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

Steven Abrams · Paul Janssen · Jan Swanepoel · Noël Veraverbeke
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

\[ (nm^{-1/2}b_n)^{1/2} \{ \hat{\lambda}_m(t_1 \mid T_2 = t_2) - \lambda_m(t_1 \mid T_2 = t_2) \} \]

\[ = \beta(t_1, t_2) + (nm^{-1/2}b_n)^{1/2} \left\{ \frac{\hat{f}_{t_2}(t_1) - E[\hat{f}_{t_2}(t_1)]}{1 - F_{t_2}(t_1)} \right\} + o_p(1). \]  

(S1)

In a similar way we have

\[ (nm^{-1/2}b_n)^{1/2} \{ \hat{\lambda}_m(t_2 \mid T_1 = t_1) - \lambda_m(t_2 \mid T_1 = t_1) \} \]

\[ = \beta^*(t_2, t_1) + (nm^{-1/2}b_n)^{1/2} \left\{ \frac{\hat{f}_{t_1}(t_2) - E[\hat{f}_{t_1}(t_2)]}{1 - F_{t_1}(t_2)} \right\} + o_p(1). \]  

(S2)

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

\[ \frac{1}{2} \left[ \beta(t_1, t_2) + \beta^*(t_2, t_1) \right] \]

\[ \lambda(t_1 \mid T_2 > t_2) + \frac{\hat{f}_{t_1}(t_2) - E[\hat{f}_{t_1}(t_2)]}{1 - F_{t_1}(t_2)} \]

and that the asymptotic variance will be that of

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

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

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

Similarly,

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

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

\[ (III) \sim \frac{1}{nb_n^2} \frac{1}{\lambda(t_1 \mid T_2 > t_2) \lambda(t_2 \mid T_1 > t_1) (F_{t_2}(t_1)[1 - F_{t_2}(t_1)] \lambda(t_1 \mid T_2 > t_2) \lambda(t_2 \mid T_1 > t_1) (1 - F_{t_2}(t_1))(1 - F_{t_2}(t_1))} \]

\[ \times \int_{-L}^{L} \int_{-L}^{L} E[T_{in}[S_1(t_1 - b_n u_1), S_2(t_2)][T_{in}^*[S_1(t_1), S_2(t_2 - b_n u_2)]] dK_0(u_1) dK_0(u_2), \]
with $T_{in}(u, v) = Z_{in}(u, v) - \tilde{Z}_{in}(u, v)$, $T^{*}_{in}(u, v) = Z^{*}_{in}(u, v) - \tilde{Z}^{*}_{in}(u, v)$, where
\[
Z_{in}(u, v) = m \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \left[ I \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) - P \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) \right] \left[ \frac{m}{m} \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \left[ I \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) - P \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) \right] \right] P_{m,k}(u) P_{m-1,l}(v),
\]
\[
\tilde{Z}_{in}(u, v) = mC^{(2)}(u, v) \sum_{k=0}^{m-1} \left[ I \left( \frac{k}{m} < \tilde{V}_i \leq \frac{k+1}{m} \right) - \frac{1}{m} \right] P_{m-1,l}(v),
\]
\[
Z^{*}_{in}(u, v) = m \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \left[ I \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) - P \left( \frac{k}{m} \leq \frac{l}{m} \leq \frac{l+1}{m} \right) \right] P_{m-1,l}(v),
\]
\[
\tilde{Z}^{*}_{in}(u, v) = mC^{(1)}(u, v) \sum_{k=0}^{m-1} \left[ I \left( \frac{k}{m} < \tilde{V}_i \leq \frac{k+1}{m} \right) - \frac{1}{m} \right] P_{m-1,l}(v).
\]

