Partially Monotone Tensor Spline Estimation of the Joint Distribution Function with Bivariate Current Status Data

Yuan Wu

Department of Epidemiology

University of Iowa

Iowa city, IA 52242

Ying Zhang

Department of Biostatistics

University of Iowa

Iowa City, IA 52242

Summary

The analysis of the joint distribution function with bivariate event time data is a challenging problem both theoretically and numerically. This paper develops a tensor spline-based sieve maximum likelihood estimation method to estimate the joint distribution function with bivariate current status data. The I-spline basis functions are used in approximating the joint distribution function in order to simplify the numerical computation of constrained maximum likelihood estimation problem. The generalized gradient projection algorithm is used to compute the constrained optimization problem. The proposed tensor spline-based nonparametric sieve maximum likelihood estimator is shown to be consistent and the rate of convergence can be as good as $n^{1/4}$ under some regularity conditions. The simulation studies with moderate sample sizes are carried out to demonstrate that the finite sample performance of the proposed estimator is generally satisfactory.

Key Words: Bivariate current status data, Constrained maximum likelihood estimation, Empirical process, Sieve maximum likelihood estimation, Tensor spline basis functions

1 Introduction

In some applications, observation of 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 1. Current status data arises naturally in many applications, see for example, in animal tumorigenicity experiments by Hoel and Walburg (1972), and Finkelstein and Wolfe (1985); in social demographic studies of the distribution of the age at weaning by Diamond, McDonald and Shah (1986), Diamond and McDonald (1991), and Grummer-Strawn (1993); and in studies of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) by Shiboski and Jewell (1992), and Jewell, Malani and Vittinghoff (1994).

The univariate current status data has been thoroughly studied in literatures. Groenoboom and Wellner (1992) and Huang and Wellner (1995) studied the asymptotic properties of the nonparametrc maximum likelihood estimator (NPMLE) of the distribution function with current status data. Huang (1996) considered Cox proportional hazards model with current status data and showed that the maximum likelihood estimator (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.

Bivariate event time data occurs in many applications as well. For example, in an Australian twin study (Duffy, Martin and Matthews, 1990), the researchers were interested in times to a certain event such as a disease or a disease-related symptom in both twins. NPMLE of the joint distribution function of the correlated event times with bivariate right censored data was studied by Dabrowska (1988), Prentice and Cai (1992), Pruitt (1991). van der Laan (1996) and Quale, van der Laan and Robins (2006). As an alternative, Kooperberg (1998) developed a tensor spline estimation of the logarithm of joint density function with bivariate right censored data. Shih and Louis (1995) proposed a two-stage semiparametric estimation procedure for the association parameter for bivariate right censored data, in which the joint distribution of the two event times 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.

For bivariate interval censored data, a nonparametric maximum likelihood estimation method can be generalized from the univariate case. For the NPMLE, one needs to design an efficient searching algorithm for the non-zero mass intersection rectangles (Betensky and Finkelstein, 1999; Wong and Yu, 1999; Gentleman and Vandal, 2001; Maathuis, 2005). Sun, Wang and Sun (2006) adopted the same idea used by Shih and Louis (1995) and proposed a two-stage method to estimate the association parameter in Copula models for bivariate interval censored data.

This paper studies bivariate current status data, a special type of bivariate interval censored data. This data structure arises in the studies of two diseases in same subject or a common disease in two correlated subjects. 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, the observation of bivariate current status data consists of

$$X = (C_1, C_2, \Delta_1 = I(T_1 \leq C_1), \Delta_2 = I(T_2 \leq C_2)), \tag{1.1}$$

where $I(\cdot)$ is the indicator function. Wang and Ding (2000) studied whether or not the onsets of hypertension and diabetes are correlated for people in Taiwan. They adopted the same idea used by Shih and Louis (1995) and Sun, Wang and Sun (2006) and proposed a two-stage estimation of the association parameter of two event times with bivariate current status data. This two-stage method facilitates an easy estimator of the joint distribution function through Copula model as a by-product and is the only available method in literatures to estimate the joint distribution function with bivariate current status data. In a study on HIV transmission, Jewell, van der Laan and Lei (2005) investigated 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 the marginal distribution functions. For 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 estimable. This paper proposes a tensor spline-based sieve maximum likelihood estimation of the joint distribution function with bivariate current status data in a general scenario in which C_1 and C_2 are allowed to be different and hence the method is more applicable in practice.

The rest of the paper is organized as follows. Section 2 characterizes the spline-based sieve MLE $\hat{\tau}_n = (\hat{F}_n, \hat{F}_{n,1}, \hat{F}_{n,2})$, where $\hat{F}_n$ is the tensor spline-based estimator of the joint distribution function, $\hat{F}_{n,1}$ and $\hat{F}_{n,2}$ are the spline-based estimators of the two corresponding marginal distribution functions. Section 3 presents two asymptotic properties (consistency and convergence rate) of the proposed spline-based sieve MLE. Section 4 discusses the computation of the spline-based estimator. Section 5 carries out a set of simulation studies to examine the finite sample performance of the proposed method and compare the proposed method with the method extended from Wang and Ding (2000)'s idea. Section 6 summarizes our findings and discusses some related problems. Section 7 provides the proofs of the lemmas and theorems stated in the early sections. Finally, some technical lemmas required by the proofs of the asymptotic properties are developed in Section 8.

2 Tensor Spline-based Sieve Maximum Likelihood Estimation Method

2.1 Spline-based Maximum Likelihood Estimation

Consider a sample of n i.i.d. bivariate current status data (1.1), $\{(c_{1,k}, \delta_{1,k}, c_{2,k}, \delta_{2,k}) : k = 1, 2, \cdots, n\}$. Suppose that (T_1, T_2) and (C_1, C_2) are independent and (C_1, C_2) are non-informative to (T_1, T_2). Then the log-likelihood for the observed data can be expressed by

$$
\begin{aligned}
l_n(\cdot;\text{data}) = \sum_{k=1}^{n}\{&\delta_{1,k}\delta_{2,k}\log P(T_1 \le c_{1,k}, T_2 \le c_{2,k}) \\
&+ \delta_{1,k}(1-\delta_{2,k})\log P(T_1 \le c_{1,k}, T_2 > c_{2,k}) \\
&+ (1-\delta_{1,k})\delta_{2,k}\log P(T_1 > c_{1,k}, T_2 \le c_{2,k}) \\
&+ (1-\delta_{1,k})(1-\delta_{2,k})\log P(T_1 > c_{1,k}, T_2 > c_{2,k})\}.
\end{aligned} \tag{2.1}
$$

Denote F the joint distribution function of event times (T_1, T_2) and F_1 and F_2 the marginal distribution functions of F, respectively, the log-likelihood (2.1) can be rewritten as

$$
\begin{aligned}
l_n(F, F_1, F_2;\text{data}) = \sum_{k=1}^{n}\{&\delta_{1,k}\delta_{2,k}\log F(c_{1,k}, c_{2,k}) \\
&+ \delta_{1,k}(1-\delta_{2,k})\log(F_1(c_{1,k}) - F(c_{1,k}, c_{2,k})) \\
&+ (1-\delta_{1,k})\delta_{2,k}\log(F_2(c_{2,k}) - F(c_{1,k}, c_{2,k})) \\
&+ (1-\delta_{1,k})(1-\delta_{2,k})\log(1 - F_1(c_{1,k}) - F_2(c_{2,k}) \\
&+ F(c_{1,k}, c_{2,k}))\}.
\end{aligned} \tag{2.2}
$$

A class of real-valued functions is defined in a bounded region $[L_1, U_1] \times [L_2, U_2]$ as

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

where F, F_1 and F_2 satisfy the following conditions in (2.3):

$$
\begin{aligned}
& 0 \leq F(s,t), \\
& F(s',t) \leq F(s'',t), \\
& F(s,t') \leq F(s,t''), \\
& [F(s'',t'') - F(s',t'')] - [(F(s'',t') - F(s',t')] \geq 0, \\
& F_1(s) - F(s,t) \geq 0 \\
& F_2(t) - F(s,t) \geq 0, \\
& [F_1(s'') - F_1(s')] - [F(s'',t) - F(s',t)] \geq 0, \\
& [F_2(t'') - F_2(t')] - [F(s,t'') - F(s,t')] \geq 0, \\
& [1 - F_1(s)] - [F_2(t) - F(s,t)] \geq 0,
\end{aligned}
\tag{2.3}
$$

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

It can be easily argued that if F is a joint distribution function and F_1 and F_2 are its two corresponding marginal distribution functions, $(F, F_1, F_2) \in \mathcal{F}$. Throughout this paper, $F_0, F_{0,1}$ and $F_{0,2}$ are denoted for the true joint and marginal distribution functions, respectively. Hence the NPMLE of $(F_0, F_{0,1}, F_{0,2})$ is defined as

$$
(\hat{F}_n, \hat{F}_{n,1}, \hat{F}_{n,2}) = \arg\max_{(F,F_1,F_2)\in\mathcal{F}} l_n(F, F_1, F_2; \text{data}). \tag{2.4}
$$

The conventional NPMLE maximizes (2.2) 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 $k = 1, \ldots, n$. The study of the conventional NPMLE with bivariate current status data is both numerically and theoretically challenging. Compare the NPMLE with its univariate counterpart developed by Groenoboom and Wellner (1992), the computation of the NPMLE for Problem (2.4) is much more involved in view of the constraints of $\mathcal{F}$ given in (2.3). The NPMLE method adopted by Maathius (2005) may be applied to Problem (2.4), but this type of methods may not necessarily produce a unique NPMLE as pointed out by Maathius (2005). While the asymptotic properties of the NPMLE with univariate current status data were thoroughly investigated by Groenoboom and Wellner (1992) and Huang and Wellner (1995), they are much harder to study for bivariate current status data, mainly due to the difficulty in evaluating the entropy of $\mathcal{F}$ (Song and Wellner, 2002).

To overcome the difficulties in Problem (2.4), the spline-based sieve maximum likelihood estimation procedure is proposed. The main idea of the spline-based sieve method is to solve Problem (2.4) in a subclass of $\mathcal{F}$ but "approximating" to $\mathcal{F}$ asymptotically, with the advantage that the estimator to be found in this subclass is easy to compute and analyze. The univariate spline-based sieve MLEs for various models were developed by Shen (1998), Lu, Zhang and Huang (2007, 2009) and Zhang, Hua and Huang (2010). In terms of estimating bivariate functions, the tensor spline (De Boor, 2001) estimation has been studied by Stone (1994) in nonparametric regression setting, by Koo (1996) and Scott (1992) in estimating a multivariate density function without censoring, and as noted in Section 1, by Kooperberg (1998) in estimating the density function of bivariate event times subject to right censoring. Recently, an application of the tensor B-spline estimation of a bivariate monotone function has also been investigated by Wang and Taylor (2004) in a biomedical study.

In this paper, we propose a partially monotone tensor spline estimation of the bivariate distribution function. To solve Problem (2.4), the unknown joint distribution function is estimated by a linear combination of the tensor spline basis functions and its two marginal distribution function are also independently estimated by linear combinations of spline basis functions in the same way given by Lu, Zhang and Huang (2007, 2009) and Zhang, Hua and Huang (2010). Then maximizing the log likelihood with respect to the unknown functions converts to maximizing the sieve log likelihood with respect to the unknown spline coefficients subjecting to corresponding inequality constraints.

2.2 B-spline-based Estimation

In this section, the spline-based sieve maximum likelihood estimation problem is reformulated as a constrained optimization problem with respect to the coefficients of the B-spline basis functions.

Suppose two sets of the normalized B-spline basis functions of order l (Schumaker, 1981), $\{N_i^{(1),l}(s)\}_{i=1}^{p_n}$ and $\{N_j^{(2),l}(t)\}_{j=1}^{q_n}$ are constructed in $[L_1, U_1] \times [L_2, U_2]$ with the knot sequence $\{u_i\}_{i=1}^{p_n+l}$ 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 $\{v_j\}_{j=1}^{q_n+l}$ 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 $0 < v < 1$.

