CentER

Discussion Paper

No. 2008–40

IMPROVING UPON THE MARGINAL EMPIRICAL DISTRIBUTION FUNCTIONS WHEN THE COPULA IS KNOWN

By Johan Segers, Ramon van den Akker, Bas J.M. Werker

April 2008

ISSN 0924-7815

Tilburg University

Improving upon the Marginal Empirical Distribution Functions When the Copula is Known

Johan Segers*, Ramon van den Akker†, and Bas J.M. Werker‡

Université catholique de Louvain and Tilburg University

Abstract

At the heart of the copula methodology in statistics is the idea of separating marginal distributions from the dependence structure. However, as shown in this paper, this separation is not to be taken for granted: in the model where the copula is known and the marginal distributions are completely unknown, the empirical distribution functions are semiparametrically efficient if and only if the copula is the independence copula. Incorporating the knowledge of the copula into a nonparametric likelihood yields an estimation procedure which by simulations is shown to outperform the empirical distribution functions, the amount of improvement depending on the copula. Although the known-copula model is arguably artificial, it provides an instructive stepping stone to the more general model of a parametrically specified copula and arbitrary margins.

Keywords: independence copula, nonparametric maximum likelihood estimator, score function, semiparametric efficiency, tangent space

JEL: C14

1 Introduction

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from a bivariate distribution function H with absolutely continuous marginal distribution functions F and G and copula C. Recall that C is the unique bivariate distribution function with uniform margins on $(0, 1)$ such that $H(x, y) = C(F(x), G(y))$ for all $x, y \in \mathbb{R}$ (Sklar, 1959; Nelsen, 1999).

The problem we consider is efficient estimation of the marginal distribution functions F and G given that the copula C is known. Admittedly, this assumption is artificial since it can never be fulfilled in practice. However, the solution of this problem is thought to be an important step towards the solution of the more realistic problem of estimating the bivariate distribution H given that its copula C belongs to a parametric family and without further knowledge of the margins except for (absolute) continuity. Indeed,

*Université catholique de Louvain, Institut de statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be

†Econometrics group, CentER, Tilburg University, PO Box 90153, NL-5000 LE Tilburg, the Netherlands. E-mail: R.vdnAkker@TilburgUniversity.nl

‡Econometrics & Finance group, Tilburg University, PO Box 90153, NL-5000 LE Tilburg, the Netherlands. E-mail: B.J.M.Werker@TilburgUniversity.nl

in Klaassen and Wellner (1997), which is the first paper to consider efficient estimation of the marginals in a copula model, the following is written:

> "*It would be very interesting to know information bounds and efficient estimators for estimation of the marginal distribution functions F and G in the bivariate normal copula model treated here, or in other copula models.*" — Klaassen and Wellner (1997, p. 72)

For instance, if $(C_\theta \mid \theta \in \Theta)$ is a well-behaved family of copulas and if an efficient estimator of θ is available, then the so-called plug-in principle produces an efficient estimator of (F, G) (Klaassen and Putter, 2005). Furthermore, the contribution of this paper can be considered as a complement to Bickel et al. (1991) and Peng and Schick (2002, 2004, 2005), in which efficient estimation of some aspect of the bivariate distribution function is considered if complete information on one or both marginals is available.

Since the whole philosophy behind the use of copulas is to disentangle the analysis of the marginal distributions from the one of the dependence structure, one may be tempted to believe that knowledge of the copula cannot in any way contribute to knowledge of the marginals. That is, even if the copula would be completely known, one cannot improve upon the marginal empirical distribution functions. The following example shows this conception to be false in a dramatic way.

EXAMPLE 1.1. Let the copula C be the distribution function corresponding to the uniform distribution on the union of the squares $(0, 1/2) \times (0, 1/2)$ and $(1/2, 1) \times (1/2, 1)$. A scatterplot of a sample from a continuous distribution with this copula consists of the union of two clouds of points, one in the bottom left of the plot and the other one in the top right. The bottom-left cloud corresponds to points to the south-west of the pair of medians and the other cloud to points to the north-east of the pair of medians. With probability tending to one, it is possible to correctly partition the data into these two clouds. For each coordinate, the median of the corresponding marginal distribution must lie between the maximal value of that coordinate in the bottom left cloud and the minimal value of that coordinate in the upper right cloud. Hence, if the marginal density functions are positive and continuous in the medians, it is possible to estimate those medians at the rate $O_p(1/n)$, in contrast to the usual nonparametric rate $O_p(1/\sqrt{n})$.

A related problem concerns the lack of efficiency of the omnibus procedure for the estimation of the copula parameter θ. Recall that this procedure consists of maximizing the copula likelihood after the unknown marginal distribution functions have been estimated by their empirical counterparts (Oakes, 1994; Genest et al., 1995; Shih and Louis, 1995; Tsukahara, 2005). Although in Klaassen and Wellner (1997) and Genest and Werker (2002), the omnibus estimator was proved to be efficient for the normal copula family, this efficiency property was shown in Genest and Werker (2002) to be "an exception rather than the norm". For fully parametric models, similar conclusions about the loss in efficiency arising from separating marginal from dependence aspects were formulated in Joe (2005).

Recall that if no model information is available whatsoever, the empirical distribution of the data can be seen as the maximizer of a nonparametric likelihood, this likelihood being simply the joint probability of observing what has been observed. Therefore, we propose to estimate the marginal distributions by maximizing a nonparametric likelihood as well, the likelihood being constructed to take into account the knowledge that the copula is C. This leads to the definition of the nonparametric maximum likelihood estimator as the solution of a nonlinear, high-dimensional optimization problem with equality and inequality constraints (Section 2). To reduce the complexity of the estimator, we propose a number of simplifications which eventually lead to a system of linear equations, the solution of which can be seen as an approximation of the nonparametric maximum likelihood estimator (Section 3). This linearized estimator is represented as an update of the marginal empirical distribution functions, the update itself being the solution to a system of estimating equations in functional form.

A comparison of the finite-sample performances of the nonparametric maximum likelihood estimator and the marginal empirical distributions in Section 4 shows that the update of the empirical distributions constitutes a modest though nonnegligeable improvement. The amount of improvement depends on the copula. For example, for the Frank copula we document efficiency increases of up to 40% of our nonparametric maximum likelihood estimator over marginal empirical distribution functions. These findings are confirmed in Section 5 where the following is shown: under standard smoothness assumptions, the marginal empirical distribution functions are semiparametrically efficient in the known-copula model if and only if the copula is the independence copula. Here, the meaning of semiparametric efficiency is intended as in Bickel et al. (1998).

It should be noted that in Chen et al. (2006), a sieve maximum likelihood estimation procedure is proposed for real-valued functionals of the full parameter (θ, F, G). By general properties of sieve estimators (Shen, 1997), this procedure is shown to be semiparametrically efficient. However, implementing the procedure is difficult as the final output depends on the choice of the sieve, which ideally should be constructed in function of the shape of the unknown distribution. Furthermore, it is unclear what primitive restrictions the assumptions in Chen et al. (2006) impose upon the copula.

This paper is based on Chapter 5 of Van den Akker (2007), in which asymptotic normality and semiparametric efficiency is proved for an estimator which can be thought of as a large-sample version of the one considered here; see Remark 3.2 below. For our estimator, the issue of its asymptotic behavior remains open.

2 Nonparametric maximum likelihood estimator

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate distribution with continuous marginal distributions and known copula C. Assume that C is absolutely continuous with positive and twice continuously differentiable density c. Denote the log copula density by $\ell = \log c$ and its first and second-order partial derivatives by $\dot{\ell}_i$ and $\ddot{\ell}_{ij}$ for $i, j \in \{1, 2\}$. Throughout $X_{i:n}$ ($Y_{i:n}$) denotes the i-th order statistic among $X_1, \ldots, X_n$ ($Y_1, \ldots, Y_n$).

Since the (unknown) joint distribution H is known to have copula C, it seems natural to impose that the estimated joint distribution has copula C as well. This leads to the idea to consider estimators (F_n^*, G_n^*) that maximize the nonparametric or empirical loglikelihood

$$\mathcal{E}_n(F, G) = \sum_{i=1}^{n} \log \mathbb{P}_{F,G}\{(X_i, Y_i)\}, \tag{2.1}$$

over $F, G \in \mathcal{F}$, the set of all distribution functions on $\mathbb{R}$; here $\mathbb{P}_{F,G} = \mathbb{P}_{F,G}^C$ denotes the probability distribution on $\mathbb{R}^2$ with marginal distribution functions F and G and copula C. For the moment, ignore the question of existence of a maximum. The next proposition shows that any maximizer concentrates on the data.

