Journal de la Société Française de Statistique
Vol. 154 No. 1 (2013)

Copula parameter estimation using Blomqvist's beta

Titre: L'emploi du bêta de Blomqvist pour l'estimation du paramètre d'une copule

Christian Genest[1], Alberto Carabarín-Aguirre[2] and Fanny Harvey[3]

Abstract: The authors consider the inversion of Blomqvist's beta as a method-of-moments estimator for a real-valued dependence parameter in a bivariate copula model. This estimator results from solving the equation $\beta = \beta_n$ for the copula parameter, where β_n is a rank-based estimate of β derived from a random sample of size n. Small- and large-sample comparisons are made between this estimator and an analogous estimator based on the inversion of Kendall's tau. While the results show that the latter is more efficient, the computation of β_n requires only $O(n)$ operations, as opposed to $O(n^2)$ for the estimation of Kendall's tau. Thus for large n, the inversion of β quickly leads to an unbiased estimator and a good starting value for canonical likelihood maximization.

Résumé : Les auteurs s'intéressent à l'inversion du beta de Blomqvist comme estimateur des moments du paramètre de dépendance réel d'un modèle de copule bivarié. Cet estimateur est obtenu en isolant le paramètre de la copule dans l'équation $\beta = \beta_n$, où β_n est un estimateur de rangs de β déduit d'un échantillon aléatoire de taille n. La performance asymptotique et à taille finie de cet estimateur est comparée à celle d'un estimateur analogue obtenu en inversant le tau de Kendall. Bien que les résultats montrent que ce dernier est plus efficace, le calcul de β_n ne requiert que $O(n)$ opérations et non $O(n^2)$ comme pour l'estimation du tau de Kendall. Pour n grand, l'inversion de β fournit donc rapidement un estimateur sans biais et une bonne valeur initiale pour la maximisation de la vraisemblance canonique.

Keywords: Archimedean copula, Blomqvist's beta, Kendall's tau, meta-elliptical copula, extreme-value copula
Mots-clés : copule Archimédienne, beta de Blomqvist, tau de Kendall, copule meta-elliptique, copule de valeurs extrêmes
AMS 2000 subject classifications: 62H12, 62G05, 62H20

[1] Department of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal (Québec) Canada H3A 0B9
E-mail: cgenest@math.mcgill.ca
[2] Department of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal (Québec) Canada H3A 0B9
E-mail: carabarin@math.mcgill.ca
[3] Département de mathématiques et de statistique, Université Laval, 1045, avenue de la Médecine, Québec (Québec) Canada G1V 0A6
E-mail: fanny.harvey.1@ulaval.ca

1. Introduction

Let X and Y be continuous random variables with joint cumulative distribution function H and margins F and G, respectively. Sklar's Theorem [Sklar, 1959] guarantees the existence of a unique function C such that, for all $x, y \in \mathbb{R}$,

$$H(x,y) = C\{F(x), G(y)\}. \tag{1}$$

The function C, called the copula of (X,Y), is the joint cumulative distribution function of the pair $(U,V) = (F(X), G(Y))$, whose components are distributed uniformly on the interval $[0,1]$.

Sklar's representation of H provides a useful way to model the joint behaviour of X and Y by choosing C, F and G from appropriate parametric families. Typically, it is assumed that $C \in (C_\theta)$, where the parameter θ is real-valued. The construction of a copula model involves several steps. For a general blueprint of the entire process, the reader is referred to [Genest and Favre, 2007]. This paper is limited to a single aspect: the estimation of θ, the parameter indexing the copula family.

The dependence between X and Y is characterized entirely by the copula C. Given that copulas are invariant by monotone increasing transformations of the margins, it is reasonable to focus on ranked-based estimators of θ. This is because the pairs of ranks in a sample form a maximally invariant set with respect to strictly increasing transformations of the margins.

Canonical maximum likelihood (CML) estimation is a recognized standard for rank-based estimation of copula parameters. Originally suggested in [Oakes, 1994], it was formalized in [Genest et al., 1995] and subsequently expanded upon in [Tsukahara, 2005]. It involves maximizing a modified version of the log-likelihood in which the unknown marginal distributions are replaced by their empirical counterparts. Under weak regularity conditions, the resulting estimator is consistent and asymptotically Gaussian; it is even semiparametrically asymptotically efficient in special cases; see, e.g., [Genest and Werker, 2002].

In practice, however, the CML method can be computationally intensive. In addition, its application is limited to cases where the copula C_θ has a density with respect to Lebesgue measure. For this reason, nonparametric analogues of the method of moments are often used. The most common choice is the inversion of Kendall's tau. Given a random sample of size n from H, this estimation technique involves solving the equation $\tau(C_\theta) = \tau_n$ for θ, where $\tau(C_\theta)$ and τ_n represent the population value and sample estimate of Kendall's tau, respectively. Inversion of Spearman's rho is another option. These two estimators are also known to be consistent and asymptotically Gaussian under weak regularity conditions; see, e.g., [Borkowf, 2002, Hoeffding, 1947, Hoeffding, 1948].

The purpose of the present paper is to investigate an even simpler rank-based estimator for the dependence parameter θ based on the method of moments. The idea is to invert the medial correlation coefficient, also known as Blomqvist's beta [Blomqvist, 1950]. Broadly speaking, the empirical version β_n of Blomqvist's beta is a suitably scaled version of the proportion of points whose components are either both smaller, or both larger, than their respective sample medians. The computation of β_n thus involves only $O(n)$ operations, as opposed to $O(n^2)$ for the empirical versions of Kendall's tau and Spearman's rho. In addition, the population version of Blomqvist's beta, $\beta(C_\theta) = -1 + 4C_\theta(1/2, 1/2)$, is available in closed form for many popular copula families. As a result, its inversion often leads to a simple, easily computed, explicit estimator for θ. These advantages become significant when dealing with large sample sizes or in situations where real-valued dependence parameters must be computed for many pairs of variables.

Consider, for instance, the problem of fitting a multivariate meta-elliptical copula to $d \geq 2$ financial risks. The copula family then involves $d(d-1)/2$ parameters which, initially at least, are often estimated by inverting Kendall's tau for each pair of variables; see, e.g., [Genest et al., 2011]. When d is large, as is often the case in risk management, inversion of Blomqvist's beta would entail substantial savings in computing time. The resulting estimates may also be sufficient as initial values for implementation of the CML method. To assess the validity of the latter statement, this paper compares the relative performance of estimators of θ based on the inversion of Kendall's tau and Blomqvist's beta. While both approaches could be implemented in arbitrary dimension by resorting to d-variate extensions of Kendall's tau [Genest et al., 2011] and Blomqvist's beta [Schmid and Schmidt, 2007], the comparison is limited here to the case $d = 2$ for simplicity.

To start things off, Section 2 gives a brief overview of the CML and standard rank-based moment estimators of a real-valued copula parameter θ. Inversion of Blomqvist's beta as a technique for the estimation of θ is then described in Section 3. Illustrations of this approach in concrete classes of copulas can be found in Section 4, along with direct comparisons to the estimator resulting from the inversion of Kendall's tau. A simulation study intended to shed more light into the effectiveness of the new estimator was carried out, and its results are reported in Section 5. Some final thoughts and comments are presented in Section 6.

2. Estimating θ: Canonical maximum likelihood and moments-based procedures

Let $(X_1,Y_1),\dots,(X_n,Y_n)$ be a random sample from a pair (X,Y) of continuous random variables with joint distribution H, copula C, and margins F and G, respectively. Further let $(R_1,S_1),\dots,$ (R_n,S_n) denote the pairs of ranks associated with the sample: that is, R_i represents the rank of X_i among $X_1,\dots,X_n$ and S_i is the rank of Y_i among $Y_1,\dots,Y_n$. Since both X and Y are assumed to be continuous, ties occur with probability zero and therefore the ranks are well-defined.

As stated in the Introduction, this paper is aimed squarely at the estimation of the dependence parameter in a copula model. In later sections, the effectiveness of the inversion of Blomqvist's beta as a method of estimation will be examined by comparing it to that of the most widely-used procedures over several copula models. For completeness, a brief review seems in order; details will be left out and references will be given instead.

