Supplementary Material for “Nonparametric estimation of the cross ratio function”

Steven Abrams · Paul Janssen · Jan Swanepoel · Noël Veraverbeke

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 Potchefstroom 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 Potchefstroom Campus, Hoffman Street 11, 2520 Potchefstroom, South Africa

1 Asymptotic normality of the symmetrized nonparametric estimator

In this supplementary part, we provide the proof of Theorem 4.

Proof of Theorem 4

From (11) – (18) in the proof of Theorem 1 we have the following asymptotic representation

$$(nm^{-1/2}b_n)^{1/2}[\widehat{\lambda}_m(t_1 \mid T_2 = t_2) - \lambda_m(t_1 \mid T_2 = t_2)]$$
$$= \beta(t_1, t_2) + (nm^{-1/2}b_n)^{1/2} \left\{ \frac{\widehat{f}_{t_2}(t_1) - E[\widehat{f}_{t_2}(t_1)]}{1 - F_{t_2}(t_1)} \right\} + o_p(1). \quad \text{(S1)}$$

In a similar way we have

$$(nm^{-1/2}b_n)^{1/2}[\widehat{\lambda}_m(t_2 \mid T_1 = t_1) - \lambda_m(t_2 \mid T_1 = t_1)]$$
$$= \beta^*(t_2, t_1) + (nm^{-1/2}b_n)^{1/2} \left\{ \frac{\widehat{f}_{t_1}(t_2) - E[\widehat{f}_{t_1}(t_2)]}{1 - F_{t_1}(t_2)} \right\} + o_p(1). \quad \text{(S2)}$$

From (S1) and (S2) it is clear that the bias expression for $\widehat{\vartheta}_m(t_1, t_2) - \theta(t_1, t_2)$ is

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

and that the asymptotic variance will be that of

$$\frac{1}{2} \left[\frac{\widehat{f}_{t_2}(t_1) - E[\widehat{f}_{t_2}(t_1)]}{\lambda(t_1 \mid T_2 > t_2)[1 - F_{t_2}(t_1)]} + \frac{\widehat{f}_{t_1}(t_2) - E[\widehat{f}_{t_1}(t_2)]}{\lambda(t_2 \mid T_1 > t_1)[1 - F_{t_1}(t_2)]} \right].$$

This variance is equal to $\frac{1}{4}[(I) + (II) + (III)]$ where, from Theorem 3,

$$(I) \sim \frac{m^{1/2}}{nb_n} \frac{\|K_0\|^2}{2\sqrt{\pi F_2(t_1)[1 - F_2(t_1)]}} \frac{\theta^2(t_1, t_2)}{f_{t_2}(t_1)}.$$

Similarly,

$$(II) \sim \frac{m^{1/2}}{nb_n} \frac{\|K_0\|^2}{2\sqrt{\pi F_1(t_2)[1 - F_1(t_2)]}} \frac{\theta^2(t_1, t_2)}{f_{t_1}(t_2)}.$$

From the proof of the Theorem in Janssen et al. (2017) it follows that the covariance term (III) is given by

$$(III) \sim \frac{1}{nb_n^2} \frac{1}{\lambda(t_1 \mid T_2 > t_2)\lambda(t_2 \mid T_1 > t_1)} \frac{1}{[1 - F_{t_2}(t_1)][1 - F_{t_1}(t_2)]}$$
$$\times \int_{-L}^{L} \int_{-L}^{L} E\left\{T_{in}[S_1(t_1 - b_n u_1), S_2(t_2)] T_{in}^*[S_1(t_1), S_2(t_2 - b_n u_2)]\right\} \mathrm{d}K_0(u_1) \mathrm{d}K_0(u_2),$$

with $T_{in}(u,v) = Z_{in}(u,v) - \widetilde{Z}_{in}(u,v)$, $T^*_{in}(u,v) = Z^*_{in}(u,v) - \widetilde{Z}^*_{in}(u,v)$, where

$$\begin{aligned}
Z_{in}(u,v) &= m\sum_{k=0}^{m}\sum_{l=0}^{m-1}\left[I\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right.\\
&\qquad \left. - P\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right] P_{m,k}(u)P_{m-1,l}(v),\\
\widetilde{Z}_{in}(u,v) &= mC^{(2)}(u,v)\sum_{k=0}^{m-1}\left[I\left(\frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right) - \frac{1}{m}\right] P_{m-1,l}(v),\\
Z^*_{in}(u,v) &= m\sum_{k=0}^{m-1}\sum_{l=0}^{m}\left[I\left(\frac{k}{m} < \widetilde{U}_i \le \frac{k+1}{m}, \widetilde{V}_i \le \frac{l}{m}\right)\right.\\
&\qquad \left. - P\left(\frac{k}{m} < \widetilde{U}_i \le \frac{k+1}{m}, \widetilde{V}_i \le \frac{l}{m}\right)\right] P_{m-1,k}(u)P_{m,l}(v),\\
\widetilde{Z}^*_{in}(u,v) &= = mC^{(1)}(u,v)\sum_{k=0}^{m-1}\left[I\left(\frac{k}{m} < \widetilde{U}_i \le \frac{k+1}{m}\right) - \frac{1}{m}\right] P_{m-1,k}(u).
\end{aligned}$$

The random vectors $(\widetilde{U}_1, \widetilde{V}_1), \ldots, (\widetilde{U}_n, \widetilde{V}_n)$, with $\widetilde{U}_i = S_1(T_{1i}), \widetilde{V}_i = S_2(T_{2i})$, are independent with copula C and with uniform marginals on $[0,1]$. The quantities above are the survival copula versions of similar quantities used for (distribution) copulas in Janssen et al. (2016, 2017).