The random vectors $(\tilde{U}_1, \tilde{V}_1), \ldots, (\tilde{U}_n, \tilde{V}_n)$, with $\tilde{U}_i = S_1(T_{1i})$, $\tilde{V}_i = S_2(T_{2i})$, 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
\[
E \left[ Z_{in}(u, v)Z^{*}_{in}(u', v') \right] - E \left[ Z_{in}(u, v)\tilde{Z}^{*}_{in}(u', v') \right]
- E \left[ \tilde{Z}_{in}(u, v)Z^{*}_{in}(u', v') \right] + E \left[ \tilde{Z}_{in}(u, v)\tilde{Z}^{*}_{in}(u', v') \right].
\]
Now $E[Z_{in}(u,v)Z_{in}^*(u',v')]$ is

\[
E \left\{ m^2 \sum_{k=0}^{m} \sum_{l=0}^{m-1} \sum_{k'=0}^{m} \sum_{l'=0}^{m-1} \left[ I \left( \frac{k}{m} \leq \frac{l}{m} < \frac{l+1}{m} \right) \right. \\
- P \left( \frac{k'}{m} < \tilde{V}_i \leq \frac{k'+1}{m}, \tilde{U}_i \leq \frac{l'}{m} \right) \right. \\
- \left. \left. P \left( \frac{k'}{m} \leq \tilde{U}_i < \frac{k'+1}{m}, \tilde{V}_i \leq \frac{l'}{m} \right) \right] P_{m,k}(u)P_{m-1,l}(v)P_{m-1,k'}(u')P_{m,l'}(v') \right. \\
= m^2 \sum_{k=0}^{m} \sum_{l=0}^{m-1} \sum_{k'=0}^{m} \sum_{l'=0}^{m-1} P \left( \frac{k'}{m} < \tilde{U}_i \leq \frac{k' + (k'+1)}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{(l+1) \land l'}{m} \right) \\
\times P_{m,k}(u)P_{m-1,l}(v)P_{m-1,k'}(u')P_{m,l'}(v') \\
- m^2 \left[ \sum_{k=0}^{m} \sum_{l=0}^{m-1} P \left( \frac{k'}{m} < \tilde{U}_i \leq \frac{k' + 1}{m}, \tilde{V}_i \leq \frac{l+1}{m} \right) P_{m,k}(u)P_{m-1,l}(v) \right] \\
\times \left[ \sum_{k'=0}^{m-1} \sum_{l'=0}^{m-1} P \left( \frac{k'}{m} < \tilde{U}_i \leq \frac{k' + 1}{m}, \tilde{V}_i \leq \frac{l'}{m} \right) P_{m-1,k'}(u')P_{m,l'}(v') \right]. \tag{S4}
\]

The first term in (S4) can be written as

\[
m^2 \sum_{k'=0}^{m-1} \sum_{l=0}^{m-1} P \left( \frac{k'}{m} < \tilde{U}_i \leq \frac{k' + 1}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{l+1}{m} \right) \\
\times P_{m-1,k'}(u')P_{m-1,l}(v) \sum_{k=0}^{m} P_{m,k}(u) \sum_{l'=l+1}^{m} P_{m,l'}(v').
\]

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

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

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

\[
E [Z_{in}(u,v)Z_{in}^*(u',v')] = O(1) = o(m^{1/2}).
\]
Furthermore, we have
\[
\mathbb{E} \left[ Z_{im}(u,v) \tilde{Z}_{im}'(u',v') \right] \\
= \mathbb{E} \left\{ m^2 C^{(1)}(u',v') \sum_{k=0}^{m} \sum_{l=0}^{m-1} \sum_{k'=0}^{l} \left[ I \left( \tilde{U}_i \leq \frac{k}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{l+1}{m} \right) \\
- P \left( \tilde{U}_i \leq \frac{k}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{l+1}{m} \right) \left[ l \leq \frac{l+1}{m} \right] \\
\times \frac{1}{m} Z_{im}(u) \right\} \\
= m^2 C^{(1)}(u',v') \sum_{k=0}^{m} \sum_{l=0}^{m-1} P \left( \frac{k'}{m} < \tilde{U}_i \leq \frac{k'+1}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{l+1}{m} \right) \\
\times P_{m,k}(u) P_{m-1,l}(v) \sum_{k=k'+1}^{m} P_{m,k}(u) \\
- m C^{(1)}(u',v') \sum_{k=0}^{m} \sum_{l=0}^{m-1} P \left( \frac{k}{m} < \tilde{U}_i \leq \frac{k+1}{m}, \frac{l}{m} < \tilde{V}_i \leq \frac{l+1}{m} \right) P_{m,k}(u) P_{m-1,l}(v) \\
\leq C^{(1)}(u',v') \left[ c(u',v) + O(m^{-1}) \right] - C^{(1)}(u',v') \left[ C^{(2)}(u,v) + O(m^{-1}) \right] \\
= O(1) = o(m^{1/2}).
\]

