

## **Sieve Estimation with Bivariate Interval Censored Data**

By: Yuan Wu and [Xiaoli Gao](#)

Wu, Y., Gao, X. (2011) Sieve Estimation with Bivariate Interval Censored Data. *Journal of Statistics: Advances in Theory and Applications* 5(1), 37-61.

**Made available courtesy of Scientific Advances Publishers:**  
<http://www.scientificadvances.co.in/artical/4/61>

### **Abstract:**

Bivariate interval censored data arises in many applications. However, both theoretical and computational investigations for this type of data are limited because of the complicated structure of bivariate censoring. In this paper, we propose a two-stage spline-based sieve estimator for the association between two event times with bivariate case 2 interval censored data. A smooth and explicit estimator for the joint distribution function is also available. The proposed estimators are shown to be asymptotically consistent and computationally efficient. We demonstrate the finite sample performances of the spline-based sieve estimators using both simulation studies and real data analysis from an AIDS clinical trial study.

**Keywords:** association parameter | bivariate interval censored data | case 2 interval censored data | copula model | semiparametric problem | spline-based sieve estimator

### **Article:**

**\*\*\*Note: Full text of article below**

## **SIEVE ESTIMATION WITH BIVARIATE INTERVAL CENSORED DATA**

**YUAN WU and XIAOLI GAO**

Department of Epidemiology  
University of Iowa  
Iowa City, Iowa 52242  
USA  
e-mail: yuan-wu@uiowa.edu

Department of Mathematics and Statistics  
Oakland University  
Rochester, Michigan 48309  
USA

### **Abstract**

Bivariate interval censored data arises in many applications. However, both theoretical and computational investigations for this type of data are limited because of the complicated structure of bivariate censoring. In this paper, we propose a two-stage spline-based sieve estimator for the association between two event times with bivariate case 2 interval censored data. A smooth and explicit estimator for the joint distribution function is also available. The proposed estimators are shown to be asymptotically consistent and computationally efficient. We demonstrate the finite sample performances of the spline-based sieve estimators using both simulation studies and real data analysis from an AIDS clinical trial study.

---

2010 Mathematics Subject Classification: 62N01.

Keywords and phrases: association parameter, bivariate interval censored data, case 2 interval censored data, copula model, semiparametric problem, spline-based sieve estimator.

Received January 25, 2011

© 2011 Scientific Advances Publishers

## 1. Introduction

In some situations, the observation of random event time  $T$  is restricted to the knowledge that whether  $T$  is left of  $U$ , between  $U$  and  $V$ , or right of  $V$ , where  $U$  and  $V$  are two monitoring times with  $U < V$ . This type of data is known as being interval censored; it is categorized as case 2 interval censored data when  $0 < U < V < \infty$  and case 1 interval censored data or current status data when  $U = 0$  or  $V = \infty$ . Interval censor data arises naturally in many applications, see, for example, animal tumorigenicity experiments in Hoel and Walburg [9], Finkelstein and Wolfe [5]; the studies of the age at weaning in Diamond et al. [3], Diamond and McDonald [4]; and acquired immunodeficiency syndrome (AIDS) studies in Shibuski and Jewell [17], and Jewell et al. [12].

The non-parametric maximum likelihood estimator (NPMLE) for the distribution function has been widely investigated for univariate interval censored data. For example, for current status data, Groeneboom and Wellner [8] and Huang and Wellner [10] studied the asymptotic properties of the NPMLE for current status data. Groeneboom and Wellner [8] also introduced a convex minorant algorithm to compute the NPMLE. For case 2 interval censored data, one may expect better estimation results than for current status data, since more information of the location of the event time is available. However, both theoretical and practical aspects of the problem are more complicated. Only iterative procedures are available for the computation of the NPMLE of the distribution function. For example, Groeneboom and Wellner [8] proposed iterative convex minorant (ICM) algorithm for this matter. Groeneboom and Wellner [8] and Wellner [23] found the convergence rate of the NPMLE for two different situations of the observation time distribution.

Bivariate case 1 or case 2 interval censored data is collected to study two diseases in same patient or one common disease for two correlated subjects. Two main interests of these studies are the calibration of the association and the joint distribution of two event times. Goggins and Finkelstein [7] studied an example of bivariate case 2 interval censored

data from an AIDS clinical trial. The two event times of interest are the times to cytomegalovirus (CMV) shedding in the blood and in the urine. Although the patients in the study were assigned prescheduled clinic visit times, many missed some of the visits and returned with a change in laboratory results for CMV shedding, thus yielding interval censored times of CMV shedding in the blood and in the urine. The data set consists of 204 patients. For time to shedding in the urine, 49 patients were left censored, 88 were right censored, and 67 were interval censored. For shedding in the blood, 7 observations were left censored, 174 were right censored, and 23 were interval censored. CD4 cell count for each patient is related to immune system deterioration. 204 CD4 counts are dichotomized into high and low based on a threshold, 93 patients have high CD4 counts and 111 patients have low CD4 counts.

To estimate the joint distribution function with bivariate case 2 interval censored data, a non-parametric maximum likelihood estimation method can be generalized from the univariate case. For computing the generalized NPMLE, one needs to design an efficient searching algorithm for the non-zero mass intersection rectangles [e.g., 1, 6, 14, 25]. Alternatively, a semiparametric approach with copula model assumption is often used in studying bivariate censored data [e.g., 18, 22]. For bivariate case 2 interval censored data, Sun et al. [19] applied this semiparametric approach and proposed a two-stage maximum pseudo-likelihood estimator for the association parameter between two event times. They computed the NPMLEs of the marginal distribution functions in the first stage, and then obtained the maximum pseudo-likelihood estimator for the association parameter.

In Sun et al. [19], bootstrap method is necessary to estimate the standard error of the estimated association parameter, which is time-consuming. So, it is worthwhile to develop a faster approach to estimate the marginal distribution functions. Moreover, the NPMLE of each marginal distribution function in Sun et al. [19] is neither smooth nor explicit.

In this paper, we investigate the association and joint distribution of two event times with case 2 interval censored data by using spline-based sieve estimation method. Very recently, the spline-based sieve maximum likelihood estimation method has often been used in survival analysis. See, for example, Lu et al. [13] and Zhang et al. [28]. Under a copula model, we adopt a two-stage approach and apply the spline-based sieve method to estimate the marginal distributions first, and then the association parameter. We make three contributions in this paper: The proposed two-stage estimators are asymptotically consistent; thanks to the spline procedure, the computation of the two-stage estimator for the association parameter is much more efficient than the two-stage semiparametric method in Sun et al. [19]; thanks to the spline procedure, the estimation for the joint distribution of two failure times is smooth and explicit.

The rest of the paper is organized as follows. Section 2 proposes the spline-based semiparametric estimators. Section 3 summarizes the consistency of the proposed estimators. Section 4 illustrates the finite sample performances of the proposed method by simulation studies and a real AIDS clinical trial study. Section 5 summarizes the strengths of the proposed method and outlines some related future work. Section 6 sketches the proof of the consistency for the spline-based sieve estimators for the marginal distribution function and the association parameter.