Define

$$\Omega_n = \{\tau_n = (F_n, F_{n,1}, F_{n,2}) : F_n(s,t) = \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t),$$

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

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

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 (2.5)$\}$,

$$\begin{aligned}
&\alpha_{1,1} \geq 0,\\
&\alpha_{1,j+1} - \alpha_{1,j} \geq 0 \text{ for } j = 1, \ldots, q_n - 1,\\
&\alpha_{i+1,1} - \alpha_{i,1} \geq 0 \text{ for } i = 1, \ldots, p_n - 1,\\
&(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1, j = 1, \ldots, q_n - 1,\\
&\beta_1 - \alpha_{1,q_n} \geq 0,\\
&(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1,\\
&\gamma_1 - \alpha_{p_n,1} \geq 0,\\
&(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \text{ for } j = 1, \ldots, q_n - 1,\\
&\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.
\end{aligned} \tag{2.5}$$

To obtain the tensor B-spline-based sieve likelihood with bivariate current status data, $(F, F_1, F_2) = (F_n, F_{n,1}, F_{n,2}) = \tau_n \in \Omega_n$ is substituted into (2.2) to result in

$$
\begin{aligned}
\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data}) = \sum_{k=1}^{n} \Bigg\{ & \delta_{1,k}\delta_{2,k} \log \sum_{i=1}^{p_n} \sum_{j=2}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \\
& + \delta_{1,k}(1-\delta_{2,k}) \log \Bigg\{ \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \Bigg\} \\
& + (1-\delta_{1,k})\delta_{2,k} \log \Bigg\{ \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k}) \\
& - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \Bigg\} \\
& + (1-\delta_{1,k})(1-\delta_{2,k}) \log \Bigg\{ 1 - \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \\
& - \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k}) + \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \Bigg\} \Bigg\} .
\end{aligned} \tag{2.6}
$$

Hence, the proposed sieve MLE with the B-spline basis functions is the maximizer of (2.6) over Ω_n.

Lemma 2.1. *Class $\Omega_n \subset \mathcal{F}$.*

Remark 2.1. *Lemma 2.1 implies that the spline-based sieve MLE in Ω_n is the MLE in a sub-class of $\mathcal{F}$. The spline-based sieve MLE may have good asymptotic properties if this sub-class "approximates" to $\mathcal{F}$ as $n \to \infty$.*

3 Asymptotic Properties

In this section, we describe the asymptotic properties of the tensor spline-based sieve MLE of the joint distribution function with bivariate current status data. The study of asymptotic properties of the proposed sieve estimator requires some regularity conditions, regarding the event times, observation times and the choice of the knot sequences . The following conditions sufficiently guarantee the results in the forthcoming theorems.

Regularity Conditions:

C1. Both $\frac{\partial F_0(s,t)}{\partial s}$ and $\frac{\partial F_0(s,t)}{\partial t}$ have positive lower bounds in $[L_1, U_1] \times [L_2, U_2]$.

C2. $\frac{\partial^2 F_0(s,t)}{\partial s \partial t}$ has a 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, $\nabla_m^p F_0 = \frac{\partial^p F_0(s,t)}{\partial s^m t^{p-m}}$ for $m = 1, 2, \ldots, p$, in $[L_1, U_1] \times [L_2, U_2]$; $F_{0,1}(s)$ 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. The observation times (C_1, C_2) follow a bivariate distribution only 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)'s distribution has a positive lower bound in $[l_1, u_1] \times [l_2, u_2]$.

C6. Knot sequences $\{u_i\}_{i=1}^{p_n+l}$ and $\{v_j\}_{j=1}^{q_n+l}$ of the B-spline basis functions $\{N_i^{(1),l}\}_{i=1}^{p_n}$ and $\{N_j^{(2),l}\}_{j=1}^{q_n}$, respectively, satisfy that both $\frac{\min_i \Delta_i^{(u)}}{\max_i \Delta_i^{(u)}}$ and $\frac{\min_j \Delta_j^{(v)}}{\max_j \Delta_j^{(v)}}$ 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 $j = l, \ldots, q_n$.

Remark 3.1. *C1 implies that* $\frac{dF_{0,1}(s)}{ds}$ *and* $\frac{dF_{0,2}(t)}{dt}$ *have positive lower bounds on* $[L_1, U_1]$ *and* $[L_2, U_2]$*, respectively. C3 implies that both* $\frac{\partial F_0(s,t)}{\partial s}$ *and* $\frac{\partial F_0(s,t)}{\partial t}$ *have positive upper bounds in* $[L_1, U_1] \times [L_2, U_2]$*;* $\frac{dF_{0,1}(s)}{ds}$ *and* $\frac{dF_{0,2}(t)}{dt}$ *have positive upper bounds on* $[L_1, U_1]$ *and* $[L_2, U_2]$*, respectively.*

Let

$$\begin{aligned}\Omega_{n,1} = \{\tau = (F_n, F_{n,1}, F_{n,2}) : & F_n(s,t) = \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t), \\ & F_{n,1}(s) = \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s), \\ & F_{n,2}(t) = \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(t), \\ & \text{with } \underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{p_n,q_n}), \underline{\beta} = (\beta_1, \cdots, \beta_{p_n}), \\ & \text{ and } \underline{\gamma} = (\gamma_1, \cdots, \gamma_{q_n}) \\ & \text{subject to the following conditions in (3.1)}\},\end{aligned}$$

$$\begin{aligned}& \alpha_{1,1} \geq 0, \\ & \alpha_{1,j+1} - \alpha_{1,j} \geq 0 \text{ for } j = 1, \ldots, q_n - 1, \\ & \alpha_{i+1,1} - \alpha_{i,1} \geq 0 \text{ for } i = 1, \ldots, p_n - 1, \\ & (\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq \frac{b_0 \min_{i_1 : l \leq i_1 \leq p_n} \Delta_{i_1}^{(u)} \min_{j_1 : l \leq j_1 \leq q_n} \Delta_{j_1}^{(v)}}{l^2} \\ & \text{for } i = 1, \ldots, p_n - 1, j = 1, \ldots, q_n - 1, \\ & \beta_1 - \alpha_{1,q_n} \geq 0, \\ & (\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \text{ for } i = 1, \ldots, p_n - 1, \\ & \gamma_1 - \alpha_{p_n,1} \geq 0, \\ & (\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \text{ for } j = 1, \ldots, q_n - 1, \\ & \beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.\end{aligned} \tag{3.1}$$

Remark 3.2. *Note that $\Omega_{n,1}$ is a sub-class of Ω_n. We propose to find the estimator in*

$\Omega_{n,1}$ mainly due to the technique convenience in justifying the asymptotic properties. In computation, the relaxation parameter b_0 can be chosen small enough that would not result in a different estimator from the one found in Ω_n defined in section 2.

We study the asymptotic properties in the feasible region for observation times: $[l_1, u_1] \times [l_2, u_2]$. Let $\Omega_n^{'} = \{\tau_n(s,t) : \tau \in \Omega_{n,1}, \text{ for } (s,t) \in [l_1, u_1] \times [l_2, u_2]\}$ and let $\tau_0(s,t) = (F_0(s,t), F_{0,1}(s), F_{0,2}(t))$ with $(s,t) \in [l_1, u_1] \times [l_2, u_2]$. Under C4, the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over $\Omega_{n,1}$ is actually the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over $\Omega_n^{'}$. Throughout the study of the asymptotic properties, we denote $\hat{\tau}_n$ as the maximizer of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over $\Omega_n^{'}$.

Suppose the $L_r(Q)$-norm associated with probability measure Q is denoted by $\|f\|_{L_r(Q)} = (Q|f|^r)^{1/r} = (\int |f|^r dQ)^{1/r}$. In the following, the $L_r(P_{C_1,C_2})$-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 the 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 the L_2-norms, the distance between $\tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}$ and $\tau_0 = (F_0, F_{0,1}, F_{0,2})$ is defined as

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

Theorem 3.1. *Suppose C2-C6 hold, and $p_n = O(n^v)$, $q_n = O(n^v)$ for $v < 1$, that is, the numbers of interior knots of knot sequences $\{u_i\}_1^{p_n+l}$ and $\{v_j\}_1^{q_n+l}$ are both in the order of n^v for $v < 1$. Then*

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

Theorem 3.2. *Suppose C1-C6 hold, and* $p_n = O(n^v)$, $q_n = O(n^v)$ *for* $v \le \frac{1}{4p}$*, that is, the numbers of interior knots of knot sequences* $\{u_i\}_1^{p_n+l}$ *and* $\{v_j\}_1^{q_n+l}$ *are both in the order of* n^v *for* $v \le \frac{1}{4p}$*. Then*

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

4 Computation of the Spline-Based Sieve MLE

We propose to compute the sieve MLE using I-splines for which the I-spline basis functions are defined by

$$I_i^l(s) = \begin{cases} 0, \ i > j, \\ \sum_{m=i}^{j}(u_{m+l+1} - u_m)M_m^{l+1}(s)/(l+1), \ j-l+1 \le i \le j, \\ 1, \ i < j-l+1, \end{cases} \tag{4.1}$$

for $u_j \le s < u_{j+1}$, where M_m^ls are the M-spline basis functions of order l studied by Curry and Schoenberg (1966) and can be calculated recursively by

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

By the relationship between the B-spline basis functions and the M-spline basis functions (Schumaker, 1981), it can be easily argued that the I-spline basis function defined by (4.1) can be expressed by a summation of the B-spline basis functions, that is

$$I_i^{l-1}(s) = \sum_{m=i}^{p_n} N_m^l(s). \tag{4.2}$$

Therefore, the spline-based sieve estimation can be re-parameterized by the I-spline basis functions. Let

$$\begin{aligned}
\Theta_n = \{\tau_n = (F_n, F_{n,1}, F_{n,2}) : F_n(s,t) = \sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(s) I_j^{(2),l-1}(t),\\
F_{n,1}(s) = \sum_{i=1}^{p_n}\{\sum_{j=1}^{q_n} \eta_{i,j} + \omega_i\} I_i^{(1),l-1}(s),\\
F_{n,2}(t) = \sum_{j=1}^{q_n}\{\sum_{i=1}^{p_n} \eta_{i,j} + \pi_j\} I_j^{(2),l-1}(t)\\
\text{with } \underline{\eta} = (\eta_{1,1}, \cdots, \eta_{p_n,q_n}), \underline{\omega} = (\omega_1, \cdots, \omega_{p_n}), \text{ and } \underline{\pi} = (\pi_1, \cdots, \pi_{q_n})\\
\text{subject to the following conditions in (4.3)}\},
\end{aligned}$$

$$\begin{aligned}
&\eta_{i,j} \geq 0 \text{ for } i = 1, \cdots, p_n, j = 1, \cdots, q_n,\\
&\omega_i \geq 0, i = 1, \ldots, p_n,\\
&\pi_j \geq 0, j = 1, \ldots, q_n,\\
&\sum_{i=1}^{p_n}\sum_{j=1}^{q_n} \eta_{i,j} + \sum_{i=1}^{p_n} \omega_i + \sum_{j=1}^{q_n} \pi_j \leq 1.
\end{aligned} \tag{4.3}$$

Then the log likelihood with the I-spline basis functions is given by

$$
\begin{aligned}
\tilde{l}_n(\underline{\eta}, \underline{\omega}, \underline{\pi}; \cdot) = \sum_{k=1}^{n} \Bigg\{ & \delta_{1,k}\delta_{2,k} \log \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \\
& + \delta_{1,k}(1-\delta_{2,k}) \log \left\{ \sum_{i=1}^{p_n} \left[\sum_{j=1}^{q_n} \eta_{i,j} + \omega_i \right] I_i^{(1),l-1}(C_{1,k}) \right. \\
& \left. - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \right\} \\
& + (1-\delta_{1,k})\delta_{2,k} \log \left\{ \sum_{j=1}^{q_n} \left[\sum_{i=1}^{p_n} \eta_{i,j} + \pi_j \right] I_j^{(2),l-1}(c_{2,k}) \right. \\
& \left. - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \right\} \\
& + (1-\delta_{1,k})(1-\delta_{2,k}) \log \left\{ 1 - \sum_{i=1}^{p_n} \left[\sum_{j=1}^{q_n} \eta_{i,j} + \omega_i \right] I_i^{(1),l-1}(c_{1,k}) \right. \\
& - \sum_{j=1}^{q_n} \left[\sum_{i=1}^{p_n} \eta_{i,j} + \pi_j \right] I_j^{(2),l-1}(c_{2,k}) \\
& \left. + \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(c_{1,k}) I_j^{(2),l-1}(c_{2,k}) \right\} \Bigg\}.
\end{aligned}
\tag{4.4}
$$

and the spline-based sieve MLE maximizes (4.4) over Θ_n.

Lemma 4.1. *Class Θ_n is equivalent to Ω_n .*

Remark 4.1. *Lemma 2 indicates that the I-spline-based sieve MLE is the same as the B-spline-based sieve MLE and it is advocated in numerical computation due to the simplicity of the constraints in class Θ_n.*

Given p_n and q_n, the proposed sieve estimation problem described above is actually a restricted parametric maximum likelihood estimation problem with respect to the coefficients of the I-spline and the tensor I-spline basis functions. Jamshidian (2004) generalized the gradient projection algorithm originally proposed by Rosen (1960) using a weighted L_2

norm $\|x\| = x'Wx$ with a positive definite matrix W for the restricted maximum likelihood estimation problems. Lu, Zhang and Huang (2007, 2009) and Zhang, Hua and Huang (2010) implemented the generalized gradient projection algorithm for the spline-based sieve maximum likelihood estimation problem with panel count data and interval censored data, respectively. The algorithm adopted by Zhang, Hua and Huang (2010) is modified to compute the proposed tensor I-spline-based sieve estimator.