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

2 MISE and bias-variance relationship

The mean integrated squared error $MISE_{\hat{\theta}_m}$ for an estimator $\hat{\theta}_m(t_1,t_2)$ of the cross ratio function $\theta(t_1,t_2)$ is defined as
\[
MISE_{\hat{\theta}_m} \equiv \mathbb{E}_{\hat{\theta}_m} \left\{ \int_{a_1}^{b_1} \int_{a_2}^{b_2} \left[ \hat{\theta}_m(t_1,t_2) - \theta(t_1,t_2) \right]^2 dt_1 dt_2 \right\}.
\]

The bias-variance relationship decomposes $MISE_{\hat{\theta}_m}$ in a bias and variance term as follows:
\[
MISE_{\hat{\theta}_m} = E \left( \int_{a_1}^{b_1} \int_{a_2}^{b_2} \left( \hat{\theta}_m(t_1,t_2) - E[\hat{\theta}_m(t_1,t_2)] \right)^2 dt_1 dt_2 \right) \\
+ \int_{a_1}^{b_1} \int_{a_2}^{b_2} \left[ E[\hat{\theta}_m(t_1,t_2)] - \theta(t_1,t_2) \right]^2 dt_1 dt_2 \\
\equiv IVAR_{\hat{\theta}_m} + ISBIAS_{\hat{\theta}_m},
\]
where $IVAR_\theta$ represents the integrated variance and $ISBIAS_\theta$ is the integrated squared bias. In order to estimate the mean integrated squared error, we use the following approximation based on $M$ simulation replications:

$$MI_{\hat{\theta}_m} = \frac{1}{M} \sum_{r=1}^{M} \left\{ \Delta_1 \Delta_2 \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \left[ \hat{\theta}^{(r)}_m(t_1[k], t_2[l]) - \hat{\theta}(t_1[k], t_2[l]) \right]^2 \right\},$$

where $\hat{\theta}^{(r)}_m$ 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 $IV_{\hat{\theta}_m}$ and $ISB_{\hat{\theta}_m}$, respectively, as follows:

$$IV_{\hat{\theta}_m} = \frac{1}{M} \sum_{r=1}^{M} \left\{ \Delta_1 \Delta_2 \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \left[ \hat{\theta}^{(r)}_m(t_1[k], t_2[l]) - \hat{\theta}^{(r)}(t_1[k], t_2[l]) \right]^2 \right\},$$

$$ISB_{\hat{\theta}_m} = \Delta_1 \Delta_2 \sum_{k=1}^{N_1} \sum_{l=1}^{N_2} \left[ \hat{\theta}^{(r)}_m(t_1[k], t_2[l]) - \theta(t_1[k], t_2[l]) \right]^2,$$

where $\hat{\theta}^{(r)}(t_1[k], t_2[l])$ is equal to

$$\hat{\theta}^{(r)}(t_1[k], t_2[l]) = \frac{1}{M} \sum_{r=1}^{M} \hat{\theta}^{(r)}_m(t_1[k], t_2[l]).$$

3 Bandwidth selection

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

$$MISE_{\hat{\theta}_m}(m, b_n) = E_b \left\{ \int \int \left[ \hat{\theta}^{(b)}_m(t_1, t_2; b_n) - \hat{\theta}_m(t_1, t_2; b_n) \right]^2 dt_1 dt_2 \right\},$$

where $\hat{\theta}^{(b)}_m(t_1, t_2; b_n)$ 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 $MISE_{\hat{\theta}_m}$ can be estimated by resampling the original data nonparametrically $B$ times, having

$$\tilde{MISE}_{\hat{\theta}_m} = \frac{1}{B} \sum_{b=1}^{B} \int \int \left[ \hat{\theta}^{(b)}_m(t_1, t_2; b_n) - \hat{\theta}_m(t_1, t_2; b_n) \right]^2 dt_1 dt_2.$$