Assuming that $C \in (C_\theta)$, the most popular method for estimating the dependence parameter $\theta \in \mathbb{R}$ is based on the principle of maximum likelihood. Suppose that the copula C_θ associated with (X,Y) has a density c_θ defined, for all $u,v \in (0,1)$, by

$$c_\theta(u,v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u,v).$$

When the margins F and G are known, the log-likelihood function for θ is given by

$$\ell(\theta) = \sum_{i=1}^{n} \ln\left[c_\theta\{F(X_i),G(Y_i)\}\right].$$

However, given that the margins are rarely known in practice, Oakes [Oakes, 1994] suggested to simply replace them with their empirical versions. Genest et al. [Genest et al., 1995] explored and suitably modified this idea, employing the rescaled versions of the empirical margins defined, for all $x, y \in \mathbb{R}$, by

$$F_n(x) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(X_i \le x), \quad G_n(y) = \frac{1}{n+1}\sum_{i=1}^{n} \mathbf{1}(Y_i \le y),$$

where $\mathbf{1}(A)$ denotes the indicator function of the event A. It is easy to check that $F_n(X_i) = R_i/(n+1)$ and $G_n(Y_i) = S_i/(n+1)$ for all $i \in \{1,\dots,n\}$, so the resulting log pseudo-likelihood function is

$$\ell^*(\theta) = \sum_{i=1}^{n} \ln\left\{c_\theta\left(\frac{R_i}{n+1}, \frac{S_i}{n+1}\right)\right\}.$$

The canonical maximum likelihood estimator is the value θ_n that maximizes ℓ^*. It was shown in [Genest et al., 1995] that under mild regularity conditions, θ_n is consistent and asymptotically Gaussian.

As explained in the Introduction, however, there are instances where the copula C_θ does not have a density, or where the maximization of the log pseudo-likelihood function is computationally intensive. In those cases, moment-based methods can provide a "quick and dirty" way of estimating θ, or at least produce an appropriate starting value for pseudo-likelihood estimation. Among them, the inversion of Kendall's tau is the most popular, mainly because its form is often explicit.

Any two pairs (X_i,Y_i), (X_j,Y_j) are said to be concordant whenever $(X_i - X_j)(Y_i - Y_j) > 0$; conversely, they are discordant if $(X_i - X_j)(Y_i - Y_j) < 0$. Kendall's tau, a famous measure of dependence based on the notion of concordance, is estimated empirically by

$$\tau_n = \frac{4}{n(n-1)} P_n - 1,$$

where P_n stands for the number of concordant pairs in the sample. Its population version can be expressed entirely in terms of the copula C_θ as

$$\tau(X,Y) = \tau(C_\theta) = -1 + 4\int_0^1\int_0^1 C_\theta(u,v)\,\mathrm{d}C_\theta(u,v). \tag{2}$$

In analogy to the method of moments in classical statistics, the inversion of Kendall's tau consists of solving the equation $\tau(C_\theta) = \tau_n$ for θ. The resulting value, $\theta_{\tau,n}$, provides an estimate of θ.

The asymptotic behaviour of $\theta_{\tau,n}$ can be derived through the properties of Kendall's tau, many of which were studied by Hoeffding [Hoeffding, 1947] using the theory of U-statistics. It is noted in [Genest et al., 2011] that as $n \to \infty$, $n^{1/2}(\tau_n - \tau)$ converges weakly to a centered Gaussian random variable with variance

$$\sigma^2_{C,\tau} = 16\,\mathrm{var}\{C(U,V) + \bar{C}(U,V)\}, \tag{3}$$

where (U,V) is a random vector with distribution C and survival function $\bar{C}$ defined, for all $u, v \in [0,1]$, by $\bar{C}(u,v) = 1 - u - v + C(u,v)$. By considering the transformation $g_\tau : \tau \to \theta$, one can write $\theta_{\tau,n} = g_\tau(\tau_n)$. An application of the Delta Method shows that as long as $g'_\tau(\tau)$ exists and is non-zero, $\theta_{\tau,n}$ is asymptotically Gaussian with mean zero and variance

$$\{g'_\tau(\tau)\}^2 \sigma^2_{C,\tau}. \tag{4}$$

The same principles outlined above can be used with Spearman's rho in place of Kendall's tau. The resulting estimator is also well-behaved asymptotically under mild regularity conditions; it is, however, far less popular than $\theta_{\tau,n}$ because copulas with a closed form population version of Spearman's rho are few and far between. For this reason, details are omitted.

3. Inversion of Blomqvist's beta

This paper considers an alternative rank-based estimator of the dependence parameter θ inspired by the method of moments. The idea is to invert the medial correlation coefficient, also known as Blomqvist's beta [Blomqvist, 1950]. This dependence measure is computed from a 2×2 contingency table constructed from the data.

Let $\tilde{X}_n$ and $\tilde{Y}_n$ be the medians of the samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. To gather information about the dependence between X and Y, Blomqvist suggested dividing the x–y plane into four regions by drawing the lines $x = \tilde{X}_n$ and $y = \tilde{Y}_n$ and comparing the following quantities:

n_1: the number of points lying in either the lower left quadrant or the upper right quadrant;
n_2: the number of points in either the upper left quadrant or the lower right quadrant.

The definition of β_n, which came to be called Blomqvist's beta, is given by

$$\beta_n = \frac{n_1 - n_2}{n_1 + n_2} = -1 + 2\frac{n_1}{n_1 + n_2}. \tag{5}$$

If n is even, then no sample point falls on either of the lines $x = \tilde{X}_n$ and $y = \tilde{Y}_n$, and it follows that both n_1 and n_2 are even. If n is odd, however, then either one or two sample points lie on the lines defined by the sample medians. In the case of a single point lying on a median, Blomqvist [Blomqvist, 1950] proposed not to count the point altogether. In the second case, one point has to fall on each line: one of them is assigned to the quadrant touched by the two points, and the other is not counted. This allows both n_1 and n_2 to remain even.

As can be deduced from Equation (5), β_n is the difference between the proportion of sample points having both components either smaller or greater than their respective medians, and the proportion of the other sample points. The population analogue of β_n is

$$\beta = \Pr\{(X - \tilde{x})(Y - \tilde{y}) > 0\} - \Pr\{(X - \tilde{x})(Y - \tilde{y}) < 0\},$$

where $\tilde{x}$ and $\tilde{y}$ denote the population medians of X and Y, respectively. Using the simple facts that

$$\Pr\{(X - \tilde{x})(Y - \tilde{y}) > 0\} = \Pr(X - \tilde{x} > 0, Y - \tilde{y} > 0) + \Pr(X - \tilde{x} < 0, Y - \tilde{y} < 0)$$

and

$$\Pr(X > \tilde{x}, Y > \tilde{y}) = \Pr(X < \tilde{x}, Y < \tilde{y})$$

together with Equation (1), it is straightforward to show that

$$\beta = -1 + 4C\{F(\tilde{x}), G(\tilde{y})\} = -1 + 4C\left(\frac{1}{2}, \frac{1}{2}\right). \tag{6}$$

As β is only a function of C, it is possible to write it in terms of θ whenever $C \in (C_\theta)$.

Equation (6) also leads to an alternative definition for β_n. It has been known since the work of Rüschendorf [Rüschendorf, 1976] that C can be estimated consistently by the so-called empirical copula. The latter is usually defined, for each $u, v \in [0,1]$, by

$$C_n(u,v) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(\frac{R_i}{n+1} \le u, \frac{S_i}{n+1} \le v \right).$$

As shown by Segers [Segers, 2012], C_n converges weakly to C as $n \to \infty$ whenever C is regular in the sense of [Genest et al., 2012]: namely, the partial derivatives $C_1(u,v) = \partial C(u,v)/\partial u$ and $C_2(u,v) = \partial C(u,v)/\partial v$ must exist everywhere on $[0,1]^2$, with the convention that one-sided derivatives are used at the boundary points; moreover, C_1 and C_2 must be continuous on $(0,1) \times [0,1]$ and $[0,1] \times (0,1)$, respectively.

