Model robust inference with two-stage maximum likelihood estimation for copulas

Vinnie Ko *, Nils Lid Hjort

Department of Mathematics, University of Oslo, PB 1053, Blindern, NO-0316, Oslo, Norway

ARTICLE INFO

ABSTRACT

This article is concerned with inference in parametric copula setups, where both the marginals and the copula have parametric forms. For such models, two-stage maximum likelihood estimation, often referred to as inference function for margins, is used as an attractive alternative to the full maximum likelihood estimation strategy. Previous studies of the two-stage maximum likelihood estimator have largely been based on the assumption that the chosen parametric model captures the true model that generated data. We study the impact of dropping this true model assumption, both theoretically and numerically. We first show that the two-stage maximum likelihood estimator is consistent for a well-defined least false parameter value, different from the analogous least false parameter associated with the full maximum likelihood procedure. Then we demonstrate limiting normality of the full vector of estimators, with concise matrix notation for the variance matrices involved. Along with consistent estimators for these, we have built a model-robust machinery for inference in parametric copula models. The special case where the parametric model is assumed to hold corresponds to situations studied earlier in the literature, with simpler formulas for variance matrices. As a numerical illustration, we perform a set of simulations. We also analyze five-dimensional Norwegian precipitation data. We find that the variance of the copula parameter estimate can both increase and decrease, by dropping the true model assumption. In addition, we observe that the two-stage maximum likelihood estimator is still highly efficient when the true model assumption is dropped and thus the model robust asymptotic variance formulas are used. Additionally, we discover that using highly misspecified models can lead to situations where the asymptotic variance of the two-stage maximum likelihood estimator is lower than that of full maximum likelihood estimator. Our results are also used to analyze the mean squared error properties for both the full and the two-stage maximum likelihood estimators of any focus parameter.

© 2019 Elsevier Inc. All rights reserved.

1. Introduction and copula models

The popularity of copula modeling and methods has increased rapidly in the last decade and now they are regularly used in fields like biostatistics, hydrology, finance and actuarial science; see Embrechts [10]. One of the more popular estimation techniques for parametric copulas with parametric margins is the two-stage maximum likelihood estimation (two-stage ML estimation), first introduced by Shih and Louis [35]. This two-stage method is also often referred to as inference function

for margins (IFM), a term coined by Xu [39]. The asymptotic behavior and efficiency of this method have been studied both theoretically and numerically in [5,19,20,24,35,39].

Although the insights and results from these studies are fruitful, they are generally based on the assumption that the chosen copula and marginal distributions are the true model that generated data. We refer to this as the 'true model assumption', and the limitations and impact of this assumption for two-stage ML estimation in copula contexts have not been studied earlier, to our knowledge. Andersen [4] has worked with model robust versions of asymptotic variances. This was in the context of composite likelihoods, however, and her work does not give attention to the inference and impact of model robustness. In this paper we develop model robust inference methods based on two-stage ML estimators, and study their theoretical and numerical properties under possible model misspecification.

Our technical setting is as follows. Let $(Y_1, \ldots, Y_d)^\top$ be a d-variate continuous random vector originating from a joint density $g(y_1, \ldots, y_d)$, and further let $y_1 = (y_{1,1}, \ldots, y_{1,d})^\top, \ldots, y_n = (y_{n,1}, \ldots, y_{n,d})^\top$ be mutually independent observations of this variable. Typically, this true joint distribution g is unknown. Let $f(y_1, \ldots, y_d, \eta)$ be our choice of parametric approximation of g, with η the parameter vector, belonging to some connected subset of the appropriate Euclidean space. Further, G and $F(\cdot, \eta)$ indicate cumulative distribution functions corresponding to g and $f(\cdot, \eta)$, respectively. In addition, $G_j(y_j)$ and $F_j(y_j, \alpha_j)$ indicate the jth marginal distribution functions corresponding to G and $F(\cdot, \eta)$, respectively, with α_j the parameter vector pertaining to modeling margin component $j \in \{1, \ldots, d\}$.

Starting from the joint distribution $F(\cdot, \eta)$ under certain regularity conditions, Sklar's theorem [36] implies that there exists a copula $C(u_1, \ldots, u_d, \theta)$ that satisfies

$$F(y_1, \ldots, y_d, \eta) = C\{F_1(y_1, \alpha_1), \ldots, F_d(y_d, \alpha_d), \theta\}, \tag{1}$$

with θ the vector of parameters for the copula. For the full parameter vector, now conveniently blocked into parameters for margins and the part for the copula, we use

$$\eta = (\alpha^\top, \theta^\top)^\top = (\alpha_1^\top, \ldots, \alpha_d^\top, \theta^\top)^\top.$$

When $F(y_1, \ldots, y_d, \eta)$ is continuous, $C(\cdot, \theta)$ is unique. If we assume that $F_1(y_1, \alpha_1), \ldots, F_d(y_d, \alpha_d)$ are absolutely continuous and strictly increasing, $C(\cdot, \theta)$ can be differentiated and (1) becomes

$$f(y_1, \ldots, y_d, \eta) = c\{F_1(y_1, \alpha_1), \ldots, F_d(y_d, \alpha_d), \theta\} \prod_{j=1}^{d} f_j(y_j, \alpha_j), \tag{2}$$

where $c(u_1, \ldots, u_d) = \partial^d C(u_1, \ldots, u_d, \theta)/\partial u_1 \cdots \partial u_d$ and $f_j(y_j, \alpha_j) = \partial F_j(y_j, \alpha_j)/\partial y_j$; see Nelsen [32]. Analogously, we can decompose the true density g into marginal densities and copula density and obtain

$$g(y_1, \ldots, y_d) = c_0\{G_1(y_1), \ldots, G_d(y_d)\} \prod_{j=1}^{d} g_j(y_j),$$

with c_0 the true copula.

One of the most important families of copulas is the Archimedean class. An Archimedean copula has the form

$$C(u_1, \ldots, u_d) = \phi^{-1}\{\phi(u_1) + \cdots + \phi(u_d)\},$$

where ϕ is its generator; see, e.g., [31]. One of the most common Archimedean copulas is the Gumbel copula with parameter $\theta \in [1, \infty)$, which is defined for all $u_1, \ldots, u_d \in (0, 1)$, as

$$C(u_1, \ldots, u_d) = \exp[-(|\ln u_1|^\theta + \cdots + |\ln u_d|^\theta)^{1/\theta}].$$

Another member of the Archimedean family is the Frank copula with parameter $\theta \in (0, \infty)$, defined for all $u_1, \ldots, u_d \in (0, 1)$, defined as

$$C(u_1, \ldots, u_d) = -\frac{1}{\theta} \ln \left[1 + \frac{\prod_{i=1}^{d}\{\exp(-\theta u_i) - 1\}}{\{\exp(-\theta - 1)\}^{d-1}} \right].$$

For further parametric copula constructions, see, e.g., [11–13,19,21,32], and with several multiparameter copula models exhibited and discussed in [9,33].

The further structure of this paper is as follows. In Section 2, we briefly describe how the two-stage ML estimation method works. In Section 3, we first derive the limit distribution for two-stage ML estimators outside model conditions, including proof of consistency towards the relevant least false parameter. After this, we examine consequences of the true model assumption for the asymptotic distribution. The resulting model non-robust asymptotic variance formula is essentially similar to that given in Joe [20], but one difference is that we choose to use quantities that are more in line with the classical ML estimation theory. Our results give rise to clear recipes for confidence intervals, confidence curves, and hypothesis tests, based on two-stage ML estimators, discussed in Section 4.

In Sections 5 and 6, we study the numerical behavior of our model robust inference methods against that of the model non-robust version by using a set of simulations and a dataset of Norwegian precipitation. Generally speaking, the two-stage

ML estimator $\widetilde{\eta}$ has a variance matrix, say Σ, for which our apparatus provides two estimators: $\widetilde{\Sigma}_A$, computed under model conditions, and $\widetilde{\Sigma}_B$ using the model-robust machinery. When the true model really is captured by the candidate model, then both variance matrix estimators tend to be in agreement. Next, when the data generating mechanism lies outside the model, the second method aims at the correct matrix, whereas the first is not consistent. The disagreement between these two variance matrix estimators gets larger as the degree of misspecification increases. Furthermore, we observe that a high degree of model misspecification may lead to situations where the limiting variance matrix of the two-stage ML estimator is smaller than the corresponding variance matrix of the full ML estimator. This cannot happen when the data generating mechanism is inside the parametric model. By classical theorems on the optimality of ML estimators, such as the Hájek–Le Cam convolution theorem [15], each competing sequence of estimators will have a limiting variance matrix at least as large as the one for ML estimation (the matrix difference is nonnegative definite), under model conditions and a few further regularity assumptions on competitors. But these optimality theorems for ML estimation do not apply when the data generating mechanism is outside the parametric model.

In Section 7, our methods and results are used to examine the mean squared error properties of both full and two-stage maximum likelihood estimators for any given focus parameter. This also leads to a new focused information criterion for copulas, when examining different candidate models. In Section 8, we offer a list of concluding remarks, some pointing to further research work. In particular, we mention model selection criteria, and explain briefly that our two-stage ML machinery may be extended to classes of conditional copula regression models.

2. Two-stage maximum likelihood

2.1. Maximum likelihood and Kullback–Leibler divergence

With observations from a model parametrized via a parameter vector η, the ML estimator $\widehat{\eta}$ is the maximizer of $\ell(\eta)$, the log-likelihood function of the model for the given observations. Properties of ML estimators are extensively covered by the classic literature in statistics, including [6,29,38]. In this paper, we use $\hat{}$ to indicate that a quantity is estimated by full or joint ML, and $\tilde{}$ to indicate that the quantity in question is estimated by two-stage ML, covered in Section 2.2.

The Kullback–Leibler (KL) divergence [28] measures the extent to which one probability distribution diverges from another. The KL divergence from g to f is defined as

$$\mathrm{KL}\{g, f(\cdot, \eta)\} = \int g(y) \ln\{g(y)/f(y, \eta)\} \, \mathrm{d}y.$$

In our technical setting, the true density g is the same for all models. Thus, minimizing the KL divergence is equivalent to maximizing $\int g(y) \ln f(y, \eta) \, \mathrm{d}y$.

Under the usual regularity assumptions and assuming that the integral is finite, the Law of Large Numbers gives

$$\frac{1}{n} \ell(\eta) = \frac{1}{n} \sum_{i=1}^{n} \ln f(y_i, \eta) \xrightarrow{p} \int g(y) \ln f(y, \eta) \, \mathrm{d}y = \mathrm{E}_G\{\ln f(Y, \eta)\}.$$

Here and later it is understood that the convergence relates to the sample size n growing beyond bounds. The above result implies under mild and standard regularity conditions that $\widehat{\eta}$, the ML estimator of η, will tend asymptotically to $\eta_{0,\mathrm{ML}}$, the least false parameter value and the maximizer of $\int g(y) \ln f(y, \eta) \, \mathrm{d}y$. Thus,

$$\widehat{\eta} \xrightarrow{p} \eta_{0,\mathrm{ML}} = \arg\min_{\eta} \mathrm{KL}\{g, f(\cdot, \eta)\} = \arg\max_{\eta} \int g(y) \ln f(y, \eta) \, \mathrm{d}y.$$

In other words, the $\eta_{0,\mathrm{ML}}$ is the parameter vector making the candidate model closest to the true model, in the sense of KL. When the parametric model is correctly specified, i.e., $g(y) = f(y, \eta_{0,\mathrm{ML}})$, for a suitable $\eta_{0,\mathrm{ML}}$, the minimum of the KL divergence is zero and $\eta_{0,\mathrm{ML}}$ is called true parameter value. By applying the decomposition (2) on our setting and $f(y, \eta)$, we have that $\widehat{\eta}$ is a consistent estimator of the least false parameter $\eta_{0,\mathrm{ML}}$, say $\eta_{0,\mathrm{ML}} = (\alpha_{0,\mathrm{ML}}^\top, \theta_{0,\mathrm{ML}}^\top)^\top$.

2.2. Two-stage maximum likelihood estimator

When the dimension d of the copula gets higher, the ML estimator becomes computationally more demanding and not even feasible when the number of parameters is high [19]. To avoid this problem, the two-stage ML is a natural alternative estimation strategy, first proposed, for the two-dimensional case, by Shih and Louis [35].

When parametric families for the copula and the margins are chosen, the two-stage ML estimator works as follows.