Proposition 2.1. *If (F_n^*, G_n^*) maximizes $\mathcal{E}_n$, then F_n^* and G_n^* are concentrated on $\{X_1, \ldots, X_n\}$ and $\{Y_1, \ldots, Y_n\}$, respectively.*

Proof. Let (F_n^1, G_n^1) be a maximizer of $\mathcal{E}_n$. Define the distribution function F_n^*, concentrated on $\{X_1, \ldots, X_n\}$, by

$$F_n^*(X_{i:n}) = \begin{cases} F_n^1(X_{i:n}) & \text{if } i \in \{1, \ldots, n-1\}, \\ 1 & \text{if } i = n. \end{cases}$$

Define G_n^* in a similar way. It is sufficient to show that for all $i, j \in \{1, \ldots, n\}$,

$$\mathbb{P}_{F_n^*, G_n^*}\{(X_{i:n}, Y_{j:n})\} \geq \mathbb{P}_{F_n^1, G_n^1}\{(X_{i:n}, Y_{j:n})\},$$

with equality for all i, j only if F_n^1 and G_n^1 are already concentrated on the data. For convenience, denote $X_{0:n} = Y_{0:n} = -\infty$. Then, for $i, j \in \{1, \ldots, n-1\}$,

$$\begin{aligned}
&\mathbb{P}_{F_n^*, G_n^*}\{(X_{i:n}, Y_{j:n})\} \\
&\quad = C\left(F_n^*(X_{i:n}), G_n^*(Y_{j:n})\right) - C\left(F_n^*(X_{i-1:n}), G_n^*(Y_{j:n})\right) \\
&\qquad - C\left(F_n^*(X_{i:n}), G_n^*(Y_{j-1:n})\right) + C\left(F_n^*(X_{i-1:n}), G_n^*(Y_{j-1:n})\right) \\
&\quad = C\left(F_n^1(X_{i:n}), G_n^1(Y_{j:n})\right) - C\left(F_n^1(X_{i-1:n}), G_n^1(Y_{j:n})\right) \\
&\qquad - C\left(F_n^1(X_{i:n}), G_n^1(Y_{j-1:n})\right) + C\left(F_n^1(X_{i-1:n}), G_n^1(Y_{j-1:n})\right) \\
&\quad = \mathbb{P}_{F_n^1, G_n^1}\left((X_{i-1:n}, X_{i:n}] \times (Y_{j-1:n}, Y_{j:n}]\right) \\
&\quad \geq \mathbb{P}_{F_n^1, G_n^1}\{(X_{i:n}, Y_{j:n})\},
\end{aligned}$$

the last inequality being an equality if and only if $\mathbb{P}_{F_n^1,G_n^1}$ does not assign any mass to the set $((X_{i-1:n}, X_{i:n}] \times (Y_{j-1:n}, Y_{j:n}]) \setminus \{(X_{i:n}, Y_{j:n})\}$. Similarly, we find for $i = n$ and $j \in \{1, \ldots, n-1\}$,

$$\begin{aligned}\mathbb{P}_{F_n^*,G_n^*}\{(X_{n:n}, Y_{j:n})\} &= \mathbb{P}_{F_n^1,G_n^1}\left((X_{n-1:n}, \infty) \times (Y_{j-1:n}, Y_{j:n}]\right) \\ &\geq \mathbb{P}_{F_n^1,G_n^1}\{(X_{n,n}, Y_{j:n})\},\end{aligned}$$

the last inequality being an equality if and only if $\mathbb{P}_{F_n^1,G_n^1}$ does not assign any mass to the set $((X_{n-1:n}, \infty) \times (Y_{j-1:n}, Y_{j:n}]) \setminus \{(X_{n:n}, Y_{j:n})\}$. Similar inequalities hold for the cases $i \in \{1, \ldots, n-1\}$, $j = n$, and $i = j = n$. □

Any maximizer of the empirical loglikelihood $\mathcal{E}_n(F, G)$ in (2.1) is thus necessarily of the form

$$F_n^*(x) = \sum_{i=1}^n p_{i,n} I(X_i \leq x), \qquad G_n^*(y) = \sum_{i=1}^n q_{i,n} I(Y_i \leq y),$$

with $p_{i,n}$ and $q_{i,n}$ the probability masses assigned by F_n^* and G_n^* to the atoms $\{X_i\}$ and $\{Y_i\}$, respectively (provided there are no ties in the data). Thinking of empirical loglikelihood $\mathcal{E}_n$ as a function of $\boldsymbol{p}_n = (p_{1,n}, \ldots, p_{n,n})'$ and $\boldsymbol{q}_n = (q_{1,n}, \ldots, q_{n,n})'$, the nonparametric maximum likelihood estimator is defined as the solution to the following optimization problem:

$$\begin{array}{rl}\max\limits_{\boldsymbol{p}_n,\boldsymbol{q}_n} & \mathcal{E}_n(\boldsymbol{p}_n, \boldsymbol{q}_n) \\ \text{such that} & p_{i,n} \geq 0, \quad q_{i,n} \geq 0, \qquad (i \in \{1, \ldots, n\}) \\ & \sum_{i=1}^n p_{i,n} = 1, \quad \sum_{i=1}^n q_{i,n} = 1.\end{array} \tag{2.2}$$

This constitutes a highly non-linear constrained optimization problem in $2n$ variables. In order to reduce its complexity we suggest two approximations.

3 Linearization and representation

As a first simplification of the optimization problem (2.2), we approximate the probabilities $\mathbb{P}_{F_n^*,G_n^*}\{(X_i, Y_i)\}$ in (2.1) by $p_{i,n} q_{i,n} c(F_n^*(X_i), G_n^*(Y_i))$, where c is the copula density. In this way, the objective function $\mathcal{E}_n$ is replaced by

$$\mathcal{L}_n(\boldsymbol{p}_n, \boldsymbol{q}_n) = \sum_{i=1}^n \log p_{i,n} + \sum_{i=1}^n \log q_{i,n} + \sum_{i=1}^n \ell\left(F_n^*(X_i), G_n^*(Y_i)\right), \tag{3.1}$$

with $\ell = \log c$.

In order to get a computationally feasible estimator, we suggest a second simplification: forget about the nonnegativity constraints on $p_{i,n}$ and $q_{i,n}$ and replace the objective function $\mathcal{L}_n$ by its quadratic expansion around $p_{i,n} = q_{i,n} = 1/n$. The motivation for this approximation is that we think of the estimators F_n^* and G_n^* as being "close to" the empirical distribution functions

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x) \qquad \text{and} \qquad G_n(y) = \frac{1}{n}\sum_{i=1}^n I(Y_i \le y),$$

respectively. The quadratic expansion is more conveniently described in terms of the variables $a_{i,n}$ and $b_{i,n}$ for $i \in \{1, \ldots, n\}$ defined implicitly by

$$p_{i,n} = \frac{1}{n}(1 + a_{i,n}) \qquad \text{and} \qquad q_{i,n} = \frac{1}{n}(1 + b_{i,n}).$$

In view of the equality constraints $\sum_{i=1}^n p_{i,n} = 1$ and $\sum_{i=1}^n q_{i,n} = 1$, these new variables must satisfy the equality constraints $\sum_{i=1}^n a_{i,n} = \sum_{i=1}^n b_{i,n} = 0$. For the sake of computational simplicity, the inequality constraints $a_{i,n} \ge -1$ and $b_{i,n} \ge -1$ will be ignored.

Think of $\mathcal{L}_n$ as a function of the $2n$ variables $(\boldsymbol{a}_n', \boldsymbol{b}_n')'$ and denote the $n \times 1$ vectors of first-order partial derivatives by

$$\begin{aligned} \dot{\boldsymbol{\mathcal{L}}}_{n,a} &= \left(\frac{\partial \mathcal{L}_n}{\partial a_{1,n}}, \ldots, \frac{\partial \mathcal{L}_n}{\partial a_{n,n}} \right)', \\ \dot{\boldsymbol{\mathcal{L}}}_{n,b} &= \left(\frac{\partial \mathcal{L}_n}{\partial b_{1,n}}, \ldots, \frac{\partial \mathcal{L}_n}{\partial b_{n,n}} \right)', \end{aligned}$$

all partial derivatives being calculated in $\boldsymbol{a}_n = \mathbf{0}$ and $\boldsymbol{b}_n = \mathbf{0}$. Similarly, denote the $n \times n$ matrices of second-order partial derivatives by

$$\begin{aligned} \ddot{\boldsymbol{\mathcal{L}}}_{n,aa} &= \left(\frac{\partial^2 \mathcal{L}_n}{\partial a_{i,n} \partial a_{j,n}} \right)_{i,j=1}^n, & \ddot{\boldsymbol{\mathcal{L}}}_{n,ab} &= \left(\frac{\partial^2 \mathcal{L}_n}{\partial a_{i,n} \partial b_{j,n}} \right)_{i,j=1}^n, \\ \ddot{\boldsymbol{\mathcal{L}}}_{n,ba} &= \left(\frac{\partial^2 \mathcal{L}_n}{\partial b_{i,n} \partial a_{j,n}} \right)_{i,j=1}^n, & \ddot{\boldsymbol{\mathcal{L}}}_{n,bb} &= \left(\frac{\partial^2 \mathcal{L}_n}{\partial b_{i,n} \partial b_{j,n}} \right)_{i,j=1}^n, \end{aligned}$$

again with all partial derivatives being calculated in $\boldsymbol{a}_n = \mathbf{0}$ and $\boldsymbol{b}_n = \mathbf{0}$. Then up to constant terms, the quadratic expansion of the objective function $\mathcal{L}_n$ around $\boldsymbol{a}_n = \mathbf{0}$ and $\boldsymbol{b}_n = \mathbf{0}$ is equal to

$$\begin{aligned} \mathcal{Q}_n(\boldsymbol{a}_n, \boldsymbol{b}_n) = {} & \boldsymbol{a}_n' \dot{\boldsymbol{\mathcal{L}}}_{n,a} + \boldsymbol{b}_n' \dot{\boldsymbol{\mathcal{L}}}_{n,b} \\ & + \frac{1}{2}\Big(\boldsymbol{a}_n' \ddot{\boldsymbol{\mathcal{L}}}_{n,aa} \boldsymbol{a}_n + \boldsymbol{a}_n' \ddot{\boldsymbol{\mathcal{L}}}_{n,ab} \boldsymbol{b}_n + \boldsymbol{b}_n' \ddot{\boldsymbol{\mathcal{L}}}_{n,ba} \boldsymbol{a}_n + \boldsymbol{b}_n' \ddot{\boldsymbol{\mathcal{L}}}_{n,bb} \boldsymbol{b}_n \Big). \end{aligned}$$

