

## Nonparametric estimation of the cross ratio function

Peer-reviewed author version

ABRAMS, Steven; JANSSEN, Paul; SWANEPOEL, Jan & VERAVERBEKE, Noel  
(2020) Nonparametric estimation of the cross ratio function. In: Annals of the Institute  
of Statistical Mathematics, 72 (3), p. 771-801.

DOI: 10.1007/s10463-019-00709-3

Handle: <http://hdl.handle.net/1942/28357>

## Nonparametric estimation of the cross ratio function

Steven Abrams · Paul Janssen · Jan  
Swanepoel · Noël Veraverbeke

Received: date / Revised: date

**Abstract** The cross ratio function (CRF) is a commonly used tool to describe local dependence between two correlated variables. Being a ratio of conditional hazards, the CRF can be rewritten in terms of (first and second derivatives of) the survival copula of these variables. Bernstein estimators for (the derivatives of) this survival copula are used to define a nonparametric estimator of the cross ratio and asymptotic normality thereof is established. We consider simulations to study the finite sample performance of our estimator for copulas with different types of local dependency. A real dataset is used to investigate the dependence between food expenditure and net income. The estimated CRF reveals that families with a low net income relative to the mean net income will spend less money to buy food compared to families with larger net incomes.

---

S. Abrams (✉)

Hasselt University, Center for Statistics, Department of Mathematics, Campus Diepenbeek,  
Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium  
E-mail: steven.abrams@uhasselt.be

University of Antwerp, Department of Epidemiology and Social Medicine, Campus Drie  
Eiken, Universiteitsplein 1, 2610, Wilrijk, Belgium  
E-mail: steven.abrams@uantwerpen.be

P. Janssen

Hasselt University, Center for Statistics, Department of Mathematics, Campus Diepenbeek,  
Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium

J. Swanepoel

North-West University, School of Computer Science, Statistics and Mathematics Potchef-  
stroom Campus, Hoffman Street 11, 2520 Potchefstroom, South Africa

N. Veraverbeke

Hasselt University, Center for Statistics, Department of Mathematics, Campus Diepenbeek,  
Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium  
North-West University, School of Computer Science, Statistics and Mathematics Potchef-  
stroom Campus, Hoffman Street 11, 2520 Potchefstroom, South Africa

**The online version of this article contains supplementary material.**

This dependence, however, disappears when the net income is large compared to the mean income.

**Keywords** Asymptotic distribution · Bernstein estimation · Copula · Cross ratio function · Hazard rate

## 1 Introduction

To analyse familial tendency in disease incidence, Clayton (1978) defined the cross ratio function

$$\theta(t_1, t_2) = \frac{\lambda(t_1 \mid T_2 = t_2)}{\lambda(t_1 \mid T_2 > t_2)} \quad (1)$$

where, for a pair  $(T_1, T_2)$  of absolutely continuous variables, e.g., infection times in infectious disease epidemiology, failure times in survival analysis, or lifetimes in reliability theory,  $\lambda(\cdot \mid T_2 = t_2)$  and  $\lambda(\cdot \mid T_2 > t_2)$  are the conditional hazard rate functions for  $T_1$  given  $T_2 = t_2$  and  $T_2 > t_2$  respectively. The function  $\theta(t_1, t_2)$  is symmetric in the sense that it is also equal to the ratio of the hazard of  $T_2$  at  $t_2$  given  $T_1 = t_1$ , to the hazard of  $T_2$  at  $t_2$  given  $T_1 > t_1$ . Independence between  $T_1$  and  $T_2$  corresponds to  $\theta(t_1, t_2) \equiv 1$  and positive association corresponds to  $\theta(t_1, t_2) > 1$ . See also Oakes (1982, 1986, 1989) for a further discussion thereon and for details on semiparametric inference.

In the disease incidence context considered in Clayton (1978),  $T_1$  is the time at which the son experiences the disease of interest and  $T_2$  the time at which the father experiences this disease. Clayton (1978) considers the case  $\theta(t_1, t_2) \equiv \theta + 1$ , for which  $\theta = 0$  corresponds to independence and  $\theta > 0$  implies positive association. In terms of copulas (see Section 2), this model corresponds to the bivariate Clayton copula  $C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$ ,  $\theta \geq 0$ .

Nowadays the cross ratio function is a commonly used measure to describe local dependence between correlated failure times (Section 4.1.4 in Duchateau and Janssen, 2008; Section 6.4 in Wienke, 2010). Classical measures of dependence are global measures (e.g., Kendall's tau). The cross ratio function, being a local dependence measure, can however detect association characteristics that cannot be captured by any global dependence measure. In this context it is interesting to note that the cross ratio can be rewritten in terms of a local version of Kendall's  $\tau$ , has a local odds ratio interpretation, and that it can be used as a diagnostic tool for testing independence; see Oakes (1982, 1989) and Section 4.2.6 in Duchateau and Janssen (2008) for details.

The cross ratio has been estimated parametrically by Nan et al. (2006) using a partition of the sample space in rectangular regions with edges parallel to the variable axes, while assuming the cross ratio to be constant in each rectangular region. Time-varying cross ratios have been estimated using copula models, thereby making strong assumptions about the functional form of the bivariate copula (see, e.g., Li and Lin, 2006; Li et al., 2008). Hu et al. (2011) considered a parametric polynomial model for the log-transformed cross ratio function applied to right-censored bivariate survival data and using a pseudo-partial likelihood approach. Hsu and Prentice (1996) proposed a crude non-parametric estimator for the cross ratio function for which the performance was explored in the constant Clayton setting. Several authors considered the cross ratio function in the context of diagnostics to assess the frailty distribution or copula function used to describe dependence in bivariate survival

data. Viswanathan and Manatunga (2001) used a kernel smoothing approach to have an implicit estimator, and Chen and Bandeen-Roche (2005) considered an empirical estimator based on binning of the survival time space producing a piecewise constant estimator, in both cases relying on the Archimedean copula model family. Glidden (2007) implicitly proposed a non-parametric estimator for the cross ratio function, albeit that the estimator is of little practical use due to its rough behavior. More specifically, he defined residuals for pairwise dependence diagnostics, depending on an estimator for the cross ratio function, and used these residuals to do model checking for both frailty and copula models.

In this paper, we propose a smooth nonparametric Bernstein based estimator for the cross ratio function without relying on a parametric functional form or specific family of copula functions. In Section 2 we rewrite the conditional hazards defining the cross ratio function in terms of the survival copula, describing the dependence between  $T_1$  and  $T_2$ , and its partial derivatives. Using Bernstein estimators for the survival copula and its derivatives, we obtain Bernstein based estimators for the conditional hazards and a nonparametric estimator for the cross ratio function  $\theta(t_1, t_2)$ . The reason for using a Bernstein copula-based estimator for the cross ratio function is motivated from the results in the papers by Sancetta and Satchell (2004), Leblanc (2012), Janssen et al. (2012, 2014, 2016) and Bouezmarni et al. (2009, 2013). Simulations in these papers show that, compared to its nonparametric competitors (including kernel estimators), Bernstein based estimators for the copula and copula derivatives are superior. These papers also show that, compared to standard kernel estimators, the asymptotic variance of Bernstein estimators has smaller order and that the bias is uniform, i.e., Bernstein estimators do not show boundary effects. The asymptotic normality of the Bernstein estimator for  $\lambda(t_1 | T_2 = t_2)$  is studied in detail in Section 3. In Section 4 the asymptotic behavior of this estimator is compared to that of existing nonparametric competitors. In Section 5 we study the asymptotic normality of the Bernstein based cross ratio estimator defined in (7) of Section 2. Given the symmetry of  $\theta(t_1, t_2)$ , a corresponding estimator should preserve this property. We therefore include an interesting asymptotic normality result for a symmetrized version of the estimator (7). Efficient computational formulas are given in Section 6 which are important for the simulation study described in Section 7 and the real data example presented in Section 8. Proofs are given in the Appendix.

## 2 The estimator

Let  $(T_1, T_2)$  be a random pair of variables ( $T_1 \geq 0, T_2 \geq 0$ ) with absolutely continuous marginal distribution functions  $F_1$  and  $F_2$  and density functions  $f_1$  and  $f_2$ . Denote the marginal survival functions by  $S_1 = 1 - F_1$  and  $S_2 = 1 - F_2$ . According to Sklar's theorem (Sklar 1959; Nelsen 2006), there exists a unique copula function  $C$  that links the joint survival function of  $(T_1, T_2)$  to the

marginal survival functions, i.e.,

$$S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2) = C[S_1(t_1), S_2(t_2)]$$

or, since  $S_1$  and  $S_2$  are continuous

$$C(u_1, u_2) = S[S_1^{-1}(u_1), S_2^{-1}(u_2)]$$

with  $S_1^{-1}$  and  $S_2^{-1}$  the inverses of  $S_1$  and  $S_2$ . The conditional hazard rate functions appearing in expression (1) for  $\theta(t_1, t_2)$  can be expressed in terms of the copula  $C$  and its derivatives  $C^{(1)}$  and  $C^{(2)}$ , where, for  $i = 1, 2$ ,

$$C^{(i)}(u_1, u_2) = \frac{\partial}{\partial u_i} C(u_1, u_2).$$

Indeed it is easily verified that with  $S^{(i)}(t_1, t_2) = \frac{\partial}{\partial t_i} S(t_1, t_2)$  for  $i = 1, 2$ , we have:

$$\begin{aligned} \lambda(t_1 | T_2 = t_2) &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} P(t_1 < T_1 \leq t_1 + \Delta | T_1 > t_1, T_2 = t_2) \\ &= \frac{-\frac{\partial}{\partial t_1} S^{(2)}(t_1, t_2)}{S^{(2)}(t_1, t_2)} = \frac{-\frac{\partial}{\partial t_1} C^{(2)}[S_1(t_1), S_2(t_2)]}{C^{(2)}[S_1(t_1), S_2(t_2)]}, \end{aligned} \quad (2)$$

$$\begin{aligned} \lambda(t_1 | T_2 > t_2) &= \lim_{\Delta \rightarrow 0} \frac{1}{\Delta} P(t_1 < T_1 \leq t_1 + \Delta | T_1 > t_1, T_2 > t_2) \\ &= \frac{-\frac{\partial}{\partial t_1} S(t_1, t_2)}{S(t_1, t_2)} = \frac{-\frac{\partial}{\partial t_1} C[S_1(t_1), S_2(t_2)]}{C[S_1(t_1), S_2(t_2)]}. \end{aligned} \quad (3)$$

Nonparametric estimation of these two quantities will be done by first constructing nonparametric estimators for the corresponding cumulative hazard functions

$$\Lambda(t_1 | T_2 = t_2) = \int_0^{t_1} \frac{-d_s C^{(2)}[S_1(s), S_2(t_2)]}{C^{(2)}[S_1(s), S_2(t_2)]}, \quad (4)$$

$$\Lambda(t_1 | T_2 > t_2) = \int_0^{t_1} \frac{-d_s C[S_1(s), S_2(t_2)]}{C[S_1(s), S_2(t_2)]}, \quad (5)$$

followed by kernel smoothing.