Stage 1: For each $j \in \{1, \ldots, d\}$, obtain $\widetilde{\alpha}_j$, the marginal ML estimate of α_j, by maximizing, with respect to α_j,

$$\ell_{f_j} = \sum_{i=1}^{n} \ln f_j(y_{i,j}, \alpha_j).$$

Stage 2: Plug in $\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_d$ from Stage 1, to get

$$\ell(\widetilde{\alpha}, \theta) = \sum_{i=1}^{n} [\ln f_1(y_{i,1}, \widetilde{\alpha}_1) + \cdots + \ln f_d(y_{i,d}, \widetilde{\alpha}_d) + \ln c\{F_1(y_{i,1}, \widetilde{\alpha}_1), \ldots, F_d(y_{i,d}, \widetilde{\alpha}_d), \theta\}].$$

Then, θ is estimated by maximizing $\ell(\widetilde{\alpha}, \theta)$ with respect to θ, yielding

$$\widetilde{\theta} = \arg\max_{\theta} \ell(\widetilde{\alpha}, \theta) = \arg\max_{\theta} \ell_c(\widetilde{\alpha}, \theta),$$

where

$$\ell_c(\alpha, \theta) = \sum_{i=1}^{n} \ln c\{F_1(y_{i,1}, \alpha_1), \ldots, F_d(y_{i,d}, \alpha_d), \theta\}.$$

The two-stage ML estimate $\widetilde{\eta} = (\widetilde{\alpha}^\top, \widetilde{\theta}^\top)^\top = (\widetilde{\alpha}_1^\top, \ldots, \widetilde{\alpha}_d^\top, \widetilde{\theta}^\top)^\top$ obtained in this way satisfies

$$\left(\frac{\partial}{\partial \widetilde{\alpha}_d} \ell_{f_1}, \ldots, \frac{\partial}{\partial \widetilde{\alpha}_d} \ell_{f_d}, \frac{\partial}{\partial \widetilde{\theta}} \ell_c \right) = (0, \ldots, 0).$$

There are $p_1 + \cdots + p_d + q$ parameters and equations here, where p_j is the dimension of α_j and q the dimension of θ. In the next section we demonstrate limiting normality for the full $(p_1 + \cdots + p_d + q)$-dimensional $\widetilde{\eta}$. Note that since the main focus of this paper is two-stage ML estimation, we will simplify the notation for the least false parameter of two-stage ML by using $\eta_0 = \eta_{0,\text{2ML}}$ throughout this paper.

3. Large-sample behavior of two-stage maximum likelihood estimators

To work out clear limit theorems for the two-stage ML estimators we are helped by the following regularity assumptions.

A1: $\eta \in \Theta$, where Θ is compact.
A2: $\ln f_j(y, \alpha_j)$ is twice differentiable with respect to α_j.
A3: $\ln c\{F_1(y_1, \alpha_1), \ldots, F_d(y_d, \alpha_d), \theta\}$ is thrice differentiable with respect to α and θ.
A4: There exists a function $R(y_j)$ such that $E_G\{R(Y_j)\} < \infty$ and $|\ln f_j(y_j, \alpha_j)| \le R(y_j)$ for all y_j, α_j.
A5: There exists a function $R(y, \alpha)$ such that $E_G\{R(Y, \alpha)\} < \infty$ and $|\ln c(y, \alpha, \theta)| \le R(y, \alpha)$ for all y, α, θ.

In Stage 1 of two-stage ML estimation, the parameters of each margin are estimated by using separate ML estimation. We may consequently use large-sample results from ML estimation theory directly, regarding the behavior of $\widetilde{\alpha}_j$. In particular, the estimator $\widetilde{\alpha}_j$ is now aiming for the least false value instead of the usual true parameter value under the true model assumption.

3.1. Large-sample results for Stage 1 of two-stage ML estimation

The model-robust asymptotic proprieties of the ML estimator are already covered by White [38]. Here, we briefly recap the result.

Lemma 1. *Let $\widetilde{\alpha}$ be the ML estimator of the margin parameter vector $\alpha = (\alpha_1, \ldots, \alpha_d)^\top$ from Stage 1 and let $\alpha_0 = (\alpha_{0,1}, \ldots, \alpha_{0,d})^\top$ be the least false value. Then we have*

$$\sqrt{n}(\widetilde{\alpha} - \alpha_0) = \sqrt{n} \mathcal{I}_\alpha^{-1} U_{n,\alpha}(\alpha_0) + o_p(1) \rightsquigarrow \mathcal{I}_\alpha^{-1} \Lambda_\alpha \sim \mathcal{N}(0, \mathcal{I}_\alpha^{-1} K_\alpha \mathcal{I}_\alpha^{-1}),$$

where

$$U_{n,\alpha}(\alpha) = \begin{pmatrix} U_{n,\alpha_1}(\alpha_1) \\ \vdots \\ U_{n,\alpha_d}(\alpha_d) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} U_{\alpha_1}(y_{i,1}, \alpha_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} U_{\alpha_d}(y_{i,d}, \alpha_d) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \alpha_1} \ln f_1(y_{i,1}, \alpha_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \alpha_d} \ln f_d(y_{i,d}, \alpha_d) \end{pmatrix},$$

$$K_\alpha = \begin{pmatrix} K_{\alpha_1} & K_{\alpha_1,\alpha_2} & \cdots & K_{\alpha_1,\alpha_d} \\ K_{\alpha_2,\alpha_1} & K_{\alpha_2} & \cdots & K_{\alpha_2,\alpha_d} \\ \vdots & \vdots & \ddots & \vdots \\ K_{\alpha_d,\alpha_1} & K_{\alpha_d,\alpha_2} & \cdots & K_{\alpha_d} \end{pmatrix}, \quad K_{\alpha_j} = K_{\alpha_j,\alpha_j},$$

$$K_{\alpha_j,\alpha_k} = \text{cov}_G\{U_{\alpha_j}(y_j, \alpha_{0,j}), U_{\alpha_k}(y_k, \alpha_{0,k})\} = E_G\{U_{\alpha_j}(y_j, \alpha_{0,j}) U_{\alpha_k}(y_k, \alpha_{0,k})^\top\},$$

$$
\mathcal{I}_\alpha = \begin{pmatrix} \mathcal{I}_{\alpha_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathcal{I}_{\alpha_d} \end{pmatrix}, \quad \mathcal{I}_{\alpha_j} = -\int g_j \frac{\partial^2}{\partial \alpha_{0,j} \partial \alpha_{0,j}^\top} \ln f_j(y_j, \alpha_{0,j}) \, dy_j = -E_{G_j}\{H_{\alpha_j}(y_j, \alpha_{0,j})\}.
$$

We have used $H_{\alpha_j}(y_j, \alpha_j)$ for the Hessian matrix operation on $\ln f_j(y_j, \alpha_j)$.

For the proof, see the proof of Theorem 3.2 in White [38].

Under the assumption that the margins are correctly specified, one has $K_{\alpha_j} = \mathcal{I}_{\alpha_j}$. When $j \neq k$, however, it holds in general that $K_{\alpha_j, \alpha_k} \neq 0$. The asymptotic variance matrix will hence keep the sandwich form even under the assumption that margins are correctly specified.

3.2. Large-sample results for Stage 2 of two-stage ML estimation

Under the assumption that all margins and the copula are correctly specified, it is well known that $\widetilde{\theta}$ is a consistent estimator of the true parameter value θ_0, i.e., $KL\{c_0, c(\cdot, \theta_0)\} = 0$. However, without this assumption, the consistency of $\widetilde{\theta}$ needs more care, also since it needs to be clarified precisely what it is aiming for.

Lemma 2. *Consider the function*

$$
M(\alpha_0, \theta) = \int g \ln c\{F_1(y_1, \alpha_{0,1}), \ldots, F_d(y_d, \alpha_{0,d}), \theta\} \, dy
$$

and assume that there exists θ_0, a unique and well-separated point of maximum of $M(\alpha_0, \theta)$, which satisfies

$$
\sup\{M(\alpha_0, \theta) : d(\theta, \theta_0) < \varepsilon\} < M(\alpha_0, \theta_0)
$$

for every $\varepsilon > 0$; here $d(\theta, \theta_0)$ refers to Euclidean distance. Let $M_n(\widetilde{\alpha}, \theta)$ be the Stage 2 sample version of $M(\alpha_0, \theta)$, viz.

$$
M_n(\widetilde{\alpha}, \theta) = \frac{1}{n} \sum_{i=1}^n \ln c\{F_1(y_{i,1}, \widetilde{\alpha}_1), \ldots, F_d(y_{i,d}, \widetilde{\alpha}_d), \theta\},
$$

where the $\widetilde{\alpha}_j$s are the estimates from the Stage 1. Then $\widetilde{\theta}$, the maximizer of $M_n(\widetilde{\alpha}, \theta)$, is a consistent estimator of θ_0, the least false parameter value.

The proof is given in the Online Supplement [26].

So, under margin and copula misspecification, $\widetilde{\theta}$ is still consistent, but for the appropriate least false parameter value, rather than for any 'true' parameter value. The divergence in question is of the KL type form

$$
\int g \ln \frac{c_0\{F_1(y_1, \alpha_{0,1}), \ldots, F_d(y_d, \alpha_{0,d})\}}{c\{F_1(y_1, \alpha_{0,1}), \ldots, F_d(y_d, \alpha_{0,d}), \theta\}} \, dy.
$$

This leads to a precise notion of the least false parameter vector $\eta_0 = (\alpha_0^\top, \theta_0^\top)^\top$ associated with two-stage ML estimation.

Based on the consistency lemma above, we now derive the model robust asymptotic distribution of the two-stage ML estimator.

Proposition 1. *With $\widetilde{\eta}$ the two-stage ML estimator of η, we have*

$$
\sqrt{n} \, (\widetilde{\eta} - \eta_0) = \sqrt{n} \begin{pmatrix} \mathcal{I}_\alpha & 0 \\ \mathcal{I}_{\alpha,\theta}^\top & \mathcal{I}_\theta \end{pmatrix}^{-1} \begin{pmatrix} U_{n,\alpha}(\alpha_0) \\ U_{n,\theta}(\alpha_0, \theta_0) \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix} \rightsquigarrow \mathcal{I}_\eta^{-1} \Lambda_\eta \sim \mathcal{N}(0, \, V_\eta),
$$

where

$$
V_\eta = \mathcal{I}_\eta^{-1} K_\eta (\mathcal{I}_\eta^{-1})^\top,
$$

$$
U_{n,\theta}(\alpha, \theta) = \frac{1}{n} \sum_{i=1}^n U_\theta(y_i, \alpha, \theta) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln c\{F_1(y_{i,1}, \alpha_1), \ldots, F_d(y_{i,d}, \alpha_d)\theta\},
$$

$$
K_\eta = \begin{pmatrix} K_\alpha & K_{\alpha,\theta} \\ K_{\alpha,\theta}^\top & K_\theta \end{pmatrix},
$$

$$
K_\theta = \mathrm{var}_G \, U_\theta(y, \alpha_0, \theta_0) = E\{U_\theta(y, \alpha_0, \theta_0) \, U_\theta(y, \alpha_0, \theta_0)^\top\},
$$

$$
K_{\alpha,\theta} = \mathrm{cov}_G\{U_\alpha(y, \alpha_0), U_\theta(y, \alpha_0, \theta_0)\} = E\{U_\alpha(y, \alpha_0) \, U_\theta(y, \alpha_0, \theta_0)^\top\},
$$

$$
\mathcal{I}_\eta = \begin{pmatrix} \mathcal{I}_\alpha & 0 \\ \mathcal{I}_{\alpha,\theta}^\top & \mathcal{I}_\theta \end{pmatrix},
$$

$$
\mathcal{I}_\theta = -\int g \frac{\partial^2}{\partial \theta_0 \partial \theta_0^\top} \ln c\{F_1(y_1, \alpha_{0,1}), \ldots, F_d(y_d, \alpha_{0,d}), \theta\} \, dy = -E_G\{H_\theta(y, \alpha_0, \theta_0)\},
$$

$$\mathcal{I}_{\alpha,\theta} = -\int g \,\frac{\partial^2}{\partial\alpha_0\partial\theta_0^\top}\,\ln c\{F_1(y_1,\alpha_{0,1}),\ldots,F_d(y_d,\alpha_{0,d}),\theta\}\,\mathrm{d}y = -\mathrm{E}_G\{H_{\alpha,\theta}(y,\alpha_0,\theta_0)\}.$$

The proof is given in the Online Supplement [26].

Compared to the results from [20,39] where margins and copula are assumed to be correctly specified, the asymptotic variance expression in Proposition 1 has a more general form. Writing it in block matrix form gives

$$V_\eta = \begin{pmatrix} V_\alpha & V_{\alpha,\theta} \\ V_{\alpha,\theta}^\top & V_\theta \end{pmatrix} = \begin{pmatrix} \mathcal{I}_\alpha^{-1} & 0 \\ -\mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^\top\mathcal{I}_\alpha^{-1} & \mathcal{I}_\theta^{-1} \end{pmatrix}\begin{pmatrix} K_\alpha & K_{\alpha,\theta} \\ K_{\alpha,\theta}^\top & K_\theta \end{pmatrix}\begin{pmatrix} \mathcal{I}_\alpha^{-1} & -\mathcal{I}_\alpha^{-1}\mathcal{I}_{\alpha,\theta}\mathcal{I}_\theta^{-1} \\ 0 & \mathcal{I}_\theta^{-1} \end{pmatrix}, \tag{3}$$

where

$$V_\alpha = \mathcal{I}_\alpha^{-1}K_\alpha\,\mathcal{I}_\alpha^{-1},$$
$$V_\theta = \mathcal{I}_\theta^{-1}K_\theta\,\mathcal{I}_\theta^{-1} + \mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^\top\,\mathcal{I}_\alpha^{-1}K_\alpha\,\mathcal{I}_\alpha^{-1}\mathcal{I}_{\alpha,\theta}\,\mathcal{I}_\theta^{-1} - \mathcal{I}_\theta^{-1}K_{\alpha,\theta}^\top\,\mathcal{I}_\alpha^{-1}\mathcal{I}_{\alpha,\theta}\,\mathcal{I}_\theta^{-1} - \mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^\top\,\mathcal{I}_\alpha^{-1}K_{\alpha,\theta}\,\mathcal{I}_\theta^{-1},$$
$$V_{\alpha,\theta} = \mathcal{I}_\alpha^{-1}K_{\alpha,\theta}\,\mathcal{I}_\theta^{-1} - \mathcal{I}_\alpha^{-1}K_\alpha\,\mathcal{I}_\alpha^{-1}\mathcal{I}_{\alpha,\theta}\,\mathcal{I}_\theta^{-1}.$$

Andersen [4] has a similar result, but in the context of a composite likelihood construction.

