know $\psi_{l,l-1} = \{\frac{\partial s}{\partial y} : s \in \psi_{l,l}\}$ and $\psi_{l-1,l} = \{\frac{\partial s}{\partial x} : s \in \psi_{l,l}\}$, (2.29) implies

$$d(g,\psi_{l,l}) \leq (l+1)|T|\{d(\frac{\partial g}{\partial x},\psi_{l-1,l}) + d(\frac{\partial g}{\partial y},\psi_{l,l-1})\}. \tag{2.30}$$

Proceeding in this way as we derive (2.30), finally we get

$$\begin{aligned}
&d(g,\psi_{l,l}) \\
\leq& c|T|^{w-1}\{d(\frac{\partial^{w-1}g}{\partial x^{w-1}},\psi_{l-w+1,l}) + d(\frac{\partial^{w-1}g}{\partial x^{w-2}\partial y},\psi_{l-w+2,l-1}) + \cdots + d(\frac{\partial^{w-1}g}{\partial y^{w-1}},\psi_{l,l-w+1})\} \\
\leq& c|T|^{w-1}\{\omega(\frac{\partial^{w-1}g}{\partial x^{w-1}},|T|) + \omega(\frac{\partial^{w-1}g}{\partial x^{w-2}\partial y},|T|) + \cdots + \omega(\frac{\partial^{w-1}g}{\partial y^{w-1}},|T|)\}
\end{aligned}$$

$$
\begin{aligned}
&\leq c|T|^w\{\|\frac{\partial^w g}{\partial x^w}\|_\infty + \|\frac{\partial^w g}{\partial x^{w-1}\partial y}\|_\infty + \cdots + \|\frac{\partial^w g}{\partial y^w}\|_\infty\} \\
&\leq c|T|^w \max_{0\leq m\leq w} \|\frac{\partial^w g}{\partial x^m \partial y^{w-m}}\|_\infty .
\end{aligned}
$$

□

In the following Lemma 2.21 and Lemma 2.22, we derive the bounds of the bracketing numbers of the set of monotone B-spline functions and the set of partially monotone tensor B-spline functions. We only give the proof for Lemma 2.22. The proof for Lemma 2.21 is similar to that for Lemma 2.22 but much simpler. Lemma 2.20 is applied in the proof of Lemma 2.22.

Lemma 2.20. *(Lemma 5, Shen and Wong (1994)) Let S be a $(n^{1/2}\sigma)$-sphere in R^n, that is, $S = \{x = (x_1, \ldots, x_n) \in R^n : \sum_{i=1}^n x_i^2 \leq n\sigma^2\}$. Let $\|\cdot\|_\infty$ be the usual L_∞-norm in R^n. Then $H(\epsilon, S, \|\cdot\|_\infty) \leq cn\log(\frac{\sigma}{\epsilon})$, for some constant $c > 0$ and $\epsilon < \sigma$.*

Lemma 2.21. $\Theta = \{\phi : \phi(s) = \sum_{i=1}^p \beta_i N_i^l(s), \|\phi\|_\infty < \delta\}$*, where each $\beta_i \geq 0$ and $\beta_{i+1} \geq \beta_i$ for $i = 1, \ldots, p-1$, $N_i^l(u)$'s are the B-spline basis functions with the knot sequence $\{u_i\}_1^{p+l}$ satisfying $L = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U$. Then $H_{[\,]}(\epsilon, \Theta, \|\cdot\|_\infty) \leq cp\log(\delta/\epsilon)$, for some constant $c > 0$ and $\epsilon < \delta$.*

Lemma 2.22. $\Theta = \{\phi : \phi(s,t) = \sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t), \|\phi\|_\infty < \delta\}$*, where $0 \leq \alpha_{1,j} \leq \alpha_{2,j} \leq \cdots \leq \alpha_{p,j}$ for $j = 1, \ldots, q$ and $0 \leq \alpha_{i,1} \leq \alpha_{i,2} \leq \cdots \leq \alpha_{i,q}$ for $i = 1, \ldots, p$, $N_i^{(1),l}(u)$'s and $N_j^{(2),l}(t)$'s are the B-spline basis functions with the knot sequence $\{u_i\}_1^{p+l}$ satisfying $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U_1$ and the knot sequence $\{v_i\}_1^{q+l}$ satisfying $L_2 = v_1 = \cdots = v_l < v_{l+1} < \cdots < v_q < v_{q+1} =$*

$\cdots = v_{q+l} = U_2$, *respectively. Then* $H_{[\,]}(\epsilon, \Theta, \|\cdot\|_\infty) \le cpq\log(\delta/\epsilon)$, *for some constant* $c > 0$ *and* $\epsilon < \delta$.

Proof. The basic idea of the proof is using the bracket number of a set in Euclidean space R^{pq} to bound the bracket number of Θ. The detailed proof is given as follows.

For any $\phi \in \Theta$, by Lemma 2.7

$$\|\phi\|_\infty^2 = (\phi(U_1, U_2))^2.$$

By property (B1) and property (B2) in Section 2.1.1

$$(\phi(U_1, U_2))^2 = (\alpha_{p,q} N_p^{(1),l}(U_1) N_q^{(2),l}(U_2))^2 = \alpha_{p,q}^2.$$

Since $\|\phi\|_\infty^2 = \alpha_{p,q}^2 \ge \frac{1}{pq}\sum_{i=1}^p \sum_{j=1}^q a_{i,j}^2$ and $\|\phi\|_\infty^2 \le \delta^2$, we have for the coefficients $(\alpha_{1,1}, \cdots, \alpha_{p,q})$ of ϕ,

$$\sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j}^2 \le pq\|\phi\|_\infty^2 \le pq\delta^2. \tag{2.31}$$

Let

$$S = \{\underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{pq}) : \quad \sum_{i=1}^p \sum_{j=1}^q \alpha_{ij}^2 \le pq\delta^2\}.$$

Lemma 2.20 indicates that there exist ϵ-balls $B_1, B_2, \cdots, B_{[(\frac{\delta}{\epsilon})^{cpq}]}$ centered at $\underline{\alpha}^{(1)} = (\alpha_{1,1}^{(1)}, \cdots, \alpha_{p,q}^{(1)})$, $\underline{\alpha}^{(2)} = (\alpha_{1,1}^{(2)}, \cdots, \alpha_{p,q}^{(2)})$, $\cdots$, $\underline{\alpha}^{([(\frac{\delta}{\epsilon})^{cpq}])} = (\alpha_{1,1}^{([(\frac{\delta}{\epsilon})^{cpq}])}, \cdots, \alpha_{p,q}^{([(\frac{\delta}{\epsilon})^{cpq}])})$, respectively, which cover S.

Let

$$\psi^{(k)}(s,t) = \sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j}^{(k)} N_i^{(1),l}(s) N_j^{(2),l}(t)$$

and

$$\Psi_1^{(k)} = \{\psi : \|\psi - \psi^{(k)}\| \le \epsilon \text{ and } \psi \in \Psi\}$$

for $k = 1, \cdots, [(\frac{\delta}{\epsilon})^{cpq}]$, where $\Psi = \{\psi : \psi(s,t) = \sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t)\}$.

Then $\{\Psi_1^{(k)} : k = 1, \cdots, [(\frac{\delta}{\epsilon})^{cpq}]\}$ constitute a set of ϵ-balls for Ψ.

In what follows, we show $\{\Psi_1^{(k)} : k = 1, \cdots, [(\frac{\delta}{\epsilon})^{cpq}]\}$ cover Θ.

For any $\psi(s,t) = \sum_{i=1}^{p} \sum_{j=1}^{q} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t) \in \Theta$, its coefficients $\underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{p,q}) \in S$ by (2.31).

By the fact that ϵ-balls $B_1, B_2, \cdots, B_{[(\frac{\delta}{\epsilon})^{cpq}]}$ cover S, there exists m with $1 \le m \le [(\frac{\delta}{\epsilon})^{cpq}]$, such that

$$\|\underline{\alpha} - \underline{\alpha}^{(m)}\|_\infty = \max_{\substack{i=1,\cdots,p \\ j=1,\cdots,q}} |\alpha_{i,j} - \alpha_{i,j}^{(m)}| \le \epsilon.$$

Then

$$\begin{aligned}
|\psi^{(m)}(s,t) - \psi(s,t)| &= |\sum_{i=1}^{p} \sum_{j=1}^{q} (\alpha_{i,j}^{(m)} - \alpha_{i,j}) N_i^{(1),k}(u) N_j^{(2),k}(t)| \\
&\le \max_{\substack{i=1,\cdots,p \\ j=1,\cdots,q}} |(\alpha_{i,j}^{(m)} - \alpha_{i,j})| \sum_{i=1}^{p} \sum_{j=1}^{q} N_i^{(1),4}(u) N_j^{(2),4}(t) \\
&= \max_{\substack{i=1,\cdots,p \\ j=1,\cdots,q}} |(\alpha_{i,j}^{(m)} - \alpha_{i,j})| \\
&\le \epsilon.
\end{aligned}$$

Hence,

$$\|\psi^{(m)} - \psi\| \le \epsilon.$$

In a word, for any $\psi \in \Theta$, there exist $\Psi_1^{(m)}$ with $1 \le m \le [(\frac{\delta}{\epsilon})^{cpq}]$, such that $\psi \in \Psi_1^{(m)}$, which means $\{\Psi_1^{(k)} : k = 1, \cdots, [(\frac{\delta}{\epsilon})^{cpq}]\}$ cover Θ.

So the ϵ-covering number of Θ is bounded by $[(\frac{\delta}{\epsilon})^{cpq}]$, or

$$H(\epsilon, \Theta, \|\cdot\|_\infty) \le cpq \log(\delta/\epsilon), \tag{2.32}$$

where $\|\cdot\|_\infty$ is the usual L_∞-norm in the tensor spline space Ψ.

We also know

$$H_{[\,]}(2\epsilon, \Theta, \|\cdot\|_\infty) \leq H(\epsilon, \Theta, \|\cdot\|_\infty). \tag{2.33}$$

(2.32) and (2.33) result in

$$H_{[\,]}(\epsilon, \Theta, \|\cdot\|_\infty) \leq cpq \log(\delta/\epsilon).$$

□

Remark 2.4. *In the proof of Theorem 4.2 (convergence rate), we use the fact that* $\delta \leq 1$*, then it is obvious that* $H_{[\,]}(\epsilon, \Theta, \|\cdot\|_\infty) \leq cpq \log(1/\epsilon)$ *by both Lemma 2.21 and Lemma 2.22.*

CHAPTER 3
TENSOR SPLINE-BASED SIEVE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION METHOD

3.1 Likelihood of Bivariate Current Status Data

Suppose we have n independent bivariate current status data $\{(c_{1,k}, \delta_{1,k}, c_{2,k}, \delta_{2,k}) : k = 1, 2, \cdots, n\}$, where $\delta_{j,k} = 1_{[t_{j,k} \le c_{j,k}]}, j = 1, 2$.

Assume $(c_{1,k}, c_{2,k})$'s constitute of a sample of n from the bivariate random censoring times (C_1, C_2) and $(t_{1,k}, t_{2,k})$'s constitute of a sample of n from the bivariate random event times (T_1, T_2). Hence $(\delta_{1,k}, \delta_{2,k})$'s constitute of a sample of n from the bivariate random indicator variables (Δ_1, Δ_2) with $\Delta_j = 1_{[T_j \le C_j]}, j = 1, 2$. We also assume (T_1, T_2) and (C_1, C_2) are independent.

The joint distribution of $(C_1, \Delta_1, C_2, \Delta_2)$ is generally given by $P(C_1 \le c_1, C_2 \le c_2, \Delta_1 = \delta_1, \Delta_2 = \delta_2)$.

If $\delta_1 = 1$ and $\delta_2 = 1$, by the independence between (T_1, T_2) and (C_1, C_2), we have

$$
\begin{aligned}
&P(C_1 \le c_1, C_2 \le c_2, \Delta_1 = 1, \Delta_2 = 1) \\
=&P(C_1 \le c_1, C_2 \le c_2, T_1 \le C_1, T_2 \le C_2) \\
=&\int_0^{c_1} \int_0^{c_2} \int_0^{C_1} \int_0^{C_2} f(T_1, T_2, C_1, C_2) dT_2 dT_1 dC_2 dC_1 \\
=&\int_0^{c_1} \int_0^{c_2} \{f(C_1, C_2) \int_0^{C_1} \int_0^{C_2} f(T_1, T_2) dT_2 dT_1\} dC_2 dC_1 \\
=&\int_0^{c_1} \int_0^{c_2} f(C_1, C_2) P(T_1 \le C_1, T_2 \le C_2) dC_2 dC_1.
\end{aligned}
$$

By taking the mixture derivative with respect to c_1 and c_2, we obtain the joint density function as

$$f(c_1, c_2, \delta_1 = 1, \delta_2 = 1) = f_{C_1,C_2}(c_1, c_2)P(T_1 \le c_1, T_2 \le c_2),$$

where $f_{C_1,C_2}(c_1, c_2)$ is the bivariate density function of (C_1, C_2). Similarly, if $\delta_1 = 1$ and $\delta_2 = 0$, then $f(c_1, c_2, \delta_1 = 1, \delta_2 = 0) = f_{C_1,C_2}(c_1, c_2)P(T_1 \le c_1, T_2 > c_2)$; if $\delta_1 = 0$ and $\delta_2 = 1$, then $f(c_1, c_2, \delta_1 = 0, \delta_2 = 1) = f_{C_1,C_2}(c_1, c_2)P(T_1 > c_1, T_2 \le c_2)$; if $\delta_1 = 0$ and $\delta_2 = 0$, then $f(c_1, c_2, \delta_1 = 1, \delta_2 = 0) = f_{C_1,C_2}(c_1, c_2)P(T_1 > c_1, T_2 > c_2)$.

Suppose the censoring times (C_1, C_2) are noninformative to the event times (T_1, T_2), then the joint density function of censoring times $f_{C_1,C_2}(x_1, x_2)$ can be ignored for computing the maximum likelihood estimate of the joint distribution function of the bivariate event times (T_1, T_2) with bivariate current status data. So

$$\begin{aligned} l_n(\cdot;\text{data}) = \sum_{k=1}^{n} \{ & \delta_{1,k}\delta_{2,k} \log P(T_1 \le c_{1,k}, T_2 \le c_{2,k}) \\ & + \delta_{1,k}(1 - \delta_{2,k}) \log P(T_1 \le c_{1,k}, T_2 > c_{2,k}) \\ & + (1 - \delta_{1,k})\delta_{2,k} \log P(T_1 > c_{1,k}, T_2 \le c_{2,k}) \\ & + (1 - \delta_{1,k})(1 - \delta_{2,k}) \log P(T_1 > c_{1,k}, T_2 > c_{2,k}) \}. \end{aligned} \tag{3.1}$$

If we denote F as the joint distribution function of event time (T_1, T_2) and F_1 and F_2 as the marginal distribution functions of F, the log-likelihood (3.1) can be rewritten as

$$
\begin{aligned}
l_n(F, F_1, F_2; \text{data}) = \sum_{k=1}^{n} \{ & \delta_{1,k}\delta_{2,k} \log F(c_{1,k}, c_{2,k}) \\
& + \delta_{1,k}(1-\delta_{2,k}) \log(F_1(c_{1,k}) - F(c_{1,k}, c_{2,k})) \\
& + (1-\delta_{1,k})\delta_{2,k} \log(F_2(c_{2,k}) - F(c_{1,k}, c_{2,k})) \\
& + (1-\delta_{1,k})(1-\delta_{2,k}) \log(1 - F_1(c_{1,k}) - F_2(c_{2,k}) \\
& + F(c_{1,k}, c_{2,k}))\}.
\end{aligned} \tag{3.2}
$$

3.2 Spline-based Maximum Likelihood Estimation

Suppose the observation times C_1 and C_2 are within intervals $[l_1, u_1]$ and $[l_2, u_2]$, respectively. For $L_1 < l_1$, $L_2 < l_2$, $u_1 < U_1$ and $u_2 < U_2$, we define a class

$$
\mathcal{F} = \{(F(s,t), F_1(s), F_2(t)) : \text{ for } (s,t) \in [L_1, U_1] \times [L_2, U_2]\},
$$

where F, F_1 and F_2 satisfy the following conditions:

$$
\begin{aligned}
& 0 \leq F(s,t), \\
& F(s^{'}, t) \leq F(s^{''}, t), \\
& F(s, t^{'}) \leq F(s, t^{''}), \\
& [F(s^{''}, t^{''}) - F(s^{'}, t^{''})] - [(F(s^{''}, t^{'}) - F(s^{'}, t^{'})] \geq 0, \\
& F_1(s) - F(s,t) \geq 0 \\
& F_2(t) - F(s,t) \geq 0, \\
& [F_1(s^{''}) - F_1(s^{'})] - [F(s^{''}, t) - F(s^{'}, t)] \geq 0, \\
& [F_2(t^{''}) - F_2(t^{'})] - [F(s, t^{''}) - F(s, t^{'})] \geq 0, \\
& [1 - F_1(s)] - [F_2(t) - F(s,t)] \geq 0,
\end{aligned} \tag{3.3}
$$

for $s' \leq s''$ with s' and s'' on $[L_1, U_1]$, and $t' \leq t''$ with t' and t'' on $[L_2, U_2]$.

Lemma 3.1. *$\mathcal{F}$ is the class of all joint distribution functions and their marginal distribution functions in the closed region $[L_1, U_1] \times [L_2, U_2]$. In other words, each element of $\mathcal{F}$ is a vector-valued function, which corresponds to a joint distribution function and its two marginal distribution functions.*

Proof. (i) For any random variable (S, T), let $F(s,t) = Pr(S \leq s, T \leq t)$, $F_1(s) = Pr(S \leq s)$ and $F_2(s) = Pr(T \leq t)$, then F, F_1 and F_2 satisfy (3.3) in $[L_1, U_1] \times [L_2, U_2]$ by the following arguments.

(a) The first three conditions given in (3.3) are obvious.

(b) $[F(s'', t'') - F(s', t'')] - [F(s'', t') - F(s', t')] = Pr(s' \leq S \leq s'', t' \leq T \leq t'') \geq 0$.

(c) $F_1(s) - F(s,t) = Pr(S \leq s, T \geq t) \geq 0$.

(d) $F_2(t) - F(s,t) = Pr(S \geq s, T \leq t) \geq 0$.

(e) $[F_1(s'') - F_1(s')] - [F(s'', t) - F(s', t)] = Pr(s' \leq S \leq s'', T \geq t) \geq 0$.

(f) $[F_2(t'') - F_2(t')] - [F(s, t'') - F(s, t')] = Pr(S \geq s, t' \leq T \leq t'') \geq 0$.

(g) $[1 - F_1(s)] - [F_2(t) - F(s,t)] = Pr(S \geq s, T \geq t) \geq 0$.

(ii) On the other hand, for any $(F, F_1, F_2) \in \Omega$, conditions given in (3.3) guarantee that F is a joint distribution in the closed region $[L_1, U_1] \times [L_2, U_2]$, and F_1 and F_2 are both possible marginal distribution functions of F on $[L_1, U_1]$ and $[L_2, U_2]$, respectively.

□

Lemma 3.1 indicates that $(F_0, F_{0,1}, F_{0,2}) \in \mathcal{F}$, when F_0 is the true joint distribution function of event times (T_1, T_2), and $F_{0,1}$ and $F_{0,1}$ are marginal distribution functions of F_0.

Based on Lemma 3.1, the nonparametric maximum likelihood estimate of $(F_0, F_{0,1}, F_{0,2})$ is defined as

$$(\hat{F}, \hat{F}_1, \hat{F}_2) = \arg\max_{(F,F_1,F_2)\in\mathcal{F}} l_n(F, F_1, F_2; \text{data}). \tag{3.4}$$

This means the nonparametric maximum likelihood estimator is sought by maximizing (3.2) over $\mathcal{F}$ with respect to $F(c_{1,k}, c_{2,k})$, $F_1(c_{1,k})$ and $F_2(c_{2,k})$ for $i = 1, \ldots, n$. This estimator is the extension of the nonparametric maximum likelihood estimator of univariate distribution function with univariate current status data. The univariate nonparametric maximum likelihood estimation problem has been thoroughly studied by Groenoboom and Wellner (1992). They studied the asymptotic properties of the nonparametric maximum likelihood estimator of the distribution function. They also showed that the NPMLE can be easily computed using convex minorant algorithm.

Problem (3.4) is much more complicated in view of number of unknown quantities needed to be estimated and the nature of the constraints. Theoretically, consistency of this estimator will be much more difficult to study, the convergence rate is even more difficult to be established due to the fact that computing the bracket number of the set of bounded partially monotone bivariate functions is still an open question (Song and Wellner (2002)). Numerically, the only available method for this problem in the literate is the nonparametric likelihood estimation procedure proposed by Betensky and Finkelstein (1999) if current status data are treated as interval-censored data case 2. For the Betensky-Finkelstein method, first the nonzero mass rectangles (or points) are determined, second the estimated mass for selected rectangles (or points) are computed through maximizing the likelihood. However, this method has some disadvantages. First, this method is designed for interval censored data case 2, and the estimation will be much worse when the method is applied for the current status data which contains less information about the event time distribution. Second, this procedure is not theoretically justified for consistency and convergence rate. Third, Yu et al. (2000) pointed out that the estimation through this procedure is not necessarily unique. Finally, if the sample size is large, the first stage of this procedure can be very time-consuming.

To overcome the theoretical and numerical difficulties in the nonparametric estimation problem, a spline-based sieve maximum likelihood estimation procedure is applied to estimate the unknown bivariate distribution function. For this procedure the unknown function in the log likelihood is approximated by a linear combination of tensor spline basis functions to form a sieve log likelihood. Then maximizing the log likelihood with respect to the unknown function converts to maximizing the spline-based sieve log likelihood with respect to the unknown coefficients of the tensor spline basis functions. The success of spline-based sieve nonparametric maximum likelihood estimation is given by Lu et al. (2007), who studied nonparametric likelihood-based estimators of the mean function of counting processes with panel count data by using monotone splines to approximate the mean function. Other applications of spline-based sieve maximum likelihood estimation can be found in Shen (1998), Zhang et al. (2009), Lu et al. (2009).

3.2.1 B-spline-based Estimation

In this section, the spline-based sieve nonparametric maximum likelihood estimation problem is represented as a constrained optimization problem with respect to the coefficients of the tensor B-splines.

B-spline basis functions $\{N_i^{(1),l}(s) : i = 1, \ldots, p_n\}$ and $\{N_j^{(2),l}(t) : j = 1, \ldots, q_n\}$ are constructed in $[L_1, U_1] \times [L_2, U_2]$ with the knot sequence $\{u_i^{p_n+l}\}$ satisfying $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_{p_n} < u_{p_n+1} = u_{p_n+l} = U_1$ and knot sequence $\{v_j^{q_n+l}\}$ satisfying $L_2 = v_1 = \cdots = v_l < v_{l+1} < \cdots < v_{q_n} < v_{q_n+1} = v_{q_n+l} = U_2$, where $p_n = O(n^v)$ and $q_n = O(n^v)$ for some $0 < v < 1$.

Let

$$\Omega_n = \{\tau = (F, F_1, F_2) : F(s,t) = \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t)\},$$

$$F_1(s) = \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s),$$

$$F_2(t) = \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(t),$$

with $\underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{p_n,q_n}), \underline{\beta} = (\beta_1, \cdots, \beta_{p_n}),$ and $\underline{\gamma} = (\gamma_1, \cdots, \gamma_{q_n})$

subject to the following conditions in (3.5)},

$$
\begin{aligned}
&\alpha_{1,1} \geq 0,\\
&\alpha_{1,j+1} - \alpha_{1,j} \geq 0 \text{ for } j = 1, \ldots, q_n - 1,\\
&\alpha_{i+1,1} - \alpha_{i,1} \geq 0 \text{ for } i = 1, \ldots, p_n - 1,\\
&(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1, j = 1, \ldots, q_n - 1,\\
&\beta_1 - \alpha_{1,q_n} \geq 0, \qquad (3.5)\\
&(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1,\\
&\gamma_1 - \alpha_{p_n,1} \geq 0,\\
&(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \text{ for } j = 1, \ldots, q_n - 1,\\
&\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.
\end{aligned}
$$

The following Lemma implies the conditions given in (3.5) are closely related to those given in (3.3).

Lemma 3.2. $\Omega_n \subseteq \mathcal{F}$.

Proof. (i) By $\alpha_{1,1} \geq 0$, it is obvious that $0 \leq F(s,t)$.

(ii) By Property (B3) in Section 2.1.1,

$$\frac{\partial F(s,t)}{\partial s} = \sum_{i=1}^{p_n-1} \sum_{j=1}^{q_n} \frac{(l-1)(\alpha_{i+1,j} - \alpha_{i,j})}{u_{i+l} - u_{i+1}} N_{i+1}^{(1),l-1}(s) N_j^{(2),l}(t),$$

and by the properties $\alpha_{i+1,1} - \alpha_{i,1} \geq 0$ and $(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0$, we have

$$\alpha_{i+1,j} - \alpha_{i,j} \geq 0.$$

It is then followed by

$$\frac{\partial F(s,t)}{\partial s} \geq 0, \text{ or } F(s^{'},t) \leq F(s^{''},t).$$

(iii) By the similar arguments as in (ii), it can be shown that $F(s,t^{'}) \leq F(s,t^{''})$.