Let $\dot{\tilde{l}}(\theta)$ and $H(\theta)$ be the gradient and Hessian matrix of the log likelihood given by (4.4) 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. Note that $H(\theta)$ may not be negative definite for every θ. During the numerical iterations, if $H(\theta)$ is negative definite, we use $W = -H(\theta)$; otherwise use $W = -H(\theta) + \gamma I$, where I is identity matrix and $\gamma > 0$ is chosen sufficiently large to guarantee W being 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, for $j = 1, 2, \cdots, r$,

(i) if $i_j \leq p_n \cdot q_n + p_n + q_n$, 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 $\mathcal{A}$ could have the following form,

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

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

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

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

Step 2. (*Forcing the updated θ to fulfill 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 adding 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 Step 1. otherwise compute $\lambda = (AW^{-1}A^T)^{-1}AW^{-1}\dot{\tilde{l}}(\theta)$.

(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 Step 1.

5 Simulation Studies

Copula models are often used in studying bivariate event time data (Shih and Louis, 1995; Wang and Ding, 2000; Sun, Wang and Sun, 2006; Zhang, Zhang, Chaloner, and Stapleton, 2010)

We consider bivariate Clayton Copula function

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

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

In the simulation studies, We compare the proposed sieve MLE to the semiparametric maximum pseudo-likelihood estimator based on the method studied by Wang and Ding (2000) under Clayton Copula model for the finite sample performance. The semiparametric maximum pseudo-likelihood estimator of the bivariate distribution function is constructed as follows: First, the NPMLEs of the two marginal distribution functions are computed using Convex Minorant Algorithm (Gnoeneboom and Wellner, 1992) and the association parameter α is estimated by the maximum pseudo-likelihood method. Then, the NPMLEs of the marginal distribution functions and the maximum pseudo-likelihood estimator of the association parameter are plugged into the Clayton Copula model to form the semiparametric maximum pseudo-likelihood estimator of the joint distribution function.

The proposed sieve MLE and the semiparametric maximum pseudo-likelihood estimator are evaluated with various combinations of Kendall's τ ($\tau = 0.25, 0.75$) and sample sizes ($n = 100, 200$). Under each of the four settings, the Monte-Carlo simulation with 500 repetitions is conducted and the cubic (l=4) I-spline basis functions are used in the proposed sieve estimation method. The event times (T_1, T_2), monitoring times (C_1, C_2), and the knots selection of the cubic I-spline basis functions are specified as follows:

(i) (*Event times*) (T_1, T_2) are generated from the Clayton copula with the two marginal distributions being exponential with the rate parameter 0.5. Under this setting, $Pr(T_i \geq 5) < 0.1$ for $i = 1, 2$ and $[L_1, U_1] \times [L_2, U_2]$ is chosen to be $[0, 5] \times [0, 5]$.

(ii) (*Censoring times*) Both C_1 and C_2 are generated independently from the uniform distribution on $[0.0201, 4.7698]$ ($Pr(0 < T_i < 0.0201) = Pr(4.7698 < T_i < 5) = 0.01$, for

$i = 1, 2$). The observation region $[l_1, u_1] \times [l_2, u_2]$=$[0.0201, 4.7698] \times [0.0201, 4.7698]$ is inside $[0, 5] \times [0, 5]$ and the distribution functions are bounded away from 0 and 1 inside the observation region.

(iii) (*Knots selection*) Theorem 3.2 implies that the proposed sieve estimator converges at a rate not faster than $n^{1/4}$, and the rate of convergence 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 number of knots is mainly of interest for the asymptotic properties when n is very large. In practice, for the number of interior knots m_n, m_n+1 is often chosen as the closest integer to $n^{1/3}$ that was used by Lu, Zhang and Huang (2007, 2009) and Zhang, Hua and Huang (2010). For moderate sample sizes, say $n = 100, 200$, our experiments show that such m_n is a reasonable choice for the number of interior knots and hence the number of spline basis functions is determined by $p_n = q_n = m_n + 4$ in our computation. Therefore, we choose 4 and 5 as the numbers of interior knots for sample size 100 and 200, respectively. Two end knots of all knot sequences are chosen to be 0 and 5. For each sample of bivariate observation times (C_1, C_2), the interior knots of $\{I_i^{(1),3}\}_{i=1}^{p_n}$ and $\{I_j^{(2),3}\}_{j=1}^{q_n}$ are allocated at the $k/(m_n+1)$ quantiles, $k = 1, \ldots, m_n$ of the sample of C_1 and the sample of C_2, respectively.

Table 1 and 2 display the estimation biases (*Bias*) and the square roots of the mean square errors (*MSE*$^{1/2}$) from the Monte-Carlo simulation of 500 repetitions for both the proposed sieve MLE and the semiparametric maximum pseudo-likelihood estimator of the bivariate distribution function at the 9 pairs of time points (s_1, s_2) with different sample sizes and different values of Kendall's τ. Table 3 calculates the average estimation bias and the average square root of the mean square error for 2209 values of (s_1, s_2) with both s_1 and s_2 uniformly taking 47 values from 0.1 to 4.7. It appears that the bias and the mean square error of the proposed sieve MLE may be a little larger at some points near the boundary than its counterpart, it outperforms its counterpart because of smaller overall bias and mean square error (Table 3). It is also noted that mean square error of the proposed sieve MLE noticeably decreases as sample size increases from 100 to 200.

For sample size $n = 200$, the estimation biases of the joint distribution function from the same Monte-Carlo simulation for both estimation methods are graphically presented in Figure 1 and 2 with Kendall's $\tau = 0.25$ and 0.75, respectively. These figures clearly indicate that the bias of the proposed sieve MLE is noticeably smaller than that of the semiparametric maximum pseudo-likelihood estimation inside the closed region $[0.1, 4.7] \times [0.1, 4.7]$, but the bias of the proposed sieve MLE near the origin increases as Kendall's τ increases. As the by-product of the estimation methods, the estimates of the marginal distribution function of T_1 from the same Monte-Carlo simulation for both the proposed sieve MLE (*Sieve*) and the NPMLE using Convex Minorant Algorithm (*Nonparametric*) are also computed and plotted in Figure 3 along with the true marginal distribution function (*True*). Figure 3 clearly indicates that the bias of the proposed sieve MLE for the marginal distribution function is generally smaller than that of the NPMLE, particularly near the two end points of interval $[0.1, 4.7]$.

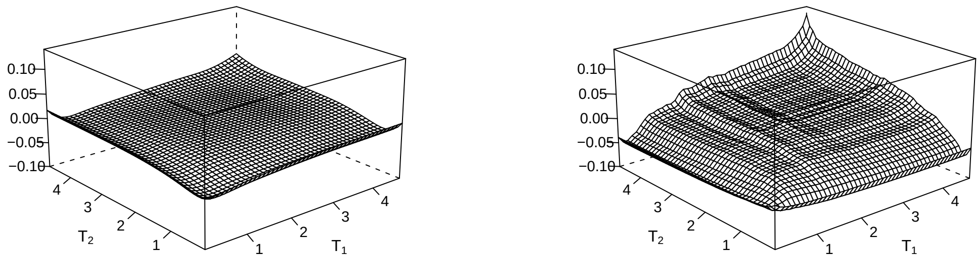

Figure 1: Comparison of the bias between the proposed spline-based sieve estimator (left) and the semiparametric maximum pseudo-likelihood estimator (right) for the joint distribution function when sample size $n = 200$, Kendall's $\tau = 0.25$

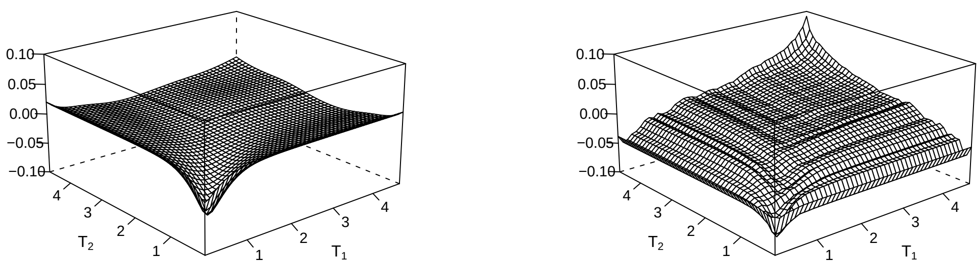

Figure 2: Comparison of the bias between the proposed spline-based sieve estimator (left) and the semiparametric maximum pseudo-likelihood estimator (right) for the joint distribution function when sample size $n = 200$, Kendall's $\tau = 0.75$

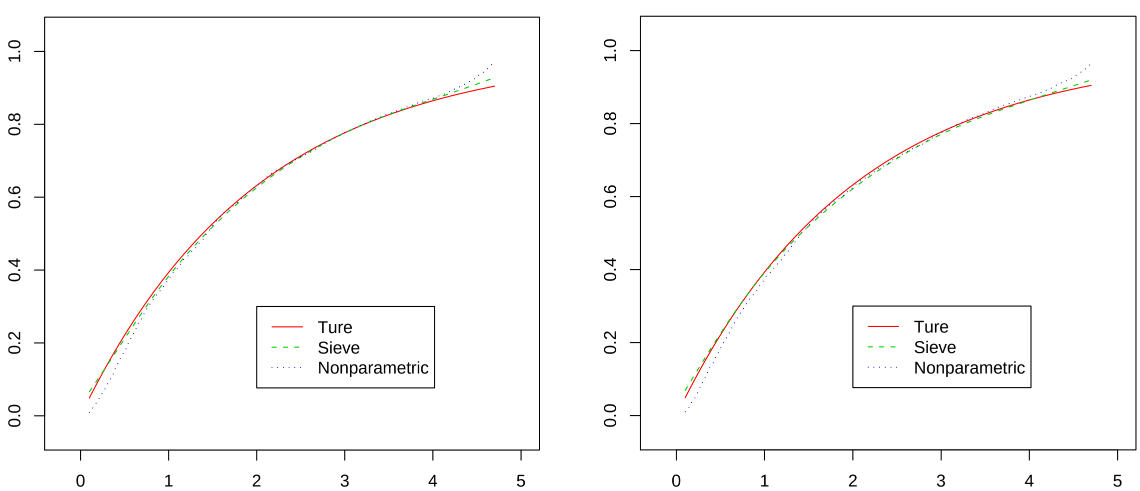

Figure 3: Comparisons of the estimated marginal distributions of T_1 between the proposed spline-based sieve estimation method and the nonparametric maximum likelihood method when sample size $n = 200$ (left: kendall's $\tau = 0.25$; right: kendall's $\tau = 0.75$)

Table 1: Comparisons of the pointwise bias and square root of mean square error between the proposed spline-based sieve estimator and the semiparametric maximum pseudo-likelihood estimator when Kendall's $\tau = 0.25$

Sample Size $n = 100$									
T_1	T_2								
	0.1			1.6			4.6		
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
0.1	Bias	5.47e-3	-1.90e-2	Bias	1.99e-2	-3.69e-2	Bias	3.11e-2	-3.95e-2
	$MSE^{1/2}$	3.21e-2	1.90e-2	$MSE^{1/2}$	6.74e-2	4.86e-2	$MSE^{1/2}$	7.81e-2	5.42e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
1.6	Bias	1.50e-2	-4.06e-2	Bias	-5.20e-2	1.02e-3	Bias	-3.08e-2	2.81e-3
	$MSE^{1/2}$	6.37e-2	4.62e-2	$MSE^{1/2}$	9.67e-2	1.02e-3	$MSE^{1/2}$	8.57e-2	1.12e-1
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
4.6	Bias	2.52e-2	-4.39e-2	Bias	-2.76e-2	1.15e-2	Bias	-4.46e-3	1.01e-1
	$MSE^{1/2}$	7.27e-2	5.06e-2	$MSE^{1/2}$	8.96e-2	1.22e-1	$MSE^{1/2}$	7.22e-2	1.32e-1
Sample Size $n = 200$									
T_1	T_2								
	0.1			1.6			4.6		
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
0.1	Bias	3.32e-3	-1.86e-2	Bias	1.14e-2	-3.66e-2	Bias	1.67e-2	-3.90e-2
	$MSE^{1/2}$	2.35e-2	1.90e-2	$MSE^{1/2}$	4.93e-2	4.56e-2	$MSE^{1/2}$	5.37e-2	5.06e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
1.6	Bias	5.39e-3	-3.69e-2	Bias	-4.83e-2	1.45e-3	Bias	-2.69e-2	1.68e-2
	$MSE^{1/2}$	4.22e-2	4.67e-2	$MSE^{1/2}$	8.29e-2	7.26e-2	$MSE^{1/2}$	6.70e-2	9.26e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
4.6	Bias	1.19e-2	-3.80e-2	Bias	-2.21e-2	1.21e-2	Bias	-8.32e-3	7.76e-2
	$MSE^{1/2}$	4.83e-2	5.29e-2	$MSE^{1/2}$	6.91e-2	9.32e-2	$MSE^{1/2}$	5.75e-2	1.08e-1