Further, since the two-stage ML estimator can be seen as a special case of the generalized method of moments (GMM) in econometrics, Proposition 1 and Lemma 2 can also be seen as the direct result of consistency and asymptotic normality of the GMM estimator with the optimal weight matrix $W = K_\eta^{-1}$. See Hall [16] for details of the GMM estimator. Similarly, Proposition 1 can also be derived by using the theories of estimating equations, like in Section 5.5 of Joe [21].

Since the main focus of our paper is finding the exact effect of the true model assumption, i.e., $f = g$, we now look at how the true model assumption simplifies the result from Proposition 1. As already mentioned in Section 3.1, the true model assumption in Stage 1 gives equality $K_{\alpha_j} = \mathcal{I}_{\alpha_j}$ for all $j \in \{1,\ldots,d\}$, where d indicates the dimension of the data.

Joe [20] implements this result by replacing each $\mathcal{I}_{\alpha_j}$ with K_{α_j} in his asymptotic variance formula. Although this method is theoretically correct, it is not the most economical way of stating results, since V_η contains more $\mathcal{I}_{\alpha_j}$s than K_{α_j}s; note that every K_{α_j} is 'sandwiched' by $\mathcal{I}_{\alpha_j}$ in (3). In addition, this is in opposition to the usual practice of maximum likelihood theory, where the asymptotic variance is defined and used as the inverse of Fisher information, as opposed to the variance of the score function vector.

Thus, under the true model assumption, we choose to simplify V_η by replacing K_α with

$$K_\alpha^{\mathrm{TMA}} = \begin{pmatrix} \mathcal{I}_{\alpha_1} & K_{\alpha_1,\alpha_2} & \cdots & K_{\alpha_1,\alpha_d} \\ K_{\alpha_2,\alpha_1} & \mathcal{I}_{\alpha_2} & \cdots & K_{\alpha_2,\alpha_d} \\ \vdots & \vdots & \ddots & \vdots \\ K_{\alpha_d,\alpha_1} & K_{\alpha_d,\alpha_2} & \cdots & \mathcal{I}_{\alpha_d} \end{pmatrix}.$$

Although our method is theoretically the same as that of Joe [20], it will make a difference in practice since it is almost never the case that the two matrices K_{α_j} and $\mathcal{I}_{\alpha_j}$, though identical as population quantities, have empirical estimates that are very close.

Lemma 3. *Under the assumption that the margins and copula are correctly specified, one has $K_{\alpha,\theta} = 0$.*

The proof is given in the Online Supplement [26].

In two-stage ML estimation, the log-likelihood functions are different in Stages 1 and 2. This implies that the true model assumption does not give $\mathcal{I}_{\alpha,\theta} = K_{\alpha,\theta}$. Instead, the true model assumption for two-stage ML estimation yields a different relationship, as follows.

Lemma 4. *Let*

$$U_\alpha^*(y,\alpha) = \begin{pmatrix} U_{\alpha_1}^*(y,\alpha_1) \\ \vdots \\ U_{\alpha_d}^*(y,\alpha_d) \end{pmatrix} = \begin{pmatrix} \dfrac{\partial}{\partial\alpha_1}\ln c\{F_1(y_1,\alpha_1),\ldots,F_d(y_d,\alpha_d),\theta\} \\ \vdots \\ \dfrac{\partial}{\partial\alpha_d}\ln c\{F_1(y_1,\alpha_1),\ldots,F_d(y_d,\alpha_d),\theta\} \end{pmatrix}$$

and

$$K_{\alpha,\theta}^* = \mathrm{cov}_G\{U_\alpha^*(y,\alpha_0),U_\theta(y,\alpha_0,\theta_0)\} = \mathrm{E}_G\{U_\alpha^*(y,\alpha_0)\,U_\theta(y,\alpha_0,\theta_0)^\top\}.$$

Under the assumption that the margins and copula are correctly specified, one has $\mathcal{I}_{\alpha,\theta} = K_{\alpha,\theta}^*$.

The proof is given in the Online Supplement [26].

Lemma 5. *Under the assumption that the margins and copula are correctly specified, one has $\mathcal{I}_\theta = K_\theta$.*

The proof is given in the Online Supplement [26].

3.3. Impact of model misspecification on the copula parameter

The most apparent consequence of margin misspecification is that the two-stage ML copula parameter estimate $\widetilde{\theta}$ is no longer consistent for the true parameter value. Another consequence is the change in asymptotic variance V_η. Applying the lemmas resulting from true model assumption to (3) gives

$$
V_\eta^{\text{TMA}} = \begin{pmatrix} V_\alpha^{\text{TMA}} & V_{\alpha,\theta}^{\text{TMA}} \\ (V_{\alpha,\theta}^{\text{TMA}})^\top & V_\theta^{\text{TMA}} \end{pmatrix} = \begin{pmatrix} \mathcal{I}_\alpha^{-1} & 0 \\ -\mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^\top \mathcal{I}_\alpha^{-1} & \mathcal{I}_\theta^{-1} \end{pmatrix} \begin{pmatrix} K_\alpha^{\text{TMA}} & 0 \\ 0 & \mathcal{I}_\theta \end{pmatrix} \begin{pmatrix} \mathcal{I}_\alpha^{-1} & -\mathcal{I}_\alpha^{-1}\mathcal{I}_{\alpha,\theta}\mathcal{I}_\theta^{-1} \\ 0 & \mathcal{I}_\theta^{-1} \end{pmatrix},
\tag{4}
$$

where

$$
V_\alpha^{\text{TMA}} = \mathcal{I}_\alpha^{-1} K_\alpha^{\text{TMA}} \mathcal{I}_\alpha^{-1}, \quad V_\theta^{\text{TMA}} = \mathcal{I}_\theta^{-1} + \mathcal{I}_\theta^{-1}\mathcal{I}_{\alpha,\theta}^\top \mathcal{I}_\alpha^{-1} K_\alpha^{\text{TMA}} \mathcal{I}_\alpha^{-1} \mathcal{I}_{\alpha,\theta} \mathcal{I}_\theta^{-1}, \quad V_{\alpha,\theta}^{\text{TMA}} = -\mathcal{I}_\alpha^{-1} K_\alpha^{\text{TMA}} \mathcal{I}_\alpha^{-1} \mathcal{I}_{\alpha,\theta} \mathcal{I}_\theta^{-1}.
$$

In Section 5 we illustrate the difference between (3) and (4) in practice by using simulated data. In Section 6, we illustrate the same difference by using real-life data.

4. Inference

Having established limiting normality of the two-stage ML estimator $\widetilde{\eta}$ in Proposition 1, we can derive limiting normality also for smooth parameter functions $\widetilde{\mu} = \mu(\widetilde{\eta})$, via the delta method. To use such results in practice we need consistent estimators of all variances.

The general formula for the variance matrix of the limiting distribution of $\widetilde{\eta}$ has the form $V_\eta = \mathcal{I}_\eta^{-1} K_\eta (\mathcal{I}_\eta^{-1})^\top$. These components are population quantities, defined as means and variance matrices of random variables. Consistent estimators for these components emerge generally speaking by using plug-in sample averages $\bar{h}_n = \sum_{i=1}^n h(y_i, \widetilde{\eta})/n$ for estimating a required $h_0 = \mathrm{E}_G\{h(y, \eta_0)\}$. For regularity conditions that secure convergence in probability of such $\bar{h}_n$ to h_0, see [22]. This general recipe leads to consistent estimators of $\mathcal{I}_\eta^{-1}, K_\eta$ above, and hence of $\widehat{V}_\eta = \widehat{\mathcal{I}}_\eta^{-1} \widehat{K}_\eta (\widehat{\mathcal{I}}_\eta^{-1})^\top$.

Consider now such a focus parameter $\mu = \mu(\eta) = \mu(\alpha, \theta)$, a parameter of primary interest, a smooth function of the model parameters. In addition to the ML estimator $\widehat{\mu} = \mu(\widehat{\alpha}, \widehat{\theta})$, we may define the associated two-stage ML estimator $\widetilde{\mu} = \mu(\widetilde{\alpha}, \widetilde{\theta})$, for which the delta method yields

$$
\sqrt{n}(\widetilde{\mu} - \mu_0) \rightsquigarrow \mathcal{N}(0, \tau^2),
\tag{5}
$$

with $\tau^2 = c^\top V_\eta c$, and $c = \partial\mu(\eta_0)/\partial\eta_0$. Then $\widetilde{\tau}^2 = \widetilde{c}^\top \widetilde{V}_\eta \widetilde{c}$ is consistent for τ^2, with $\widetilde{c} = \partial\mu(\widetilde{\eta})/\widetilde{\eta}$. This leads to confidence intervals, for any parameter of interest, as with

$$
\widetilde{\mu} \pm 1.96\, \widetilde{\tau}/\sqrt{n}
\tag{6}
$$

for the approximate 95% interval. There are different versions of (6), corresponding to different ways of applying the variance matrix formula V_η above. One may use components for $\widetilde{V}_\eta$ in a model-robust fashion, or the version where the parametric model assumption is trusted, leading to say $\widetilde{V}_\eta^{\text{TMA}}$ and to a consequent $\widetilde{\tau}^{\text{TMA}}$. Also, results similar to and in fact more familiar than those of (5) and (6) are easy to write down and use for the full ML estimation method, leading to $\widehat{\mu} \pm 1.96\, \widehat{\tau}/\sqrt{n}$ etc.

We note that full confidence distributions can be computed and displayed, to supplement the two-stage ML based point estimate $\widetilde{\mu}$ and estimated standard deviation $\widetilde{\tau}/\sqrt{n}$. These are random curves $\mathrm{cc}(\mu)$, one such for each focus parameter, constructed post data, with the property that $\Pr\{\mathrm{cc}(\mu) \le \alpha\}$ is equal to or approximately equal to α, for all confidence levels α. In addition to the easy to use first-order large-sample confidence curve

$$
\mathrm{cc}(\mu) = \Phi\{\sqrt{n}(\mu - \widetilde{\mu})/\widetilde{\tau}\},
$$

somewhat more elaborate and better approximations may be constructed via the methods in Chapters 3–4 of Schweder and Hjort [34], involving generalizations of the Wilks type theorems. In Section 6 such confidence curves are computed and displayed for relevant parameters pertaining to Norwegian precipitation data.

Our methodology lends itself nicely also to hypothesis testing. If $\mu = \mu(\alpha, \theta)$ is a parameter where a certain null value μ_{null} is of interest, then we may test $\mathcal{H}_{\text{null}}: \mu = \mu_{\text{null}}$ via inspection of the associated confidence interval, or via some fully or nearly equivalent route. This in particular applies for testing independence in the model structure, for the full d-dimensional vector or for a subset, if this corresponds to a null value for the θ parameter.

We may also use the developed machinery to test whether aspects of two or more sets of data are identical or different. If one has data from two groups thought to be not very different, say A and B, one may fit the same copula model to both, yielding parameter estimates for $\eta_A = (\alpha_A, \theta_A)$ and $\eta_B = (\alpha_B, \theta_B)$, along with variance estimators. With this taken care as explained above, one may then test the hypothesis $\mathcal{H}_0: \theta_A = \theta_B$. Specifically, with sample sizes n_A and n_B, we would have

$$
\widetilde{\theta}_A - \widetilde{\theta}_B \approx_d \mathcal{N}(\theta_A - \theta_B, W),
$$

say, with $W = \Sigma_A/n_A + \Sigma_B/n_B$. Here Σ_A and Σ_B are the variance matrices appearing in the limit distributions for $\sqrt{n_A}\,(\widetilde{\theta}_A - \theta_A)$ and $\sqrt{n_B}\,(\widetilde{\theta}_B - \theta_B)$. The test statistic

$$
Z = (\widetilde{\theta}_A - \widetilde{\theta}_B)^\top \widetilde{W}^{-1} (\widetilde{\theta}_A - \widetilde{\theta}_B),
$$

with $\widetilde{W}$ involving consistent estimators for Σ_A and Σ_B, would then follow an approximate χ_q^2 null distribution, with q the dimension of θ.

Table 1
Description of the models used in simulation 1.

	Copula	Margin 1	Margin 2	$\widehat{\text{KL}}(g, f)$
Data generating model	Gaussian $\theta = 0.3$	Weibull $\alpha_1 = (1.5, 4)^\top$ (shape, scale)	Gamma $\alpha_2 = (2, 1)^\top$ (shape, rate)	
Model 1	Gaussian	Weibull	Gamma	0
Model 2	Gaussian	Weibull	Generalized gamma	0
Model 3	Gaussian	Log-normal	Log-normal	0.140
Model 4	Student t	Weibull	Gamma	0.020
Model 5	Student t	Log-normal	Log-normal	0.166
Model 6	Frank	Weibull	Gamma	0.004
Model 7	Frank	Log-normal	Log-normal	0.139

Table 2
Description of the models used in simulation 2.