(iv) By Property (B3) in Section 2.1.1,

$$\frac{\partial^2 F(s,t)}{\partial s \partial t} = \sum_{i=1}^{p_n-1}\sum_{j=1}^{q_n-1}(l-1)^2 \frac{\alpha_{i+1,j+1}-\alpha_{i,j+1}-\alpha_{i+1,j}+\alpha_{i,j}}{(s_{i+l}-s_{i+1})(t_{j+l}-t_{j+1})} N_{i+1}^{(1),l-1}(s) N_{j+1}^{(2),l-1}(t).$$

Then by the property $(\alpha_{i+1,j+1}-\alpha_{i+1,j})-(\alpha_{i,j+1}-\alpha_{i,j}) \geq 0$,

$$\frac{\partial^2 F(s,t)}{\partial s \partial t} \geq 0, \text{ or } F(s^{''},t^{'}) - F(s^{'},t^{'}) \leq F(s^{''},t^{''}) - F(s^{'},t^{''}).$$

(v) By $\beta_1 - \alpha_{1,q_n} \geq 0$ and $(\beta_{i+1}-\beta_i)-(\alpha_{i+1,q_n}-\alpha_{i,q_n}) \geq 0$, $\beta_i - \alpha_{i,q_n} \geq 0$. Then $F(s,t) \leq F_1(s)$.

(vi) By the similar arguments as in (v), it can be shown that $F(s,t) \leq F_2(t)$.

(vii) By Property (B3) in Section 2.1.1,

$$\frac{dF_1(s)}{ds} = \sum_{i=1}^{p_n-1} \frac{(l-1)(\beta_{i+1}-\beta_i)}{u_{i+l}-u_{i+1}} N_{i+1}^{(1),l-1}(s).$$

Then by the property $(\beta_{i+1}-\beta_i)-(\alpha_{i+1,q_n}-\alpha_{i,q_n}) \geq 0$,

$$\frac{\partial(F_1(s)-F(s,t))}{\partial s} \geq 0, \text{ or } F(s^{''},t) - F(s^{'},t) \leq F_1(s^{''}) - F_1(s^{'}).$$

(viii) By the similar arguments as in (vii), it can be shown that $F(s,t^{''}) - F(s,t^{'}) \leq F_2(t^{''}) - F_2(t^{'})$.

(ix) Since $F_1(U_1) = \beta_{p_n} N_{p_n}^{(1),l}(U_1) = \beta_{p_n}$, $F_2(U_2) = \gamma_{q_n} N_{q_n}^{(2),l}(U_2) = \gamma_{q_n}$, and $F(U_1,U_2) = \alpha_{p_n,q_n} N_{p_n}^{(1),l}(U_1) N_{q_n}^{(2),l}(U_2) = \alpha_{p_n,q_n}$, then

$$\begin{aligned} F_2(U_2) - F(U_1,U_2) &= \gamma_{q_n} - \alpha_{p_n,q_n} \\ &\leq 1 - \beta_{p_n} \\ &= 1 - F_1(U_1). \end{aligned}$$

Moreover, $\frac{dF_1(s)}{ds} \geq \frac{\partial F(s,t)}{\partial s}$ and $\frac{dF_2(t)}{dt} \geq \frac{\partial F(s,t)}{\partial t}$ guarantee $F_1(U_1) - F_1(s) \geq F(U_1, U_2) - F(s, U_2)$ and $F_2(U_2) - F_2(t) \geq F(U_1, U_2) - F(U_1, t)$, respectively.

Then

$$
\begin{aligned}
& 1 - F_1(s) - F_2(t) + F(s,t) \\
=& \{1 - F_1(U_1) + F_1(U_1) - F_1(s)\} \\
& - \{F_2(t) - F(U_1, t) + F(U_1, t) - F(s,t)\} \\
\geq& \{F_2(U_2) - F(U_1, U_2) + F(U_1, U_2) - F(s, U_2)\} \\
& - \{F_2(t) - F(U_1, t) + F(U_1, t) - F(s,t)\} \\
=& \{F_2(U_2) - F_2(t) - F(U_1, U_2) + F(U_1, t)\} \\
& + \{F(U_1, U_2) - F(s, U_2) - F(U_1, t) + F(s,t)\} \\
\geq& \, 0.
\end{aligned}
$$

□

To obtain the tensor B-spline-based sieve likelihood with bivariate current status data, $\tau = (F, F_1, F_2) \in \Omega_n$ is substituted into (3.2) which results in

$$
\begin{aligned}
\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot) = \sum_{k=1}^{n} \{ & \delta_{1,k}\delta_{2,k} \log \sum_{i=1}^{p_n} \sum_{j=2}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \\
& + \delta_{1,k}(1-\delta_{2,k}) \log\{ \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \} \\
& + (1-\delta_{1,k})\delta_{2,k} \log\{ \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \} \\
& + (1-\delta_{1,k})(1-\delta_{2,k}) \log\{ 1 - \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \\
& - \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k}) + \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \} \}.
\end{aligned} \tag{3.6}
$$

Hence, for the partially monotone tensor B-spline-based sieve nonparametric maximum likelihood estimation problem, we look for $\hat{\tau} = (\hat{F}, \hat{F}_1, \hat{F}_2)$ that maximizes $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot)$ given by (3.6) over Ω_n.

3.2.2 I-spline-based Estimation

According to Lemma 2.3, the I-splines can be expressed by the sum of the B-splines as follows.

$$
I_i^{(1),l-1}(s) = \sum_{m=i}^{p_n} N_m^{(1),l}(s), \tag{3.7}
$$

and

$$
I_i^{(2),l-1}(t) = \sum_{m=i}^{q_n} N_m^{(2),l}(t). \tag{3.8}
$$

Let

$$\Theta_n = \{\tau = (F, F_1, F_2) : F(s,t) = \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(s) I_j^{(2),l-1}(t)\},$$

$$F_1(s) = \sum_{i=1}^{p_n}\{\sum_{j=1}^{q_n} \eta_{i,j} + \omega_i\} I_i^{(1),l-1}(s),$$

$$F_2(t) = \sum_{j=1}^{q_n}\{\sum_{i=1}^{p_n} \eta_{i,j} + \pi_j\} I_j^{(2),l-1}(t)$$

with $\underline{\eta} = (\eta_{1,1}, \cdots, \eta_{p_n,q_n}), \underline{\omega} = (\omega_1, \cdots, \omega_{p_n}),$ and $\underline{\pi} = (\pi_1, \cdots, \pi_{q_n})$

subject to the following conditions (3.9)},

$$\begin{aligned} &\eta_{i,j} \geq 0 \text{ for } i = 1, \cdots, p_n, j = 1, \cdots, q_n, \\ &\omega_i \geq 0, i = 1, \ldots, p_n, \\ &\pi_j \geq 0, j = 1, \ldots, q_n, \\ &\sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \eta_{i,j} + \sum_{i=1}^{p_n} \omega_i + \sum_{j=1}^{q_n} \pi_j \leq 1. \end{aligned} \tag{3.9}$$

The relationship between the constraints given in (3.9) and those given in (3.5) is summarized in the following lemma.

Lemma 3.3. *Set Ω_n is equivalent to set Θ_n.*

Proof. Let

$$\alpha_{i,j} = \sum_{m=1}^{i}\sum_{n=1}^{j} \eta_{m,n},$$

$$\beta_i = \sum_{m=1}^{i}\{\sum_{j=1}^{q_n} \eta_{m,j} + \omega_m\},$$

and

$$\gamma_j = \sum_{n=1}^{j}\{\sum_{i=1}^{p_n} \eta_{i,n} + \pi_n\}.$$

By (3.7) and (3.8) (the relationships between the B-splines and the I-splines), it follows that

$$\sum_{i=1}^{p_n}\sum_{j=1}^{q_n}\alpha_{i,j}N_i^{(1),l}(s)N_j^{(2),l}(t)=\sum_{i=1}^{p_n}\sum_{j=1}^{q_n}\eta_{i,j}I_i^{(1),l-1}(s)I_j^{(2),l-1}(t)\},$$

$$\sum_{i=1}^{p_n}\beta_i N_i^{(1),l}(s)=\sum_{i=1}^{p_n}\{\sum_{j=1}^{q_n}\eta_{i,j}+\omega_i\}I_i^{(1),l-1}(s),$$

and

$$\sum_{j=1}^{q_n}\gamma_j N_j^{(2),l}(t)=\sum_{j=1}^{q_n}\{\sum_{i=1}^{p_n}\eta_{i,j}+\pi_j\}I_j^{(2),l-1}(t).$$

In what follows, we verify that condition (3.5) and (3.9) are equivalent.

(i) $\alpha_{1,1}=\eta_{1,1}$, then $\eta_{1,1}\geq 0$ is equivalent to $\alpha_{1,1}\geq 0$.

(ii) $\alpha_{1,j+1}-\alpha_{1,j}=\eta_{1,j+1}$, then $\eta_{1,j}\geq 0$ for $j=2,\ldots,q_n$ is equivalent to $\alpha_{1,j+1}-\alpha_{1,j}\geq 0$ for $j=1,\ldots,q_n-1$.

(iii) $\alpha_{i+1,1}-\alpha_{i,1}=\eta_{i+1,1}$, then $\eta_{i,1}\geq 0$ for $i=2,\ldots,p_n$ is equivalent to $\alpha_{i+1,1}-\alpha_{i,1}\geq 0$ for $i=1,\ldots,p_n-1$.

(iv) $(\alpha_{i+1,j+1}-\alpha_{i+1,j})-(\alpha_{i,j+1}-\alpha_{i,j})=\eta_{i+1,j+1}$, then $\eta_{i,j}\geq 0$ for $i=2,\ldots,p_n$ and $j=2,\ldots,q_n$, is equivalent to $(\alpha_{i+1,j+1}-\alpha_{i+1,j})-(\alpha_{i,j+1}-\alpha_{i,j})\geq 0$ for $i=1,\ldots,p_n-1$ and $j=1,\ldots,q_n-1$.

(v) $\beta_1-\alpha_{1,q_n}=(\sum_{j=1}^{q_n}\eta_{1,j}+\omega_1)-(\sum_{j=1}^{q_n}\eta_{1,j})=\omega_1$, then $\omega_1\geq 0$ is equivalent to $\beta_1-\alpha_{1,q_n}\geq 0$.

(vi)

$$
\begin{aligned}
(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) &= (\sum_{m=1}^{i+1}\{\sum_{j=1}^{q_n} \eta_{m,j} + \omega_m\}) - \sum_{m=1}^{i}\{\sum_{j=1}^{q_n} \eta_{m,j} + \omega_m\}) \\
&- (\sum_{m=1}^{i+1}\sum_{j=1}^{q_n} \eta_{m,j} - \sum_{m=1}^{i}\sum_{j=1}^{q_n} \eta_{m,j}) \\
&= \omega_{i+1} \geq 0,
\end{aligned}
$$

then $\omega_i \geq 0$ for $i = 2, \ldots, p_n$ is equivalent to $(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0$ for $i = 1, \ldots, p_n - 1$.

(vii) $\gamma_1 - \alpha_{p_n,1} = (\sum_{i=1}^{p_n} \eta_{i,1} + \pi_1) - (\sum_{i=1}^{p_n} \eta_{i,1}) = \pi_1$, then $\pi_1 \geq 0$ is equivalent to $\gamma_1 - \alpha_{p_n,1} \geq 0$.

(viii)

$$
\begin{aligned}
(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) &= (\sum_{n=1}^{j+1}\{\sum_{i=1}^{p_n} \eta_{i,n} + \pi_n\}) - \sum_{n=1}^{j}\{\sum_{i=1}^{p_n} \eta_{i,n} + \pi_n\}) \\
&- (\sum_{n=1}^{j+1}\sum_{i=1}^{p_n} \eta_{i,n} - \sum_{n=1}^{j}\sum_{i=1}^{p_n} \eta_{i,n}) \\
&= \pi_{j+1} \geq 0.
\end{aligned}
$$

then $\pi_j \geq 0$ for $j = 2, \ldots, q_n$ is equivalent to $(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0$ for $j = 1, \ldots, q_n - 1$.

(ix)

$$
\begin{aligned}
\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} &= \sum_{i=1}^{p_n}\{\sum_{j=1}^{q_n} \eta_{i,j} + \omega_i\} + \sum_{j=1}^{q_n}\{\sum_{i=1}^{p_n} \eta_{i,j} + \pi_j\} \\
&- \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \eta_{i,j}
\end{aligned}
$$

$$= \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} + \sum_{i=1}^{p_n} \omega_i + \sum_{j=1}^{q_n} \pi_j,$$

which implies $\sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} + \sum_{i=1}^{p_n} \omega_i + \sum_{j=1}^{q_n} \pi_j \leq 1$ is equivalent to $\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1$.

The proof is complete. □

Lemma 3.3 indicates that the tensor spline-based sieve nonparametric maximum likelihood estimation problem can be equivalently represented as a constrained optimization problem with respect to the coefficients of the tensor B-splines or the tensor I-splines, with the constraints given by (3.5) and (3.9), respectively. Numerically, the constrained optimization problem with the tensor I-splines apparently has the advantage of simplicity in the constraints. Hence, we compute the tensor spline-based nonparametric estimate using the partially monotone tensor I-splines.

Similar to obtaining the likelihood in terms of the B-splines, $\tau = (F, F_1, F_2) \in \Theta_n$ is substituted into (3.2) to result in the tensor I-spline-based sieve likelihood of bivariate current status data as

$$
\begin{aligned}
\tilde{l}_n(\underline{\eta}, \underline{\omega}, \underline{\pi}; \cdot) = \sum_{k=1}^{n} \{ & \delta_{1,k}\delta_{2,k} \log \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \\
& + \delta_{1,k}(1-\delta_{2,k}) \log \{ \sum_{i=1}^{p_n} \{ \sum_{j=1}^{q_n} \eta_{i,j} + \omega_i \} I_i^{(1),l-1}(C_{1,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \} \\
& + (1-\delta_{1,k})\delta_{2,k} \log \{ \sum_{j=1}^{q_n} \{ \sum_{i=1}^{p_n} \eta_{i,j} + \pi_j \} I_j^{(2),l-1}(c_{2,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \} \\
& + (1-\delta_{1,k})(1-\delta_{2,k}) \log \{ 1 - \sum_{i=1}^{p_n} \{ \sum_{j=1}^{q_n} \eta_{i,j} + \omega_i \} I_i^{(1),l-1}(c_{1,k}) \\
& - \sum_{j=1}^{q_n} \{ \sum_{i=1}^{p_n} \eta_{i,j} + \pi_j \} I_j^{(2),l-1}(c_{2,k}) \\
& + \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \} \}.
\end{aligned}
\tag{3.10}
$$

Hence, for the partially monotone tensor I-spline-based sieve nonparametric maximum likelihood estimation problem, we look for $\hat{\tau} = (\hat{F}, \hat{F}_1, \hat{F}_2)$ that maximizes $\tilde{l}_n(\underline{\eta}, \underline{\omega}, \underline{\pi}; \cdot)$ given by (3.10) over Θ_n.

CHAPTER 4
ASYMPTOTIC PROPERTIES

In this chapter, we describe and prove the asymptotic results of the proposed tensor spline-based sieve nonparametric maximum likelihood estimator of the joint distribution function with bivariate current status data. The study of asymptotic properties of the tensor spline-based maximum likelihood estimator requires some regularity conditions for the event times and censoring times. Suppose the bivariate event times have joint distribution function $F_0(s,t)$ and marginal distribution functions $F_{0,1}(s)$ and $F_{0,2}(t)$. The following conditions sufficiently guarantee the results in the forthcoming theorems.

4.1 Regularity Conditions

(C1) $\frac{\partial F_0(s,t)}{\partial s}$, $\frac{\partial F_0(s,t)}{\partial t}$, $\frac{dF_{0,1}(s)}{ds}$ and $\frac{dF_{0,2}(t)}{dt}$ all have positive lower and upper bound in $[L_1, U_1] \times [L_2, U_2]$.

(C2) $\frac{\partial^2 F_0(s,t)}{\partial s \partial t}$ has positive lower bound b_0 in $[L_1, U_1] \times [L_2, U_2]$.

(C3) $F_0(s,t)$ has continuous mixed derivatives of order p, $\nabla^p_m g = \frac{\partial^p F_0(s,t)}{\partial s^m t^{p-m}}$ for $m = 1, 2, \ldots, p$, in the bounded region $[L_1, U_1] \times [L_2, U_2]$; $F_{0,1}(s)$ has continuous derivative $\frac{d^p F_{0,1}(s)}{ds^p}$ on $[L_1, U_1]$; and $F_{0,2}(t)$ has continuous derivative $\frac{d^p F_{0,2}(t)}{dt^p}$ on $[L_2, U_2]$.

(C4) Censoring times (C_1, C_2) are bivariate random variable taking values in $[l_1, u_1] \times [l_2, u_2]$, with $l_1 > L_1, u_1 < U_1, l_2 > L_2,$ and $u_2 < U_2$.

(C5) The density of (C_1, C_2) has positive lower bound at every point in $[l_1, u_1] \times [l_2, u_2]$.

Let

$$\begin{aligned}\Omega_{n,1} = \{\tau = (F, F_1, F_2) : & F(s,t) = \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t)\}, \\ & F_1(s) = \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s), \\ & F_2(t) = \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(t), \\ & \text{with } \underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{p_n,q_n}), \underline{\beta} = (\beta_1, \cdots, \beta_{p_n}), \\ & \text{ and } \underline{\gamma} = (\gamma_1, \cdots, \gamma_{q_n}) \\ & \text{subject to the following conditions in (4.1),} \\ & \text{and the following condition (Sp1) holds}\},\end{aligned}$$

$$\begin{aligned}& \alpha_{1,1} \geq 0, \\ & \alpha_{1,j+1} - \alpha_{1,j} \geq 0 \text{ for } j = 1, \ldots, q_n - 1, \\ & \alpha_{i+1,1} - \alpha_{i,1} \geq 0 \text{ for } i = 1, \ldots, p_n - 1, \\ & (\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq \frac{b_0 \min_{i_1} \Delta_{i_1}^{(u)} \min_{j_1} \Delta_{j_1}^{(v)}}{l^2} \\ & \text{for } i = 1, \ldots, p_n - 1, j = 1, \ldots, q_n - 1, \\ & \beta_1 - \alpha_{1,q_n} \geq 0, \\ & (\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1, \\ & \gamma_1 - \alpha_{p_n,1} \geq 0, \\ & (\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \text{ for } j = 1, \ldots, q_n - 1, \\ & \beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.\end{aligned} \tag{4.1}$$

(Sp1) Knot sequences $\{u_i\}_1^{p_n+l}$ and $\{v_j\}_1^{q_n+l}$ of the B-spline basis functions $\{N_i^{(1),l}\}_1^{p_n}$ and

$\{N_j^{(2),l}\}_1^{q_n}$, respectively, satisfy that both $\frac{\min_{i_1}\Delta_{i_1}^{(u)}}{\max_{i_1}\Delta_{i_1}^{(u)}}$ and $\frac{\min_{j_1}\Delta_{j_1}^{(v)}}{\max_{j_1}\Delta_{j_1}^{(v)}}$ have positive lower bounds which are not greater than 1, where $\Delta_i^{(u)} = u_{i+1} - u_i$ for $i = l, \ldots, p_n$ and $\Delta_j^{(v)} = v_{j+1} - v_j$ for $j = l, \ldots, q_n$.

In this chapter, the constraints set $\Omega_{n,1}$ is a subset of the constraints set Ω_n discussed in Chapter 3. We only discuss the asymptotic properties in $[l_1, u_1] \times [l_2, u_2]$ and let $\Omega_n^{'} = \{\tau(s,t) : \tau \in \Omega_{n,1},\ \text{for } (s,t) \in [l_1, u_1] \times [l_2, u_2]\}$.

Under (C4), the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot)$ over $\Omega_{n,1}$ is actually the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot)$ over $\Omega_n^{'}$, where $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot)$ is defined by (3.6). Let $\hat{\tau_n}$ maximizes $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \cdot)$ over $\Omega_n^{'}$. Then the asymptotic properties of $\hat{\tau_n}$ in terms of consistency and convergence rate are stated in the theorems given in the next sections.

In this chapter, the $L_r(Q)$-norm associated with probability measure Q is defined by

$$\|f\|_{L_r(Q)} = (Q|f|^r)^{1/r} = (\int |f|^r dQ)^{1/r}. \tag{4.2}$$

Then according to (4.2), $L_r(P_{C_1,C_2})$-norm, $L_r(P_{C_1})$-norm and $L_r(P_{C_2})$-norm are denoted as L_r-norm associated with the joint and marginal probability measures of observation times (C_1, C_2), and $L_r(P)$-norm is denoted as L_r-norm associated with the joint probability measure P of observation and event times (T_1, T_2, C_1, C_2).

Based on L_2-norm, the distance between $\tau_n = (F_n, F_{n,1}, F_{n,2})$ with $\tau_n \in \Omega_n^{'}$ and $\tau_0(s,t) = (F_0(s,t), F_{0,1}(s), F_{0,2}(t))$ with $(s,t) \in [l_1, u_1] \times [l_2, u_2]$ can be defined by

$$d(\tau_0, \tau_n) = (\|F_0 - F_n\|^2_{L_2(P_{C_1,C_2})} + \|F_{0,1} - F_{n,1}\|^2_{L_2(P_{C_1})} + \|F_{0,2} - F_{n,2}\|^2_{L_2(P_{C_2})})^{1/2}. \tag{4.3}$$

In this chapter, Empirical measure $\mathbb{P}_n$ of a sample of random elements $x_1, \ldots, x_n$ is defined by

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(x_i). \tag{4.4}$$

Throughout the technical proofs of Chapter 4, K is denoted as a universal positive constant that may be different from place to place.

4.2 Consistency

Theorem 4.1. *Suppose (C2), (C3) and (C4) hold, and $p_n = q_n = n^v$ for $v < 1$, that is, the numbers of subintervals made from knot sequences $\{u_i\}_1^{p_n+l}$ and $\{v_j\}_1^{q_n+l}$, respectively, are both equal to $O(n^v)$ for $v < 1$. Then*

$$d(\hat{\tau_n}, \tau_0) \to_p 0, \text{ as } n \to \infty.$$

Proof. We will show $\hat{\tau_n}$ is consistent estimator by verifying the conditions of Lemma 2.10.

Before verifying the three conditions of Lemma 2.10, we define Ω which contains both τ_0 and Ω_n' as follows.

$$\Omega = \{\tau = (F, F_1, F_2) : \tau \text{ satisfies the following properties (a) and (b) }\}.$$

(a) $F(s,t)$ is nondecreasing in both s and t directions, $F_1(s) - F(s,t)$ is nondecreasing in s and nonincreasing in t direction, $F_2(t) - F(s,t)$ is nondecreasing in t and nonincreasing in s direction, and $1 - F_1(s) - F_2(t) + F(s,t)$ is nonincreasing in both s and t directions.

(b) $F(s,t) \geq b_1$, $F_1(s) - F(s,t) \geq b_2$, $F_2(t) - F(s,t) \geq b_3$, and $1 - F_1(s) - F_2(t) + F(s,t) \geq b_4$, with $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and $b_4 > 0$ small enough, such that

$\tau_0 = (F_0, F_{0,1}, F_{0,2}) \in \Omega$ and $\Omega_n^{'} \subseteq \Omega$.

Remark 4.1. *For the true distribution functions $\tau_0 = (F_0, F_{0,1}, F_{0,2})$, under (C2), Lemma 4.1 given in section 4.3 guarantee that $\tau_0 \in \Omega$ and that for any $\tau_n \in \Omega_n^{'}$, $\tau_n \in \Omega$.*

The class of functions made by the log of density for single observation (s, t) is defined as

$$\mathcal{L} = \{l(\tau) : \tau \in \Omega\},$$

where

$$\begin{aligned} l(\tau)|_{(s,t)} =& \delta_1\delta_2 \log F(s,t) + \delta_1(1-\delta_2)\log[F_1(s) - F(s,t)] \\ &+ (1-\delta_1)\delta_2 \log[F_2(t) - F(s,t)] \\ &+ (1-\delta_1)(1-\delta_2)\log[1 - F_1(s) - F_2(t) + F(s,t)], \end{aligned} \tag{4.5}$$

with $\delta_1 = 1_{[T_1 \leq s]}, \delta_2 = 1_{[T_2 \leq t]}$.

For the rest of Chapter 4, let

$$\mathbb{M}(\tau) = Pl(\tau) \text{ and } \mathbb{M}_n(\tau) = \mathbb{P}_n(l(\tau)), \tag{4.6}$$

where $Pl(\tau)$ and $\mathbb{P}_n(l(\tau))$ are given according to (4.2), (4.4) and (4.5).