Our linearized estimator is now defined as the solution of the following optimization problem:

$$\begin{aligned} \max_{\boldsymbol{a}_n, \boldsymbol{b}_n} \quad & \mathcal{Q}_n(\boldsymbol{a}_n, \boldsymbol{b}_n) \\ \text{such that} \quad & \textstyle\sum_{i=1}^n a_{i,n} = 0, \\ & \textstyle\sum_{i=1}^n b_{i,n} = 0. \end{aligned} \tag{3.2}$$

By the method of Lagrange multipliers, the solution to (3.2) can be found by solving the following linear system of $2n+2$ equations in $2n+2$ unknowns:

$$\begin{aligned} \dot{\boldsymbol{\mathcal{L}}}_{n,a} + \ddot{\boldsymbol{\mathcal{L}}}_{n,aa}\boldsymbol{a}_n + \ddot{\boldsymbol{\mathcal{L}}}_{n,ab}\boldsymbol{b}_n + \kappa\mathbf{1} &= \mathbf{0}, \\ \dot{\boldsymbol{\mathcal{L}}}_{n,b} + \ddot{\boldsymbol{\mathcal{L}}}_{n,ba}\boldsymbol{a}_n + \ddot{\boldsymbol{\mathcal{L}}}_{n,bb}\boldsymbol{b}_n + \lambda\mathbf{1} &= \mathbf{0}, \\ \mathbf{1}'\boldsymbol{a}_n &= 0, \\ \mathbf{1}'\boldsymbol{b}_n &= 0. \end{aligned} \tag{3.3}$$

Here, κ and λ are the Lagrange multipliers arising from the equality constraints on $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$, respectively, and $\mathbf{0}$ and $\mathbf{1}$ denote $n \times 1$ vectors of zeros and ones, respectively. Convenient matrix expressions for the above system are derived in the appendix.

Although (3.3) constitutes a representation of the linearized nonparametric maximum likelihood estimator that is convenient for computational purposes, it provides little insight in the nature of the estimator. The remainder of this section is devoted to a representation of the estimator as the solution to a system of estimating equations defined in an appropriate function space.

In order to describe these equations, we need some more notation. To the solutions $\boldsymbol{a}_n$ and $\boldsymbol{b}_n$ to (3.3), there correspond functions $a_n, b_n : (0,1] \to \mathbb{R}$ defined by

$$\begin{aligned} a_n(u) &= a_{i,n} \qquad \text{if } F_n(X_i) - \tfrac{1}{n} < u \le F_n(X_i), \\ b_n(v) &= b_{i,n} \qquad \text{if } G_n(X_i) - \tfrac{1}{n} < v \le G_n(Y_i). \end{aligned}$$

Further, for $u, v \in [0,1]$, put

$$A_n(u) = \int_0^u a_n(z)\,\mathrm{d}z \qquad \text{and} \qquad B_n(v) = \int_0^v b_n(z)\,\mathrm{d}z.$$

The (right-continuous with left-hand limits) empirical copula is defined by

$$C_n(u,v) = \frac{1}{n}\sum_{i=1}^n I\{F_n(X_i) \le u,\, G_n(Y_i) \le v\}$$

for $u, v \in [0,1]^2$. Note that $\int f\,\mathrm{d}C_n = n^{-1}\sum_{i=1}^n f(F_n(X_i), G_n(Y_i))$ for $f : [0,1]^2 \to \mathbb{R}$. For $h_1, h_2 : [0,1] \to \mathbb{R}$, define new functions by

$$\begin{aligned} &\Psi_{1,n}(h_1,h_2)(u) \\ &\quad = -h_1(u) + \int (x \wedge u - xu)\{h_1(x)\ddot{\ell}_{11}(x,y) + h_2(y)\ddot{\ell}_{12}(x,y)\}\,\mathrm{d}C_n(x,y), \\ &\Psi_{2,n}(h_1,h_2)(v) \\ &\quad = -h_2(v) + \int (y \wedge v - yv)\{h_2(y)\ddot{\ell}_{22}(x,y) + h_1(x)\ddot{\ell}_{12}(x,y)\}\,\mathrm{d}C_n(x,y), \end{aligned}$$

for $u, v \in [0,1]$, the integrals extending over the unit square. The following result is proved in the appendix.

PROPOSITION 3.1. *For* $\boldsymbol{a}_n$ *and* $\boldsymbol{b}_n$ *solving* (3.3)*, the corresponding estimators* F_n^* *and* G_n^* *can be written as*

$$\begin{aligned} F_n^*(x) &= F_n(x) + A_n(F_n(x)), \\ G_n^*(y) &= G_n(y) + B_n(G_n(y)), \end{aligned} \tag{3.4}$$

for $x, y \in [0,1]$*, the functions* A_n *and* B_n *solving for all* $u, v \in [0,1]$ *the equations*

$$\begin{aligned} \Psi_{1,n}(A_n, B_n)(u) &= -\int (x \wedge u - xu)\dot{\ell}_1(x,y)\,\mathrm{d}C_n(x,y), \\ \Psi_{2,n}(A_n, B_n)(v) &= -\int (y \wedge v - yv)\dot{\ell}_2(x,y)\,\mathrm{d}C_n(x,y). \end{aligned} \tag{3.5}$$

REMARK 3.2. If on the right-hand sides of (3.5), the empirical copula C_n is replaced by the true copula C, then, under Conditions 5.1–5.2 below, the integrals on the right-hand side are zero. Based on results on the empirical copula process $\sqrt{n}(C_n - C)$, we conjecture that the right-hand sides of (3.5) multiplied by $\sqrt{n}$ converge jointly to a pair of Gaussian process on $[0,1]$. If, in addition, the operator sequence $\Psi_n = (\Psi_{1,n}, \Psi_{2,n})$ converges in an appropriate way to a limiting operator $\Psi = (\Psi_1, \Psi_2)$ (just replace C_n by C in the definitions of $\Psi_{i,n}$), an operator which moreover is continuously invertible, then the sequence of processes $\sqrt{n}(A_n, B_n)$ must converge to a Gaussian process as well. This approach is the subject of a future paper; for more details, we refer to Van den Akker (2007, Chapter 5).

4 Comparison with marginal empirical distribution functions

This section compares the finite-sample performances of the linearized nonparametric maximum likelihood estimator and the marginal empirical distribution function as estimators for the marginal distribution functions in the model where the copula of a bivariate distribution is known and the marginal distributions are arbitrary.

4.1 Set-up of the experiments

We generated samples from the bivariate Gaussian, Plackett, and Frank copula families; see Table 1 for an overview of the precise settings. We refer to Nelsen (1999), in particular for the parametrization, for details and references on these copula families. For each sample, we estimated the marginal distribution of the first component, $F(u) = u$, by the empirical distribution function and by the linearized nonparametric maximum likelihood estimator. Only uniform marginals need to be considered as the results for other margins are just a "transformation of the axes" (this since A_n and B_n only

Table 1: Settings for the simulation experiments.

Figure 1	Gaussian copula at $\rho \in \{\pm 0.2, \pm 0.5, \pm 0.8\}$ $M = 10\,000$ samples of size $n = 100$
Figure 2	Gaussian copula at $\rho = 0.75$ (top) and $\rho = 0.90$ (bottom) $M = 10\,000$ samples of size $n = 100$ (left) $M = 5\,000$ samples of size $n = 1\,000$ (right)
Figure 3	Plackett copula at $\theta \in \{2, 5\} \cup \{1/2, 1/5\}$ $M = 10\,000$ samples of size $n = 100$
Figure 4	Plackett copula at $\theta \in \{10, 20\} \cup \{1/10, 1/20\}$ $M = 10\,000$ samples of size $n = 100$
Figure 5	Frank copula at $a = -\log\theta \in \{\pm 1, \pm 2, \pm 3\}$ $M = 5\,000$ samples of size $n = 500$
Figure 6	Frank copula at $a = -\log\theta \in \{\pm 5, \pm 10, \pm 20\}$ $M = 5\,000$ samples of size $n = 100$

depend on the ranks of the observations). In each experiment and for each u in a grid of values on $[0, 1]$, we estimated the normalized mean squared error $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ by its empirical counterpart over the M samples. Note that in case of the empirical distribution function, the normalized mean squared error is equal to $u(1 - u)$.

4.2 Results

The figures are organized at the back. For the *Gaussian copula* (Figures 1–2), the nonparametric likelihood estimator was always at least as good as the empirical distribution function. The difference in efficiency becomes larger for ρ farther away from zero. Figure 2 shows that for large ρ, the improvement in efficiency is much more pronounced for $n = 1000$ than for $n = 100$; a tentative explanation is that for these values of ρ, the sample size needs to be rather larger in order for the various approximations (linearization; omission of inequality constraints) to be effective.

For the *Plackett copula* (Figures 3–4), the results indicate a large potential of efficiency gain by incorporating knowledge of the copula. The increase in efficiency becomes larger when the odds-ratio parameter θ moves farther away from one, i.e., from independence. The results for θ and $1/\theta$ are virtually the same; this is to be expected, for if the distribution of (U, V) is the Plackett copula with parameter θ, then the distribution of $(1 - U, V)$ is the Plackett copula with parameter $1/\theta$.