## 2. Spline-Based Semiparametric Estimation

### 2.1. Model and likelihood

Let  $(T_1, T_2)$  be the two event times of interest and  $(U_1, V_1)$  and  $(U_2, V_2)$  are the two pairs of random monitoring times for  $T_1$  and  $T_2$ , respectively. In this setting, the observation of bivariate case 2 interval censored data consists of

$$(U_1, V_1, \Delta_1^{(1)} = I(T_1 \leq U_1), \Delta_1^{(2)} = I(U_1 < T_1 \leq V_1), \Delta_1^{(3)} = I(T_1 > V_1)),$$

$$(U_2, V_2, \Delta_2^{(1)} = I(T_2 \leq U_2), \Delta_2^{(2)} = I(U_2 < T_2 \leq V_2), \Delta_2^{(3)} = I(T_2 > V_2)), \quad (1)$$

where  $I(\cdot)$  is the indicator function. The joint distribution of the two event times is assumed to follow a copula model, that is, the joint survival function

$$\begin{aligned} S(t_1, t_2) &= P(T_1 > t_1, T_2 > t_2) = C_\alpha(S_1(t_1), S_2(t_2)) \\ &= C_\alpha(P(T_1 > t_1), P(T_2 > t_2)). \end{aligned}$$

Different copula models can be used to construct the above joint distribution function [15]. For example, we can choose the Clayton copula

$$C_\alpha(u, v) = (u^{-\alpha} + v^{-\alpha} - 1)^{1/(-\alpha)}, \quad \alpha > 0, \quad (2)$$

where  $\alpha$  is an association parameter with a larger  $\alpha$  corresponding to a stronger positive association between the two marginal distributions and  $\alpha \rightarrow 0$  corresponding to independence between the two event times. The association parameter  $\alpha$  and Kendall's  $\tau$  for the Clayton copula, are related by  $\tau = \alpha / (\alpha + 2)$ .

Consider  $n$  pairs of bivariate current status data (1),  $\{u_{1,k}, v_{1,k}, \delta_{1,k}^{(1)}, \delta_{1,k}^{(2)}, \delta_{1,k}^{(3)} : k = 1, 2, \dots, n\}$  and  $\{u_{2,k}, v_{2,k}, \delta_{2,k}^{(1)}, \delta_{2,k}^{(2)}, \delta_{2,k}^{(3)} : k = 1, 2, \dots, n\}$ . Suppose that  $(T_1, T_2)$  and  $(U_1, V_1, U_2, V_2)$  are independent and  $(U_1, V_1, U_2, V_2)$  are non-informative to  $(T_1, T_2)$ . Then, under the Clayton copula model (2), the log-likelihood for the observed data can be expressed as

$$\begin{aligned} l_n(\cdot; \text{data}) &= \sum_{k=1}^n \{\delta_{1,k}^{(1)} \delta_{2,k}^{(1)} \log S_{1,1}(\alpha, x_k) + \delta_{1,k}^{(1)} \delta_{2,k}^{(2)} \log S_{1,2}(\alpha, x_k) \\ &\quad + \delta_{1,k}^{(1)} \delta_{2,k}^{(3)} \log S_{1,3}(\alpha, x_k) + \delta_{1,k}^{(2)} \delta_{2,k}^{(1)} \log S_{2,1}(\alpha, x_k) \\ &\quad + \delta_{1,k}^{(2)} \delta_{2,k}^{(2)} \log S_{2,2}(\alpha, x_k) + \delta_{1,k}^{(2)} \delta_{2,k}^{(3)} \log S_{2,3}(\alpha, x_k)\} \end{aligned}$$

$$\begin{aligned}
& + \delta_{1,k}^{(3)} \delta_{2,k}^{(1)} \log S_{3,1}(\alpha, x_k) + \delta_{1,k}^{(3)} \delta_{2,k}^{(2)} \log S_{3,2}(\alpha, x_k) \\
& + \delta_{1,k}^{(3)} \delta_{2,k}^{(3)} \log S_{3,3}(\alpha, x_k) \},
\end{aligned} \tag{3}$$

where

$$\begin{aligned}
S_{1,1}(\alpha, x_k) &= P(T_1 \leq u_{1,k}, T_2 \leq u_{2,k}) = 1 - S_1(u_{1,k}) \\
&\quad - S_2(u_{2,k}) + C_\alpha(S_1(u_{1,k}), S_2(u_{2,k})), \\
S_{1,2}(\alpha, x_k) &= P(T_1 \leq u_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) = S_2(u_{2,k}) - S_2(v_{2,k}) \\
&\quad + C_\alpha(S_1(u_{1,k}), S_2(v_{2,k})) - C_\alpha(S_1(u_{1,k}), S_2(u_{2,k})), \\
S_{1,3}(\alpha, x_k) &= P(T_1 \leq u_{1,k}, T_2 > v_{2,k}) = S_2(v_{2,k}) - C_\alpha(S_1(u_{1,k}), S_2(v_{2,k})), \\
S_{2,1}(\alpha, x_k) &= P(u_{1,k} < T_1 \leq v_{1,k}, T_2 \leq u_{2,k}) = S_1(u_{1,k}) - S_1(v_{1,k}) \\
&\quad + C_\alpha(S_1(v_{1,k}), S_2(u_{2,k})) - C_\alpha(S_1(u_{1,k}), S_2(u_{2,k})), \\
S_{2,2}(\alpha, x_k) &= P(u_{1,k} < T_1 \leq v_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) \\
&= C_\alpha(S_1(u_{1,k}), S_2(u_{2,k})) - C_\alpha(S_1(u_{1,k}), S_2(v_{2,k})) \\
&\quad - C_\alpha(S_1(v_{1,k}), S_2(u_{2,k})) + C_\alpha(S_1(v_{1,k}), S_2(v_{2,k})), \\
S_{2,3}(\alpha, x_k) &= P(u_{1,k} < T_1 \leq v_{1,k}, T_2 > v_{2,k}) \\
&= C_\alpha(S_1(u_{1,k}), S_2(v_{2,k})) - C_\alpha(S_1(v_{1,k}), S_2(v_{2,k})), \\
S_{3,1}(\alpha, x_k) &= P(T_1 > v_{1,k}, T_2 \leq u_{2,k}) = S_1(v_{1,k}) - C_\alpha(S_1(v_{1,k}), S_2(u_{2,k})), \\
S_{3,2}(\alpha, x_k) &= P(T_1 > v_{1,k}, u_{2,k} < T_2 \leq v_{2,k}) = C_\alpha(S_1(v_{1,k}), S_2(u_{2,k})) \\
&\quad - C_\alpha(S_1(v_{1,k}), S_2(v_{2,k})), \\
S_{3,3}(\alpha, x_k) &= P(T_1 > v_{1,k}, T_2 > v_{2,k}) = C_\alpha(S_1(v_{1,k}), S_2(v_{2,k})).
\end{aligned}$$

To estimate the association parameter  $\alpha$ , we maximize  $l_n(\cdot; \text{data})$  in (3) by using a two-stage procedure. In the first stage, we estimate two marginal distribution functions. In the second stage, we substitute the estimated marginal distribution functions into the likelihood function to get a pseudo-likelihood function and maximize the pseudo-likelihood function to obtain an estimate for  $\alpha$ .

## 2.2. Spline-based sieve estimation for the marginal distribution functions

We denote the marginal distribution functions of  $T_1$  and  $T_2$  as  $F_1$  and  $F_2$  with the real ones  $F_{0,1}$  and  $F_{0,2}$ , respectively. The likelihoods for the two univariate samples are given by:

$$\begin{aligned} l_{n,r}(\cdot; \text{data}) = & \sum_{k=1}^n \{\delta_{r,k}^{(1)} \log F_r(u_{r,k}) + \delta_{r,k}^{(2)} \log(F_r(v_{r,k}) - F_r(u_{r,k})) \\ & + \delta_{r,k}^{(3)} \log(1 - F_r(v_{r,k}))\}, \end{aligned} \quad (4)$$

for  $r = 1, 2$ . For simplification, we only provide the estimation for  $F_{0,1}$ .