Table 2: Comparisons of the pointwise bias and square root of mean square error between the proposed spline-based sieve estimator and the semiparametric maximum pseudo-likelihood estimator when Kendall's $\tau = 0.75$

	Sample Size $n = 100$								
T_1	T_2 0.1			1.6			4.6		
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
0.1	Bias	-1.07e-2	-4.34e-2	Bias	2.86e-2	-4.19e-2	Bias	3.00e-2	-4.19e-2
	$MSE^{1/2}$	4.24e-2	4.34e-2	$MSE^{1/2}$	7.85e-2	5.37e-2	$MSE^{1/2}$	7.95e-2	5.38e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
1.6	Bias	2.80e-2	-4.01e-2	Bias	-5.63e-2	-4.72e-2	Bias	-8.88e-3	-2.09e-2
	$MSE^{1/2}$	7.92e-2	5.66e-2	$MSE^{1/2}$	9.65e-2	1.07e-1	$MSE^{1/2}$	7.64e-2	1.20e-1
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
4.6	Bias	3.00e-2	-4.00e-2	Bias	-1.02e-2	-1.98e-2	Bias	-3.13e-2	7.84e-2
	$MSE^{1/2}$	8.09e-2	5.67e-2	$MSE^{1/2}$	7.59e-2	1.15e-1	$MSE^{1/2}$	7.01e-2	1.10e-1
	Sample Size $n = 200$								
T_1	T_2 0.1			1.6			4.6		
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
0.1	Bias	-1.19e-2	-4.18e-2	Bias	1.95e-2	-3.82e-2	Bias	2.02e-2	-3.82e-2
	$MSE^{1/2}$	3.35e-2	4.41e-2	$MSE^{1/2}$	5.50e-2	5.34e-2	$MSE^{1/2}$	5.54e-2	5.35e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
1.6	Bias	2.02e-2	-3.77e-2	Bias	-5.11e-2	-2.26e-2	Bias	-1.09e-2	-3.81e-3
	$MSE^{1/2}$	5.60e-2	5.20e-2	$MSE^{1/2}$	7.72e-2	7.60e-2	$MSE^{1/2}$	5.97e-2	8.82e-2
		Sieve	Pseudo		Sieve	Pseudo		Sieve	Pseudo
4.6	Bias	2.09e-2	-3.77e-2	Bias	-1.18e-2	-4.64e-3	Bias	-3.39e-2	5.31e-2
	$MSE^{1/2}$	5.62e-2	5.20e-2	$MSE^{1/2}$	5.99e-2	9.30e-2	$MSE^{1/2}$	6.07e-2	8.38e-2

Table 3: Comparisons of the overall bias and square root of mean square error between the proposed spline-based sieve estimator and the semiparametric maximum pseudo-likelihood estimator

kendall's τ	Sample Size 100			200		
		Sieve	Pseudo		Sieve	Pseudo
0.25	Bias	-1.08e-3	-7.71e-3	Bias	-2.20e-3	-7.57e-3
	$MSE^{1/2}$	7.72e-2	1.04e-1	$MSE^{1/2}$	5.98e-2	7.93e-2
		Sieve	Pseudo		Sieve	Pseudo
0.75	Bias	-5.17e-3	-2.81e-2	Bias	-4.63e-3	-1.88e-2
	$MSE^{1/2}$	7.42e-2	1.08e-1	$MSE^{1/2}$	5.74e-2	8.27e-2

6 Final Remarks

The estimation of the joint distribution function with bivariate event time data is a challenging problem in survival analysis. Development of sophisticated methods for this type of problems is much needed for practice. In this paper, we develop a tensor spline-based sieve maximum likelihood method for estimating the joint distribution function with bivariate current status data. This sieve estimation approach reduces the dimensionality of the nonparametric maximum likelihood estimation problem substantially which makes the nonparametric maximum likelihood estimation tractable numerically. Under mild regularity conditions, we also show that the proposed spline-based sieve estimator is consistent and could converge to the true joint distribution function at a rate of $n^{1/4}$ if the true joint distribution function is smooth enough. The simulation studies indicate that the finite sample performance of this proposed sieve estimation method is generally satisfactory and even better than the semiparametric maximum pseudo-likelihood estimation method with the Clayton copula model. It is also worth noting from our simulation studies that, for estimating the marginal distribution function in this bivariate event time setting, using the joint estimation method as proposed in this article may yield a better estimator than the NPMLE with only the marginal event time data. This fact may be general true as the joint estimation method implicitly takes the potential correlation between two event times into consideration.

The proposed spline-based sieve estimation method can be readily extended to bivariate right censored and bivariate interval censored data studied by for example, Dabrowska (1988) and Kooperberg (1998), and by Maathuis (2005) and Sun, Wang and Sun (2006), respectively.

7 Technical Proofs

For the rest of this paper, we denote K as a universal positive constant that may be different from place to place and $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$, the empirical process indexed by $f(X)$.

Proof of Lemma 2.1. (i) Since $\alpha_{1,1} \geq 0$, it is obvious that $0 \leq F_n(s,t)$.

(ii) By Theorem 5.9 in Schumaker (1981), we have

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

Then by the constraints $\alpha_{i+1,1} - \alpha_{i,1} \geq 0$ and $(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0$, we have $\alpha_{i+1,j} - \alpha_{i,j} \geq 0$. Hence $\frac{\partial F_n(s,t)}{\partial s} \geq 0$ and it is followed by

$$F_n(s',t) \leq F_n(s'',t). \tag{7.1}$$

(iii) By the similar arguments as in (ii), it can be shown that

$$F_n(s,t') \leq F_n(s,t''). \tag{7.2}$$

(iv) By Theorem 5.9 in Schumaker (1981), we can derive

$$\frac{\partial^2 F_n(s,t)}{\partial s \partial t} = \sum_{i=1}^{p_n-1}\sum_{j=1}^{q_n-1} (l-1)^2 \frac{\alpha_{i+1,j+1}-\alpha_{i,j+1}-\alpha_{i+1,j}+\alpha_{i,j}}{(s_{i+l}-s_{i+1})(t_{j+l}-t_{j+1})} N_{i+1}^{(1),l-1}(s) N_{j+1}^{(2),l-1}(t). \tag{7.3}$$

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

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

(v) Since $\beta_1 - \alpha_{1,q_n} \geq 0$ and $(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0$, then $\beta_i - \alpha_{i,q_n} \geq 0$. Hence,

$$F_n(s,t) \leq F_n(s,U_2) \leq F_{n,1}(s). \tag{7.4}$$

(vi) By the similar arguments as in (v), it can be shown that

$$F_n(s,t) \leq F_n(U_1,t) \leq F_{n,2}(t). \tag{7.5}$$

(vii) By Theorem 5.9 in Schumaker (1981) again, we have

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

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

$$\frac{\partial(F_{n,1}(s) - F_n(s,t))}{\partial s} \geq 0, \text{ or } F_n(s'',t) - F_n(s',t) \leq F_{n,1}(s'') - F_{n,1}(s').$$

(viii) By the similar arguments as in (vii), it can be shown that $F_n(s,t'') - F_n(s,t') \leq F_{n,2}(t'') - F_{n,2}(t')$.

(ix) Since B-spline basis functions sum to one and their supports only cover a part of the knot intervals, then $F_{n,1}(U_1) = \beta_{p_n} N_{p_n}^{(1),l}(U_1) = \beta_{p_n}$, $F_{n,2}(U_2) = \gamma_{q_n} N_{q_n}^{(2),l}(U_2) = \gamma_{q_n}$, and $F_n(U_1, U_2) = \alpha_{p_n,q_n} N_{p_n}^{(1),l}(U_1) N_{q_n}^{(2),l}(U_2) = \alpha_{p_n,q_n}$. Hence,

$$F_{n,2}(U_2) - F_n(U_1, U_2) = \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1 - \beta_{p_n} = 1 - F_{n,1}(U_1).$$

Moreover, $\frac{dF_{n,1}(s)}{ds} \geq \frac{\partial F_n(s,t)}{\partial s}$ and $\frac{dF_{n,2}(t)}{dt} \geq \frac{\partial F_n(s,t)}{\partial t}$ guarantee $F_{n,1}(U_1) - F_{n,1}(s) \geq F_n(U_1, U_2) - F_n(s, U_2)$ and $F_{n,2}(U_2) - F_{n,2}(t) \geq F_n(U_1, U_2) - F_n(U_1, t)$, respectively. Hence,

$$\begin{aligned}
1 - F_{n,1}(s) - F_{n,2}(t) + F_n(s,t) =& \{1 - F_{n,1}(U_1) + F_{n,1}(U_1) - F_{n,1}(s)\} \\
& - \{F_{n,2}(t) - F_n(U_1, t) + F_n(U_1, t) - F_n(s, t)\} \\
\geq& \{F_{n,2}(U_2) - F_n(U_1, U_2) + F_n(U_1, U_2) - F_n(s, U_2)\} \\
& - \{F_{n,2}(t) - F_n(U_1, t) + F_n(U_1, t) - F_n(s, t)\} \\
=& \{F_{n,2}(U_2) - F_{n,2}(t) - F_n(U_1, U_2) + F_n(U_1, t)\} \\
& + \{F_n(U_1, U_2) - F_n(s, U_2) - F_n(U_1, t) + F_n(s, t)\} \\
\geq& 0.
\end{aligned} \tag{7.6}$$

□

Proof of Theorem 3.1. We show $\hat{\tau}_n$ is an consistent estimator by verifying the three conditions of Theorem 5.7 in van der Vaart (1998).

For $(s,t) \in [l_1, u_1] \times [l_2, u_2]$ we define Ω by

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

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

(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 $1 - F_1(s) - F_2(t) + F(s,t) \geq b_4$, for $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and $b_4 > 0$.

Lemma 8.1 indicates that there exist $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and $b_4 > 4$ small enough to guarantee that $\tau_0 \in \Omega$ and $\Omega_n^{'} \in \Omega$ under C2 and C6. We suppose b_1, b_2, b_3 and b_4 in above Condition (b) are chosen small enough such that Ω contains both τ_0 and $\Omega_n^{'}$.

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

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

with $\delta_1 = 1_{[T_1 \leq s]}, \delta_2 = 1_{[T_2 \leq t]}$. We denote $\mathbb{M}(\tau) = Pl(\tau)$ and $\mathbb{M}_n(\tau) = \mathbb{P}_n(l(\tau))$.

(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, since

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

Let $A_1 = \{\frac{\log F(s,t)}{\log b_1} : \tau = (F, F_1, F_2) \in \Omega\}$, and $\mathcal{G}_1 = \{1_{[l_1,s]\times[l_2,t]}, l_1 \leq s \leq u_1, l_2 \leq t \leq u_2\}$. By Conditions (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 nonincreasing in both s and t. Therefore $A_1 \subseteq \overline{sconv}(\mathcal{G}_1)$, the closure of the symmetric convex hull of $\mathcal{G}_1$ (van der Vaart and Wellner, 1996). Hence Theorem 2.6.7 in van der Vaart and Wellner (1996) implies that

$$N(\epsilon, \mathcal{G}_1, L_2(Q_{C_1,C_2})) \leq K \left(\frac{1}{\epsilon}\right)^4, \tag{7.7}$$

for any probability measure Q_{C_1,C_2} for (C_1, C_2), by the fact that $V(\mathcal{G}_1) = 3$ and the envelop function of $\mathcal{G}_1$ is 1. (7.7) is followed by

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

using the result of Theorem 2.6.9 in van der Vaart and Wellner (1996). Hence

$$\log N(\epsilon, A_1, L_2(Q_{C_1,C_2}) \leq K \left(\frac{1}{\epsilon}\right)^{4/3}. \tag{7.8}$$

Let

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

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)

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

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 $A_1^{'}$. Hence by (7.8) we have $\log N(\epsilon\hat{b}_1, A_1^{'}, L_2(Q)) \leq K\left(\frac{1}{\epsilon}\right)^{4/3}$, and it follows that

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon\hat{b}_1, A_1^{'}, L_2(Q))} d\epsilon \leq \int_0^1 \sqrt{K}\left(\frac{1}{\epsilon}\right)^{2/3} d\epsilon < \infty.$$

It is obvious that the envelop function of $A_1^{'}$ is $\hat{b}_1$, therefore $A_1^{'}$ is a P-Donsker, by Theorem 2.5.2 in van der Vaart and Wellner (1996).