For the *Frank copula* (Figures 5–6), the results are quite favorable for the nonparametric maximum likelihood estimator. Within the range $-20 \leq a \leq 20$ covered by the simulations, the relative efficiency of the nonparametric maximum likelihood estimator with respect to the empirical distribution function was found to be increasing in $|a|$, with a maximum of about $0.25/0.15 = 5/3$ for $|a| = 20$ and in $u = 0.5$. In this setting, the nonparametric maximum likelihood estimator achieves the same accuracy as the empirical distribution function with only about 60% of the observations.

5 Asymptotic inefficiency of the marginal empirical distribution functions

The previous section shows that the linearized nonparametric maximum likelihood estimator (3.3) of the marginal distribution improves upon the empirical distribution function, the size of the improvement depending on the underlying copula. Intuitively it is clear that no improvement will be possible under the independence copula $C(u, v) = uv$. Theorem 5.4 below shows that the converse is true as well: the marginal empirical distribution functions are efficient only under the independence copula.

In order to prove this result rigorously, we derive the tangent space for our copula model. For details on this concept and asymptotic efficiency we refer to Bickel et al. (1998) and Van der Vaart (2000, Chapter 25). First we describe the regularity assumptions on the copula C we need.

Condition 5.1. *C is absolutely continuous with respect to Lebesgue measure. There is a version of its density, c, which is strictly positive on $(0,1)^2$ and which is two times continuously differentiable on $(0,1)^2$.*

As before, denote the partial derivatives of the log density $\ell = \log c$ by $\dot{\ell}_i$ and $\ddot{\ell}_{ij}$ for $i, j \in \{1, 2\}$. For $x, y \in (0, 1)$, also define

$$I_{11}(x) = \int_0^1 \dot{\ell}_1^2(x,y)c(x,y)\,\mathrm{d}y \quad \text{and} \quad I_{22}(y) = \int_0^1 \dot{\ell}_2^2(x,y)c(x,y)\,\mathrm{d}x.$$

Condition 5.2. *For some constant $M > 0$ and for all $i \in \{1, 2\}$ and $u \in (0, 1)$,*

$$I_{ii}(u) \leq \frac{M}{u^2(1-u)^2}.$$

Moreover, for $i \in \{1, 2\}$ and $u_i \in (0, 1)$,

$$\begin{aligned} \int_0^1 \dot{\ell}_i(u_1,u_2)c(u_1,u_2)\,\mathrm{d}u_{3-i} &= 0, \\ \int_0^1 \ddot{\ell}_{ii}(u_1,u_2)c(u_1,u_2)\,\mathrm{d}u_{3-i} &= -I_{ii}(u_i). \end{aligned}$$

Note that the second part of Condition 5.2 is similar to the standard smoothness conditions in parametric models: scores have mean zero, and the Fisher-equality holds: the expectation of the outerproduct of the scores equals minus the expectation of the derivative of the score. The first part of Condition 5.2 puts a condition for copulas that are "exploding" on the boundary of the unit square. The conditions are standard in the semiparametric literature on copulas; see Section 4.7 of Bickel et al. (1998).

Recall that $\mathbb{P}_{F,G} = \mathbb{P}^C_{F,G}$ is the probability distribution on $\mathbb{R}^2$ with marginal distribution functions F and G and copula C; expectations with respect to $\mathbb{P}_{F,G}$ are denoted by $\mathbb{E}_{F,G}$. Let $\mathcal{F}_{\text{ac}}$ be the set of absolutely continuous distribution functions. The model of interest is the known-copula model $\mathcal{P}(C) = (\mathbb{P}_{F,G} \mid F, G \in \mathcal{F}_{\text{ac}})$.

Fix $F_0, G_0 \in \mathcal{F}_{\text{ac}}$. We describe how to construct a tangent space for our model $\mathcal{P}(C)$ at $\mathbb{P}_0 = \mathbb{P}_{F_0,G_0}$. Let $k : \mathbb{R} \to \mathbb{R}$ be defined by $k(z) = 2/\{1 + \exp(-2z)\}$. Let $v, w \in L_2^0(\text{Un}[0,1])$, the subset of elements a from $L_2(\text{Un}[0,1])$ that satisfy $\int_0^1 a(u)\mathrm{d}u = 0$. Define densities, for $t \in \mathbb{R}$, by

$$\begin{aligned} f_t^v(x) &= c_f^v(t)k(tv(F_0(x)))f_0(x), \\ g_t^w(y) &= c_g^w(t)k(tw(G_0(y)))g_0(y), \end{aligned} \tag{5.1}$$

where $c_f^v(t)$ and $c_g^w(t)$ are normalizing constants ensuring that f_t^v and g_t^w are indeed densities. [It is a matter of routine to check that $0 < k \le 2$, $0 < k' \le 4$, $0 < k'/k \le 2$ and $k(0) = k'(0) = 1$, that $c_f^v(0) = c_g^w(0) = 1$, and that $t \mapsto c_f^v(t)$ and $t \mapsto c_g^w(t)$ are continuously differentiable with $(c_f^v)'(0) = (c_g^w)'(0) = 0$.] The densities f_t^v and g_t^w induce distribution functions $F_t^v, G_t^w \in \mathcal{F}_{\text{ac}}$, and the paths $t \mapsto F_t^v$, $t \mapsto G_t^w$ pass F_0 and G_0 at $t = 0$.

Next we introduce the score operators $\dot{\ell}_F$ and $\dot{\ell}_G$, which are mappings from $L_2^0(\text{Un}[0,1])$ into $L_2(\mathbb{P}_{\text{Un}[0,1],\text{Un}[0,1]})$, by

$$\begin{aligned} \dot{\ell}_F v(x,y) &= v(x) + \dot{\ell}_1(x,y)\int_0^x v(z)\mathrm{d}z, \\ \dot{\ell}_G w(x,y) &= w(y) + \dot{\ell}_2(x,y)\int_0^y w(z)\mathrm{d}z; \end{aligned}$$

see Bickel et al. (1998, Proposition 4.7.5). The following proposition yields a tangent space at $\mathbb{P}_0$.

Lemma 5.3. *Let C satisfy Conditions 5.1–5.2. For $v, w \in L_2^0(Un[0,1])$, the path $t \mapsto \mathbb{P}_{F_t^v,G_t^w}$ in $\mathcal{P}(C)$, as induced by (5.1), has the following score at $t = 0$:*

$$\dot{\ell}^{v,w}_{F_0,G_0}(x,y) = \dot{\ell}_F v(F_0(x), G_0(y)) + \dot{\ell}_G w(F_0(x), G_0(y));$$

that is, writing $p_t(x,y) = p_{F_t^v,G_t^w}(x,y)$, as $t \to 0$,

$$\iint_{\mathbb{R}^2} \left(\frac{\sqrt{p_t(x,y)} - \sqrt{p_0(x,y)}}{t} - \frac{1}{2}\dot{\ell}^{v,w}_{F_0,G_0}(x,y)\sqrt{p_0(x,y)} \right)^2 \mathrm{d}x\,\mathrm{d}y \to 0.$$

This yields the following tangent space at $\mathbb{P}_0$*:*

$$\mathcal{T}_0 = \mathcal{T}(\mathbb{P}_0 \mid \mathcal{P}(C)) = \big\{\dot\ell^{v,w}_{F_0,G_0} \mid v, w \in L_2^0(\mathrm{Un}[0,1])\big\},$$

which is a closed linear subspace of $L_2(\mathbb{P}_0)$*.*

Proof. The part on the score (essentially) follows from Bickel et al. (1998, Proposition 4.7.4), while the closedness is a consequence of Theorem A.4.2.B and (an easy generalization of) Proposition 4.7.6 in Bickel et al. (1998). □

Our parameter of interest is described by the mapping $\nu : \mathcal{P}(C) \to \ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$ defined by $\nu(\mathbb{P}_{F,G}) = (F, G)$. Fix $F_0, G_0 \in \mathcal{F}_{\mathrm{ac}}$. We need the pathwise derivative of ν along the paths that generate the tangent space $\mathcal{T}_0$. For a path $t \mapsto \mathbb{P}_{F_t^v, G_t^w}$ as in (5.1), it is an easy exercise to show that in $\ell^\infty(\mathbb{R}) \times \ell^\infty(\mathbb{R})$ and as $t \to 0$,

$$\begin{aligned} t^{-1}\{\nu(\mathbb{P}_{F_t^v, G_t^w}) - \nu(\mathbb{P}_0)\} &\to \nu'_{\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0}) \\ &= \Big(\nu^{1'}_{\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0}), \nu^{2'}_{\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0})\Big), \end{aligned}$$

where, for $x, y \in \mathbb{R}$,

$$\begin{aligned} \nu^{1'}_{\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0})(x) &= \int_0^{F_0(x)} v(z)\,\mathrm{d}z, \\ \nu^{2'}_{\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0})(y) &= \int_0^{G_0(y)} w(z)\,\mathrm{d}z. \end{aligned}$$

For $x, y \in \mathbb{R}$ there exist, by the Riesz representation theorem, unique elements $\nu^{1*}_{x,\mathbb{P}_0}$ and $\nu^{2*}_{y,\mathbb{P}_0}$ in $\mathcal{T}_0$ such that, for all $v, w \in L_2^0(\mathrm{Un}[0,1])$,

$$\begin{aligned} \nu'_{1,\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0})(x) &= \mathbb{E}_{F_0,G_0}\nu^{1*}_{x,\mathbb{P}_0}\dot\ell^{v,w}_{F_0,G_0}(X,Y), \\ \nu'_{2,\mathbb{P}_0}(\dot\ell^{v,w}_{F_0,G_0})(y) &= \mathbb{E}_{F_0,G_0}\nu^{2*}_{y,\mathbb{P}_0}\dot\ell^{v,w}_{F_0,G_0}(X,Y). \end{aligned}$$

These elements are called efficient influence functions and they are given by

$$\begin{aligned} \nu^{1*}_{x,\mathbb{P}_0} &= \Pi(1\{X \le x\} - F_0(x) \mid \mathcal{T}_0), \\ \nu^{2*}_{y,\mathbb{P}_0} &= \Pi(1\{Y \le y\} - G_0(y) \mid \mathcal{T}_0), \end{aligned}$$

where Π denotes the projection operator. Unfortunately it seems to be impossible to obtain explicit expressions for these projections. One reason for this complication is that the tangent space is the sum of two non-orthogonal spaces.