The estimation for  $F_{0,2}$  can be obtained by using the same approach.

Suppose all observation times for  $T_1$  are inside a bounded region  $[L_1, R_1]$ .

Let

$$\mathcal{F}_1 = \{F_1(t) : F_1(t) \geq 0, F_1(t) \leq 1, F_1(t') \leq F_1(t''), t' < t'', \{t, t', t''\} \subset [L_1, R_1]\}.$$

Hence, the NPMLE of  $F_{0,1}$  is defined as

$$\hat{F}_1 = \arg \max_{F_1 \in \mathcal{F}_1} l_{n,1}(\cdot; \text{data}). \quad (5)$$

The main idea of the spline-based sieve method is to find  $\hat{F}_1$  in (5) in a sub-class of  $\mathcal{F}_1$ , but “approximating” to  $\mathcal{F}_1$  asymptotically. In this paper, the spline-based sieve method greatly reduces the unknown variables in the maximization problem, which makes the computation much faster. At the same time, the spline-based sieve method produces smooth and explicit estimates for the marginal distribution functions. In the proposed spline-based sieve estimation procedure, the two marginal

distribution functions are independently estimated by linear combinations of spline basis functions. Thus, maximizing the log likelihood with respect to the unknown functions is converted to maximizing the sieve log likelihood with respect to the unknown spline coefficients subjecting to corresponding inequality constraints.

Consider the normalized  $B$ -spline basis functions of order  $l$  [16]:  $\{N_i^l(t)\}_{i=1}^{p_n}$  are constructed on  $[L_1, R_1]$  with the knot sequence  $\{u_i\}_{i=1}^{p_n+l}$  satisfying  $L_1 = u_1 = \dots = u_l < u_{l+1} < \dots < u_{p_n} < u_{p_n+1} = \dots = u_{p_n+l} = R_1$ , where  $p_n = O(n^\nu)$  for some  $0 < \nu < 1$ . If the number of interior knots of the knot sequence is  $m_n$ , it is obvious that  $p_n = m_n + l$ .

Define

$$\Omega_{n,1} = \left\{ F_{n,1}(t) : F_{n,1}(t) = \sum_{i=1}^{p_n} \beta_i N_i^l(t), 0 \leq \beta_1 \leq \beta_2 \leq \dots \leq \beta_{p_n} \leq 1 \right\}. \quad (6)$$

To obtain the  $B$ -spline-based sieve likelihood with case 2 interval censored data, we replace  $F_1$  in (4) by  $F_{n,1}$  in (6). Then,

$$\begin{aligned} \tilde{l}_{n,1}(\underline{\beta}; \text{data}) = \sum_{k=1}^n & \left\{ \delta_{1,k}^{(1)} \log \sum_{i=1}^{p_n} \beta_i N_i^l(u_{1,k}) + \delta_{1,k}^{(2)} \log \left( \sum_{i=1}^{p_n} \beta_i N_i^l(v_{1,k}) \right) \right. \\ & \left. - \sum_{i=1}^{p_n} \beta_i N_i^l(u_{1,k}) + \delta_{1,k}^{(3)} \log \left( 1 - \sum_{i=1}^{p_n} \beta_i N_i^l(v_{1,k}) \right) \right\}. \end{aligned} \quad (7)$$

Hence, the proposed sieve maximum likelihood estimator (MLE) with the  $B$ -spline basis functions is given by  $\hat{F}_{n,1}$  that maximizes (7) over  $\Omega_n$ . It can be easily argued that  $\Omega_n \subset \mathcal{F}_1$  by the properties of  $B$ -spline basis functions [16]. Then the spline-based sieve MLE  $\hat{F}_{n,1}$  will be inside  $\mathcal{F}_1$ , and hence can be viewed as an approximation for  $\hat{F}_1$  in (5). In Section 3, we show that  $\hat{F}_{n,1}$  converges to the true target function under some regularity conditions.

### 2.3. Estimators of the association parameter and the joint distribution function

Suppose  $\hat{F}_{n,1}$  and  $\hat{F}_{n,2}$  are spline-based sieve MLEs of  $F_{0,1}$  and  $F_{0,2}$ , respectively. By replacing  $S_1$  and  $S_2$  in (3) by  $\hat{S}_1 = 1 - \hat{F}_{n,1}$  and  $\hat{S}_2 = 1 - \hat{F}_{n,2}$ , and maximizing the resulting likelihood function with respect to  $\alpha$ , we can obtain a semiparametric estimator  $\hat{\alpha}$  for the real association parameter  $\alpha_0$ , and hence estimator  $\hat{\tau} = \hat{\alpha} / (\hat{\alpha} + 2)$  for the real Kendall's  $\tau_0$ .

If the Clayton copula model (2) is chosen in the joint distribution, the proposed two-stage semiparametric estimator for the joint distribution function  $C_{\alpha_0}(F_{0,1}(t_1), F_{0,2}(t_2))$  is given by

$$C_{\hat{\alpha}}(\hat{F}_{n,1}(t_1), \hat{F}_{n,2}(t_2)) = (\hat{F}_{n,1}(t_1)^{-\hat{\alpha}} + \hat{F}_{n,2}(t_2)^{-\hat{\alpha}} - 1)^{-1/\hat{\alpha}}. \quad (8)$$

### 3. Asymptotic Properties

In this section, we study the asymptotic properties of the two-stage spline-based sieve estimator  $\hat{\alpha}$ . Once we establish the consistency of the spline-based sieve MLEs of  $F_{0,1}$  and  $F_{0,2}$ , the consistency of two-stage spline-based sieve estimators for the association parameter  $\alpha_0$  and the joint distribution function follow directly.

The following conditions (C1)-(C5) sufficiently guarantee the asymptotic results for  $\hat{F}_{n,1}$  described in Theorem 1:

(C1)  $dF_{0,1}(t)/dt$  has a positive lower bound  $b_0$  on  $[L_1, R_1]$ .

(C2)  $F_{0,1}(t)$  has continuous derivative  $d^p F_{0,1}(t)/dt^p$  on  $[L_1, R_1]$ ,

which implies that  $dF_{0,1}(t)/dt$  has a positive upper bound on  $[L_1, R_1]$ .

(C3) The observation times  $U_1$  and  $V_1$  both follow distributions only taking values within  $[l_1, r_1]$  with  $l_1 > L_1$  and  $r_1 < R_1$ ; there exists a positive number  $\eta$  such that  $P(V_1 - U_1 \geq \eta) = 1$ .

(C4) The probability densities of distributions of  $U_1$  and  $V_1$  have positive lower bounds at every point on  $[l_1, r_1 - \eta]$  and on  $[l_1 + \eta, r_1]$ , respectively.

(C5) The knot sequence  $\{u_i\}_{i=1}^{p_n+1}$  of the  $B$ -spline basis functions  $\{N_i^l\}_{i=1}^{p_n}$ , satisfies that  $\min_i \Delta_i^{(u)} / \max_i \Delta_i^{(u)}$  has positive lower bound, which is not greater than 1, where  $\Delta_i^{(u)} = u_{i+1} - u_i$  for  $i = l, \dots, p_n$ .