Let

$$A_2^{'} = \{\delta_1(1-\delta_2)\log(F_1(s) - F(s,t)) : \tau = (F, F_1, F_2) \in \Omega\}$$

$$A_3^{'} = \{(1-\delta_1)\delta_2\log(F_2(t) - F(s,t)) : \tau = (F, F_1, F_2) \in \Omega\}$$

and

$$A_4^{'} = \{(1-\delta_1)(1-\delta_2)\log(1 - F_1(s) - F_2(t) - F(s,t)) : \tau = (F, F_1, F_2) \in \Omega\}$$

By the similar arguments in showing $A_1^{'}$ to be a P-Donsker, it can be shown that $A_2^{'}, A_3^{'}$ and $A_4^{'}$ are all P-Donsker classes. So $\mathcal{L}$ is P-Donsker as well. Since P-Donsker is also P-Clivenko-Cantelli, it then follows that $\sup_{l(\tau)\in\mathcal{L}} |(\mathbb{P}_n - P)l(\tau)| \rightarrow_p 0$.

(ii) Second, we verify

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

for any $\tau \in \Omega$. Note that

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

it follows that

$$\begin{aligned}\mathbb{M}(\tau_0) - \mathbb{M}(\tau) =& P_{C_1,C_2}\left\{Fm\left(\frac{F_0}{F}\right) + (F_1 - F)m\left(\frac{F_{0,1} - F_0}{F_1 - F}\right)\right. \\ &+ (F_2 - F)m\left(\frac{F_{0,2} - F_0}{F_2 - F}\right) \\ &\left.+(1 - F_1 - F_2 + F)m\left(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right)\right\},\end{aligned} \tag{7.9}$$

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

Since F has positive upper bound,

$$\begin{aligned}P_{C_1,C_2}\left\{Fm\left(\frac{F_0}{F}\right)\right\} &\geq P_{C_1,C_2}\left\{F\left(\frac{F_0}{F} - 1\right)^2/4\right\} \geq KP_{C_1,C_2}(F_0 - F)^2 \\ &= K\|F_0 - F\|^2_{L_2(P_{C_1,C_2})}.\end{aligned} \tag{7.10}$$

Similarly, we can easily show that

$$P_{C_1,C_2}\left\{(F_1 - F)m\left(\frac{F_{0,1} - F_0}{F_1 - F}\right)\right\} \geq K\|(F_{0,1} - F_1) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})}, \tag{7.11}$$

$$P_{C_1,C_2}\left\{(F_2 - F)m\left(\frac{F_{0,2} - F_0}{F_2 - F}\right)\right\} \geq K\|(F_{0,2} - F_2) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})}, \tag{7.12}$$

and

$$\begin{aligned}&P_{C_1,C_2}\left\{(1 - F_1 - F_2 + F)m\left(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right)\right\} \\ &\geq K\|(1 - F_{0,1} - F_{0,2} + F_0) - (1 - F_1 - F_2 + F)\|^2_{L_2(P_{C_1,C_2})}.\end{aligned} \tag{7.13}$$

So combining (7.10), (7.11), (7.12) and (7.13) results in

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

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

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K f_1 \geq (K/3)(f_1 + f_2 + f_3). \tag{7.14}$$

If f_2 is the largest, then

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K[f_1 + (f_2 - f_1)] \geq K f_2 \geq (K/3)(f_1 + f_2 + f_3). \tag{7.15}$$

If f_3 is the largest, then

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K[f_1 + (f_3 - f_1)] \geq K f_3 \geq (K/3)(f_1 + f_2 + f_3). \tag{7.16}$$

Therefore, by (7.14), (7.15) and (7.16), it follows that

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

(iii) Finally, we verify $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \geq -o_p(1)$.

Lemma 8.3 indicates that there exists $\tau_n = (F_n, F_{n,1}, F_{n,2})$ in Ω_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 $\|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv})$. Since $\hat{\tau}_n$ maximizes $\mathbb{M}_n(\tau)$ in $\Omega_n^{'}$, $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) > 0$. Hence,

$$\begin{aligned}\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) =& \mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) + \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) \geq \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) \\ =& \mathbb{P}_n(l(\tau_n)) - \mathbb{P}_n(l(\tau_0)) = (\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\} + P\{l(\tau_n) - l(\tau_0)\}.\end{aligned} \tag{7.17}$$

Define

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

Since $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$, then for any $l(\tau_n) \in \mathcal{L}_n$, we have

$$\begin{aligned}P\{l(\tau_n) - l(\tau_0)\}^2 \leq& 4P\left(\delta_1\delta_2 \log\frac{F_n}{F_0}\right)^2 + 4P\left(\delta_1(1-\delta_2)\log\frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right)^2 \\ &+ 4P\left((1-\delta_1)\delta_2\log\frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right)^2 \\ &+ 4P\left((1-\delta_1)(1-\delta_2)\log\frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right)^2 \\ \leq& 4P_{C_1,C_2}\left(\log\frac{F_n}{F_0}\right)^2 + 4P_{C_1,C_2}\left(\log\frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right)^2 \\ &+ 4P_{C_1,C_2}\left(\log\frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right)^2 + 4P_{C_1,C_2}\left(\log\frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right)^2\end{aligned} \tag{7.18}$$

The facts that $\|F_n - F_0\|_\infty \leq K(n^{-pv})$ and that F_0 has a positive lower bound result in $1/2 < \frac{F_n}{F_0} < 2$ for large n. It can be easily shown that if $1/2 \leq x \leq 2$, $|\log(x)| \leq K|x - 1|$.

Hence $\left|\log \frac{F_n}{F_0}\right| \leq K \left|\frac{F_n}{F_0} - 1\right|$ and it follows that

$$\begin{aligned} P_{C_1,C_2}\left|\log \frac{F_n}{F_0}\right|^2 \leq & K P_{C_1,C_2}\left|\frac{F_n}{F_0} - 1\right|^2 \leq K P_{C_1,C_2}|F_n - F_0|^2 \\ \leq & K(n^{-pv})^2 \to 0. \end{aligned} \tag{7.19}$$

The similar arguments yield to

$$\begin{aligned} P_{C_1,C_2}\left|\log \frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right|^2 \leq & K P_{C_1,C_2}|(F_{n,1} - F_n) - (F_{0,1} - F_0)|^2 \\ \leq & K(n^{-pv})^2 \to 0, \end{aligned} \tag{7.20}$$

$$\begin{aligned} P_{C_1,C_2}\left|\log \frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right|^2 \leq & K P_{C_1,C_2}|(F_{n,2} - F_n) - (F_{0,2} - F_0)|^2 \\ \leq & K(n^{-pv})^2 \to 0, \end{aligned} \tag{7.21}$$

and

$$P_{C_1,C_2}\left|\log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right|^2 \to 0. \tag{7.22}$$

Combining (7.18), (7.19), (7.20), (7.21) and (7.22) results in $P\{l(\tau_n) - l(\tau_0)\}^2 \to 0, \text{ as } n \to \infty$. Hence

$$\rho_P\{l(\tau_n) - l(\tau_0)\} = \{var_P[l(\tau_n) - l(\tau_0)]\}^{1/2} \leq \{P\{[l(\tau_n) - l(\tau_0)]^2\}^{1/2} \to 0, \text{ as } n \to \infty. \tag{7.23}$$

Since $\mathcal{L}$ is shown a P-Donsker in the proof of (i), Corollary 2.3.12 of van der Vaart and

Wellner (1996) indicates that

$$(\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\} = o_p\left(n^{-1/2}\right), \tag{7.24}$$

by the fact that both $l(\tau_n)$ and $l(\tau_0)$ are in $\mathcal{L}$ and (7.23).

In addition,

$$\begin{aligned} |P\{l(\tau_n) - l(\tau_0)\}| \leq & P|l(\tau_n) - l(\tau_0)| \\ \leq & K\left\{P[l(\tau_n) - l(\tau_0)]^2\right\}^{1/2} \to 0, \text{ as } n \to \infty. \end{aligned}$$

Therefore $P(l(\tau_n) - l(\tau_0)) \geq -o(1)$ as $n \to \infty$. Hence,

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

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

Proof of Theorem 3.2. We derive the rate of convergence by verifying the conditions of Theorem 3.4.1 of van der Vaart and Wellner (1996). To apply the theorem to this problem, we denote $M_n(\tau) = \mathbb{M}(\tau) = Pl(\tau)$ and $d_n(\tau_1, \tau_2) = d(\tau_1, \tau_2)$. The maximizer of $\mathbb{M}(\tau)$ is $\tau_0 = (F_0, F_{0,1}, F_{0,2})$.

(i) We first verify that for $\delta > 0$,

$$\sup_{\delta/2 < d(\tau,\tau_0) \leq \delta, \tau \in \Omega'_n} (\mathbb{M}(\tau) - \mathbb{M}(\tau_0)) \leq -K\delta^2$$

By the proof of consistency, we have already established that for any $\tau \in \Omega$, $\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq Kd^2(\tau, \tau_0)$ and it directly results in the above inequality.

(ii) We will find a function $\psi(\cdot)$ such that

$$E\left\{\sup_{\delta/2<d(\tau,\tau_0)\leq\delta,\tau\in\Omega_n'} \mathbb{G}_n(\tau - \tau_0)\right\} \leq K\frac{\psi(\delta)}{\sqrt{n}}$$

and $\delta \to \psi(\delta)/\delta^\alpha$ is decreasing on δ, for some $\alpha < 2$, and for $r_n \leq \delta^{-1}$, it satisfies

$$r_n^2\psi\left(1/r_n\right) \leq K\sqrt{n} \quad \text{for every} \quad n.$$

Let

$$\mathcal{L}_{n,\delta} = \{l(\tau) - l(\tau_0) : \tau \in \Omega_n^{'} \text{ and } \delta/2 < d(\tau, \tau_0) \leq \delta\}.$$

First, we evaluate the bracketing number of $\mathcal{L}_{n,\delta}$.

Let $\mathcal{L}_n^* = \{l(\tau) : \tau \in \Omega_n^{'}\}$, $\mathcal{F}_n = \{F : \tau = (F, F_1, F_2) \in \Omega_n^{'}\}$, $\mathcal{F}_{n,1} = \{F_1 : \tau = (F, F_1, F_2) \in \Omega_n^{'}\}$, and $\mathcal{F}_{n,2} = \{F_2 : \tau = (F, F_1, F_2) \in \Omega_n^{'}\}$.

Lemma 8.5 indicates that there exist ϵ-brackets $[F_i^L, F_i^U], i = 1, 2, \ldots, [(1/\epsilon)^{Kp_nq_n}]$ to cover $\mathcal{F}_n$. Lemma 8.6 indicates 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

$$\begin{aligned} l_{i,j,k}^U =& \delta_1\delta_2 \log F_i^U + \delta_1(1-\delta_2)\log(F_j^{(1),U} - F_i^L) + (1-\delta_1)\delta_2\log(F_k^{(2),U} - F_i^L) \\ &+ (1-\delta_1)(1-\delta_2)\log(1 - F_j^{(1),L} - F_k^{(2),L} + F_i^U), \end{aligned}$$

and

$$l_{i,j,k}^L = \delta_1\delta_2 \log F_i^L + \delta_1(1-\delta_2)\log(F_j^{(1),L} - F_i^U) + (1-\delta_1)\delta_2\log(F_k^{(2),L} - F_i^U)$$
$$+ (1-\delta_1)(1-\delta_2)\log(1 - F_j^{(1),U} - F_k^{(2),U} + F_i^L).$$

Then for any $l(\tau) \in \mathcal{L}_n^*$,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 $l_{i,j,k}^L \leq l(\tau) \leq l_{i,j,k}^U$ and the number of brackets $[l_{i,j,k}^L, l_{i,j,k}^U]'s$ is bounded by $(1/\epsilon)^{Kp_nq_n} \cdot (1/\epsilon)^{Kp_n} \cdot (1/\epsilon)^{Kq_n}$.