	Copula	Margin 1	Margin 2	Margin 3	Margin 4	$\widehat{\text{KL}}(g, f)$
Data generating model	Gumbel $\theta = 3$	Weibull $\alpha_1 = (1.5, 4)^\top$ (shape, scale)	Weibull $\alpha_2 = (2, 3)^\top$ (shape, scale)	Gamma $\alpha_3 = (2, 1)^\top$ (shape, rate)	Gamma $\alpha_4 = (3, 1)^\top$ (shape, rate)	
Model 1	Gumbel	Weibull	Weibull	Gamma	Gamma	0
Model 2	Gumbel	Weibull	Weibull	Generalized Gamma	Generalized Gamma	0
Model 3	Gumbel	Log-normal	Log-normal	Log-normal	Log-normal	0.298
Model 4	Survival Clayton	Weibull	Weibull	Gamma	Gamma	0.433
Model 5	Survival Clayton	Log-normal	Log-normal	Log-normal	Log-normal	0.508
Model 6	Frank	Weibull	Weibull	Gamma	Gamma	0.263
Model 7	Frank	Log-normal	Log-normal	Log-normal	Log-normal	0.558

5. Simulation study

To study the impact of the true model assumption on two-stage ML estimation, we have performed a set of simulations. In simulation 1, we study the impact of the true model assumption on a number of 2-dimensional copula models. In simulation 2, we do the same with 4-dimensional copula models. In simulation 3, we look at the case where the degree of misspecification increases or decreases gradually.

5.1. Simulation 1 (2-dimensional case) and simulation 2 (4-dimensional case)

Tables 1 and 2 contain descriptions of the models that were used to generate the data and the models that were used to fit the data in simulation 1 and simulation 2, respectively. The candidate models are chosen in such a way that they illustrate different types and degrees of misspecification. The column $\widehat{\text{KL}}(g, f)$ indicates estimated Kullback–Leibler divergence with ML estimated least false parameter values. We used two sample sizes $n = 100$ and $n = 1000$. For each sample size, we drew 1000 samples and the results were averaged. Some of the candidate models are easily distinguishable from the data generating model for a sample size 1000, based on likelihood-based model selection criteria such as AIC. However, since our goal is to investigate different degrees of model misspecification, we included them in our simulation.

For each model, ML and two-stage ML estimations were performed both assuming and not assuming that the model is the true model that generated data. The column 'True model assumption' in Tables 3–6 indicates the presence of this assumption. Dropping the true model assumption leads to the model robust asymptotic variance formulas. For the ML estimator, this is the so-called 'sandwich estimator'. For a copula model, this estimator can be obtained straightforwardly by applying classical theory covered in [8,17,38]. When we make the true model assumption, the sandwich estimator simplifies to the inverse of the Fisher information. For two-stage ML estimators, dropping the true model assumption yields (3) as asymptotic variance, and assuming the true model, this simplifies into (4).

Below, $\widehat{\theta}$ and $\widetilde{\theta}$ indicate the ML and the two-stage ML estimate of the copula parameter, respectively. Also, $\sqrt{n}\,\text{SE}(\widehat{\theta})$ is the square root of the estimated asymptotic variance of $\widehat{\theta}$. It is estimated by choosing the relevant formula, depending on the presence of true model assumption, and, in general terms, replacing $\text{E}_G\{h(y, \eta_0)\}$ by $\sum_{i=1}^n h(y_i, \widehat{\eta})/n$ or $\sum_{i=1}^n h(y_i, \widetilde{\eta})/n$; see the discussion in Section 4.

Next, let $q_{0.8}$ indicate the vector that contains the 0.8-quantile value of each margin according to the data generating model, i.e., $q_{0.8} = (G_1^{-1}(0.8), \ldots, G_d^{-1}(0.8))$. With $\Pr(q_{0.8} < Y)$ we mean the joint probability that each marginal variable has larger value than the corresponding 0.8-quantile value. This joint upper probability is sensitive to the tail behavior of both margins and copula. The notations $\widehat{\Pr}(q_{0.8} < Y)$ and $\widetilde{\Pr}(q_{0.8} < Y)$ refer to ML and two-stage ML estimates of this joint probability, respectively. Also, $\sqrt{n}\,\text{SE}(\widehat{\Pr}(q_{0.8} < Y))$ is the square root of the estimated asymptotic variance of $\widehat{\Pr}(q_{0.8} < Y)$, and $\sqrt{n}\,\text{SE}(\widetilde{\Pr}(q_{0.8} < Y))$ is the two-stage ML estimator analog. These asymptotic variances are obtained by

Table 3
Result from simulation 1 with sample size 100. For clarification of the column labels, see Section 5.1.

Simulation 1 ($n = 100$)

	True model assumption	MLE $\widehat{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widehat{\theta})$	$\widehat{\Pr}(q_{0.8} < Y)$	(95% CI)	Coverage $\Pr(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widehat{\Pr}(q_{0.8} < Y))$
Model 1	No	0.298	(0.107, 0.489)	0.951	0.974	0.066	(0.048, 0.085)	0.959	0.094
	Yes		(0.121, 0.475)	0.939	0.902		(0.049, 0.083)	0.939	0.087
Model 2	No	0.299	(0.109, 0.489)	0.952	0.970	0.066	(0.048, 0.085)	0.954	0.094
	Yes		(0.122, 0.476)	0.939	0.904		(0.049, 0.083)	0.937	0.087
Model 3	No	0.298	(0.095, 0.500)	0.961	1.032	0.066	(0.047, 0.086)	0.963	0.100
	Yes		(0.121, 0.474)	0.929	0.902		(0.049, 0.083)	0.930	0.087
Model 4	No	0.297	(0.098, 0.496)	n/a	1.016	0.069	(0.050, 0.088)	0.943	0.096
	Yes		(0.102, 0.491)	n/a	0.992		(0.050, 0.087)	0.942	0.094
Model 5	No	0.303	(0.091, 0.514)	n/a	1.079	0.069	(0.049, 0.089)	0.947	0.103
	Yes		(0.102, 0.503)	n/a	1.022		(0.050, 0.088)	0.938	0.098
Model 6	No	1.844	(0.504, 3.183)	n/a	6.835	0.065	(0.047, 0.083)	0.946	0.092
	Yes		(0.597, 3.090)	n/a	6.358		(0.048, 0.082)	0.932	0.086
Model 7	No	2.003	(0.527, 3.480)	n/a	7.535	0.067	(0.047, 0.087)	0.950	0.101
	Yes		(0.682, 3.325)	n/a	6.742		(0.049, 0.085)	0.921	0.090
	True model assumption	Two-stage MLE $\widetilde{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widetilde{\theta})$	$\widetilde{\Pr}(q_{0.8} < Y)$	(95% CI)	Coverage $\Pr(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widetilde{\Pr}(q_{0.8} < Y))$
Model 1	No	0.298	(0.107, 0.488)	0.950	0.973	0.066	(0.048, 0.085)	0.957	0.094
	Yes		(0.121, 0.475)	0.939	0.903		(0.049, 0.083)	0.938	0.087
Model 2	No	0.297	(0.108, 0.487)	0.951	0.967	0.066	(0.048, 0.085)	0.955	0.093
	Yes		(0.120, 0.475)	0.941	0.905		(0.049, 0.083)	0.938	0.087
Model 3	No	0.298	(0.095, 0.500)	0.961	1.032	0.066	(0.047, 0.086)	0.963	0.100
	Yes		(0.117, 0.478)	0.936	0.923		(0.049, 0.084)	0.936	0.089
Model 4	No	0.296	(0.101, 0.491)	n/a	0.995	0.069	(0.050, 0.087)	0.943	0.094
	Yes		(0.103, 0.490)	n/a	0.986		(0.050, 0.087)	0.942	0.094
Model 5	No	0.302	(0.094, 0.510)	n/a	1.062	0.069	(0.049, 0.089)	0.950	0.101
	Yes		(0.098, 0.505)	n/a	1.038		(0.049, 0.088)	0.947	0.099
Model 6	No	1.828	(0.506, 3.150)	n/a	6.746	0.065	(0.047, 0.083)	0.946	0.091
	Yes		(0.583, 3.072)	n/a	6.350		(0.048, 0.082)	0.935	0.086
Model 7	No	1.974	(0.534, 3.415)	n/a	7.351	0.067	(0.047, 0.086)	0.947	0.099
	Yes		(0.638, 3.311)	n/a	6.819		(0.049, 0.085)	0.930	0.092

writing $\Pr(q_{0.8} < Y)$ as a function of η and applying the delta method. With those standard errors, we also constructed confidence intervals. The reported confidence intervals in Tables 3–6 are obtained by averaging the confidence intervals from 1000 samples.

In the cases where the true value of θ or $\Pr(q_{0.8} < Y)$ was known, we also computed coverage to supplement the confidence interval. The coverage is computed as the number of times that the confidence interval captured the true value divided by the total number of simulations.

When estimating the above mentioned quantities, the main computational bottleneck is estimating the K-matrices. For instance, the matrix

$$\mathcal{I}_\theta = -\int g\,\frac{\partial^2}{\partial\theta_0\partial\theta_0^\top}\,\ln c\{F_1(y_1, \alpha_{0,1}), \ldots, F_d(y_d, \alpha_{0,d}), \theta\}\,\mathrm{d}y$$

is estimated by

$$\widetilde{\mathcal{I}}_\theta = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta\partial\theta^\top}\,\ln c\{F_1(y_{i,1}, \widetilde{\alpha}_1), \ldots, F_d(y_{i,d}, \widetilde{\alpha}_d), \widetilde{\theta}\}.$$

Table 4
Result from simulation 1 with sample size 1000. For clarification of the column labels, see Section 5.1.

Simulation 1 ($n = 1000$)

	True model assumption	MLE							
		$\widehat{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widehat{\theta})$	$\widehat{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	Coverage $\mathrm{Pr}(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widehat{\mathrm{Pr}}(q_{0.8} < Y))$
Model 1	No	0.298	(0.241, 0.356)	0.950	0.930	0.066	(0.060, 0.072)	0.954	0.090
	Yes		(0.242, 0.355)	0.947	0.908		(0.061, 0.071)	0.948	0.088
Model 2	No	0.299	(0.241, 0.356)	0.953	0.929	0.066	(0.060, 0.072)	0.954	0.090
	Yes		(0.242, 0.355)	0.947	0.909		(0.061, 0.071)	0.949	0.088
Model 3	No	0.291	(0.230, 0.352)	0.942	0.987	0.065	(0.059, 0.071)	0.938	0.095
	Yes		(0.234, 0.347)	0.927	0.914		(0.060, 0.071)	0.919	0.088
Model 4	No	0.299	(0.241, 0.356)	n/a	0.930	0.067	(0.061, 0.072)	0.950	0.089
	Yes		(0.241, 0.356)	n/a	0.930		(0.061, 0.072)	0.951	0.090
Model 5	No	0.294	(0.233, 0.355)	n/a	0.985	0.066	(0.060, 0.072)	0.939	0.094
	Yes		(0.234, 0.353)	n/a	0.964		(0.061, 0.072)	0.936	0.092
Model 6	No	1.806	(1.407, 2.204)	n/a	6.422	0.064	(0.059, 0.070)	0.914	0.088
	Yes		(1.417, 2.194)	n/a	6.269		(0.059, 0.070)	0.905	0.086
Model 7	No	1.958	(1.520, 2.397)	n/a	7.076	0.067	(0.061, 0.073)	0.944	0.096
	Yes		(1.544, 2.372)	n/a	6.674		(0.061, 0.072)	0.932	0.091
	True model assumption	Two-stage MLE							
		$\widetilde{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widetilde{\theta})$	$\widetilde{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	Coverage $\mathrm{Pr}(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widetilde{\mathrm{Pr}}(q_{0.8} < Y))$
Model 1	No	0.298	(0.241, 0.356)	0.950	0.929	0.066	(0.060, 0.072)	0.954	0.090
	Yes		(0.242, 0.355)	0.947	0.909		(0.061, 0.071)	0.948	0.088
Model 2	No	0.298	(0.241, 0.356)	0.952	0.929	0.066	(0.060, 0.072)	0.954	0.090
	Yes		(0.242, 0.355)	0.948	0.909		(0.061, 0.071)	0.949	0.088
Model 3	No	0.291	(0.230, 0.352)	0.942	0.987	0.065	(0.059, 0.071)	0.938	0.095
	Yes		(0.233, 0.349)	0.931	0.938		(0.060, 0.071)	0.926	0.090
Model 4	No	0.298	(0.241, 0.356)	n/a	0.929	0.067	(0.061, 0.072)	0.949	0.089
	Yes		(0.241, 0.356)	n/a	0.930		(0.061, 0.072)	0.950	0.090
Model 5	No	0.294	(0.233, 0.354)	n/a	0.983	0.066	(0.060, 0.072)	0.940	0.094
	Yes		(0.232, 0.355)	n/a	0.988		(0.060, 0.072)	0.942	0.095
Model 6	No	1.801	(1.405, 2.197)	n/a	6.385	0.064	(0.059, 0.070)	0.913	0.088
	Yes		(1.413, 2.189)	n/a	6.260		(0.059, 0.070)	0.905	0.086
Model 7	No	1.944	(1.511, 2.376)	n/a	6.982	0.066	(0.060, 0.072)	0.946	0.095
	Yes		(1.525, 2.362)	n/a	6.748		(0.061, 0.072)	0.940	0.092

When calculating this matrix, a for-loop can be avoided by swapping the order of summation and differentiation. When computing

$$\widetilde{K}_\theta = \frac{1}{n} \sum_{i=1}^{n} U_\theta(y_i, \widetilde{\alpha}, \widetilde{\theta}) U_\theta(y_i, \widetilde{\alpha}, \widetilde{\theta})^\top,$$

however, the same trick cannot be used and we are forced to use a time-consuming for-loop.