Consider a sub-class of  $\Omega_{n,1}$  defined by (6) as

$$\begin{aligned} \Omega'_{n,1} = \left\{ F_{n,1}(t) : F_{n,1}(t) = \sum_{i=1}^{p_n} \beta_i N_i^l(t), \beta_1 \geq 0, \beta_{p_n} \leq 1, \right. \\ \left. \beta_{j+1} - \beta_j \geq \frac{\beta_0 \min_{i_1:l \leq i_1 \leq p_n} \Delta_{i_1}^{(u)}}{l}, j = 1, \dots, p_n - 1 \right\}. \end{aligned}$$

We propose to find the estimator in  $\Omega'_{n,1}$  mainly due to the technique convenience in justifying the asymptotic properties. In computation, the parameter  $b_0$  can be chosen small enough so that  $\Omega_{n,1}$  and  $\Omega'_{n,1}$  would result in the same estimate.

We study the asymptotic properties in the feasible region for observation times:  $[l_1, r_1]$ . Let  $\Omega''_{n,1} = \{F_{n,1}(t) : F_{n,1} \in \Omega_{n,1}, t \in [l_1, r_1]\}$ . Under (C3), the maximization of  $\tilde{l}_{n,1}(\underline{\beta}; \text{data})$  over  $\Omega'_{n,1}$  is actually the maximization of  $\tilde{l}_{n,1}(\underline{\beta}; \text{data})$  over  $\Omega''_{n,1}$ . Throughout the study of the asymptotic properties, we denote  $\hat{F}_{n,1}$  as the maximizer of  $\tilde{l}_{n,1}(\underline{\beta}; \text{data})$  over  $\Omega''_{n,1}$ .

Let  $\|f\|_{L_r(Q)} = (Q|f|^r)^{1/r} = (\int |f|^r dQ)^{1/r}$  be the  $L_r(Q)$ -norm associated with probability measure  $Q$ . In the following,  $L_r(P_{U_1, V_1})$ -norm,  $L_r(P_{U_1})$ -norm, and  $L_r(P_{V_1})$ -norm denote  $L_r$ -norms associated with the joint and marginal probability measures of observation times  $(U_1, V_1)$ . Similarly,  $L_r(P)$ -norm is denoted as the  $L_r$ -norm associated with the joint probability measure  $P$  of observation and event times  $(T_1, U_1, V_1)$ .

Thus, the  $L_2$ -norms distance between  $F_{n,1} \in \Omega''_{n,1}$  and  $F_{0,1}$  is defined as

$$d_1(F_{n,1}, F_{0,1}) = (\|F_{n,1} - F_{0,1}\|_{L_2(P_{U_1})}^2 + \|F_{n,1} - F_{0,1}\|_{L_2(P_{V_1})}^2)^{1/2}.$$

**Theorem 1.** Suppose (C1)-(C5) hold. If  $p_n = O(n^\nu)$  for  $\nu \leq 1/(4p)$ , that is, the number of interior knots of knot sequence  $\{u_i\}_1^{p_n+l}$  is in the order of  $n^\nu$  for  $\nu \leq 1/(4p)$ , then

$$d_1(\hat{F}_{n,1}, F_{0,1}) = O_p(n^{-\min\{pv, (1-\nu)/3\}}).$$

Theorem 1 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  $p \geq 1$  and  $\nu = 1/(4p)$ . If  $p = 1$ , then  $\nu = 1/4$  and the number of interior knots could be chosen as  $n^{1/4}$ . This choice 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  $n^{1/3}$ . For moderate sample sizes, say  $n = 200$  or  $400$ , 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  $p_n = m_n + 4$  in our computation.

**Remark 1.** With the similar regularity conditions as (C1)-(C5), the spline-based sieve MLE  $\hat{F}_{n,2}$  for  $F_{0,2}$  has the same convergence rate as described in Theorem 1.

The following two conditions are used to prove the consistency of  $\hat{\alpha}$  :

(D1) Let  $\lambda = (u_1, v_1, u_2, v_2)$  and  $\delta = (\delta_1^{(1)}, \delta_1^{(2)}, \delta_1^{(3)}, \delta_2^{(1)}, \delta_2^{(2)}, \delta_2^{(3)})$ .

Suppose  $l_1(\alpha, S_1, S_2; \lambda, \delta)$  be the log-likelihood defined in (3) for observed data of size 1. Then  $\partial^3 l_1(\alpha, S_1, S_2; \lambda, \delta) / \partial \alpha^3$ ,  $\partial^3 l_1(\alpha, S_1, S_2; \lambda, \delta) / \partial \alpha^2 \partial S_1$ , and  $\partial^3 l_1(\alpha, S_1, S_2; \lambda, \delta) / \partial \alpha^2 \partial S_2$  are continuous and bounded for a compact neighborhood of  $\alpha_0$ .

(D2) Suppose  $g(\alpha; \lambda, \delta)$  is the probability density function of  $(U_1, V_1, U_2, V_2, \Delta)$ . Then  $-\sum_{\delta} \int [\partial l_1(\alpha_0, S_1, S_2; \lambda, \delta) / \partial \alpha]^2 g(\alpha_0; \lambda, \delta) d\lambda$  is negative, where the summation is over all possible  $\delta$  for  $\alpha_0$ .

**Theorem 2.** Under (C1)-(C5), 5 similar conditions as (C1)-(C5) for establishing the asymptotic properties of  $\hat{F}_{n,2}$ , and (D1), (D2).  $\hat{\alpha}$  described in Subsection 2.3 is consistent estimator, that is,

$$\hat{\alpha} \rightarrow_p \alpha_0.$$

**Remark 2.** By continuous mapping theorem, we can easily see the estimated joint distribution function described in Subsection 2.3 is a consistent estimator.

## 4. Numerical Experiments

### 4.1. Computation of the proposed spline-based sieve estimators

Given  $p_n$ , the proposed sieve estimation problem described in Subsection 2.2 becomes a restricted parametric maximum likelihood estimation problem with respect to the coefficients of the  $B$ -spline basis functions. In this section, we apply the generalized gradient projection algorithm [11] to compute the spline-based sieve MLEs for two marginal distribution functions.

The standard error of  $\hat{\alpha}$  can be estimated by bootstrap method. We draw bootstrap samples of size  $n$  with replacement from the observed data  $\{u_{1,k}, v_{1,k}, \delta_{1,k}^{(1)}, \delta_{1,k}^{(2)}, \delta_{1,k}^{(3)} : k = 1, 2, \dots, n\}$  and  $\{u_{2,k}, v_{2,k}, \delta_{2,k}^{(1)}, \delta_{2,k}^{(2)}, \delta_{2,k}^{(3)} : k = 1, 2, \dots, n\}$  independently  $M$  times, where  $M$  is a fixed integer. Then, we have an estimate of  $\alpha_0$  from each of the  $M$  bootstrap samples, the standard deviation of these  $M$  estimates is the bootstrap standard error of  $\hat{\alpha}$ . We estimate the standard error of  $\hat{\tau}$  from the bootstrap standard error of  $\hat{\alpha}$  by using delta method.

#### 4.2. Simulation studies

In the simulation studies, we first compare the proposed spline-based sieve estimators  $\hat{\alpha}$  and  $\hat{\tau}$  to the two-stage semiparametric maximum pseudo-likelihood estimators  $\tilde{\alpha}$  and  $\tilde{\tau}$  studied by Sun et al. [19]. Then, we compute the estimated joint distribution function proposed in Subsection 2.3.

We compute  $\tilde{\alpha}$  by using the ICM algorithm, which is justified by Zhang and Jamshidian [27] to be the fastest algorithm for compute the NPMLE for the distribution function with case 2 interval censored data.