Replacing C with C_n in Equation (6) gives

$$\beta_n^* = -1 + 4C_n\left(\frac{1}{2}, \frac{1}{2}\right). \tag{7}$$

If n is even, then β_n and β_n^* are equivalent. When n is odd, however, (5) and (7) can never coincide. The reason is that, under this scenario, one or two of the sample points lie on the lines defined by the sample medians. The original method proposed by Blomqvist then removes a sample point from the final tally, while the formula based on the empirical copula does not. In any case, they are asymptotically equivalent, as stated next and proved in Appendix A.

Lemma 1. *Let β_n and β_n^* be defined as in (5) and (7), respectively. If n is even, then $\beta_n = \beta_n^*$. If n is odd, then*

$$\beta_n - \beta_n^* = \begin{cases} \dfrac{4}{n-1} C_n\left(\dfrac{1}{2}, \dfrac{1}{2}\right), & \text{if there exist } i, j \in \{1, \ldots, n\} \text{ such that} \\ & X_i = \tilde{X}_n,\ Y_i > \tilde{Y}_n,\ X_j > \tilde{X}_n, \text{ and } Y_j = \tilde{Y}_n; \\ \dfrac{4}{n-1} \left\{ C_n\left(\dfrac{1}{2}, \dfrac{1}{2}\right) - 1 \right\}, & \text{otherwise.} \end{cases}$$

The asymptotic behaviour of β_n was studied by Schmid & Schmidt [Schmid and Schmidt, 2007] in a multivariate context. In the bivariate case, their findings show that, as $n \to \infty$,

$$\sqrt{n}(\beta_n - \beta) \rightsquigarrow \mathscr{N}(0, \sigma_{C,\beta}^2)$$

as long as the copula C is regular. The asymptotic variance $\sigma_{C,\beta}^2$ can then be written as

$$\begin{aligned} \sigma_{C,\beta}^2 &= 16 \Bigg[C\left(\frac{1}{2}, \frac{1}{2}\right) \left\{ 1 - C\left(\frac{1}{2}, \frac{1}{2}\right) \right\} + \frac{1}{4} \left\{ C_1\left(\frac{1}{2}, \frac{1}{2}\right) - C_2\left(\frac{1}{2}, \frac{1}{2}\right) \right\}^2 \\ &+ C\left(\frac{1}{2}, \frac{1}{2}\right) \left\{ -C_1\left(\frac{1}{2}, \frac{1}{2}\right) - C_2\left(\frac{1}{2}, \frac{1}{2}\right) + 2C_1\left(\frac{1}{2}, \frac{1}{2}\right) C_2\left(\frac{1}{2}, \frac{1}{2}\right) \right\} \Bigg]. \end{aligned} \tag{8}$$

This expression is derived in Appendix B.

Now suppose once again that $C \in (C_\theta)$ and set $\beta(C_\theta) = \beta_n$. The value $\theta_{\beta,n}$ that solves the equation yields an estimate for θ. This estimator can be expressed in the form $\theta_{\beta,n} = g_\beta(\beta_n)$ for an appropriate choice of map $g_\beta : \beta \to \theta$. Assume that $g'_\beta(\beta)$ exists and is non-zero. Analogously to $\theta_{\tau,n}$ in the previous section, the Delta Method then ensures that $\theta_{\beta,n}$ is consistent and asymptotically Gaussian with mean zero and variance

$$\{g'_\beta(\beta)\}^2 \sigma^2_{C,\beta}. \tag{9}$$

This result is formally recorded below.

Proposition 1. *If C_θ is a regular copula with $\beta(C_\theta) = \beta$, then $n^{1/2}(\theta_{\beta,n} - \theta)$ converges in distribution, as $n \to \infty$, to a Gaussian random variable with mean zero and variance $\{g'_\beta(\beta)\}^2 \sigma^2_{C,\beta}$.*

4. Application to different classes of copulas

How does $\theta_{\beta,n}$ perform as an estimator? A simple way to answer this question is to compare its behaviour to that of $\theta_{\tau,n}$, which is already known [Kojadinovic and Yan, 2010a]. For most classical families of copulas, the asymptotic variance of both estimators can be determined, either explicitly or by numeric integration. A few examples of such calculations are presented below.

4.1. Farlie–Gumbel–Morgenstern family

The bivariate Farlie–Gumbel–Morgenstern (FGM) family of copulas with parameter $\theta \in [-1, 1]$ is defined, for all $u, v \in (0, 1)$, by

$$C_\theta(u, v) = uv + \theta(1-u)(1-v)uv.$$

The density of C_θ is given by $c_\theta(u, v) = 1 + \theta(1-2u)(1-2v)$ and so, by Equation (2), a simple calculation yields $\tau(C_\theta) = 2\theta/9$. It is, of course, even easier to use (6) to get $\beta(C_\theta) = \theta/4$. These two results show the limited range of dependence that can be modelled with this family, as the values for τ are restricted to $[-2/9, 2/9]$, while $\beta \in [-1/4, 1/4]$. The resulting estimators for θ obtained by inversion of Kendall's tau and Blomqvist's beta are, respectively,

$$\theta_{\tau,n} = 9\tau_n/2, \quad \theta_{\beta,n} = 4\beta_n.$$

In order to compute their asymptotic variances, one must find $\sigma^2_{C,\tau}$ and $\sigma^2_{C,\beta}$. By Equation (3),

$$\begin{aligned} \sigma^2_{C,\tau} &= 16 \left[\int_0^1 \int_0^1 \{C_\theta(u,v) + \bar{C}_\theta(u,v)\}^2 \mathrm{d}C_\theta(u,v) \right. \\ &\quad \left. - \left[\int_0^1 \int_0^1 \{C_\theta(u,v) + \bar{C}_\theta(u,v)\} \, \mathrm{d}C_\theta(u,v) \right]^2 \right]. \end{aligned}$$

A simple calculation shows that $C_\theta(u,v) + \bar{C}_\theta(u,v) = 2uv + 2\theta uv(1-u)(1-v) + 1 - u - v$, and after some integration, one finds

$$\sigma^2_{C,\tau} = \frac{4}{9} - \frac{184}{2025}\theta^2.$$

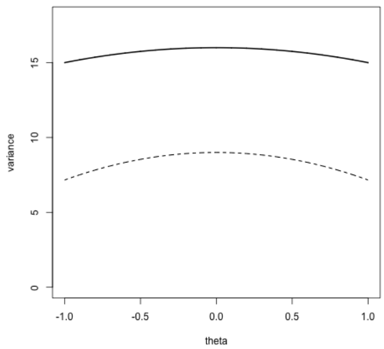

FIGURE 1. *Asymptotic variance of $\theta_{\tau,n}$ (dotted line) and $\theta_{\beta,n}$ (solid line) for the FGM copula.*

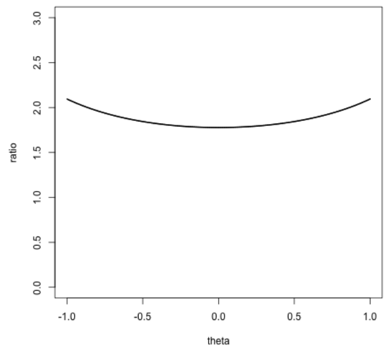

FIGURE 2. *Ratio between the asymptotic variances of $\theta_{\beta,n}$ and $\theta_{\tau,n}$ for the FGM copula.*

Furthermore, given that $C_\theta(1/2,1/2) = 1/4+\theta/16$ and $C_{\theta,1}(1/2,1/2) = C_{\theta,2}(1/2,1/2) = 1/2$, Equation (8) yields

$$\sigma^2_{C,\beta} = 1-\frac{\theta^2}{16}.$$

Given that $g_\tau(\tau) = 9\tau/2$, the asymptotic variance of $\theta_{\tau,n}$ is found to be

$$\{g'_\tau(\tau)\}^2\sigma^2_{C,\tau} = \left(\frac{9}{2}\right)^2\left(\frac{4}{9}-\frac{184}{2025}\theta^2\right) = 9-\frac{46}{25}\theta^2.$$

Similarly, using the fact that $g_\beta(\beta) = 4\beta$, one finds that the asymptotic variance of $\theta_{\beta,n}$ is

$$\{g'_\beta(\beta)\}^2\sigma^2_{C,\beta} = 4^2\left(1-\frac{\theta^2}{16}\right) = 16-\theta^2.$$

Figure 1 shows a plot of these variances. Clearly, the asymptotic variance of $\theta_{\beta,n}$ is larger than that of $\theta_{\tau,n}$ for every admissible value of θ. The ratio between $\theta_{\beta,n}$ and $\theta_{\tau,n}$ is smallest at independence (i.e., at $\theta = 0$), where it stands at $16/9 \approx 1.78$. See Figure 2 for an illustration.