Note that

$$\| l_{i,j,k}^U - l_{i,j,k}^L \|_\infty \leq \left\| \log \frac{F_i^U}{F_i^L} \right\|_\infty + \left\| \log \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} \right\|_\infty + \left\| \log \frac{F_k^{(2),U} - F_i^L}{F_k^{(2),L} - F_i^U} \right\|_\infty$$
$$+ \left\| \log \frac{1 - F_j^{(1),L} - F_j^{(2),L} + F_i^U}{1 - F_j^{(1),U} - F_j^{(2),U} + F_i^L} \right\|_\infty$$

Since for any $\tau \in \Omega_n'$, F has positive lower bound, then for small ϵ, F_i^L can be made to have positive lower bound as well. Combining with the fact that $F_i^U(s,t)$ is close to $F_i^L(s,t)$ guarantees that $0 \leq \frac{F_i^U}{F_i^L} - 1 \leq 1$ for $i = 1, 2, \ldots, [(1/\epsilon)^{Kp_nq_n}]$. Note that by $\log x \leq (x-1)$ for $0 \leq (x-1) \leq 1$, therefore $\log \frac{F_i^U}{F_i^L} \leq \frac{F_i^U}{F_i^L} - 1$.

Hence,

$$\left\| \log \frac{F_i^U}{F_i^L} \right\|_\infty \leq \left\| \frac{F_i^U}{F_i^L} - 1 \right\|_\infty \leq \left\| \frac{1}{F_i^L}(F_i^U - F_i^L) \right\|_\infty \leq K \left\| F_i^U - F_i^L \right\|_\infty \leq K\epsilon.$$

Similarly, by the definition of Ω_n', we can easily show that

$$\left\| \log \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} \right\|_\infty \le K\epsilon,$$

$$\left\| \log \frac{F_k^{(2),U} - F_i^L}{F_k^{(2),L} - F_i^U} \right\|_\infty \le K\epsilon,$$

and

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

Hence, it follows that

$$N_{[\,]}\{\epsilon, \mathcal{L}_n^*, \| \, \|_\infty\} \le (1/\epsilon)^{Kp_nq_n + Kp_n + Kq_n} \le (1/\epsilon)^{Kp_nq_n}$$

and $N_{[\,]}\{\epsilon, \mathcal{L}_n^*, L_2(P)\} \le (1/\epsilon)^{Kp_nq_n}$, by the fact that L_2-norm is bounded by L_∞-norm. Finally, by $(\mathcal{L}_{n,\delta} + l(\tau_0)) \subset \mathcal{L}_n^*$,

$$N_{[\,]}\{\epsilon, \mathcal{L}_{n,\delta}, L_2(P)\} \le (1/\epsilon)^{Kp_nq_n}. \tag{7.25}$$

Second, we show that $P\{L(\tau) - L(\tau_0)\}^2 \le K\delta^2$ for any $L(\tau) - L(\tau_0) \in \mathcal{L}_{n,\delta}$. Since for any $\tau = (F, F_1, F_2)$ with $d(\tau, \tau_0) < \delta$, $\|F - F_0\|_{L_2(P_{C_1,C_2})} \le d(F, F_0) \le \delta$. Then under C1, C3 and C5, Lemma 8.7 indicates that for very small $\delta > 0$, F 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 $1/2 < \frac{F}{F_0} < 2$.

Hence $\left|\log \frac{F}{F_0}\right| \leq K \left|\frac{F}{F_0} - 1\right|$ and it follows that

$$P_{C_1,C_2} \left|\log \frac{F}{F_0}\right|^2 \leq K P_{C_1,C_2} \left|\frac{F}{F_0} - 1\right|^2 \leq K P_{C_1,C_2} \left|F - F_0\right|^2 \leq K\delta^2.$$

Again by the definition of Ω_n', we can similarly show that given a small $\delta > 0$, when n is large enough, the following inequalities are true,

$$P_{C_1,C_2} \left|\log \frac{F_1 - F}{F_{0,1} - F_0}\right|^2 \leq K\delta^2,$$

$$P_{C_1,C_2} \left|\log \frac{F_2 - F}{F_{0,2} - F_0}\right|^2 \leq K\delta^2,$$

and

$$P_{C_1,C_2} \left|\log \frac{1 - F_1 - F_2 + F}{1 - F_{0,1} - F_{0,2} + F_0}\right|^2 \leq K\delta^2.$$

Hence for any $l(\tau) - l(\tau_0) \in \mathcal{L}_{n,\delta}$, we have $P\{l(\tau) - l(\tau_0)\}^2 \leq K\delta^2$. It is obvious that $\mathcal{L}_{n,\delta}$ is uniformly bounded by the structure of the log likelihood, Lemma 3.4.2 of van der Vaart and Wellner (1996) indicates that

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

where

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

by (7.25). This gives $\psi(\delta) = (p_n q_n)^{1/2}\delta^{1/2} + (p_n q_n)/(\delta n^{1/2})$. It is easy to see that $\psi(\delta)/\delta$ is decreasing function of δ. Note that for $p_n = q_n = n^v$,

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

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

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

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

(iii) Finally, we need to show that $\mathbb{M}(\hat{\tau}_n) - \mathbb{M}(\tau_0) \geq -O_p\left(r_n^{-2}\right)$. Note that by (7.17), $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \geq I_{1,n} + I_{2,n}$, 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 (7.24), $I_{1,n} = o_p(n^{-1/2})$. Then if $v \leq \frac{1}{4p}$, we have $I_{1,n} = o_p(n^{-2pv})$. In what follows, we show that $\mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) \leq O(n^{-2pv})$.

By (7.9),

$$\begin{aligned} \mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) =& P_{C_1,C_2}\left\{F_n m\left(\frac{F_0}{F_n}\right) + (F_{n,1} - F_n) m\left(\frac{F_{0,1} - F_0}{F_{n,1} - F_n}\right)\right. \\ &+ (F_{n,2} - F_n) m\left(\frac{F_{0,2} - F_0}{F_{n,2} - F_n}\right) \\ &\left.+ (1 - F_{n,1} - F_{n,2} + F_n) m\left(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_{n,1} - F_{n,2} + F_n}\right)\right\}. \end{aligned} \tag{7.26}$$

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

$$\begin{aligned} P_{C_1,C_2}\left\{F_n m\left(\frac{F_0}{F_n}\right)\right\} \leq & K P_{C_1,C_2}\left\{F_n^2\left(\frac{F_0}{F_n}-1\right)^2\right\} = K P_{C_1,C_2}(F_0-F_n)^2 \\ \leq & K\|F_0-F_n\|_\infty^2 = O\left(n^{-2pv}\right). \end{aligned} \tag{7.27}$$

similarly, we can show that

$$\begin{aligned} P_{C_1,C_2}\left\{(F_{n,1}-F_n)m\left(\frac{F_{0,1}-F_0}{F_{n,1}-F_n}\right)\right\} \leq & K\|F_0-F_n\|_\infty^2 + K\|F_{0,1}-F_{n,1}\|_\infty^2 \\ = & O\left(n^{-2pv}\right), \end{aligned} \tag{7.28}$$

$$\begin{aligned} P_{C_1,C_2}\left\{(F_{n,2}-F_n)m\left(\frac{F_{0,2}-F_0}{F_{n,2}-F_n}\right)\right\} \leq & K\|F_0-F_n\|_\infty^2 + K\|F_{0,2}-F_{n,2}\|_\infty^2 \\ = & O\left(n^{-2pv}\right), \end{aligned} \tag{7.29}$$

and

$$\begin{aligned} & P_{C_1,C_2}\left\{(1-F_{n,1}-F_{n,2}+F_n)m\left(\frac{1-F_{0,1}-F_{0,2}+F_0}{1-F_{n,1}-F_{n,2}+F_n}\right)\right\} \\ & \leq K\|F_0-F_n\|_\infty^2 + K\|F_{0,2}-F_{n,2}\|_\infty^2 + K\|F_{0,1}-F_{n,1}\|_\infty^2 = O\left(n^{-2pv}\right). \end{aligned} \tag{7.30}$$

Combining (7.26), (7.27), (7.28), (7.29) and (7.30) results in $\mathbb{M}(\tau_0)-\mathbb{M}(\tau_n) \leq O(n^{-2pv})$. and it then follows that

$$\begin{aligned} \mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) & \geq -O\left(n^{-2pv}\right) + o_p\left(n^{-2pv}\right) = -O_p\left(n^{-2pv}\right) \\ & \geq -O_p\left(n^{-2\min\{pv,(1-2v)/3\}}\right) = -O_p\left(r_n^{-2}\right) \end{aligned}$$

Therefore, it follows by Theorem 3.4.1 in van der Vaart and Wellner (1996) that

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

$\square$

Proof of Lemma 4.1. Let $\alpha_{i,j} = \sum_{m=1}^{i} \sum_{n=1}^{j} \eta_{m,n}$, $\beta_i = \sum_{m=1}^{i} \left\{ \sum_{j=1}^{q_n} \eta_{m,j} + \omega_m \right\}$ and $\gamma_j = \sum_{n=1}^{j} \{ \sum_{i=1}^{p_n} \eta_{i,n} + \pi_n \}$. It can be easily argued that condition (2.5) and (4.3) are equivalent. By the relationship between the B-spline basis functions and the I-spline basis functions given by (4.2), it follows that

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

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

and

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

$\square$

8 Technical Lemmas

Lemma 8.1. *Suppose $\tau = \tau_0$ or $\tau \in \Omega_n'$, then under C2 and C6, 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. $F_1(s) - F(s,t)$ is nondecreasing in s and

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

(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 $\tau = \tau_0$.

Property (1) is obviously true by the properties of any joint distribution.

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

$$\begin{aligned} F(s,t) &= F_0(s,t) = P(T_1 \le s, T_2 \le t) \ge P(L_1 < T_1 \le s, L_2 < T_2 \le t) \\ &\ge (s - L_1)(t - L_2) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \ge (l_1 - L_1)(l_2 - L_2) b_0, \end{aligned}$$

$$\begin{aligned} F_1(s) - F(s,t) &= F_{0,1}(s) - F_0(s,t) = P(T_1 \le s, T_2 > t) \ge P(L_1 < T_1 \le s, t < T_2 \le U_2) \\ &\ge (s - L_1)(U_2 - t) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \ge (l_1 - L_1)(U_2 - u_2) b_0, \end{aligned}$$

$$\begin{aligned} F_2(t) - F(s,t) &= F_{0,2}(t) - F_0(s,t) = P(T_1 > s, T_2 \le t) \ge P(s < T_1 \le U_1, L_2 < T_2 \le t) \\ &\ge (U_1 - s)(t - L_2) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \ge (U_1 - u_1)(l_2 - L_2) b_0, \end{aligned}$$

and

$$
\begin{aligned}
1 - F_1(s) - F_2(t) + F(s,t) &= 1 - F_{0,1}(s) - F_{0,2}(t) + F_0(s,t) \\
&= P(T_1 > s, T_2 > t) \geq P(s < T_1 \leq U_1, t < T_2 \leq U_2) \\
&\geq (U_1 - s)(U_2 - t) \min_{s,t} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} \\
&\geq (U_1 - u_1)(U_2 - u_2) b_0.
\end{aligned}
$$

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

Lemma 2.1 indicates that $\Omega_n^{'} \subset \mathcal{F}$ in $[l_1, u_1] \times [l_2, u_2]$, hence any $\tau = (F, F_1, F_2) \in \Omega_n^{'}$ satisfies property (1).