We simulate the data with combinations of  $\alpha_0$  values ( $\alpha_0 = 1, 2$ ) and sample sizes ( $n = 200, 400$ ). Under each of these four settings, we conduct the Monte-Carlo simulation with 500 repetitions and 200 bootstrapping samples. We use the cubic ( $l = 4$ )  $B$ -spline basis functions in the proposed spline-based sieve method.  $(T_1, T_2)$  are generated from the Clayton copula with two marginal distributions being exponential with the rate parameter 0.5. We set  $\Pr(T_i \geq 5) < 0.1$  for  $i = 1, 2$  and  $[L_1, R_1] \times [L_2, R_2]$  to be  $[0, 5] \times [0, 5]$ . For  $i = 1, 2$ ,  $U_i$  is generated from uniform distribution on  $[0.0201, 3]$  and  $V_i = U_i + W_i$  with  $W$  being generated from uniform distribution on  $[0.05, 1.7698]$ . Thus, we have  $\Pr(0 < T_i < 0.0201) = \Pr(4.7698 < T_i < 5) = 0.01$ . The observation region

$[l_1, r_1] \times [l_2, r_2] = [0.0201, 4.7698] \times [0.0201, 4.7698]$  is inside  $[0, 5] \times [0, 5]$ .

Based on the discussion on the knots selection in Section 3, we choose 5 and 6 as the numbers of interior knots for sample size 200 and 400, respectively. Two end knots of the knot sequences are chosen to be 0 and 5, the interior knots are allocated at the  $w / (m_n + 1)$  quantiles,

$w = 1, \dots, m_n$  of  $\{u_{i,k}\}_{k:\delta_{i,k}^{(3)}=0} \cup \{v_{i,k}\}_{k:\delta_{i,k}^{(1)}=0}$ , where  $\{u_{i,k}\}_{k=1}^n$  and

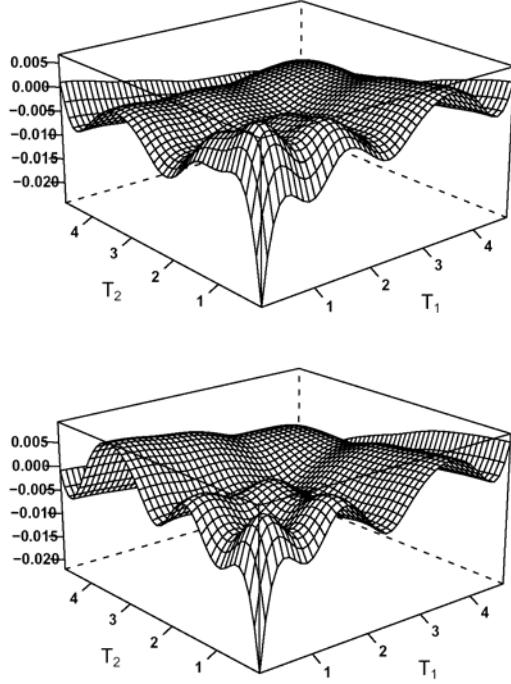
$\{v_{i,k}\}_{k=1}^n$  are samples of size  $n = 200$  or  $400$  from  $U_i$  and  $V_i$  for  $i = 1, 2$ , respectively.

Table 1 compares the spline-based method with the method proposed by Sun et al. [19] for estimating association parameter  $\alpha_0$  and Kendall's  $\tau_0$  in terms of bias, sample standard deviation (SD), bootstrap standard error (SE), and bootstrap coverage probability (CP). Our proposed spline-based sieve method works at least as good as the method in Sun et al. [19] in terms of the estimation bias, or works a little better in terms of the estimation standard error. Moreover, the computing speed for the proposed spline-based method is at least 20 times faster than the speed for computing the estimated association parameter in Sun et al. [19]. It is worthwhile to note that the coverage probability (CP) values in all settings for the proposed spline-based sieve estimators are very close to 95%, which give some numerical evidence for the asymptotic normality for  $\hat{\alpha}$  and  $\hat{\tau}$ . This observation can be used to construct some hypothesis tests based on asymptotic normality.

**Table 1.** The comparisons between the proposed spline-based sieve estimators ( $\hat{\alpha}$ ,  $\hat{\tau}$ ) and the two-stage estimators ( $\tilde{\alpha}$ ,  $\tilde{\tau}$ ) in Sun et al. [19] in their finite sample performances

$\alpha_0 = 1, n = 200$				$\alpha_0 = 1, n = 400$				
	Bias	SD	SE	CP	Bias	SD	SE	CP
$\hat{\alpha}$	0.0909	0.2526	0.2891	97%	0.0088	0.1736	0.1866	95%
$\tilde{\alpha}$	0.1534	0.6100	0.4949	96%	0.0446	0.2383	0.3727	98%
$\hat{\tau}$	0.0152	0.0549	0.0602	97%	-0.0003	0.0386	0.0411	95%
$\tilde{\tau}$	0.0186	0.0787	0.1026	96%	-0.0147	0.0584	0.0874	100%
$\alpha_0 = 2, n = 200$				$\alpha_0 = 2, n = 400$				
	Bias	SD	SE	CP	Bias	SD	SE	CP
$\hat{\alpha}$	0.0961	0.4257	0.4937	98%	0.0555	0.3476	0.3288	92%
$\tilde{\alpha}$	-0.0190	0.4650	0.7055	100%	-0.0360	0.4344	0.6475	98%
$\hat{\tau}$	0.0066	0.0506	0.0585	96%	0.0033	0.0424	0.0400	94%
$\tilde{\tau}$	-0.0091	0.0586	0.0903	100%	-0.0115	0.0654	0.0891	100%

Figure 1 graphically presents the bias of the estimated joint distribution function from the same Monte-Carlo simulation discussed above for sample size 200. We can see the bias inside region  $[0.1, 4.7] \times [0.1, 4.7]$  is promising.



**Figure 1.** The bias of estimated joint survival function in the region  $[0.1, 4.7] \times [0.1, 4.7]$  when  $\alpha_0 = 1$  (top) and  $\alpha_0 = 2$  (bottom) with sample size  $n = 200$ .

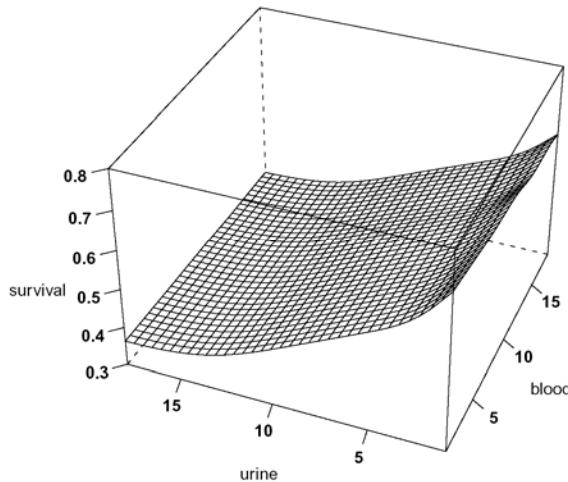
In Table 2, we calculate the average estimation bias and the average square root of the mean square error for 2209 values of  $(s_1, s_2)$ , where both  $s_1$  and  $s_2$  uniformly take 47 values from 0.1 to 4.7.