For (III) it will be shown that the expectation in the double integral is of order $o(m^{1/2})$ and hence (III) will not contribute to the final asymptotic variance in Theorem 4. We evaluate $E\left[T_{in}(u,v)T^*_{in}(u',v')\right]$ which equals

$$\begin{aligned}
&E\left[Z_{in}(u,v)Z^*_{in}(u',v')\right] - E\left[Z_{in}(u,v)\widetilde{Z}^*_{in}(u',v')\right]\\
&- E\left[\widetilde{Z}_{in}(u,v)Z^*_{in}(u',v')\right] + E\left[\widetilde{Z}_{in}(u,v)\widetilde{Z}^*_{in}(u',v')\right]. \qquad \text{(S3)}
\end{aligned}$$

Now $E\left[Z_{in}(u,v)Z^{*}_{in}(u',v')\right]$ is

$$
\begin{aligned}
& E\Bigg\{ m^2 \sum_{k=0}^{m}\sum_{l=0}^{m-1}\sum_{k'=0}^{m-1}\sum_{l'=0}^{m} \left[I\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right.\\
& \quad \left. - P\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right] \left[I\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}, \widetilde{V}_i \le \frac{l'}{m}\right)\right.\\
& \quad \left. - P\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}, \widetilde{V}_i \le \frac{l'}{m}\right)\right] P_{m,k}(u)P_{m-1,l}(v)P_{m-1,k'}(u')P_{m,l'}(v') \Bigg\}\\
& = m^2 \sum_{k=0}^{m}\sum_{l=0}^{m-1}\sum_{k'=0}^{m-1}\sum_{l'=0}^{m} P\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k \wedge (k'+1)}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{(l+1)\wedge l'}{m}\right)\\
& \qquad\qquad \times\ P_{m,k}(u)P_{m-1,l}(v)P_{m-1,k'}(u')P_{m,l'}(v')\\
& - m^2 \left[\sum_{k=0}^{m}\sum_{l=0}^{m-1} P\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right) P_{m,k}(u)P_{m-1,l}(v)\right]\\
& \times \left[\sum_{k'=0}^{m-1}\sum_{l'=0}^{m} P\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}, \widetilde{V}_i \le \frac{l'}{m}\right) P_{m-1,k'}(u')P_{m,l'}(v')\right]. \qquad \text{(S4)}
\end{aligned}
$$

The first term in (S4) can be written as

$$
\begin{aligned}
& m^2 \sum_{k'=0}^{m-1}\sum_{l=0}^{m-1} P\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\\
& \times\ P_{m-1,k'}(u')P_{m-1,l}(v) \sum_{k=k'+1}^{m} P_{m,k}(u) \sum_{l'=l+1}^{m} P_{m,l'}(v').
\end{aligned}
$$

Bounding the inner sums by one gives that this is smaller than

$$
\begin{aligned}
& m^2 \sum_{k'=0}^{m-1}\sum_{l=0}^{m-1} \left[C\left(\frac{k'+1}{m}, \frac{l+1}{m}\right) - C\left(\frac{k'+1}{m}, \frac{l}{m}\right) - C\left(\frac{k'}{m}, \frac{l+1}{m}\right) + C\left(\frac{k'}{m}, \frac{l}{m}\right)\right]\\
& \qquad \times\ P_{m-1,k'}(u')P_{m-1,l}(v)\\
& = c(u',v) + O(m^{-1}),
\end{aligned}
$$

where c is the density of copula C.
The second term in (S4) is equal to

$$
\left[C^{(1)}(u,v) + O(m^{-1})\right]\left[C^{(2)}(u',v') + O(m^{-1})\right].
$$

We conclude that

$$
E\left[Z_{in}(u,v)Z^{*}_{in}(u',v')\right] = O(1) = o(m^{1/2}).
$$

Furthermore, we have

$$
\begin{aligned}
&E\left[Z_{in}(u,v)\widetilde{Z}^*_{in}(u',v')\right]\\
&= E\Bigg\{m^2C^{(1)}(u',v')\sum_{k=0}^{m}\sum_{l=0}^{m-1}\sum_{k'=0}^{m-1}\left[I\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right.\\
&\qquad - \left. P\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\right]\left[I\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}\right) - \frac{1}{m}\right]\\
&\qquad \times\ P_{m,k}(u)P_{m-1,l}(v)P_{m-1,k'}(u')\Bigg\}\\
&= m^2C^{(1)}(u',v')\sum_{k'=0}^{m-1}\sum_{l=0}^{m-1}P\left(\frac{k'}{m} < \widetilde{U}_i \le \frac{k'+1}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)\\
&\qquad\qquad\qquad \times\ P_{m-1,k'}(u')P_{m-1,l}(v)\sum_{k=k'+1}^{m}P_{m,k}(u)\\
&\quad - mC^{(1)}(u',v')\sum_{k=0}^{m}\sum_{l=0}^{m-1}P\left(\widetilde{U}_i \le \frac{k}{m}, \frac{l}{m} < \widetilde{V}_i \le \frac{l+1}{m}\right)P_{m,k}(u)P_{m-1,l}(v)\\
&\le C^{(1)}(u',v')\left[c(u',v) + O(m^{-1})\right] - C^{(1)}(u',v')\left[C^{(2)}(u,v) + O(m^{-1})\right]\\
&= O(1) = o(m^{1/2}).
\end{aligned}
$$

The other two terms in (S3) can be handled in the same way. We therefore conlcude that (III) is of lower order than (I) and (II). □

2 MISE and bias-variance relationship

The mean integrated squared error $MISE_{\widehat{\theta}_m}$ for an estimator $\widehat{\theta}_m(t_1,t_2)$ of the cross ratio function $\theta(t_1,t_2)$ is defined as

$$
MISE_{\widehat{\theta}_m} \equiv E_{\widehat{\theta}_m}\left\{\int_{a_1}^{b_1}\int_{a_2}^{b_2}\left[\widehat{\theta}_m(t_1,t_2) - \theta(t_1,t_2)\right]^2 \mathrm{d}t_1\mathrm{d}t_2\right\}.
$$

The bias-variance relationship decomposes $MISE_{\widehat{\theta}_m}$ in a bias and variance term as follows:

$$
\begin{aligned}
MISE_{\widehat{\theta}_m} &= E\left(\int_{a_1}^{b_1}\int_{a_2}^{b_2}\left\{\widehat{\theta}_m(t_1,t_2) - E[\widehat{\theta}_m(t_1,t_2)]\right\}^2 \mathrm{d}t_1\mathrm{d}t_2\right)\\
&\quad + \int_{a_1}^{b_1}\int_{a_2}^{b_2}\left\{E[\widehat{\theta}_m(t_1,t_2)] - \theta(t_1,t_2)\right\}^2 \mathrm{d}t_1\mathrm{d}t_2\\
&\equiv IVAR_{\widehat{\theta}_m} + ISBIAS_{\widehat{\theta}_m},
\end{aligned}
$$

where $IVAR_{\widehat{\theta}_m}$ represents the integrated variance and $ISBIAS_{\widehat{\theta}_m}$ is the integrated squared bias. In order to estimate the mean integrated squared error, we use the following approximation based on M simulation replications:

$$MI_{\widehat{\theta}_m} = \frac{1}{M}\sum_{r=1}^{M}\left\{\Delta_1\Delta_2\sum_{k=1}^{N_1}\sum_{l=1}^{N_2}\left[\widehat{\theta}_m^{(r)}(t_{1[k]},t_{2[l]}) - \theta(t_{1[k]},t_{2[l]})\right]^2\right\},$$