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
Supplementary Materials – Nonparametric estimation of the cross ratio function

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 $\text{MISE}^{\hat{\theta}_m}$ as follows:

$$\text{MISE}^{\hat{\theta}_m} = \frac{1}{B} \sum_{b=1}^{B} \int \int [\hat{\theta}_m^{(b)}(t_1, t_2; b_n) - \hat{\theta}_m(t_1, t_2; b_n)]^2 dt_1 dt_2,$$

depending on an initial choice $m_0$ and $b_{n_0}$ for the bandwidths. Note that the bandwidth selection procedure for the estimator $\hat{\theta}^*_m(t_2, t_1)$ is performed based on the proposed criterion and replacing the estimates $\hat{\theta}_m(t_1, t_2)$ 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 $\text{MISE}^{\hat{\theta}_m}$. 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 $\hat{\theta}_m(t_1, t_2)$ (black lines) and $\hat{\theta}^*_m(t_2, t_1)$ (red lines). The selected optimal values are those minimizing the criterion and are equal to $b_n = 0.2$ and $1.5$ for $\hat{\theta}_m(t_1, t_2)$ and $\hat{\theta}^*_m(t_2, t_1)$, respectively, and $m = 25$ in both cases. We use these selected bandwidths to obtain $\hat{\theta}_m(t_1, t_2)$ 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 $\hat{\theta}_m(t_1, t_2)$ and $\hat{\theta}^*_m(t_2, t_1)$. Recall from Remark 4 in the main text that the scale ratio of $T_1$ and $T_2$ is 0.6. Simulations with bandwidth $b_{n_1} = b_n$ and $b_{n_2} = cb_n$ with $c$ chosen as the estimated scale ratio of $T_1$ and $T_2$, reveal that the optimal (in terms of $M1$) choice for $(m, b_n)$ remains the same (see Tables 4–6 for details).
Fig. 1 $\hat{MISE}_{\hat{\theta}_m}$ (black lines) and $\hat{MISE}_{\hat{\theta}_m^*}$ (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 estimates $\hat{\vartheta}_m$, $\hat{\vartheta'}_m$ and $\hat{\vartheta''}_m$, and different choices of $b_m$, $m$ and $n$ and Clayton copula function with parameter $\theta = 0.5$. Minimum MI-values are highlighted in bold.

<table>
<thead>
<tr>
<th>$m/\hat{b}_m$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta}_m(t_1,t_2)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>4.207</td>
<td>2.640</td>
</tr>
<tr>
<td>50</td>
<td>$\text{M} I$</td>
<td>45.160</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>42.691</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>2.469</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>100</td>
<td>$\text{M} I$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/\hat{b}_m$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta'}_m(t_2,t_1)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>3.312</td>
<td>1.928</td>
</tr>
<tr>
<td>50</td>
<td>$\text{M} I$</td>
<td>34.282</td>
<td>8.477</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>30.729</td>
<td>5.328</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>3.553</td>
<td>3.149</td>
</tr>
<tr>
<td>100</td>
<td>$\text{M} I$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/\hat{b}_m$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta''}_m(t_1,t_2)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>2.159</td>
<td>1.362</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>3.419</td>
<td>3.628</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>19.141</td>
<td>6.574</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>3.294</td>
<td>3.250</td>
</tr>
<tr>
<td>100</td>
<td>$\text{M} I$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{IV}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>$\text{ISB}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>
Table 2 M1, IV and ISB for the different estimators $\hat{\theta}_m, \hat{\theta}_m, \hat{\theta}_m$, and different choices of $b_n$, $m$ and $n$ and Gumbel copula function with parameter $\theta = 1.5$. Minimum M1-values are highlighted in bold.