Now we look at the results from the simulations. From the result of simulations 1 and 2, we can observe that models that are correctly specified or have a small degree of misspecification (e.g., Student t-copula instead of Gaussian copula and Generalized Gamma distribution instead of Gamma distribution), the presence of the true model assumption makes virtually no difference. This confirms the finding from Section 3 that when the model is correctly specified, the model robust variance formula (3) and the non-robust version (4) coincide. Further, for the ML estimation part, it echoes the classical finding that the 'sandwich estimator' becomes the inverse of the Fisher information when the model is correctly specified.

When the degree of misspecification increases (e.g., models 5 and 7 in simulation 2 where the log-normal density has a heavier tail than the data generating margins and the Frank copula has lighter tails than the Gumbel copula), we see that assuming the true model (i.e., using (4) instead of (3)) results in a larger difference in estimated the asymptotic variance. Often, assuming the true model results in a smaller asymptotic variance. This sounds intuitive because assuming the true model entails ignoring the model uncertainty and thus reduces the variance. In other words, ignoring the model uncertainty often leads to over-confident confidence intervals. The degree of over-confidence will be larger when the sample size is small, because the asymptotic variance is divided by the sample size when a confidence interval is computed. This also implies that

Table 5
Result from simulation 2 with sample size 100. For clarification of the column labels, see Section 5.1.

Simulation 2 ($n = 100$)

	True model assumption	MLE $\overline{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widehat{\theta})$	$\widehat{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	Coverage $\mathrm{Pr}(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widehat{\mathrm{Pr}}(q_{0.8} < Y))$
Model 1	No	3.022	(2.582, 3.462)	0.95	2.246	0.132	(0.122, 0.142)	0.943	0.050
	Yes		(2.571, 3.473)	0.954	2.302		(0.122, 0.142)	0.953	0.051
Model 2	No	3.024	(2.552, 3.496)	0.95	2.407	0.132	(0.122, 0.143)	0.944	0.054
	Yes		(2.570, 3.478)	0.953	2.315		(0.122, 0.142)	0.944	0.051
Model 3	No	2.869	(2.447, 3.291)	0.863	2.153	0.128	(0.118, 0.139)	0.905	0.053
	Yes		(2.469, 3.268)	0.858	2.039		(0.119, 0.138)	0.905	0.050
Model 4	No	2.867	(2.010, 3.724)	n/a	4.374	0.123	(0.105, 0.140)	0.855	0.089
	Yes		(2.108, 3.626)	n/a	3.874		(0.107, 0.138)	0.797	0.078
Model 5	No	3.907	(2.793, 5.021)	n/a	5.683	0.139	(0.125, 0.154)	0.755	0.072
	Yes		(2.970, 4.843)	n/a	4.779		(0.127, 0.151)	0.674	0.061
Model 6	No	11.160	(8.667, 13.652)	n/a	12.715	0.107	(0.092, 0.122)	0.136	0.077
	Yes		(9.375, 12.944)	n/a	9.106		(0.096, 0.118)	0.040	0.056
Model 7	No	13.839	(10.242, 17.436)	n/a	18.350	0.121	(0.104, 0.138)	0.815	0.086
	Yes		(11.421, 16.258)	n/a	12.340		(0.110, 0.132)	0.545	0.058

	True model assumption	Two-stage MLE $\widetilde{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widetilde{\theta})$	$\widetilde{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	Coverage $\mathrm{Pr}(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widetilde{\mathrm{Pr}}(q_{0.8} < Y))$
Model 1	No	3.001	(2.514, 3.487)	0.931	2.484	0.131	(0.121, 0.142)	0.941	0.056
	Yes		(2.508, 3.493)	0.936	2.512		(0.120, 0.143)	0.941	0.057
Model 2	No	2.989	(2.502, 3.475)	0.935	2.482	0.131	(0.120, 0.142)	0.939	0.056
	Yes		(2.493, 3.484)	0.944	2.527		(0.120, 0.142)	0.942	0.058
Model 3	No	2.980	(2.524, 3.435)	0.947	2.323	0.131	(0.121, 0.141)	0.957	0.053
	Yes		(2.549, 3.411)	0.938	2.199		(0.121, 0.141)	0.941	0.050
Model 4	No	2.619	(1.852, 3.386)	n/a	3.912	0.117	(0.100, 0.135)	0.671	0.090
	Yes		(1.893, 3.346)	n/a	3.706		(0.101, 0.134)	0.621	0.085
Model 5	No	3.180	(2.378, 3.983)	n/a	4.094	0.129	(0.115, 0.143)	0.940	0.072
	Yes		(2.426, 3.934)	n/a	3.846		(0.116, 0.142)	0.913	0.067
Model 6	No	9.884	(7.994, 11.775)	n/a	9.646	0.099	(0.085, 0.112)	0	0.068
	Yes		(8.322, 11.447)	n/a	7.974		(0.088, 0.110)	0	0.056
Model 7	No	9.637	(7.652, 11.623)	n/a	10.130	0.097	(0.083, 0.111)	0.001	0.073
	Yes		(8.109, 11.165)	n/a	7.797		(0.086, 0.108)	0	0.056

when hypothesis testing is carried out on the copula parameter or on a function that takes the copula parameter as an input, the choice between the model-robust and model non-robust variance formula can lead to different outcomes of the test.

Although the true model assumption often results in a smaller variance of the copula parameter, there are cases when this does not hold, e.g., model 1 from simulation 3 in Section 5.2 and the model for precipitation data in Section 6. This is in line with the fact that we could not find any analytical evidence for the inequality $V_\theta^{\mathrm{TMA}} < V_\theta$. In addition, the true model assumption changes the interpretation of $\widetilde{\theta}$ from 'estimate of the least false parameter value' to 'estimate of the true value that generated the data'.

Joe [20] carried out an extensive study on the asymptotic efficiency for two-stage ML estimators and concluded that two-stage estimation is highly efficient in most cases. By comparing the results from ML and two-stage ML estimations, we can largely confirm that this is also the case when the model robust variance formula is used, particularly when the sample size n is large, as the two variance matrix estimators involved are estimating the same quantity.

Another notable result regarding efficiency is that the Hájek–Le Cam convolution theorem, and related theorems on the optimality of ML estimation, do not apply when the model is incorrectly specified. The Hájek–Le Cam convolution theorem [15], combined with the Cramér–Rao lower bound [30] theory, states that no asymptotically unbiased and normal competing estimator can have a smaller limiting variance matrix than that of the full ML estimator. Models 5 and 7 in both the two- and four-dimensional cases have a relatively large degree of model misspecification. For these models, it frequently happens that the asymptotic variance of the two-stage ML estimator is smaller than that of the full ML estimator. This occurs both when the model-robust and model non-robust variance formulas are used. The largest difference occurs for model 7 from simulation 2, where the degree of model misspecification is high (the copula and all four margins are misspecified).

Table 6

Result from simulation 2 with sample size 1000. For clarification of the column labels, see Section 5.1.

Simulation 2 ($n = 1000$)

	True model assumption	MLE							
		$\widehat{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widehat{\theta})$	$\widehat{\Pr}(q_{0.8} < Y)$	(95% CI)	Coverage $\Pr(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widehat{\Pr}(q_{0.8} < Y))$
Model 1	No	3.002	(2.861, 3.143)	0.947	2.271	0.132	(0.129, 0.135)	0.946	0.051
	Yes		(2.861, 3.143)	0.949	2.278		(0.129, 0.135)	0.949	0.051
Model 2	No	3.002	(2.848, 3.156)	0.950	2.482	0.132	(0.129, 0.135)	0.945	0.056
	Yes		(2.860, 3.144)	0.951	2.287		(0.129, 0.135)	0.948	0.051
Model 3	No	2.853	(2.713, 2.993)	0.460	2.255	0.128	(0.125, 0.132)	0.500	0.056
	Yes		(2.728, 2.978)	0.382	2.019		(0.125, 0.132)	0.409	0.050
Model 4	No	2.785	(2.516, 3.055)	n/a	4.346	0.122	(0.116, 0.128)	0.050	0.092
	Yes		(2.551, 3.019)	n/a	3.780		(0.117, 0.127)	0.024	0.080
Model 5	No	3.862	(3.487, 4.237)	n/a	6.049	0.140	(0.135, 0.144)	0.133	0.078
	Yes		(3.569, 4.155)	n/a	4.728		(0.136, 0.143)	0.064	0.061
Model 6	No	11.067	(10.234, 11.900)	n/a	13.436	0.107	(0.102, 0.112)	0	0.083
	Yes		(10.509, 11.624)	n/a	8.997		(0.103, 0.110)	0	0.056
Model 7	No	13.777	(12.619, 14.934)	n/a	18.677	0.121	(0.116, 0.127)	0.019	0.088
	Yes		(13.019, 14.535)	n/a	12.227		(0.118, 0.125)	0.002	0.057

	True model assumption	Two-stage MLE							
		$\widetilde{\theta}$	(95% CI)	Coverage θ	$\sqrt{n}\,\mathrm{SE}(\widetilde{\theta})$	$\widetilde{\Pr}(q_{0.8} < Y)$	(95% CI)	Coverage $\Pr(q_{0.8} < Y)$	$\sqrt{n}\,\mathrm{SE}(\widetilde{\Pr}(q_{0.8} < Y))$
Model 1	No	2.999	(2.842, 3.156)	0.960	2.532	0.132	(0.128, 0.135)	0.962	0.057
	Yes		(2.842, 3.156)	0.964	2.537		(0.128, 0.135)	0.960	0.057
Model 2	No	2.997	(2.840, 3.153)	0.961	2.527	0.132	(0.128, 0.135)	0.961	0.057
	Yes		(2.839, 3.154)	0.961	2.539		(0.128, 0.135)	0.960	0.057
Model 3	No	2.987	(2.838, 3.136)	0.943	2.399	0.132	(0.128, 0.135)	0.945	0.054
	Yes		(2.846, 3.128)	0.932	2.277		(0.128, 0.135)	0.938	0.052
Model 4	No	2.586	(2.339, 2.833)	n/a	3.985	0.117	(0.112, 0.123)	0	0.093
	Yes		(2.357, 2.816)	n/a	3.703		(0.112, 0.123)	0	0.086
Model 5	No	3.178	(2.915, 3.441)	n/a	4.248	0.129	(0.125, 0.134)	0.839	0.074
	Yes		(2.937, 3.419)	n/a	3.887		(0.125, 0.134)	0.791	0.068
Model 6	No	9.856	(9.254, 10.459)	n/a	9.723	0.099	(0.095, 0.103)	0	0.069
	Yes		(9.358, 10.355)	n/a	8.040		(0.095, 0.102)	0	0.057
Model 7	No	9.647	(9.007, 10.288)	n/a	10.337	0.097	(0.093, 0.102)	0	0.075
	Yes		(9.153, 10.142)	n/a	7.980		(0.094, 0.101)	0	0.058

When it comes to the asymptotic variance of the upper tail probability $\Pr(q_{0.8} < Y)$, we often observe that the true model assumption decreases the asymptotic variance. Yet, in some cases, we observe that the true model assumption increases the asymptotic variance. This happens mostly when the true model assumption increases the asymptotic variance of θ, but there are also cases where the true model assumption increases the asymptotic $\Pr(q_{0.8} < Y)$ while it decreases the asymptotic variance of θ. We suspect that this is due to the fact that the whole V_η, including its non-diagonal elements, is used through the delta method formula to compute $V_{\Pr(q_{0.8}<Y)}$. To check whether this also happens in non-copula models, we simulated a large number of univariate data and fitted well-known distributions and computed upper tail probability from them. We could observe that the true model assumption can increase the limiting variance of the upper tail probability while the limiting variance of parameters decreases. Further, from the same univariate simulation, we also observed that using the true model assumption sometimes increases the limiting variance of parameters.

5.2. Simulation 3 (gradual change in misspecification)

To study the difference in the numerical behavior between (4) and (3) in more detail, including the cases where the degree of misspecification is very small, we ran a simulation with gradual change in misspecification. There are three scenarios and for each scenario, we generated a dataset of size 100. The data generating algorithm generated $(1 − \delta) \times 100\%$ of data points from model 1 and $\delta \times 100\%$ data points from model 2. So, δ can be seen as the degree of 'contamination' to model 1 by model 2. Then, we fitted both model 1 and model 2 to these data by using two-stage ML. This process was repeated 100,000 times and the results were averaged. We used 100 different values of δ, which were spread evenly between 0 and 1. Table 7 describes the model 1 and model 2 that were used in each scenario. In scenario 1, model 1 gets contaminated only in the

Table 7
Description of the scenarios used in simulation 3.