where $\widehat{\theta}_m^{(r)}$ is the cross ratio estimator based on the r-th simulated dataset, $\Delta_1 = (b_1 - a_1)/(N_1 - 1)$, $\Delta_2 = (b_2 - a_2)/(N_2 - 1)$, $t_{1[k]} = a_1 + (b_1 - a_1)(k - 1)/(N_1 - 1)$ and $t_{2[l]} = a_2 + (b_2 - a_2)(l - 1)/(N_2 - 1)$, for $k = 1, \ldots, N_1$, $l = 1, \ldots, N_2$. Consequently, we can write down similar approximations for the integrated variance and integrated squared bias, say $IV_{\widehat{\theta}_m}$ and $ISB_{\widehat{\theta}_m}$, respectively, as follows:

$$IV_{\widehat{\theta}_m} = \frac{1}{M}\sum_{r=1}^{M}\left\{\Delta_1\Delta_2\sum_{k=1}^{N_1}\sum_{l=1}^{N_2}\left[\widehat{\theta}_m^{(r)}(t_{1[k]},t_{2[l]}) - \widehat{\theta}^{(\cdot)}(t_{1[k]},t_{2[l]})\right]^2\right\},$$

$$ISB_{\widehat{\theta}_m} = \Delta_1\Delta_2\sum_{k=1}^{N_1}\sum_{l=1}^{N_2}\left[\widehat{\theta}_m^{(\cdot)}(t_{1[k]},t_{2[l]}) - \theta(t_{1[k]},t_{2[l]})\right]^2,$$

where $\widehat{\theta}^{(\cdot)}(t_{1[k]},t_{2[l]})$ is equal to

$$\widehat{\theta}^{(\cdot)}(t_{1[k]},t_{2[l]}) = \frac{1}{M}\sum_{r=1}^{M}\widehat{\theta}_m^{(r)}(t_{1[k]},t_{2[l]}).$$

3 Bandwidth selection

In order to select the bandwidths b_n and m we first consider the bootstrap version of the $MISE_{\widehat{\theta}_m}$ given by

$$MISE_{\widehat{\theta}_m}(m,b_n) = E_b\left\{\int\int\left[\widehat{\theta}_m^{(b)}(t_1,t_2;b_n) - \widehat{\theta}_m(t_1,t_2;b_n)\right]^2 \mathrm{d}t_1\mathrm{d}t_2\right\},$$

where $\widehat{\theta}_m^{(b)}(t_1,t_2;b_n)$ is the estimated cross-ratio function based on the b-th bootstrap sample (and stressing the dependence on b_n as compared to the notation in the main text). The $MISE_{\widehat{\theta}_m}$ can be estimated by resampling the original data nonparametrically B times, having

$$\widehat{MISE}_{\widehat{\theta}_m} = \frac{1}{B}\sum_{b=1}^{B}\int\int[\widehat{\theta}_m^{(b)}(t_1,t_2;b_n) - \widehat{\theta}_m(t_1,t_2;b_n)]^2\mathrm{d}t_1\mathrm{d}t_2.$$

Hall (1990) noted, however, that this quantity is not useful to select the bandwidth (in various uni-dimensional settings) due to *vanishing bootstrap bias estimates*, thereby leading to a quantity which is mainly dominated by the variance. As the variance decreases with increasing bandwidths b_n, we would always select a potentially oversmoothed estimated cross ratio function.

Alternatively, we will consider a generalisation of the bootstrap method proposed by Sen and Xu (2015) to obtain an estimate of the $MISE_{\widehat{\theta}_m}$ as follows:

$$\widehat{MISE}_{\widehat{\theta}_m} = \frac{1}{B}\sum_{b=1}^{B}\int\int[\widehat{\theta}_m^{(b)}(t_1,t_2;b_n) - \widehat{\theta}_{m_0}(t_1,t_2;b_{n_0})]^2 \mathrm{d}t_1\mathrm{d}t_2,$$

depending on an initial choice m_0 and b_{n_0} for the bandwidths. Note that the bandwidth selection procedure for the estimator $\widehat{\theta}_m^*(t_2,t_1)$ is performed based on the proposed criterion and replacing the estimates $\widehat{\theta}_m(t_1,t_2)$ by the counterparts obtained from reversing the roles of T_1 and T_2. Since the criterion depends on an initial choice of m_0 and b_{n_0} for the bandwidths, we used the values $m_0 = 50$ and $b_{n_0} = 0.178$. The optimal values for b_n and m are then obtained by minimizing the criterion $\widehat{MISE}_{\widehat{\theta}_m}$. We illustrate this here for the food expenditure and net income example introduced in detail in the main text. More specifically, in Figure 1, we present these criterion values for $m = 25$ (dashed lines) and $m = 50$ (dash-dotted lines) for both $\widehat{\theta}_m(t_1,t_2)$ (black lines) and $\widehat{\theta}_m^*(t_2,t_1)$ (red lines). The selected optimal values are those minimizing the criterion and are equal to $b_n = 0.2$ and 1.5 for $\widehat{\theta}_m(t_1,t_2)$ and $\widehat{\theta}_m^*(t_2,t_1)$, respectively, and $m = 25$ in both cases. We use these selected bandwidths to obtain $\widehat{\vartheta}_m(t_1,t_2)$ and to construct the figures in the real data application in the main text. Needless to say, more research is required in terms of bandwidth selection for both the kernel bandwidth b_n as well as the Bernstein order m.

4 Additional simulation results

In this section, we provide additional simulation results for sample sizes $n = 100$ and $n = 300$ under the different simulation settings described in detail in the main text (Tables 1–3). In general, the proposed estimators seem to work well in case $n = 300$, and show a lot of variability when the sample size is small ($n = 100$). However, if the cross ratio function is flat (i.e., in the Clayton setting), the proposed estimator still seems to perform well. Whenever the shape of the surface is more irregular, a larger sample size is required in order to estimate the true underlying cross ratio function appropriately. Therefore, based on the considered simulation settings, we advice to use the Bernstein-based estimator when the sample size is at least equal to 300.

In Tables 1–3 we use the same bandwidth for $\widehat{\theta}_m(t_1,t_2)$ and $\widehat{\theta}_m^*(t_2,t_1)$. Recall from Remark 4 in the main text that the scale ratio of T_1 and T_2 is 0.6. Simulations with bandwidth $b_{n1} = b_n$ and $b_{n2} = cb_n$ with – in each simulation run – c choosen as the estimated scale ratio of T_1 and T_2, reveal that the optimal (in terms of MI) choice for (m, b_n) remains the same (see Tables 4–6 for details).