(i) First, we verify the condition:

$$\sup_{\tau \in \Omega} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau)| \to_p 0.$$

It suffices to show that $\mathcal{L}$ is a P-Clivenko-Cantelli, by the fact

$$\sup_{l(\tau) \in \mathcal{L}} |(\mathbb{P}_n - P)l(\tau)| = \sup_{\tau \in \Omega} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau)| \to_p 0.$$

Let $A_1 = \{\frac{\log F(s,t)}{\log b_1} : \tau = (F, F_1, F_2) \in \Omega\}$, and $\mathcal{G}_1 = \{1_{[s_1,t_1]\times[s_2,t_2]}, s_1 \leq t_1 \leq m_1, s_2 \leq t_2 \leq m_2\}$. By Property (a) and (b), we know $0 \leq \frac{\log F(s,t)}{\log b_1} \leq 1$ and $\frac{\log F(s,t)}{\log b_1}$ is monotone nonincreasing in both s and t directions. Therefore $A_1 \subseteq \overline{sconv}(\mathcal{G}_1)$. Then Lemma 2.11 implies

$$N(\epsilon, \mathcal{G}_1, L_2(Q_{C_1,C_2})) \leq K(\frac{1}{\epsilon})^4, \tag{4.7}$$

for any probability measure Q_{C_1,C_2}, by the fact that $V(\mathcal{G}_1) = 3$ and the envelop function of $\mathcal{G}_1$ is 1. Furthermore (4.7) is followed by

$$\log N(\epsilon, \overline{sconv}(\mathcal{G}_1), L_2(Q_{C_1,C_2}) \leq K(\frac{1}{\epsilon})^{4/3},$$

using the result of Lemma 2.12. Hence

$$\log N(\epsilon, A_1, L_2(Q_{C_1,C_2}) \leq K(\frac{1}{\epsilon})^{4/3}. \tag{4.8}$$

Let

$$A_1^{'} = \{\delta_1\delta_2 \log F(s,t) : \tau = (F, F_1, F_2) \in \Omega_n^{'}\}.$$

suppose the centers of ϵ-balls of A_1 are $f_i, i = 1, 2, \ldots, [K(\frac{1}{\epsilon})^{4/3}]$, then for any joint probability measure Q for (T_1, T_2, C_1, C_2)

$$\begin{aligned}
&\|\delta_1\delta_2 \log F - \delta_1\delta_2 \log b_1 f_i\|^2_{L_2(Q)} \\
&= Q[\delta_1\delta_2 \log b_1(\frac{\log F}{\log b_1} - f_i)]^2 \\
&= E[1_{[T_1<C_1,T_2<C_2]} \log b_1(\frac{\log F(C_1,C_2)}{\log b_1} - f_i(C_1,C_2))]^2 \\
&= E\{E\{[1_{[T_1<C_1,T_2<C_2]} \log b_1(\frac{\log F(C_1,C_2)}{\log b_1} - f_i(C_1,C_2)]^2|C_1,C_2\}\} \\
&= E[F_0(C_1,C_2) \log b_1(\frac{\log F(C_1,C_2)}{\log b_1} - f_i(C_1,C_2))]^2
\end{aligned}$$

$$= E_{C_1,C_2}[F_0(C_1,C_2)\log b_1(\frac{\log F(C_1,C_2)}{\log b_1} - f_i(C_1,C_2))]^2$$
$$\leq E_{C_1,C_2}[\log b_1(\frac{\log F(C_1,C_2)}{\log b_1} - f_i(C_1,C_2))]^2$$
$$= Q_{C_1,C_2}[\log b_1(\frac{\log F}{\log b_1} - f_i)]^2$$
$$= \log^2 b_1 \|\frac{\log F}{\log b_1} - f_i\|^2_{L_2(Q_{C_1,C_2})},$$

let $\hat{b_1} = -\log b_1$ then $\delta_1\delta_2 \log b_1 f_i, i = 1, 2, \ldots, [K(\frac{1}{\epsilon})^{4/3}]$ are the centers of $\epsilon\hat{b_1}$-balls of $A_1^{'}$ and by (4.8)

$$\log N(\epsilon\hat{b_1}, A_1^{'}, L_2(Q)) \leq K(\frac{1}{\epsilon})^{4/3}, \tag{4.9}$$

and it follows that

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon\hat{b_1}, A_1^{'}, L_2(Q))} d\epsilon \leq \int_0^1 \sqrt{K}(\frac{1}{\epsilon})^{2/3} d\epsilon < \infty. \tag{4.10}$$

It is obvious that the envelop function of $A_1^{'}$ is $\hat{b_1}$, therefore $A_1^{'}$ is a P-Donsker, by Lemma 2.13.

Let $A_2 = \{\frac{\log(F_1(s)-F(s,t))}{\log b_2} : \tau = (F, F_1, F_2) \in \Omega\}$, and $\mathcal{G}_2 = \{1_{[s_1,t_1]\times[t_2,m_2]}, s_1 \leq t_1 \leq m_1, s_2 \leq t_2 \leq m_2\}$, and $A_2^{'} = \{\delta_1(1-\delta_2)\log(F_1(s) - F(s,t)) : \tau = (F, F_1, F_2) \in \Omega_n^{'}\}$. By the similar arguments in showing $A_1^{'}$ to be a P-Donsker, it can be shown that $A_2^{'}$ is P-Donsker.

Similarly, it can be shown that $A_3^{'} = \{(1-\delta_1)\delta_2 \log(F_2(t)-F(s,t)) : \tau = (F, F_1, F_2) \in \Omega\}$ and $A_4^{'} = \{(1-\delta_1)(1-\delta_2)\log(1-F_1(s)-F_2(t)-F(s,t)) : \tau = (F, F_1, F_2) \in \Omega\}$ are P-Donsker classes as well.

By Lemma 2.14, pairwise sums $A_1^{'} + A_2^{'}$ and $A_3^{'} + A_4^{'}$ are both P-Donsker classes. Then the pairwise sum $(A_1^{'}+A_2^{'})+(A_3^{'}+A_4^{'})$ is also P-Donsker. It is obvious that $\mathcal{L} \subset$

$(A_1'+A_2')+(A_3'+A_4')$, hence $\mathcal{L}$ is P-Donsker. By Slusky's Theorem, it can be showed that a P-Donsker is a P-Clivenko-Cantelli and hence $\sup_{l(\tau)\in\mathcal{L}} |(\mathbb{P}_n - P)l(\tau)| \to_p 0$.

(ii) Second, we verify

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq Kd(\tau_0, \tau)^2,$$

for any $\tau \in \Omega$.

By (4.6) and $E(E(X|Y)) = E(X)$,

$$\begin{aligned}
\mathbb{M}(\tau_0) - \mathbb{M}(\tau) =& P\{l(\tau_0) - l(\tau)\} \\
=& P\{\delta_1\delta_2 \log \frac{F_0}{F} + \delta_1(1-\delta_2) \log \frac{F_{0,1} - F_0}{F_1 - F} \\
&+ (1-\delta_1)\delta_2 \log \frac{F_{0,2} - F_0}{F_2 - F} \\
&+ (1-\delta_1)(1-\delta_2) \log \frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\}, \\
=& P_{C_1,C_2}\{F_0 \log \frac{F_0}{F} + (F_{0,1} - F_0) \log \frac{F_{0,1} - F_0}{F_1 - F} \\
&+ (F_{0,2} - F_0) \log \frac{F_{0,2} - F_0}{F_2 - F} \\
&+ (1 - F_{0,1} - F_{0,2} + F_0) \log \frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{M}(\tau_0) - \mathbb{M}(\tau) =& P_{C_1,C_2}\{Fm(\frac{F_0}{F}) + (F_1 - F)m(\frac{F_{0,1} - F_0}{F_1 - F}) \\
&+ (F_2 - F)m(\frac{F_{0,2} - F_0}{F_2 - F}) \\
&+ (1 - F_1 - F_2 + F)m(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F})\},
\end{aligned} \tag{4.11}$$

where $m(x) = x\log(x) - x + 1 \geq (x-1)^2/4$ for $0 \leq x \leq 5$.

By F has positive upper bound,

$$\begin{aligned} P_{C_1,C_2}\{Fm(\frac{F_0}{F})\} \geq & P_{C_1,C_2}\{\frac{F^2}{K}(\frac{F_0}{F}-1)^2/4\} \\ = & KP_{C_1,C_2}(F_0-F)^2 \\ = & K\|F_0-F\|^2_{L_2(P_{C_1,C_2})}. \end{aligned} \tag{4.12}$$

Also, since $F_1 - F$ has positive upper bound,

$$\begin{aligned} & P_{C_1,C_2}\{(F_1-F)m(\frac{F_{0,1}-F_0}{F_1-F})\} \\ \geq & P_{C_1,C_2}\{\frac{(F_1-F)^2}{K}(\frac{F_{0,1}-F_0}{F_1-F}-1)^2/4\} \\ \geq & K\|(F_{0,1}-F_0)-(F_1-F)\|^2_{L_2(P_{C_1,C_2})} \\ = & K\|(F_{0,1}-F_1)-(F_0-F)\|^2_{L_2(P_{C_1,C_2})}. \end{aligned} \tag{4.13}$$

Similarly,

$$\begin{aligned} & P_{C_1,C_2}\{(F_2-F)m(\frac{F_{0,2}-F_0}{F_2-F})\} \\ \geq & K\|(F_{0,2}-F_0)-(F_2-F)\|^2_{L_2(P_{C_1,C_2})} \\ = & K\|(F_{0,2}-F_2)-(F_0-F)\|^2_{L_2(P_{C_1,C_2})}, \end{aligned} \tag{4.14}$$

and

$$\begin{aligned} & P_{C_1,C_2}\{(1-F_1-F_2+F)m(\frac{1-F_{0,1}-F_{0,2}+F_0}{1-F_1-F_2+F})\} \\ \geq & K\|(1-F_{0,1}-F_{0,2}+F_0)-(1-F_1-F_2+F)\|^2_{L_2(P_{C_1,C_2})}. \end{aligned} \tag{4.15}$$

By (4.12), (4.13), (4.14) and (4.15), (4.11) results in

$$\begin{aligned} \mathbb{M}(\tau_0)-\mathbb{M}(\tau) \geq & K(\|F_0-F\|^2_{L_2(P_{C_1,C_2})} + \|(F_{0,1}-F_1)-(F_0-F)\|^2_{L_2(P_{C_1,C_2})} \\ & + \|(F_{0,2}-F_2)-(F_0-F)\|^2_{L_2(P_{C_1,C_2})}) \end{aligned}$$

Let $f_1 = \|F_0-F\|^2_{L_2(P_{C_1,C_2})}$, $f_2 = \|F_{0,1}-F_1\|^2_{L_2(P_{C_1})}$, and $f_3 = \|F_{0,2}-F_2\|^2_{L_2(P_{C_2})}$.

If f_1 is the largest among f_1, f_2, f_3, then

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K f_1 \geq (K/3)(f_1 + f_2 + f_3), \tag{4.16}$$

if f_2 is the largest, then

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K[f_1 + (f_2 - f_1)] \geq K f_2 \geq (K/3)(f_1 + f_2 + f_3), \tag{4.17}$$

if f_3 is the largest, then

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K[f_1 + (f_3 - f_1)] \geq K f_3 \geq (K/3)(f_1 + f_2 + f_3). \tag{4.18}$$

Hence

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K d(\tau_0, \tau)^2,$$

by (4.16), (4.17) and (4.18).

(iii) Finally, we verify

$$\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \geq -o_p(1).$$

Since $\hat{\tau}_n$ maximizes $\mathbb{M}_n(\tau_n)$ in $\Omega_n^{'}$, $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) > 0$. Hence,

$$\begin{aligned} \mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) =& \mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) + \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) \\ \geq& \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) \\ =& \mathbb{P}_n(l(\tau_n)) - \mathbb{P}_n(l(\tau_0)) \\ =& (\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\} + P\{l(\tau_n) - l(\tau_0)\}. \end{aligned} \tag{4.19}$$

Under (C3) and suppose p_n and q_n are both equal to $O(n^v)$, then Lemma 4.2 guarantees that there exists $\tau_n = (F_n, F_{n,1}, F_{n,2})$ in $\Omega_n^{'}$ such that for $\tau_0 = (F_0, F_{0,1}, F_{0,2})$, $\|F_n - F_0\|_\infty \leq K(n^{-pv})$, $\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv})$, and $\|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv})$.

Define

$$\mathcal{L}_n =\{l(\tau_n) : \tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}, \|F_n - F_0\|_\infty \leq K(n^{-pv}),$$
$$\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv}), \|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv})\}$$

Since $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$, then in $\mathcal{L}_n$

$$\begin{aligned} P\{l(\tau_n) - l(\tau_0)\}^2 \leq & 4P(\delta_1\delta_2 \log \frac{F_n}{F_0})^2 + 4P(\delta_1(1-\delta_2) \log \frac{F_{n,1} - F_n}{F_{0,1} - F_0})^2 \\ & + 4P((1-\delta_1)\delta_2 \log \frac{F_{n,2} - F_n}{F_{0,2} - F_0})^2 \\ & + 4P((1-\delta_1)(1-\delta_2) \log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0})^2 \\ \leq & 4P_{C_1,C_2}(\log \frac{F_n}{F_0})^2 + 4P_{C_1,C_2}(\log \frac{F_{n,1} - F_n}{F_{0,1} - F_0})^2 \\ & + 4P_{C_1,C_2}(\log \frac{F_{n,2} - F_n}{F_{0,2} - F_0})^2 \\ & + 4P_{C_1,C_2}(\log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0})^2 \end{aligned} \tag{4.20}$$

The fact that $\|F_n - F_0\|_{L_2(P_{C_1,C_2})} \leq K(n^{-pv})$ and that F_0 has a positive lower bound results in $1/2 < \frac{F_n}{F_0} < 2$ for large n.

Since it can be easily showed that if $1/2 \leq x \leq 2$, $|\log(x)| \leq K|x - 1|$. It follows that $|\log \frac{F_n}{F_0}| \leq K|\frac{F_n}{F_0} - 1|$, then

$$\begin{aligned} P_{C_1,C_2}|\log \frac{F_n}{F_0}|^2 \leq & KP_{C_1,C_2}|\frac{F_n}{F_0} - 1|^2 \\ \leq & KP_{C_1,C_2}(\frac{F_0^2}{K}|\frac{F_n}{F_0} - 1|^2) \\ = & KP_{C_1,C_2}|F_n - F_0|^2 \\ \leq & K(n^{-pv})^2 \rightarrow 0. \end{aligned} \tag{4.21}$$

Similarly, since $\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv})$, then $F_{n,1} - F_n$ and $F_{0,1} - F_0$ are very close at every point in their domain. By the fact that $F_{0,1} - F_0$ has a positive lower bound, $1/2 < \frac{F_{n,1}-F_n}{F_{0,1}-F_0} < 2$ and $|\log \frac{F_{n,1}-F_n}{F_{0,1}-F_0}| \leq K|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|$. Therefore

$$\begin{aligned} P_{C_1,C_2}|\log \frac{F_{n,1}-F_n}{F_{0,1}-F_0}|^2 \leq & KP_{C_1,C_2}|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|^2 \\ \leq & KP_{C_1,C_2}((F_{0,1}-F_0)^2|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|^2) \\ \leq & KP_{C_1,C_2}|(F_{n,1}-F_n) - (F_{0,1}-F_0)|^2 \\ \leq & K(n^{-pv})^2 \to 0. \end{aligned} \tag{4.22}$$

Similarly,

$$P_{C_1,C_2}|\log \frac{F_{n,2}-F_n}{F_{0,2}-F_0}|^2 \to 0, \tag{4.23}$$

and

$$P_{C_1,C_2}|\log \frac{1-F_{n,1}-F_{n,2}+F_n}{1-F_{0,1}-F_{0,2}+F_0}|^2 \to 0. \tag{4.24}$$

By (4.20), (4.21), (4.22), (4.23) and (4.24),

$$P|l(\tau_n) - l(\tau_0)|^2 \to 0, \text{ as } n \to \infty. \tag{4.25}$$

By the similar arguments,

$$P|l(\tau_n) - l(\tau_0)| \to 0, \text{ as } n \to \infty. \tag{4.26}$$

Hence, by (4.25) and (4.26),

$$\rho_P\{l(\tau_n) - l(\tau_0)\} = \{P\{[l(\tau_n) - l(\tau_0)] - P[l(\tau_n) - l(\tau_0)]\}^2\}^{1/2} \to 0, \text{ as } n \to \infty.$$

Since under (C2) Lemma 4.1 indicates $\Omega_n' \subset \Omega$ and $\tau_0 \in \Omega$, then $\mathcal{L}_n \subset \mathcal{L}$ and $l(\tau_0) \in \mathcal{L}$. So both $l(\tau_n)$ and $l(\tau_0)$ are in $\mathcal{L}$. In addition, $\mathcal{L}$ is a P-Donsker.

Hence

$$(\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\} = o_p(n^{-1/2}), \tag{4.27}$$

by Lemma 2.15.

And

$$P\{l(\tau_n) - l(\tau_0)\} \le P|l(\tau_n) - l(\tau_0)| \to 0, \text{ as } n \to \infty.$$

So $P(l(\tau_n) - l(\tau_0)) \ge -o(1)$ as $n \to \infty$. Then by (4.19),

$$\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \ge o_p(n^{-1/2}) - o(1) \ge -o_p(1).$$

This completes the proof of $d(\hat{\tau}_n, \tau_0) \to 0$ in probability. □

4.3 Convergence Rate

Theorem 4.2. *Suppose (C1)-(C5) hold, and $p_n = q_n = n^v$ for $v \le \frac{1}{4p}$, that is, the numbers of subintervals made from knot sequences $\{u_i\}_1^{p_n+l}$ and $\{v_j\}_1^{q_n+l}$, respectively, are both equal to $O(n^v)$ for $v \le \frac{1}{4p}$. Then*

$$d(\hat{\tau}_n, \tau_0) = O_p(n^{-\min\{pv,(1-2v)/3\}}).$$

Proof. By (4.19), $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \ge I_{1,n} + I_{2,n}$, where $I_{1,n} = (\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\}$ and $I_{2,n} = \mathbb{M}(\tau_n) - \mathbb{M}(\tau_0)$ for any $l(\tau_n) \in \mathcal{L}_n$. Given by (4.27), $I_{1,n} = o_p(n^{-1/2})$. Then if $v \le \frac{1}{4p}$

$$I_{1,n} = o_p(n^{-2pv}). \tag{4.28}$$

In what follows, it is showed that

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) \le O(n^{-2pv}). \tag{4.29}$$

Let $\tau = \tau_n$ in (4.11), then

$$\begin{aligned}\mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) =& P_{C_1,C_2}\{F_n m(\frac{F_0}{F_n}) + (F_{n,1} - F_n)m(\frac{F_{0,1} - F_0}{F_{n,1} - F_n}) \\ &+ (F_{n,2} - F_n)m(\frac{F_{0,2} - F_0}{F_{n,2} - F_n}) \\ &+ (1 - F_{n,1} - F_{n,2} + F_n)m(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_{n,1} - F_{n,2} + F_n})\}.\end{aligned} \tag{4.30}$$

By the fact that $m(x) = x\log -x + 1 \le (x-1)^2$ in the neighborhood of $x = 1$ and the definition of $\mathcal{L}_n$,

$$\begin{aligned}P_{C_1,C_2}\{F_n m(\frac{F_0}{F_n})\} \le& K P_{C_1,C_2}\{F_n^2(\frac{F_0}{F_n} - 1)^2\} \\ =& K P_{C_1,C_2}(F_0 - F_n)^2 \\ \le& K\|F_0 - F_n\|_\infty^2 \\ =& O(n^{-2pv}).\end{aligned} \tag{4.31}$$

similarly,

$$\begin{aligned}P_{C_1,C_2}\{(F_{n,1} - F_n)m(\frac{F_{0,1} - F_0}{F_{n,1} - F_n})\} \le& K\|F_0 - F_n\|_\infty^2 + K\|F_{0,1} - F_{n,1}\|_\infty^2 \\ =& O(n^{-2pv}),\end{aligned} \tag{4.32}$$

$$\begin{aligned}P_{C_1,C_2}\{(F_{n,2} - F_n)m(\frac{F_{0,2} - F_0}{F_{n,2} - F_n})\} \le& K\|F_0 - F_n\|_\infty^2 + K\|F_{0,2} - F_{n,2}\|_\infty^2 \\ =& O(n^{-2pv}),\end{aligned} \tag{4.33}$$

and

$$P_{C_1,C_2}\{(1-F_{n,1}-F_{n,2}+F_n)m(\frac{1-F_{0,1}-F_{0,2}+F_0}{1-F_{n,1}-F_{n,2}+F_n})\} \leq K\|F_0-F_n\|_\infty^2 + K\|F_{0,2}-F_{n,2}\|_\infty^2 + K\|F_{0,1}-F_{n,1}\|_\infty^2 = O(n^{-2pv}). \tag{4.34}$$

So (4.30), (4.31), (4.32), (4.33) and (4.34) result in $\mathbb{M}(\tau_0)-\mathbb{M}(\tau_n) \leq O(n^{-2pv})$. Then

$$I_{2,n} = \mathbb{M}(\tau_n)-\mathbb{M}(\tau_0) \geq -O(n^{-2pv}). \tag{4.35}$$

Hence

$$\mathbb{M}(\hat{\tau}_n)-\mathbb{M}(\tau_0) \geq -O(n^{-2pv}) + o_p(n^{-2pv}) = -O_p(n^{-2pv}),$$

by (4.28) and (4.35).

Let

$$\mathcal{L}_{n,\eta} = \{l(\tau_n)-l(\tau_0) : \tau_n \in \Omega_n^{'} \text{ and } d(\tau_n,\tau_0) \leq \eta\}.$$

In what follows, the bracket number of $\mathcal{L}_{n,\eta}$ is evaluated.

Let $\mathcal{L}_{n,1} = \{l(\tau_n) : \tau_n \in \Omega_n^{'}\}$, $\mathcal{F}_n = \{F_n : \tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}\}$, $\mathcal{F}_{n,1} = \{F_{n,1} : \tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}\}$, and $\mathcal{F}_{n,2} = \{F_{n,2} : \tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}\}$.

By Lemma 2.22, there exist ϵ-brackets $[F_i^L, F_i^U], i = 1,2,\ldots,[(1/\epsilon)^{Kp_nq_n}]$ to cover $\mathcal{F}_n$. By Lemma 2.21, there exist ϵ-brackets $[F_j^{(1),L}, F_j^{(1),U}], j = 1,2,\ldots,[(1/\epsilon)^{Kp_n}]$ to cover $\mathcal{F}_{n,1}$, and there exist ϵ-brackets $[F_k^{(2),L}, F_k^{(2),U}], k = 1,2,\ldots,[(1/\epsilon)^{Kq_n}]$ to cover $\mathcal{F}_{n,2}$.

Let

$$l^U_{i,j,k} = \delta_1\delta_2 \log F^U_i + \delta_1(1-\delta_2)\log(F^{(1),U}_j - F^L_i)$$
$$+ (1-\delta_1)\delta_2\log(F^{(2),U}_k - F^L_i)$$
$$+ (1-\delta_1)(1-\delta_2)\log(1 - F^{(1),L}_j - F^{(2),L}_k + F^U_i),$$

and

$$l^L_{i,j,k} = \delta_1\delta_2 \log F^L_i + \delta_1(1-\delta_2)\log(F^{(1),L}_j - F^U_i)$$
$$+ (1-\delta_1)\delta_2\log(F^{(2),L}_k - F^U_i)$$
$$+ (1-\delta_1)(1-\delta_2)\log(1 - F^{(1),U}_j - F^{(2),U}_k + F^L_i).$$

Then for any $l(\tau_n) \in \mathcal{L}_{n,1}$,there exist i, j, k, for $i = 1, 2, \ldots, [(1/\epsilon)^{Kp_nq_n}]$, $j = 1, 2, \ldots, [(1/\epsilon)^{Kp_n}]$ and $k = 1, 2, \ldots, [(1/\epsilon)^{Kq_n}]$, such that $l^L_{i,j,k} \le l(\tau_n) \le l^U_{i,j,k}$ and the number of brackets $[l^L_{i,j,k}, l^U_{i,j,k}]'s$ is bounded by $(1/\epsilon)^{Kp_nq_n} \cdot (1/\epsilon)^{Kp_n} \cdot (1/\epsilon)^{Kq_n}$.