Suppose now that  $(T_{11}, T_{21}), \dots, (T_{1n}, T_{2n})$  is a random sample from  $(T_1, T_2)$ . As an estimator for the copula function  $C(u_1, u_2)$  we will use the Bernstein estimator. In Janssen et al. (2012), the Bernstein estimator is defined as

$$C_{m,n}(u_1, u_2) = \sum_{k=0}^m \sum_{\ell=0}^m C_n \left( \frac{k}{m}, \frac{\ell}{m} \right) P_{m,k}(u_1) P_{m,\ell}(u_2)$$

with, for  $k = 0, 1, \dots, m$  and  $0 \leq u \leq 1$ ,

$$P_{m,k}(u) = \binom{m}{k} u^k (1-u)^{m-k}$$

representing binomial probabilities. The natural number  $m$  is called the order and we typically assume that  $m \rightarrow \infty$  as  $n \rightarrow \infty$ .  $C_n$  is the empirical copula given by

$$C_n(u_1, u_2) = S_n[S_{1n}^{-1}(u_1), S_{2n}^{-1}(u_2)],$$

where, with  $I$  the indicator function,

$$\begin{aligned} S_n(t_1, t_2) &= \frac{1}{n} \sum_{i=1}^n I(T_{1i} > t_1, T_{2i} > t_2), \\ S_{1n}(t_1) &= \frac{1}{n} \sum_{i=1}^n I(T_{1i} > t_1), \\ S_{2n}(t_2) &= \frac{1}{n} \sum_{i=1}^n I(T_{2i} > t_2). \end{aligned}$$

The Bernstein estimator for  $C^{(2)}(u_1, u_2)$  needed in (4) is defined as

$$C_{m,n}^{(2)}(u_1, u_2) = \frac{\partial}{\partial u_2} C_{m,n}(u_1, u_2).$$

Therefore, with  $S_{1n}(s-)$  the left hand limit of  $S_{1n}$  at  $s$ , appropriate estimators for the quantities in (4) and (5) are

$$\begin{aligned} \hat{A}_m(t_1 | T_2 = t_2) &= \int_0^{t_1} \frac{-d_s C_{m,n}^{(2)}[S_{1n}(s), S_{2n}(t_2)]}{C_{m,n}^{(2)}[S_{1n}(s-), S_{2n}(t_2)]} ds, \\ \hat{A}_m(t_1 | T_2 > t_2) &= \int_0^{t_1} \frac{-d_s C_{m,n}[S_{1n}(s), S_{2n}(t_2)]}{C_{m,n}[S_{1n}(s-), S_{2n}(t_2)]} ds, \end{aligned} \quad (6)$$

and by smoothing with a given probability density kernel  $K_0$  and a bandwidth sequence  $b_n \rightarrow 0$ , we obtain estimators for  $\lambda(t_1 | T_2 = t_2)$  and  $\lambda(t_1 | T_2 > t_2)$ :

$$\begin{aligned} \hat{\lambda}_m(t_1 | T_2 = t_2) &= \frac{1}{b_n} \int_0^\infty K_0\left(\frac{t_1 - s}{b_n}\right) \hat{A}_m(ds | T_2 = t_2) \\ &= \frac{1}{b_n} \int_0^\infty K_0\left(\frac{t_1 - s}{b_n}\right) \frac{-d_s C_{m,n}^{(2)}[S_{1n}(s), S_{2n}(t_2)]}{C_{m,n}^{(2)}[S_{1n}(s-), S_{2n}(t_2)]} ds, \\ \hat{\lambda}_m(t_1 | T_2 > t_2) &= \frac{1}{b_n} \int_0^\infty K_0\left(\frac{t_1 - s}{b_n}\right) \hat{A}_m(ds | T_2 > t_2) \\ &= \frac{1}{b_n} \int_0^\infty K_0\left(\frac{t_1 - s}{b_n}\right) \frac{-d_s C_{m,n}[S_{1n}(s), S_{2n}(t_2)]}{C_{m,n}[S_{1n}(s-), S_{2n}(t_2)]} ds. \end{aligned}$$

Finally, the proposed estimator for the cross ratio function  $\theta(t_1, t_2)$  in (1) is given by

$$\hat{\theta}_m(t_1, t_2) = \frac{\hat{\lambda}_m(t_1 | T_2 = t_2)}{\hat{\lambda}_m(t_1 | T_2 > t_2)}. \quad (7)$$

In Section 3 we deal with the asymptotic normality for the numerator in (7) which is a Bernstein based estimator for the conditional hazard rate function  $\lambda(t_1 | T_2 = t_2)$ . From this result and using our considerations in Section 4 on the order behavior of the denominator in (7) we obtain the asymptotic normality of  $\hat{\theta}_m(t_1, t_2)$ .

### 3 Asymptotic normality of the conditional hazard rate function estimator

The function  $\lambda(t_1 | T_2 = t_2)$  given in (2) has a more familiar interpretation. It is indeed easily checked that

$$C^{(2)}[S_1(t_1), S_2(t_2)] = P(T_1 > t_1 | T_2 = t_2) = 1 - F_{t_2}(t_1) = S_{t_2}(t_1)$$

where  $F_{t_2}(t_1) = P(T_1 \leq t_1 | T_2 = t_2)$  is the conditional distribution function of  $T_1$  given  $T_2 = t_2$ . Similarly,

$$\frac{\partial}{\partial t_1} C^{(2)}[S_1(t_1), S_2(t_2)] = -f_{t_2}(t_1)$$

where  $f_{t_2}(t_1)$  is the conditional density function of  $T_1$  given  $T_2 = t_2$ . Note that we use throughout  $F_{t_2}(t_1)$  and  $f_{t_2}(t_1)$  as shorthand notation for the more common notation  $F_{T_1|T_2}(t_1 | t_2)$  and  $f_{T_1|T_2}(t_1 | t_2)$  respectively. Hence,

$$\lambda(t_1 | T_2 = t_2) = \frac{f_{t_2}(t_1)}{1 - F_{t_2}(t_1)}.$$

Furthermore,

$$C_{m,n}^{(2)}[S_{1n}(t_1), S_{2n}(t_2)] = 1 - \hat{F}_{t_2}(t_1)$$

and

$$d_{t_1} C_{m,n}^{(2)}[S_{1n}(t_1), S_{2n}(t_2)] = -d\hat{F}_{t_2}(t_1)$$

where  $\hat{F}_{t_2}(t_1)$  is precisely the Bernstein estimator studied in Janssen et al. (2016). Hence, expression (6) can be written as

$$\hat{A}_m(t_1 | T_2 = t_2) = \int_0^{t_1} \frac{d\hat{F}_{t_2}(s)}{1 - \hat{F}_{t_2}(s-)}$$



and

$$\widehat{\lambda}_m(t_1 | T_2 = t_2) = \frac{1}{b_n} \int_0^\infty K_0 \left( \frac{t_1 - s}{b_n} \right) \frac{d\widehat{F}_{t_2}(s)}{1 - \widehat{F}_{t_2}(s-)}.$$

**Theorem 1.** Assume that the following conditions hold.

- (a) The copula  $C$  has 4th order partial derivatives in  $(0, 1)^2$ .
- (b)  $f_1(t_1)$  is twice continuously differentiable in an open neighborhood  $U$  of  $t_1$  and  $\frac{\partial^4}{\partial^3 u \partial v} C(u, v)$  exists and is continuous on  $(u, v) \in S_1(U) \times (0, 1)$ .
- (c)  $K_0$  is a continuous probability density function of bounded variation with bounded support  $[-L, L]$ ,  $K_0(-L) = K_0(L) = 0$  and

$$\mu_1(K_0) = \int_{-L}^L t K_0(t) dt = 0.$$

- (d) The order  $m$  and the bandwidth  $b_n$  satisfy the relations, for  $n \rightarrow \infty$ ,

$$\begin{aligned} n^{1/2} m^{-5/4} b_n^{-1/2} &\rightarrow 0 \\ n^{-1/2} m^{5/6} b_n^{-1/2} (\ln n)^{1/2} (\ln \ln n)^{1/2} &\rightarrow 0 \\ n m^{-1/2} b_n^5 &\rightarrow C_1 \geq 0, \end{aligned}$$

where  $C_1$  is some finite constant.

Then, for all  $(t_1, t_2)$  such that  $0 < F_1(t_1) < 1$ ,  $0 < F_2(t_2) < 1$ ,  $0 < F_{t_2}(t_1) < 1$ ,

$$\begin{aligned} &(nm^{-1/2}b_n)^{1/2} [\widehat{\lambda}_m(t_1 | T_2 = t_2) - \lambda(t_1 | T_2 = t_2)] \\ &\xrightarrow{d} N \left( \beta(t_1, t_2); \frac{\|K_0\|^2}{2\sqrt{\pi F_2(t_2)(1 - F_2(t_2))}} \frac{\lambda(t_1 | T_2 = t_2)}{1 - F_{t_2}(t_1)} \right), \end{aligned}$$

where

$$\beta(t_1, t_2) = \frac{1}{2} C_1^{1/2} \lambda''(t_1 | T_2 = t_2) \mu_2(K_0), \quad (8)$$

$$\|K_0\|^2 = \int K_0^2(t) dt \text{ and } \mu_2(K_0) = \int t^2 K_0(t) dt.$$

The reader is referred to the Appendix for the proof of Theorem 1.

#### 4 Discussion on Theorem 1

In this section we discuss several aspects of the estimator in Theorem 1: a comparison with other nonparametric estimators for the conditional hazard function (Section 4.1), a remark on the boundary effect (Section 4.2), a comparison of the required assumptions (Section 4.3) and an expression for the mean squared error (Section 4.4).

#### 4.1 Nonparametric estimators for the conditional hazard function

There are several results in the literature dealing with nonparametric estimation of the conditional hazard rate function based on complete or censored data. References can be found in Van Keilegom and Veraverbeke (2001) and Spierdijk (2008). These two papers are also representative for the two typical methods of constructing an estimator: (1) replacing the conditional density and conditional survival function by nonparametric estimators, or (2) smoothing a nonparametric estimator for the cumulative conditional hazard function. The two just mentioned papers deal with censored data, but it is easy to take out the censoring in order to compare with the results in Theorem 1. In Spierdijk (2008) an estimator of type (1) with local linear smoothing of numerator and denominator is studied while the estimator in Van Keilegom and Veraverbeke (2001) is of type (2). These authors applied kernel smoothing to a nonparametric cumulative conditional hazard estimator. Our new estimator for  $\lambda(t_1 | T_2 = t_2)$  is also of type (2). The Bernstein method is used to estimate the cumulative conditional hazard. The result of Van Keilegom and Veraverbeke (2001) is formulated for a fixed covariate design. Replacing the Gasser-Müller weights by the Nadaraya-Watson weights results in the appearance of the design density  $f_2(t_2)$  in the denominator of the expression for the asymptotic variance. Hence, the asymptotic variance of the estimators in Van Keilegom and Veraverbeke (2001) and Spierdijk (2008) is

$$\frac{1}{nh_nb_n} \frac{\|K_0\|^2 \|K\|^2}{f_2(t_2)} \frac{\lambda(t_1 | T_2 = t_2)}{1 - F_{t_2}(t_1)}, \quad (9)$$

with  $K$  the kernel used for the cumulative hazard estimator, while the asymptotic variance of our new estimator in Theorem 1 is given by

$$\frac{1}{nm^{-1/2}b_n} \frac{\|K_0\|^2}{2\sqrt{\pi F_2(t_2)} [1 - F_2(t_2)]} \frac{\lambda(t_1 | T_2 = t_2)}{1 - F_{t_2}(t_1)}. \quad (10)$$

Further note that the order  $O((nm^{-1/2}b_n)^{-1})$  in (10) is smaller than the order  $O((nh_nb_n)^{-1})$  in (9), when making the usual identification  $h_n = m^{-1}$  as proposed by Sancetta and Satchell (2004) or Leblanc (2012). This improved order is due to the Bernstein method.

A further remark is that the density  $f_2(t_2)$  in the denominator of (9) is replaced by  $\sqrt{F_2(t_2)} [1 - F_2(t_2)]$  in (10). A comparison of these two factors can be made using a result of Parzen (1979); see also Remark 5 in Janssen et al. (2016). It follows that  $\sqrt{F_2(t_2)} [1 - F_2(t_2)]$  is asymptotically (as  $t_2 \rightarrow \infty$ ) larger than  $f_2(t_2)$  for all  $T_2$  with medium tails (e.g., exponential, Weibull, and normal) and long tails (e.g., Cauchy and Pareto).

Note that asymptotically ( $t_2 \rightarrow \infty$ ) the scale of  $T_2$  does not play any role in this comparison. Indeed, with  $\sigma^2 = \text{Var}(T_2)$ ,  $\tilde{T}_2 = T_2/\sigma$ , we can easily see that the distribution function of  $\tilde{F}_2$  and the density function  $\tilde{f}_2$  of  $\tilde{T}_2$  satisfy

$$\frac{\tilde{f}_2(\tilde{t}_2)}{\sqrt{\tilde{F}_2(\tilde{t}_2)[1 - \tilde{F}_2(\tilde{t}_2)]}} = \frac{\sigma f_2(t_2)}{\sqrt{F_2(t_2)[1 - F_2(t_2)]}},$$

with  $\tilde{t}_2 = t_2/\sigma$ . Since  $t_2 \rightarrow \infty$  is equivalent to  $\tilde{t}_2 \rightarrow \infty$ , we have that the right hand side tends to zero whenever the left hand side does. So asymptotically as  $t_2 \rightarrow \infty$ , the value of  $\sigma$  has no influence.

Non-asymptotically, however, there is of course the question of ‘how large’  $t_2$  has to be in order to have that  $f(t_2)$  is smaller than  $\sqrt{F_2(t_2)[1 - F_2(t_2)]}$ . This can only be checked in concrete cases. For instance, if  $T_2$  is exponentially distributed with parameter  $1/\sigma$ , then the above inequality is satisfied for all  $t_2 > l = \sigma \ln(1 + \frac{1}{\sigma^2})$ . For  $\sigma = 1$ :  $l = 0.7$ , for  $\sigma = 10$ :  $l = 0.1$ . Note that  $l$  decreases with  $\sigma$ .