		Copula	Margin 1	Margin 2
Scenario 1	Model 1	Gaussian $\theta = 0.3$	Weibull $\alpha_1 = (1, 3)^\top$ (shape, scale)	Gamma $\alpha_2 = (2, 1)^\top$ (shape, rate)
	Model 2	Gaussian $\theta = 0.3$	Log-normal $\alpha_1 = (0.5, 1.3)^\top$ (mean, SD)	Log-normal $\alpha_2 = (0.4, 0.8)^\top$ (mean, SD)
Scenario 2	Model 1	Gaussian $\theta = 0.3$	Weibull $\alpha_1 = (1, 3)^\top$ (shape, scale)	Gamma $\alpha_2 = (2, 1)^\top$ (shape, rate)
	Model 2	Frank $\theta = 3$	Weibull $\alpha_1 = (1, 3)^\top$ (shape, scale)	Gamma $\alpha_2 = (2, 1)^\top$ (shape, rate)
Scenario 3	Model 1	Gaussian $\theta = 0.3$	Weibull $\alpha_1 = (1, 3)^\top$ (shape, scale)	Gamma $\alpha_2 = (2, 1)^\top$ (shape, rate)
	Model 2	Frank $\theta = 3$	Log-normal $\alpha_1 = (0.5, 1.3)^\top$ (mean, SD)	Log-normal $\alpha_2 = (0.4, 0.8)^\top$ (mean, SD)

Table 8
Description of the model for precipitation data.

	Copula	Margin 1 (Vestby)	Margin 2 (Ski)	Margin 3 (Lørenskog)	Margin 4 (Nannestad)	Margin 5 (Hurdal)
Model	Gumbel	Gamma	Gamma	Log-normal	Gamma	Log-normal

margins. In scenario 2, model 1 gets contaminated only in the copula. In scenario 3, model 1 gets contaminated both in copula and margins. For each model and scenario, we estimated the asymptotic variance of θ and $\Pr(q_{0.8} < Y)$ by assuming the true model and using Eq. (4) and by not assuming the true model and using Eq. (3).

Figs. 1, 2 and 3 show the results of scenarios 1, 2 and 3, respectively. We can observe that the true model assumption can both decrease and increase the estimated asymptotic variance of θ and $\Pr(q_{0.8} < Y)$. And it seems like it is model-specific. For example, model 1, which is same across all three scenarios, always shows increased variance of copula parameter θ under the true model assumption, regardless of what kind of contamination it gets.

A general pattern that we can clearly observe in all scenarios is that the difference between the two asymptotic variance formulas (4) and (3), regardless of whether it is positive or negative, gets smaller as the degree of misspecification decreases. As expected, this pattern is most clearly visible in scenario 3 where both copula and margins are misspecified.

6. Precipitation data

The precipitation data consist of daily measurements of precipitation in millimeters at five different meteorological stations in Norway (Vestby, Ski, Lørenskog, Nannestad and Hurdal) from January 1, 1990 to December 31, 2006. These data were provided by the Norwegian Meteorological Institute and used previously in Aas and Berg [1] (with one station less) and Hobæk Haff [18]. Following the example of two previous papers that used these data, we modeled only positive precipitation by removing all observations for which at least one of the stations has recorded zero precipitation, resulting in 5536 observations. The main advantage of this is that we remove time dependence.

To choose an adequate model, we first fitted a set of well-known distributions for each margin by using ML estimation and evaluated them by AIC. After the marginal distributions were chosen, we performed a probability integral transformation for each margin and fitted a set of well-known copulas by using ML estimation and evaluated these by AIC. The best model obtained in this fashion is described in Table 8. The choice of the Gumbel copula corresponds well with the fact that there are indications of strong upper, but not lower, tail dependence. This is visible in Figure 2 from the Online Supplement [26], which contains simulated scatter plots between the first two variables of the 5-dimensional Gumbel copula. Since the Gumbel copula is a member of the Archimedean copula family, the scatter plots between other possible combinations among the five variables will be virtually the same as the one between variable 1 and variable 2. They are therefore not displayed.

Figure 1 from the Online Supplement [26] shows that the marginal distributions of the fitted model are highly non-normal and can differ between meteorological stations.

Table 9 shows the result from the model described in Table 8. With ML estimation, ignoring model uncertainty gives smaller values of the asymptotic variances. However, with two-stage ML estimators, this is not the case. This phenomenon was discussed earlier in Section 5.
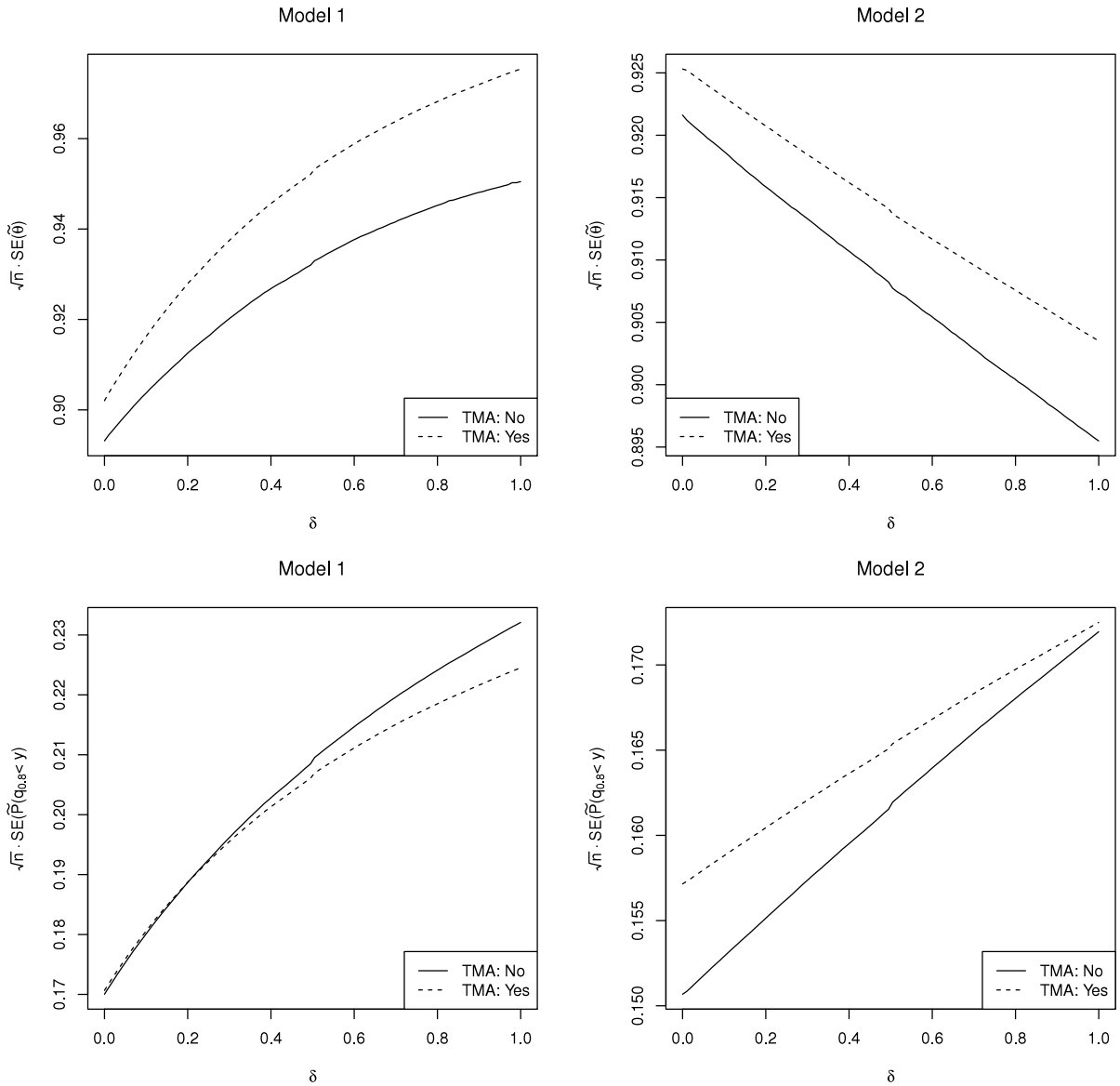
Fig. 1. Result of simulation 3, scenario 1. (Models 1 and 2 have the same copula, but different margins.) TMA stands for 'true model assumption'. The dashed lines are computed with the model non-robust asymptotic variance formula (4) and the solid lines are computed with the model robust asymptotic variance formula (3).

Furthermore, we observe that the asymptotic variance of the two-stage ML estimator is smaller than that of the ML estimator. As already discussed in Section 5, this happens when the degree of misspecification is high and consequently the Hájek–Le Cam convolution theorem is not applicable. In our case, we used the five-dimensional Gumbel copula, which is a member of the Archimedean copula family, to model the precipitation in five locations. The advantage of the Archimedean family copulas is that we can model high-dimensional dependency by using only one parameter. The disadvantage, especially when dimension is high, is that the dependency structure and strength is the same between all different pairs of variables. The dependency relationship between u_1 and u_2, for example, is the same as the dependency relationship between u_3 and u_5, in such models. Here $u_j = F_j(y_j)$, for $j \in \{1, \ldots, 5\}$, and these are uniform. The pairwise pseudo-observations plot (Figure 3 in Supplement B of Hobæk Haff [18]), however, shows that the pairwise dependencies among the stations are not the same.

This rigidness of Archimedean copulas is causing a misspecification, and therefore we see that the asymptotic variance of two-stage ML estimators is smaller than that of the full ML. Aas and Berg [1] suggest pair-copula constructions as a method to overcome this limitation.

Fig. 4 visualizes Table 9 by using confidence curves, see Section 4. For details about confidence curves, see Schweder and Hjort [34]. In Fig. 4, the confidence intervals from the model non-robust formula (dashed curves) and the confidence
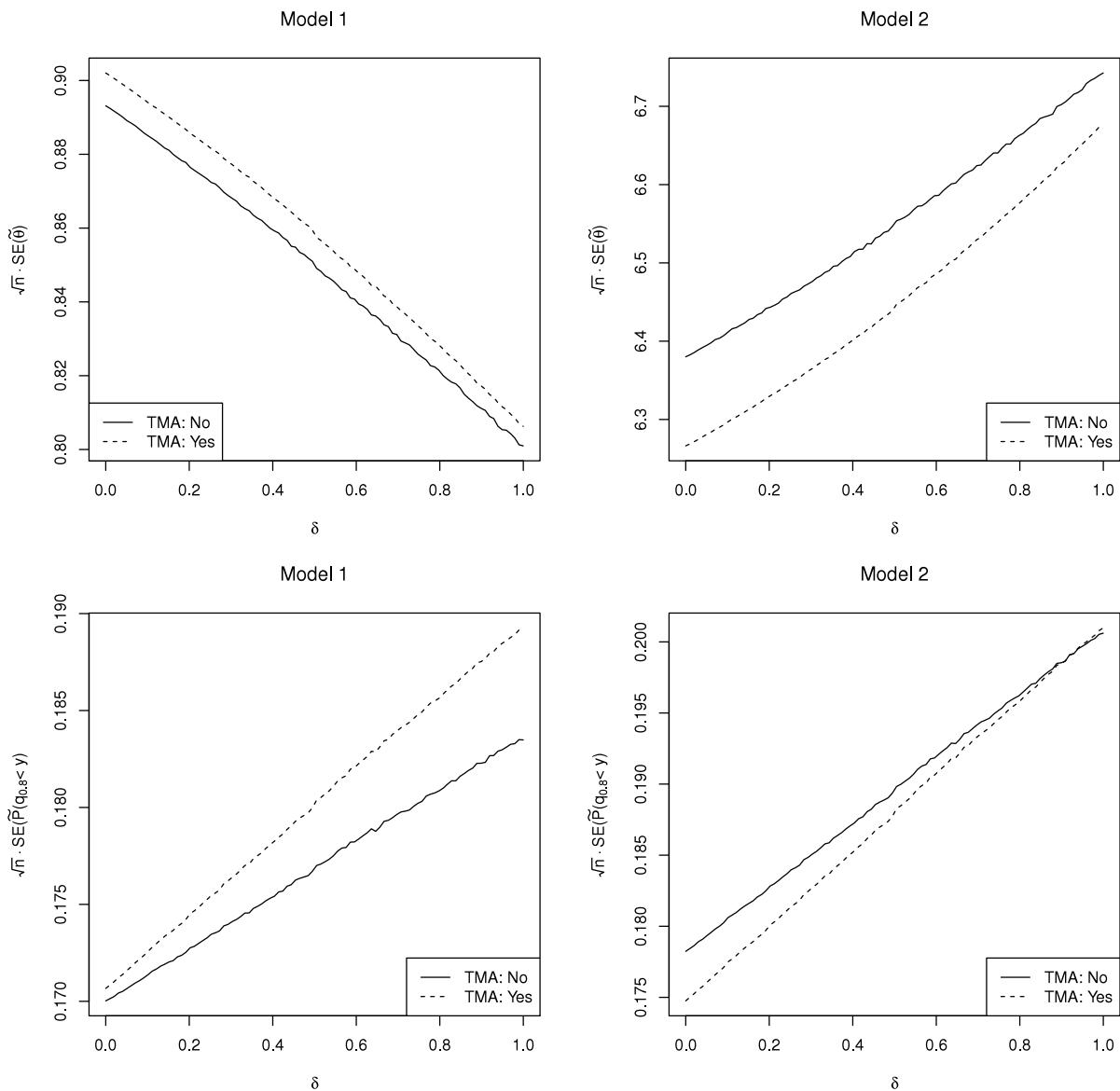
Fig. 2. Result of simulation 3, scenario 2. (Models 1 and 2 have the same margins, but different copulas.) TMA stands for 'true model assumption'. The dashed lines are computed with the model non-robust asymptotic variance formula (4) and the solid lines are computed with the model robust asymptotic variance formula (3).

intervals from the model robust formula (non-dashed lines) are quite close to each other. This is partly due to the fact that the sample size ($n = 5536$) is relatively large and thus scales down the asymptotic variance when a confidence interval is computed. For datasets with smaller number of observations, the choice of model (non-)robust formula will make a bigger difference.