It is left to show that any bracket $[l^L_{i,j,k}, l^U_{i,j,k}]$ is a $K\epsilon$-bracket.

$$\begin{aligned}
&\| l^U_{i,j,k} - l^L_{i,j,k} \|_\infty \\
\le& \| \delta_1\delta_2 \log \frac{F^U_i}{F^L_i} \|_\infty + \| \delta_1(1-\delta_2)\log \frac{F^{(1),U}_j - F^L_i}{F^{(1),L}_j - F^U_i} \|_\infty \\
&+ \| (1-\delta_1)\delta_2 \log \frac{F^{(2),U}_k - F^L_i}{F^{(2),L}_k - F^U_i} \|_\infty \\
&+ \| (1-\delta_1)(1-\delta_2)\log \frac{1 - F^{(1),L}_j - F^{(2),L}_j + F^U_i}{1 - F^{(1),U}_j - F^{(2),U}_j + F^L_i} \|_\infty \\
\le& \| \log \frac{F^U_i}{F^L_i} \|_\infty + \| \log \frac{F^{(1),U}_j - F^L_i}{F^{(1),L}_j - F^U_i} \|_\infty + \| \log \frac{F^{(2),U}_k - F^L_i}{F^{(2),L}_k - F^U_i} \|_\infty
\end{aligned}$$

$$+ \parallel \log \frac{1 - F_j^{(1),L} - F_j^{(2),L} + F_i^U}{1 - F_j^{(1),U} - F_j^{(2),U} + F_i^L} \parallel_\infty$$

Since F_n has positive lower bound, then for small ϵ, F_i^L has positive lower bound. That F_i^L has positive lower bound and $F_i^U(s,t)$ is close to $F_i^L(s,t)$ guarantee that $0 \leq \frac{F_i^U}{F_i^L} - 1 \leq 1$ for $i = 1, 2, \ldots, [(1/\epsilon)^{Kp_nq_n}]$. Then by $\log x \leq (x-1)$ for $0 \leq (x-1) \leq 1$, $\log \frac{F_i^U}{F_i^L} \leq \frac{F_i^U}{F_i^L} - 1$.

So

$$\parallel \log \frac{F_i^U}{F_i^L} \parallel_\infty \leq \parallel \frac{F_i^U}{F_i^L} - 1 \parallel_\infty \leq \parallel \frac{1}{F_i^L}(F_i^U - F_i^L) \parallel_\infty \leq K \parallel F_i^U - F_i^L \parallel_\infty \leq K\epsilon. \quad (4.36)$$

Since $F_{n,1} - F_n$ has positive lower bound, then for small ϵ, $F_j^{(1),L} - F_i^U$ has positive lower bound. $F_j^{(1),U}$ is close to $F_j^{(1),L}$ and F_i^U is close to F_i^L result in $F_j^{(1),U} - F_i^L$ is close to $F_j^{(1),L} - F_i^U$. That $F_j^{(1),L} - F_i^U$ has positive lower bound and $F_j^{(1),U} - F_i^L$ is close to $F_j^{(1),L} - F_i^U$, result in $0 \leq \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} - 1 \leq 1$ for $i = 1, 2, \ldots, [(1/\epsilon)^{Kp_nq_n}]$ and $j = 1, 2, \ldots, [(1/\epsilon)^{Kp_n}]$.

So

$$\begin{aligned} \parallel \log \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} \parallel_\infty \leq & \parallel \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} - 1 \parallel_\infty \\ & \leq K \parallel (F_j^{(1),U} - F_i^L) - (F_j^{(1),L} - F_i^U) \parallel_\infty \\ & \leq K(\parallel F_j^{(1),U} - F_j^{(1),L} \parallel_\infty + \parallel F_i^U - F_i^L \parallel_\infty) \\ & \leq K\epsilon. \end{aligned} \quad (4.37)$$

It can be similarly shown that

$$\parallel \log \frac{F_k^{(2),U} - F_i^L}{F_k^{(2),L} - F_i^U} \parallel_\infty \leq K\epsilon, \quad (4.38)$$

and

$$\| \log \frac{1 - F_j^{(1),L} - F_j^{(2),L} + F_i^U}{1 - F_j^{(1),U} - F_j^{(2),U} + F_i^L} \|_\infty \leq K\epsilon. \tag{4.39}$$

Then $\| l_{i,j,k}^U - l_{i,j,k}^L \|_\infty \leq K\epsilon$, by (4.36), (4.37), (4.38) and (4.39). Hence it follows that

$$N_{[\,]}\{\epsilon, \mathcal{L}_{n,1}, \| \, \|_\infty\} \leq (1/\epsilon)^{Kp_nq_n + Kp_n + Kq_n} \leq (1/\epsilon)^{Kp_nq_n}.$$

So $N_{[\,]}\{\epsilon, \mathcal{L}_{n,1}, L_2(P)\} \leq (1/\epsilon)^{Kp_nq_n}$, by the fact that L_2-norm is bounded by L_∞-norm. Hence

$$N_{[\,]}\{\epsilon, \mathcal{L}_{n,\eta}, L_2(P)\} \leq (1/\epsilon)^{Kp_nq_n}, \tag{4.40}$$

by $(\mathcal{L}_{n,\eta} + l(\tau_0)) \subset \mathcal{L}_{n,1}$.

In Theorem 1, it has been showed that if $\|\tau_n - \tau_0\|_{L_2(P)} \leq Kn^{-pv}$ then $P\{L(\tau_n) - L(\tau_0)\}^2 \leq K(n^{-pv})^2$. In what follows, we show that $P\{L(\tau_n) - L(\tau_0)\}^2 \leq K\eta^2$ for any $L(\tau_n) - L(\tau_0) \in \mathcal{L}_{n,\eta}$ by similar arguments.

Since $\|F_n - F_0\|_{L_2(P_{C_1,C_2})} \leq d(F_n, F_0) \leq \eta$, then under (C1) and (C5), Lemma 4.3 indicates that for large n, F_n and F_0 are very close at every point in $[l_1, u_1] \times [l_2, u_2]$. Then the fact that F_0 has a positive lower bound results in $1/2 < \frac{F_n}{F_0} < 2$.

Since it can be easily showed that if $1/2 \leq x \leq 2$, $|\log(x)| \leq K|x - 1|$. It follows that $|\log \frac{F_n}{F_0}| \leq K|\frac{F_n}{F_0} - 1|$, then

$$\begin{aligned} P_{C_1,C_2}|\log \frac{F_n}{F_0}|^2 \leq & KP_{C_1,C_2}|\frac{F_n}{F_0} - 1|^2 \\ \leq & KP_{C_1,C_2}(\frac{F_0^2}{K}|\frac{F_n}{F_0} - 1|^2) \\ = & KP_{C_1,C_2}|F_n - F_0|^2 \\ \leq & K\eta^2. \end{aligned} \tag{4.41}$$

Similarly, since $\|F_{n,1} - F_{0,1}\|_{L_2(P_{C_1})} \leq \eta$, then under (C1) and (C5), Lemma 4.4 indicates that for large n, $F_{n,1}$ and $F_{0,1}$ are very close at every point on $[l_1, u_1]$, then $F_{n,1} - F_n$ and $F_{0,1} - F_0$ are very close at every point in their domain. By the fact that $F_{0,1} - F_0$ has a positive lower bound, $1/2 < \frac{F_{n,1}-F_n}{F_{0,1}-F_0} < 2$ and $|\log \frac{F_{n,1}-F_n}{F_{0,1}-F_0}| \leq K|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|$. Therefore

$$\begin{aligned} P_{C_1,C_2}|\log \frac{F_{n,1}-F_n}{F_{0,1}-F_0}|^2 \leq & KP_{C_1,C_2}|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|^2 \\ \leq & KP_{C_1,C_2}((F_{0,1}-F_0)^2|\frac{F_{n,1}-F_n}{F_{0,1}-F_0} - 1|^2) \\ \leq & KP_{C_1,C_2}|(F_{n,1}-F_n) - (F_{0,1}-F_0)|^2 \\ \leq & K\eta^2. \end{aligned} \tag{4.42}$$

Similarly,

$$P_{C_1,C_2}|\log \frac{F_{n,2}-F_n}{F_{0,2}-F_0}|^2 \leq K\eta^2, \tag{4.43}$$

and

$$P_{C_1,C_2}|\log \frac{1-F_{n,1}-F_{n,2}+F_n}{1-F_{0,1}-F_{0,2}+F_0}|^2 \leq K\eta^2. \tag{4.44}$$

By (4.20), (4.41), (4.42), (4.43) and (4.44),

$$P\{l(\tau_n) - l(\tau_0)\}^2 \leq K\eta^2. \tag{4.45}$$

Also it is obvious that $\mathcal{L}_{n,\eta}$ is uniformly bounded. Hence by Lemma 2.16,

$$E_P^* \parallel \mathbb{G}_n \parallel_{\mathcal{L}_{n,\eta}} \leq K\tilde{J}_{[\,]}\{\eta, \mathcal{L}_{n,\eta}, L_2(P)\}[1 + \frac{\tilde{J}_{[\,]}\{\eta, \mathcal{L}_{n,\eta}, L_2(P)\}}{\eta^2\sqrt{n}}].$$

By (4.40),

$$\begin{aligned}\tilde{J}_{[\,]}\{\eta, \mathcal{L}_{n,\eta}, L_2(P)\} &= \int_0^\eta \sqrt{1 + \log N_{[\,]}\{\epsilon, \mathcal{L}_{n,\eta}, L_2(P)\}} d\epsilon \\ &= \int_0^\eta \sqrt{1 + K p_n q_n \log(1/\epsilon)} d\epsilon \\ &\leq \int_0^\eta \{K(p_n q_n)^{1/2}(1/\epsilon)^{1/2}\} d\epsilon \\ &= K(p_n q_n)^{1/2}\eta^{1/2}\end{aligned}$$

Let $\psi(\eta) = (p_n q_n)^{1/2}\eta^{1/2} + (p_n q_n)/(\eta n^{1/2})$, then it is easy to see that $\psi(\eta)/\eta$ is decreasing function of η. Note that for $p_n = q_n = n^v$,

$$\begin{aligned}n^{2pv}\psi(1/n^{pv}) &= n^{2pv}n^v n^{(-1pv)/2} + n^{2pv}n^{2v}n^{-1/2}n^{1pv} \\ &= n^{1/2}\{n^{(3pv)/2-(1-2v)/2} + n^{3pv-(1-2v)}\}.\end{aligned}$$

Therefore, if $pv \leq (1-2v)/3$, then $n^{2pv}\psi(1/n^{pv}) \leq n^{1/2}$. Also

$$\begin{aligned}n^{2(1-2v)/3}\psi(1/n^{(1-2v)/3}) &= n^{2(1-2v)/3}n^v n^{-(1-2v)/6} + n^{2(1-2v)/3}n^{2v}n^{-1/2}n^{(1-2v)/3} \\ &= 2n^{1/2}.\end{aligned}$$

This implies if $r_n = n^{\min\{pv,(1-2v)/3\}}$, then

$$r_n^2\psi(1/r_n) \leq Kn^{1/2},$$

and

$$\mathbb{M}(\hat{\tau}_n) - \mathbb{M}(\tau_0) \geq -O_p(n^{-2pv}) \geq -O_p(n^{-2\min\{pv,(1-2v)/3\}}) = -O_p(r_n^{-2}).$$

Hence it follows by Lemma 2.17 that

$$r_n d(\hat{\tau}_n, \tau_0) = O_p(1).$$

□

4.4 Technical Lemmas

Lemma 4.1. *Suppose $\tau = \tau_0$ or $\tau \in \Omega_n^{'}$, then under (C2), the following two properties hold for $F(s,t)$, $F_1(s)$ and $F_2(t)$ with $\tau(s,t) = (F(s,t), F_1(s), F_2(t))$.*

(1) $F(s,t)$ is nondecreasing in both s and t directions. $F_1(s) - F(s,t)$ is nondecreasing in s direction and nonincreasing in t direction. $F_2(t) - F(s,t)$ is nondecreasing in t direction and nonincreasing in s direction. $1 - F_1(s) - F_2(t) + F(s,t)$ is nonincreasing in both s and t directions.

(2) $F(s,t)$, $F_1(s) - F(s,t)$, $F_2(t) - F(s,t)$ and $1 - F_1(s) - F_2(t) + F(s,t)$ all have positive lower bounds.

Proof. (i) First, we verify the two properties for $\tau = \tau_0$.

Since $(F, F_1, F_2) = (F_0, F_{0,1}, F_{0,2})$ is a vector-valued function, corresponding to the joint distribution function of (T_1, T_2) and their two marginal distribution functions, then by Lemma 3.1

$$F(s^{'}, t) \leq F(s^{''}, t), \tag{4.46}$$

$$F(s, t^{'}) \leq F(s, t^{''}), \tag{4.47}$$

$$[F_1(s^{''}) - F(s^{''}, t)] - [F_1(s^{'}) - F(s^{'}, t)] \geq 0, \tag{4.48}$$

$$[F_2(t^{''}) - F(s, t^{''})] - [F_2(t^{'}) - F(s, t^{'})] \geq 0. \tag{4.49}$$

By (4.46) and (4.47), $F(s,t)$ is nondecreasing in both s and t directions. By (4.48) and (4.47), $F_1(s) - F(s,t)$ is nondecreasing in s direction, and nonincreasing in t direction. By (4.49) and (4.46), $F_2(t) - F(s,t)$ is nondecreasing in t direction, and nonincreasing in s direction. By (4.48) and (4.49), $1 - F_1(s) - F_2(t) - F(s,t)$ is nonincreasing in both s and t directions.

Under (C2) and $(s,t) \in [l_1, u_1] \times [l_2, u_2]$,

$$F(s,t) = F_0(s,t) \geq (s - L_1)(t - L_2) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \geq (l_1 - L_1)(l_2 - L_2) b_0,$$

$$\begin{aligned} F_1(s) - F(s,t) = F_{0,1}(s) - F_0(s,t) &\geq (s - L_1)(U_2 - t) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \\ &\geq (l_1 - L_1)(U_2 - u_2) b_0, \end{aligned}$$

$$\begin{aligned} F_2(t) - F(s,t) = F_{0,2}(t) - F_0(s,t) &\geq (t - L_2)(U_1 - s) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \\ &\geq (l_2 - L_2)(U_1 - u_1) b_0, \end{aligned}$$

and

$$\begin{aligned} 1 - F_1(s) - F_2(t) + F(s,t) &= 1 - F_{0,1}(s) - F_{0,2}(t) + F_0(s,t) \\ &\geq (U_1 - s)(U_2 - t) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \\ &\geq (U_1 - u_1)(U_2 - u_2) b_0. \end{aligned}$$

(ii) Second, we verify the two properties for $\tau \in \Omega_n^{'}$.

By Lemma 3.2, $\Omega_n^{'} \subset \mathcal{F}$ in $[l_1, u_1] \times [l_2, u_2]$, where $\mathcal{F}$ is defined in Section 3.2. Then for $\tau = (F, F_1, F_2) \in \Omega_n^{'}$, (4.46), (4.47), (4.48) and (4.49) hold. Hence, $\tau = (F, F_1, F_2) \in \Omega_n^{'}$ satisfy property (1).

By Property (B3) in section 2.1.1,

$$\frac{\partial^2 F(s,t)}{\partial s \partial t} = \sum_{i=1}^{p_n - 1} \sum_{j=1}^{q_n - 1} (l-1)^2 \frac{\alpha_{i+1,j+1} - \alpha_{i,j+1} - \alpha_{i+1,j} + \alpha_{i,j}}{(s_{i+l} - s_{i+1})(t_{j+l} - t_{j+1})} N_{i+1}^{(1),l-1}(s) N_{j+1}^{(2),l-1}(t).$$

Then by (Sp1) and 4th condition given in (4.1)

$$\begin{aligned}\frac{\partial^2 F(s,t)}{\partial s \partial t} \geq & \min_{i,j} \frac{\alpha_{i+1,j+1} - \alpha_{i,j+1} - \alpha_{i+1,j} + \alpha_{i,j}}{\max_{i_1} \Delta_{i_1}^{(u)} \max_{j_1} \Delta_{j_1}^{(v)}} \\ \geq & \min_{i,j} \frac{\alpha_{i+1,j+1} - \alpha_{i,j+1} - \alpha_{i+1,j} + \alpha_{i,j}}{\frac{\min_{i_1} \Delta_{i_1}^{(u)}}{l} \frac{\min_{j_1} \Delta_{j_1}^{(v)}}{l}} \frac{\frac{\min_{i_1} \Delta_{i_1}^{(u)}}{l} \frac{\min_{j_1} \Delta_{j_1}^{(v)}}{l}}{\max_{i_1} \Delta_{i_1}^{(u)} \max_{j_1} \Delta_{j_1}^{(v)}} \\ \geq & K. \end{aligned} \tag{4.50}$$

Then for $(s,t) \in [l_1, u_1] \times [l_2, u_2]$ by (4.50)

$$\begin{aligned} F(s,t) &\geq F(s,t) - F(s,L_2) - F(L_1,t) + F(L_1,L_2) \\ &\geq (s - L_1)(t - L_2) \min_{s,t} \frac{\partial^2 F(s,t)}{\partial s \partial t} \\ &\geq (s - L_1)(t - L_2) K \\ &\geq (l_1 - L_1)(l_2 - L_2) K \end{aligned}$$

By Property (B2) and $\alpha_{i,j} \leq \beta_i$, we have

$$\begin{aligned} F(s,U_2) =& \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(U_2) \\ \leq& \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \beta_i N_i^{(1),l}(s) N_j^{(2),l}(U_2) \\ =& (\sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s))(\sum_{j=1}^{q_n} N_j^{(2),l}(U_2)) \\ =& \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s) = F_1(s), \end{aligned} \tag{4.51}$$

then it follows that

$$\begin{aligned} F_1(s) - F(s,t) \geq& F(s,U_2) - F(s,t) \\ \geq& F(s,U_2) - F(s,t) - F(L_1,U_2) + F(L_1,t) \end{aligned}$$

$$
\begin{aligned}
&\geq (s - L_1)(U_2 - t)K \\
&\geq (l_1 - L_1)(U_2 - u_2)K.
\end{aligned}
$$

Similar to (4.51), it can be showed that

$$F(U_1, t) \leq F_2(t), \tag{4.52}$$

then by (4.52)

$$
\begin{aligned}
F_2(t) - F(s,t) \geq& F(U_1, t) - F(s,t) \\
\geq& F(U_1, t) - F(s,t) - F(U_1, S_2) + F(s, L_2) \\
\geq& (U_1 - s)(t - L_2)K \\
\geq& (U_1 - u_1)(l_2 - L_2)K.
\end{aligned}
$$

By Property (B1) in section 2.1.1,

$$F_1(U_1) = \beta_{p_n} N_{p_n}^{(1),l}(U_1) = \beta_{p_n},$$

$$F_2(U_2) = \gamma_{q_n} N_{q_n}^{(2),l}(U_2) = \gamma_{q_n},$$

and

$$F(U_1, U_2) = \alpha_{p_n,q_n} N_{p_n}^{(1),l}(U_1) N_{q_n}^{(2),l}(U_2) = \alpha_{p_n,q_n}.$$

Then under the last condition given in (4.1)

$$F_2(U_2) - F(U_1, U_2) = \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1 - \beta_{p_n} = 1 - F_1(U_1). \tag{4.53}$$

In the proof of Lemma 3.2, it is showed that $\frac{dF_1(s)}{ds} \geq \frac{\partial F(s,t)}{\partial s}$ and $\frac{dF_2(t)}{dt} \geq \frac{\partial F(s,t)}{\partial t}$, then

$$F_1(U_1) - F_1(s) \geq F(U_1, U_2) - F(s, U_2), \tag{4.54}$$

and

$$F_2(U_2) - F_2(t) \geq F(U_1, U_2) - F(U_1, t). \tag{4.55}$$

Hence by (4.53), (4.54) and (4.55)

$$\begin{aligned}
& 1 - F_1(s) - F_2(t) + F(s,t) \\
=& \{1 - F_1(U_1) + F_1(U_1) - F_1(s)\} \\
& - \{F_2(t) - F(U_1,t) + F(U_1,t) - F(s,t)\} \\
\geq& \{F_2(U_2) - F(U_1,U_2) + F(U_1,U_2) - F(s,U_2)\} \\
& - \{F_2(t) - F(U_1,t) + F(U_1,t) - F(s,t)\} \\
=& \{F_2(U_2) - F_2(t) - F(U_1,U_2) + F(U_1,t)\} \\
& + \{F(U_1,U_2) - F(s,U_2) - F(U_1,t) + F(s,t)\} \\
\geq& F(U_1,U_2) - F(s,U_2) - F(U_1,t) + F(s,t) \\
\geq& (U_1 - s)(U_2 - t)K \\
\geq& (U_1 - u_1)(U_2 - u_1)K.
\end{aligned}$$

□

Lemma 4.2. *Suppose (C3) holds and $p_n = q_n = n^v$. Then there exists $\tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n'$, such that*

$$\|F_n - F_0\|_\infty \leq K(n^{-pv}),$$

$$\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv}),$$

and

$$\|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv}).$$

Proof. Actually if the spline coefficients of F_n, $F_{n,1}$ and $F_{n,2}$ are chosen as $\alpha_{i,j} = F_0(\tau_i, \xi_j)$, $\beta_i = F_{0,1}(\tau_i)$ and $\gamma_j = F_{0,2}(\xi_j)$, where $\tau_i,\ i = 1, \ldots, p_n$ and $\xi_j,\ j = 1, \ldots, q_n$ are defined by (2.24) and (2.25) in the proof of Lemma 2.19 . Then under (C3) and (Sp1), Lemma 2.18 and Lemma 2.19 indicate that $\|F_n - F_0\|_\infty \leq C(n^{-pv})$, $\|F_{n,1} - F_{0,1}\|_\infty \leq C(n^{-pv})$, and $\|F_{n,2} - F_{0,2}\|_\infty \leq C(n^{-pv})$.

To complete the proof, in what follows $\alpha_{i,j}$, β_i and γ_j are showed to satisfy the conditions given by (4.1).

(i) $\alpha_{1,1} = F_0(\tau_1, \xi_1) \geq 0$;

(ii) $\alpha_{1,j+1} - \alpha_{1,j} = F_0(\tau_1, \xi_{j+1}) - F_0(\tau_1, \xi_j) \geq 0$;

(iii) $\alpha_{i+1,1} - \alpha_{i,1} = F_0(\tau_{i+1}, \xi_1) - F_0(\tau_i, \xi_1) \geq 0$;

(iv) $\frac{(\alpha_{i+1,j+1}-\alpha_{i+1,j})-(\alpha_{i,j+1}-\alpha_{i,j})}{\frac{\min_{i_1}\Delta_{i_1}^{(u)}}{l}\frac{\min_{j_1}\Delta_{j_1}^{(v)}}{l}} \geq \frac{\alpha_{i+1,j+1}-\alpha_{i,j+1}-\alpha_{i+1,j}+\alpha_{i,j}}{(\tau_{i+1}-\tau_i)(\xi_{j+1}-\xi_i)}$
$= \frac{F_0(\tau_{i+1},\xi_{j+1})-F_0(\tau_i,\xi_{j+1})-F_0(\tau_{i+1},\xi_j)+F_0(\tau_i,\xi_j)}{(\tau i+1-\tau_i)(\xi_{j+1}-\xi_i)} \geq \min_{\substack{s\in[L_1,U_1]\\ t\in[L_2,U_2]}} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} = b_0$,
then $(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq \frac{b_0 \min_{i_1}\Delta_{i_1}^{(u)} \min_{j_1}\Delta_{j_1}^{(v)}}{l^2}$, here l is necessary due to the definitions (2.24) and (2.25) of τ_i's and ξ_j's, respectively;

(v) $\beta_1 - \alpha_{1,q_n} = F_{0,1}(\tau_1) - F_0(\tau_1, \xi_{q_n}) \geq 0$;

(vi) $\beta_{i+1} - \beta_i - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) = F_{0,1}(\tau_{i+1}) - F_{0,1}(\tau_i) - (F_0(\tau_{i+1}, \xi_{q_n} - F_0(\tau_i, \xi_{q_n})) \geq 0$;

(vii) $\gamma_1 - \alpha_{p_n,1} = F_{0,2}(\xi_1) - F_0(\tau_{p_n}, \xi_1) \geq 0$;

(viii) $\gamma_{j+1} - \gamma_j - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) = F_{0,2}(\xi_{j+1}) - F_{0,1}(\xi_j) - (F_0(\tau_{p_n}, \xi_{j+1} - F_0(\tau_{p_n}, \xi_j)) \geq 0$;

(ix) $1 - \beta_{p_n} - \gamma_{q_n} + \alpha_{p_n,q_n} = 1 - F_{0,1}(\tau_{p_n}) - F_{0,2}(\xi_{q_n}) + F_0(\tau_{p_n}, \xi_{q_n}) \geq 0$.

The proof is complete. □

Lemma 4.3. *$\Lambda_0(s,t)$ and $\Lambda(s,t)$ are both partially nondecreasing functions with domain $[L_1,U_1]\times[L_2,U_2]$ and they satisfy $\|\Lambda-\Lambda_0\|_{L_2(\mu)}\le\eta$. If the following conditions (1) and (2) hold, then there exists constant K independent of Λ such that*

$$\sup_{(s,t)\in[L_1,U_1]\times[L_2,U_2]}|\Lambda(s,t)-\Lambda_0(s,t)|\le(\eta/c)^{1/K}.$$

(1) $\Lambda_0(s,t)$ is differentiable in both s and t directions and there exists a constant $0<f_0<\infty$ such that $1/f_0\le\partial\Lambda_0(s,t)/\partial s\le f_0$ and $1/f_0\le\partial\Lambda_0(s,t)/\partial t\le f_0$ for $\forall(s,t)\in[L_1,U_1]\times[L_2,U_2]$.