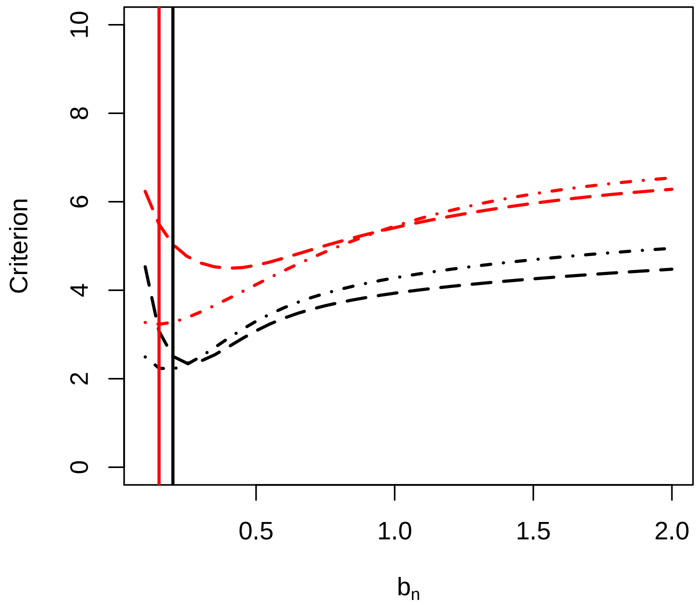

Fig. 1 $\widehat{MISE}_{\widehat{\theta}_m}$ (black lines) and $\widehat{MISE}_{\widehat{\theta}^*_m}$ (red lines) values for different values of the bandwidths b_n and m (dashed lines: $m = 25$ and dash-dotted lines: $m = 50$).

Table 1 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_n, m and n and Clayton copula function with parameter $\theta = 0.5$. Minimum MI-values are highlighted in bold.

		Clayton copula ($\theta = 0.5$) - $\widehat{\theta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	7.636	7.837	6.415	5.479	**4.856**	4.650	4.504	4.381	4.220	4.208
	$IV_{\widehat{\theta}_m}$	4.207	4.325	2.640	1.427	0.646	1.141	0.974	0.785	0.496	0.262
	$ISB_{\widehat{\theta}_m}$	3.429	3.512	3.775	4.052	4.210	3.508	3.530	3.597	3.724	3.946
50	$MI_{\widehat{\theta}_m}$	45.160	> 100	32.752	7.228	5.048	7.371	6.225	5.292	4.375	**3.879**
	$IV_{\widehat{\theta}_m}$	42.691	> 100	29.670	3.733	1.515	4.475	3.650	2.762	1.649	0.831
	$ISB_{\widehat{\theta}_m}$	2.469	> 100	3.082	3.495	3.533	2.896	2.575	2.531	2.726	3.048
100	$MI_{\widehat{\theta}_m}$	> 100	> 100	> 100	15.222	6.498	23.345	16.965	11.796	7.098	4.416
	$IV_{\widehat{\theta}_m}$	> 100	> 100	> 100	11.854	3.822	17.576	13.422	9.500	5.279	2.569
	$ISB_{\widehat{\theta}_m}$	> 100	> 100	> 100	3.368	2.676	5.768	3.543	2.296	1.819	1.846

		Clayton copula ($\theta = 0.5$) - $\widehat{\theta}^*_m(t_2, t_1)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	6.832	6.215	5.618	4.943	4.616	4.538	4.411	4.299	4.189	4.335
	$IV_{\widehat{\theta}^*_m}$	3.312	2.696	1.928	0.967	0.434	1.036	0.857	0.668	0.425	0.230
	$ISB_{\widehat{\theta}^*_m}$	3.520	3.519	3.690	3.976	4.182	3.502	3.553	3.631	3.764	4.105
50	$MI_{\widehat{\theta}^*_m}$	34.282	13.733	8.477	5.443	**4.298**	6.811	5.821	4.984	4.104	3.931
	$IV_{\widehat{\theta}^*_m}$	30.729	10.748	5.328	2.121	0.911	4.167	3.246	2.257	1.122	0.548
	$ISB_{\widehat{\theta}^*_m}$	3.553	2.985	3.149	3.322	3.387	2.644	2.575	2.727	2.982	3.383
100	$MI_{\widehat{\theta}^*_m}$	> 100	> 100	82.291	8.418	4.762	25.811	13.269	8.864	5.128	**3.805**
	$IV_{\widehat{\theta}^*_m}$	> 100	> 100	78.230	5.602	2.158	21.462	11.017	6.985	3.201	1.4883
	$ISB_{\widehat{\theta}^*_m}$	> 100	> 100	4.061	2.816	2.604	4.350	2.251	1.879	1.927	2.317

		Clayton copula ($\theta = 0.5$) - $\widehat{\vartheta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	5.578	5.319	4.990	4.677	4.441	4.135	4.087	4.0650	4.045	4.180
	$IV_{\widehat{\vartheta}^*_m}$	2.159	1.834	1.362	0.772	0.359	0.635	0.549	0.454	0.303	0.162
	$ISB_{\widehat{\vartheta}^*_m}$	3.419	3.485	3.628	3.905	4.082	3.500	3.538	3.611	3.742	4.018
50	$MI_{\widehat{\vartheta}^*_m}$	22.435	31.975	9.554	5.048	**4.038**	5.206	4.566	4.143	3.729	3.644
	$IV_{\widehat{\vartheta}^*_m}$	19.141	29.240	6.574	1.798	0.766	2.479	2.025	1.533	0.885	0.445
	$ISB_{\widehat{\vartheta}^*_m}$	3.294	2.735	2.980	3.250	3.272	2.727	2.541	2.610	2.844	3.199
100	$MI_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	8.365	4.386	15.771	9.767	6.868	4.394	**3.252**
	$IV_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	5.477	1.915	10.810	6.998	4.874	2.587	1.228
	$ISB_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	2.888	2.471	4.961	2.769	1.994	1.807	2.024

Table 2 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_n, m and n and Gumbel copula function with parameter $\theta = 1.5$. Minimum MI-values are highlighted in bold.