These efficient influence functions characterize efficient estimators through the infinite-dimensional version of the famous Hájek-Le Cam convolution theorem; see, for example, Van der Vaart (1991, Theorem 2.1) or Bickel et al. (1998, Theorem 5.2.1). Using this theorem, the next proposition shows that, amongst the copulas satisfying Conditions 5.1–5.2, the independence copula is the only one for which (F_n, G_n) constitutes an efficient estimator of (F, G).

THEOREM 5.4. *Let the copula C satisfy Conditions 5.1–5.2. Then (F_n, G_n) is an efficient estimator of (F, G) in the model $\mathcal{P}(C)$ if and only if $C(u, v) = uv$.*

Proof. Let $F_0, G_0 \in \mathcal{F}_{\mathrm{ac}}$. Using Bickel et al. (1998, Corollary 5.2.1) and the "transformation of axes" structure of the tangent space, it is easy to see that (F_n, G_n) is efficient at $\mathbb{P}_0$ if and only if (F_n, G_n) is efficient at $\mathbb{P}_{\mathrm{Un}[0,1],\mathrm{Un}[0,1]}$. Therefore we only consider uniform margins in the sequel of the proof. Since no confusion can arise we drop subscripts related to the margins.

Sufficiency. If $C(u, v) = uv$, it is easy to check that $\nu^{1*}_\alpha = 1\{X \le \alpha\} - \alpha$ and $\nu^{2*}_\beta = 1\{Y \le \beta\} - \beta$ for all $\alpha, \beta \in [0, 1]$. Efficiency now follows directly from Bickel et al. (1998, Corollary 5.2.1).

Necessity. Since F_n is an efficient estimator of F, for all $\alpha \in [0, 1]$, the influence function of $F_n(\alpha)$, which is $x \mapsto 1\{x \le \alpha\} - \alpha$, belongs to the tangent space $\mathcal{T}(\mathbb{P}_{\mathrm{Un}[0,1],\mathrm{Un}[0,1]} \mid \mathcal{P}(C))$, i.e. there exists $a^\alpha, b^\alpha \in L^0_2(\mathrm{Un}[0, 1])$ such that

$$1\{X \le \alpha\} - \alpha = \dot{\ell}_F a^\alpha(X, Y) + \dot{\ell}_G b^\alpha(X, Y) \qquad \text{a.s.} \tag{5.2}$$

By partial integration, for $a, b \in L^0_2(\mathrm{Un}[0, 1])$,

$$\begin{aligned} \mathbb{E}[\dot{\ell}_F a(X, Y) + \dot{\ell}_G b(X, Y) \mid X] &= a(X), \\ \mathbb{E}[\dot{\ell}_F a(X, Y) + \dot{\ell}_G b(X, Y) \mid Y] &= b(Y), \end{aligned}$$

whence

$$\begin{aligned} a^\alpha(x) &= 1\{x \le \alpha\} - \alpha, \\ b^\alpha(y) &= \int_0^\alpha c(z, y)\,\mathrm{d}z - \alpha. \end{aligned}$$

In combination with (5.2), the above equations yield, for all $x, y, \alpha \in (0, 1)$,

$$-\dot{\ell}_1(x, y)(x \wedge \alpha - x\alpha) = \int_0^\alpha c(z, y)\,\mathrm{d}z - \alpha + \dot{\ell}_2(x, y)\{C(\alpha, y) - \alpha y\}; \tag{5.3}$$

since all functions involved are continuous the "a.s." disappears. In case $x < \alpha$ differentiating both sides of (5.3) with respect to x yields

$$-(1 - \alpha)\{x\ddot{\ell}_{11}(x, y) + \dot{\ell}_1(x, y)\} = \ddot{\ell}_{12}(x, y)\{C(\alpha, y) - \alpha y\}, \tag{5.4}$$

and in case $x > \alpha$ we have

$$-\alpha\{(1 - x)\ddot{\ell}_{11}(x, y) - \dot{\ell}_1(x, y)\} = \ddot{\ell}_{12}(x, y)\{C(\alpha, y) - \alpha y\}. \tag{5.5}$$

Fix $x, y \in (0, 1)$. Since all objects involved are continuous, we obtain, by letting $\alpha \downarrow x$ in (5.4) and $\alpha \uparrow x$ in (5.5),

$$(1 - x)\{x\ddot{\ell}_{11}(x, y) + \dot{\ell}_1(x, y)\} = x\{(1 - x)\ddot{\ell}_{11}(x, y) - \dot{\ell}_1(x, y)\}.$$

Trivially, this yields $\dot{\ell}_1(x, y) = 0$. Hence $c_1(x, y) = 0$. So $x \mapsto c(x, y)$ is constant. This yields $c(x, y) = 1$. $\square$

Remark 5.5. From the proof we see that actually a stronger result holds: if the copula is known to be C, then F_n (G_n) is an efficient estimator of F (G) only if C is the independence copula. In other words, we only need efficiency of one marginal empirical distribution function to be able to conclude that the copula must be the independence copula.

Acknowledgements

Helpful discussions with John H. J. Einmahl (Tilburg University) and Christian Genest (Université Laval) are gratefully acknowledged. The research by Johan Segers was supported financially by a VENI grant of the Netherlands Science Organisation (NWO) and by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).

A Derivation and representation of the new estimator

In this appendix we prove the computationally more convenient characterization of the nonparametric maximum likelihood estimator (3.3) and provide the proof of Proposition 3.1.

Matrix formulation of (3.3). The system (3.3) can be written as

$$\begin{pmatrix} \ddot{\boldsymbol{\mathcal{L}}}_{n,aa} & \ddot{\boldsymbol{\mathcal{L}}}_{n,ab} & \mathbf{1} & \mathbf{0} \\ \ddot{\boldsymbol{\mathcal{L}}}_{n,ba} & \ddot{\boldsymbol{\mathcal{L}}}_{n,bb} & \mathbf{0} & \mathbf{1} \\ \mathbf{1}' & \mathbf{0}' & 0 & 0 \\ \mathbf{0}' & \mathbf{1}' & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_n \\ \boldsymbol{b}_n \\ \kappa \\ \lambda \end{pmatrix} = - \begin{pmatrix} \dot{\boldsymbol{\mathcal{L}}}_{n,a} \\ \dot{\boldsymbol{\mathcal{L}}}_{n,b} \\ 0 \\ 0 \end{pmatrix}. \tag{A.1}$$

The entries in (A.1) involving $\mathcal{L}_n$ can be computed as follows. The first-order partial derivatives are given by

$$\begin{aligned} \frac{\partial \mathcal{L}_n}{\partial a_{i,n}} &= 1 + \frac{1}{n} \sum_{k=1}^n \dot{\ell}_1(F_n(X_k), G_n(Y_k)) I(X_i \le X_k), \\ \frac{\partial \mathcal{L}_n}{\partial b_{i,n}} &= 1 + \frac{1}{n} \sum_{k=1}^n \dot{\ell}_2(F_n(X_k), G_n(Y_k)) I(Y_i \le Y_k), \end{aligned}$$

for $i \in \{1, \ldots, n\}$. If the log copula density ℓ is unbounded on some part of the boundary of the unit square, the empirical distribution functions F_n and G_n can be replaced by the rescaled versions $\tilde{F}_n(x) = (n+1)^{-1} \sum_{l=1}^n I(X_l \le x)$ and $\tilde{G}_n(y) = (n+1)^{-1} \sum_{l=1}^n I(Y_l \le y)$. The second-order partial derivatives of $\mathcal{L}_n$ are

