Estimation in copula models with two-piece skewed margins using the inference for margins method ⋆

Jonas Baillien[a], Irène Gijbels[a], Anneleen Verhasselt[b,*]

[a]*Department of Mathematics and Leuven statistics Research Center (LStat) , Celestijnenlaan 200 B, Heverlee, 3001, Belgium*
[b]*Center for Statistics, Data Science Institute, Agoralaan Building D, Diepenbeek, 3590, Belgium*

Abstract

Copulas provide a versatile tool in the modelling of multivariate distributions. With an increased awareness for possible asymmetry in data, skewed copulas in combination with classical margins have been employed to appropriately model these data. The reverse, skewed margins with a (classical) copula has also been considered, but mainly with classical skew-symmetrical margins. An alternative approacch is to rely on a large family of asymmetric two-piece distributions for the univariate marginal distributions. Together with any copula this family of asymmetric univariate distributions provides a powerful tool for skewed multivariate distributions. Maximum likelihood estimation of all parameters involved is discussed. A key step in achieving statistical inference results is an extension of the theory available for generalized method of moments, under non-standard conditions. This together with the inference results for the family of univariate distributions, allows to establish consistency and asymptotic normality of the estimators obtained through the method of 'inference functions for margins'. The theoretical results are complemented by a simulation study and the practical use of the method is demonstrated on real data examples.

Keywords: asymptotic normality, consistency, Fisher information, generalized method of moments, skewed distributions.

1. Introduction

Copulas are an interesting tool for modelling multidimensional data. With an ever increasing popularity by practitioners in fields such as finance, insurance, biology, . . . , they have become a cornerstone in the field of multivariate modelling. Classical examples include the family of Archimedean copulas, which are favored because of their relative simplicity,

⋆This paper comes with Supplementary Material.

*Corresponding author. E-mail address: anneleen.verhasselt@uhasselt.be (A. Verhasselt).

and the elliptical copulas. Of the latter class, the most well known examples are the Gaussian and Student's t-copulas. As for the univariate margins, these are usually chosen based on the application at hand, and one often falls back on well known distributions such as the normal, Weibull, etc., or in many instances the empirical margins as this reduces the risk of misspecification. Skewness can be introduced in a multivariate copula-based distribution in three ways: by using a skewed copula (see [43] and [44], among others), by using skewed margins (see [20], [39], among others), or as a consequence, by combining both. Skewed copulas are mostly based on some multivariate skewed distribution, much like elliptical copulas are constructed. The main advantage in this is that asymmetric dependencies can be incorporated in the distribution ([38], [42]). This can be combined with skewed margins to provide the most flexible approach, but can also be considered overkill because it drastically increases the complexity of the model.

More precisely, consider a d-dimensional random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^\top \in \mathbb{R}^d$ with joint and marginal cumulative distribution functions respectively $F_{\mathbf{Z}}$ and F_{Z_j} (for $j = 1, \ldots, d$). By Sklar's theorem (see [36]), there exists a d-dimensional function $C : [0,1]^d \mapsto [0,1]$, called a copula, such that

$$F_{\mathbf{Z}}(\mathbf{z}) = C(F_{Z_1}(z_1), \ldots, F_{Z_d}(z_d)) \qquad \forall\, \boldsymbol{z} = (z_1, \ldots, z_d)^\top \in \mathbb{R}^d.$$

The copula function C 'couples' the univariate marginal cumulative distribution functions $F_{Z_1}, \ldots, F_{Z_d}$ to the joint cumulative distribution function $F_{\mathbf{Z}}$. The function C is unique in case all F_{Z_j} are continuous, and then fully characterizes the dependence structure within $(Z_1, \ldots, Z_d)^\top$. A copula is a d-variate distribution function with uniform margins. For extensive introductions to copulas and their properties, the reader is referred to [29] or [18].

Suppose now that, for $j = 1, \ldots, d$, the marginal distribution F_{Z_j} (short the margin) depends on a parameter $\boldsymbol{\eta}_j$, with $\boldsymbol{\eta}_j \in \mathcal{H}_j \subset \mathbb{R}^{q_j}$, and that the copula function C depends on a parameter $\boldsymbol{\theta}_C \in \boldsymbol{\Theta}_C \subset \mathbb{R}^p$. We thus write the copula function as $C(\cdot; \boldsymbol{\theta}_C)$. Denoting $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^\top, \ldots, \boldsymbol{\eta}_d^\top)^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\eta}^\top, \boldsymbol{\theta}_C^\top)^\top$, we then have

$$F_{\mathbf{Z}}(\mathbf{z}; \boldsymbol{\theta}) = C(F_{Z_1}(z_1; \boldsymbol{\eta}_1), \ldots, F_{Z_d}(z_d; \boldsymbol{\eta}_d); \boldsymbol{\theta}_C). \tag{1}$$