#### 4.2 Boundary effect

The definition of  $\hat{\lambda}(t_1|T_2 = t_2)$  is obtained by kernel smoothing of the empirical cumulative hazard function. It is a well known fact that such estimator can suffer from some boundary effect at  $t_1 = 0$ . The remedies are well described in the literature. We mention: (1) the use of a boundary kernel (Müller and Wang, 1994), (2) mirror-reflection (Gijbels and Mielniczuk, 1990), (3) shrinkage of the bandwidth (Omelka et al., 2009), and (4) various transformation methods (Ruppert and Cline, 1994, Swanepoel and van Graan, 2005). To avoid the related technicalities, we restrict our attention to interior points.

#### 4.3 Required conditions

The conditions of Theorem 1 are in line with the conditions in Van Keilegom and Veraverbeke (2001) and Spierdijk (2008). Conditions (a) and (b) of Theorem 1 are very close to the assumptions (iv) and (v) of the parallel result in Spierdijk (2008). Starting from the relation  $C^{(2)}[S_1(t_1), S_2(t_2)] = 1 - F_{t_2}(t_1)$ , it is easily seen that there is a correspondence between the third order derivative of  $F_{t_2}(t_1)$  and the fourth order partial derivatives of the copula  $C$ , together with the second derivative of  $f_1$ . Condition (c) is of course standard in kernel smoothing. Finally, the relations between the order  $m$  and the bandwidth  $b_n$  are needed to take care of a remainder term and also to control a bias term. Similar relations appear, for example, in Theorem 4 of Van Keilegom and Veraverbeke (2001).

#### 4.4 Asymptotic mean squared error

From (11) and (17), both in the proof of Theorem 1 in the Appendix, we have an explicit expression for the asymptotic mean squared error of  $\hat{\lambda}_m(t_1 | T_2 =$

$t_2$ ), namely

$$\frac{1}{nm^{-1/2}b_n} \frac{\|K_0\|^2}{2\sqrt{\pi F_2(t_2)[1-F_2(t_2)]}} \frac{\lambda(t_1 | T_2 = t_2)}{1 - F_{t_2}(t_1)} + \left[ \frac{1}{2} b_n^2 \lambda''(t_1 | T_2 = t_2) \mu_2(K_0) + \frac{1}{2} m^{-1} \phi(t_1, t_2) \right]^2,$$

with  $\phi(t_1, t_2)$  defined in (17) of the proof of Theorem 1. Minimizing with respect to  $m$  and  $b_n$  provides that the optimal choices are of the form  $b_n = O(n^{-1/6})$  and  $m = O(n^{1/3})$  and that the asymptotic mean squared error has order  $O(n^{-2/3})$ . Furthermore,  $\hat{\lambda}_m(t_1 | T_2 = t_2) - \lambda(t_1 | T_2 = t_2) = O_p(n^{-1/3})$ .

## 5 Asymptotic normality of the cross ratio function estimator

We first prove a result on the behavior of the denominator in the estimator  $\hat{\theta}_m(t_1, t_2)$  given in (7).

**Theorem 2.** Assume that the conditions of Theorem 1 hold, together with the extra relation  $m^{1/2}b_n \rightarrow \infty$ . Then, for all  $(t_1, t_2)$  such that  $C(S_1(t_1), S_2(t_2)) > 0$ ,

$$(nm^{-1/2}b_n)^{1/2} [\hat{\lambda}_m(t_1 | T_2 > t_2) - \lambda(t_1 | T_2 > t_2)] \xrightarrow{P} 0.$$

Combining the results of Theorem 1 and Theorem 2 leads to the asymptotic normality for the cross ratio estimator  $\hat{\theta}_m(t_1, t_2)$ .

**Theorem 3.** Assume that the conditions of Theorem 2 hold.

Then, for all  $(t_1, t_2)$  such that  $0 < F_1(t_1) < 1$ ,  $0 < F_2(t_2) < 1$ ,  $0 < F_{t_2}(t_1) < 1$ ,  $0 < C[S_1(t_1), S_2(t_2)]$ ,

$$\xrightarrow{d} N \left( \frac{\beta(t_1, t_2)}{\lambda(t_1 | T_2 > t_2)}; \frac{\|K_0\|^2}{2\sqrt{\pi F_2(t_2)[1-F_2(t_2)]}} \frac{\theta^2(t_1, t_2)}{f_{t_2}(t_1)} \right).$$

Note that the estimator  $\hat{\theta}_m(t_1, t_2)$  in (7) is not symmetric in the pair  $(T_1, T_2)$ . However, this is in contrast to the function  $\theta(t_1, t_2)$  defined in (1) which has the symmetry property that  $\theta(t_1, t_2) = \theta^*(t_2, t_1)$  where

$$\theta^*(t_2, t_1) = \frac{\lambda(t_2 | T_1 = t_1)}{\lambda(t_2 | T_1 > t_1)},$$

which is a direct application of Bayes' rule. Since our estimator  $\hat{\theta}_m(t_1, t_2)$  does not preserve this property, we define the following symmetrized version

$$\hat{\vartheta}_m(t_1, t_2) = \frac{1}{2} [\hat{\theta}_m(t_1, t_2) + \hat{\theta}_m^*(t_2, t_1)],$$

where  $\theta_m^*(t_2, t_1)$  and its estimator  $\hat{\theta}_m^*(t_2, t_1)$  are obtained by reversing the roles of  $T_1$  and  $T_2$  in the definition of  $\theta(t_1, t_2)$  and the estimator in (7) respectively. The next theorem provides the asymptotic normality result for  $\hat{\vartheta}_m(t_1, t_2)$ .

**Theorem 4.** Assume that the conditions of Theorem 2 hold and

- (b\*)  $f_2(t_2)$  is twice continuously differentiable in an open neighborhood  $V$  of  $t_2$  and  $\frac{\partial^4}{\partial^3 u \partial v} C(u, v)$  exists and is continuous on  $(u, v) \in (0, 1) \times S_2(V)$ .

Then, we have for all  $(t_1, t_2)$  such that  $0 < F_1(t_1) < 1$ ,  $0 < F_2(t_2) < 1$ ,  $0 < C[S_1(t_1), S_2(t_2)] < 1$ ,

$$(nm^{-1/2}b_n)^{1/2}[\hat{\vartheta}_m(t_1, t_2) - \theta(t_1, t_2)] \xrightarrow{d} N(B(t_1, t_2); V(t_1, t_2)),$$

where

$$B(t_1, t_2) = \frac{1}{2} \left[ \frac{\beta(t_1, t_2)}{\lambda(t_1 | T_2 > t_2)} + \frac{\beta^*(t_2, t_1)}{\lambda(t_2 | T_1 > t_1)} \right],$$

$$V(t_1, t_2) = \frac{1}{4} \|K_0\|^2 \theta^2(t_1, t_2) \left[ \frac{1}{2f_{t_2}(t_1) \sqrt{\pi F_2(t_2) [1 - F_2(t_2)]}} + \frac{1}{2f_{t_1}(t_2) \sqrt{\pi F_1(t_1) [1 - F_1(t_1)]}} \right],$$

with  $\beta(t_1, t_2)$  defined in (8) and

$$\beta^*(t_2, t_1) = \frac{1}{2} C_1^{1/2} \lambda''(t_2 | T_1 = t_1) \mu_2(K_0).$$

The proof of this theorem is given in the Supplementary Material.

**Remark 1.** Note that the asymptotic variance of  $\hat{\vartheta}_m(t_1, t_2)$  is essentially 1/4 of the sum of the asymptotic variances of the two estimators  $\hat{\theta}_m(t_1, t_2)$  and  $\hat{\theta}_m^*(t_2, t_1)$ . The proof indeed shows that the covariance terms are of lower order.

**Remark 2.** We have the following expression for the asymptotic mean squared error of  $\hat{\vartheta}_m(t_1, t_2)$ :

$$\frac{V(t_1, t_2)}{nm^{-1/2}b_n} + \frac{1}{4} \left\{ \frac{\frac{1}{2}b_n^2 \lambda''(t_1 | T_2 = t_2) \mu_2(K_0) + \frac{1}{2}m^{-1} \phi(t_1, t_2)}{\lambda(t_1 | T_2 > t_2)} + \frac{\frac{1}{2}b_n^2 \lambda''(t_2 | T_1 = t_1) \mu_2(K_0) + \frac{1}{2}m^{-1} \phi^*(t_1, t_2)}{\lambda(t_2 | T_1 > t_1)} \right\}^2,$$

with  $\phi(t_1, t_2)$  defined in (17) of the proof of Theorem 1 and

$$\phi^*(t_1, t_2) = \frac{b^{*(2)}[S_1(t_1), S_2(t_2)]}{1 - F_{t_1}(t_2)} f_2(t_2),$$

$$b^*(u, v) = (1 - 2u)C^{(1,1)}(u, v) + u(1 - u)C^{(1,1,1)}(u, v) + v(1 - v)C^{(1,2,2)}(u, v).$$

**Remark 3.** In Theorem 4 we take the same bandwidth  $b_n$  for  $\hat{\theta}_m(t_1, t_2)$  (smoothing over  $t_1$ ) and  $\hat{\theta}_m^*(t_2, t_1)$  (smoothing over  $t_2$ ), which is a natural choice when  $T_1$  and  $T_2$  have a similar scale. In Section 7 (simulations), we briefly discuss the non-similar scale case.

## 6 Computational formulas

For the random sample  $(T_{11}, T_{21}), \dots, (T_{1n}, T_{2n})$  let, for  $i = 1, 2$ ,  $T_{i(1)} \leq T_{i(2)} \leq \dots \leq T_{i(n)}$  denote the ordered  $T_{ij}$ -values; and let  $R_{ij}$  denote the rank of  $T_{ij}$ ,  $j = 1, \dots, n$ . For computational convenience, the empirical version of the marginal survival functions used in the computational formulas are

$$\tilde{S}_{1n}(t_1) = \frac{n}{n+1} S_{1n}(t_1) \text{ and } \tilde{S}_{2n}(t_2) = \frac{n}{n+1} S_{2n}(t_2).$$

The empirical estimator for the survival copula of  $(T_1, T_2)$  is:

$$\begin{aligned} C_n(u_1, u_2) &= \frac{1}{n} \sum_{j=1}^n I[\tilde{S}_{1n}(T_{1j}) < u_1, \tilde{S}_{2n}(T_{2j}) < u_2] \\ &= \frac{1}{n} \sum_{j=1}^n I\left(\frac{R_{1j}}{n+1} > \frac{n}{n+1} - u_1, \frac{R_{2j}}{n+1} > \frac{n}{n+1} - u_2\right). \end{aligned}$$

In terms of  $C_{m,n}$ ,  $C_{m,n}^{(1)}$ ,  $C_{m,n}^{(2)}$ ,  $C_{m,n}^{(1,2)}$ , the Bernstein estimators for the copula  $C$  and its first order and second order partial derivatives, the proposed estimator for the cross ratio function given in (7) can be written as:

$$\hat{\theta}_m(t_1, t_2) = \frac{\hat{\lambda}_m(t_1 \mid T_2 = t_2)}{\hat{\lambda}_m(t_1 \mid T_2 > t_2)},$$

where

$$\begin{aligned} \hat{\lambda}_m(t_1 \mid T_2 = t_2) &= \frac{1}{(n+1)b_n} \sum_{j=1}^n K_0\left(\frac{t_1 - T_{1(j)}}{b_n}\right) \frac{C_{m,n}^{(1,2)}\left[\frac{n-j}{n+1}, \tilde{S}_{2n}(t_2)\right]}{C_{m,n}^{(2)}\left[\frac{n-j+1}{n+1}, \tilde{S}_{2n}(t_2)\right]}, \\ \hat{\lambda}_m(t_1 \mid T_2 > t_2) &= \frac{1}{(n+1)b_n} \sum_{j=1}^n K_0\left(\frac{t_1 - T_{1(j)}}{b_n}\right) \frac{C_{m,n}^{(1)}\left[\frac{n-j}{n+1}, \tilde{S}_{2n}(t_2)\right]}{C_{m,n}^{(1)}\left[\frac{n-j+1}{n+1}, \tilde{S}_{2n}(t_2)\right]}, \end{aligned}$$

with

$$\begin{aligned}
C_{m,n}(u_1, u_2) &= \sum_{k=0}^m \sum_{l=0}^m C_n \left( \frac{k}{m}, \frac{l}{m} \right) P_{m,k}(u_1) P_{m,l}(u_2), \\
C_{m,n}^{(1)}(u_1, u_2) &= m \sum_{k=0}^{m-1} \sum_{l=0}^m \left[ C_n \left( \frac{k+1}{m}, \frac{l}{m} \right) - C_n \left( \frac{k}{m}, \frac{l}{m} \right) \right] P_{m-1,k}(u_1) P_{m,l}(u_2), \\
C_{m,n}^{(2)}(u_1, u_2) &= m \sum_{k=0}^m \sum_{l=0}^{m-1} \left[ C_n \left( \frac{k}{m}, \frac{l+1}{m} \right) - C_n \left( \frac{k}{m}, \frac{l}{m} \right) \right] P_{m,k}(u_1) P_{m-1,l}(u_2), \\
C_{m,n}^{(1,2)}(u_1, u_2) &= m^2 \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} P_{m-1,k}(u_1) P_{m-1,l}(u_2) \\
&\quad \times \left[ C_n \left( \frac{k+1}{m}, \frac{l+1}{m} \right) - C_n \left( \frac{k}{m}, \frac{l+1}{m} \right) - C_n \left( \frac{k+1}{m}, \frac{l}{m} \right) + C_n \left( \frac{k}{m}, \frac{l}{m} \right) \right].
\end{aligned}$$

See Janssen et al. (2012, 2014, 2016) for more details.