4.2. *Meta-elliptical copulas*

A random pair $\mathbf{X}$ is said to have an elliptical distribution with mean zero and positive-definite dispersion matrix $\Sigma = (\sigma_{ij})$ if it can be represented as

$$\mathbf{X} = RA\mathbf{U},$$

where R is a strictly positive random variable, $\mathbf{U}$ is a random pair independent of R that is uniformly distributed on the unit circle, and A is a 2×2 constant matrix such that $AA^\top = \Sigma$. In particular, if R admits a density, then it is proportional to $q(\mathbf{X}^\top\Sigma^{-1}\mathbf{X})$, where q is a scale function completely determined by R. The associated copula is then called meta-elliptical with

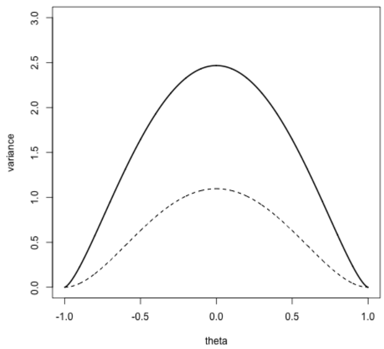

FIGURE 3. *Asymptotic variance of $\theta_{\tau,n}$ (dotted line) and $\theta_{\beta,n}$ (solid line) for the Gaussian copula.*

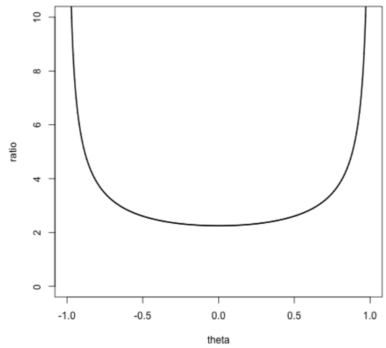

FIGURE 4. *Ratio between the asymptotic variances of $\theta_{\beta,n}$ and $\theta_{\tau,n}$ for the Gaussian copula.*

radial variable R and correlation matrix $\Upsilon = (\theta_{ij})$, where $\theta_{ij} = \sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{1/2}$ for $i,j \in \{1,2\}$; see [Abdous et al., 2005, Fang et al., 2005] for more details.

Hult & Lindskog [Hult and Lindskog, 2002] showed that the value of Kendall's tau is the same for all bivariate meta-elliptical copulas with correlation θ; in other words, it does not depend on q. As pointed out in [Schmid and Schmidt, 2007], the same is true for Blomqvist's beta. In fact, the identity

$$\tau(C_\theta) = \beta(C_\theta) = \frac{2}{\pi}\arcsin(\theta)$$

is valid for any bivariate meta-elliptical copula C_θ with correlation θ. In light of these results, the inversion of τ and β leads to the estimators

$$\theta_{\tau,n} = \sin\left(\frac{\pi}{2}\tau_n\right), \quad \theta_{\beta,n} = \sin\left(\frac{\pi}{2}\beta_n\right).$$

Schmid & Schmidt [Schmid and Schmidt, 2007] also report a concise expression for the asymptotic variance of β that remains invariant in the class of meta-elliptical copulas, namely

$$\sigma^2_{C,\beta} = 1 - \frac{4\arcsin^2(\theta)}{\pi^2}.$$

Unfortunately, a closed form expression for the asymptotic variance of τ is unavailable, except in special cases. When C_θ is the bivariate Gaussian copula, the form of the scale function q does make an explicit computation possible; a direct comparison can then be made.

Example 1. *The Gaussian copula with parameter $\theta \in (-1,1)$ is defined, for all $u,v \in (0,1)$, by*

$$C_\theta(u,v) = \Phi_\theta\{\Phi^{-1}(u), \Phi^{-1}(v)\},$$

where Φ_θ represents the cumulative distribution function of the bivariate standard Normal law with correlation θ and Φ stands for its univariate margin. Given that $\tau(C_\theta) = \beta(C_\theta) =$

$2\arcsin(\theta)/\pi$ for all meta-elliptical copulas, both τ and β can range over the full interval $[-1,1]$. Genest et al. [Genest et al., 2011] showed that

$$\sigma^2_{C,\tau} = \frac{4}{9} - \frac{16}{\pi^2}\arcsin^2\left(\frac{\theta}{2}\right).$$

Using the fact that $g'_\tau(\tau) = g'_\beta(\beta) = \pi\sqrt{1-\theta^2}/2$, Equations (4) and (9) yield

$$\begin{aligned} \{g'_\tau(\tau)\}^2\sigma^2_{C,\tau} &= \left(\frac{\pi}{2}\sqrt{1-\theta^2}\right)^2\left\{\frac{4}{9} - \frac{16}{\pi^2}\arcsin^2\left(\frac{\theta}{2}\right)\right\}, \\ \{g'_\beta(\beta)\}^2\sigma^2_{C,\beta} &= \left(\frac{\pi}{2}\sqrt{1-\theta^2}\right)^2\left\{1 - \frac{4\arcsin^2(\theta)}{\pi^2}\right\}. \end{aligned}$$

A comparison between these asymptotic variances produces similar results to those already reported for the FGM family. Here again, $\theta_{\tau,n}$ is more efficient than $\theta_{\beta,n}$ for all possible values of θ; see Figure 3. The ratio between the asymptotic variances of $\theta_{\beta,n}$ and $\theta_{\tau,n}$ at independence is $9/4 = 2.25$, and it only grows larger as θ approaches ± 1 as can be seen in Figure 4. Note that if $|\theta| = 1$, both variances are zero; this is a consequence of the fact that $\theta = \pm 1$ correspond to the Fréchet–Hoeffding bounds.

4.3. Archimedean copulas

A bivariate copula C_θ is said to be Archimedean [Genest and MacKay, 1986] if there exists a strictly decreasing, convex function $\phi_\theta : [0,1] \to \mathbb{R}_+$ such that $\phi_\theta(1) = 0$ and, for all $u, v \in (0,1]$,

$$C_\theta(u,v) = \psi_\theta\{\phi_\theta(u) + \phi_\theta(v)\}. \tag{10}$$

Here ψ_θ denotes the generalized inverse of ϕ_θ. The map ϕ_θ is called the generator of the copula. For convenience, it is assumed here that $\phi_\theta(t) \to \infty$ as $t \to 0$, in which case ψ_θ is the standard inverse and C_θ is absolutely continuous with respect to Lebesgue measure.

For copulas of the form (10), Blomqvist's beta is given by

$$\beta(C_\theta) = -1 + 4\psi_\theta\left\{2\phi_\theta\left(\frac{1}{2}\right)\right\}. \tag{11}$$

Obviously, an expression for $\sigma^2_{C,\beta}$ involves ψ, ϕ and their first derivatives, which exist almost everywhere. One has

$$C_1\left(\frac{1}{2},\frac{1}{2}\right) = C_2\left(\frac{1}{2},\frac{1}{2}\right) = \psi'_\theta\left\{2\phi_\theta\left(\frac{1}{2}\right)\right\} \times \phi'_\theta\left(\frac{1}{2}\right).$$

In some cases, it is not possible to invert β explicitly, as will be seen later; hence for those families direct computation of g'_β is out of the question. However, the exact asymptotic variance of the estimator could still be found by exploiting (11). Indeed, one might alternatively compute

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\beta = 4\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\psi_\theta\left\{2\phi_\theta\left(\frac{1}{2}\right)\right\} + 2\psi'_\theta\left\{2\phi_\theta\left(\frac{1}{2}\right)\right\} \times \frac{\mathrm{d}}{\mathrm{d}\theta}\phi_\theta\left(\frac{1}{2}\right)\right] \tag{12}$$

and then $g'_\beta(\beta) = (\mathrm{d}\beta/\mathrm{d}\theta)^{-1}$.