(2) The probability measure μ associated with L_2-norm has mixed derivative $\frac{\partial^2\mu(s,t)}{\partial s\partial t}$ satisfying $\frac{\partial^2\mu(s,t)}{\partial s\partial t}\ge c_0$ for some positive c_0.

Proof. Suppose that $(s^{'},t^{'})\in[L_1,U_1]\times[L_2,U_2]$ satisfies

$$|\Lambda(s^{'},t^{'})-\Lambda_0(s^{'},t^{'})|\ge(1/2)\sup_{(s,t)\in[L_1,U_1]\times[L_2,U_2]}|\Lambda(s,t)-\Lambda_0(s,t)|\equiv\xi/2.$$

Then either $\Lambda(s^{'},t^{'})\ge\Lambda_0(s^{'},t^{'})+\xi/2$ or $\Lambda_0(s^{'},t^{'})\ge\Lambda(s^{'},t^{'})+\xi/2$.

In the first case, there exist h satisfying $(s^{'}+h,t^{'}+h)\equiv(s^{''},t^{''})$, such that $\Lambda_0(s^{''},t^{''})=\Lambda_0(s^{'},t^{'})+\xi/2$, then

$$\begin{aligned}
\eta^2 &\ge \int\{\Lambda(s,t)-\Lambda(s,t)\}^2d\mu(s,t)\\
&= \int\int_{(s,t)\in[L_1,U_1]\times[L_2,U_2]}\{\Lambda(s,t)-\Lambda_0(s,t)\}^2\frac{\partial^2\mu(s,t)}{\partial s\partial t}dsdt\\
&\ge \int_{t^{'}}^{t^{''}}\int_{s^{'}}^{s^{''}}\{\Lambda(s,t)-\Lambda_0(s,t)\}^2\frac{\partial^2\mu(s,t)}{\partial s\partial t}dsdt\\
&\ge \int_{t^{'}}^{t^{''}}\int_{s^{'}}^{s^{''}}\{\Lambda_0(s^{''},t^{''})-\Lambda_0(s,t)\}^2\frac{\partial^2\mu(s,t)}{\partial s\partial t}dsdt
\end{aligned}$$

$$\geq c_0 \int_{t'}^{t''} \int_{s'}^{s''} \{\Lambda_0(s'', t'') - \Lambda_0(s,t)\}^2 ds dt$$
$$= c_0 \int_{t'}^{t''} \int_{\Lambda_0(s',t)}^{\Lambda_0(s'',t)} \{\Lambda_0(s'', t'') - x)\}^2 \frac{1}{\partial \Lambda_0(\Lambda_0^{-1}(x)|_t, t)/\partial s} dx dt$$
$$\geq (c_0/f_0) \int_{t'}^{t''} \int_{\Lambda_0(s',t)}^{\Lambda_0(s'',t)} \{\Lambda_0(s'', t'') - x)\}^2 dx dt$$
$$= (c_0/f_0) \int_{t'}^{t''} \{(\Lambda_0(s'', t'') - \Lambda_0(s', t))^3/3 - (\Lambda_0(s'', t'') - \Lambda_0(s'', t))^3/3\} dt.$$

Then by $x^3 - y^3 = (x-y)(x^2 + xy + y^2)$,

$$\eta^2 \geq \frac{c_0}{3f_0} \int_{t'}^{t''} (\Lambda_0(s'', t) - \Lambda_0(s', t)) \\ [(\Lambda_0(s'', t'') - \Lambda_0(s', t))^2 + (\Lambda_0(s'', t'') - \Lambda_0(s'', t))^2] dt. \tag{4.56}$$

Using Taylor expansion, there exists a $w \in (s', s'')$, such that

$$\Lambda_0(s'', t) - \Lambda_0(s', t) = (\partial \Lambda_0(w,t)/\partial s) h \geq h/f_0. \tag{4.57}$$

Using Taylor expansion along s and t directions, respectively, we have

$$\begin{aligned} \xi/2 &= \Lambda_0(s'', t'') - \Lambda_0(s', t') \\ &= \Lambda_0(s'', t'') - \Lambda_0(s'', t') + \Lambda_0(s'', t') - \Lambda_0(s', t') \\ &\leq 2hf_0. \end{aligned} \tag{4.58}$$

Combining (4.57) and (4.58) yields,

$$\Lambda_0(s'', t) - \Lambda_0(s', t) \geq \frac{\xi}{4f_0^2}. \tag{4.59}$$

Finally, substituting (4.59) into (4.56), we obtain

$$\begin{aligned}
\eta^2 &\geq \frac{c_0\xi}{12f_0^3}\int_{t'}^{t''}[(\Lambda_0(s'',t'')-\Lambda_0(s',t))^2+(\Lambda_0(s'',t'')-\Lambda_0(s'',t))^2]dt \\
&\geq \frac{c_0\xi}{12f_0^3}\int_{t'}^{t''}(\Lambda_0(s'',t'')-\Lambda_0(s'',t))^2dt \\
&= \frac{c_0\xi}{12f_0^3}\int_{\Lambda_0(s'',t')}^{\Lambda_0(s'',t'')}(\Lambda_0(s'',t'')-x)^2\frac{1}{\partial\Lambda_0(\Lambda_0^{-1}(x)|_{s''},s'')/\partial t}dx \\
&\geq \frac{c_0\xi}{12f_0^4}\int_{\Lambda_0(s'',t')}^{\Lambda_0(s'',t'')}(\Lambda_0(s'',t'')-x)^2dx \\
&= \frac{c_0\xi}{12f_0^4}(\Lambda_0(s'',t'')-\Lambda_0(s'',t'))^3/3 \\
&\geq \frac{c_0\xi^4}{2304f_0^{10}}.
\end{aligned}$$

This yields the stated conclusion with $K \equiv \sqrt{c_0/(2304f_0^{10})}$.

In the second case the same conclusion holds by a similar argument. □

Lemma 4.4. *(Lemma 7.1 in Wellner and Zhang (2007))* $\Lambda_0(s)$ *and* $\Lambda(s)$ *are both nondecreasing functions with domain* $[L_1, U_1]$ *and they satisfy* $\|\Lambda - \Lambda_0\|_{L_2(\mu)} \leq \eta$. *If the following conditions (1) and (2) hold, then there exists constant* K *independent of* Λ *such that*

$$\sup_{s\in[L_1,U_1]}|\Lambda(s)-\Lambda_0(s)| \leq (\eta/K)^{2/3}.$$

(1) $\Lambda_0(s)$ *is differentiable and there exists a constant* $0 < f_0 < \infty$ *such that* $1/f_0 \leq \frac{d\Lambda_0(s)}{ds} \leq f_0$ *for any* $s \in [L_1, U_1]$.

(2) The probability measure μ *associated with* L_2*-norm has derivative* $\dot{\mu}$ *satisfying* $\dot{\mu}(s) \geq c_0$ *for some positive* c_0.

CHAPTER 5
NUMERICAL STUDIES

In this chapter the finite sample performance of the proposed sieve maximum likelihood estimation method is justified. It has been mentioned in Chapter 3 that the constrained optimization problem with the I-splines has the advantage of simplicity in the constraints. Hence, the partially monotone tensor I-splines are used to compute the spline-based nonparametric estimate in the numerical studies.

5.1 Computing Algorithm

Given p_n and q_n, the partially monotone tensor I-splines-based nonparametric maximum likelihood estimation problem described in Section 3.2.2 is actually a restricted parametric maximum likelihood estimation problem for the coefficients of the I-splines or the tensor I-splines. For restricted parametric maximum likelihood estimation problems, Jamshidian (2004) generalized the gradient projection algorithm originally proposed by Rosen (1960) using the generalized Euclidean metric $\|x\| = x^T W x$, where W is a positive definite matrix and possibly varies from iteration to iteration. Zhang et al. (2009) implemented the generalized gradient projection algorithm for the spline-based maximum likelihood estimation for the Cox model with interval-censored data. In the following, we extend the algorithm steps adopted by Zhang et al. (2009) to compute the proposed tensor I-splines-based sieve nonparametric estimate.

Let $\frac{\partial \tilde{l}_n(\theta;\cdot)}{\partial \theta}$ and $H(\theta;\cdot)$ be the gradient and Hessian matrix of the log likelihood given by (3.10) with respect to $\theta = (\theta_1, \theta_2, \cdots, \theta_{p_n \cdot q_n + p_n + q_n}) = (\underline{\eta}, \underline{\omega}, \underline{\pi})$, respectively.

Since $H(\theta;\cdot)$ could be singular, we let $W = -H(\theta;\cdot) + \gamma I$, where I is identity matrix and $\gamma > 0$ is chosen sufficiently large so that W is positive definite. During the numerical computation, the index set of active constraints is denoted as $\mathcal{A} = \{i_1, i_2, \cdots, i_r\}$, that is, during the numerical computation, for $j = 1, 2, \cdots, r$,

(i) if $i_j \le p_n \cdot q_n + p_n + q_n$, then $\theta_{i_j} = 0$,

(ii) if $i_j = p_n \cdot q_n + p_n + q_n + 1$, then $\sum_{i=1}^{p_n \cdot q_n + p_n + q_n} \theta_i = 1$.

Suppose the indexes in $\mathcal{A}$ are in ascending order and $i_r = p_n \cdot q_n + p_n + q_n + 1$, then the working matrix corresponding to set $\mathcal{A}$ has the following form,

$$A = \begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 \end{bmatrix}_{r \times (p_n \cdot q_n + p_n + q_n)}.$$

The generalized gradient projection method is implemented in the following steps:

Step 1: (**Computing the feasible search direction**) Compute

$$\underline{d} = (d_1, d_2, \cdots, d_{p_n \cdot q_n + p_n + q_n}) = \{I - W^{-1}A^T(AW^{-1}A^T)^{-1}A\}W^{-1}\frac{\partial \tilde{l}_n(\theta;\cdot)}{\partial \theta}.$$

Step 2: (**Forcing the updated θ fulfills the constraints**) Compute

$$\gamma = \begin{cases} \min\{\min_{i:d_i<0}\{-\frac{\theta_i}{d_i}\}, \frac{1-\sum_{i=1}^{p_n\cdot q_n+p_n+q_n}\theta_i}{\sum_{i=1}^{p_n\cdot q_n+p_n+q_n}d_i}\}, & \text{if} \sum_{i=1}^{p_n\cdot q_n+p_n+q_n} d_i > 0, \\ \min_{i:d_i<0}\{-\frac{\theta_i}{d_i}\}, & \text{else.} \end{cases}$$

Doing so guarantees that $\theta_i + \gamma d_i \geq 0$ for $i = 1, 2, \cdots, p_n \cdot q_n + p_n + q_n$, and $\sum_{i=1}^{p_n\cdot q_n+p_n+q_n}(\theta_i + \gamma d_i) \leq 1$.

Step 3: (**Updating the solution by Step-Halving line search**) Find the smallest integer k starting from 0 such that

$$\tilde{l}_n(\theta + (1/2)^k\gamma\underline{d};\cdot) \geq \tilde{l}_n(\theta;\cdot).$$

Replace θ by $\tilde{\theta} = \theta + \min\{(1/2)^k\gamma, 0.5\}\underline{d}$.

Step 4: (**Updating the active constraint set and working matrix**) If $k = 0$ and $\gamma \leq 0.5$, modify $\mathcal{A}$ by add indexes of all the newly active constraints to $\mathcal{A}$ and accordingly modify the working matrix A.

Step 5: (**Checking the stopping criterion**) If $\|\underline{d}\| \geq \epsilon$, for small ϵ, go to Step1. otherwise compute $\lambda = (AW^{-1}A^T)^{-1}AW^{-1}\frac{\partial\tilde{l}_n(\theta;\cdot)}{\partial\theta}$.

(i) If $\lambda_j \geq 0$ for all j, set $\hat{\theta} = \theta$ and stop.

(ii) If there is at least one j such that $\lambda_j < 0$, let $j^* = \arg\min_{j:\lambda_j<0}\{\lambda_j\}$, then remove the index i_{j^*} from $\mathcal{A}$ and remove the j^*th row from A and go to Step1.

5.2 Simulation Studies

5.2.1 Design of the Studies

Copula functions are developed to model the joint distribution function through marginal distribution functions and associations among the individual event times. Recently, the Copula models have become very popular in bivariate survival analysis (Wang and Ding (2000), Jewell et al. (2005), Zhang (2008)). Nelson (2006) gave an extensive review of Copula functions. We consider two bivariate Copula functions in the simulation studies. One is the bivariate Gumbel Copula function

$$C_\alpha(u, v) = \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}\}$$

with $\alpha \geq 1$, the other is the bivariate Clayton Copula function

$$C_\alpha(u, v) = (u^{(1-\alpha)} + v^{(1-\alpha)} - 1)^{\frac{1}{1-\alpha}}$$

with $\alpha > 1$. For both Copula functions, a larger α corresponds to stronger positive association between two marginal distributions and $\alpha = 1$ corresponds to independence between the two event times. The association parameter α and Kendall's τ for Gumbel Copula and Clayton Copula, are related by $\tau = \frac{\alpha-1}{\alpha}$ and $\tau = \frac{\alpha-1}{\alpha+1}$, respectively.

Suppose the joint distribution of the bivariate data (T_1, T_2) is $C_\alpha(F_1, F_2)$, with F_1 and F_2 being the marginal distribution functions of T_1 and T_2, respectively. Then a sample of (T_1, T_2) can be generated based on the conditional distribution function by the following steps.

(i) Generate a random sample t_1 from distribution F_1;

(ii) Generate a random sample z_1 from uniform-$[0,1]$ distribution;

(iii) Solve equation $\frac{C_\alpha(F_1(t_1),F_2(t_2))}{F_1(t_1)} = z_1$ for t_2 (Usually the equation is solved numerically), then (t_1, t_2) is a sample of (T_1, T_2).

In simulation studies, we compare the performance of the proposed sieve maximum likelihood estimation to that of a three-stage semiparametric maximum pseudo likelihood estimation. The maximum pseudo likelihood estimation of the joint distribution function follows the idea of the semiparametric method proposed by Wang and Ding (2000) in which the main goal is to estimate the association parameter. However, Wang and Ding (2000) only discussed the case when censoring times are common for two event times. Their method is described as follows.

(i) Use the nonparametric method proposed by Groenoboom and Wellner (1992) to estimate the two marginal distribution functions.

(ii) Compute the maximum pseudo likelihood estimate of the association parameter based on the specified bivariate Copula function.

(iii) Estimate the bivariate distribution function by the specified bivariate Copula function with the unknown parameters estimated in (i) and (ii).

To justify the finite sample performance of the proposed sieve maximum likelihood estimation, two simulation studies are carried out. They are described as follows.

Study 1 The bivariate random event times are generated from bivariate Clayton Copula model. The proposed sieve maximum likelihood estimate is computed. The maximum pseudo likelihood estimate is also computed with the correctly specified Clayton Copula model.

Study 2 The bivariate random event times are generated from bivariate Gumbel Copula model. The proposed sieve maximum likelihood estimate is computed. The maximum pseudo likelihood estimate is also computed with the incorrectly specified Clayton Copula model.

In both studies, the performances of the proposed sieve maximum likelihood estimate and the maximum pseudo likelihood estimate are evaluated with various combinations of Kendall's τ $(\tau = 0.25, 0.5, 0.75)$ and sample size $(n = 100, 200)$. Under each of the six settings, the Monte-Carlo simulation study with 500 repetitions is conducted and the cubic I-splines are applied in the proposed sieve maximum likelihood. In what follows, the specifications of marginal distributions of event times T_1, T_2 and censoring times C_1, C_2, and the knots selection of cubic I-splines are described.

(i) (**Event times**) T_1 and T_2 in the each study are both chosen to be exponentially distributed with the rate parameter equal to 0.5. Since $Pr(T_i \geq 5) < 0.1$ for $i = 1, 2$, in the studies our interest is only the estimation inside the closed region $[0, 5] \times [0, 5]$.

(ii) (**Censoring times**) C_1 and C_2 in each study are both chosen to be independent and uniformly distributed on $[0.0201, 4.7698]$ in order to ensure that the bivariate samples of C_1 and C_2 are contained in $[0, 5] \times [0, 5]$.

(iii) (**Knots selection**) Theorem 2 in Chapter 4 implies that the proposed tensor spline-based estimator converges at a rate not faster than $n^{1/4}$, and convergence rate reaches

$n^{1/4}$ for $p \geq 2$ and $v = \frac{1}{4p}$. If $p = 2$, then $v = 1/8$ and the number of subintervals made of the knot sequence could be chosen as $n^{1/8}$. This choice of the knot sequences is mainly of interest in asymptotic properties when n is assumed very large. However, in practice the sample size is usually not very large. For example if the sample size is 100 and hence $100^{1/8} < 2$, which makes the spline-based method infeasible. Our experiments show that $n^{1/3}$ is a reasonable choice for the number of subintervals made of the knot sequence when sample size is 100 and 200. In other words, we choose 5 and 6 as the numbers of subintervals made of the knot sequence for sample size 100 and 200, respectively. Using the percentile of the distribution of T_1 or T_2 inside $[0, 5]$, the knots of cubic I-spline basis functions are chosen as $[0, 0.41, 0.91, 1.60, 2.65, 5]$ for sample size 100 and chosen as $[0, 0.33, 0.73, 1.23, 1.89, 2.90, 5]$ for sample size 200.

5.2.2 Results of the Studies

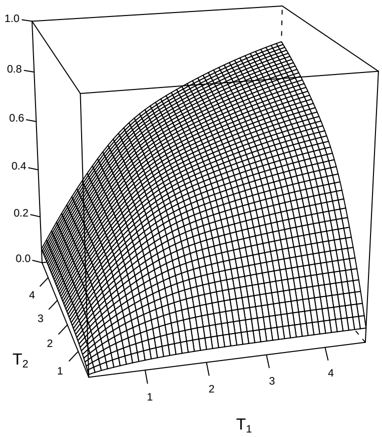
The estimation bias of the joint distribution function with the Monte-Carlo studies of 500 repetitions is graphically presented through all odd-numbered figures (Figure 5.1, Figure 5.3, $\cdots$, Figure 5.23) for various cases in two studies. And these figures indicate that the bias of the proposed sieve maximum likelihood estimation is noticeably smaller than that of the maximum pseudo likelihood estimation near the boundary of closed region $[0.1, 4.7] \times [0.1, 4.7]$, the bias of the proposed sieve maximum likelihood estimation near the origin increases as Kendall's τ increases, and the bias of the proposed sieve maximum likelihood estimation decreases as sample size increases from 100 to 200.

Moreover, the mean estimate of the marginal distribution function of T_1 from the same Monte-Carlo studies is plotted in all even-numbered figures (Figure 5.2, Figure 5.4, $\cdots$, Figure 2.24). And these figures indicate that the bias of the proposed sieve maximum likelihood estimation for the marginal distribution function of T_1 is much smaller than that of the maximum pseudo likelihood estimation near two end points of interval $[0.1, 4.7]$, and in general, the bias of the proposed sieve maximum likelihood estimation for the marginal distribution function of T_1 increases as Kendall's τ increases or sample size decreases from 200 to 100.

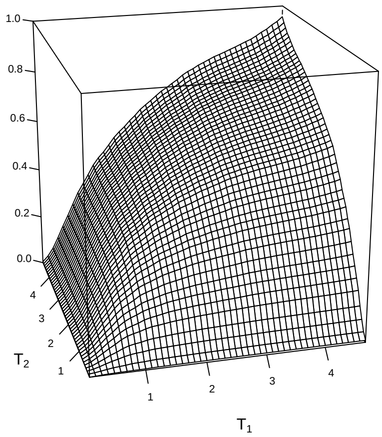
Table 5.1-Table 5.12 display the squared estimation bias (*Bias*^{2}) and mean square error (*MSE*) based on 500 repetitions for both proposed sieve maximum likelihood estimation (*Sieve*) and the maximum pseudo likelihood estimation (*Pseudo*) at the 12 pairs of time points (T_1, T_2) for different sample sizes, different kendall's τ in two simulation studies. At the bottom of each table *Average Bias*^{2} calculates the average of squared estimation bias at 2209 values of (T_1, T_2) with both T_1 and T_2 uniformly taking 47 values from 0.1 to 4.7. *Average MSE* gives the average of mean square error at the same 2209 points. It appears that in terms of both *Average Bias*^{2} and *Average MSE*, the performance of the sieve maximum likelihood estimation is better than the maximum pseudo likelihood estimation on different settings on the sample size and the association level in both simulation studies. The bias of the proposed sieve maximum likelihood estimation may be a little larger far from boundary of region $[0, 1, 4.7] \times [0.1, 4.7]$ than the maximum pseudo likelihood estimation. The mean square error of the proposed sieve maximum likelihood estimation may be a little larger near the origin than the maximum pseudo likelihood estimation. Both

overall bias and overall mean square error of the proposed sieve maximum likelihood estimation are smaller compared to its counterpart. Both estimation bias and mean square error of the proposed sieve maximum likelihood estimation are noticeably decreased as sample size increases from 100 to 200.

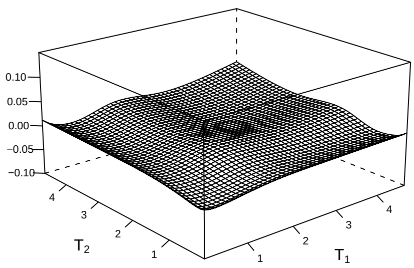
Figure 5.1: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)



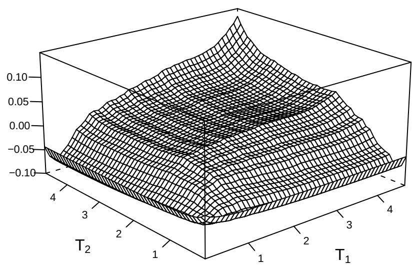
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.2: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)

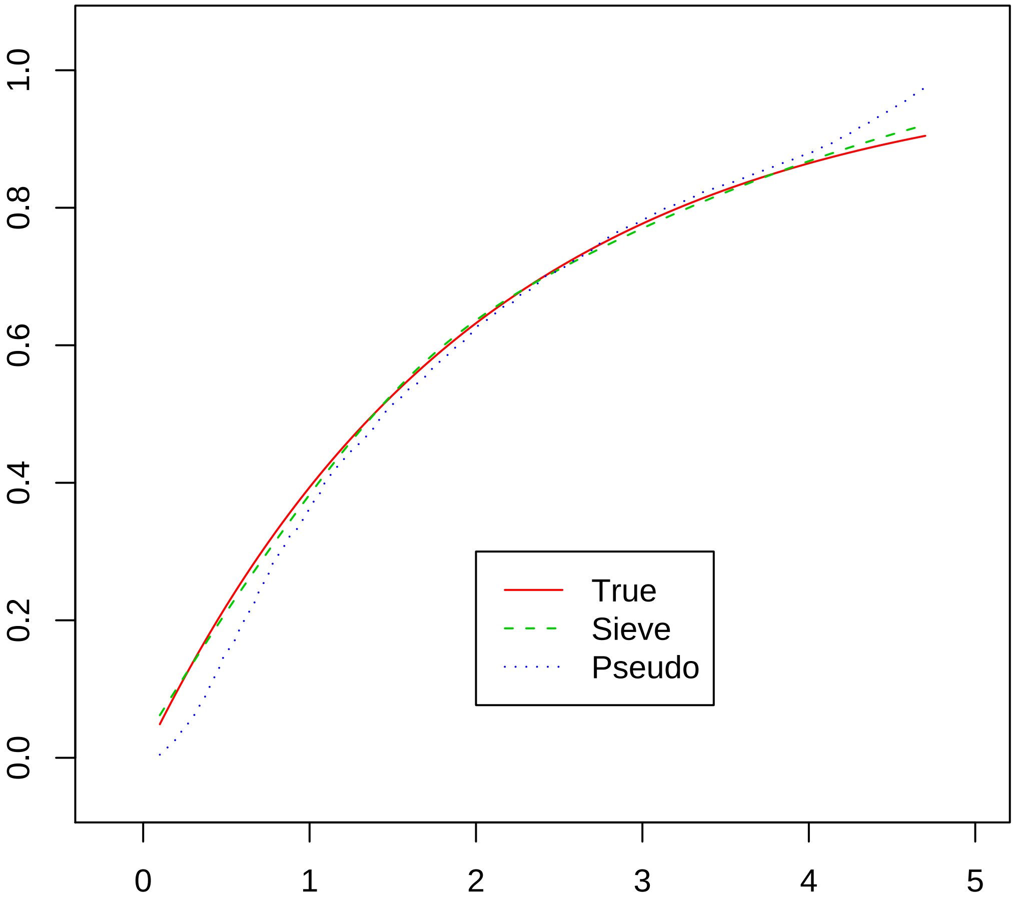
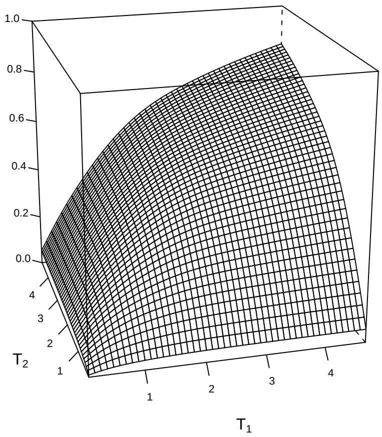
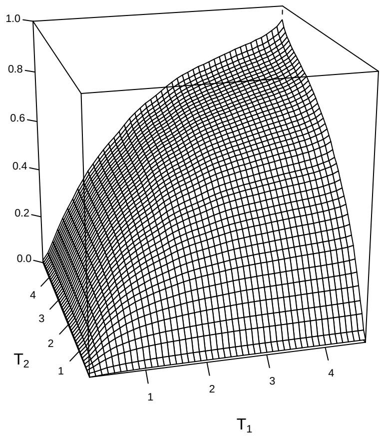