**Table 2.** The overall bias and square root of mean square error of estimated joint distribution function when sample size  $n = 200$

	Bias	MSE <sup>1/2</sup>
$\alpha_0 = 1$	$-2.03e-3$	$3.80e-2$
$\alpha_0 = 2$	$2.60e-4$	$4.23e-2$

#### 4.3. A real example

The proposed sieve semiparametric method is illustrated by the data from the AIDS clinical trial introduced in Section 1 [7]. In computation, we assume that the joint distribution of the real data follows the Clayton copula and the association parameter is  $\alpha$ . We use cubic spline basis functions and perform knots selection as the way in simulation studies. We also draw 200 independent bootstrap samples. For the 93 patients with high CD4 counts, the estimated association parameter between times to CMV shedding in the blood and in the urine is 1.62 with bootstrap standard error 0.700. For the 111 patients with low CD4 counts, the estimated association parameter is 1.71 with bootstrap standard error 0.598. As mentioned in simulation studies, we can test whether blood shedding and urine shedding are correlated by testing  $\alpha = 0$  based on asymptotic normality. The  $p$ -values for testing  $\alpha = 0$  for high and low CD4 counts patients are 0.01039 and 0.00215, respectively. The  $p$ -value for testing whether association parameters for two groups of patients are equal, is 0.46093. This implies neither of the two association parameters is equal to 0, and the two association parameters are not significantly different. Figure 2 sketches the estimated joint survival function of two event times in months in the region  $[1, 18] \times [1, 18]$ .



**Figure 2.** The estimated joint survival function of times in months to CMV blood shedding and urine shedding in the region  $[1, 18] \times [1, 18]$ .

Table 3 shows the values of the estimated joint distribution at some time points, the first column represents time to blood shedding in months and the first row represents time to urine shedding in months. We can see that the estimated survival function of time to blood shedding decreases very slow over time. It may be due to a large portion of right censored data being observed.

**Table 3.** The estimated joint survival function of times in months to CMV blood shedding and urine shedding at some time points

		Urine shedding						
		Months	3	6	9	12	15	18
Blood shedding	3	0.6007	0.5276	0.4635	0.4037	0.3701	0.3621	
	6	0.5913	0.5212	0.4592	0.4009	0.3678	0.3600	
	9	0.5858	0.5174	0.4566	0.3992	0.3665	0.3588	
	12	0.5748	0.5098	0.4513	0.3956	0.3638	0.3562	
	15	0.5660	0.5036	0.4471	0.3927	0.3615	0.3541	
	18	0.5587	0.4985	0.4435	0.3903	0.3596	0.3523	

## 5. Discussion

Bivariate case 2 interval censored data arises in many applications. The main interest is to estimate the association parameter and the joint distribution of the two event times. In this paper, we propose a two-stage spline-based sieve semiparametric maximum pseudo-likelihood estimator for the association parameter, and hence, a smooth and explicit estimator for the joint distribution of two case 2 interval censored data. We have proved that the proposed estimators are consistent and simulation studies give some evidences for the asymptotic normality for the estimated association parameter. Simulation studies also show that the finite sample performances of the estimated association parameter is at least as good as the performances of Sun et al. [19]'s estimator, but with much shorter computing time compared to its counterpart, and that the estimated joint distribution is generally satisfactory with sample size 200. Our proposed method is also illustrated by an AIDS clinical trial study, with a result that two event times of interest seem to be dependent based on the observed data.

If an easy-to-compute standard error of the proposed estimated association parameter could be discovered based on asymptotic normality theory, the computing time in this paper will be further reduced. We can try to use the spline-based sieve estimation method to solve problems with bivariate case 1 interval censored data or bivariate right censored data.

## References

- [1] R. A. Betensky and D. M. Finkelstein, A non-parametric maximum likelihood estimator for bivariate interval censored data, *Stat. Med.* 18 (1999), 3089-3100.
- [2] C. De Boor, *A Practical Guide to Splines*, Springer, New York, 2001.
- [3] I. D. Diamond, J. W. McDonald and I. H. Shah, Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan, *Demography* 23 (1986), 607-620.
- [4] I. D. Diamond and J. W. McDonald, Analysis of current status data, In *Demographic Applications of Event History Analysis* (J. Trussell, R. Hankinson and J. Tilton, eds.), Oxford University Press, (1991), 231-252.
- [5] D. M. Finkelstein and R. A. Wolfe, A semiparametric model for regression analysis of interval-censored failure time data, *Biometrics* 41 (1985), 933-945.
- [6] R. Gentleman and A. C. Vandal, Computational algorithms for censored-data problems using intersection graphs, *J. Comput. Graph. Statist.* 10 (1985), 403-421.
- [7] W. B. Goggins and D. M. Finkelstein, A proportional hazards model for multivariate interval-censored failure time data, *Biometrics* 56 (1985), 940-943.
- [8] P. Groeneboom and J. A. Wellner, *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhäuser, Basel, 1992.
- [9] D. G. Hoel and H. E. Walburg, Statistical analysis of survival experiments, *Journal of the National Cancer Institute* 49 (1972), 361-372.
- [10] J. Huang and J. A. Wellner, Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1, *Statist. Neer.* 49 (1995), 153-163.
- [11] M. Jamshidian, On algorithms for restricted maximum likelihood estimation, *Comp. Statist. Data Anal.* 45 (2004), 137-157.
- [12] N. P. Jewell, H. Malani and E. Vittinghoff, Nonparametric estimation for a form of doubly censored data with application to two problems in AIDS, *J. Amer. Statist. Assoc.* 89 (1994), 7-18.
- [13] M. Lu, Y. Zhang and J. Huang, Estimation of the mean function with panel count data using monotone polynomial splines, *Biometrika* 94 (2007), 705-718.

- [14] M. H. Maathuis, Reduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data, *J. Comput. Graph. Statist.* 14 (2005), 352-362.
- [15] R. B. Nelsen, *An Introduction to Copulas*, Springer, New York, 2006.
- [16] L. Schumaker, *Spline Functions: Basic Theory*, Wiley, New York, 1981.
- [17] S. C. Shiboski and N. P. Jewell, Statistical analysis of the time dependence of HIV infectivity based on partner study data, *J. Amer. Statist. Assoc.* 87 (1992), 360-372.
- [18] J. H. Shih and T. A. Louis, Inference on the association parameter in copula models for bivariate survival data, *Biometrics* 51 (1995), 1384-1399.
- [19] L. Sun, L. Wang and J. Sun, Estimation of the association for bivariate interval censored failure time data, *Scand. J. Statist.* 33 (2006), 637-649.
- [20] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer, New York, 1996.
- [21] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
- [22] W. Wang and A. A. Ding, On assessing the association for bivariate current status data, *Biometrika* 87 (2000), 879-893.
- [23] J. A. Wellner, Interval censoring case 2: Alternative hypotheses, In *Analysis of Censored Data* (H. L. Koul and J. V. Deshpande, eds.) (1995), 271-291.
- [24] J. A. Wellner and Y. Zhang, Likelihood-based semiparametric estimation methods for panel count data with covariates, *Ann. Statist.* 28 (2007), 2106-2142.
- [25] G. Y. Wong and Q. Yu, Generalized MLE of a joint distribution function with multivariate interval-censored data, *J. Multivariate Anal.* 69 (1999), 155-166.
- [26] Y. Wu and Y. Zhang, Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data, Tech. Rep., Department of Biostatistics, University of Iowa, USA, 2010.
- [27] Y. Zhang and M. Jamshidian, On algorithms for the non-parametric maximum likelihood estimator of the failure function with censored data, *J. Comput. Graph. Statist.* 13 (2004), 123-140.
- [28] Y. Zhang, L. Hua and J. Huang, A spline-based semiparametric maximum likelihood estimation for the cox model with interval-censored data, *Scand. J. Statist.* 37 (2010), 338-354.

## Appendix

**Proof of Theorem 1.** In this proof, 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  $f(X)$ .

Before deriving the convergence rate, we need to show that the sieve estimator  $\hat{F}_n$  is consistent. This can be accomplished by verifying the conditions of Theorem 5.7 of van der Vaart [21].