The joint upper tail probability $\Pr(q_{0.8} < Y)$ can be interpreted as the probability that, given that there is precipitation in all five locations, there is precipitation higher than the 0.8-quantile values in all five locations within the same day.

7. Comparing estimation schemes

Methods and results reached in the present paper may be exploited for examining and assessing the impact of model misspecification for different types of outcomes. Some quantities will be more affected by misspecification than others, depending on the particularities of the models entertained, the underlying data generating mechanism, and also the sample size. It is also a consequence of our results that the two-stage based estimator may have lower variance than the full ML
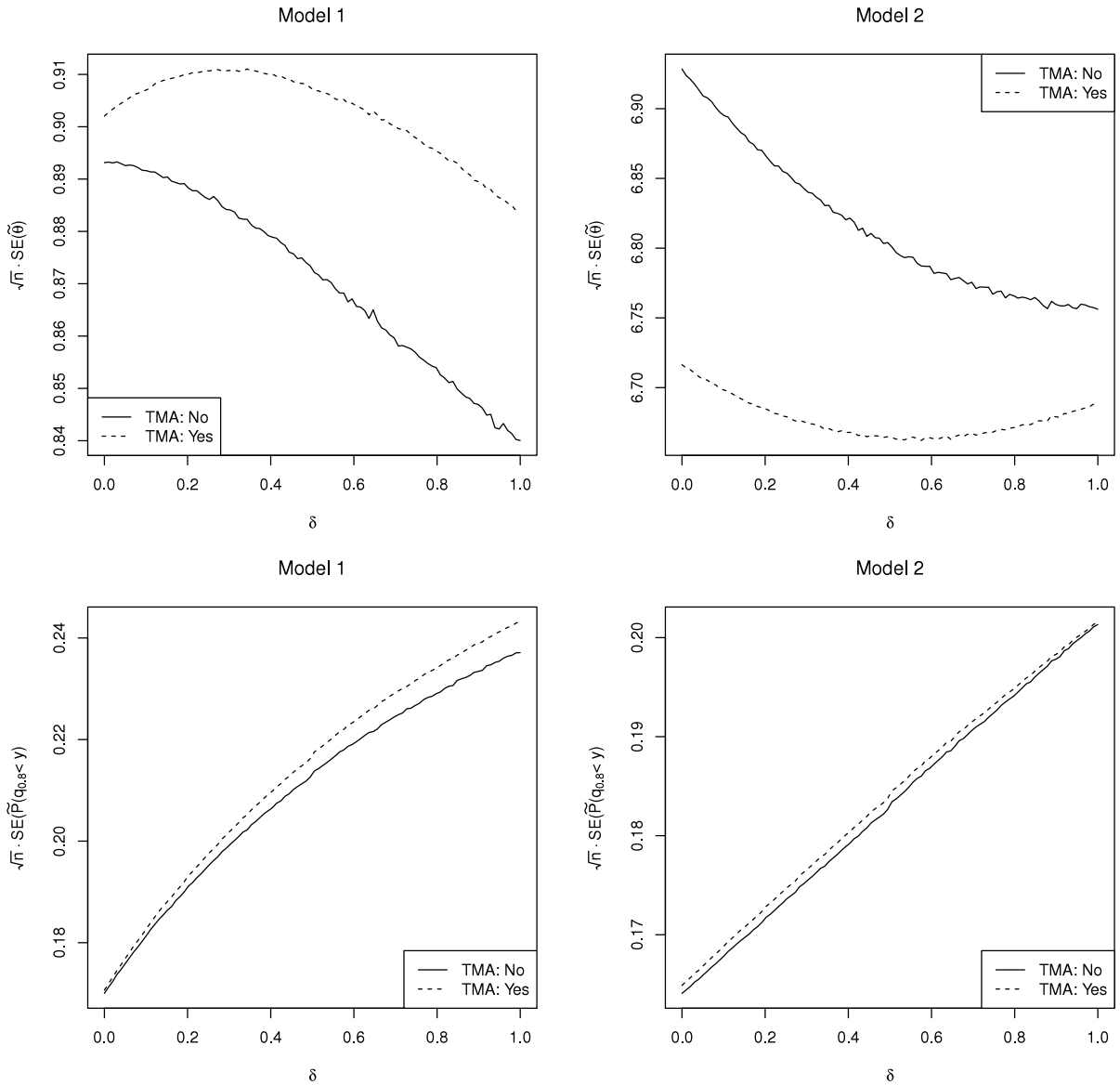
Fig. 3. Result of simulation 3, scenario 3. (Models 1 and 2 have different margins and different copulas.) TMA stands for 'true model assumption'. The dashed lines are computed with the model non-robust asymptotic variance formula (4) and the solid lines are computed with the model robust asymptotic variance formula (3).

Table 9
Result from the model described in Table 8 fitted with the precipitation data. For the description of column labels, see Section 5.1.

True model assumption	MLE					
	$\widehat{\theta}$	(95% CI)	$\sqrt{n}\,\mathrm{SE}(\widehat{\theta})$	$\widehat{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	$\sqrt{n}\,\mathrm{SE}(\widehat{\mathrm{Pr}}(q_{0.8} < Y))$
No	2.351	(2.299, 2.404)	1.996	0.108	(0.101, 0.114)	0.246
Yes		(2.304, 2.398)	1.796		(0.102, 0.114)	0.226
True model assumption	Two-stage MLE					
	$\widetilde{\theta}$	(95% CI)	$\sqrt{n}\,\mathrm{SE}(\widetilde{\theta})$	$\widetilde{\mathrm{Pr}}(q_{0.8} < Y)$	(95% CI)	$\sqrt{n}\,\mathrm{SE}(\widetilde{\mathrm{Pr}}(q_{0.8} < Y))$
No	2.196	(2.154, 2.238)	1.578	0.101	(0.096, 0.107)	0.207
Yes		(2.151, 2.241)	1.696		(0.095, 0.107)	0.225
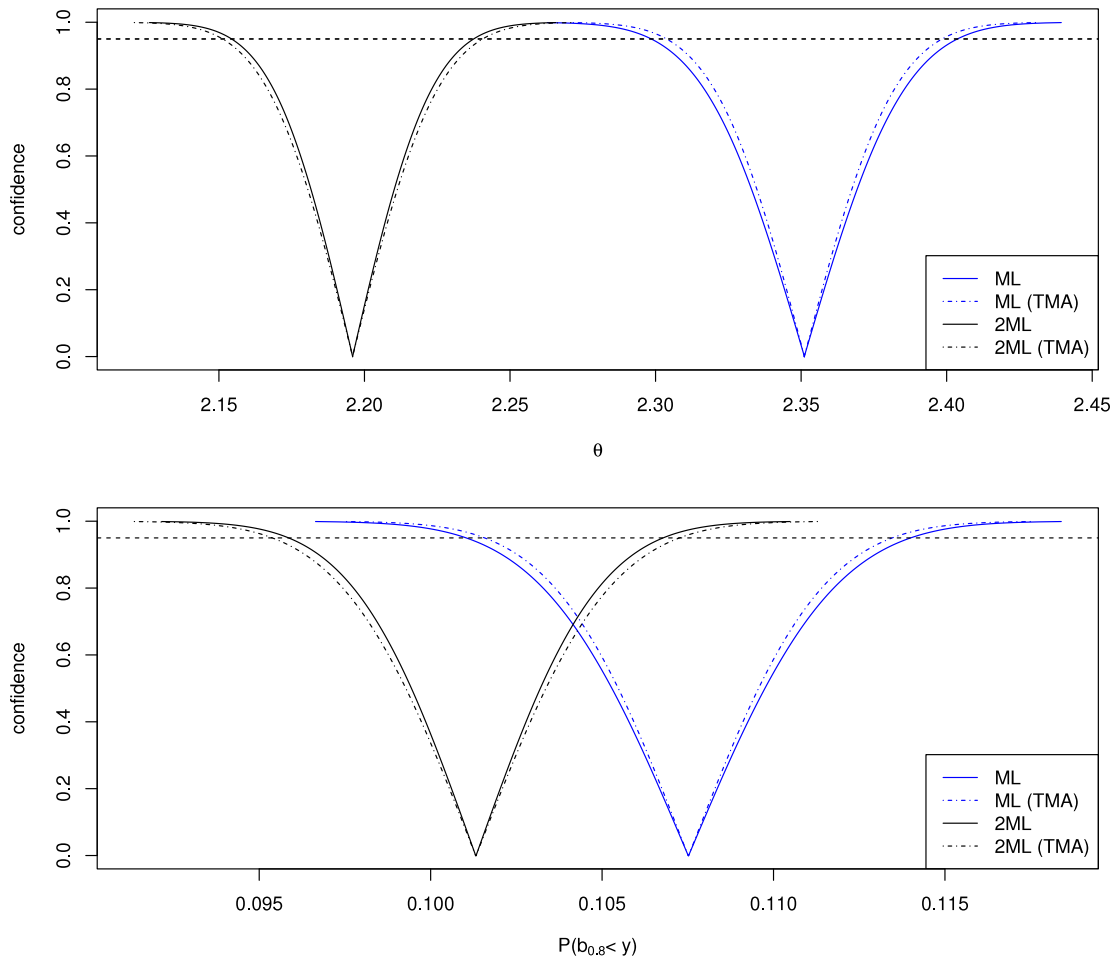
Fig. 4. Confidence curves for θ and $\Pr(q_{0.8} < Y)$. Blue curves are from the model fitted with ML and black curves are from the model fitted with two-stage ML. Dashed curves correspond to the true model assumption. The dashed horizontal lines indicate point-wise 95% confidence . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimator, outside model conditions. The extent to which this is an advantage or not then depends on the implicit biases, when used to estimate well-defined population quantities.

Consider any parameter of interest, say $\mu = \mu(G)$, expressed in terms of the underlying data generating mechanism G, the joint cumulative distribution function for $y = (y_1, \ldots, y_d)^\top$. When G is modeled as in (1)–(2), the parameter of interest may be expressed in terms of the parameters of that model, i.e., $\mu = \mu(\eta) = \mu(\alpha, \theta)$. Under mild regularity conditions, also regarding the smoothness of μ, methods and results obtained in Section 4 lead to

$$\sqrt{n}\left(\widehat{\mu}_{\text{ML}} - \mu_{0,\text{ML}}\right) \rightsquigarrow \mathcal{N}\left(0, \tau_{\text{ML}}\right), \quad \sqrt{n}\left(\widetilde{\mu}_{\text{2ML}} - \mu_{0,\text{2ML}}\right) \rightsquigarrow \mathcal{N}\left(0, \tau_{\text{2ML}}\right),$$

for the ML and the two-stage ML method, respectively. Here $\mu_{0,\text{ML}} = \mu(\eta_{0,\text{ML}})$ and $\mu_{0,\text{2ML}} = \mu(\eta_{0,\text{2ML}})$ are the appropriate least false parameter values for $\mu = \mu(\eta)$, involving the least false parameter values that the ML and the two-stage ML method are aiming for, as in Section 4. Also, $\tau_{\text{ML}} = c_{\text{ML}}^\top V_{\eta,\text{ML}} c_{\text{ML}}$ and $\tau_{\text{2ML}} = c_{\text{2ML}}^\top V_{\eta,\text{2ML}} c_{\text{2ML}}$, with gradient vectors $c = \partial\mu(\eta)/\partial\eta$ evaluated at respectively the $\eta_{0,\text{ML}}$ and $\eta_{0,\text{2ML}}$ positions in the $\eta = (\alpha^\top, \theta^\top)^\top$ parameter space. There is also a third natural estimator to be reckoned with here, namely the nonparametric $\widehat{\mu}_{\text{np}} = \mu(G_n)$, exploiting the direct empirical distribution of the observations, and for which a similar result $\sqrt{n}\left(\widehat{\mu}_{\text{np}} - \mu\right) \rightsquigarrow \mathcal{N}(0, \tau_{\text{np}})$ holds. This essentially requires the existence of an influence function for $\mu(G)$ with a finite variance.