When C_θ is Archimedean, it is well known [Genest and MacKay, 1986] that

$$\tau(C_\theta) = 1 + 4\int_0^1 \frac{\phi_\theta(t)}{\phi'_\theta(t)}\,\mathrm{d}t.$$

Unfortunately, the general expression for the asymptotic variance of τ_n is often unwieldy, so any direct comparison to the estimator based on β can only be made on a case-by-case basis. To that effect, the following example will be examined.

Example 2. *The Clayton copula with parameter $\theta > 0$ is defined, for all $u, v \in (0,1)$, by*

$$C_\theta(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

The value of $\tau(C_\theta)$ for this family is well-known and $\beta(C_\theta)$ is easy to compute directly. The expressions are

$$\tau(C_\theta) = \frac{\theta}{\theta+2}, \quad \beta(C_\theta) = -1 + 4(2^{\theta+1} - 1)^{-1/\theta}.$$

Values of θ greater than zero lead to $\tau, \beta \in (0,1)$. It is also possible to take $\theta \in (-1,0)$, leading to negative dependence and values of τ and β in $(-1,0)$. The resulting estimators are then

$$\theta_{\tau,n} = \frac{2\tau_n}{1-\tau_n}, \quad \theta_{\beta,n} = g_\beta(\beta_n),$$

where g is an implicit function. The asymptotic variances of τ and β can be given explicitly, although the former involves Spearman's rho. Once again, these expressions have been presented in [Genest et al., 2011] and [Schmid and Schmidt, 2007], respectively. They are

$$\begin{aligned}
\sigma^2_{C,\tau} &= \frac{8}{3} \times \frac{-\theta^3 + \theta^3\rho - 7\theta^2 + 7\theta^2\rho - 6\theta + 16\theta\rho + 2 + 12\rho}{12 + 16\theta + 7\theta^2 + \theta^3}, \\
\sigma^2_{C,\beta} &= 16\{\eta(1-\eta) + \eta(-4\eta^{\theta+1}2^\theta + 8\eta^{2\theta+2}2^{2\theta})\},
\end{aligned}$$

where $\eta = (2^{\theta+1} - 1)^{-1/\theta}$. Note that ρ has to be calculated by numerical integration as it has no closed form expression for the Clayton copula.

It is easy to see that, for this family, the function g'_τ is given by $g'_\tau(\tau) = (\theta+2)^2/2$ while (12) can be used to obtain

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\beta = 4(2^{\theta+1} - 1)^{-1/\theta}\left\{\frac{\ln(2^{\theta+1} - 1)}{\theta^2} - \frac{2^{\theta+1}\ln(2)}{\theta(2^{\theta+1} - 1)}\right\}$$

and thus $g'_\beta(\beta) = \mathrm{d}\theta/\mathrm{d}\beta$. Consequently,

$$\begin{aligned}
\{g'_\tau(\tau)\}^2\sigma^2_{C,\tau} &= \frac{2(\theta+2)^4}{3} \times \frac{-\theta^3 + \theta^3\rho - 7\theta^2 + 7\theta^2\rho - 6\theta + 16\theta\rho + 2 + 12\rho}{12 + 16\theta + 7\theta^2 + \theta^3}, \\
\{g'_\beta(\beta)\}^2\sigma^2_{C,\beta} &= \{g'_\beta(\beta)\}^2 \times 16\{\eta(1-\eta) + \eta(-4\eta^{\theta+1}2^\theta + 8\eta^{2\theta+2}2^{2\theta})\}.
\end{aligned}$$

It can be shown that these asymptotic variances tend to infinity as $\theta \to \infty$, although much more quickly so for $\theta_{\beta,n}$ than for $\theta_{\tau,n}$. This becomes evident by looking at Figure 5.

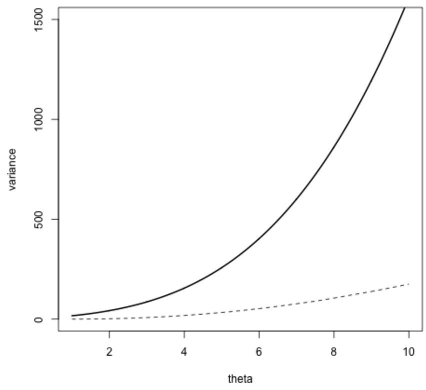

FIGURE 5. *Asymptotic variance of $\theta_{\tau,n}$ (dotted line) and $\theta_{\beta,n}$ (solid line) for the Clayton copula.*

4.4. Extreme-value copulas

A bivariate copula is said to be of the extreme-value type if there exists a convex map $A : [0,1] \to [1/2,1]$ such that, for all $u, v \in (0,1)$,

$$C_A(u,v) = \exp\left[\ln(uv) A\left\{\frac{\ln(v)}{\ln(uv)}\right\}\right]. \tag{13}$$

The map A is called the Pickands dependence function and in order for C_A to be a copula, it must satisfy $\max(t, 1-t) \le A(t) \le 1$ for all $t \in [0,1]$.

Ghoudi et al. [Ghoudi et al., 1998] showed that if A_θ is a parametric Pickands dependence function with real-valued parameter θ, the population value of Kendall's tau is given by

$$\tau(C_{A_\theta}) = \int_0^1 \frac{t(1-t)}{A_\theta} \, \mathrm{d}A'_\theta(t).$$

In a similar vein, it is easily found that

$$\beta(C_{A_\theta}) = -1 + \left(\frac{1}{4}\right)^{A_\theta(1/2)-1}. \tag{14}$$

The latter expression may not be overly helpful for computing the value of β for a given extreme-value copula; many times it is just as easy to apply Equation (6) directly. However, it proves handy when it comes to simplifying the calculation for $\sigma^2_{C,\beta}$, because it provides a starting point to express Equation (8) in terms of A. Indeed, for $k = 1, 2$,

$$C_\theta\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{1}{4}\right)^{A_\theta(1/2)}, \quad C_k\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{1}{4}\right)^{A_\theta(1/2)} \left\{2A_\theta\left(\frac{1}{2}\right) + (-1)^k A'_\theta\left(\frac{1}{2}\right)\right\},$$

and so the asymptotic variance can be written in terms of $A_\theta(t)$ and $A'_\theta(t) = \partial A_\theta(t)/\partial t$ as

$$\sigma^2_{C,\beta} = \left(\frac{1}{4}\right)^{A_\theta(1/2)-2} + \left(\frac{1}{4}\right)^{2A_\theta(1/2)-2}\left[\left\{A'_\theta\left(\frac{1}{2}\right)\right\}^2 - 4A_\theta\left(\frac{1}{2}\right) - 1\right] + \left(\frac{1}{4}\right)^{3A_\theta(1/2)-2}\left[\left\{8A_\theta\left(\frac{1}{2}\right)\right\}^2 - 2\left\{A'_\theta\left(\frac{1}{2}\right)\right\}^2\right]. \tag{15}$$

Whenever β cannot be inverted explicitly, one could proceed as previously described for the Archimedean family by computing

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\beta = \left(\frac{1}{4}\right)^{A_\theta\left(\frac{1}{2}\right)-1}\ln\left(\frac{1}{4}\right) \times \frac{\mathrm{d}}{\mathrm{d}\theta}A_\theta\left(\frac{1}{2}\right),$$

because $g'_\beta(\beta) = (\mathrm{d}\beta/\mathrm{d}\theta)^{-1}$, and so the asymptotic variance of $\theta_{\beta,n}$ can be found through (9).