For  $\{t, t', t''\} \subset [l_1, r_1]$  with  $t'' - t' \geq h$ , and  $b_1 > 0, b_2 > 0$ , and  $b_3 > 0$ , we define  $\Omega_1$  by  $\Omega_1 = \{F_1(t) : F_1(t) \text{ is nondecreasing; } F_1(t') \geq b_1, 1 - F_1(t'') \geq b_2, F_1(t'') - F_1(t') \geq b_3\}$ . If  $b_1, b_2$ , and  $b_3$  are small enough, then under (C1) and (C5), it can be shown that  $F_{0,1} \in \Omega_1$  and  $\Omega''_{n,1} \subset \Omega_1$ .

Let  $\mathcal{L}_1 = \{l(F_1) : F_1 \in \Omega_1\}$ , where  $l(F_1) = \delta_1^{(1)} \log F_1(u) + \delta_1^{(2)} [\log F_1(v) - \log F_1(u)] + \delta_1^{(3)} \log[1 - F_1(v)]$ , with  $\delta_1^{(1)} = 1_{[T_1 \leq u]}, \delta_1^{(2)} = 1_{[u < T_1 \leq v]}$ , and  $\delta_1^{(3)} = 1_{[T_1 > v]}$ . Denote  $\mathbb{M}(F_1) = Pl(F_1)$  and  $\mathbb{M}_n(F_1) = \mathbb{P}_n l(F_1)$ , hence  $\mathbb{M}_n(F_1) - \mathbb{M}(F_1) = (\mathbb{P}_n - P)l(F_1)$ .

By Theorem 2.7.5 of van der Vaart and Wellner [20], there exist brackets  $[F_i^l, F_i^u]$ ,  $i = 1, \dots, [(1/\epsilon)^K]$  to cover  $\Omega_1$  and satisfy  $P_{U_1}|F_i^u - F_i^l| \leq \epsilon$ ; and there exist brackets  $[G_j^l, G_j^u]$ ,  $j = 1, \dots, [(1/\epsilon)^K]$  to cover  $\Omega_1$  and satisfy  $P_{V_1}|G_j^u - G_j^l| \leq \epsilon$ .

Hence, we can construct brackets  $[l_{i,j}^l, l_{i,j}^u]$ ,  $i = 1, 2, \dots, [(1/\epsilon)^K]$ ,  $j = 1, 2, \dots, [(1/\epsilon)^K]$  such that for any  $l(F_1) \in \mathcal{L}_1$ , there exist  $i, j$  such that  $l_{i,j}^l \leq l(F_1) \leq l_{i,j}^u$ , where

$$l_{i,j}^l = \delta_1^{(1)} \log F_i^l + \delta_1^{(2)} [\log G_j^l - \log F_i^u] + \delta_1^{(3)} [1 - \log G_j^u],$$

and

$$l_{i,j}^u = \delta_1^{(1)} \log F_i^u + \delta_1^{(2)} [\log G_j^u - \log F_i^l] + \delta_1^{(3)} [1 - \log G_j^l].$$

Using Taylor's expansion along with the property of  $\Omega_1$ , we can demonstrate that  $P|l_{i,j}^u - l_{i,j}^l| \leq K\epsilon$  for all  $i, j$ . So the  $\epsilon$ -bracket number for  $\mathcal{L}_1$  with  $L_1(P)$ -norm is less than  $(1/\epsilon)^K$ , which is bounded. As  $N(\epsilon, \mathcal{L}_1, L_1(P)) \leq N_{[]} (2\epsilon, \mathcal{L}_1, L_1(P))$ ,  $\mathcal{L}_1$  is Glivenko-Cantelli by Theorem 2.4.3 of van der Vaart and Wellner [20]. Therefore,

$$\sup_{F_1 \in \Omega_1} |\mathbb{M}_n(F_1) - \mathbb{M}(F_1)| \rightarrow_P 0.$$

For any  $F_1 \in \Omega_1$ , we have

$$\begin{aligned} \mathbb{M}(F_{0,1}) - \mathbb{M}(F_1) &= P\{l(F_{0,1}) - l(F_1)\} \\ &= P\left\{\delta_1^{(1)} \log \frac{F_{0,1}(u)}{F_1(u)} + \delta_1^{(2)} \log \frac{F_{0,1}(v) - F_{0,1}(u)}{F_1(v) - F_1(u)} + \delta_1^{(3)} \log \frac{1 - F_{0,1}(v)}{1 - F_1(v)}\right\} \\ &= P_{U_1, V_1} \left\{ F_{0,1}(u) \log \frac{F_{0,1}(u)}{F_1(u)} + (F_{0,1}(v) - F_{0,1}(u)) \log \frac{F_{0,1}(v) - F_{0,1}(u)}{F_1(v) - F_1(u)} \right. \\ &\quad \left. + (1 - F_{0,1}(v)) \log \frac{1 - F_{0,1}(v)}{1 - F_1(v)} \right\} \\ &= P_{U_1, V_1} \left\{ F_1(u)m\left(\frac{F_{0,1}(u)}{F_1(u)}\right) + (F_1(v) - F_1(u))m\left(\frac{F_{0,1}(v) - F_{0,1}(u)}{F_1(v) - F_1(u)}\right) \right. \\ &\quad \left. + (1 - F_1(v))m\left(\frac{1 - F_{0,1}(v)}{1 - F_1(v)}\right) \right\}, \end{aligned}$$

where  $m(x) = x \log x - x + 1 \geq (x-1)^2 / 4$  for  $0 \leq x \leq 5$ . Then by the definition of  $\Omega_1$ , we can argue that  $P_{U_1, V_1}\{F_1(u)m\left(\frac{F_{0,1}(u)}{F_1(u)}\right)\} \geq KP_{U_1}\{F_{0,1}(u) - F_1(u)\}^2$  and  $P_{U_1, V_1}\{(1 - F_1(v))m\left(\frac{1 - F_{0,1}(v)}{1 - F_1(v)}\right)\} \geq KP_{V_1}\{F_{0,1}(v) - F_1(v)\}^2$ . Hence,

$$\mathbb{M}(F_{0,1}) - \mathbb{M}(F_1) \geq K[P_{U_1}\{F_{0,1} - F_1\}^2 + P_{V_1}\{F_{0,1} - F_1\}^2] = Kd(F_{0,1}, F_1)^2.$$

$$\text{Then } \sup_{F_1: d(F_1, F_{0,1}) \geq \epsilon, F_1 \in \Omega_1} \mathbb{M}(F_1) \leq \mathbb{M}(F_{0,1}) - K\epsilon^2 < \mathbb{M}(F_{0,1}).$$