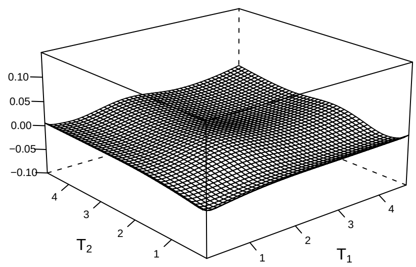
Figure 5.3: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)



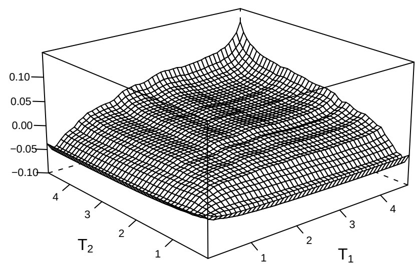
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.4: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)

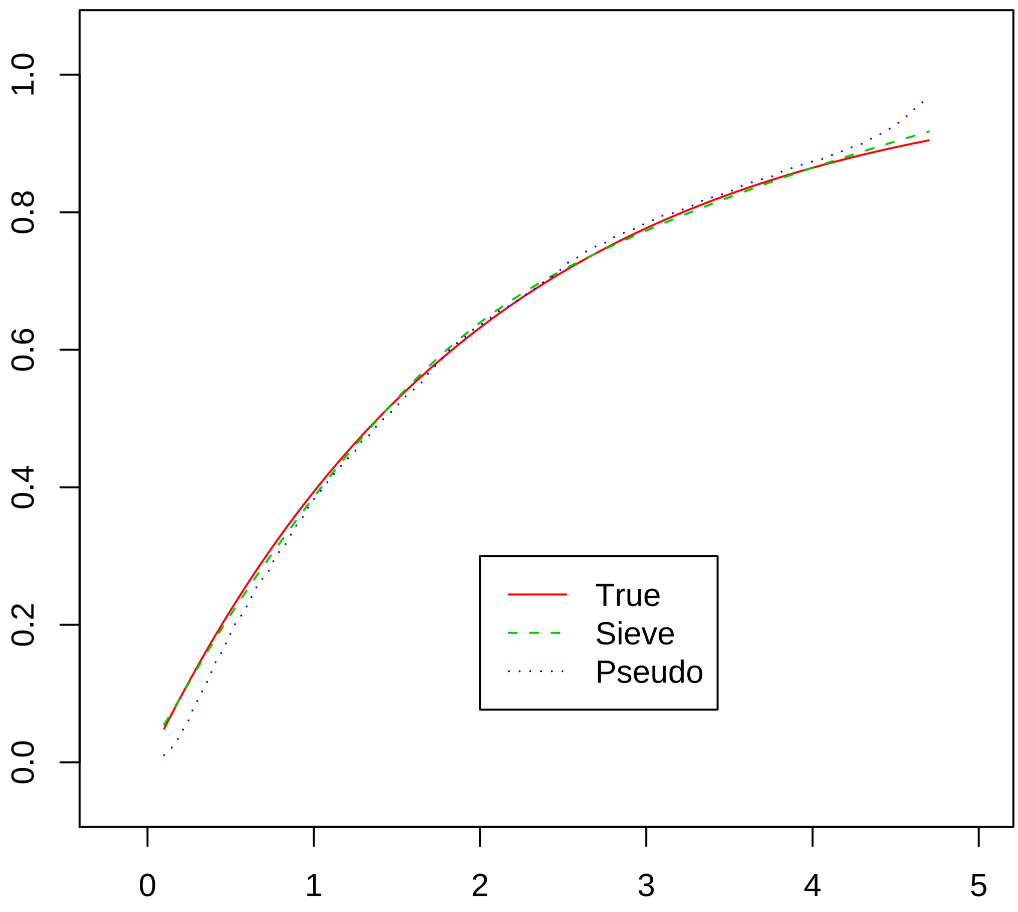
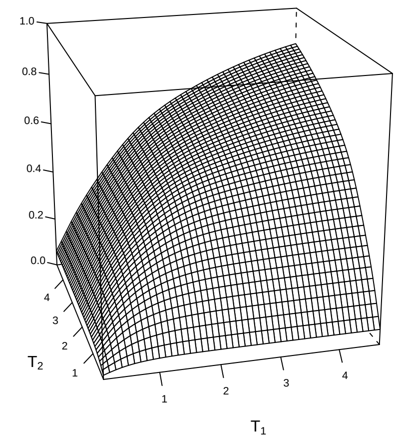
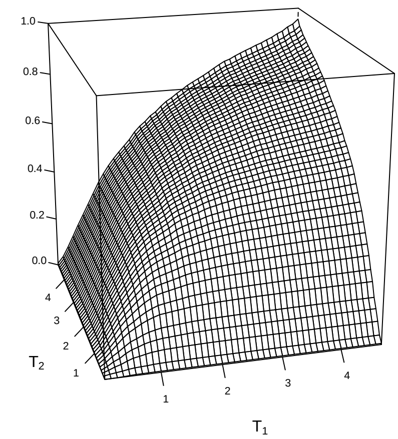


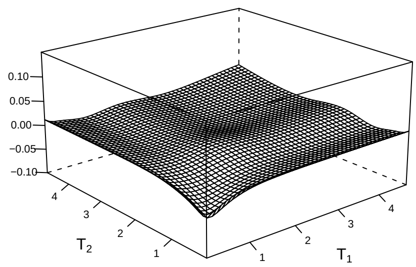
Figure 5.5: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$)



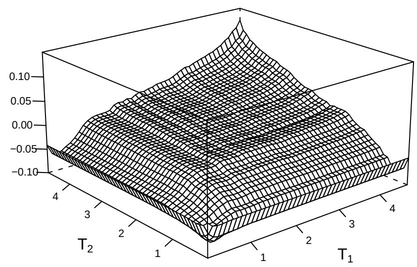
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.6: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$)

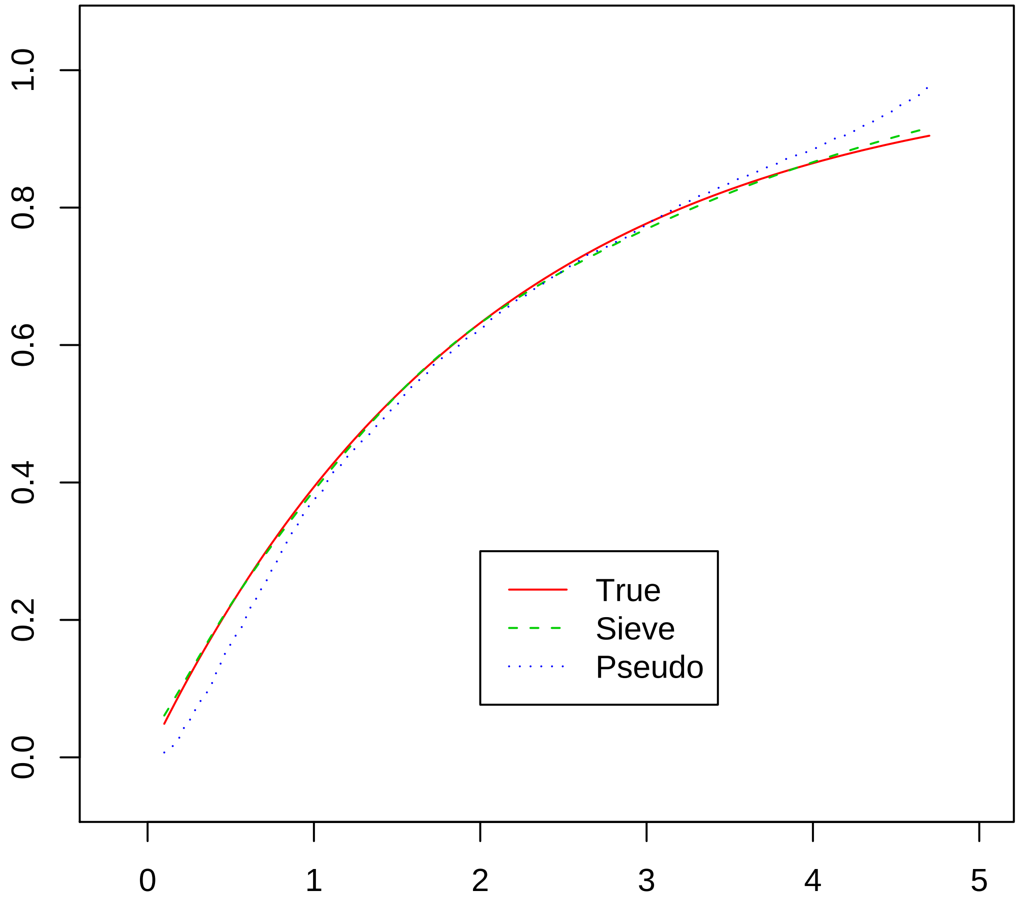
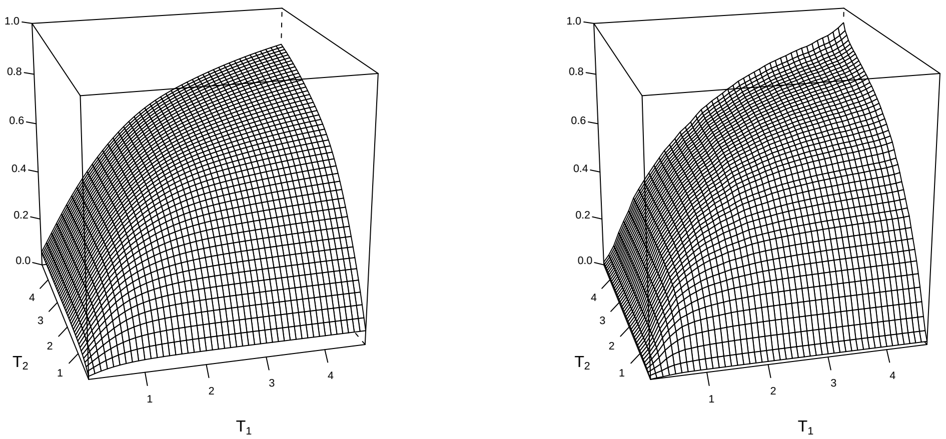
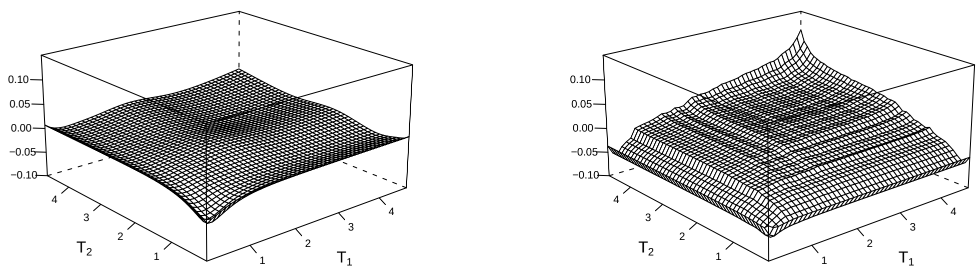


Figure 5.7: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$)



(a) Estimation of the sieve method

(b) Estimation of the pseudo method



(c) Bias of the sieve method

(d) Bias of the pseudo method

Figure 5.8: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$)

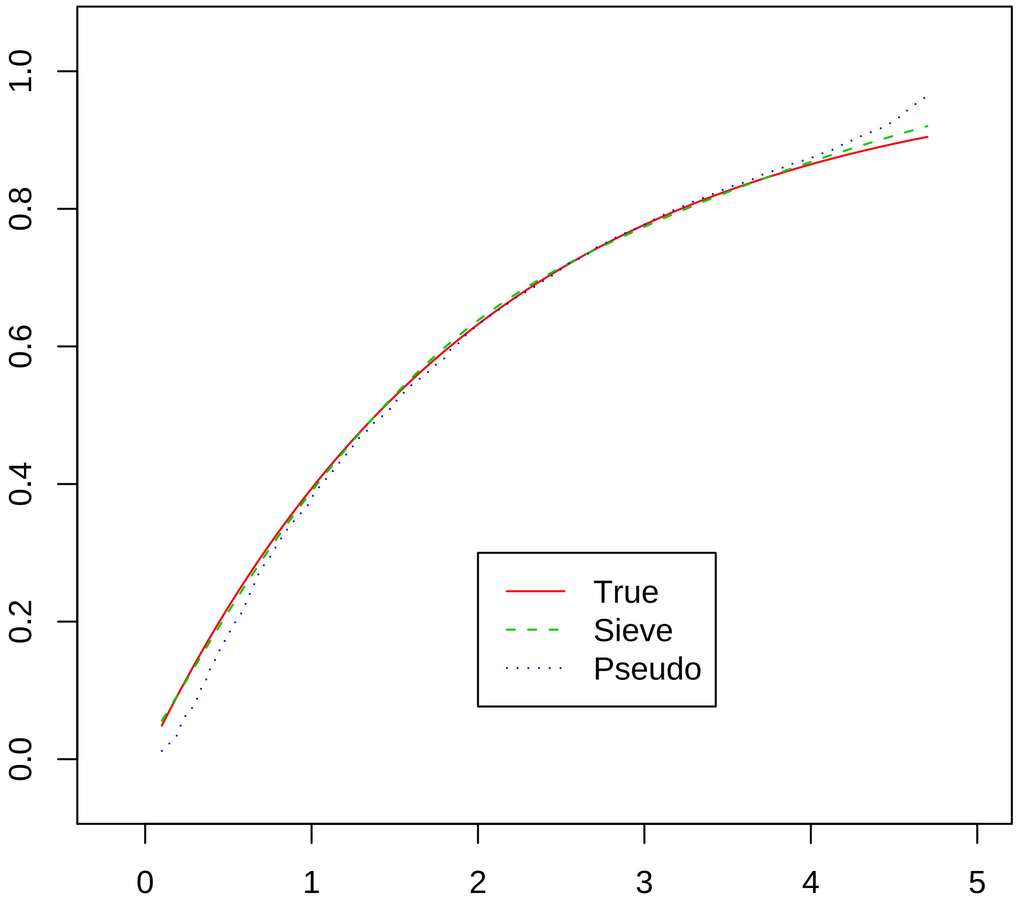
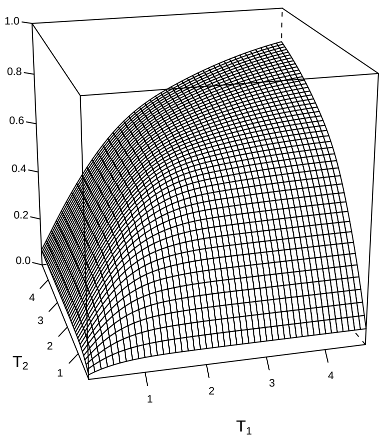
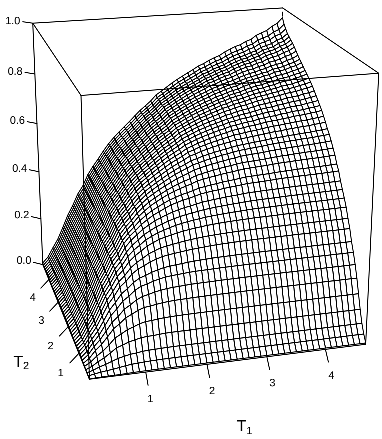


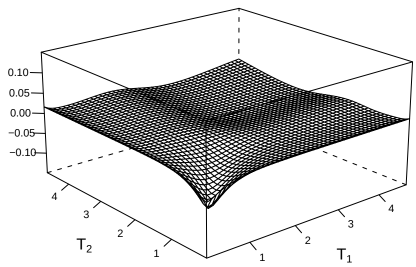
Figure 5.9: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)



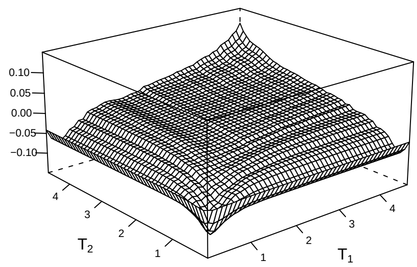
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.10: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)

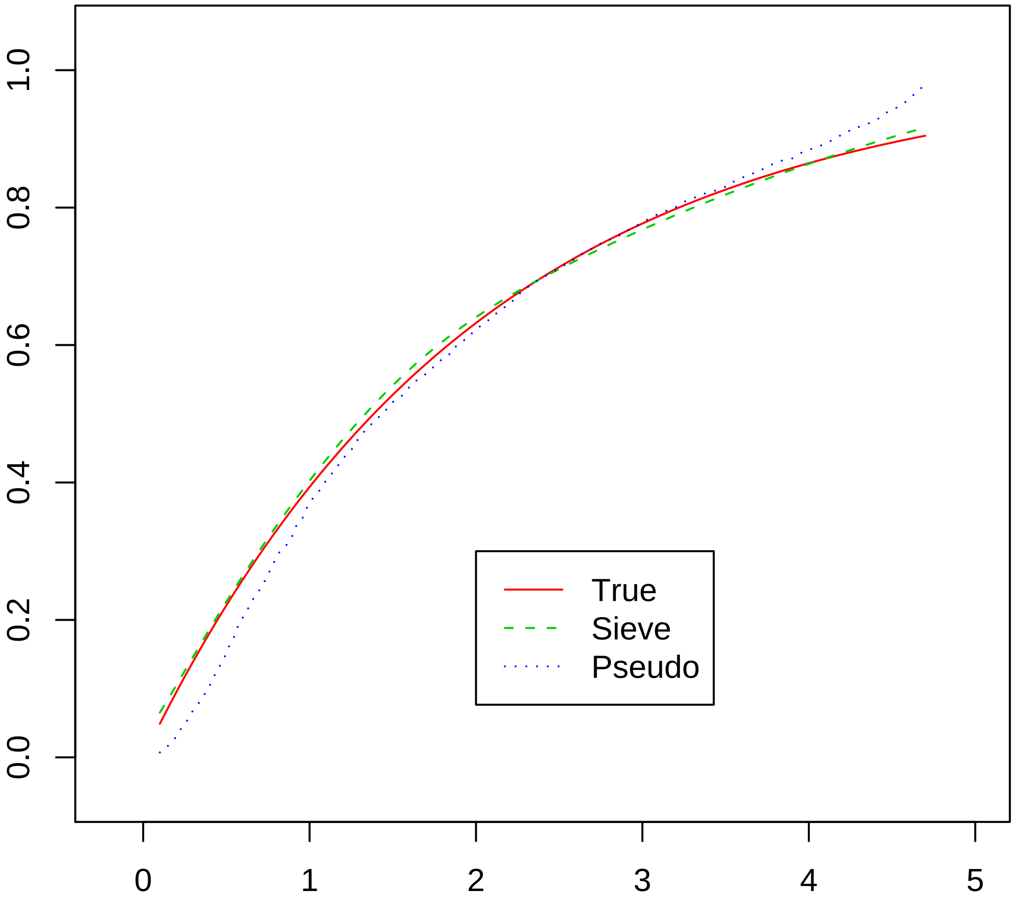
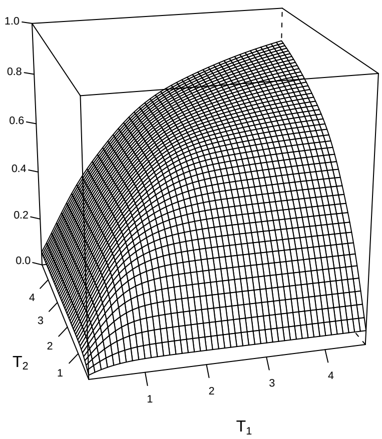
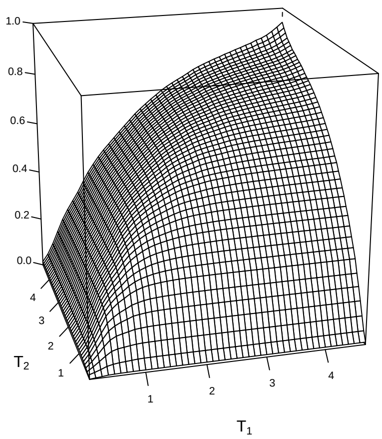


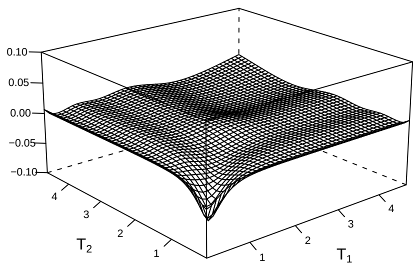
Figure 5.11: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)



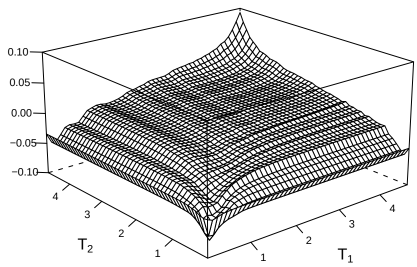
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.12: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Clayton Copula model and $n = 200$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)

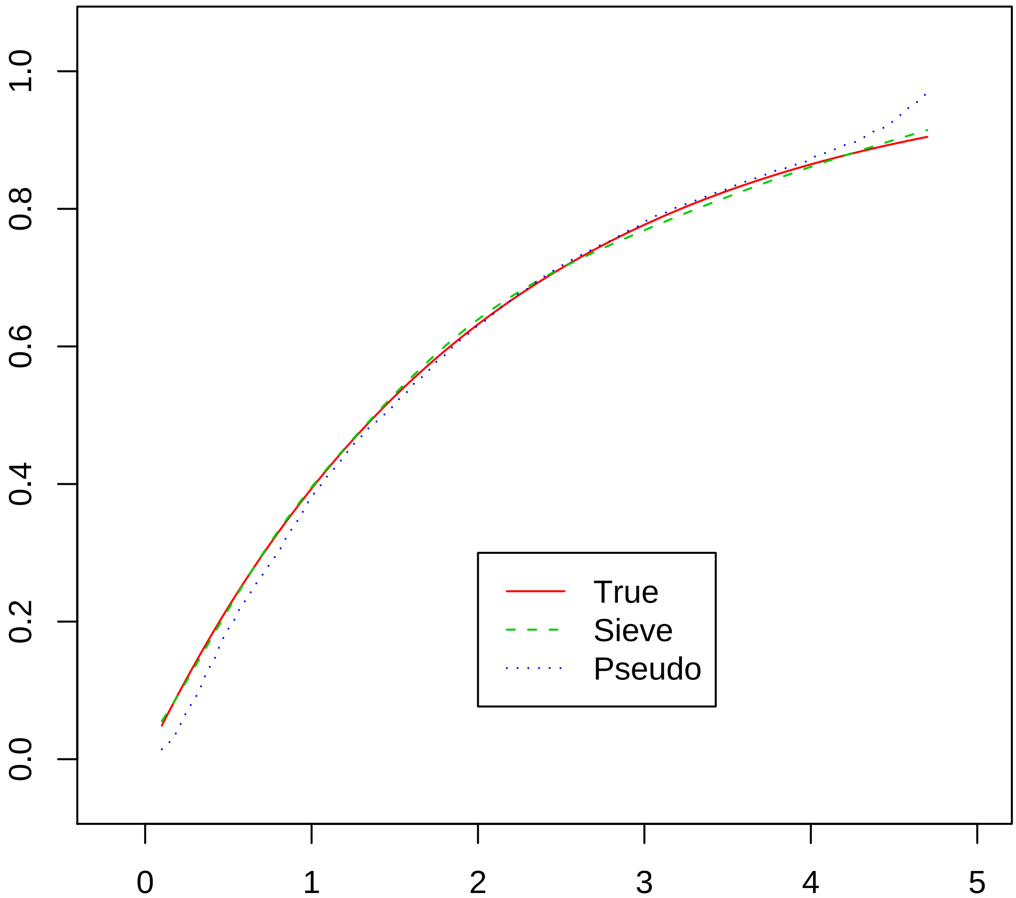
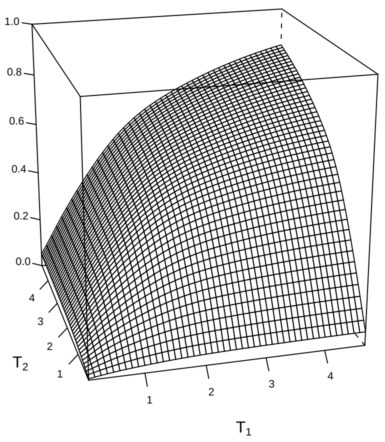
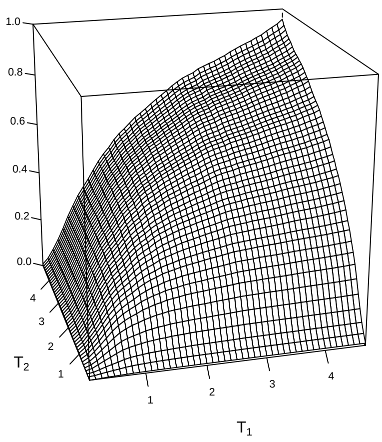