In order to have an estimator which is also smooth in the  $t_2$ -direction, we replace in the simulations and the data example the empirical survival function  $\tilde{S}_{2n}(t_2)$  in the Bernstein estimator for  $\theta(t_1, t_2)$  by:

$$\tilde{\tilde{S}}_{2n}(t_2) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{K}_0 \left( \frac{T_{2(i)} - t_2}{b_n} \right),$$

where  $\mathbb{K}_0$  is the integrated kernel

$$\mathbb{K}_0(t) = \int_{-\infty}^t K_0(u) du.$$

For computational efficiency we use the following equivalent formulas (see Janssen et al., 2017, Remark 5), for similar formulas in the context of conditional density estimation):

$$\begin{aligned}
C_{m,n}(u_1, u_2) &= \frac{1}{n} \sum_{j=1}^n \sum_{k=\lfloor \frac{m(n-R_{1j})}{n+1} \rfloor + 1}^m P_{m,k}(u_1) \times \sum_{l=\lfloor \frac{m(n-R_{2j})}{n+1} \rfloor + 1}^m P_{m,l}(u_2), \\
C_{m,n}^{(1)}(u_1, u_2) &= \frac{m}{n} \sum_{j=1}^n P_{m-1, \lfloor \frac{m(n-R_{1j})}{n+1} \rfloor}(u_1) \times \sum_{l=\lfloor \frac{m(n-R_{2j})}{n+1} \rfloor + 1}^m P_{m,l}(u_2), \\
C_{m,n}^{(2)}(u_1, u_2) &= \frac{m}{n} \sum_{j=1}^n \sum_{k=\lfloor \frac{m(n-R_{1j})}{n+1} \rfloor + 1}^m P_{m,k}(u_1) \times P_{m, \lfloor \frac{m(n-R_{2j})}{n+1} \rfloor}(u_2), \\
C_{m,n}^{(1,2)}(u_1, u_2) &= \frac{m^2}{n} \sum_{j=1}^n P_{m, \lfloor \frac{m(n-R_{1j})}{n+1} \rfloor}(u_1) \times P_{m, \lfloor \frac{m(n-R_{2j})}{n+1} \rfloor}(u_2).
\end{aligned}$$

## 7 Simulations

Based on simulations we show the finite sample performance of  $\hat{\theta}_m(t_1, t_2)$ ,  $\hat{\theta}_m^*(t_2, t_1)$  and  $\hat{\vartheta}_m(t_1, t_2)$ , our unsymmetrized and symmetrized estimators of the cross ratio function  $\theta(t_1, t_2)$ . To quantify the impact of the choice of the kernel bandwidth  $b_n$  and the Bernstein order  $m$ , we use an approximation of the  $MISE_{\hat{\theta}_m}$ , the mean integrated squared error for which a formal definition can be found in the Supplementary Material. We now define for  $\hat{\theta}_m(t_1, t_2)$ , the quantities needed to study this impact. For  $\hat{\theta}_m^*(t_2, t_1)$  and  $\hat{\vartheta}_m(t_1, t_2)$  similar expressions are obtained by replacing  $\hat{\theta}_m(t_1, t_2)$  by the estimator under consideration. The integrated squared error on the rectangle  $[a_1^*, b_1^*] \times [a_2^*, b_2^*] \subset [0, 1] \times [0, 1]$ ,  $ISE_{\hat{\theta}_m}$ , is defined as

$$ISE_{\hat{\theta}_m} = \int_{a_1^*}^{b_1^*} \int_{a_2^*}^{b_2^*} \left\{ \hat{\theta}_m[F_1^{-1}(u_1), F_2^{-1}(u_2)] - \theta[F_1^{-1}(u_1), F_2^{-1}(u_2)] \right\}^2 \\ \times dF_1^{-1}(u_1) dF_2^{-1}(u_2).$$

Note that, with  $[a_1^*, b_1^*] \times [a_2^*, b_2^*] \equiv [F_1(a_1), F_1(b_1)] \times [F_2(a_2), F_2(b_2)]$

$$ISE_{\hat{\theta}_m} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \left[ \hat{\theta}_m(t_1, t_2) - \theta(t_1, t_2) \right]^2 dt_1 dt_2.$$

Using a bivariate grid of  $N_1 \times N_2$  equally-spaced grid points in  $[a_1^*, b_1^*] \times [a_2^*, b_2^*]$ , we approximate  $ISE_{\hat{\theta}_m}$  by

$$I_{\hat{\theta}_m} = \Delta_1 \Delta_2 \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} w_{kl} \left\{ \hat{\theta}_m[F_1^{-1}(u_{1[k]}), F_2^{-1}(u_{2[l]})] - \theta[F_1^{-1}(u_{1[k]}), F_2^{-1}(u_{2[l]})] \right\}^2,$$

where  $\Delta_1 = (b_1^* - a_1^*)/(N_1 - 1)$ ,  $\Delta_2 = (b_2^* - a_2^*)/(N_2 - 1)$ ,  $u_{1[k]} = a_1^* + (b_1^* - a_1^*)(k-1)/(N_1 - 1)$  and  $u_{2[l]} = a_2^* + (b_2^* - a_2^*)(l-1)/(N_2 - 1)$ ,  $k = 1, \dots, N_1$ ,  $l = 1, \dots, N_2$ ; and the weights  $w_{kl}$  are equal to  $(dF_1^{-1}(u_1)/du_1)(dF_2^{-1}(u_2)/du_2)$  evaluated in  $(u_{1[k]}, u_{2[l]})$ . The  $MISE_{\hat{\theta}_m}$  is then approximated by averaging over  $M = 100$  simulation runs (denoted by  $MI_{\hat{\theta}_m}$ ):

$$MI_{\hat{\theta}_m} = \frac{1}{M} \sum_{r=1}^M I_{\hat{\theta}_m}^{(r)},$$

where  $I_{\hat{\theta}_m}^{(r)}$  is the approximation of the  $ISE$  based on the  $r$ -th simulated dataset. In our simulations we take  $[a_1^*, b_1^*] = [a_2^*, b_2^*] = [0.01, 0.99]$  and  $N_1 = N_2 = 99$ .



On top of that, since the mean integrated squared error can be decomposed into variance and (squared) bias components (see Supplementary Material for more details), we provide approximations of the integrated variance  $IVAR_{\hat{\theta}_m}$  and the integrated squared bias  $ISBIAS_{\hat{\theta}_m}$ , which will be denoted by  $IV_{\hat{\theta}_m}$  and  $ISB_{\hat{\theta}_m}$  respectively from this point onwards, and for which expressions are provided in the Supplementary Material.

### 7.1 Simulation procedure

We generate  $n$  pairs of event times  $(t_{1j}, t_{2j})$ ,  $j = 1, \dots, n$  using the ‘copula’ package in R version 3.3.2. More specifically, random samples  $(u_{1j}, u_{2j})$  are drawn from three different copula functions with various tail dependencies (Clayton, Gumbel and Frank copulae) after which dependent exponential event times are obtained with constant hazard functions  $\lambda_1 = 0.03$  and  $\lambda_2 = 0.05$ , as follows:

$$t_{ij} = -\frac{\ln(1 - u_{ij})}{\lambda_i}.$$

The Clayton copula captures lower tail dependence, while the Gumbel copula captures upper tail dependence; the Frank copula family has no (upper or lower) tail dependence. In our simulation study, we generate simulation sets of sample size  $n = 500$  and  $n = 800$ . Additional simulation results for varying sample sizes are provided in the Supplementary Material. The proposed estimator works globally well for sample sizes of at least 300 pairs. Depending on the shape of the cross ratio surface, smaller sample sizes can still provide a useful estimator (see Supplementary Material for more details).

For  $K_0$  we take the standard normal density. To investigate the effect of the kernel bandwidth  $b_n$  and the Bernstein order  $m$  on the performance of our estimators  $\hat{\theta}_m(t_1, t_2)$ ,  $\hat{\theta}_m^*(t_2, t_1)$  and  $\hat{\vartheta}_m(t_1, t_2)$ , we explore a range of values  $B = \{1, 2.7, 7.4, 20.1, 54.6\}$  for  $b_n$ , i.e., the bandwidths values are chosen equidistant on a logarithmic scale, and a range  $A = \{25, 50, 100\}$  for the Bernstein order  $m$ . The bandwidths chosen in  $B$  are equidistant on the log-scale to have more small candidate values than large ones, since large bandwidths typically yield very similar smooth estimates. The order  $m$  is typically chosen as a fraction of  $n$  (see Janssen et al., 2016). Furthermore, note that  $T_1$  takes values in the range  $[0, 100]$  with probability 0.95 thereby explaining the maximal range of the grid.

**Remark 4.** Note that, although  $T_1$  and  $T_2$  have different scales (the scale ratio is  $\lambda_1/\lambda_2 = 0.6$ ), we use in the simulations the same bandwidth for  $\hat{\theta}_m(t_1, t_2)$  and  $\hat{\theta}_m^*(t_2, t_1)$ . Simulations (for  $n = 300$ , see Tables 4–6 in the Supplementary Material) in which we choose bandwidths  $b_{n1}$  for  $\hat{\theta}_m(t_1, t_2)$  and  $b_{n2} = cb_{n1}$  for  $\hat{\theta}_m^*(t_2, t_1)$ , where, in each simulation run, we take for  $c$  the estimated scale ratio for  $T_1$  and  $T_2$ , give the same optimal choices for the smoothing parameter.

Hence, the selection of the bandwidth is not sensitive to this size of the scale difference.

Also note that, if working with two different bandwidths,  $b_{n2} = cb_{n1}$  is a simple and workable choice in relation to condition (d). Theorem 4 can easily be adapted to this choice.

## 7.2 Clayton copula function

First, we consider the Clayton copula function ( $\theta > 0$ ):

$$C_\theta(u_1, u_2) = \left\{ \max[u_1^{-\theta} + u_2^{-\theta} - 1, 0] \right\}^{-1/\theta}$$

with parameter  $\theta = 0.5$ . The true underlying cross ratio function takes constant value  $1 + \theta = 1.5$ . In Table 1, we show the approximated mean integrated squared errors, integrated variances and integrated squared biases for the symmetrized estimator  $\hat{\vartheta}_m(t_1, t_2)$  for various choices of  $b_n$ ,  $m$  and  $n$ . As the performances are very similar, the simulation results for the unsymmetrical estimators  $\hat{\theta}_m(t_1, t_2)$  and  $\hat{\theta}_m^*(t_2, t_1)$  can be found in the Supplementary Material. Since the true cross ratio function is constant, large bandwidths  $b_n$  are preferred based on the minimization of  $MI_{\hat{\vartheta}_m}$ . In Figure 1, we graphically show a heat plot of the difference between the estimated cross ratio values  $\hat{\vartheta}_m(t_1, t_2)$  averaged over the  $M$  replications and the true values  $\theta(t_1, t_2)$  (left upper panel), and the estimated cross ratio function  $\hat{\vartheta}_m(t_1, t_2)$  (black solid lines in the other panels) as a function of one time component by fixing the other ( $t_1 = F_1^{-1}(0.5)$  in the right upper panel, or  $t_2 = F_2^{-1}(0.5)$  in the left lower panel, respectively, and  $\hat{b}_n = 54.6$ ,  $m = 100$  and  $n = 800$ ). Pointwise 95% simulation-based confidence intervals (gray dash-dotted lines) and true cross ratio values (red dashed lines) are included as well. In the right lower panel, we plot the cross ratio function  $\hat{\vartheta}_m[F_1^{-1}(u), F_2^{-1}(u)]$  against  $u \in (0, 1)$ . In the left panel of Figure 2, we illustrate the fact that the asymptotic variance of the symmetrized estimator  $\hat{\vartheta}_m(t_1, t_2)$  is essentially equal to  $(1/4)$  times the sum of the asymptotic variances of  $\hat{\theta}_m(t_1, t_2)$  and  $\hat{\theta}_m^*(t_2, t_1)$  based on the simulation results.