Example 3. *Consider the Gumbel copula of type A with parameter $\theta \in [0,1]$. The Pickands dependence function for this family is given, for all $t \in [0,1]$, by $A_\theta(t) = \theta t^2 - \theta t + 1$. Upon substitution in (13), this choice of A_θ yields, for all $u, v \in (0,1)$,*

$$C_\theta(u,v) = uv\exp\left\{-\theta\,\frac{\ln(u)\ln(v)}{\ln(uv)}\right\}.$$

Computation of Blomqvist's beta for the type A Gumbel is equally straightforward with either (6) or (14), as it is easy to find that $\beta(\theta) = -1 + 2^{\theta/2}$. As in the case of the FGM copula, this model restricts the set of possible values for β, which is $[0, \sqrt{2}-1]$. The inversion of β for this family leads to

$$\theta_{\beta,n} = \frac{2\ln(\beta_n + 1)}{\ln 2},$$

which in turn means that

$$g'_\beta(\beta) = \frac{2}{2^{\theta/2}\ln 2}.$$

The calculation of the asymptotic variance only involves the values of A_θ and A'_θ at $1/2$. Given that $A_\theta(1/2) = 1 - \theta/4$ and $A'_\theta(1/2) = 0$, substitution into (15) yields

$$\sigma^2_{C,\beta} = 2^{2+\theta/2} + (\theta - 5) \times 2^\theta + (\theta - 4)^2 \times 2^{3\theta/2-3},$$

and so the asymptotic variance of $\theta_{\beta,n}$ is given by

$$\{g'_\beta(\beta)\}^2\sigma^2_{C,\beta} = \left(\frac{2}{\ln 2}\right)^2\{2^{2-\theta/2} + (\theta-5) + (\theta-4)^2 \times 2^{\theta/2-3}\}.$$

The inversion of τ is more challenging. From [Ghoudi et al., 1998]*, Kendall's tau for this family is given by*

$$\tau(\theta) = -2 + \frac{8}{\sqrt{\theta(4-\theta)}}\arctan\left(\sqrt{\frac{\theta}{4-\theta}}\right),$$

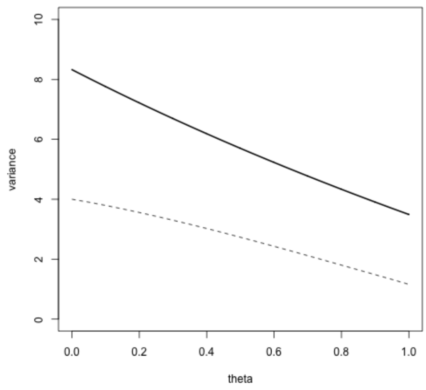

FIGURE 6. *Asymptotic variance of $\theta_{\tau,n}$ (dotted line) and $\theta_{\beta,n}$ (solid line) for the type A Gumbel copula.*

and hence $\tau \in [0, 0.418]$. Unfortunately, it is clear that solving for θ is not viable, while (3) does not produce an explicit expression. The first issue can be solved by differentiating $\tau(\theta)$ directly with respect to θ as done in earlier sections; after some manipulation, this process results in

$$g'_\tau(\tau) = \frac{\theta^2(4-\theta)^2}{4\left\{\theta(4-\theta)-(4-2\theta)\sqrt{\theta(4-\theta)}\arctan\left(\sqrt{\dfrac{\theta}{4-\theta}}\right)\right\}}.$$

Computation of $\sigma^2_{C,\tau}$ has to be done numerically, which ultimately dooms any attempt to write down a closed form expression for the asymptotic variance of $\theta_{\tau,n}$. To contrast it with the asymptotic variance of $\theta_{\beta,n}$, a graph is provided in Figure 6. Once again, the difference between the two estimators is fairly noticeable.

5. Simulation study

As there are few families of copulas for which the asymptotic variances of $\theta_{\tau,n}$ and $\theta_{\beta,n}$ can be expressed in closed form, a simulation study was conducted to compare the efficiency of these estimators. As stated in the Introduction and later confirmed through examples in Section 4, it is to be expected that $\theta_{\tau,n}$ will outperform $\theta_{\beta,n}$.

In order to facilitate comparisons between $\theta_{\beta,n}$ and methods currently used, the same experimental design was used as in the Monte Carlo study of Kojadinovic & Yan [Kojadinovic and Yan, 2010a], where the estimators based on the inversion of Kendall's tau and Spearman's rho were considered along with the canonical maximum likelihood method. The same six copula families were also considered, namely the Clayton, Frank, Gaussian, Gumbel–Hougaard, Plackett, and Student's t with 4 degrees of freedom. For each family, the dependence was set to 5 different levels, corresponding to $\tau = 0.1, 0.2, 0.4, 0.6$, and 0.8. Sample sizes were set to $n = 50, 100, 200$, and 400; 1000 repetitions were generated for each combination of copula family, dependence level and sample size.

Table 1 presents a basic summary of the results of the simulation corresponding to $n = 50$. Listed alongside the level of dependence and copula family is the true value of θ for each combination, the mean and standard deviation of $\theta_{\tau,n}$ and $\theta_{\beta,n}$, as well as the percentage relative bias (PRB) of each. Table 2 provides the same summary corresponding to $n = 200$. A direct comparison to the CML estimator was not included, given that the results in [Kojadinovic and Yan, 2010a] support the conclusion that it outperforms the estimator based on Kendall's tau in overall terms.

It can be seen from Tables 1 and 2 that $\theta_{\tau,n}$ is consistently less biased than $\theta_{\beta,n}$. The difference is especially marked for smaller samples: although $\theta_{\tau,n}$ remains superior as n increases, $\theta_{\beta,n}$ does not perform too badly, as can be gathered from Table 2. Neither estimator appears to do well for the Plackett family, a phenomenon already documented in [Kojadinovic and Yan, 2010a], even for the CML estimator. However, this problem is far more accute for $\theta_{\beta,n}$: in the case of strong dependence, the estimator loses all semblance of effectiveness in small samples.

Note that there are no results reported for the Clayton, Frank, and Gumbel–Hougaard copulas

TABLE 1. *Simulation summary for $n = 50$: The mean of the estimates, the standard deviation and the percentage relative bias is reported both for $\theta_{\tau,n}$ and $\theta_{\beta,n}$; C stands for Clayton, G for Gumbel–Hougaard, F for Frank, N for Gaussian, t for Student's t, and P for Plackett. An asterisk (*) means a figure greater than 10,000.*

τ	Copula	θ	μ_τ	s_τ	PRB$_\tau$	μ_β	s_β	PRB$_\beta$
0.1	C	0.222	0.240	0.251	7.78	0.262	0.374	18.04
	G	1.111	1.131	0.116	1.81	1.162	0.174	4.54
	F	0.907	0.963	0.916	6.13	0.931	1.183	2.56
	N	0.156	0.148	0.150	−5.65	0.148	0.217	−5.53
	t	0.156	0.159	0.160	1.46	0.148	0.219	−5.24
	P	1.564	1.699	0.776	8.65	1.868	1.302	19.49
0.2	C	0.500	0.525	0.306	4.96	0.544	0.453	8.86
	G	1.250	1.270	0.165	1.61	1.293	0.240	3.46
	F	1.861	1.887	0.902	1.40	1.854	1.257	−0.35
	N	0.309	0.311	0.138	0.80	0.306	0.201	−1.13
	t	0.309	0.302	0.153	−2.33	0.294	0.208	−5.02
	P	2.479	2.680	1.288	8.09	2.916	2.007	17.61
0.4	C	1.333	1.372	0.493	2.89	1.442	0.783	8.13
	G	1.667	1.709	0.256	2.52	1.736	0.415	4.18
	F	4.161	4.219	1.132	1.38	4.249	1.839	2.12
	N	0.588	0.589	0.105	0.27	0.577	0.167	−1.84
	t	0.588	0.580	0.114	−1.26	0.565	0.168	−3.94
	P	6.580	7.353	3.363	11.75	8.132	6.643	23.59
0.6	C	3.000	3.081	0.903	2.71	3.314	1.923	10.48
	G	2.500	2.546	0.436	1.84	2.671	0.940	6.83
	F	7.930	8.044	1.513	1.44	8.487	3.758	7.03
	N	0.809	0.803	0.058	−0.74	0.786	0.111	−2.79
	t	0.809	0.804	0.064	−0.59	0.790	0.114	−2.36
	P	21.132	24.635	11.738	16.58	*	*	*
0.8	C	8.000	——	——	——	——	——	——
	G	5.000	——	——	——	——	——	——
	F	18.192	——	——	——	——	——	——
	N	0.951	0.950	0.017	−0.13	0.939	0.048	−1.32
	t	0.951	0.949	0.020	−0.21	0.937	0.050	−1.52
	P	114.963	131.747	72.702	14.60	*	*	*