<table>
<thead>
<tr>
<th>$m/\hat{b}_n$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m(t_1, t_2)$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>12.577 9.161 4.262 2.630</td>
<td>2.230</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>10.852 7.797 2.965 1.120 0.534</td>
<td>1.229 0.979 0.755 0.490 0.286</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>1.725 1.364 1.297 1.510 1.696</td>
<td>0.871 0.950 1.112 1.399 1.696</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 5.004 3.248</td>
<td>5.549 4.247 3.503 2.866 2.469</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 3.229 1.545</td>
<td>4.214 3.133 2.287 1.365 0.765</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>22.896 19.932 4.991 1.775 1.704</td>
<td>1.335 1.115 1.217 1.501 1.704</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 12.454 4.688</td>
<td>&gt; 100 61.982 13.643 6.505 4.212</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 10.260 3.071</td>
<td>&gt; 100 57.950 10.958 3.974 1.994</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 79.348 2.195 1.617</td>
<td>10.704 4.092 2.686 2.532 2.218</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/\hat{b}_n$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m^*(t_1, t_2)$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m^*(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>13.506 5.529 2.997 2.302</td>
<td>1.393 1.885 1.902 1.969 2.097</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>12.117 4.230 1.605 0.750 0.391</td>
<td>1.106 0.853 0.655 0.422 0.249</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>1.389 1.299 1.463 1.632 1.840</td>
<td>0.878 1.032 1.247 1.547 1.848</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 19.097 5.217 1.917 0.901</td>
<td>4.837 3.467 2.085 1.126 0.619</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>9.701 2.216 1.733 1.725 1.822</td>
<td>1.285 1.212 1.421 1.668 1.821</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 22.642 6.989 3.869</td>
<td>20.283 11.741 8.475 5.506 3.786</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 19.225 4.673 2.082</td>
<td>15.811 8.893 5.885 3.045 1.659</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 3.417 2.316 1.787</td>
<td>4.474 2.848 2.590 2.462 2.128</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/\hat{b}_n$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m^*(t_1, t_2)$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m^*(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 100$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>7.291 4.368 2.530 2.030</td>
<td>1.943</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>5.784 3.086 1.166 0.464 0.184</td>
<td>0.626 0.499 0.388 0.252 0.147</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>1.507 1.282 1.364 1.566 1.759</td>
<td>0.871 0.987 1.175 1.468 1.767</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 73.405 3.285 2.439</td>
<td>3.817 3.016 2.575 2.325 2.176</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 70.900 1.551 0.703</td>
<td>2.531 1.871 1.270 0.750 0.421</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>13.828 7.283 2.505 1.734 1.763</td>
<td>1.286 1.145 1.305 1.575 1.756</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$M1_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 6.073 3.086</td>
<td>&gt; 100 20.383 7.236 4.525 3.253</td>
</tr>
<tr>
<td>$IV_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 &gt; 100 3.890 1.435</td>
<td>&gt; 100 17.183 4.645 2.055 1.098</td>
</tr>
<tr>
<td>$ISB_{\hat{\theta}_m}$</td>
<td>&gt; 100 &gt; 100 21.915 2.183 1.651</td>
<td>6.261 3.200 2.592 2.471 2.155</td>
</tr>
</tbody>
</table>
Supplementary Materials – Nonparametric estimation of the cross ratio function

Table 3  
MI, IV and ISB for the different estimators $\hat{\theta}_m$, $\hat{\theta}_m^*$ and $\hat{\theta}_m^*$, 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.