## 7.3 Gumbel copula function

Second, a Gumbel copula function is considered with parameter  $\theta = 1.5$  to induce time-varying association among the event times, i.e., for  $\theta \in [1, \infty)$ :

$$C_\theta(u_1, u_2) = \exp \left( - \left\{ [-\ln(u_1)]^\theta + [-\ln(u_2)]^\theta \right\}^{1/\theta} \right).$$

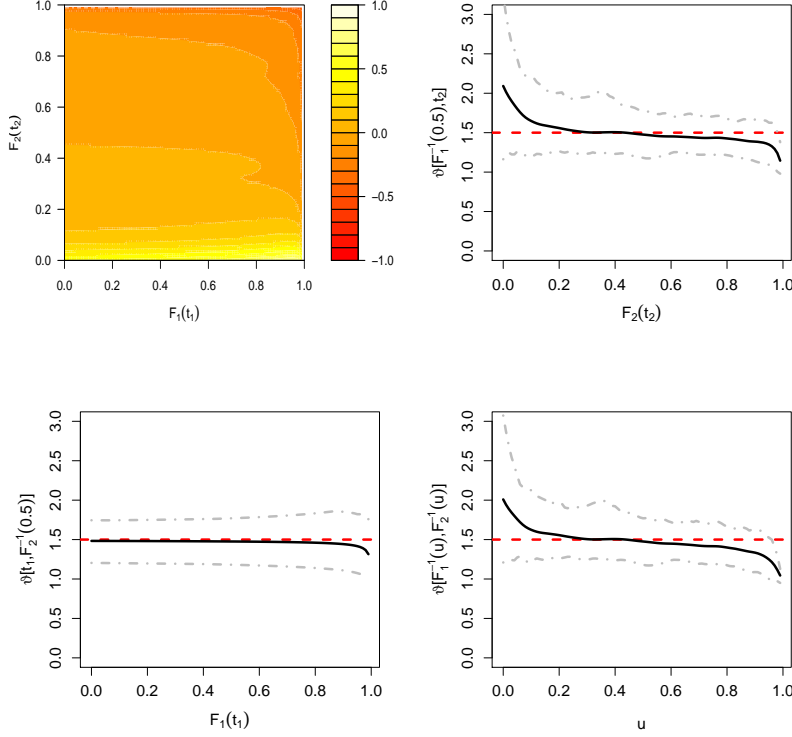
The expression for the true cross ratio function is given by:

$$\theta(t_1, t_2) = 1 + (\theta - 1) \left( \{-\ln[S_1(t_1)]\}^\theta + \{-\ln[S_2(t_2)]\}^\theta \right)^{-1/\theta}.$$

**Table 1**  $MI_{\hat{\vartheta}_m}$ ,  $IV_{\hat{\vartheta}_m}$  and  $ISB_{\hat{\vartheta}_m}$  for different choices of  $b_n$ ,  $m$  and  $n$  and different assumptions regarding the dependence between the event times; Clayton, Gumbel and Frank copula functions with parameters  $\theta = 0.5, 1.5$  and  $3.0$ , respectively. Minimum  $MI_{\hat{\vartheta}_m}$ -values are highlighted in bold.

Clayton copula ( $\theta = 0.5$ )											
$m/\hat{b}_n$		$n = 500$					$n = 800$				
		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\hat{\vartheta}_m}$	3.928	3.914	3.903	3.929	4.156	3.826	3.823	3.825	3.861	4.120
	$IV_{\hat{\vartheta}_m}$	0.382	0.333	0.269	0.176	0.094	0.236	0.206	0.172	0.120	0.067
	$ISB_{\hat{\vartheta}_m}$	3.546	3.581	3.634	3.753	4.062	3.590	3.617	3.653	3.741	4.053
50	$MI_{\hat{\vartheta}_m}$	3.828	3.541	3.357	<b>3.146</b>	3.266	3.161	3.033	2.955	2.881	3.126
	$IV_{\hat{\vartheta}_m}$	1.600	1.333	1.059	0.649	0.347	0.972	0.819	0.664	0.443	0.242
	$ISB_{\hat{\vartheta}_m}$	2.228	2.208	2.298	2.497	2.919	2.189	2.214	2.291	2.438	2.884
100	$MI_{\hat{\vartheta}_m}$	7.880	5.816	4.504	3.420	2.983	5.115	4.122	3.560	2.874	<b>2.640</b>
	$IV_{\hat{\vartheta}_m}$	5.392	4.102	2.887	1.683	0.841	3.439	2.735	2.141	1.330	0.688
	$ISB_{\hat{\vartheta}_m}$	2.488	1.714	1.617	1.737	2.142	1.676	1.387	1.419	1.544	1.952
Gumbel copula ( $\theta = 1.5$ )											
$m/\hat{b}_n$		$n = 500$					$n = 800$				
		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\hat{\vartheta}_m}$	<b>1.209</b>	1.275	1.403	1.623	1.878	<b>1.006</b>	1.122	1.282	1.539	1.844
	$IV_{\hat{\vartheta}_m}$	0.365	0.294	0.232	0.151	0.086	0.197	0.163	0.131	0.085	0.048
	$ISB_{\hat{\vartheta}_m}$	0.844	0.981	1.171	1.472	1.792	0.809	0.959	1.151	1.454	1.796
50	$MI_{\hat{\vartheta}_m}$	2.258	2.008	1.981	1.999	2.029	1.696	1.640	1.713	1.833	1.954
	$IV_{\hat{\vartheta}_m}$	1.255	0.959	0.745	0.463	0.258	0.845	0.664	0.528	0.340	0.194
	$ISB_{\hat{\vartheta}_m}$	1.003	1.049	1.236	1.536	1.771	0.851	0.976	1.185	1.493	1.760
100	$MI_{\hat{\vartheta}_m}$	8.086	5.692	4.608	3.538	2.746	4.629	3.732	3.388	2.931	2.470
	$IV_{\hat{\vartheta}_m}$	4.829	3.417	2.435	1.325	0.687	2.827	2.123	1.633	0.995	0.542
	$ISB_{\hat{\vartheta}_m}$	3.257	2.275	2.173	2.213	2.059	1.802	1.609	1.755	1.936	1.928
Frank copula ( $\theta = 3.0$ )											
$m/\hat{b}_n$		$n = 500$					$n = 800$				
		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\hat{\vartheta}_m}$	0.395	0.315	<b>0.272</b>	0.304	0.517	0.249	0.206	<b>0.190</b>	0.258	0.502
	$IV_{\hat{\vartheta}_m}$	0.321	0.257	0.203	0.132	0.075	0.199	0.162	0.131	0.088	0.053
	$ISB_{\hat{\vartheta}_m}$	0.075	0.058	0.068	0.172	0.442	0.049	0.044	0.059	0.169	0.449
50	$MI_{\hat{\vartheta}_m}$	1.799	1.294	1.028	0.796	0.752	1.061	0.816	0.681	0.588	0.679
	$IV_{\hat{\vartheta}_m}$	1.256	0.947	0.721	0.446	0.260	0.792	0.626	0.497	0.328	0.197
	$ISB_{\hat{\vartheta}_m}$	0.543	0.346	0.3073	0.349	0.491	0.269	0.190	0.184	0.261	0.483
100	$MI_{\hat{\vartheta}_m}$	14.612	5.023	3.602	2.396	1.608	3.786	2.685	2.185	1.547	1.159
	$IV_{\hat{\vartheta}_m}$	12.019	3.558	2.460	1.423	0.792	2.635	1.951	1.525	0.916	0.501
	$ISB_{\hat{\vartheta}_m}$	2.593	1.465	1.142	0.974	0.816	1.151	0.734	0.660	0.631	0.658

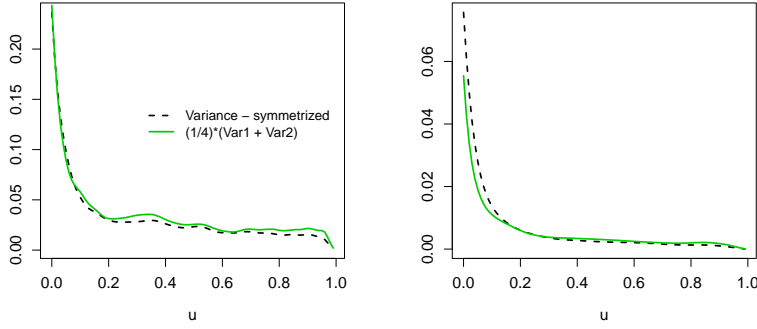
The results in Table 1 show that small values for the bandwidth  $b_n$  are preferred in the Gumbel setting in order to minimize the mean integrated squared error. Since the surface is peaked in the lower left corner, smaller bandwidths are needed to capture the peak. In Figure 3, we graphically depict the difference between the average estimated cross ratio function  $\hat{\vartheta}_m(t_1, t_2)$  and the true cross ratio function (heatplot in left upper panel). Intersections of the averaged estimated cross ratio function (black solid lines) are shown together with pointwise 95% simulation-based confidence intervals (gray dash-dotted



**Fig. 1** Clayton copula with  $\hat{b}_n = 54.6$ ,  $m = 100$  and  $n = 800$ : heatplot representing the difference between the estimated cross ratio function  $\hat{\vartheta}_m(t_1, t_2)$  averaged over the  $M$  replications and the true cross ratio function  $\theta(t_1, t_2)$  (left upper panel) and intersections of the estimated cross ratio surface given  $t_1 = F_1^{-1}(0.5)$  (right upper panel; black solid line),  $t_2 = F_2^{-1}(0.5)$  (left lower panel) and  $F_1^{-1}(u) = F_2^{-1}(u)$  (right lower panel) with pointwise 95% simulation-based confidence intervals (gray dash-dotted lines). True cross ratio curves are graphically depicted in red dashed lines.

lines) for  $\hat{b}_n = 1$ ,  $m = 25$  and  $n = 800$ . Although the nonparametric estimator generally performs well,  $\theta(t_1, t_2)$  is slightly underestimated for small values of  $(t_1, t_2)$  in the lower left corner of the surface.

In the right panel of Figure 2, we again show the relation between the asymptotic variance of the symmetrized estimator  $\hat{\vartheta}_m(t_1, t_2)$  (black dashed line) and the asymptotic variances of  $\hat{\theta}_m(t_1, t_2)$  and  $\hat{\theta}_m^*(t_2, t_1)$  based on simulation results.



**Fig. 2** Empirical variance of  $\hat{\vartheta}_m[F_1^{-1}(u), F_2^{-1}(u)]$  (black dashed line) compared to  $(1/4)$  times the sum of the asymptotic variances of  $\hat{\theta}_m[F_1^{-1}(u), F_2^{-1}(u)]$  and  $\hat{\theta}_m^*[F_1^{-1}(u), F_2^{-1}(u)]$  for the Clayton simulation setting with  $\hat{b}_n = 54.6$ ,  $m = 100$  and  $n = 800$  (left panel) and the Gumbel simulation setting with  $\hat{b}_n = 1$ ,  $m = 25$  and  $n = 800$  (left panel).