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}_n}{\partial a_{i,n} \partial a_{j,n}} &= -\delta_{ij} + \frac{1}{n^2} \sum_{k=1}^n \ddot{\ell}_{11}(F_n(X_k), G_n(X_k)) I(X_i \le X_k) I(X_j \le X_k), \\
\frac{\partial^2 \mathcal{L}_n}{\partial a_{i,n} \partial b_{j,n}} &= \quad + \frac{1}{n^2} \sum_{k=1}^n \ddot{\ell}_{12}(F_n(X_k), G_n(X_k)) I(X_i \le X_k) I(Y_j \le Y_k), \\
\frac{\partial^2 \mathcal{L}_n}{\partial b_{i,n} \partial a_{j,n}} &= \quad + \frac{1}{n^2} \sum_{k=1}^n \ddot{\ell}_{21}(F_n(X_k), G_n(X_k)) I(Y_i \le Y_k) I(X_j \le X_k), \\
\frac{\partial^2 \mathcal{L}_n}{\partial b_{i,n} \partial b_{j,n}} &= -\delta_{ij} + \frac{1}{n^2} \sum_{k=1}^n \ddot{\ell}_{22}(F_n(X_k), G_n(X_k)) I(Y_i \le Y_k) I(Y_j \le Y_k),
\end{aligned}$$

for $i, j \in \{1, \ldots, n\}$. These identities can be written down more succinctly in matrix notation as follows. Define the $n \times n$ matrices

$$\boldsymbol{A}_n = \frac{1}{n} \big(I(X_i \le X_j) \big)_{i,j=1}^n \qquad \text{and} \qquad \boldsymbol{B}_n = \frac{1}{n} \big(I(Y_i \le Y_j) \big)_{i,j=1}^n.$$

For $r \in \{1, 2\}$, put the $n \times 1$ vectors of first-order partial derivatives

$$\dot{\boldsymbol{\ell}}_{n,r} = \Big(\dot{\ell}_r(F_n(X_1), G_n(Y_1)), \ldots, \dot{\ell}_r(F_n(X_n), G_n(Y_n)) \Big)'$$

and for $r, s \in \{1, 2\}$, put the $n \times n$ diagonal matrices of second-order partial derivatives

$$\ddot{\boldsymbol{\ell}}_{n,rs} = \begin{pmatrix} \ddot{\ell}_{rs}(F_n(X_1), G_n(Y_1)) & & 0 \\ & \ddots & \\ 0 & & \ddot{\ell}_{rs}(F_n(X_n), G_n(Y_n)) \end{pmatrix}.$$

Using these notations, we can write the blocks in (A.1) involving $\mathcal{L}_n$ as

$$\begin{pmatrix} \dot{\boldsymbol{\mathcal{L}}}_{n,a} \\ \dot{\boldsymbol{\mathcal{L}}}_{n,b} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_n \dot{\boldsymbol{\ell}}_{n,1} + \boldsymbol{1} \\ \boldsymbol{B}_n \dot{\boldsymbol{\ell}}_{n,2} + \boldsymbol{1} \end{pmatrix}, \tag{A.2}$$

$$\begin{pmatrix} \ddot{\boldsymbol{\mathcal{L}}}_{n,aa} & \ddot{\boldsymbol{\mathcal{L}}}_{n,ab} \\ \ddot{\boldsymbol{\mathcal{L}}}_{n,ba} & \ddot{\boldsymbol{\mathcal{L}}}_{n,bb} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_n \ddot{\boldsymbol{\ell}}_{n,11} \boldsymbol{A}_n' & \boldsymbol{A}_n \ddot{\boldsymbol{\ell}}_{n,12} \boldsymbol{B}_n' \\ \boldsymbol{B}_n \ddot{\boldsymbol{\ell}}_{n,21} \boldsymbol{A}_n' & \boldsymbol{B}_n \ddot{\boldsymbol{\ell}}_{n,22} \boldsymbol{B}_n' \end{pmatrix} - \boldsymbol{I}_{2n}. \tag{A.3}$$

It is the representation of the estimator in (A.1), (A.2) and (A.3) that has been used in the simulation study in Section 4. □

Proof of Proposition 3.1. Since the function a_n is constant on intervals of the form $((r-1)/n, r/n]$, we have $A_n(k/n) = n^{-1} \sum_{r=1}^k a_n(r/n)$ for all $k \in \{1, \ldots, n\}$. Now $a_n(r/n) = a_{i,n}$ if and only if $F_n(X_i) = r/n$, that is, the rank of X_i among $X_1, \ldots, X_n$ is r. Therefore,

$$A_n(k/n) = \frac{1}{n} \sum_{i=1}^n a_{i,n} I\{F_n(X_i) \le k/n\}$$

for all $k \in \{1, \ldots, n\}$. Since $F_n(X_i) \le F_n(X_j)$ if and only if $X_i \le X_j$, we get

$$A_n(F_n(X_j)) = \frac{1}{n} \sum_{i=1}^n a_{i,n} I(X_i \le X_j) = F_n^*(X_j) - F_n(X_j) \tag{A.4}$$

for all $j \in \{1, \ldots, n\}$. Since F_n^* and F_n are distribution functions supported by $\{X_1, \ldots, X_n\}$, we conclude that $A_n(F_n(x)) = F_n^*(x) - F_n(x)$ for all $x \in \mathbb{R}$. Similarly $B_n(G_n(y)) = G_n^*(y) - G_n(y)$ for all $y \in \mathbb{R}$.

Write $\dot\ell_{n,1}(k) = \dot\ell_1(F_n(X_k), G_n(X_k))$ etc. The first n equations in (3.3) read

$$\begin{aligned} 0 &= 1 + \frac{1}{n} \sum_{k=1}^n \dot\ell_{n,1}(k) I(X_i \le X_k) \\ &\quad - a_{i,n} + \sum_{j=1}^n \frac{1}{n^2} \sum_{k=1}^n \ddot\ell_{n,11}(k) I(X_i \le X_k) I(X_j \le X_k) a_{j,n} \\ &\quad + \sum_{j=1}^n \frac{1}{n^2} \sum_{k=1}^n \ddot\ell_{n,12}(k) I(X_i \le X_k) I(Y_j \le Y_k) b_{j,n} + \kappa \end{aligned}$$

for $i \in \{1, \ldots, n\}$. In view of $n^{-1} \sum_{j=1}^n a_{j,n} I(X_j \le X_k) = A_n(F_n(X_k))$ and $n^{-1} \sum_{j=1}^n b_{j,n} I(Y_j \le Y_k) = B_n(G_n(Y_k))$, see (A.4), we obtain

$$\begin{aligned} 0 &= 1 + \frac{1}{n} \sum_{k=1}^n \dot\ell_{n,1}(k) I(X_i \le X_k) - a_{i,n} \\ &\quad + \frac{1}{n} \sum_{k=1}^n I(X_i \le X_k) \{\ddot\ell_{n,11}(k) A_n(F_n(X_k)) + \ddot\ell_{n,12}(k) B_n(G_n(Y_k))\} + \kappa. \end{aligned}$$

Sum these equations over $i \in \{1, \ldots, n\}$ and use the constraint $\sum_i a_{i,n} = 0$ to solve for the Lagrange multiplier κ:

$$\begin{aligned} \kappa &= -1 - \frac{1}{n} \sum_{k=1}^n \dot\ell_{n,1}(k) F_n(X_k) \\ &\quad - \frac{1}{n} \sum_{k=1}^n F_n(X_k) \{\ddot\ell_{n,11}(k) A_n(F_n(X_k)) + \ddot\ell_{n,12}(k) B_n(G_n(Y_k))\}. \end{aligned}$$

Substitute this expression back into the original equations: for all $i \in \{1, \ldots, n\}$,

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{k=1}^n \dot\ell_{n,1}(k) \{I(X_i \le X_k) - F_n(X_k)\} - a_{i,n} \\ &\quad + \frac{1}{n} \sum_{k=1}^n \{I(X_i \le X_k) - F_n(X_k)\} \\ &\qquad\qquad \times \{\ddot\ell_{n,11}(k) A_n(F_n(X_k)) + \ddot\ell_{n,12}(k) B_n(G_n(Y_k))\}. \end{aligned}$$

Now $I(X_i \le X_k) = I\{F_n(X_i) \le F_n(X_k)\}$ while $a_{i,n} = a_n(F_n(X_i))$. Therefore, we can rewrite the previous equation in terms of integrals with respect to C_n: for all $i \in \{1, \dots, n\}$,

$$\begin{aligned} 0 &= \int \dot{\ell}_1(x,y)[I\{F_n(X_i) \le x\} - x]\,\mathrm{d}C_n(x,y) - a_n(F_n(X_i)) \\ &\quad + \int [I\{F_n(X_i) \le x\} - x]\{\ddot{\ell}_{11}(x,y)A_n(x) + \ddot{\ell}_{12}B_n(y)\}\,\mathrm{d}C_n(x,y). \end{aligned}$$

If the previous equation holds for some $i \in \{1, \dots, n\}$, then it also holds for all z such that $F_n(X_i) - 1/n < z \le F_n(X_i)$. Therefore, for all $z \in (0,1]$,

$$\begin{aligned} 0 &= \int \dot{\ell}_1(x,y)\{I(z \le x) - x\}\,\mathrm{d}C_n(x,y) - a_n(z) \\ &\quad + \int \{I(z \le x) - x\}\{\ddot{\ell}_{11}(x,y)A_n(x) + \ddot{\ell}_{12}B_n(y)\}\,\mathrm{d}C_n(x,y). \end{aligned}$$

For $u \in (0,1]$, integrate the previous identity over $z \in (0,u]$, apply Fubini's theorem and use the fact that $\int_0^u \{I(z \le x) - x\}\,\mathrm{d}z = x \wedge u - xu$ to arrive at the first equation of (3.5). For $u = 0$, both sides of the first equation of (3.5) are automatically zero.

The proof of the second equation in (3.5) is entirely similar. □

References

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1998). *Efficient and adaptive estimation for semiparametric models* (2nd ed.). Springer.

Bickel, P., Y. Ritov, and J. Wellner (1991). Efficient estimation of linear functionals of a probability measure P with known marginal distributions. *Annals of Statistics 19*, 1316–1346.