TABLE 2. *Simulation summary for $n = 200$: The mean of the estimates, the standard deviation and the percentage relative bias is reported both for $\theta_{\tau,n}$ and $\theta_{\beta,n}$; C stands for Clayton, G for Gumbel–Hougaard, F for Frank, N for Gaussian, t for Student's t, and P for Plackett.*

τ	Copula	θ	μ_τ	s_τ	PRB$_\tau$	μ_β	s_β	PRB$_\beta$
0.1	C	0.222	0.224	0.120	0.62	0.233	0.184	4.89
	G	1.111	1.117	0.058	0.53	1.123	0.083	1.07
	F	0.907	0.917	0.451	1.01	0.924	0.600	1.80
	N	0.156	0.154	0.073	−1.60	0.153	0.111	−2.31
	t	0.156	0.155	0.082	−0.95	0.153	0.110	−1.96
	P	1.564	1.572	0.339	0.56	1.611	0.482	3.05
0.2	C	0.500	0.511	0.147	2.24	0.521	0.219	4.22
	G	1.250	1.255	0.079	0.43	1.257	0.114	0.55
	F	1.861	1.881	0.435	1.10	1.894	0.605	1.76
	N	0.309	0.309	0.070	0.06	0.308	0.105	−0.35
	t	0.309	0.305	0.073	−1.22	0.302	0.102	−2.29
	P	2.479	2.528	0.553	1.96	2.581	0.724	4.10
0.4	C	1.333	1.336	0.238	0.19	1.332	0.355	−0.10
	G	1.667	1.672	0.115	0.30	1.681	0.184	0.85
	F	4.161	4.196	0.527	0.85	4.218	0.806	1.36
	N	0.588	0.587	0.050	−0.18	0.582	0.084	−0.93
	t	0.588	0.587	0.056	−0.17	0.584	0.084	−0.63
	P	6.580	6.782	1.477	3.07	6.951	2.326	5.64
0.6	C	3.000	3.019	0.422	0.61	3.054	0.719	1.78
	G	2.500	2.504	0.196	0.15	2.519	0.356	0.78
	F	7.930	7.977	0.729	0.59	8.091	1.518	2.04
	N	0.809	0.809	0.027	−0.05	0.803	0.055	−0.78
	t	0.809	0.807	0.031	−0.29	0.803	0.053	−0.76
	P	21.132	21.715	4.836	2.76	22.358	9.234	5.80
0.8	C	8.000	8.074	0.943	0.92	8.343	2.530	4.29
	G	5.000	5.043	0.450	0.85	5.201	1.245	4.01
	F	18.192	18.307	1.324	0.64	19.170	5.391	5.38
	N	0.951	0.951	0.008	−0.01	0.948	0.021	−0.31
	t	0.951	0.950	0.009	−0.14	0.946	0.022	−0.52
	P	114.963	118.266	28.426	2.87	134.740	89.240	17.20

for $\tau = 0.8$ and $n = 50$. This is because when $\beta \approx 1$, its inversion often led to dependence parameter values that exceeded the numerical accuracy of the R `copula` package [Kojadinovic and Yan, 2010b]. Consequently, the results for the inversion of Kendall's tau were omitted since no comparison could be drawn. The same also happened for the Frank family with $\tau = 0.8$ and $n = 100$.

Table 3 displays the percentage relative efficiency (PRE$_{\tau/\beta}$) of $\theta_{\beta,n}$ with respect to $\theta_{\tau,n}$ for each dependence level, copula family, and sample size considered. The interpretation of PRE$_{\tau/\beta}$ is analogous to [Kojadinovic and Yan, 2010a]: the ratio of the estimated mean square error of $\theta_{\tau,n}$ over that of $\theta_{\beta,n}$. Here, the superiority of $\theta_{\tau,n}$ is displayed in full force: at best, $\theta_{\beta,n}$ is only between 50–60% as effective. In worst-case scenarios, such as when the dependence is high, the sample size is small and the wrong copula family is being considered, the relative efficiency attained is alarmingly low.

TABLE 3. *Percentage relative efficiency (PRE) of $\theta_{\beta,n}$ with respect to $\theta_{\tau,n}$.*

τ	Copula	θ	$\text{PRE}_{\tau/\beta}$ for			
			$n=50$	$n=100$	$n=200$	$n=400$
0.1	C	0.222	0.447	0.395	0.425	0.433
	G	1.111	0.427	0.400	0.480	0.468
	F	0.907	0.602	0.583	0.565	0.531
	N	0.156	0.474	0.483	0.437	0.414
	t	0.156	0.531	0.575	0.558	0.553
	P	1.564	0.347	0.520	0.491	0.554
0.2	C	0.500	0.453	0.419	0.449	0.454
	G	1.250	0.465	0.452	0.480	0.412
	F	1.861	0.516	0.500	0.517	0.502
	N	0.309	0.469	0.479	0.446	0.430
	t	0.309	0.538	0.567	0.509	0.542
	P	2.479	0.402	0.421	0.577	0.558
0.4	C	1.333	0.390	0.448	0.449	0.440
	G	1.667	0.380	0.373	0.388	0.425
	F	4.161	0.379	0.391	0.428	0.442
	N	0.588	0.396	0.385	0.347	0.398
	t	0.588	0.454	0.486	0.443	0.458
	P	6.580	0.256	0.325	0.401	0.460
0.6	C	3.000	0.217	0.337	0.343	0.348
	G	2.500	0.211	0.277	0.304	0.317
	F	7.930	0.160	0.202	0.229	0.258
	N	0.809	0.267	0.278	0.238	0.241
	t	0.809	0.303	0.355	0.330	0.359
	P	21.132	0.000	0.197	0.273	0.341
0.8	C	8.000	——	0.084	0.137	0.177
	G	5.000	——	0.126	0.129	0.140
	F	18.192	——	——	0.059	0.077
	N	0.951	0.121	0.127	0.137	0.130
	t	0.951	0.150	0.173	0.173	0.160
	P	114.963	0.000	0.035	0.098	0.164

For low dependence levels ($\tau \leq 0.2$), $\theta_{\beta,n}$ seems to reach a plateau of efficiency relative to $\theta_{\tau,n}$ at sample sizes of around $n = 200$; for $\tau \geq 0.4$, larger sample sizes appear to continue helping in that regard. In any case, it is abundantly clear that $\theta_{\tau,n}$ is a better estimator of θ than $\theta_{\beta,n}$.

6. Conclusions

The findings described in Sections 4 and 5 make it evident that the estimator resulting from the inversion of Kendall's tau outperforms $\theta_{\beta,n}$ by a wide margin. Given that Kojadinovic & Yan [Kojadinovic and Yan, 2010a] have already shown the advantages that CML estimation enjoys in comparison to the inversion of Kendall's tau and Spearman's rho, little room is left for Blomqvist's beta.

However, all is not lost. Though not nearly as precise as $\theta_{\tau,n}$, $\theta_{\beta,n}$ does provide a quick and easily computed approximation to the dependence parameter that is readily available for any explicit family of copulas. Whenever a formal expression for Kendall's tau cannot be found for a given family of copulas, or canonical maximum likelihood becomes too cumbersome, the inversion of Blomqvist's beta presents a decent alternative. It could also find an application in tandem with more refined methods: for instance, it could be used in the selection of a starting value for CML estimation.