#### 7.4 Frank copula function

Finally, a Frank copula function is considered with parameter  $\theta = 3$ . The copula function is defined as

$$C_\theta(u_1, u_2) = -\theta^{-1} \ln \left\{ 1 + \frac{[\exp(-\theta u_1) - 1][\exp(-\theta u_2) - 1]}{\exp(-\theta) - 1} \right\},$$

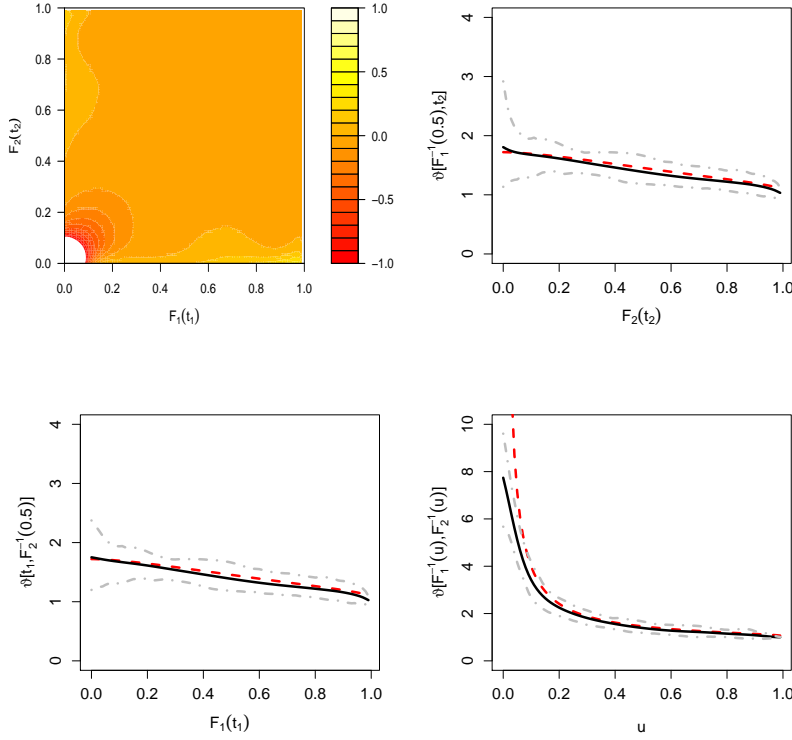
for  $(-\infty, \infty) \setminus \{0\}$ . After some calculus, the expression for the true cross ratio function takes the form:

$$\theta(t_1, t_2) = \phi \frac{\ln(1 + \phi^{-1} \{ \exp[-\theta S_1(t_1)] - 1 \} \{ \exp[-\theta S_2(t_2)] - 1 \})}{\{ \exp[-\theta S_1(t_1)] - 1 \} \{ \exp[-\theta S_2(t_2)] - 1 \}},$$

where  $\phi = [\exp(-\theta) - 1]$ . In Figure 4, similar plots are presented for  $\hat{b}_n = 7.4$ ,  $m = 25$  and  $n = 800$ . Compared to the Clayton copula setting, smaller bandwidth values are selected for both the Gumbel and Frank copulae based on the minimisation of  $MI_\theta$ . This is mainly due to the shape of the cross ratio surface which is non-constant for the latter copulae, thereby requiring a smaller kernel bandwidth to allow for more flexibility when estimating the surface.

### 8 Real data examples

A dataset on recurrent asthma attacks in children (Duchateau et al., 2003) and a dataset on the relationship between food expenditure and net income (Family Expenditure Survey, 1968–1983, Härdle, 1990) illustrate the use of our novel nonparametric estimator for the cross ratio function. The analysis

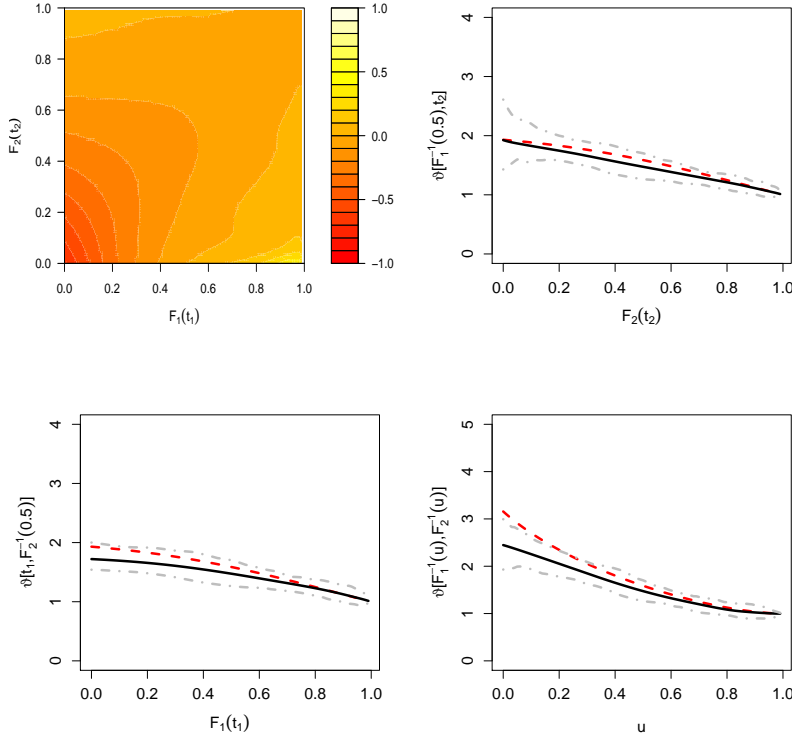


**Fig. 3** Gumbel copula with  $\hat{b}_n = 1$ ,  $m = 25$  and  $n = 800$ : heatmap representing the difference between the estimated cross ratio function  $\hat{\vartheta}_m(t_1, t_2)$  averaged over the  $M$  replications and the true cross ratio function  $\theta(t_1, t_2)$  (left upper panel) and intersections of the estimated cross ratio surface given  $t_1 = F_1^{-1}(0.5)$  (right upper panel; black solid line),  $t_2 = F_2^{-1}(0.5)$  (left lower panel) and  $F_1^{-1}(u) = F_2^{-1}(u)$  (right lower panel) with pointwise 95% simulation-based confidence intervals (gray dash-dotted lines). True cross ratio curves are graphically depicted in red dashed lines.

of the asthma data (event times) is given in the Supplementary Material. In this section we discuss the food expenditure dataset.

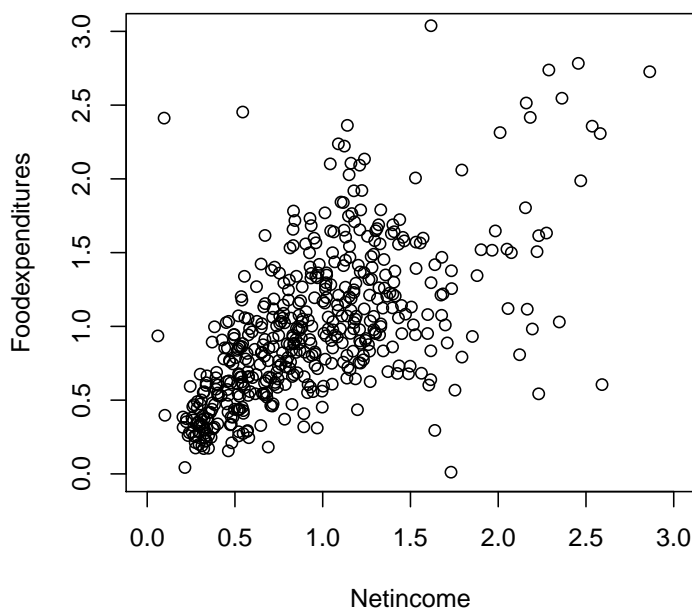
A random subsample of size  $n = 500$  is selected, this dataset is used in our analysis. In Figure 5 we graphically depict food expenditure ( $T_2$ ) versus net income ( $T_1$ ) in multiples of the expenditure sample mean, respectively, the net income sample mean. Summary statistics for the selected subsample are in Table 7 of Section 5 of the Supplementary Material, which also contains a link to the used dataset. From this Table 7 it follows that  $T_1$  and  $T_2$  have similar scales.

In order to select appropriate values for the bandwidths  $b_n$  and  $m$ , we use an ad-hoc procedure inspired by the work of Sen and Xu (2015). For more details, the reader is referred to the Supplementary Material. Based on the selected



**Fig. 4** Frank copula with  $\hat{b}_n = 7.4$ ,  $m = 25$  and  $n = 800$ : heatmap representing the difference between the estimated cross ratio function  $\hat{\vartheta}_m(t_1, t_2)$  averaged over the  $M$  replications and the true cross ratio function  $\theta(t_1, t_2)$  (left upper panel) and intersections of the estimated cross ratio surface given  $t_1 = F_1^{-1}(0.5)$  (right upper panel; black solid line),  $t_2 = F_2^{-1}(0.5)$  (left lower panel) and  $F_1^{-1}(u) = F_2^{-1}(u)$  (right lower panel) with pointwise 95% simulation-based confidence intervals (gray dash-dotted lines). True cross ratio curves are graphically depicted in red dashed lines.

values, a heatmap of the estimated cross ratio surface  $\hat{\vartheta}_m(t_1, t_2)$  is plotted in the upper panel of Figure 6. To visualize the variability of the estimates we include bootstrap based pointwise 95% bootstrap-percentile confidence intervals for the cross ratio function at values of  $T_1$  corresponding to the empirical quartiles (25%, 50% and 75%). Figure 6 shows that the degree of dependence is stronger for low net incomes relative to the mean net income. More specifically, when families have a low relative net income  $T_1 = t_1$  they tend to spend less money on food as compared to families with a net income exceeding  $T_1 > t_1$ . For families with a net income which is large relative to the average ( $t_1$ -values larger than 1), the cross ratio values are approximately equal to one, implying that there is no longer dependence between food expenditure and net income.



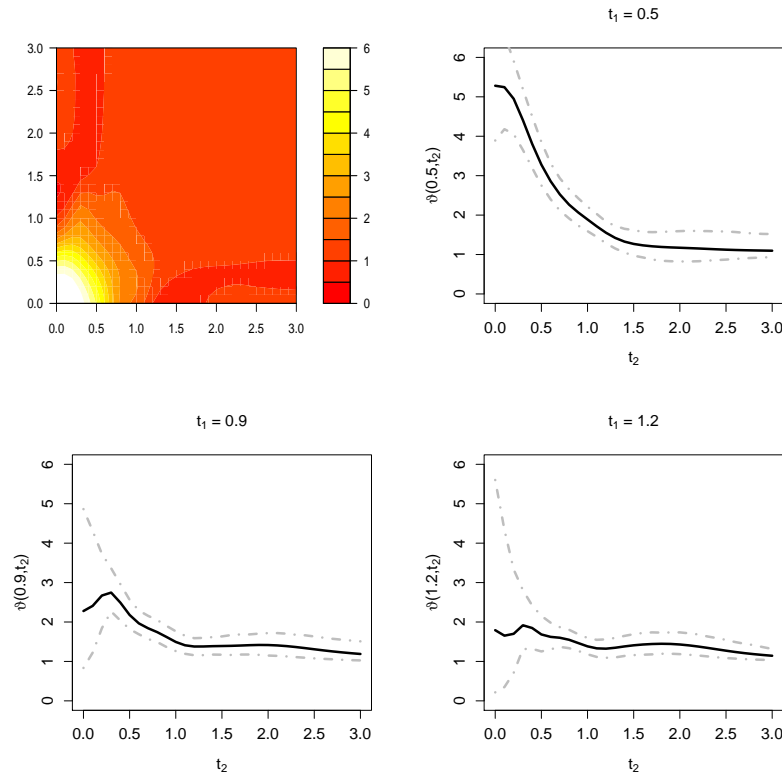
**Fig. 5** Food expenditure versus net income data application: scatterplot of the food expenditure ( $y$ -axis) versus net income ( $x$ -axis) expressed in multiples of the expenditure sample mean, respectively, the net income sample mean. Family Expenditure Survey (1968–1983).

## 9 Discussion

In this paper, we propose a Bernstein-based nonparametric estimator for the cross ratio function which is commonly used to describe local dependence between two correlated variables. As explained in Section 1 there is some earlier work on cross ratio estimation, but the study of a fully nonparametric estimator is new. Our estimator uses Bernstein estimators for the survival copula and its derivatives to estimate conditional cumulative hazards in a nonparametric way. Kernel smoothing is then used to obtain nonparametric estimators for the conditional hazards in the numerator and the denominator of the cross ratio function. The asymptotic distributional behavior of the new estimator and of its symmetrized version is studied. Simulations show the overall good performance of the proposed estimators for  $n \geq 300$ . An example on the relation between net income and food expenditure shows how the estimator gives local information on the relation between these two variables.

A couple of interesting further questions emerge from our findings. On the theoretical side it would be nice to study the bootstrap consistency. An





**Fig. 6** Food expenditure and net income data application: heatplot of the estimated cross ratio surface  $\hat{\vartheta}_m(t_1, t_2)$  and cross ratio curves for  $t_1$  equal to the empirical quartiles (25%, 50% and 75%, i.e., 0.5, 0.9 and 1.2, respectively) (black solid lines) with pointwise 95% bootstrap-percentile confidence intervals (gray dash-dotted lines).

interesting problem on the applied side is to obtain a data-driven method to determine the kernel bandwidth  $b_n$  and the Bernstein order  $m$ . These are challenging open problems. Even for more simple Bernstein-based estimators, e.g. for the conditional density, these problems are unsolved.

### Appendix: Proofs of Theorems 1–3

In this appendix, we present the proofs of Theorems 1–3 in the main text.

#### Proof of Theorem 1

$$\begin{aligned}
& \hat{\lambda}_m(t_1 | T_2 = t_2) - \lambda(t_1 | T_2 = t_2) \\
&= \frac{1}{b_n} \int \left[ \hat{\Lambda}_m(t_1 - b_n u | T_2 = t_2) - \Lambda(t_1 - b_n u | T_2 = t_2) \right] dK_0(u) \\
&\quad + \frac{1}{b_n} \int \Lambda(t_1 - b_n u | T_2 = t_2) dK_0(u) - \lambda(t_1 | T_2 = t_2) \\
&= (A) + (B).
\end{aligned}$$

The non-random term (B) equals

$$(B) = \frac{1}{2} b_n^2 \lambda''(t_1 | T_2 = t_2) \mu_2(K_0) + o(b_n^2) \quad (11)$$

where  $\mu_2(K_0) = \int t^2 K_0(t) dt$ . This is because  $\lambda(t_1 | T_2 = t_2)$  is twice continuously differentiable with respect to  $t_1$ .