		Gumbel copula ($\theta = 1.5$) - $\widehat{\theta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	12.577	9.161	4.262	2.630	**2.230**	2.100	1.929	**1.867**	1.889	1.982
	$IV_{\widehat{\theta}_m}$	10.852	7.797	2.965	1.120	0.534	1.229	0.979	0.755	0.490	0.286
	$ISB_{\widehat{\theta}_m}$	1.725	1.364	1.297	1.510	1.696	0.871	0.950	1.112	1.399	1.696
50	$MI_{\widehat{\theta}_m}$	> 100	> 100	> 100	5.004	3.248	5.549	4.247	3.503	2.866	2.469
	$IV_{\widehat{\theta}_m}$	> 100	> 100	> 100	3.229	1.545	4.214	3.133	2.287	1.365	0.765
	$ISB_{\widehat{\theta}_m}$	22.806	19.932	4.991	1.775	1.704	1.335	1.115	1.217	1.501	1.704
100	$MI_{\widehat{\theta}_m}$	> 100	> 100	> 100	12.454	4.688	> 100	61.982	13.643	6.505	4.212
	$IV_{\widehat{\theta}_m}$	> 100	> 100	> 100	10.260	3.071	> 100	57.950	10.958	3.974	1.994
	$ISB_{\widehat{\theta}_m}$	> 100	> 100	79.348	2.195	1.617	10.704	4.032	2.686	2.532	2.218

		Gumbel copula ($\theta = 1.5$) - $\widehat{\theta}^*_m(t_2, t_1)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	13.506	5.529	2.997	2.302	**2.139**	1.983	**1.885**	1.902	1.969	2.097
	$IV_{\widehat{\theta}^*_m}$	12.117	4.230	1.605	0.750	0.391	1.106	0.853	0.655	0.422	0.249
	$ISB_{\widehat{\theta}^*_m}$	1.389	1.299	1.463	1.632	1.840	0.878	1.032	1.247	1.547	1.848
50	$MI_{\widehat{\theta}^*_m}$	> 100	21.313	6.950	3.642	2.723	6.122	4.679	3.506	2.793	2.440
	$IV_{\widehat{\theta}^*_m}$	> 100	19.097	5.217	1.917	0.901	4.837	3.467	2.085	1.126	0.619
	$ISB_{\widehat{\theta}^*_m}$	9.701	2.216	1.733	1.725	1.822	1.285	1.212	1.421	1.668	1.821
100	$MI_{\widehat{\theta}^*_m}$	> 100	> 100	22.642	6.989	3.869	20.283	11.741	8.475	5.506	3.786
	$IV_{\widehat{\theta}^*_m}$	> 100	> 100	19.225	4.673	2.082	15.811	8.893	5.885	3.045	1.659
	$ISB_{\widehat{\theta}^*_m}$	> 100	> 100	3.417	2.316	1.787	4.474	2.848	2.590	2.462	2.128

		Gumbel copula ($\theta = 1.5$) - $\widehat{\vartheta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	7.291	4.368	2.530	2.030	**1.943**	1.497	**1.487**	1.563	1.720	1.914
	$IV_{\widehat{\vartheta}^*_m}$	5.784	3.086	1.166	0.464	0.184	0.626	0.499	0.388	0.252	0.147
	$ISB_{\widehat{\vartheta}^*_m}$	1.507	1.282	1.364	1.566	1.759	0.871	0.987	1.175	1.468	1.767
50	$MI_{\widehat{\vartheta}^*_m}$	> 100	> 100	73.405	3.285	2.439	3.817	3.016	2.575	2.325	2.176
	$IV_{\widehat{\vartheta}^*_m}$	> 100	> 100	70.900	1.551	0.703	2.531	1.871	1.270	0.750	0.421
	$ISB_{\widehat{\vartheta}^*_m}$	13.828	7.283	2.505	1.734	1.736	1.286	1.145	1.305	1.575	1.756
100	$MI_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	6.073	3.086	> 100	20.383	7.236	4.525	3.253
	$IV_{\widehat{\vartheta}^*_m}$	> 100	> 1001	> 100	3.890	1.435	> 100	17.183	4.645	2.055	1.098
	$ISB_{\widehat{\vartheta}^*_m}$	> 100	> 100	21.915	2.183	1.651	6.261	3.200	2.592	2.471	2.155

Table 3 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_n, m and n and Frank copula function with parameter $\theta = 3.0$. Minimum MI-values are highlighted in bold.

		Frank copula ($\theta = 3.0$) - $\widehat{\theta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	7.038	3.958	2.651	1.398	**0.896**	1.365	1.020	0.764	**0.577**	0.602
	$IV_{\widehat{\theta}_m}$	6.058	3.458	2.377	1.176	0.582	1.164	0.916	0.690	0.442	0.246
	$ISB_{\widehat{\theta}_m}$	0.981	0.500	0.274	0.222	0.314	0.201	0.104	0.074	0.135	0.356
50	$MI_{\widehat{\theta}_m}$	54.063	39.912	12.865	3.581	1.805	5.674	3.671	2.583	1.630	1.131
	$IV_{\widehat{\theta}_m}$	48.411	36.855	11.732	2.998	1.402	4.486	3.036	2.207	1.296	0.698
	$ISB_{\widehat{\theta}_m}$	5.653	3.058	1.132	0.583	0.403	1.188	0.635	0.377	0.334	0.433
100	$MI_{\widehat{\theta}_m}$	> 100	> 100	> 100	7.854	3.362	20.485	12.872	7.737	4.220	2.388
	$IV_{\widehat{\theta}_m}$	> 100	> 100	> 100	6.531	2.812	16.644	10.961	6.520	3.325	1.663
	$ISB_{\widehat{\theta}_m}$	> 100	> 100	53.211	1.323	0.550	3.841	1.911	1.218	0.896	0.725

		Frank copula ($\theta = 3.0$) - $\widehat{\theta}^*_m(t_2, t_1)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	5.609	3.517	1.996	1.152	**0.929**	1.221	0.929	0.726	**0.614**	0.739
	$IV_{\widehat{\theta}^*_m}$	4.765	3.095	1.706	0.843	0.435	1.074	0.813	0.598	0.359	0.210
	$ISB_{\widehat{\theta}^*_m}$	0.844	0.422	0.289	0.309	0.497	0.148	0.116	0.128	0.255	0.529
50	$MI_{\widehat{\theta}^*_m}$	> 100	50.682	8.315	2.301	1.456	4.925	3.322	2.366	1.571	1.237
	$IV_{\widehat{\theta}^*_m}$	> 100	47.657	7.515	1.842	0.952	3.865	2.740	1.905	1.080	0.638
	$ISB_{\widehat{\theta}^*_m}$	9.946	3.025	0.800	0.459	0.504	1.061	0.583	0.460	0.491	0.599
100	$MI_{\widehat{\theta}^*_m}$	> 100	> 100	63.273	4.602	2.385	15.119	9.319	6.180	3.476	2.214
	$IV_{\widehat{\theta}^*_m}$	> 100	> 100	60.647	3.874	1.930	12.280	7.745	4.922	2.460	1.397
	$ISB_{\widehat{\theta}^*_m}$	> 100	37.901	2.626	0.728	0.455	2.838	1.574	1.258	1.016	0.817