Under C6 and the 4th condition in (3.1), (7.3) in the proof of Lemma 2.1 results in

$$
\begin{aligned}
\frac{\partial^2 F(s,t)}{\partial s \partial t} &\geq \min_{i,j} \frac{\alpha_{i+1,j+1} - \alpha_{i,j+1} - \alpha_{i+1,j} + \alpha_{i,j}}{\max_{i_1} \Delta_{i_1}^{(u)} \max_{j_1} \Delta_{j_1}^{(v)}} \\
&\geq \min_{i,j} \frac{\alpha_{i+1,j+1} - \alpha_{i,j+1} - \alpha_{i+1,j} + \alpha_{i,j}}{\frac{\min_{i_1} \Delta_{i_1}^{(u)}}{l} \frac{\min_{j_1} \Delta_{j_1}^{(v)}}{l}} \frac{\frac{\min_{i_1} \Delta_{i_1}^{(u)}}{l} \frac{\min_{j_1} \Delta_{j_1}^{(v)}}{l}}{\max_{i_1} \Delta_{i_1}^{(u)} \max_{j_1} \Delta_{j_1}^{(v)}} \qquad (8.1) \\
&\geq K.
\end{aligned}
$$

(7.1) in the proof of Lemma 2.1 and (8.1) imply that for $(s,t) \in [l_1, u_1] \times [l_2, u_2]$

$$
\begin{aligned}
F(s,t) \geq F(s,t) - F(s, L_2) - F(L_1, t) + F(L_1, L_2) &= \int_s^{L_1} \int_t^{L_2} \frac{\partial^2 F(x,y)}{\partial x \partial y} dy dx \\
\geq (s - L_1)(t - L_2) \min_{s,t} \frac{\partial^2 F(s,t)}{\partial s \partial t} &\geq (l_1 - L_1)(l_2 - L_2) K.
\end{aligned}
$$

It follows by (7.4) and (7.2) in the proof of Lemma 2.1 and (8.1) that

$$\begin{aligned} F_1(s) - F(s,t) \geq & F(s,U_2) - F(s,t) \geq F(s,U_2) - F(s,t) - F(L_1,U_2) + F(L_1,t) \\ \geq & (l_1 - L_1)(U_2 - u_2)K. \end{aligned}$$

Similarly, (7.5) and (7.1) in the proof of Lemma 2.1 and (8.1) result in

$$F_2(t) - F(s,t) \geq (U_1 - u_1)(l_2 - L_2)K.$$

Finally, (7.6) in the proof of Lemma 2.1 and (8.1) result in

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

□

Lemma 8.2. *Suppose $g(x,y)$ is a bivariate function in closed region $[L_1,U_1] \times [L_2,U_2]$ with the continuous mixed derivatives of order w, $\nabla_m^w g = \frac{\partial^w g(x,y)}{\partial x^m y^{w-m}}$ for $m = 1,2,\ldots,w$. Then there exists a bivariate function made of a linear combination of tensor B-spline basis functions, $Ag(x,y) = \sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j} N_i^{(1),l}(x) N_j^{(2),l}(y)$, with order $l \geq w+1$ for every B-spline basis function and $\{N_i^{(1),l}\}_{i=1}^p$ having knot sequence $\{u_i\}_{i=1}^{p+l}$ satisfying $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U_1$, $\{N_j^{(2),l}\}_{j=1}^q$ having knot sequence $\{v_j\}_{j=1}^{q+l}$ satisfying $L_2 = v_1 = \cdots = v_l < v_{l+1} < \cdots < v_q < v_{q+1} = \cdots = v_{q+l} = U_2$, such that for some* *constant* $K > 0$

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

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

Proof. The proof of this lemma closely follows the arguments for justifying Jackson Theorem in De Boor (2001, p149). 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 automatically true 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

$$\begin{aligned} |g(x_1,y_1)-g(x_2,y_2)| \le & |g(x_1,y_1)-g(x_3,y_3)| + |g(x_3,y_3)-g(x_2,y_2)| \\ \le & \max_{\substack{|x_1-x_3|\le h_1\\ |y_1-y_3|\le h_1}} |g(x_1,y_1)-g(x_3,y_3)| + \max_{\substack{|x_2-x_3|\le h_2\\ |y_2-y_3|\le h_2}} |g(x_3,y_3)-g(x_2,y_2)| \\ = & \omega(g;h_1)+\omega(g;h_2). \end{aligned} \tag{8.2}$$

By (8.2), $\omega(g;h_1+h_2) \le \omega(g;h_1)+\omega(g;h_2)$ 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 linear combination of the tensor B-spline basis functions Ag to approximate the smooth function g on $[L_1, U_1] \times [L_2, U_2]$ as follows.

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

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

$$Ag(\hat{x}, \hat{y}) = \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} g(\tau_i, \xi_j) N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}), \tag{8.3}$$

due to the fact that the supports of the B-spline basis functions only cover a part of the knot intervals. Since the B-spline basis functions sum to one, we have

$$\begin{aligned} g(\hat{x}, \hat{y}) &= g(\hat{x}, \hat{y}) \sum_{i=j_1+1-l}^{j_1} N_i^{(1),l}(\hat{x}) \\ &= g(\hat{x}, \hat{y}) \sum_{i=j_1+1-l}^{j_1} \left\{ \sum_{j=j_2+1-l}^{j_2} N_j^{(2),l}(\hat{y}) \right\} N_i^{(1),l}(\hat{x}) \\ &= g(\hat{x}, \hat{y}) \sum_{i=j_1+1-l}^{j_1} \sum_{j=j_2+1-l}^{j_2} N_i^{(1),l}(\hat{x}) N_j^{(2),l}(\hat{y}). \end{aligned} \tag{8.4}$$

Subtracting (8.3) from (8.4) yields,

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

Hence,

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

$$\le \max_{\substack{j_1+1-l\le i\le j_1 \\ j_2+1-l\le j\le j_2}} |g(\hat{x},\hat{y}) - g(\tau_i,\xi_j)|.$$

We specifically choose the sequences of $\{\tau_i\}_{i=1}^p$ and $\{\xi_j\}_{j=1}^q$ as follows

$$\tau_i = \begin{cases} u_1 + \frac{(i-1)(u_{l+1}-u_l)}{l}, \ i = 1,\dots,l, \\ \\ u_i, \ i = l+1,\dots,p. \end{cases} \tag{8.5}$$

$$\xi_j = \begin{cases} v_1 + \frac{(j-1)(v_{l+1}-v_l)}{l}, \ j = 1,\dots,l, \\ \\ v_j, \ j = l+1,\dots,q, \end{cases} \tag{8.6}$$

Then (8.5) and (8.6) 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 $\hat{y} \in [v_{j_2}, v_{j_2+1}]$.

Then 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 $\hat{y} \in [v_{j_2}, v_{j_2+1}]$

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

Hence,

$$\begin{aligned}\max_{\substack{j_1+1-l\leq i\leq j_1\\ j_2+1-l\leq j\leq j_2}} |g(\hat{x},\hat{y}) - g(\tau_i,\xi_j)| \leq &\max\{|g(x_1,y_1) - g(x_2,y_2)| : \\ & |x_1 - x_2| \leq (l+1)|T|, |y_1 - y_2| \leq (l+1)|T|\} \\ = &\omega(g;(l+1)|T|) = (l+1)\omega(g;|T|),\end{aligned} \tag{8.7}$$

where the last inequality is due to the subadditivity property of $\omega(g;h)$.

(8.7) implies that

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

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

$$d(g,\psi_{l,l}) = \inf_{f\in\psi_{l,l}} \|g - f\| \leq (l+1)\omega(g;|T|), \tag{8.8}$$

where $\psi_{l,l}$ denotes the set of all linear combinations of the tensor B-spline basis functions with order l for every basis function. Because the distance of function g from $\psi_{l,l}$ is the same as the distance of the function $g - f$ from $\psi_{l,l}$ for $f \in \psi_{l,l}$, (8.8) implies

$$d(g,\psi_{l,l}) = d(g - f,\psi_{l,l}) \leq (l+1)\omega(g - f,|T|). \tag{8.9}$$

Furthermore, since g has bounded partial derivatives, then

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

Therefore, by (8.9)

$$d(g,\psi_{l,l}) \le (l+1)|T| \left(\left\|\frac{\partial(g-f)}{\partial y}\right\|_\infty + \left\|\frac{\partial(g-f)}{\partial x}\right\|_\infty \right). \tag{8.10}$$

Since $\psi_{l,l-1} = \left\{\frac{\partial f}{\partial y} : f \in \psi_{l,l}\right\}$ and $\psi_{l-1,l} = \left\{\frac{\partial f}{\partial x} : f \in \psi_{l,l}\right\}$, (8.10) implies

$$d(g,\psi_{l,l}) \le (l+1)|T| \left\{ d\left(\frac{\partial g}{\partial x}, \psi_{l-1,l}\right) + d\left(\frac{\partial g}{\partial y}, \psi_{l,l-1}\right) \right\}. \tag{8.11}$$

Iterating the same derivation for (8.11) leads to

$$\begin{aligned}&d(g,\psi_{l,l}) \\ &\quad\le K|T|^{w-1} \left\{ d\left(\frac{\partial^{w-1} g}{\partial x^{w-1}}, \psi_{l-w+1,l}\right) + d\left(\frac{\partial^{w-1} g}{\partial x^{w-2}\partial y}, \psi_{l-w+2,l-1}\right) + \cdots + d\left(\frac{\partial^{w-1} g}{\partial y^{w-1}}, \psi_{l,l-w+1}\right) \right\} \\ &\quad\le K|T|^{w-1} \left\{ \omega\left(\frac{\partial^{w-1} g}{\partial x^{w-1}}, |T|\right) + \omega\left(\frac{\partial^{w-1} g}{\partial x^{w-2}\partial y}, |T|\right) + \cdots + \omega\left(\frac{\partial^{w-1} g}{\partial y^{w-1}}, |T|\right) \right\} \\ &\quad\le K|T|^{w} \left\{ \left\|\frac{\partial^w g}{\partial x^w}\right\|_\infty + \left\|\frac{\partial^w g}{\partial x^{w-1}\partial y}\right\|_\infty + \cdots + \left\|\frac{\partial^w g}{\partial y^w}\right\|_\infty \right\} \\ &\quad\le K|T|^{w} \max_{0\le m\le w} \left\|\frac{\partial^w g}{\partial x^m \partial y^{w-m}}\right\|_\infty .\end{aligned}$$

□

Lemma 8.3. *Let $p_n = O(n^v)$ and $q_n = O(n^v)$. If C2, C3 and C6 hold, there exists $\tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n^{'}$, such that $\|F_n - F_0\|_\infty \le K(n^{-pv})$, $\|F_{n,1} - F_{0,1}\|_\infty \le K(n^{-pv})$ and $\|F_{n,2} - F_{0,2}\|_\infty \le K(n^{-pv})$.*

Proof. Suppose 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 (8.5) and (8.6) in the proof of Lemma 8.2. With C3, C6, Jackson Theorem in De boor (2001, p149) and Lemma 8.2, it is easily seen that that $\|F_n - F_0\|_\infty \le K(n^{-pv})$, $\|F_{n,1} - F_{0,1}\|_\infty \le K(n^{-pv})$, and $\|F_{n,2} - F_{0,2}\|_\infty \le K(n^{-pv})$.

To complete the proof, it remains to show that $\alpha_{i,j}$, β_i and γ_j satisfy the conditions in (3.1).

(i) $\alpha_{1,1} = F_0(\tau_1, \xi_1) \ge 0$.

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

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

(iv) $\frac{(\alpha_{i+1,j+1}-\alpha_{i+1,j})-(\alpha_{i,j+1}-\alpha_{i,j})}{\frac{\min_{i_1}\Delta_{i_1}^{(u)}}{l}\frac{\min_{j_1}\Delta_{j_1}^{(v)}}{l}} \ge \frac{\alpha_{i+1,j+1}-\alpha_{i,j+1}-\alpha_{i+1,j}+\alpha_{i,j}}{(\tau_{i+1}-\tau_i)(\xi_{j+1}-\xi_i)}$
$= \frac{F_0(\tau_{i+1},\xi_{j+1})-F_0(\tau_i,\xi_{j+1})-F_0(\tau_{i+1},\xi_j)+F_0(\tau_i,\xi_j)}{(\tau i+1-\tau_i)(\xi_{j+1}-\xi_i)} \ge \min_{\substack{s\in[L_1,U_1]\\ t\in[L_2,U_2]}} \frac{\partial^2 F_0(s,t)}{\partial s \partial t} = b_0$, by C2.

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

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

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

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

□

Lemma 8.4. *Let S be a sphere in R^n with radius $(n^{1/2}\sigma)$, 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 $\log N(\epsilon, S, \|\cdot\|_\infty) \leq Kn\log(\sigma/\epsilon)$, for some constant $K > 0$ and $\epsilon < \sigma$.*

Proof. The proof follows along the same lines as for the proof of Lemma 5 in Shen and Wong (1994). □

Lemma 8.5. $\Theta_\delta = \{\phi : \phi(s,t) = \sum_{i=1}^p \sum_{j=1}^q \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t), \|\phi\|_\infty < \delta\}$, *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 $i = 1, \ldots, p$, $\{N_i^{(1),l}\}_{i=1}^p$ and $\{N_j^{(2),l}\}_{j=1}^q$ are two sets of B-spline basis functions with the knot sequence $\{u_i\}_{i=1}^{p+l}$ 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 $\{v_j\}_{1=1}^{q+l}$ 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 $\log N_{[\,]}(\epsilon, \Theta, \|\cdot\|_\infty) \leq Kpq\log(\delta/\epsilon)$, for some constant $K > 0$ and $\epsilon < \delta$.*

Proof. For any $\phi \in \Theta_\delta$, we have

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

since the B-spline basis functions sum to one and their supports only cover a part of the knot intervals. Then $\|\phi\|_\infty^2 = \alpha_{p,q}^2 \geq \frac{1}{pq}\sum_{i=1}^p \sum_{j=1}^q a_{i,j}^2$ and $\|\phi\|_\infty^2 \leq \delta^2$. Hence, for the coefficients

$(\alpha_{1,1}, \cdots, \alpha_{p,q})$ of ϕ,

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_{i,j}^2 \le pq\|\phi\|_\infty^2 \le pq\delta^2. \tag{8.12}$$

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 8.4 indicates that there exist ϵ-balls $B_1, B_2, \cdots, B_{[(\frac{\delta}{\epsilon})^{Kpq}]}$ centered at $\underline{\alpha}^{(1)} = (\alpha_{1,1}^{(1)}, \cdots, \alpha_{p,q}^{(1)})$, $\underline{\alpha}^{(2)} = (\alpha_{1,1}^{(2)}, \cdots, \alpha_{p,q}^{(2)})$, $\cdots$, $\underline{\alpha}^{([(\frac{\delta}{\epsilon})^{Kpq}])} = (\alpha_{1,1}^{([(\frac{\delta}{\epsilon})^{Kpq}])}, \cdots, \alpha_{p,q}^{([(\frac{\delta}{\epsilon})^{Kpq}])})$, 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 $k = 1, \cdots, [(\frac{\delta}{\epsilon})^{Kpq}]$, 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 $\{\Psi_1^{(k)} : k = 1, \cdots, [(\frac{\delta}{\epsilon})^{Kpq}]\}$ constitute a set of ϵ-balls for Ψ.