Importantly, these results give rise to mean squared error (MSE) approximations, for the three natural estimators of the focus parameter. We may write

$$\widehat{\mu}_{\text{np}} = \mu + n^{-1/2}\tau_{\text{np}}^{1/2} Z_{n,\text{np}}, \quad \widehat{\mu}_{\text{ML}} = \mu_{0,\text{ML}} + n^{-1/2}\tau_{\text{ML}}^{1/2} Z_{n,\text{ML}}, \quad \widetilde{\mu}_{\text{2ML}} = \mu_{0,\text{2ML}} + n^{-1/2}\tau_{\text{2ML}}^{1/2} Z_{n,\text{2ML}},$$

with the three Z_n variables tending in distribution to the standard normal $\mathcal{N}(0, 1)$. These results invite the first-order large-sample approximations

$$\text{MSE}_{\text{np}} = 0^2 + \tau_{\text{np}}/n, \quad \text{MSE}_{\text{ML}} = b_{\text{ML}}^2 + \tau_{\text{ML}}/n, \quad \text{MSE}_{\text{2ML}} = b_{\text{2ML}}^2 + \tau_{\text{2ML}}/n,$$

involving in addition to the approximate variances also the biases $b_{\mathrm{ML}} = \mu_{0,\mathrm{ML}} - \mu$ and $b_{2\mathrm{ML}} = \mu_{0,2\mathrm{ML}} - \mu$. The nonparametric estimator has zero asymptotic bias.

Depending on the specifics of the models studied, and the focus parameter examined, each of the three estimators might win, in the sense of having the smallest MSE. Particularly with a complicated μ, the parametric variances might be significantly smaller than the nonparametric ones, so either the ML or the two-stage method might win, for small or moderate sample sizes, if the biases are not large. It is also clear that if the parametric model is not accurate, leading to non-zero biases b_{ML} and $b_{2\mathrm{ML}}$, the nonparametric estimation method will win for sufficiently large n. What is best in practice, for different focus parameters, when working with a given dataset, becomes a matter of balancing bias with variance.

In work flowing from the present paper, the authors have pursued these ideas, by constructing asymptotically unbiased estimators of the quantities involved in the MSE expressions. Incidentally, this requires the full simultaneous limit distribution of the Z_n variables above, in that limit distributions of $\widehat{b}_{\mathrm{ML}} = \widehat{\mu}_{\mathrm{ML}} - \widehat{\mu}_{\mathrm{np}}$ and $\widetilde{b}_{2\mathrm{ML}} = \widetilde{\mu}_{2\mathrm{ML}} - \widehat{\mu}_{\mathrm{np}}$ also play a role in the construction of estimators of the squared biases involved. The end result is a focused information criterion (FIC) for copula, in the tradition of the earlier developed focused information criterion (see Claeskens and Hjort [7,8] and Jullum and Hjort [22,23]), but now involving at least three estimators for each focus parameter. If the statistician examines say k different parametric copula models, each of type (1)–(2), then $2k+1$ different estimators are evaluated, assessed, and ranked, via the FIC for copula. See Ko et al. [27] for details, discussion, and applications.

8. Conclusions and further research

In our paper we have reached precise results for the behavior of two-stage ML estimators for parametric copula models, both under and outside model conditions. This has then led to a full and flexible machinery for inference, including confidence intervals and curves for focus parameters, testing hypotheses, etc. Here we offer some concluding remarks, some pointing to further research.

The assumption that the chosen parametric model captures the true model that generated data leads to a set of simplifications in our apparatus. For instance, the model robust asymptotic variance (3) simplifies to the model non-robust asymptotic variance formula (4), which is essentially the same as the asymptotic variance formula from Joe [20]. The difference is that we choose to use the inverse of the Fisher information matrix instead of the covariance of score vectors. We believe that this is more concordant with the current practice of ML theory. Although these two definitions of asymptotic variances are theoretically the same under the true model assumption, the practical difference between them can be non-negligible when a real-life dataset is used. A further study on the impact of making one vs. the other choice for estimating the population Fisher information matrix, either in the usual fashion based on the Hessian matrix at the ML position, or using the variance matrix for the score vectors, would be useful not only for the copula community, but also more generally for statisticians applying ML theory in other contexts.

The methods in this paper require that the first and second derivatives of the density function with respect to the model parameters can be obtained. When it is not feasible to obtain the derivatives, one can consider to use the resampling methods mentioned in Joe [20,21] as alternatives.

Our simulation study shows that assuming that the model is true (i.e., using the model non-robust variance formula) can both decrease and increase the asymptotic variance of the copula parameter estimate, depending on the parametric form of the model. The degree of decrease or increase gets larger as the degree of misspecification increases.

Since the asymptotic standard deviation is divided by $\sqrt{n}$ when computing a confidence interval, the overconfidence or underconfidence coming from the true model assumption gets larger as sample size decreases. When carrying out hypothesis testing for the copula parameter, or on a function that takes the copula parameter as an input, the choice of model-robust variance formula or not can therefore lead to different decisions.

Joe [20] compares asymptotic relative efficiency (ARE) of the two-stage ML estimator with that of the full ML and concludes that two-stage ML method typically has good ARE. We could observe that the two-stage ML estimator is still highly efficient when the true model assumption is dropped and the model robust asymptotic variance formulas are used.

When models are highly misspecified, however, we see that the asymptotic variance of two-stage ML estimators can be smaller than that of the ML estimator. This happens both when the model-robust and model non-robust variance formulas are used. This effect increases as the degree of model misspecification and dimension increases. This relates to the fact that the Hájek–Le Cam convolution theorem, along with theorems of a similar nature for the optimality of ML estimation for parametric models, does not apply when the models are misspecified.

When the true model assumption decreases the variance of the estimator of the copula parameter, we sometimes observe that the asymptotic variance of the joint upper tail probability increases. A possible reason is that this asymptotic variance is computed with the delta method by using the whole variance matrix of $\widetilde{\eta}$ (or $\widehat{\eta}$), including the covariances between copula and margin parameters. A further theoretical and numerical study on the property of variance transformation through the delta method would be fruitful, especially considering the fact that this also happens in various other types of models.

Our general results concerning misspecified models involve both the notion of least false parameters (those aimed at by respectively the ML and the two-stage ML methods) and ensuing limit variance matrices. It appears possible to study how much these vary from their true counterparts, under model conditions, in terms of functional derivatives, in certain types of model neighborhoods. The $O(1/\sqrt{n})$ model misspecification from Chapters 6 and 7 of Claeskens and Hjort [8] can be used for such purposes. Such considerations could then be used to pinpoint which parameter estimands $\mu = \mu(\eta)$ are more robust

than others, with respect to moderate misspecification. It is outside the scope of the present paper to pursue this issue, however. We also consider the development of Section 7, with concrete analysis for each focus parameter $\mu = \mu(\eta)$, to be more directly informative in this regard.

The authors of this study are currently developing model selection criteria for two-stage ML estimators that utilize the model robust large-sample distribution from Section 3 of this article [25]. These model selection criteria will complement earlier model selection methodology efforts, such as the Copula Information Criterion (CIC) developed by Grønneberg and Hjort [14]. One particular goal of these extended efforts will be to aid model building and selection for pair-copula constructions [1,2].

The methodology for two-stage ML estimation for copula models, inside and outside model conditions, has been developed in this paper, primarily aiming for the case of iid sequences of vector observations. Importantly, in the presence of relevant covariate information, the large-sample results and ensuing inference methods can be extended to various classes of conditional copula regression models, with the required extra efforts. This will then lead to further estimation and inference tools for models worked in, e.g., [3,37]. The covariates may influence the margins, the copula mechanism, or both. As one such example, we might fit the Norwegian precipitation data of Section 6 using the Gumbel copula, but now with

$$\theta_i = \theta \exp\{\gamma_1(x_i - 1990) + \gamma_2(x_i - 1990)^2\},$$

where x_i is the calendar year for observation $i \in \{1, \ldots, n\}$. This leads to estimates and confidence curves for γ_1, γ_2 and related parameters, via the two-stage construction

$$\ell_c(\widetilde{\alpha}, \theta, \gamma) = \sum_{i=1}^{n} \ln c\{F_1(y_{i,1}, \widetilde{\alpha}_1), \ldots, F_d(y_{i,d}, \widetilde{\alpha}_d), \theta_i\},$$

and allows one the opportunity to assess any changes of the copula mechanism over time.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2019.01.004.

References

[1] K. Aas, D. Berg, Models for construction of multivariate dependence: A comparison study, Eur. J. Finance 15 (2009) 639–659.
[2] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, Insurance Math. Econom. 44 (2009) 182–195.
[3] E.F. Acar, R.V. Craiu, F. Yao, Statistical testing of covariate effects in conditional copula models, Electron. J. Statist. 7 (2013) 2822–2850.
[4] E.W. Andersen, Composite likelihood and two-stage estimation in family studies, Biostatistics 5 (2004) 15–30.
[5] E.W. Andersen, Two-stage estimation in copula models used in family studies, Lifetime Data Anal. 11 (2005) 333–350.
[6] G. Casella, R.L. Berger, Statistical Inference, Duxbury Pacific Grove, CA, 2002.
[7] G. Claeskens, N.L. Hjort, The focused information criterion [with discussion contributions and a rejoinder], J. Amer. Statist. Assoc. 98 (2003) 900–916.
[8] G. Claeskens, N.L. Hjort, Model Selection and Model Averaging, Cambridge University Press, Cambridge, 2008.
[9] S.G. Coles, J. Heffernan, J.A. Tawn, Dependence measures for extreme value analyses, Extremes 2 (1999) 339–365.
[10] P. Embrechts, Copulas: A personal view, J. Risk Insur. 76 (2009) 639–650.
[11] C. Genest, A.-C. Favre, Everything you always wanted to know about copula modeling but were afraid to ask, J. Hydrol. Eng. 12 (2007) 347–368.
[12] C. Genest, J.G. Nešlehová, L.-P. Rivest, The class of multivariate max-id copulas with L_1-norm symmetric exponent measure, Bernoulli J. 24 (2018) 3751–3790.
[13] C. Genest, L.-P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, J. Amer. Statist. Assoc. 88 (1993) 1034–1043.
[14] S. Grønneberg, N.L. Hjort, The copula information criteria, Scand. J. Stat. 41 (2014) 436–459.
[15] J. Hájek, A characterization of limiting distributions of regular estimates, Probab. Theory Related Fields 14 (1970) 323–330.
[16] A.R. Hall, Generalized Method of Moments, Oxford University Press, Oxford, 2005.
[17] J.W. Hardin, The sandwich estimate of variance, in: T.B. Fomby, R.C. Hill (Eds.), Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later, Emerald Group Publishing Limited, 2003, pp. 45–73.
[18] I. Hobæk Haff, Parameter estimation for pair-copula constructions, Bernoulli 19 (2013) 462–491.
[19] H. Joe, Multivariate Models and Multivariate Dependence Concepts, Chapman & Hall, London, 1997.
[20] H. Joe, Asymptotic efficiency of the two-stage estimation method for copula-based models, J. Multivariate Anal. 94 (2005) 401–419.
[21] H. Joe, Dependence Modeling with Copulas, CRC Press, Boca Raton, FL, 2014.
[22] M. Jullum, N.L. Hjort, Parametric or nonparametric: The FIC approach, Statist. Sinica 27 (2017) 951–981.
[23] M. Jullum, N.L. Hjort, What price semiparametric Cox regression? Lifetime Data Anal. (2018) http://dx.doi.org/10.1007/s10985-018-9450-7, (in press).
[24] G. Kim, M.J. Silvapulle, P. Silvapulle, Comparison of semiparametric and parametric methods for estimating copulas, Comput. Statist. Data Anal. 51 (2007) 2836–2850.
[25] V. Ko, N.L. Hjort, Copula information criterion for model selection with two-stage maximum likelihood estimation, Econometr. Statist. (2019) (in press).

[26] V. Ko, N.L. Hjort, Supplement to "Model robust inference with two-stage maximum likelihood estimation for copulas", J. Multivariate Anal. (2019).
[27] V. Ko, N.L. Hjort, I. Hobæk Haff, Focused information criterion for copulas, Scand. J. Stat. (2019) in press.
[28] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[29] L. Le Cam, Maximum likelihood: An introduction, Internat. Statist. Rev. 58 (1990) 153–171.
[30] E.L. Lehmann, Elements of Large-Sample Theory, Springer Science & Business Media, 2004.
[31] A.J. McNeil, J. Nešlehová, Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions, Ann. Statist. 37 (2009) 3059–3097.
[32] R.B. Nelsen, An Introduction to Copulas, second ed., Springer, New York, 2006.
[33] A.K. Nikoloulopoulos, H. Joe, H. Li, Vine copulas with asymmetric tail dependence and applications to financial return data, Comput. Statist. Data Anal. 56 (2012) 3659–3673.
[34] T. Schweder, N.L. Hjort, Confidence, Likelihood, Probability: Inference With Confidence Distributions, Cambridge University Press, Cambridge, 2016.
[35] J.H. Shih, T.A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[36] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.
[37] N. Veraverbeke, M. Omelka, I. Gijbels, Estimation of a conditional copula and association measures, Scand. J. Stat. 38 (2011) 766–780.
[38] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1982) 1–25.
[39] J.J. Xu, Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data (Ph.D. thesis), University of British Columbia, Vancouver, BC, Canada, 1996.