		Frank copula ($\theta = 3.0$) - $\widehat{\vartheta}_m(t_1, t_2)$									
		$n = 100$					$n = 300$				
$m/\widehat{b}_n$		1	2.7	7.4	20.1	54.6	1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	3.816	2.266	1.443	0.871	**0.710**	0.765	0.579	0.454	**0.410**	0.558
	$IV_{\widehat{\vartheta}^*_m}$	2.925	1.823	1.176	0.616	0.318	0.603	0.478	0.359	0.221	0.122
	$ISB_{\widehat{\vartheta}^*_m}$	0.891	0.443	0.268	0.255	0.392	0.162	0.101	0.096	0.189	0.436
50	$MI_{\widehat{\vartheta}^*_m}$	61.789	23.597	5.860	1.816	1.070	3.358	2.201	1.584	1.097	0.902
	$IV_{\widehat{\vartheta}^*_m}$	55.058	20.813	4.932	1.312	0.640	2.256	1.613	1.180	0.696	0.393
	$ISB_{\widehat{\vartheta}^*_m}$	6.731	2.784	0.928	0.505	0.430	1.102	0.588	0.405	0.402	0.508
100	$MI_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	3.992	1.825	10.946	6.710	4.343	2.572	1.650
	$IV_{\widehat{\vartheta}^*_m}$	> 100	> 100	> 100	3.015	1.372	7.699	5.040	3.142	1.633	0.891
	$ISB_{\widehat{\vartheta}^*_m}$	> 100	> 100	16.048	0.977	0.453	3.246	1.670	1.201	0.939	0.759

Table 4 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_{n1}, m and n, $b_{n2} = cb_{n1}$ ($c = 0.6$), and Clayton copula function with parameter $\theta = 0.5$ ($n = 300$). Minimum MI-values are highlighted in bold.

		Clayton copula ($\theta = 0.5$) - $\widehat{\theta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	4.650	4.504	4.381	4.220	4.208
	$IV_{\widehat{\theta}_m}$	1.141	0.974	0.785	0.496	0.262
	$ISB_{\widehat{\theta}_m}$	3.508	3.530	3.597	3.724	3.946
50	$MI_{\widehat{\theta}_m}$	7.371	6.225	5.292	4.375	**3.879**
	$IV_{\widehat{\theta}_m}$	4.475	3.650	2.762	1.649	0.831
	$ISB_{\widehat{\theta}_m}$	2.896	2.575	2.531	2.726	3.048
100	$MI_{\widehat{\theta}_m}$	23.345	16.965	11.796	7.098	4.416
	$IV_{\widehat{\theta}_m}$	17.576	13.422	9.500	5.279	2.569
	$ISB_{\widehat{\theta}_m}$	5.768	3.543	2.296	1.819	1.846

		Clayton copula ($\theta = 0.5$) - $\widehat{\theta}^*_m(t_2, t_1)$				
$m/\widehat{b}_{n2} = cb_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	4.547	4.4750	4.369	4.229	4.223
	$IV_{\widehat{\theta}^*_m}$	1.082	0.954	0.788	0.542	0.322
	$ISB_{\widehat{\theta}^*_m}$	3.496	3.521	3.582	3.687	3.901
50	$MI_{\widehat{\theta}^*_m}$	7.162	6.324	5.508	4.488	**3.928**
	$IV_{\widehat{\theta}^*_m}$	4.431	3.751	2.893	1.628	0.785
	$ISB_{\widehat{\theta}^*_m}$	2.731	2.573	2.615	2.861	3.143
100	$MI_{\widehat{\theta}^*_m}$	30.864	19.091	11.291	6.644	4.219
	$IV_{\widehat{\theta}^*_m}$	25.147	16.149	9.277	4.764	2.153
	$ISB_{\widehat{\theta}^*_m}$	5.717	2.941	2.014	1.881	2.066

		Clayton copula ($\theta = 0.5$) - $\widehat{\vartheta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}/\widehat{b}_{n2}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	4.109	4.094	4.071	4.044	4.119
	$IV_{\widehat{\vartheta}^*_m}$	0.640	0.572	0.483	0.339	0.196
	$ISB_{\widehat{\vartheta}^*_m}$	3.499	3.523	3.587	3.705	3.923
50	$MI_{\widehat{\vartheta}^*_m}$	5.330	4.715	4.271	3.832	3.626
	$IV_{\widehat{\vartheta}^*_m}$	2.546	2.164	1.713	1.047	0.534
	$ISB_{\widehat{\vartheta}^*_m}$	2.784	2.552	2.557	2.785	3.092
100	$MI_{\widehat{\vartheta}^*_m}$	17.359	11.518	7.588	4.900	**3.416**
	$IV_{\widehat{\vartheta}^*_m}$	11.716	8.342	5.477	3.078	1.473
	$ISB_{\widehat{\vartheta}^*_m}$	5.644	3.176	2.111	1.822	1.943

Table 5 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_{n1}, m and n, $b_{n2} = cb_{n1}$ ($c = 0.6$), and Gumbel copula function with parameter $\theta = 1.5$ ($n = 300$). Minimum MI-values are highlighted in bold.

		Gumbel copula ($\theta = 1.5$) - $\widehat{\theta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	2.100	1.929	**1.867**	1.889	1.982
	$IV_{\widehat{\theta}_m}$	1.229	0.979	0.755	0.490	0.286
	$ISB_{\widehat{\theta}_m}$	0.871	0.950	1.112	1.399	1.696
50	$MI_{\widehat{\theta}_m}$	5.549	4.247	3.503	2.866	2.469
	$IV_{\widehat{\theta}_m}$	4.214	3.133	2.287	1.365	0.765
	$ISB_{\widehat{\theta}_m}$	1.335	1.115	1.217	1.501	1.704
100	$MI_{\widehat{\theta}_m}$	> 100	61.982	13.643	6.505	4.212
	$IV_{\widehat{\theta}_m}$	> 100	57.950	10.958	3.974	1.994
	$ISB_{\widehat{\theta}_m}$	10.704	4.032	2.686	2.532	2.218

		Gumbel copula ($\theta = 1.5$) - $\widehat{\theta}^*_m(t_2, t_1)$				
$m/\widehat{b}_{n2} = cb_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	2.068	1.913	**1.883**	1.933	2.019
	$IV_{\widehat{\theta}^*_m}$	1.207	0.972	0.772	0.530	0.328
	$ISB_{\widehat{\theta}^*_m}$	0.861	0.942	1.112	1.403	1.691
50	$MI_{\widehat{\theta}^*_m}$	7.040	5.397	4.114	3.125	2.565
	$IV_{\widehat{\theta}^*_m}$	5.602	4.212	2.831	1.555	0.833
	$ISB_{\widehat{\theta}^*_m}$	1.439	1.185	1.284	1.570	1.732
100	$MI_{\widehat{\theta}^*_m}$	25.147	15.387	10.296	6.805	4.483
	$IV_{\widehat{\theta}^*_m}$	19.271	12.041	7.624	4.259	2.216
	$ISB_{\widehat{\theta}^*_m}$	5.880	3.347	2.672	2.547	2.267