<table>
<thead>
<tr>
<th>$m/b_n$</th>
<th>Frank copula ($\theta = 3.0$) - $\hat{\theta}_m(t_1, t_2)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2.7</td>
<td>7.4</td>
</tr>
<tr>
<td>25</td>
<td>MI$^{\hat{\theta}_m}$</td>
<td>7.038</td>
<td>3.958</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m}$</td>
<td>6.058</td>
<td>3.458</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m}$</td>
<td>0.981</td>
<td>0.500</td>
</tr>
<tr>
<td>50</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>54.063</td>
<td>39.912</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>48.411</td>
<td>36.855</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>5.653</td>
<td>3.058</td>
</tr>
<tr>
<td>100</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/b_n$</th>
<th>Frank copula ($\theta = 3.0$) - $\hat{\theta}_m^*(t_2, t_1)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2.7</td>
<td>7.4</td>
</tr>
<tr>
<td>25</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>5.699</td>
<td>3.517</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>4.765</td>
<td>3.095</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>0.844</td>
<td>0.422</td>
</tr>
<tr>
<td>50</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>50.682</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>47.657</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>9.946</td>
<td>3.025</td>
</tr>
<tr>
<td>100</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>37.901</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/b_n$</th>
<th>Frank copula ($\theta = 3.0$) - $\hat{\theta}_m^*(t_1, t_2)$</th>
<th>n = 100</th>
<th>n = 300</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2.7</td>
<td>7.4</td>
</tr>
<tr>
<td>25</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>3.816</td>
<td>2.266</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>2.925</td>
<td>1.823</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>0.891</td>
<td>0.443</td>
</tr>
<tr>
<td>50</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>61.789</td>
<td>23.597</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>55.058</td>
<td>20.813</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>6.731</td>
<td>2.784</td>
</tr>
<tr>
<td>100</td>
<td>MI$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>IV$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
<tr>
<td></td>
<td>ISB$^{\hat{\theta}_m^*}$</td>
<td>&gt; 100</td>
<td>&gt; 100</td>
</tr>
</tbody>
</table>
Table 4  MI, IV and ISB for the different estimators $\hat{\theta}_m$, $\hat{\theta}^*_m$ and $\hat{\vartheta}_m$, and different choices of $b_{n1}$, $m$ and $n$, $b_{n2} = c b_{n1}$ ($c = 0.6$), and Clayton copula function with parameter $\theta = 0.5$ ($n = 300$). Minimum MI-values are highlighted in bold.

<table>
<thead>
<tr>
<th>$m/b_{n1}$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta}_m(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>3.508</td>
</tr>
<tr>
<td>50</td>
<td>7.371</td>
</tr>
<tr>
<td></td>
<td>4.475</td>
</tr>
<tr>
<td></td>
<td>2.896</td>
</tr>
<tr>
<td>100</td>
<td>23.345</td>
</tr>
<tr>
<td></td>
<td>17.576</td>
</tr>
<tr>
<td></td>
<td>5.768</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/b_{n2} = cb_{n1}$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta}^*_m(t_2, t_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>4.547</td>
</tr>
<tr>
<td></td>
<td>1.082</td>
</tr>
<tr>
<td></td>
<td>3.496</td>
</tr>
<tr>
<td>50</td>
<td>7.102</td>
</tr>
<tr>
<td></td>
<td>4.431</td>
</tr>
<tr>
<td></td>
<td>2.731</td>
</tr>
<tr>
<td>100</td>
<td>30.864</td>
</tr>
<tr>
<td></td>
<td>5.717</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m/b_{n1}/b_{n2}$</th>
<th>Clayton copula ($\theta = 0.5$) - $\hat{\vartheta}^*_m(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.640</td>
</tr>
<tr>
<td></td>
<td>3.499</td>
</tr>
<tr>
<td>50</td>
<td>5.530</td>
</tr>
<tr>
<td></td>
<td>2.546</td>
</tr>
<tr>
<td></td>
<td>2.784</td>
</tr>
<tr>
<td>100</td>
<td>17.359</td>
</tr>
<tr>
<td></td>
<td>11.716</td>
</tr>
<tr>
<td></td>
<td>5.644</td>
</tr>
</tbody>
</table>
Table 5 MI, IV and ISB for the different estimators $\hat{\theta}_m$, $\hat{\theta}_m^*$ and $\hat{\vartheta}_m$, 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.

<table>
<thead>
<tr>
<th>$m/b_{n1}$</th>
<th>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m(t_1, t_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_{n2} = cb_{n1}$</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>50</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>100</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>$m/b_{n2} = cb_{n1}$</td>
<td>Gumbel copula ($\theta = 1.5$) - $\hat{\theta}_m^*(t_2, t_1)$</td>
</tr>
<tr>
<td>25</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>50</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>100</td>
<td>$MI_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\theta}_m}$</td>
</tr>
<tr>
<td>$m/b_{n2}$</td>
<td>Gumbel copula ($\theta = 1.5$) - $\hat{\vartheta}_m(t_1, t_2)$</td>
</tr>
<tr>
<td>25</td>
<td>$MI_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td>50</td>
<td>$MI_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td>100</td>
<td>$MI_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}$</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}$</td>
</tr>
</tbody>
</table>
Table 6 MI, IV and ISB for the different estimators $\hat{\theta}_m$, $\hat{\theta}_m^*$ and $\hat{\vartheta}_m$, 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.