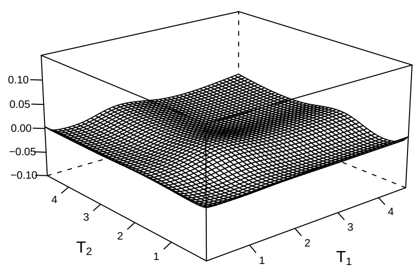
Figure 5.13: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Clayton Copula model with the underlying distribution being correctly specified when sample size $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)



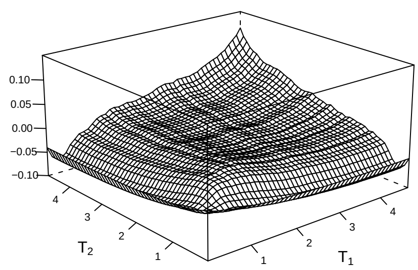
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.14: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)

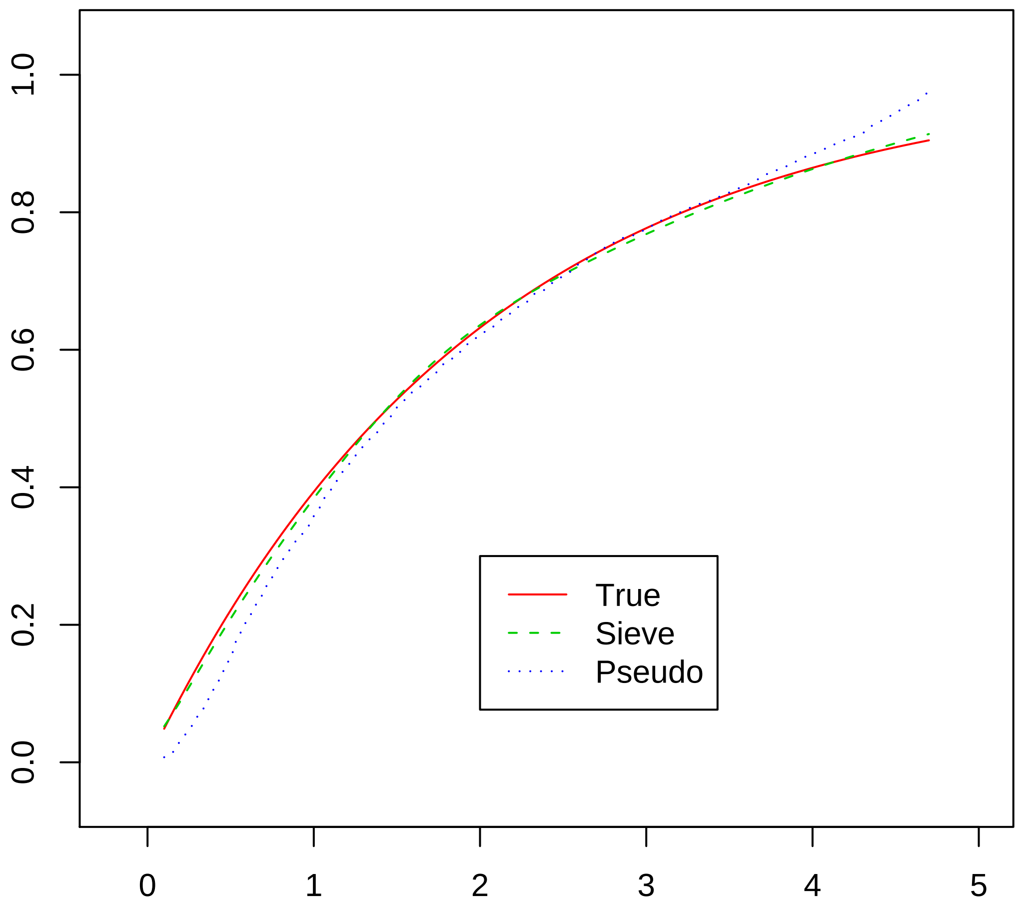
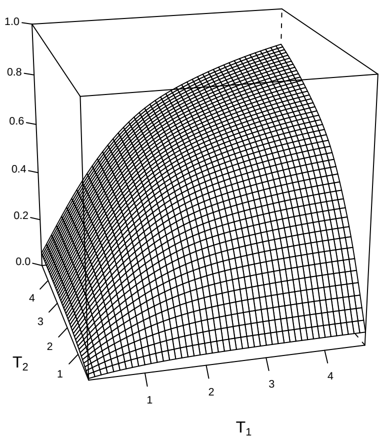
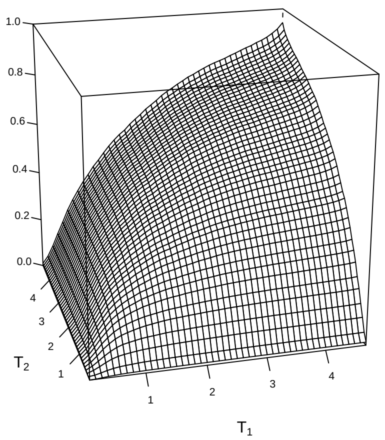


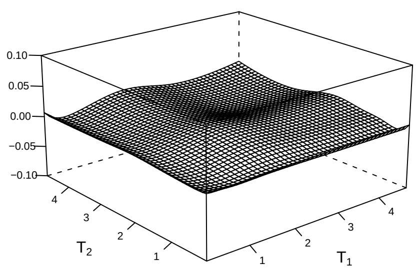
Figure 5.15: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Gumbel Copula model with the underlying distribution being incorrectly specified as bivariate Clayton Copula model when sample size $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)



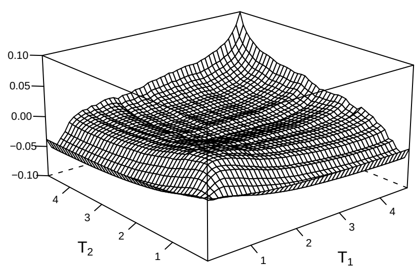
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.16: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)

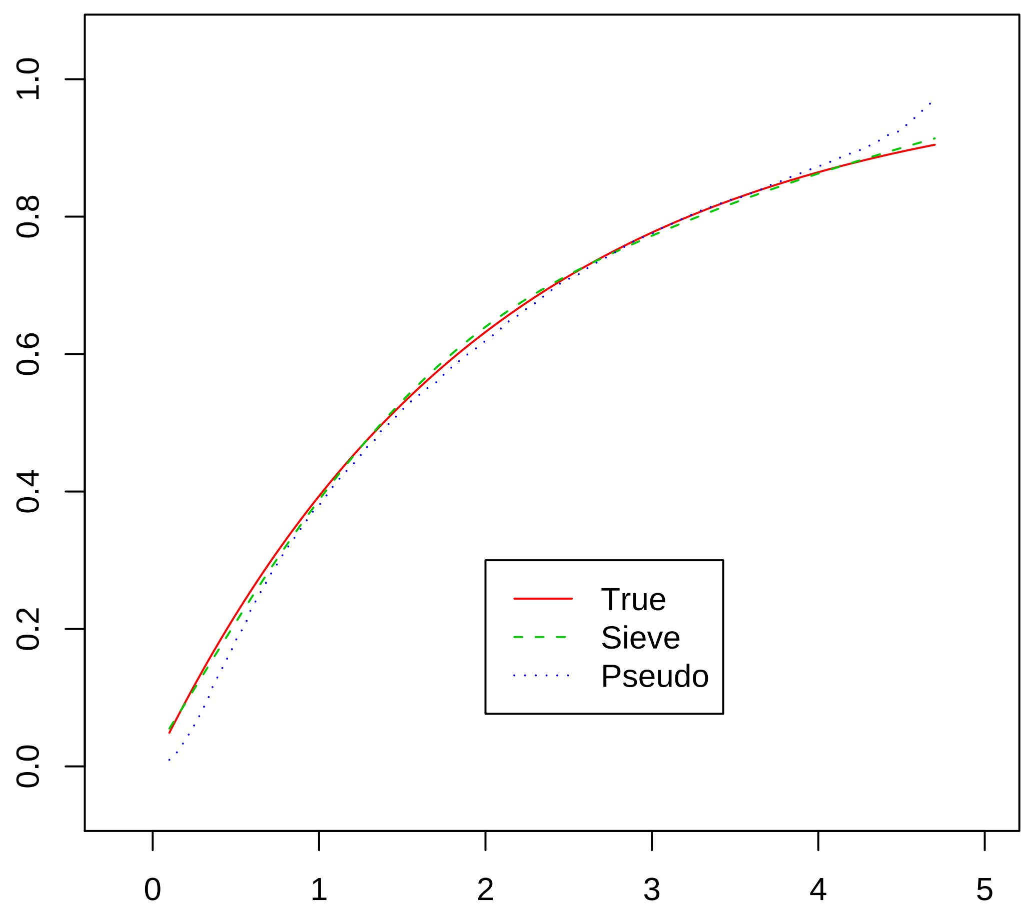
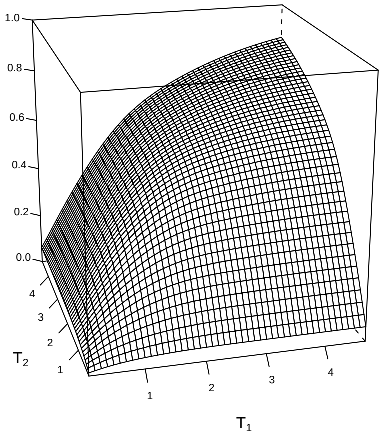
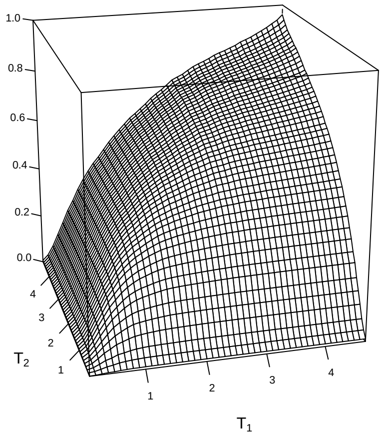


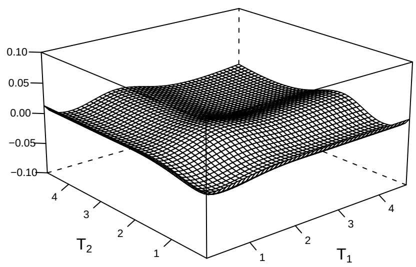
Figure 5.17: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Gumbel Copula model with the underlying distribution being incorrectly specified as bivariate Clayton Copula model when sample size $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)



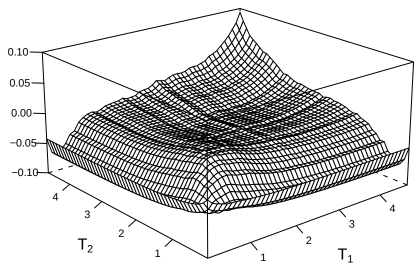
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.18: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)

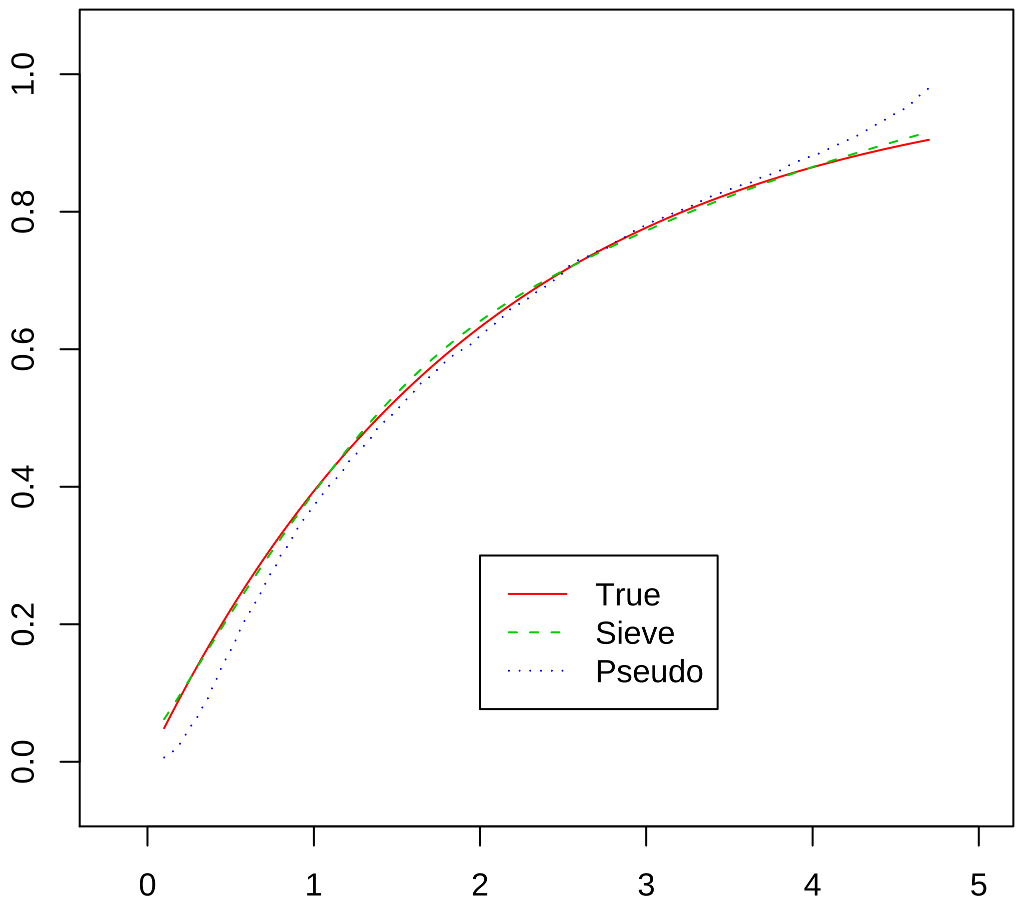
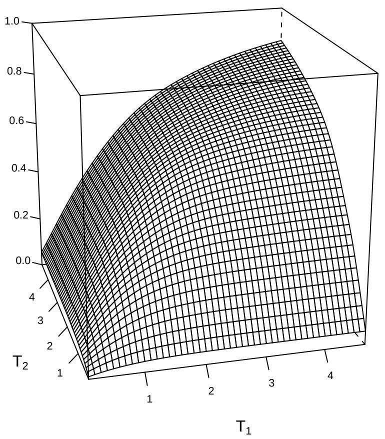
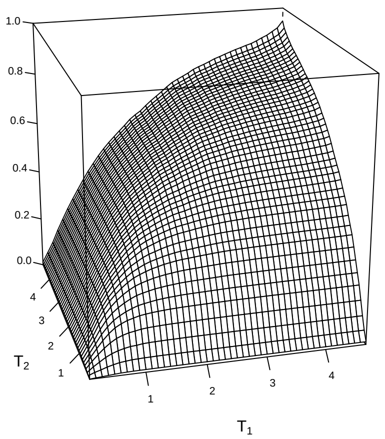


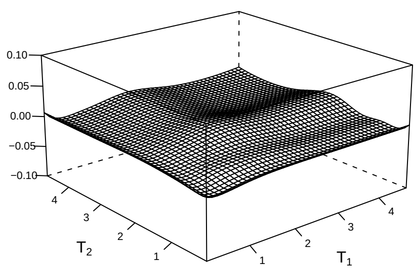
Figure 5.19: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Gumbel Copula model with the underlying distribution being incorrectly specified as bivariate Clayton Copula model when sample size $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)



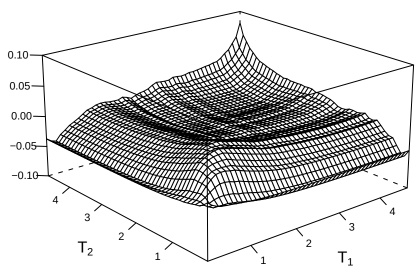
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.20: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)

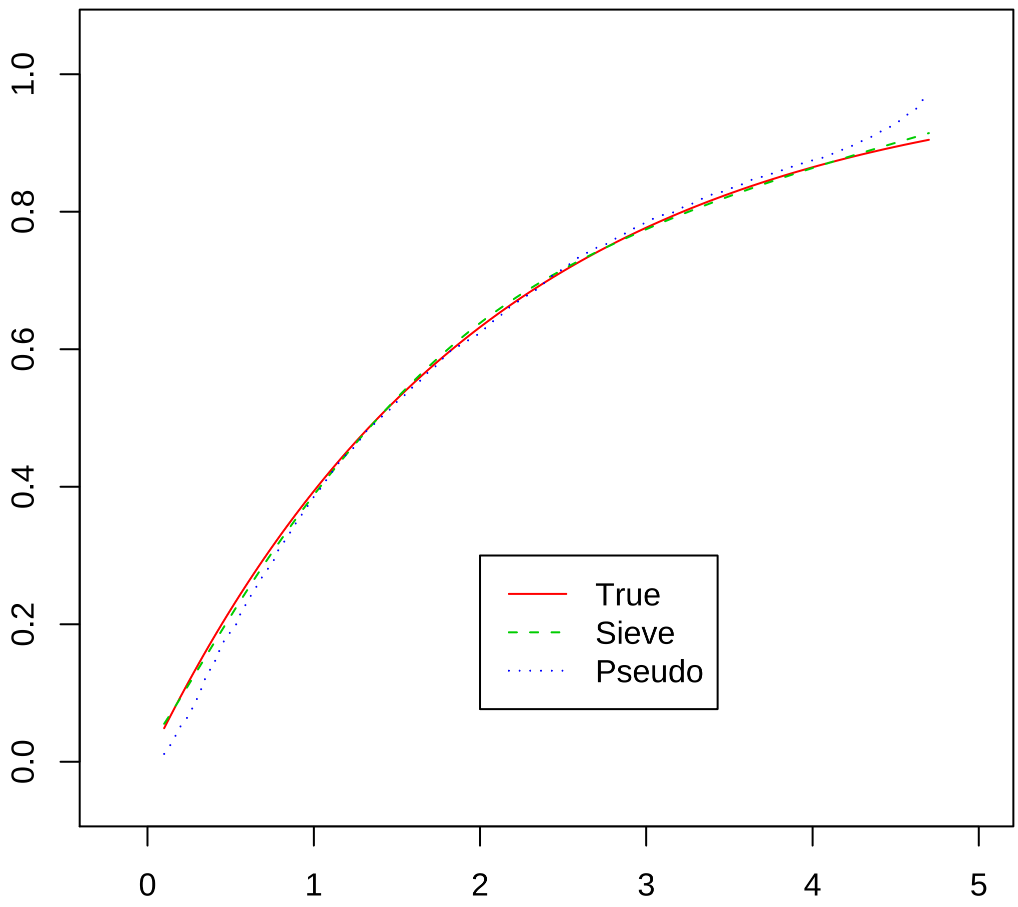
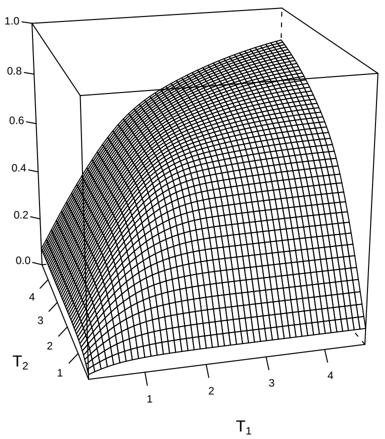
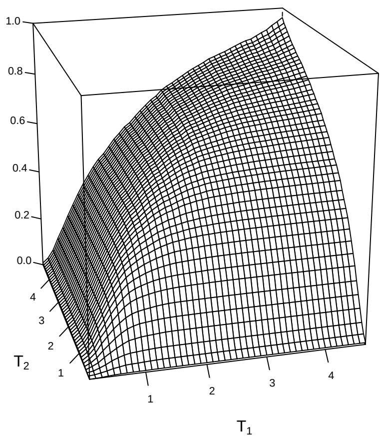


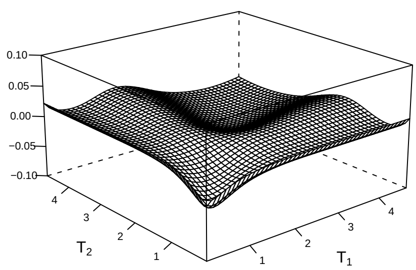
Figure 5.21: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Gumbel Copula model with the underlying distribution being incorrectly specified as bivariate Clayton Copula model when sample size $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)



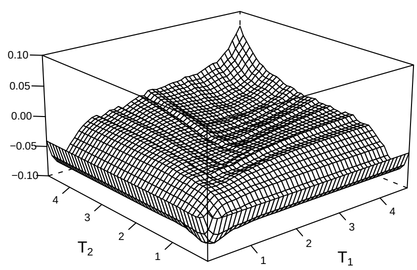
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.22: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)

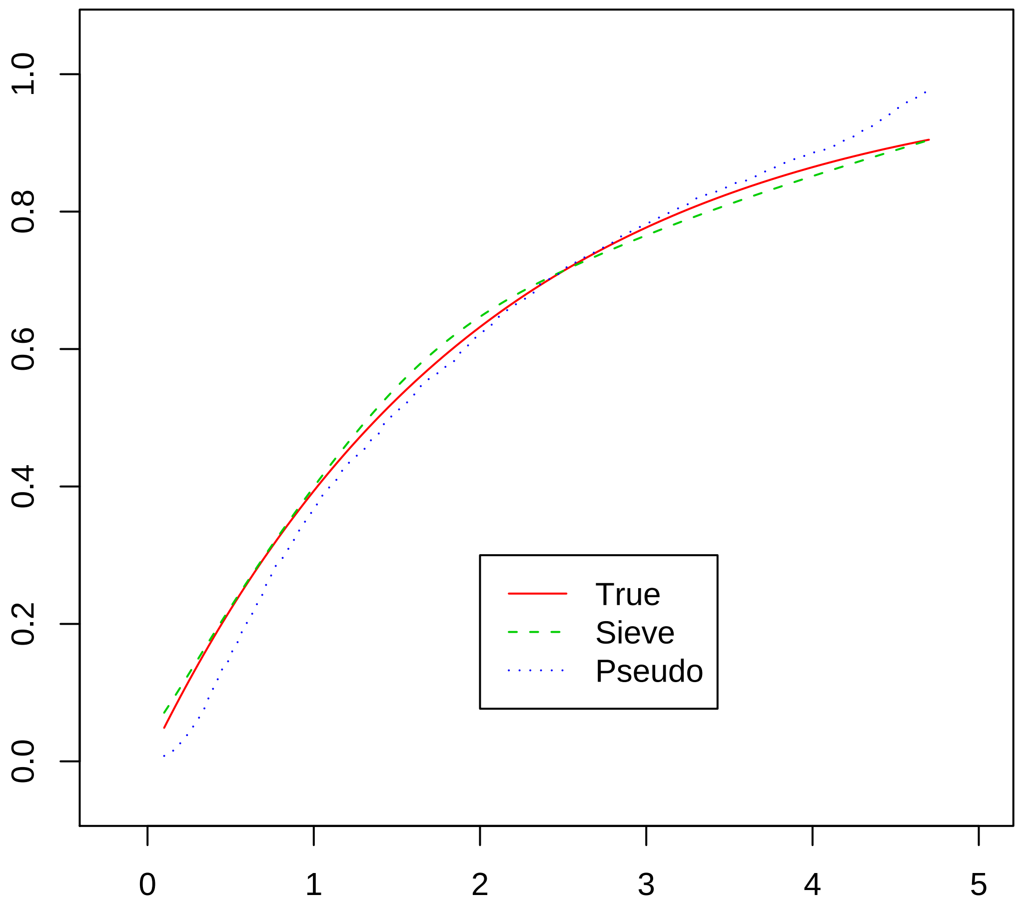
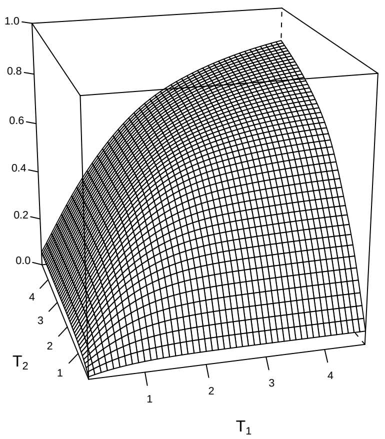
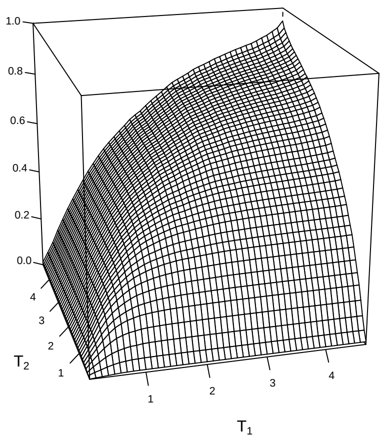