		Gumbel copula ($\theta = 1.5$) - $\widehat{\vartheta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}/\widehat{b}_{n2}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	1.514	**1.474**	1.530	1.681	1.863
	$IV_{\widehat{\vartheta}^*_m}$	0.650	0.530	0.419	0.281	0.170
	$ISB_{\widehat{\vartheta}^*_m}$	0.864	0.944	1.110	1.400	1.693
50	$MI_{\widehat{\vartheta}^*_m}$	4.088	3.205	2.706	2.405	2.203
	$IV_{\widehat{\vartheta}^*_m}$	2.724	2.071	1.466	0.876	0.489
	$ISB_{\widehat{\vartheta}^*_m}$	1.364	1.134	1.239	1.529	1.714
100	$MI_{\widehat{\vartheta}^*_m}$	> 100	21.505	7.766	4.915	3.506
	$IV_{\widehat{\vartheta}^*_m}$	> 100	18.019	5.118	2.392	1.271
	$ISB_{\widehat{\vartheta}^*_m}$	7.043	3.486	2.648	2.524	2.235

Table 6 MI, IV and ISB for the different estimators $\widehat{\theta}_m$, $\widehat{\theta}^*_m$ and $\widehat{\vartheta}_m$, and different choices of b_{n1}, m and n, $b_{n2} = cb_{n1}$ ($c = 0.6$), and Frank copula function with parameter $\theta = 3.0$ ($n = 300$). Minimum MI-values are highlighted in bold.

		Frank copula ($\theta = 3.0$) - $\widehat{\theta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}_m}$	1.365	1.020	0.764	**0.577**	0.602
	$IV_{\widehat{\theta}_m}$	1.164	0.916	0.690	0.442	0.246
	$ISB_{\widehat{\theta}_m}$	0.201	0.104	0.074	0.135	0.356
50	$MI_{\widehat{\theta}_m}$	5.674	3.671	2.583	1.630	1.131
	$IV_{\widehat{\theta}_m}$	4.486	3.036	2.207	1.296	0.698
	$ISB_{\widehat{\theta}_m}$	1.188	0.635	0.377	0.334	0.433
100	$MI_{\widehat{\theta}_m}$	20.485	12.872	7.737	4.220	2.388
	$IV_{\widehat{\theta}_m}$	16.644	10.961	6.520	3.325	1.663
	$ISB_{\widehat{\theta}_m}$	3.841	1.911	1.218	0.896	0.725

		Frank copula ($\theta = 3.0$) - $\widehat{\theta}^*_m(t_2, t_1)$				
$m/\widehat{b}_{n2} = cb_{n1}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\theta}^*_m}$	1.341	1.071	0.849	**0.644**	0.6579
	$IV_{\widehat{\theta}^*_m}$	1.167	0.946	0.732	0.472	0.278
	$ISB_{\widehat{\theta}^*_m}$	0.174	0.125	0.116	0.172	0.380
50	$MI_{\widehat{\theta}^*_m}$	5.471	4.062	2.889	1.892	1.348
	$IV_{\widehat{\theta}^*_m}$	4.190	3.293	2.385	1.433	0.820
	$ISB_{\widehat{\theta}^*_m}$	1.275	0.769	0.504	0.459	0.529
100	$MI_{\widehat{\theta}^*_m}$	17.807	11.949	7.968	4.638	2.738
	$IV_{\widehat{\theta}^*_m}$	14.073	9.938	6.560	3.496	1.843
	$ISB_{\widehat{\theta}^*_m}$	3.734	2.012	1.409	1.142	0.896

		Frank copula ($\theta = 3.0$) - $\widehat{\vartheta}_m(t_1, t_2)$				
$m/\widehat{b}_{n1}/\widehat{b}_{n2}$		1	2.7	7.4	20.1	54.6
25	$MI_{\widehat{\vartheta}^*_m}$	0.804	0.620	0.487	**0.406**	0.508
	$IV_{\widehat{\vartheta}^*_m}$	0.624	0.511	0.396	0.255	0.142
	$ISB_{\widehat{\vartheta}^*_m}$	0.179	0.109	0.091	0.151	0.366
50	$MI_{\widehat{\vartheta}^*_m}$	3.550	2.448	1.741	1.190	0.926
	$IV_{\widehat{\vartheta}^*_m}$	2.334	1.757	1.310	0.801	0.450
	$ISB_{\widehat{\vartheta}^*_m}$	1.215	0.690	0.430	0.389	0.476
100	$MI_{\widehat{\vartheta}^*_m}$	11.921	7.556	4.875	2.923	1.824
	$IV_{\widehat{\vartheta}^*_m}$	8.187	5.633	3.588	1.916	1.021
	$ISB_{\widehat{\vartheta}^*_m}$	3.734	1.923	1.288	1.007	0.803

5 Food expenditure and net income dataset

In the main text, we illustrate the use of our novel nonparametric estimator for the cross ratio function by analyzing data on the relationship between food expenditure and net income (Family Expenditure Survey, 1968–1983, Härdle, 1990). A random subsample of size $n = 500$ is selected from the aforementioned dataset (see Figure 5 in the main text) to be analysed and these data are available on https://ibiostat.be/online-resources.

The summary statistics for the selected subsample are given in Table 7.

Table 7 Summary statistics for the food expenditure and net income data application: Q1 and Q3 represent the estimated first and third quartile, respectively, and SD is the estimated standard deviation.

	Food expenditure	
Measure	T_1	T_2
Min	0.061	0.012
Q1	0.545	0.607
Median	0.885	0.919
Mean	0.978	0.994
Q3	1.236	1.293
Max	4.625	3.498
SD	0.585	0.544

6 Asthma data application

In this section, we illustrate the use of our novel nonparametric estimator for the cross ratio function by analyzing data on recurrent asthma attacks in children (Duchateau et al., 2003 and Section 8 of the main text). Children having high risk of asthma (but with no attacks so far) enter the prevention trial at the age of six months. The follow up period is 18 months. After randomisation a placebo is administered to the control group ($n_1 = 119$) and an anti-allergic drug to the treated group ($n_2 = 113$).
Two subsequent (first and second) asthma events per patient are considered in this analysis; the times at risk (expressed in days) before experiencing event 1 and 2 (i.e., time to an attack) are denoted as T_1 and T_2, respectively. The summary statistics are given in Table 8.