Under (C2), by Jackson type theorem on page 149 of De Boor [2], for  $l \geq p + 2$ , there exists  $F_{n,1} \in \Omega''_{n,1}$ , such that

$$\|F_{n,1} - F_{0,1}\|_\infty \leq Kp_n^{-p} = O(n^{-pv}).$$

Let  $\mathcal{L}_{n,1} = \{l(F_{n,1}) : F_{n,1} \in \Omega''_{n,1}, \|F_{n,1} - F_{0,1}\|_\infty \geq K(n^{-pv})\}$ . Using similar argument to that for showing  $\mathcal{L}_1$  is a Glivenko-Cantelli, we can prove the  $\epsilon$ -bracketing number for  $\mathcal{L}_{n,1}$  associated with  $L_2(P)$ -norm is finite, hence  $\mathcal{L}_{n,1}$  is a  $P$ -Donsker. By the dominated convergence theorem, it is easy to see that  $P\{l(F_{n,1}) - l(F_{0,1})\} \rightarrow 0$  as  $n \rightarrow \infty$ , hence

$$(\mathbb{P}_n - P)\{l(F_{n,1}) - l(F_{0,1})\} = O_p(n^{-1/2}),$$

by the fact that both  $l(F_{n,1})$  and  $l(F_{0,1})$  are in  $\mathcal{L}_1$  and the relationship between Donsker and asymptotic equicontinuity given by Corollary 2.3.12 of van der Vaart and Wellner [20]. By the dominated convergence theorem again,  $P\{l(F_{n,1}) - l(F_{0,1})\} > -o(1)$ . Therefore,

$$\begin{aligned} \mathbb{M}_n(\hat{F}_{n,1}) - \mathbb{M}_n(F_{0,1}) &= \mathbb{M}_n(\hat{F}_{n,1}) - \mathbb{M}_n(F_{n,1}) + \mathbb{M}_n(F_{n,1}) - \mathbb{M}_n(F_{0,1}) \\ &\geq \mathbb{P}_n l(F_{n,1}) - \mathbb{P}_n l(F_{0,1}) \\ &= (\mathbb{P}_n - P)\{l(F_{n,1}) - l(F_{0,1})\} + P\{l(F_{n,1}) - l(F_{0,1})\} \\ &\geq o_p(n^{-1/2}) - o(1) = -o_p(1). \end{aligned}$$

This completes the proof for  $d(\hat{F}_{n,1}, F_{0,1}) \rightarrow_p 0$ .

Next, we verify the conditions of Theorem 3.4.1 of van der Vaart and Wellner [20] to derive the convergence rate. First, we already showed in the proof of consistency that  $P\{l(F_{0,1}) - l(F_{n,1})\} \geq Kd^2(F_{0,1}, F_{n,1})$ .

Second, in the proof of consistency, we know that  $\mathbb{M}_n(\hat{F}_{n,1}) - \mathbb{M}_n(F_{0,1}) \geq I_{1,n} + I_{2,n}$ , where  $I_{1,n} = (\mathbb{P}_n - P)\{l(F_{n,1}) - l(F_{0,1})\}$  and  $I_{2,n} = P\{l(F_{n,1}) - l(F_{0,1})\}$ . If  $v \leq 1/(4p)$ , then by  $I_{1,n} = o_p(n^{-1/2})$ , we have

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

Using the fact that  $m(x) = x \log x - x + 1 \leq (x-1)^2$  for  $x$  close to 1, it is easy to see that  $P\{l(F_{0,1}) - l(F_{n,1})\} \leq K\|F_{0,1} - F_{n,1}\|_\infty^2 = O(n^{-2pv})$ , which means that

$$I_{2,n} \geq -O(n^{-2pv}).$$

Thus, we conclude that

$$\mathbb{M}_n(\hat{F}_{n,1}) - \mathbb{M}_n(F_{0,1}) \geq -O_p(n^{-2pv}) = -O_p(n^{-2\min\{pv, (1-v)/3\}}).$$

Let  $\mathcal{L}_{n,\delta} = \{l(F_{n,1}) - l(F_{0,1}) : F_{n,1} \in \Omega''_{n,1}, d(F_{n,1}, F_{0,1}) \leq \delta\}$ . By Lemma 8.6 of Wu and Zhang [26], for small  $\epsilon > 0$ , there exist brackets  $[f_i^l, f_i^u]$ ,  $i = 1, 2, \dots, [(1/\epsilon)^{Kp_n}]$  to cover  $\Omega''_{n,1}$  and satisfy  $\|f_i^u - f_i^l\|_\infty \leq \epsilon$ . Using the similar argument as that in the proof of consistency, we obtain that  $\log N[\{\epsilon, \mathcal{L}_{n,\delta}, L_\infty\}]$  is bounded by  $Kp_n \log(1/\epsilon)$ , and hence by the fact that the  $L_2(P)$ -norm is smaller than the  $L_\infty$ -norm,  $\log N[\{\epsilon, \mathcal{L}_{n,\delta}, L_2(P)\}]$  is bounded by  $Kp_n \log(1/\epsilon)$ . This leads to

$$J[\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\}] = \int_0^\delta \sqrt{(1 + \log N[\{\epsilon, \mathcal{L}_{n,\delta}, L_2(P)\}])} d\epsilon \leq Kp_n^{1/2} \delta^{1/2}.$$

Lemma 7.1 of Wellner and Zhang [24] indicates that under some conditions, the  $L_\infty$ -norm can be bounded by the  $L_2$ -norm. Therefore, under (C1), (C2), and (C4), if  $d(F_{n,1}, F_{0,1}) \leq \delta$  for very small  $\delta > 0$ , then  $F_{n,1}$  and  $F_{0,1}$  are very close to each other on  $[l_1, r_1]$ . Thus, for

$l(F_{n,1}) - l(F_{0,1})$  in  $\mathcal{L}_{n,\delta}$ , we can easily show that  $P\{l(F_{n,1}) - l(F_{0,1})\} \leq K\delta^2$ . The structure of the log-likelihood guarantee that uniform boundedness of  $l(F_{n,1})$ . The key function  $\phi_n(\delta)$  in Theorem 3.4.1 of van der Vaart and Wellner [20] is given by  $\psi_n(\delta) = p_n^{1/2}\delta^{1/2} + p_n / (\delta n^{1/2})$ . Note that

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

Therefore, if  $pv < (1-v)/3$ ,  $n^{2pv}\psi_n(1/n^{pv})$ . This implies that if we choose  $r_n = \min\{pv, (1-v)/3\}$ , then  $r_n^2\psi_n(1/r_n) \leq n^{1/2}$  and  $\mathbb{M}_n(\hat{F}_{n,1}) - \mathbb{M}_n(F_{0,1}) \geq -O_p(r_n^{-2})$ . Hence,

$$r_n d_1(\hat{F}_{n,1}, F_{0,1}) = O_p(1). \quad \square$$

**Proof of Theorem 2.** Theorem 1 states that  $d_1(\hat{F}_{n,1}, F_{0,1}) = O_p(n^{-\min\{pv, (1-v)/3\}})$ . As discussed in the proof of Theorem 1, under (C1), (C2), and (C4), Theorem 7.1 of Wellner and Zhang [24] implies that  $\sup_{t \in [l_1, r_1]} |\hat{F}_{n,1}(t) - F_{0,1}(t)| \rightarrow_p 0$ . Similarly, by  $d_2(\hat{F}_{n,2}, F_{0,2}) = O_p(n^{-\min\{pv, (1-v)/3\}})$  with  $d_2(F_{n,2}, F_{0,2}) = (\|F_{n,2} - F_{0,2}\|_{L_2(P_{U_2})}^2 + \|F_{n,2} - F_{0,2}\|_{L_2(P_{V_2})}^2)^{1/2}$ , we have  $\sup_{t \in [l_2, r_2]} |\hat{F}_{n,2}(t) - F_{0,2}(t)| \rightarrow_p 0$ . For  $i = 1, 2$ , let  $\hat{S}_{n,i} = 1 - \hat{F}_{n,i}$  and  $S_{0,i} = 1 - F_{0,i}$ . Then,  $\sup_{t \in [l_i, r_i]} |\hat{S}_{n,i}(t) - S_{0,i}(t)| \rightarrow_p 0$ . Next, under (D1) and (D2), by following the proof of Theorem 1 of Sun et al. [19], we can easily argue that  $\hat{\alpha}$  is consistent estimator of  $\alpha_0$ .

■