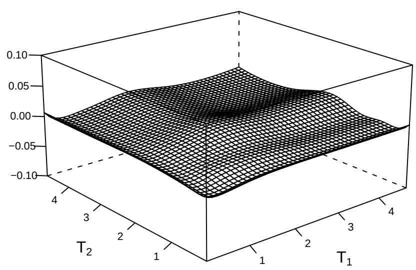
Figure 5.23: Comparison of the proposed sieve maximum likelihood estimation ((a), (c)) and the maximum pseudo likelihood estimation ((b), (d)) for the joint distribution function of bivariate Gumbel Copula model with the underlying distribution being incorrectly specified as bivariate Clayton Copula model when sample size $n = 200$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)



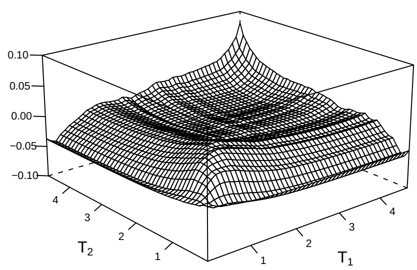
(a) Estimation of the sieve method



(b) Estimation of the pseudo method



(c) Bias of the sieve method



(d) Bias of the pseudo method

Figure 5.24: Comparison of the proposed sieve maximum likelihood estimation (Sieve) and the nonparametric maximum likelihood estimation (Pseudo) for the marginal distribution function of T_1 when (T_1, T_2) is distributed according to Gumbel Copula model and $n = 200$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)

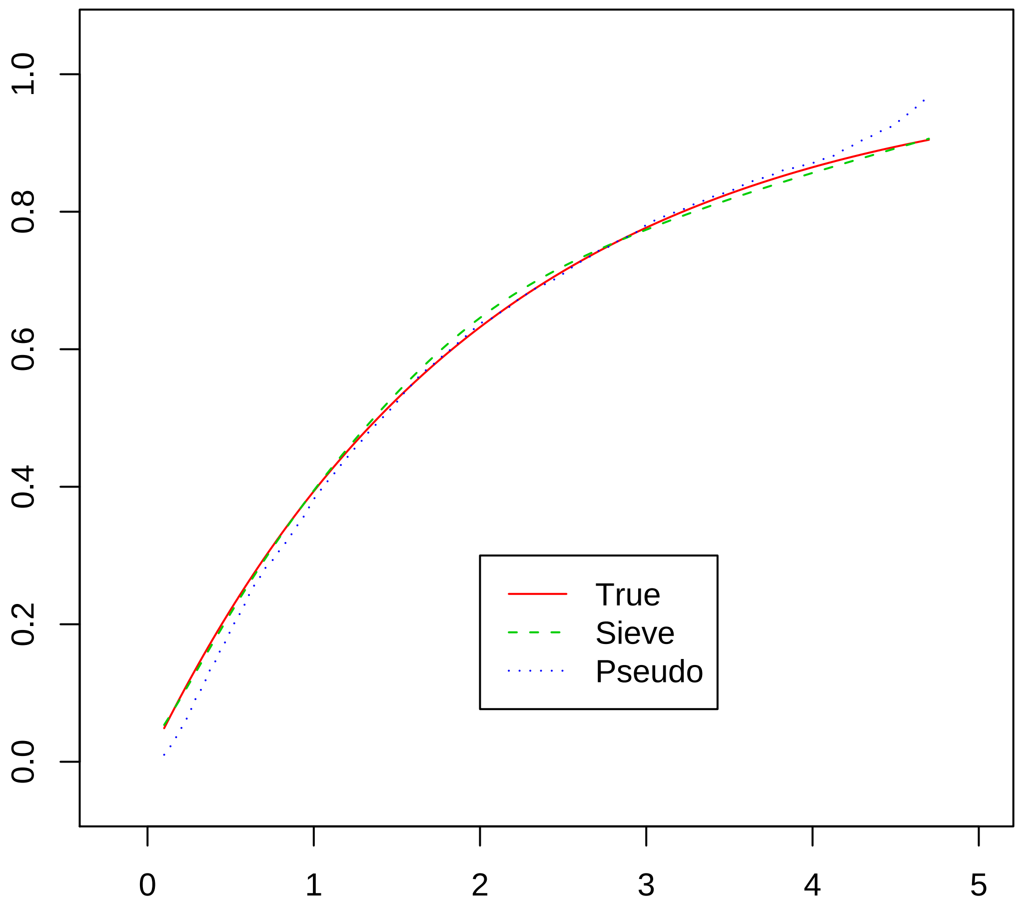


Table 5.1: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	7.63e-05	7.58e-04	3.66e-04	3.66e-04
(0.1, 1.6)	3.36e-05	4.84e-03	1.65e-03	2.03e-03
(0.1, 4.6)	1.46e-04	5.88e-03	1.90e-03	2.47e-03
(1.6, 0.1)	5.50e-06	3.26e-03	1.41e-03	2.40e-03
(1.6, 1.6)	1.63e-05	6.32e-03	6.83e-07	8.65e-03
(1.6, 4.6)	8.09e-05	7.17e-03	1.37e-04	1.22e-02
(4.6, 0.1)	6.95e-05	3.94e-03	1.61e-03	2.99e-03
(4.6, 1.6)	8.09e-05	6.62e-03	2.45e-04	1.46e-02
(4.6, 4.6)	7.04e-04	5.83e-03	1.09e-02	1.76e-02
Average	6.13e-05	6.48e-03	1.22e-03	1.09e-02

Table 5.2: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 5/3$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	1.17e-04	4.54e-04	3.56e-04	3.65e-04
(0.1, 1.6)	1.63e-08	2.18e-03	1.26e-03	2.11e-03
(0.1, 4.6)	2.99e-05	2.75e-03	1.43e-03	2.58e-03
(1.6, 0.1)	2.16e-08	2.20e-03	1.28e-03	2.16e-03
(1.6, 1.6)	4.68e-05	4.31e-03	1.42e-05	4.89e-03
(1.6, 4.6)	9.73e-05	3.91e-03	9.97e-05	8.08e-03
(4.6, 0.1)	1.27e-05	2.49e-03	1.41e-03	2.74e-03
(4.6, 1.6)	3.02e-05	3.79e-03	1.72e-04	7.90e-03
(4.6, 4.6)	3.31e-04	3.41e-03	6.16e-03	1.15e-02
Average	5.64e-05	3.70e-03	5.71e-04	6.32e-03

Table 5.3: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$))

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	3.49e-04	1.55e-03	1.18e-03	1.19e-03
(0.1, 1.6)	8.56e-05	4.99e-03	1.78e-03	3.08e-03
(0.1, 4.6)	1.42e-04	5.35e-03	1.74e-03	3.28e-03
(1.6, 0.1)	8.99e-05	4.71e-03	1.93e-03	2.63e-03
(1.6, 1.6)	2.17e-06	5.01e-03	2.30e-04	8.66e-03
(1.6, 4.6)	6.16e-06	5.88e-03	4.27e-06	1.29e-02
(4.6, 0.1)	1.26e-04	4.82e-03	1.93e-03	2.68e-03
(4.6, 1.6)	5.56e-05	5.64e-03	8.11e-06	1.34e-02
(4.6, 4.6)	3.95e-04	4.91e-03	7.96e-03	1.44e-02
Average	5.83e-05	5.93e-03	1.54e-03	1.18e-02

Table 5.4: The comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 3$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	5.06e-04	9.71e-04	1.15e-03	1.19e-03
(0.1, 1.6)	2.49e-05	2.79e-03	1.37e-03	2.66e-03
(0.1, 4.6)	4.99e-05	2.90e-03	1.35e-03	2.75e-03
(1.6, 0.1)	1.57e-05	2.81e-03	1.33e-03	2.62e-03
(1.6, 1.6)	8.89e-08	3.82e-03	9.61e-05	5.33e-03
(1.6, 4.6)	2.28e-05	4.24e-03	2.73e-06	7.72e-03
(4.6, 0.1)	3.69e-05	2.92e-03	1.32e-03	2.70e-03
(4.6, 1.6)	2.66e-06	3.84e-03	1.55e-06	8.58e-03
(4.6, 4.6)	3.01e-04	2.97e-03	5.01e-03	9.75e-03
Average	6.25e-05	3.71e-03	6.92e-04	6.76e-03

Table 5.5: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	6.63e-04	2.02e-03	1.85e-03	1.89e-03
(0.1, 1.6)	2.30e-04	6.19e-03	1.73e-03	2.97e-03
(0.1, 4.6)	2.65e-04	6.38e-03	1.73e-03	2.99e-03
(1.6, 0.1)	1.55e-04	5.40e-03	1.90e-03	2.76e-03
(1.6, 1.6)	1.02e-04	4.83e-03	1.56e-03	1.03e-02
(1.6, 4.6)	1.50e-04	5.11e-03	1.39e-04	1.37e-02
(4.6, 0.1)	1.89e-04	5.56e-03	1.89e-03	2.76e-03
(4.6, 1.6)	5.61e-05	5.49e-03	3.83e-05	1.39e-02
(4.6, 4.6)	9.97e-05	3.90e-03	5.55e-03	1.15e-02
Average	1.33e-04	6.05e-03	1.71e-03	1.25e-02

Table 5.6: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the pseudo maximum likelihood estimation of bivariate Clayton Copula distributed event times when the underlying distribution is correctly specified with sample size $n = 200$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 7$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	7.48e-04	1.61e-03	1.81e-03	1.90e-03
(0.1, 1.6)	3.65e-05	3.23e-03	1.18e-03	3.01e-03
(0.1, 4.6)	4.17e-05	3.29e-03	1.18e-03	3.02e-03
(1.6, 0.1)	3.60e-05	3.13e-03	1.53e-03	2.83e-03
(1.6, 1.6)	1.20e-04	3.62e-03	6.74e-04	6.50e-03
(1.6, 4.6)	2.09e-05	3.33e-03	8.12e-05	9.22e-03
(4.6, 0.1)	4.04e-05	3.16e-03	1.53e-03	2.84e-03
(4.6, 1.6)	4.93e-06	3.07e-03	3.06e-05	8.61e-03
(4.6, 4.6)	1.88e-04	2.45e-03	4.08e-03	8.23e-03
Average	6.07e-05	3.45e-03	6.52e-04	7.12e-03

Table 5.7: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model with sample size $n = 100$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	5.28e-07	3.30e-04	3.66e-05	4.99e-05
(0.1, 1.6)	4.68e-06	2.78e-03	9.98e-04	1.90e-03
(0.1, 4.6)	1.26e-05	3.62e-03	1.63e-03	2.86e-03
(1.6, 0.1)	2.12e-06	2.69e-03	9.36e-04	2.03e-03
(1.6, 1.6)	6.52e-05	6.80e-03	1.80e-05	9.39e-03
(1.6, 4.6)	5.66e-05	7.62e-03	6.68e-06	1.35e-02
(4.6, 0.1)	3.30e-05	3.75e-03	1.52e-03	3.14e-03
(4.6, 1.6)	3.48e-05	6.09e-03	8.19e-06	1.42e-02
(4.6, 4.6)	3.50e-05	4.65e-03	6.93e-03	1.40e-02
Average	6.19e-05	6.09e-03	7.65e-04	1.07e-02

Table 5.8: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model with sample size $n = 200$, Kendall's $\tau = 0.25$ (the association parameter $\alpha = 4/3$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	1.83e-08	2.64e-04	3.52e-05	4.56e-05
(0.1, 1.6)	2.03e-05	2.17e-03	8.62e-04	1.69e-03
(0.1, 4.6)	4.23e-05	2.92e-03	1.45e-03	2.53e-03
(1.6, 0.1)	5.75e-06	1.92e-03	7.32e-04	1.91e-03
(1.6, 1.6)	5.35e-05	4.75e-03	2.06e-06	5.34e-03
(1.6, 4.6)	7.54e-05	4.34e-03	1.76e-05	7.29e-03
(4.6, 0.1)	1.75e-05	2.53e-03	1.25e-03	2.81e-03
(4.6, 1.6)	4.21e-05	4.19e-03	1.17e-05	9.32e-03
(4.6, 4.6)	5.51e-05	2.60e-03	3.45e-03	9.07e-03
Average	3.14e-05	3.81e-03	4.42e-04	6.32e-03

Table 5.9: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model with sample size $n = 100$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	2.73e-06	9.39e-04	1.77e-04	2.48e-04
(0.1, 1.6)	1.11e-04	4.55e-03	1.59e-03	2.38e-03
(0.1, 4.6)	1.71e-04	4.95e-03	1.80e-03	2.63e-03
(1.6, 0.1)	3.01e-05	4.13e-03	1.30e-03	3.06e-03
(1.6, 1.6)	5.64e-05	5.87e-03	2.20e-05	9.26e-03
(1.6, 4.6)	8.57e-05	6.28e-03	1.47e-04	1.35e-02
(4.6, 0.1)	7.29e-05	4.52e-03	1.47e-03	3.42e-03
(4.6, 1.6)	1.19e-04	5.51e-03	2.08e-04	1.44e-02
(4.6, 4.6)	6.86e-06	3.45e-03	4.14e-03	1.06e-02
Average	8.04e-05	5.96e-03	1.03e-03	1.16e-02

Table 5.10: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model and sample size $n = 200$, Kendall's $\tau = 0.50$ (the association parameter $\alpha = 2$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	1.88e-05	4.52e-04	1.74e-04	2.33e-04
(0.1, 1.6)	2.41e-05	2.71e-03	1.22e-03	2.77e-03
(0.1, 4.6)	4.13e-05	3.00e-03	1.39e-03	3.02e-03
(1.6, 0.1)	6.11e-06	2.41e-03	1.25e-03	2.78e-03
(1.6, 1.6)	6.26e-06	4.46e-03	6.74e-05	5.35e-03
(1.6, 4.6)	8.68e-06	4.03e-03	3.53e-05	7.66e-03
(4.6, 0.1)	1.57e-05	2.50e-03	1.43e-03	3.02e-03
(4.6, 1.6)	1.84e-06	3.48e-03	7.15e-05	8.50e-03
(4.6, 4.6)	6.64e-06	1.78e-03	1.49e-03	6.11e-03
Average	3.77e-05	3.55e-03	3.96e-04	6.79e-03

Table 5.11: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and the maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model with sample size $n = 100$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	9.19e-05	1.48e-03	7.38e-04	7.74e-04
(0.1, 1.6)	4.23e-04	7.07e-03	1.67e-03	2.96e-03
(0.1, 4.6)	4.83e-04	7.28e-03	1.67e-03	2.96e-03
(1.6, 0.1)	1.71e-04	5.69e-03	1.73e-03	2.94e-03
(1.6, 1.6)	1.99e-05	4.93e-03	1.36e-03	1.07e-02
(1.6, 4.6)	3.36e-04	5.59e-03	3.38e-04	1.34e-02
(4.6, 0.1)	2.08e-04	5.83e-03	1.73e-03	2.95e-03
(4.6, 1.6)	1.56e-04	5.65e-03	2.07e-04	1.44e-02
(4.6, 4.6)	4.16e-04	3.49e-03	2.32e-03	7.72e-03
Average	2.40e-04	6.08e-03	1.64e-03	1.25e-02

Table 5.12: Comparison of Bias^2 and MSE of the proposed sieve maximum likelihood estimation and maximum pseudo likelihood estimation of bivariate Gumbel Copula distributed event times when the underlying distribution is incorrectly specified as bivariate Clayton Copula model with sample size $n = 200$, Kendall's $\tau = 0.75$ (the association parameter $\alpha = 4$)

(T_1, T_2)	Sieve		Pseudo	
	Bias^2	MSE	Bias^2	MSE
(0.1, 0.1)	1.60e-04	7.86e-04	7.29e-04	7.78e-04
(0.1, 1.6)	1.89e-05	2.82e-03	1.49e-03	2.76e-03
(0.1, 4.6)	2.29e-05	2.84e-03	1.49e-03	2.76e-03
(1.6, 0.1)	8.80e-05	3.26e-03	1.55e-03	2.75e-03
(1.6, 1.6)	1.56e-05	4.01e-03	2.01e-04	5.52e-03
(1.6, 4.6)	1.07e-04	3.43e-03	4.83e-07	7.87e-03
(4.6, 0.1)	9.80e-05	3.33e-03	1.55e-03	2.75e-03
(4.6, 1.6)	1.02e-04	3.44e-03	1.36e-04	8.47e-03
(4.6, 4.6)	2.68e-04	1.78e-03	6.09e-04	4.69e-03
Average	8.89e-05	3.43e-03	6.11e-04	7.14e-03

CHAPTER 6
DISCUSSION AND FUTURE WORK

The estimation of the joint distribution function with bivariate current status data is a very challenging problem in survival analysis. There is no sophisticated method available to solve this problem in the literature. In this thesis, we proposed a tensor spline-based sieve nonparametric maximum likelihood method for estimating the joint distribution function with bivariate current status data. This tensor spline-based sieve approach reduces the dimensionality of the nonparametric maximum likelihood estimation problem substantially which makes the nonparametric maximum likelihood estimation tractable numerically. We also show that the proposed tensor spline-based estimator is consistent and converges to the true joint distribution function at a rate of $n^{1/4}$ if the objective joint distribution function is smooth enough. The simulation studies indicate that the finite-sample performance of this proposed sieve maximum likelihood estimation is better than a semiparametric maximum pseudo likelihood approach whether or not the copula model is correctly specified in the maximum pseudo likelihood approach.

Based on the proposed sieve maximum likelihood estimation of the joint distribution function and its two marginal distribution functions, a nonparametric test of the association of two event times can be constructed by summing over the difference between the product of the two estimated marginal distribution functions and the estimated joint distribution function over all the monitoring time points. This test will be robust against the underlying bivariate distribution which, otherwise is required in Wang and Ding's method (2000). However, the asymptotic property of the test statistic still remains a challenging task for future research.

The proposed sieve maximum likelihood estimation approach can be extended to other bivariate censored data such as bivariate right censored data (Dabrowska (1988), Kooperberg (1998)) and bivariate interval censored data case 2 (Betensky and Finkelstein (1999) and Yu, et al. (2000)).

Parallel to bivariate current status data, the association test of the two event times for bivariate right censored and bivariate interval censored data case 2 can also be similarly studied.

REFERENCES

[1] Betensky, R. A. and Finkelstein, D. M. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine*, **18**, 3089-3100.

[2] Curry, H. B. and Schoenberg, I. J. (1966). On Polya frequency functions. IV. The fundamental spline functions and their limits. *Journal d′analyse mathéatique*. **17**, 71-107.

[3] Dabrowska, D. M. (1988). Kaplan-Meier estimate on the Plane. *The Annals of Statistics*, **16**, 1475-1489.

[4] Diamond, I. D. and McDonald, J. W. (1991). Analysis of current status data. In *Demographic Applications of Event History Analysis* (J. Trussell, R. Hankinson and J. Tilton, eds.) 231-252. Oxford Univ. Press.

[5] Diamond, I. D., McDonald, J. W. and Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in Pakistan. *Demography*, **23**, 607-620.

[6] Ding, A. A. and Wang, W. (2004). Testing independence for bivariate current status data. *Journal of the American Statistical Association*, **99**, 145-155.

[7] De Boor, C. (2001). *A Practical Guide to Splines*. New York: Spring-Verlag.

[8] Duffy, D. L., Martin, N. G. and Matthews, J. D. (1990). Appendectomy in australian twins. *The American Journal of Human Genetics*, **47**, 590-592.

[9] Finkelstein, D. M. (1986). A proportional hazards model for interval-cenored failure time data. *Biometrics*, **42**, 845-854.

[10] Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric medel for regression analysis of interval-censored failure time data. *Biometrics*, **41**, 933-945.

[11] Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser, Basel.

[12] Grummer-Strawn, L. M. (1993). Regression analysis of current status data: An application to breat feeding. *Journal of American Statistical Association*, **88**, 758-765.

[13] Hoel, D. G. and Walburg, H. E. (1972). Statistical analysis of survival experiments. *Journal of the National Cancer Institute*, **49**, 361-372.

[14] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, **24**, 540-568.

[15] Huang, J. and Wellner, J. A. (1997). Interval Censored Survival Data: A Review of Recent Progress. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis* (D. Lin and T. Fleming, eds.) 123-169. New York: Springer-Verlag.

[16] Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. Comput. *Computational Statistics and Data Analysis*, **45**, 137-157.

[17] Jewell, N. P., Malani, H. and Vittinghoff, E. (1994). Nonparametric estimation for a form of doubly censored data with application to two problems in AIDS. *Journal of American Statistical Association*, **89**, 7-18.

[18] Jewell, N. P., van der Laan, M. J. and Lei, X. (2005). Bivariate current status data with univariate monitoring times. *Biometrika*, **92**, 847-862.

[19] Kooperberg, C. (1998). Bivariate density estimation with an application to survival analysis. *Journal of Computational and Graphical Statistics*, **7**, 322-341.

[20] Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika*, **94**, 705-18.

[21] Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric Estimation Methods for Panel Count Data Using Monotone B-Splines. *Journal of American Statistical Association*, **104**, 1060-1070.

[22] Nelsen, R. B. (2006). *An Introduction to Copulas*. New York: Springer-Verlag.

[23] Ramsay, J. O. (1988). Monotone regression splines in action. *Statist Science*, **3**, 425-41.

[24] Rosen, J. B. (1960), The Gradient Projection Method for Nonlinear Programming. *Journal of the Society for Industrial and Applied Mathematics*, **8**, 181-217.

[25] Schumaker, L. (1981). *Spline Functions: Basic Theory*. New York: Wiley.

[26] Shen, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika*. **85**, 165-77.

[27] Shen, X. and Wong, W. H. (1994), Convergence Rate of Sieve Estimates. *The Annals of Statistics*. **22**, 580-615.

[28] Shiboski, S. C. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *Journal of American Statistical Association*, **87**, 360-372.

[29] van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

[30] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

[31] Wang, W. and Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika*, **87**, 879-893.

[32] Wellner, J. A. and Song, S. (2002). An upper bound for uniform entropy numbers. *Technical Report, Department of Biostatistics, The University of Washington.*

[33] Wellner, J. A. and Zhang, Y. (2007). Likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics.* **28**, 2106-2142.

[34] Yu, Q., Wong, G. Y. C. and He, Q. (2000). Estimation of a joint distribution function with multivariate interval-censored data when the nonparametric MLE is not unique. *Biometrical Journal*, **6**, 747-763.

[35] Zhang, S. (2008). Inference on the association measure for bivariate survival data with hybrid censoring and applications to an HIV study. *Ph.D Dissertation, Department of Biostatistics, The University of Iowa.*

[36] Zhang, Y., Huang, J. and Hua, L. (2009). An Adaptive Spline-Based Sieve Semiparametric Maximum Likelihood Estimation for the Cox Model with Interval-Censored Data. *Technical Report, Department of Biostatistics, The University of Iowa.*

APPENDIX

(i) The explicit form of the cubic I-spline is given as

$$I_i^3(x) = \begin{cases} 0, & x < t_i \\ \frac{(x-t_i)^3}{(t_{i+1}-t_i)(t_{i+2}-t_i)(t_{i+3}-t_i)}, & t_i \le x < t_{i+1} \\ \frac{(t_{i+1}-t_i)^2}{(t_{i+2}-t_i)(t_{i+3}-t_i)} & \\ +\frac{-(x^3-t_{i+1}^3)+\frac{3}{2}(t_i+t_{i+2})(x^2-t_{i+1}^2)-3t_it_{i+2}(x-t_{i+1})}{(t_{i+2}-t_{i+1})(t_{i+2}-t_i)(t_{i+3}-t_i)} & \\ +\frac{-(x^3-t_{i+1}^3)+\frac{3}{2}(t_{i+1}+t_{i+3})(x^2-t_{i+1}^2)-3t_{i+1}t_{i+3}(x-t_{i+1})}{(t_{i+2}-t_{i+1})(t_{i+3}-t_{i+1})(t_{i+3}-t_i)}, & t_{i+1} \le x < t_{i+2} \\ 1-\frac{(t_{i+3}-x)^3}{(t_{i+3}-t_{i+2})(t_{i+3}-t_{i+1})(t_{i+3}-t_i)}, & t_{i+2} \le x < t_{i+3} \\ 1, & x \ge t_{i+3}. \end{cases}$$

(ii) The following steps are used to construct cubic I-splines through B-splines in statistics package R, higher order I-splines can be constructed similarly.

(1) Loading *spline* package in R.

(2) Inputting knot sequence $knot = \{u_i\}_1^{p+4}$, and value x which is inside $knot$.

(3) Obtaining the value of $I_m^3(x)$ with $m = 1, \cdots, p$ by the follow function.

```
isp3 = function(m, x)
{
    a = splineDesign(knot, x)
    res = 0
    for(i in m : p )
        res = res + a[i]
    res
},
```

where $splineDesign(knot, x)$ returns the value of the vector

$(N_1^4(x), N_2^4(x), \ldots, N_p^4(x))$.