In this study interest is in the cross ratio function $\widehat{\vartheta}_m(t_1, t_2)$. Given that the estimated scale ratio is 0.71 (control) an 0.82 (treatment), we use the same bandwidth for smoothing in the t_1 and the t_2 direction. Given the rather smooth shape of the cross ratio function our estimator, based on these moderate sample sizes, will provide useful information.

Heatplots of the estimated cross ratio surfaces $\widehat{\vartheta}_m^{(1)}(t_1, t_2)$ and $\widehat{\vartheta}_m^{(2)}(t_1, t_2)$ for members of the placebo and drug group, respectively, are plotted in the upper panel of Figure 2, which also graphically shows the difference $\widehat{\delta}_m(t_1, t_2) =$

Table 8 Summary statistics for the asthma data application: Q1 and Q3 represent the estimated first and third quartile, respectively, and SD is the estimated standard deviation.

	Asthma data			
	Control		Treatment	
Measure	T_1	T_2	T_1	T_2
Min	0.508	0.825	1.370	1.013
Q1	3.076	1.243	4.907	2.192
Median	73.642	38.936	72.384	43.343
Mean	110.639	65.556	106.172	78.922
Q3	393.781	247.217	381.783	307.485
Max	511.646	292.654	549.187	397.359
SD	103.446	73.069	104.857	86.063

$\widehat{\vartheta}_m^{(2)}(t_1,t_2) - \widehat{\vartheta}_m^{(1)}(t_1,t_2)$ between both estimates (right upper panel). To visualise the variability of the estimates we include bootstrap based pointwise 95% bootstrap-percentile confidence intervals for the cross ratio function at three different values of T_1 (lower panels). The cross ratio describes the degree of dependence at point (t_1,t_2) (Hougaard 2000, Section 4.6). Figure 2 shows that the degree of dependence is stronger in the control group when compared to the treated group. Children in the control group having a first asthma attack at time $T_1 = t_1$ have, for a broad range of t_2 values, higher instantaneous risk to develop a second asthma attack compared to children (in the control group) having $T_1 > t_1$ (i.e., children who did not yet develop a first asthma at time t_1). In the treated group, there is no pronounced degree of dependence which means that the anti-allergic drug disconnects the dependence between the time to first attack and the time to second attack. Our findings are in line with the results obtained in Çiğşar and Lawless (2012), and Meyer and Romeo (2015).

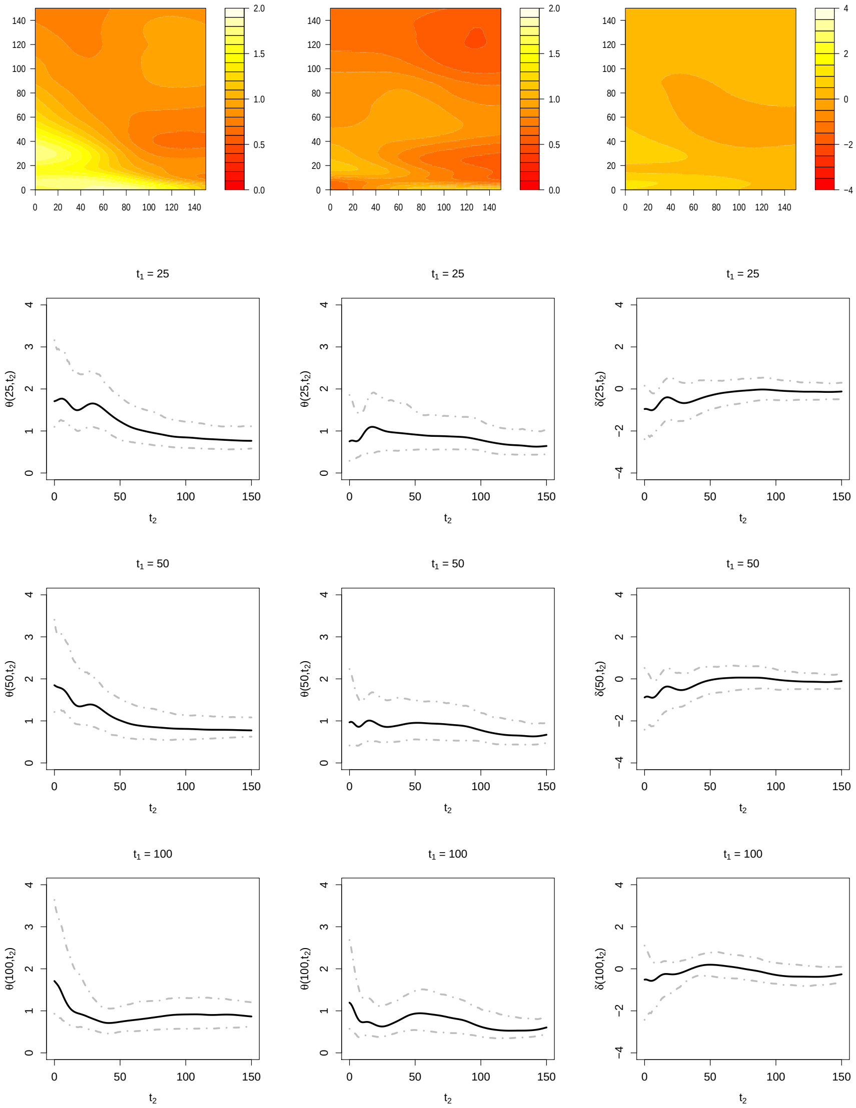

Fig. 2 Asthma data application: heatplots of the estimated cross ratio surfaces $\widehat{\vartheta}_m^{(1)}(t_1,t_2)$ and $\widehat{\vartheta}_m^{(2)}(t_1,t_2)$ and cross ratio curves for $t_1 = 25$, 50 and 100 (black solid lines) with pointwise 95% bootstrap-percentile confidence intervals (gray dash-dotted lines) for members of the placebo (left panels) and drug group (middle panels), respectively, and the difference between both estimates (right panels).

References

Çiğşar, C., Lawless, J. F. (2012). Assessing transient carryover effects in recurrent event processes, with application to chronic health conditions. *The Annals of Applied Statistics*, 6, 1641–1663.

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.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32, 177–203.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

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.

Meyer, R., Romeo, J. S. (2015). Bayesian semiparametric analysis of recurrent failure time data using copulas. *Biometrical Journal*, 57, 982–1001.

Sen, B., Xu, G. (2015). Model based bootstrap methods for interval censored data. *Computational Statistics & Data Analysis*, 81, 121–129.