For the integrand in (A) we first note that

$$\begin{aligned}
& \hat{\Lambda}_m(t_1 | T_2 = t_2) - \Lambda(t_1 | T_2 = t_2) \\
&= \int_0^{t_1} \frac{d\hat{F}_{t_2}(s)}{1 - \hat{F}_{t_2}(s)} - \int_0^{t_1} \frac{dF_{t_2}(s)}{1 - F_{t_2}(s)} + O\left(\frac{m}{n}\right) \text{ a.s.}
\end{aligned}$$

Indeed, for  $n$  sufficiently large,

$$\left| \int_0^{t_1} \frac{d\hat{F}_{t_2}(s)}{1 - \hat{F}_{t_2}(s)} - \int_0^{t_1} \frac{d\hat{F}_{t_2}(s)}{1 - \hat{F}_{t_2}(s)} \right| \leq \frac{4}{[1 - F_{t_2}(t_1)]^2} O\left(\frac{m}{n}\right)$$

because the maximal jump of  $\hat{F}_{t_2}(\cdot)$  is  $O\left(\frac{m}{n}\right)$  a.s. (see Janssen et al., 2016) and because  $\hat{F}_{t_2}(t_1)$  converges to  $F_{t_2}(t_1)$ . Hence the term (A) can be written as

$$(A) = \frac{1}{b_n} \int_{t_1 - b_n L}^{t_1 + b_n L} \ln \left[ \frac{1 - \hat{F}_{t_2}(s)}{1 - F_{t_2}(s)} \right] dK_0\left(\frac{t_1 - s}{b_n}\right) + O\left(\frac{m}{nb_n}\right) \text{ a.s.}$$

By the mean value theorem we obtain that

$$\ln \left[ \frac{1 - \hat{F}_{t_2}(s)}{1 - F_{t_2}(s)} \right] = -\frac{\hat{F}_{t_2}(s) - F_{t_2}(s)}{1 - F_{t_2}(s)} - \frac{1}{2} \frac{[\hat{F}_{t_2}(s) - F_{t_2}(s)]^2}{[1 - \theta_n(s)]^2}$$

for some  $\theta_n(s)$  between  $F_{t_2}(s)$  and  $\widehat{F}_{t_2}(s)$ . Hence,

$$(A) = \frac{1}{b_n} \int_{-L}^L \frac{\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)}{1 - F_{t_2}(t_1 - b_n u)} dK_0(u) \\ + R_n(t_1, t_2) + O\left(\frac{m}{nb_n}\right) \text{ a.s.}, \quad (12)$$

where

$$R_n(t_1, t_2) = \frac{1}{2b_n} \int_{-L}^L \frac{[\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)]^2}{[1 - \theta_n(t_1 - b_n u)]^2} dK_0(u).$$

From Theorem 3 of Janssen et al. (2016) and the first part of the proof of Lemma 7 available in the Electronic Supplementary Material provided by Janssen et al. (2016) we conclude that

$$\sup_s |\widehat{F}_{t_2}(s) - F_{t_2}(s)| = O\left(m^{1/4} n^{-1/2} (\ln n)^{1/2}\right) + O(n^{-1/2} (\ln n)^{1/2} + m^{-1}) \\ + m^{1/2} n^{-3/4} (\ln n)^{1/2} (\ln \ln n)^{1/4} + m^{13/12} n^{-1} (\ln n)^{1/2} (\ln \ln n)^{1/2} \text{ a.s.} \\ = O\left(m^{1/4} n^{-1/2} (\ln n)^{1/2}\right) \text{ a.s.}, \quad (13)$$

by applying the assumptions in condition (d) of the theorem.

Since we assume that  $K_0$  is a continuous density function of bounded variation, there exist two nondecreasing bounded and continuous functions  $K_{01}$  and  $K_{02}$  such that  $K_0(u) = K_{01}(u) - K_{02}(u)$ . Assume that  $K_{01}$  and  $K_{02}$  are supported on  $[-L, L_1]$  and  $[L_1, L]$  respectively, for some  $-L \leq L_1 \leq L$ . Hence,  $K_{01}(-L) = K_{02}(-L) = 0 = K_{01}(L) = K_{02}(L)$  and  $K_{01}(L_1) = -K_{02}(L_1)$ . Therefore,

$$|R_n(t_1, t_2)| \leq \frac{1}{2b_n} \int_{-L}^{L_1} \frac{[\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)]^2}{[1 - \theta_n(t_1 - b_n u)]^2} dK_{01}(u) \\ + \frac{1}{2b_n} \int_{L_1}^L \frac{[\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)]^2}{[1 - \theta_n(t_1 - b_n u)]^2} dK_{02}(u).$$

Furthermore, since  $\sup_s |\widehat{F}_{t_2}(s) - F_{t_2}(s)| \rightarrow 0$  a.s., we have that  $\sup_s |\theta_n(s) - F_{t_2}(s)| \rightarrow 0$  a.s.; hence for some constant  $C > 0$ ,

$$|R_n(t_1, t_2)| \leq \frac{2C}{b_n [1 - F_{t_2}(t_1)]^2} \left[ \sup_s |\widehat{F}_{t_2}(s) - F_{t_2}(s)| \right]^2 K_{01}(L_1) \\ = O\left(\frac{m^{1/2}}{nb_n} \ln n\right) \text{ a.s.},$$

by using (13). Therefore, under the conditions in (d) we conclude that

$$(nm^{-1/2}b_n)^{1/2}R_n(t_1, t_2) \rightarrow 0 \text{ a.s.} \quad (14)$$

By the mean value theorem the first term in the expression of (A) given in (12) becomes

$$\begin{aligned} & \frac{1}{b_n [1 - F_{t_2}(t_1)]} \int_{-L}^L [\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)] dK_0(u) \\ & + \int_{-L}^L \frac{u [\widehat{F}_{t_2}(t_1 - b_n u) - F_{t_2}(t_1 - b_n u)] f_{t_2}[\theta(u)]}{\{1 - F_{t_2}[\theta(u)]\}^2} dK_0(u) \\ & =: (A_{11}) + \tilde{R}_n(t_1, t_2), \end{aligned} \quad (15)$$

for some  $\theta(u)$  between  $t_1$  and  $t_1 - b_n u$ .

As above, we have that for some constant  $C > 0$ ,

$$\begin{aligned} |\tilde{R}_n(t_1, t_2)| & \leq \frac{C f_{t_2}(t_1)}{[1 - F_{t_2}(t_1)]^2} \left[ \sup_s |\widehat{F}_{t_2}(s) - F_{t_2}(s)| \right] \int_{-L}^L |u| d[K_{01}(u) + K_{02}(u)] \\ & = O\left(m^{1/4} n^{-1/2} (\ln n)^{1/2}\right) \text{ a.s.} \end{aligned}$$

Under the conditions in (d), we have

$$(nm^{-1/2}b_n)^{1/2}\tilde{R}_n(t_1, t_2) \rightarrow 0 \text{ a.s.} \quad (16)$$

For  $(A_{11})$  in equation (15), we write

$$\begin{aligned} (A_{11}) & = \frac{1}{b_n [1 - F_{t_2}(t_1)]} \int_{-L}^L \left\{ \widehat{F}_{t_2}(t_1 - b_n u) - E[\widehat{F}_{t_2}(t_1 - b_n u)] \right\} dK_0(u) \\ & + \frac{1}{b_n [1 - F_{t_2}(t_1)]} \int_{-L}^L \left\{ E[\widehat{F}_{t_2}(t_1 - b_n u)] - F_{t_2}(t_1 - b_n u) \right\} dK_0(u) \\ & =: (A_{111}) + (A_{112}). \end{aligned}$$

For  $(A_{112})$ , which contributes to the bias, note that  $E[\widehat{F}_{t_2}(t_1 - b_n u)] - F_{t_2}(t_1 - b_n u) = -\{E[\widehat{S}_{t_2}(t_1 - b_n u)] - S_{t_2}(t_1 - b_n u)\}$ . In line with Remark 3 in Janssen et al. (2016), we have

$$E[\widehat{S}_{t_2}(t_1 - b_n u)] - S_{t_2}(t_1 - b_n u) = -\frac{1}{2} m^{-1} b [S_1(t_1 - b_n u), S_2(t_2)] + o(m^{-1}),$$

where

$$b(u, v) = (1 - 2v)C^{(2,2)}(u, v) + u(1 - u)C^{(1,1,2)}(u, v) + v(1 - v)C^{(2,2,2)}(u, v).$$

Using partial integration we obtain that

$$(A_{112}) = \frac{1}{2}m^{-1}\phi(t_1, t_2) + o(m^{-1}), \quad (17)$$

where

$$\phi(t_1, t_2) = \frac{b^{(1)}[S_1(t_1), S_2(t_2)]}{1 - F_{t_2}(t_1)} f_1(t_1),$$

with  $b^{(1)}(u, v) = \frac{\partial}{\partial u} b(u, v)$ . For the first term we have, after partial integration,

$$(A_{111}) = \frac{1}{1 - F_{t_2}(t_1)} \left\{ \widehat{f}_{t_2}(t_1) - E[\widehat{f}_{t_2}(t_1)] \right\}, \quad (18)$$

where  $\widehat{f}_{t_2}(t_1)$  is precisely the Bernstein estimator for a conditional density function studied in Janssen et al. (2017).

The proof of the theorem follows directly from (11) – (18) and the Theorem in Janssen et al. (2017) by simply replacing  $Y$  by  $T_1$  and  $X$  by  $T_2$  in the aforementioned paper.

Also note that the term  $\frac{1}{2}m^{-1}\phi(t_1, t_2)$  in the bias vanishes after multiplication with  $(nm^{-1/2}b_n)^{1/2}$ . This is because  $(nm^{-1/2}b_n)^{1/2}m^{-1} \leq n^{1/2}m^{-5/4}b_n^{-1/2} \rightarrow 0$  by the first relation in (d). This proves Theorem 1.  $\square$

**Proof of Theorem 2**

Write

$$\begin{aligned}
& \hat{\lambda}_m(t_1 \mid T_2 > t_2) - \lambda(t_1 \mid T_2 > t_2) \\
&= \frac{1}{b_n} \int \left[ \hat{\Lambda}_m(t_1 - b_n u \mid T_2 > t_2) - \Lambda(t_1 - b_n u \mid T_2 > t_2) \right] dK_0(u) \\
&\quad + \frac{1}{b_n} \int \Lambda(t_1 - b_n u \mid T_2 > t_2) dK_0(u) - \lambda(t_1 \mid T_2 > t_2) \\
&= (\tilde{A}) + (\tilde{B}).
\end{aligned}$$

For the non-random term  $(\tilde{B})$  we have, similar to (11),

$$(\tilde{B}) = O(b_n^2). \quad (19)$$

For  $(\tilde{A})$  we perform analogous operations as we did in the proof of Theorem 1. This gives, in analogy with (12),

$$\begin{aligned}
(\tilde{A}) &= \frac{1}{b_n} \int_{-L}^L \frac{C_{m,n}[S_{1n}(t_1 - b_n u), S_{2n}(t_2)] - C[S_1(t_1 - b_n u), S_2(t_2)]}{C[S_1(t_1 - b_n u), S_2(t_2)]} dK_0(u) \\
&\quad + \tilde{R}_n(t_1, t_2) + O\left(\frac{m^{1/2}}{nb_n}\right) \text{ a.s.}, \quad (20)
\end{aligned}$$

where

$$\begin{aligned}
& \tilde{R}_n(t_1, t_2) = \\
& \frac{1}{2b_n} \int_{-L}^L \frac{\{C_{m,n}[S_{1n}(t_1 - b_n u), S_{2n}(t_2)] - C[S_1(t_1 - b_n u), S_2(t_2)]\}^2}{[1 - \tilde{\theta}_n(t_1 - b_n u)]^2} dK_0(u)
\end{aligned}$$

for some  $\tilde{\theta}_n(t_1 - b_n u)$  between  $C_{m,n}[S_{1n}(t_1 - b_n u), S_{2n}(t_2)]$  and  $C[S_1(t_1 - b_n u), S_2(t_2)]$ . The  $O\left(\frac{m^{1/2}}{nb_n}\right)$  term in (20) comes from the replacement of  $S_{1n}(s-)$  by  $S_{1n}(s)$ .