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

For any $\psi(s,t) = \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_{i,j}N_i^{(1),l}(s)N_j^{(2),l}(t) \in \Theta_\delta$, its coefficients $\underline{\alpha} = (\alpha_{1,1}, \cdots, \alpha_{p,q}) \in S$ by (8.12). By the fact that ϵ-balls $B_1, B_2, \cdots, B_{[(\frac{\delta}{\epsilon})^{Kpq}]}$ cover S, there exists m with $1 \le m \le [(\frac{\delta}{\epsilon})^{Kpq}]$, such that

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

Hence, for any $(s,t) \in [L_1, U_1] \times [L_2, U_2]$,

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

And it follows that

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

This implies that $\{\Psi_1^{(k)} : k = 1, \cdots, [(\frac{\delta}{\epsilon})^{Kpq}]\}$ cover Θ_δ. Hence the ϵ-covering number of Θ_δ is bounded by $[(\frac{\delta}{\epsilon})^{Kpq}]$, or $\log N(\epsilon, \Theta_\delta, \|\cdot\|_\infty) \leq Kpq \log(\delta/\epsilon)$. It is obvious that

$$N_{[\,]}(2\epsilon, \Theta_\delta, \|\cdot\|_\infty) \leq N(\epsilon, \Theta_\delta, \|\cdot\|_\infty).$$

Therefore, it follows that

$$\log N_{[\,]}(\epsilon, \Theta_\delta, \|\cdot\|_\infty) \leq Kpq \log(\delta/\epsilon).$$

□

Lemma 8.6. *$\Theta_\delta = \{\phi : \phi(s) = \sum_{i=1}^{p} \beta_i N_i^l(s), \|\phi\|_\infty < \delta\}$, where $0 \leq \beta_1 \leq \beta_2 \leq \cdots \beta_p$, $\{N_i^l\}_{i=1}^p$ are the B-spline basis functions with the knot sequence $\{u_i\}_{i=1}^{p+l}$ satisfying $L = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_p < u_{p+1} = \cdots = u_{p+l} = U$. Then $\log N_{[\,]}(\epsilon, \Theta_\delta, \|\cdot\|_\infty) \leq Kp \log(\delta/\epsilon)$, for some constant $K > 0$ and $\epsilon < \delta$.*

Proof. The proof is exactly along the same lines of those for Lemma 8.5, and thus is omitted.

□

Remark 8.1. *In the proof of Theorem 3.2 (convergence rate), we use the fact that* $\delta \leq 1$, *then it is obvious that* $\log N_{[\,]}(\epsilon, \Theta_\delta, \|\cdot\|_\infty) \leq Kpq \log(1/\epsilon)$ *by both Lemma 8.5 and Lemma 8.6.*

Lemma 8.7. $\Lambda_0(s,t)$ *and* $\Lambda(s,t)$ *are both partially nondecreasing functions in the domain* $[L_1, U_1] \times [L_2, U_2]$ *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*

$$\sup_{(s,t)\in[L_1,U_1]\times[L_2,U_2]} |\Lambda(s,t) - \Lambda_0(s,t)| \leq (\eta/K)^{1/2}.$$

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

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

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

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

Then either $\Lambda(s',t') \geq \Lambda_0(s',t') + \xi/2$ or $\Lambda_0(s',t') \geq \Lambda(s',t') + \xi/2$. In the following, we only show the inequality for the first case, $\Lambda(s',t') \geq \Lambda_0(s',t') + \xi/2$, as the arguments are parallel for the second case.

There exists h satisfying $(s'+h, t'+h) \equiv (s'', t'')$, such that $\Lambda_0(s'', t'') = \Lambda_0(s', t') + \xi/2$. Then

$$\begin{aligned}
\eta^2 &\geq \int \{\Lambda(s,t) - \Lambda_0(s,t)\}^2 d\mu(s,t) \\
&= \int\int_{(s,t)\in[L_1,U_1]\times[L_2,U_2]} \{\Lambda(s,t) - \Lambda_0(s,t)\}^2 \frac{\partial^2\mu(s,t)}{\partial s \partial t} ds dt \\
&\geq \int_{t'}^{t''}\int_{s'}^{s''} \{\Lambda(s,t) - \Lambda_0(s,t)\}^2 \frac{\partial^2\mu(s,t)}{\partial s \partial t} ds dt \\
&\geq \int_{t'}^{t''}\int_{s'}^{s''} \{\Lambda_0(s'',t'') - \Lambda_0(s,t)\}^2 \frac{\partial^2\mu(s,t)}{\partial s \partial t} ds dt \\
&\geq c_0 \int_{t'}^{t''}\int_{s'}^{s''} \{\Lambda_0(s'',t'') - \Lambda_0(s,t)\}^2 ds dt \\
&= c_0 \int_{t'}^{t''}\int_{\Lambda_0(s',t)}^{\Lambda_0(s'',t)} \{\Lambda_0(s'',t'') - x\}^2 \frac{1}{\partial\Lambda_0(s,t)/\partial s|_{s=f_t^{-1}(x)}} dx dt \\
&\geq (c_0/f_0) \int_{t'}^{t''}\int_{\Lambda_0(s',t)}^{\Lambda_0(s'',t)} \{\Lambda_0(s'',t'') - x\}^2 dx dt \\
&= (c_0/f_0) \int_{t'}^{t''} \{(\Lambda_0(s'',t'') - \Lambda_0(s',t))^3/3 - (\Lambda_0(s'',t'') - \Lambda_0(s'',t))^3/3\} dt,
\end{aligned}$$

where $x = f_t(s) \equiv \Lambda_0(s,t)$. Therefore, by $a^3 - b^3 = (a-b)(a^2+ab+b^2) \geq (a-b)(a^2+b^2)$ for $ab(a-b) \geq 0$, it follows that

$$\begin{aligned}
\eta^2 \geq \frac{c_0}{3f_0}\int_{t'}^{t''} &(\Lambda_0(s'',t) - \Lambda_0(s',t)) \\
&[(\Lambda_0(s'',t'') - \Lambda_0(s',t))^2 + (\Lambda_0(s'',t'') - \Lambda_0(s'',t))^2] dt.
\end{aligned} \tag{8.13}$$

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

$$\Lambda_0(s'',t) - \Lambda_0(s',t) = (\partial\Lambda_0(w,t)/\partial s)h \geq h/f_0. \tag{8.14}$$

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

$$\begin{aligned}\xi/2 &= \Lambda_0(s'', t'') - \Lambda_0(s', t') \\ &= \Lambda_0(s'', t'') - \Lambda_0(s'', t') + \Lambda_0(s'', t') - \Lambda_0(s', t') \\ &\leq 2hf_0.\end{aligned} \tag{8.15}$$

Combining (8.14) and (8.15) yields,

$$\Lambda_0(s'', t) - \Lambda_0(s', t) \geq \frac{\xi}{4f_0^2}. \tag{8.16}$$

Finally, substituting (8.16) into (8.13), we obtain

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

where $x = g_{s''}(t) \equiv \Lambda_0(s'', t)$. This yields the stated conclusion with $K \equiv \sqrt{c_0/(2304f_0^{10})}$. $\square$

References

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

[2] Curry, H. B. and Schoenberg, I. J. (1966). On Polya frequency functions. IV. The fundamental spline functions and their limits. *J. Analyse math.* **17** 71–107.

[3] Dabrowska, D. M. (1988). Kaplan-Meier estimate on the Plane. *Ann. Statist.* **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] De Boor, C. (2001). *A Practical Guide to Splines.* Springer, New York.

[7] Duffy, D. L., Martin, N. G. and Matthews, J. D. (1990). Appendectomy in australian twins. *Am. J. Hum. Genet.* **47** 590–592.

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

[9] Gentleman, R. and Vandal, A. C. (2001). Computational Algorithms for Censored-Data Problems Using Intersection Graphs. *J. Comput. Graph. Statist.* **10** 403–421.

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

[11] Grummer-Strawn, L. M. (1993). Regression analysis of current status data: An application to breat feeding. *J. Amer. Statist. Assoc.* **88** 758–765.

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

[13] Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568.

[14] Huang, J. and Wellner, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, Case 1. *Statist. Neer.* **49** 153–163.

[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. Springer, new York.

[16] Jamshidian, M. (2004). On algorithms for restricted maximum likelihood estimation. Comput. *Comp. Statist. Data Anal.* **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. *J Amer. Statist. Assoc.*, **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] KOO, J.-Y. (1996). Bivariate B-splines for Tensor Logspline Density Estimation. *Comp. Statist. Data Anal.* **21** 31–42.

[20] KOOPERBERG, C. (1998). Bivariate density estimation with an application to survival analysis. *J. Comput. Graph. Statist.* **7** 322–341.

[21] LU, M., ZHANG, Y. and HUANG, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94** 705–718.

[22] LU, M., ZHANG, Y. and HUANG, J. (2009). Semiparametric Estimation Methods for Panel Count Data Using Monotone B-Splines. *J. Amer. Statist. Assoc.* **104** 1060–1070.

[23] MAATHUIS, M. H. (2005). Reduction Algorithm for the NPMLE for the Distribution Function of Bivariate Interval-Censored Data. *J. Comput. Graph. Statist.* **14** 352–362.

[24] NELSEN, R. B. (2006). *An Introduction to Copulas.* Springer, New York.

[25] PRENTICE, R. L. and CAI, J. (1992). Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika.* **79** 495–512.

[26] PRUITT, R. C. (1991). Strong Consistency of Self-Consistent Estimators: General Theory and an Application to Bivariate Survival Analysis. *Technical Report, Department of Statisticcs, University of Minnesota.*

[27] QUALE, C. M., VAN DER LAAN, M. J. and JAMES R. ROBINS, J. R. (2006). Locally Efficient Estimation With Bivariate Right-Censored Data. *J. Amer. Statist. Assoc.* **101** 1076–1084.

[28] RAMSAY, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3** 425–441.

[29] Rosen, J. B. (1960), The Gradient Projection Method for Nonlinear Programming. *J. SIAM* **8** 181-217.

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

[31] Scott, D. W. (1992). *Multivariate Density Estimation.* Wiley, New York.

[32] Shen, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika* **85** 165–177.

[33] Shen, X. and Wong, W. H. (1994). Convergence Rate of Sieve Estimates. *Ann. Statist.* **22** 580–615.

[34] Shiboski, S. C. and Jewell, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Amer. Statist. Assoc.* **87** 360–372.

[35] Shih, J. H. and Louis, T. A. (1995). Inference on the association parameter in copula models for bivariate survival data. *Biometrics* **51** 1384–1399.

[36] Stone, C. J. (1994). The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation (with discussion). *Ann. Statist.* **22** 118–184.

[37] Sun, L., Wang, L. and Sun, J. (2006). Estimation of the Association for Bivariate Interval-censored Failure Time Data. *Scand. J. Statist.* **33** 637–649.

[38] ver der Laan, M. J. (1996). Efficient Estimation of the Bivariate Censoring Model and Repairing NPMLE. *Ann. Statist.* **24** 596–627.

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

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

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

[42] Wang, Y. and Taylor, J. (2004). Monotone Constrained Tensor-product B-spline with applicaiton to screen studies *Technical Report, Department of Biostatistics, University of Michigan*.

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

[44] Wong, G. Y. and Yu, Q. (1999). Generalized MLE of a Joint Distribution Function with Multivariate Interval-Censored Data. *J. Multivariate Anal.* **69** 155–166.

[45] Zhang, S., Zhang, Y., Chaloner, K. and Stapleton, J. T. (2010). A copula model for bivariate hybrid censored survival data with application to the MACS study. *Lifetime Data Analysis* **16,** 231-249.

[46] Zhang, Y., Hua, L. and Huang, J. (2010). A Spline-Based Semiparametric Maximum Likelihood Estimation for the Cox Model with Interval-Censored Data. *Scand. J. Statist.* **37** 338–354.