Appendix A: Proof of Lemma 1

There are two cases, depending on whether n is even or odd. If n is even, then the sample medians do not touch any sample points, which means that every point is counted and so $n = n_1 + n_2$. Moreover, by the definition of sample median, the number of points lying in the lower left quadrant is always equal to that in the upper right quadrant, and so n_1 can be written as

$$n_1 = 2\sum_{i=1}^{n} \mathbf{1}\left(\frac{R_i}{n+1} \leq \frac{1}{2}, \frac{S_i}{n+1} \leq \frac{1}{2}\right) = 2nC_n\left(\frac{1}{2}, \frac{1}{2}\right).$$

When n is odd, one of the sample points is not counted for β_n and therefore $n_1 + n_2 = n - 1$. Let k be the number of sample points whose ranks in both X and Y are less than or equal to their respective mid-rank. Then

$$k = nC_n\left(\frac{1}{2}, \frac{1}{2}\right) = \sum_{i=1}^{n} \mathbf{1}\left(R_i \leq \frac{n+1}{2}, S_i \leq \frac{n+1}{2}\right).$$

This quantity can also be viewed as the number of sample points in the lower left quadrant defined by the lines $x = \tilde{X}_n$ and $y = \tilde{Y}_n$. However, k includes the points on the boundaries of the quadrant, while n_1 eliminates one of those points if they exist. Therefore, n_1 and $2k$ are only equal if there is a sample point on each of the lines $x = \tilde{X}_n$ and $y = \tilde{Y}_n$ and if both of these points touch the upper right quadrant; otherwise, $2k - n_1 = 2$. Now

$$\beta_n - \beta_n^* = \left(-1 + \frac{2n_1}{n_1 + n_2}\right) - \left(-1 + \frac{2 \times 2k}{n}\right) = \frac{2n_1 n - 2 \times 2k \times n + 2 \times 2k}{n(n-1)}. \tag{16}$$

Thus if $2k = n_1$, this expression reduces to

$$\frac{k}{n} \times \frac{4}{n-1} = \frac{4}{n-1} C_n\left(\frac{1}{2}, \frac{1}{2}\right),$$

while if $2k - n_1 = 2$, (16) becomes

$$\frac{4k}{n(n-1)} - \frac{4}{n-1} = \frac{4}{n-1} C_n\left(\frac{1}{2}, \frac{1}{2}\right) - \frac{4}{n-1}.$$

Appendix B: Proof of Equation (8)

Let $\sigma^2_{C,\beta}$ be the asymptotic variance of Blomqvist's beta with underlying copula C. Under the conditions set by [Segers, 2012], the value of $\sigma^2_{C,\beta}$ is given by

$$\frac{\sigma^2_{C,\beta}}{16} = \operatorname{var}\left\{\mathbb{C}(1/2,1/2) - C_1(1/2,1/2)\mathbb{C}(1/2,1) - C_2(1/2,1/2)\mathbb{C}(1,1/2)\right\}, \tag{17}$$

where $C_1(u,v) = \partial C(u,v)/\partial u$ and $C_2(u,v) = \partial C(u,v)/\partial v$. Note that in (17), $\mathbb{C}(1/2,1/2)$, $\mathbb{C}(1/2,1)$ and $\mathbb{C}(1,1/2)$ are random variables, while $a = C_1(1/2,1/2)$ and $b = C_2(1/2,1/2)$ are constants.

In order to arrive at (8), it is necessary to expand the right-hand side of (17) as follows:

$$\begin{aligned}\frac{\sigma^2_{C,\beta}}{16} = {} & \mathrm{var}\{\mathbb{C}(1/2,1/2)\} + a^2\mathrm{var}\{\mathbb{C}(1/2,1)\} + b^2\mathrm{var}\{\mathbb{C}(1,1/2)\} \\ & -2a\,\mathrm{cov}\{\mathbb{C}(1/2,1/2),\mathbb{C}(1/2,1)\} - 2b\,\mathrm{cov}\{\mathbb{C}(1/2,1/2),\mathbb{C}(1,1/2)\} \\ & +2ab\,\mathrm{cov}\{\mathbb{C}(1/2,1),\mathbb{C}(1,1/2)\}. \end{aligned} \tag{18}$$

It only remains to calculate each of the variance and covariance terms in the expression above. This can be accomplished by noting that the limiting process $\mathbb{C}$ has covariance function

$$\mathrm{cov}\{\mathbb{C}(u,v),\mathbb{C}(s,t)\} = C(u\wedge s, v\wedge t) - C(u,v)C(s,t)$$

for all $u,v,s,t \in [0,1]$, where $a \wedge b = \min(a,b)$. Let $C(1/2,1/2) = c$. Direct calculations yield

$$\mathrm{var}\{\mathbb{C}(1/2,1/2)\} = c(1-c), \quad \mathrm{var}\{\mathbb{C}(1/2,1)\} = \mathrm{var}\{\mathbb{C}(1,1/2)\} = 1/4,$$

$$\mathrm{cov}\{\mathbb{C}(1/2,1/2),\mathbb{C}(1/2,1)\} = \mathrm{cov}\{\mathbb{C}(1/2,1/2),\mathbb{C}(1,1/2)\} = c/2$$

and

$$\mathrm{cov}\{\mathbb{C}(1/2,1),\mathbb{C}(1,1/2)\} = c - 1/4.$$

Substituting these terms into (18) yields

$$\frac{\sigma^2_{C,\beta}}{16} = c(1-c) + c(-a-b+2ab) + (a-b)^2/4.$$

Acknowledgments

The authors would like to thank Professor Johanna Nešlehová for stimulating discussion. This research was supported by the Canada Research Chairs Program and grants from the Natural Sciences and Engineering Research Council of Canada and the Fonds de recherche du Québec – Nature et technologies. The second author is also grateful to the Groupe d'études et de recherche en analyse des décisions for partial support.

References

[Abdous et al., 2005] Abdous, B., Genest, C., and Rémillard, B. (2005). Dependence properties of meta-elliptical distributions. In *Statistical Modeling and Analysis for Complex Data Problems*, volume 1, pages 1–15. Springer, New York.

[Blomqvist, 1950] Blomqvist, N. (1950). On a measure of dependence between two random variables. *Ann. Math. Statist.*, 21:593–600.

[Borkowf, 2002] Borkowf, C. (2002). Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation. *Comput. Statist. Data Anal.*, 39:271–286.

[Fang et al., 2005] Fang, H.-B., Fang, K.-T., and Kotz, S. (2002 [Corrig.: *J. Multivariate Anal.*, 94:222–223, 2005]). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.*, 82:1–16.

[Genest and Favre, 2007] Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12:347–368.

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

[Genest and MacKay, 1986] Genest, C. and MacKay, R. (1986). Copules archimédiennes et familles de lois bidimensionnelles dont les marges sont données. *Canad. J. Statist.*, 14:145–159.

[Genest et al., 2011] Genest, C., Nešlehová, J., and Ben Ghorbal, N. (2011). Estimators based on Kendall's tau in multivariate copula models. *Aust. N. Z. J. Stat.*, 53:157–177.

[Genest et al., 2012] Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Ann. Inst. Statist. Math.*, 64:811–834.

[Genest and Werker, 2002] Genest, C. and Werker, B. (2002). Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In *Distributions with given marginals and statistical modelling*, pages 103–112. Kluwer, Dordrecht.

[Ghoudi et al., 1998] Ghoudi, K., Khoudraji, A., and Rivest, L.-P. (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles. *Canad. J. Statist.*, 26:187–197.

[Hoeffding, 1947] Hoeffding, W. (1947). On the distribution of the rank correlation coefficient τ when the variates are not independent. *Biometrika*, 34:183–196.

[Hoeffding, 1948] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, 19:293–325.

[Hult and Lindskog, 2002] Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Adv. Appl. Probab.*, 34:587–608.

[Kojadinovic and Yan, 2010a] Kojadinovic, I. and Yan, J. (2010a). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance Math. Econom.*, 47:52–63.

[Kojadinovic and Yan, 2010b] Kojadinovic, I. and Yan, J. (2010b). Modeling multivariate distributions with continuous margins using the copula R package. *J. Statist. Software*, 34:1–20.

[Oakes, 1994] Oakes, D. (1994). Multivariate survival distributions. *J. Nonparametr. Statist.*, 3:343–354.

[Rüschendorf, 1976] Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.*, 4:912–923.

[Schmid and Schmidt, 2007] Schmid, F. and Schmidt, R. (2007). Nonparametric inference on multivariate versions of Blomqvist's beta and related measures of tail dependence. *Metrika*, 66:323–354.

[Segers, 2012] Segers, J. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli*, 18:764–782.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231.

[Tsukahara, 2005] Tsukahara, H. (2005). Semiparametric estimation in copula models. *Canad. J. Statist.*, 33:357–375.