Indeed, for  $n$  sufficiently large, we have for some constant  $M > 0$ :

$$\begin{aligned}
& \left| \int_0^{t_1} \left\{ \frac{1}{C_{m,n}[S_{1n}(s-), S_{2n}(t_2)]} - \frac{1}{C_{m,n}[S_{1n}(s), S_{2n}(t_2)]} \right\} d_s C_{m,n}[S_{1n}(s), S_{2n}(t_2)] \right| \\
& \leq \frac{M}{C^2[S_1(t_1), S_2(t_2)]} \sup_s \sum_{k=1}^m \sum_{\ell=1}^m C_n \left( \frac{k}{m}, \frac{\ell}{m} \right) \frac{P_{m,\ell}[S_{2n}(t_2)]}{n} \left| P'_{m,k} \left( S_{1n}(s) + \frac{1}{n} \right) \right| \\
& = O\left(\frac{m^{1/2}}{n}\right) \text{ a.s., using Lemma 1 in Janssen et al. (2014).}
\end{aligned}$$

Using that  $C_{m,n}[S_{1n}(s), S_{2n}(t_2)] \rightarrow C[S_1(s), S_2(t_2)]$  a.s. and the fact that  $K_0$  is of bounded variation we can make an argument completely analogous to the one used for  $R_n(t_1, t_2)$  in (12). This give the following bound for  $\tilde{R}_n(t_1, t_2)$ :

$$\tilde{R}_n(t_1, t_2) = O\left(\frac{1}{b_n} \left\{ \sup_s |C_{m,n}[S_{1n}(s), S_{2n}(t_2)] - C[S_1(s), S_2(t_2)]| \right\}^2\right) \text{ a.s.}$$

Now,

$$\begin{aligned} & |C_{m,n}[S_{1n}(s), S_{2n}(t_2)] - C[S_1(s), S_2(t_2)]| \\ & \leq |C_{m,n}[S_{1n}(s), S_{2n}(t_2)] - C[S_{1n}(s), S_{2n}(t_2)]| \\ & \quad + |S_{1n}(s) - S_1(s)| + |S_{2n}(s) - S_2(s)| \end{aligned}$$

by the Lipschitz continuity of  $C$  (see Nelsen 2006).

The supremum of the first term on the right hand side is

$O(n^{-1/2}(\ln \ln n)^{1/2} + m^{-1/2})$  a.s. (see the proof of Theorem 1 in Janssen et al., 2012) and the supremum of the other two terms is  $O(n^{-1/2}(\ln \ln n)^{1/2})$

a.s. So the bound for  $\tilde{R}_n(t_1, t_2)$  is

$$O\left(\frac{n^{-1} \ln \ln n}{b_n} + \frac{m^{-1}}{b_n}\right) \text{ a.s.}$$

Combine this with (19) and (20) to obtain

$$\begin{aligned} & \hat{\lambda}_m(t_1 | T_2 > t_2) - \lambda(t_1 | T_2 > t_2) \\ &= \frac{1}{b_n} \int_{-L}^L \frac{C_{m,n}[S_{1n}(t_1 - b_n u), S_{2n}(t_2)] - C[S_1(t_1 - b_n u), S_2(t_2)]}{C[S_1(t_1 - b_n u), S_2(t_2)]} dK_0(u) \\ &+ O\left(\frac{n^{-1}}{b_n} \ln \ln n + \frac{m^{-1}}{b_n} + \frac{m^{1/2}}{nb_n} + b_n^2\right) \text{ a.s.} \end{aligned}$$

For the first term in the right hand side we write

$$\begin{aligned} & C_{m,n}[S_{1n}(t_1), S_{2n}(t_2)] - C[S_1(t_1), S_2(t_2)] \\ &= \{C_{m,n}[S_1(t_1), S_2(t_2)] - C[S_1(t_1), S_2(t_2)]\} \\ & \quad + C_{m,n}^{(1)}[\theta_{1n}(t_1), \theta_{2n}(t_2)] [S_{1n}(t_1) - S_1(t_1)] \\ & \quad + C_{m,n}^{(2)}[\theta_{1n}(t_1), \theta_{2n}(t_2)] [S_{2n}(t_2) - S_2(t_2)], \end{aligned}$$

with  $(\theta_{1n}(t_1), \theta_{2n}(t_2))$  denoting an intermediate point between  $(S_{1n}(t_1), S_{2n}(t_2))$  and  $(S_1(t_1), S_2(t_2))$ . Now using similar ideas as in Lemma 3 of Janssen et al. (2012) and the convergence rate of the Bernstein approximation given in (5) of the same paper, we obtain

$$\begin{aligned} & C_{m,n}[S_{1n}(t_1), S_{2n}(t_2)] - C[S_1(t_1), S_2(t_2)] \\ &= \frac{1}{n} \sum_{i=1}^n Y_{mi}[S_1(t_1), S_2(t_2)] + O_P(m^{-1}) + o_P(n^{-1/2}) \end{aligned}$$

where the  $Y_m(u_1, u_2)$  are independent zero mean random variables which are bounded. With this

$$\begin{aligned} & \hat{\lambda}_m(t_1 \mid T_2 > t_2) - \lambda(t_1 \mid T_2 > t_2) \\ &= \sum_{i=1}^n W_{in} + O_P \left( \frac{n^{-1}}{b_n} \ln \ln n + \frac{m^{-1}}{b_n} + \frac{m}{nb_n} + b_n^2 + \frac{n^{-1/2}}{b_n} \right), \end{aligned}$$

where

$$W_{in} = \frac{1}{nb_n} \int \frac{Y_{mi}[S_1(t_1 - b_n u), S_2(t_2)]}{C[S_1(t_1 - b_n u), S_2(t_2)]} dK_0(u).$$

Now

$$\begin{aligned} & \text{Var} \left( \sum_{i=1}^n W_{in} \right) = \\ & \int \int \frac{E\{Y_{mi}[S_1(t_1 - b_n u_1), S_2(t_2)]Y_{mi}[S_1(t_1 - b_n u_2), S_2(t_2)]\}}{nb_n^2 C[S_1(t_1 - b_n u_1), S_2(t_2)]C[S_1(t_1 - b_n u_2), S_2(t_2)]} dK_0(u_1) dK_0(u_2) \\ &= O \left( \frac{1}{nb_n^2} \right) \end{aligned}$$

by the boundedness of the  $Y_{mi}$  and the fact that  $K_0$  is of bounded variation. Hence,

$$\sum_{i=1}^n W_{in} = O_P \left( \frac{n^{-1/2}}{b_n} \right)$$

and

$$\begin{aligned} & \hat{\lambda}_m(t_1 \mid T_2 > t_2) - \lambda(t_1 \mid T_2 > t_2) \\ &= O_P \left( \frac{n^{-1/2}}{b_n} + \frac{m^{-1}}{b_n} + \frac{m^{1/2}}{nb_n} + b_n^2 \right). \end{aligned}$$

The imposed conditions in (d) of Theorem 1 and the extra condition  $m^{1/2}b_n \rightarrow \infty$  imply that all the terms in the right hand side vanish after multiplication with  $(nm^{-1/2}b_n)^{1/2}$ .  $\square$



**Proof of Theorem 3**

Linearization of the ratio gives that  $\widehat{\theta}_m(t_1, t_2) - \theta(t_1, t_2)$  has the same limiting distribution as

$$\begin{aligned} & \frac{1}{\lambda(t_1 \mid T_2 > t_2)} [\widehat{\lambda}_m(t_1 \mid T_2 = t_2) - \lambda(t_1 \mid T_2 = t_2)] \\ & - \frac{\lambda(t_1 \mid T_2 = t_2)}{\lambda^2(t_1 \mid T_2 > t_2)} [\widehat{\lambda}_m(t_1 \mid T_2 > t_2) - \lambda(t_1 \mid T_2 > t_2)]. \end{aligned}$$

Multiplication with  $(nm^{-1/2}b_n)^{1/2}$  gives that the second term is  $o_P(1)$  (by Theorem 2) and that the first term is asymptotically normal (by Theorem 1).  $\square$

**Acknowledgements** The authors thank the Editor and two referees for their valuable comments that have led to an improved version of the manuscript. The work was supported by the IAP Research Network P7/13 of the Belgian State (Belgian Science Policy). The third author thanks the National Science Foundation of South Africa for financial support. The fourth author is also extraordinary professor at the North-West University, Potchefstroom, South Africa.

## References

- Bouezmarni, T., Rombouts, J., Taamouti, A. (2009). Asymptotic properties of the Bernstein density copula estimator for  $\alpha$ -mixing data. *Journal of Multivariate Analysis*, 101, 1–10.
- Bouezmarni, T., El Ghouch, A., Taamouti, A. (2013). Bernstein estimator for unbounded copula densities. *Statistics and Risk Modeling*, 30, 343–360.
- Chen, M.-C., Bandeen-Roche, K. (2005). A diagnostic for association in bivariate survival models. *Lifetime Data Analysis*, 11, 245–264.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141–151.
- Duchateau, L., Janssen, P., Kezic, I., Fortpied, C. (2003). Evolution of recurrent asthma event rate over time in frailty models. *Journal of the Royal Statistical Society (Series C)*, 52, 355–363.
- Duchateau, L., Janssen, P. (2008). *The Frailty Model*. New York: Springer.
- Gijbels, I., Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics, Theory and Methods*, 19, 445–464.
- Glidden, D. V. (2007). Pairwise dependence diagnostics for clustered failure-time data. *Biometrika*, 94, 371–385.
- Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.
- Hsu, L., Prentice, R. L. (1996). On assessing the strength of dependency between failure time variables. *Biometrika*, 83, 491–506.
- Hu, T., Nan, B., Lin, X., Robins, J. M. (2011). Time-dependent cross ratio estimation for bivariate failure times. *Biometrika*, 98, 341–354.
- Janssen, P., Swanepoel, J., Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, 142, 1189–1197.
- Janssen, P., Swanepoel, J., Veraverbeke, N. (2014). A note on the asymptotic behavior of the Bernstein estimator of the copula density. *Journal of Multivariate Analysis*, 124, 480–487.
- Janssen, P., Swanepoel, J., Veraverbeke, N. (2016). Bernstein estimation for a copula derivative with application to conditional distribution and regression functionals. *Test*, 25, 351–374.
- Janssen, P., Swanepoel, J., Veraverbeke, N. (2017). Smooth copula-based estimation of the conditional density function with a single covariate. *Journal of Multivariate Analysis*, 159, 39–48.
- Leblanc, A. (2012). On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, 64, 919–943.
- Li, Y., Lin, X. (2006). Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101, 591–603.
- Li, Y., Prentice, R. L., Lin, X. (2008). Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data. *Biometrika*, 95, 947–960.
- Müller, H. G., Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 50, 61–76.
- Nan, B., Lin, X., Lisabeth, L. D., Harlow, S. D. (2006). Piecewise constant cross-ratio estimation for association of age at a marker event and age at menopause. *Journal of the American Statistical Association*, 101, 65–77.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Second edition. New York: Springer.

- Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society Series B – Statistical Methodology*, 44, 414–422.
- Oakes, D. (1986). Semi-parametric inference in a model for association in bivariate survival data. *Biometrika*, 73, 353–361.
- Oakes, D. (1989). Bivariate survival data induced by frailties. *Journal of the American Statistical Association*, 84, 487–493.
- Omelka, M., Gijbels, I., Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Annals of Statistics*, 37, 3023–3058.
- Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74, 105–121.
- Ruppert, D., Cline, D. H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *Annals of Statistics*, 22, 185–210.
- Sancetta, A., Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Economic Theory*, 20, 535–562.
- Sen, B., Xu, G. (2015). Model based bootstrap methods for interval censored data. *Computational Statistics & Data Analysis*, 81, 121–129.
- Sklar, A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'institut de statistique de l'Université de Paris*, 8, 229–231.
- Spierdijk, L. (2008). Nonparametric conditional hazard rate estimation: a local linear approach. *Computational Statistics & Data Analysis*, 52, 2419–2434.
- Swanepoel, J. W. H., van Graan, F. C. (2005). A new kernel distribution function estimator based on a nonparametric transformation of the data. *Scandinavian Journal of Statistics*, 32, 551–562.
- Van Keilegom, I., Veraverbeke, N. (2001). Hazard rate estimation in nonparametric regression with censored data. *Annals of the Institute of Statistical Mathematics*, 53, 730–745.
- Viswanathan, B., Manatunga, A. (2001). Diagnostic plots for assessing the frailty distribution in multivariate survival data. *Lifetime Data Analysis*, 7, 143–155.
- Wienke, A. (2010). *Frailty Models in Survival Analysis*. Boca Raton: Chapman and Hall/CRC.