Chen, X., Y. Fan, and V. Tsyrennikov (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association 101*, 1228–1240.

Genest, C., K. Ghoudi, and L.-P. Rivest (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika 82*, 543–552.

Genest, C. and B. Werker (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In C. Cuadras and A. Rodríguez-Lallena (Eds.), *Distributions with given marginals and statistical modelling*, Dordrecht, pp. 103–112. Kluwer Academic Publishers.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis 94*, 401–419.

Klaassen, C. and H. Putter (2005). Efficient estimation of banach parameters in semiparametric models. *Annals of Statistics 33*, 307–346.

Klaassen, C. and J. Wellner (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli 3*, 55–77.

Nelsen, R. (1999). *An introduction to copulas* (1st ed.). New York: Springer-Verlag.

Oakes, D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics 3*, 343–354.

Peng, H. and A. Schick (2002). On efficient estimation of linear functionals of a bivariate distribution with known marginals. *Statistics & Probability Letters 59*, 83–91.

Peng, H. and A. Schick (2004). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown, marginals: the minimum chi-square approach. *Statistics & Decisions 22*, 301–318.

Peng, H. and A. Schick (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown marginals: the least-squares approach. *Journal of Multivariate Analysis 95*, 385–409.

Shen, X. (1997). On methods of sieves and penalization. *Annals of Statistics 25*, 2555–2591.

Shih, J. and T. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics 51*, 1384–1399.

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.

Tsukahara, H. (2005). Semiparametric estimation in copula models. *The Canadian Journal of Statistics 33*, 357–375.

Van den Akker, R. (2007). *Integer-Valued Time Series*. Ph. D. thesis, Tilburg University.

Van der Vaart, A. (1991). Efficiency and Hadamard differentiability. *Scandinavian Journal of Statistics 18*, 63–75.

Van der Vaart, A. (2000). *Asymptotic Statistics* (1st ed.). Cambridge: Cambridge University Press.

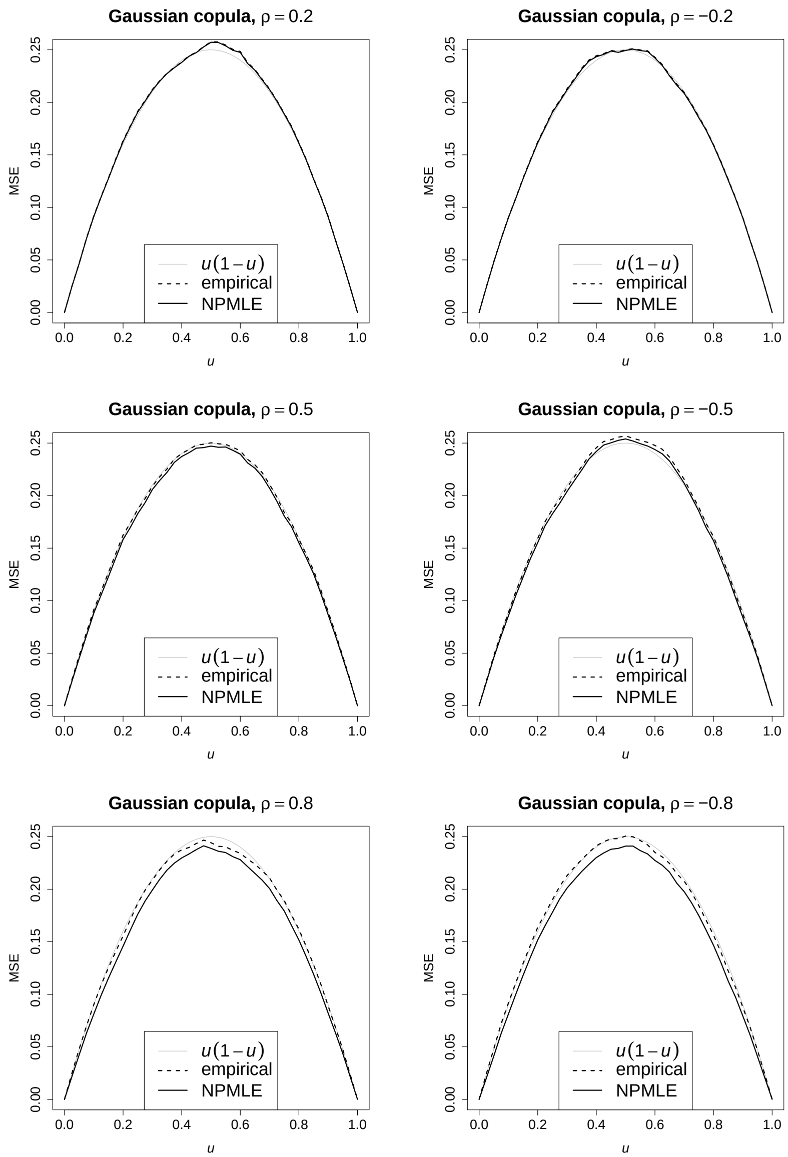

Figure 1: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 10\,000$ *samples of size* $n = 100$ *from the bivariate Gaussian copula with correlation parameter* $\rho \in \{\pm 0.2, \pm 0.8, \pm 0.5\}$ *together with the function* $u \mapsto u(1-u)$*. Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*

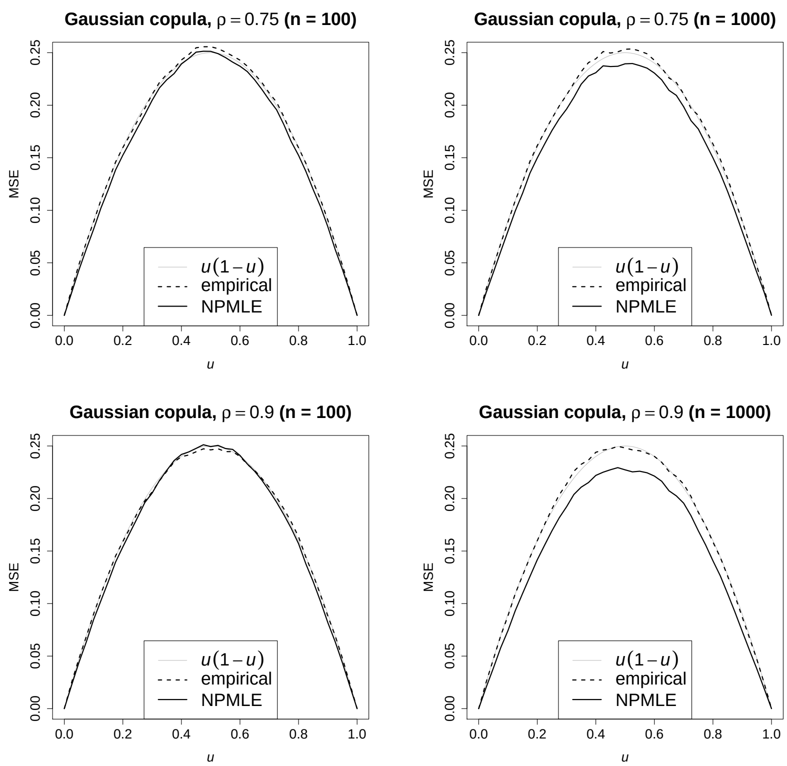

Figure 2: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 10\,000$ *samples of size* $n = 100$ *(left) or* $M = 5\,000$ *samples of size* $n = 1\,000$ *(right) from the bivariate Gaussian copula with correlation parameter* $\rho = 0.75$ *(top) and* $\rho = 0.90$ *(bottom) together with the function* $u \mapsto u(1-u)$. *Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*

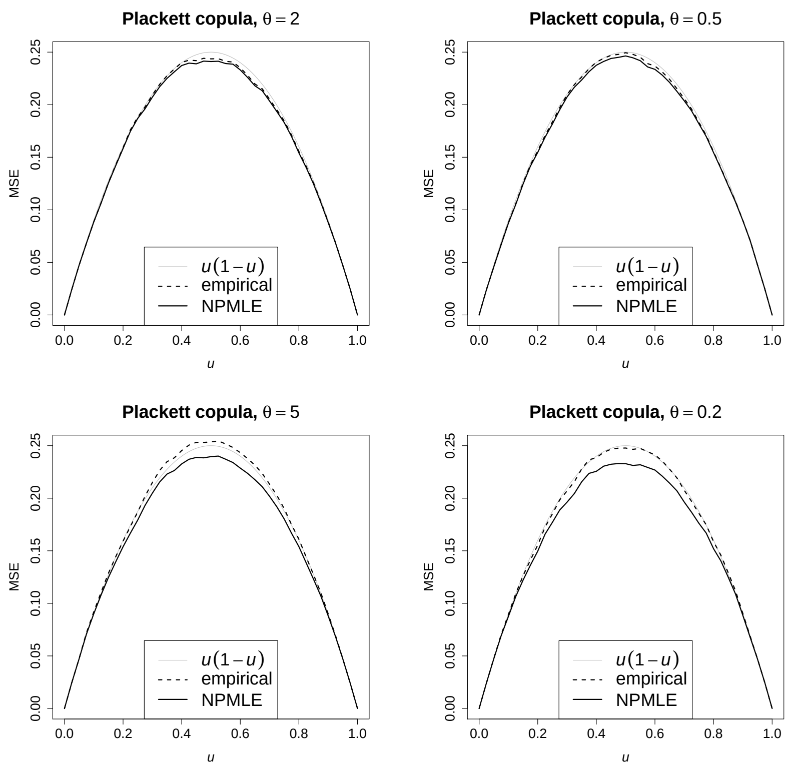

Figure 3: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 10\,000$ *samples of size* $n = 100$ *from the bivariate Plackett copula with odds-ratio parameter* $\theta \in \{2, 5\} \cup \{1/2, 1/5\}$ *together with the function* $u \mapsto u(1-u)$*. Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*