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

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

<table>
<thead>
<tr>
<th>$m/b_{n1}/b_{n2}$</th>
<th>Frank copula ($\theta = 3.0$) - $\hat{\vartheta}_m(t_1, t_2)$</th>
<th>1</th>
<th>2.7</th>
<th>7.4</th>
<th>20.1</th>
<th>54.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$MI_{\hat{\vartheta}_m}^*$</td>
<td>0.804</td>
<td>0.620</td>
<td>0.487</td>
<td><strong>0.306</strong></td>
<td>0.508</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}^*$</td>
<td>0.624</td>
<td>0.511</td>
<td>0.396</td>
<td>0.255</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}^*$</td>
<td>0.179</td>
<td>0.109</td>
<td>0.091</td>
<td>0.151</td>
<td>0.366</td>
</tr>
<tr>
<td>50</td>
<td>$MI_{\hat{\vartheta}_m}^*$</td>
<td>3.550</td>
<td>2.448</td>
<td>1.741</td>
<td>1.190</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}^*$</td>
<td>2.334</td>
<td>1.757</td>
<td>1.310</td>
<td>0.801</td>
<td>0.450</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}^*$</td>
<td>1.215</td>
<td>0.690</td>
<td>0.430</td>
<td>0.389</td>
<td>0.476</td>
</tr>
<tr>
<td>100</td>
<td>$MI_{\hat{\vartheta}_m}^*$</td>
<td>11.921</td>
<td>7.556</td>
<td>4.875</td>
<td>2.923</td>
<td>1.824</td>
</tr>
<tr>
<td></td>
<td>$IV_{\hat{\vartheta}_m}^*$</td>
<td>8.187</td>
<td>5.633</td>
<td>3.588</td>
<td>1.916</td>
<td>1.021</td>
</tr>
<tr>
<td></td>
<td>$ISB_{\hat{\vartheta}_m}^*$</td>
<td>3.734</td>
<td>1.923</td>
<td>1.288</td>
<td>1.007</td>
<td>0.803</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Food expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( T_1 )</td>
</tr>
<tr>
<td>Min</td>
<td>0.061</td>
</tr>
<tr>
<td>Q1</td>
<td>0.545</td>
</tr>
<tr>
<td>Median</td>
<td>0.885</td>
</tr>
<tr>
<td>Mean</td>
<td>0.978</td>
</tr>
<tr>
<td>Q3</td>
<td>1.236</td>
</tr>
<tr>
<td>Max</td>
<td>4.625</td>
</tr>
<tr>
<td>SD</td>
<td>0.585</td>
</tr>
</tbody>
</table>

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 \( \hat{\vartheta}_m(t_1, t_2) \). 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 \( \hat{\vartheta}_m^{(1)}(t_1, t_2) \) and \( \hat{\vartheta}_m^{(2)}(t_1, t_2) \) for members of the placebo and drug group, respectively, are plotted in the upper panel of Figure 2, which also graphically shows the difference \( \hat{\delta}_m(t_1, t_2) = \)
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.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.508</td>
<td>0.825</td>
</tr>
<tr>
<td>Q1</td>
<td>3.076</td>
<td>1.243</td>
</tr>
<tr>
<td>Median</td>
<td>73.642</td>
<td>38.936</td>
</tr>
<tr>
<td>Mean</td>
<td>110.639</td>
<td>65.556</td>
</tr>
<tr>
<td>Q3</td>
<td>393.781</td>
<td>247.217</td>
</tr>
<tr>
<td>Max</td>
<td>511.646</td>
<td>292.654</td>
</tr>
<tr>
<td>SD</td>
<td>103.446</td>
<td>73.069</td>
</tr>
</tbody>
</table>

\[ \hat{\vartheta}_m^{(2)}(t_1, t_2) - \hat{\vartheta}_m^{(1)}(t_1, t_2) \] 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 \( \hat{\vartheta}^{(1)}_{m}(t_1,t_2) \) and \( \hat{\vartheta}^{(2)}_{m}(t_1,t_2) \) 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