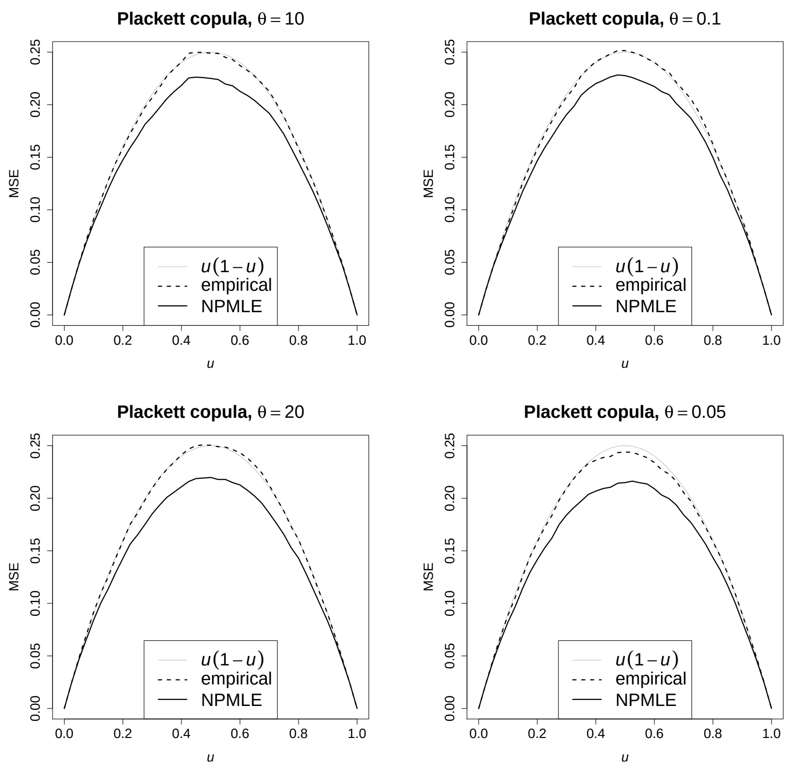

Figure 4: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 10\,000$ *samples of size* $n = 100$ *from the bivariate Plackett copula with odds-ratio parameter* $\theta \in \{10, 20\} \cup \{1/10, 1/20\}$ *together with the function* $u \mapsto u(1-u)$*. Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*

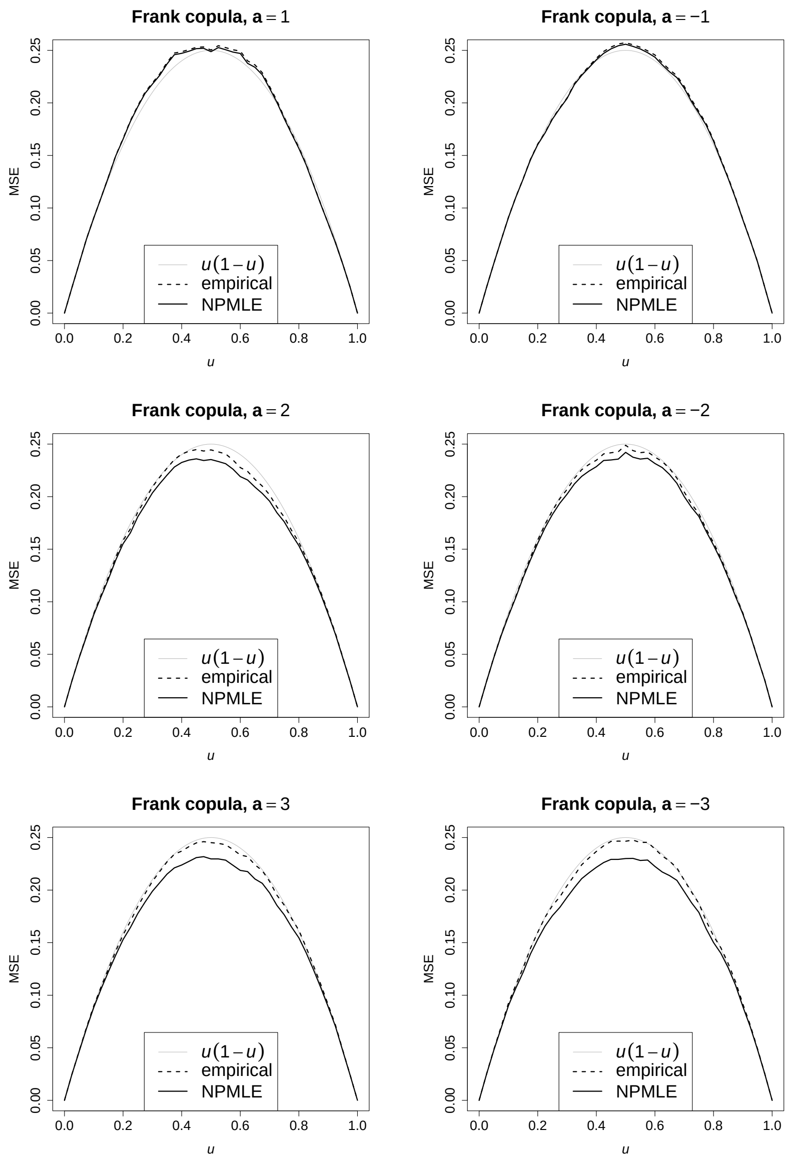

Figure 5: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 5\,000$ *samples of size* $n = 500$ *from the bivariate Frank copula with parameter* $a = -\log\theta \in \{\pm 1, \pm 2, \pm 3\}$ *together with the function* $u \mapsto u(1-u)$*. Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*

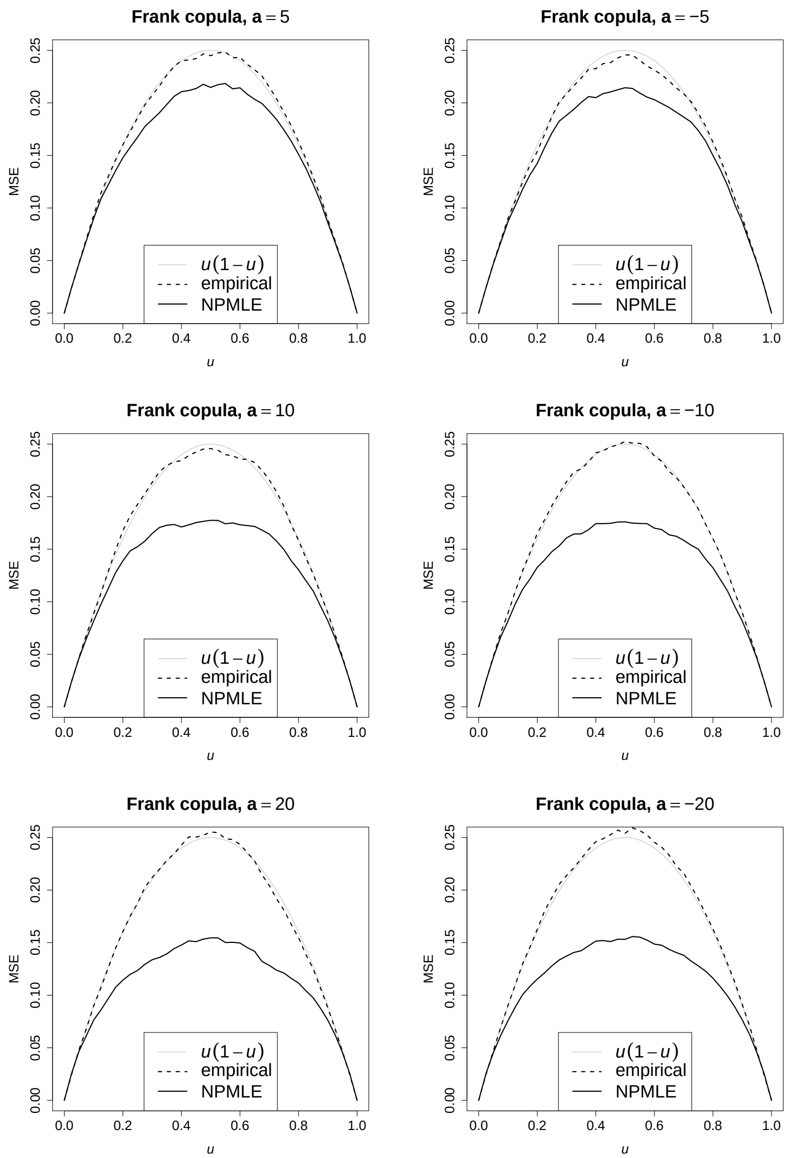

Figure 6: *Estimates of the normalized mean squared error* $\mathrm{E}[n\{\hat{F}_n(u) - u\}^2]$ *based on* $M = 5\,000$ *samples of size* $n = 100$ *from the bivariate Frank copula with parameter* $a = -\log\theta \in \{\pm 5, \pm 10, \pm 20\}$ *together with the function* $u \mapsto u(1-u)$*. Here* $\hat{F}_n$ *is the empirical distribution function (dashed) or the nonparametric maximum likelihood estimator if the copula is known (solid).*