Throughout the paper the copula function C is assumed to be absolutely continuous, i.e. the copula density

$$c(u_1,\dots,u_d;\boldsymbol{\theta}_C)=\frac{\partial^d C(u_1,\dots,u_d;\boldsymbol{\theta}_C)}{\partial u_1\dots\partial u_d},$$

with $u_1,\dots,u_d\in[0,1]$, exists. By differentiating (1) with respect to all z_j, we obtain the expression of the multivariate density of $\boldsymbol{Z}$ in terms of the copula density and the marginal densities $f_{Z_j}(\cdot;\boldsymbol{\eta}_j)$ (where $f_{Z_j}=F'_{Z_j}$):

$$f_{\mathbf{Z}}(\mathbf{z};\boldsymbol{\theta})=c\big(F_{Z_1}(z_1;\boldsymbol{\eta}_1),\dots,F_{Z_d}(z_d;\boldsymbol{\eta}_d);\boldsymbol{\theta}_C\big)\prod_{j=1}^{d}f_{Z_j}(z_j;\boldsymbol{\eta}_j). \tag{2}$$

Parameter estimation for copulas exists, like for most distributions, in many forms. In a fully parametric setting, one can distinguish two major approaches. See [24] and [5] for some discussions. A first parametric estimation approach is classical maximum likelihood estimation (MLE) over all parameters (the parameters of the margins and these of the copula), simultaneously. This is convenient when multiple margins share common parameters and provides the most efficient estimator in case the margins and copula are specified correctly. However, this yields a potentially large optimization problem, which can be computationally impractical and time consuming. The second parametric estimation approach is a two-step MLE also known as Inference Functions for Margins (IFM, see [22] and [21]). In the first step only the marginal parameters $\boldsymbol{\eta}_j$ (for $j=1,\dots,d$), are estimated. In the second step, the copula parameters are estimated using the estimated marginal parameters $\boldsymbol{\eta}_j$ obtained in the first step. The advantage over the first approach, the full MLE, is that one does not have a single large problem, but multiple smaller optimization problems which are easier and faster to solve. It has been shown in [21] that full MLE is in theory the most efficient procedure, but IFM mostly performs very similar except in some exceptional situations. In [22] asymptotic normality of both the MLE and the IFM estimator is proven, under certain mild regularity conditions. Due to a similar performance but a more tractable and easier formulation, IFM is widely preferred over full MLE. Often also, the empirical distribution of the margins is used instead of parametric margins, this results in the Maximum Pseudo Likelihood. We will however consider parametric marginal distributions throughout the paper. For additional information on parametric and nonparametric margins, the reader is referred to, among others, [18].

Most results on parametric statistical inference for copulas are derived under certain smoothness assumptions. These assumptions surely do not hold for all distributions, especially for certain classes of margins. A broad class of interesting skew univariate distributions are the two-piece distributions, but these are not always, or everywhere, differentiable. Dating back as early as [9], the class of two-piece distributions contains well known examples such as the split-normal distribution (see [11]) or the Fernandez-Steel distribution (see [10]). An extensive review on two-piece distributions is given in [40]. We focus on a more recent family of two-piece distributions, namely the quantile-based asymmetric (QBA) family of distributions studied in [12]. A univariate skew density f_{Z_j} within this family is defined as follows. Consider f_j a unimodal symmetric around 0 density (called (symmetric) reference density hereafter), and define the density function of the j-th margin of $F_{\mathbf{Z}}$ as

$$f_{Z_j}(z_j;\boldsymbol{\eta}_j) = \frac{2\alpha_j(1-\alpha_j)}{\phi_j}\begin{cases} f_j\left((1-\alpha_j)\frac{\mu_j - z_j}{\phi_j};\boldsymbol{\kappa}_j\right) & \text{if } z_j \leq \mu_j \\ f_j\left(\alpha_j\frac{z_j-\mu_j}{\phi_j};\boldsymbol{\kappa}_j\right) & \text{if } z_j > \mu_j. \end{cases} \tag{3}$$

Herein $\mu_j \in \mathbb{R}$ and $\phi_j \in (0,\infty)$ denote respectively a location and scale parameter. The parameter $\alpha_j \in (0,1)$ denotes the skewness parameter: for $\alpha = 0.5$ the density f_{Z_j} is symmetric around the mode μ_j; whereas for $\alpha_j > 0.5$ (respectively $\alpha_j < 0.5$) the density is left-skewed (respectively right-skewed). In expression (3), $\boldsymbol{\eta}_j = (\alpha_j, \mu_j, \phi_j, \boldsymbol{\kappa}_j^\top)^\top$, where $\boldsymbol{\kappa}_j$ contains all other parameters from the unimodal symmetric reference density f_j. An example where an extra parameter is present is when f_j is a Student's t-distribution with degrees of freedom ν_j. [12] established an asymptotic normality result for parameter estimators of f_{Z_j}, under certain conditions on the reference density f_j.

Combining a given copula with marginal distributions that are member of a very broad family of skew distributions, allows for a broad class of skew multivariate distributions. The family of distributions in (3) is very versatile, englobing for each given reference density f_j, symmetric and skewed univariate distributions (depending on whether α_j equals 0.5 or not). The advantage of working with the specific two-pieces distributions in (3) is twofold: (i) in the various dimensions, different behaviours in skewness can be considered by allowing for different densities f_j, $j = 1, \ldots, d$; and (ii) statistical inference for each of the parameters in the margins is feasible. To validate this advantage however one needs to find a way to combine techniques for estimating parameters in the copula functions, with techniques for parameter estimation of the margins. A crucial point is that two-pieces distributions can violate classical regularity conditions. Hence, we prefer IFM over MLE as for these non-regular margins, derivative free optimization is required, which comes at an increased computational cost. Under this setting, consistency of the IFM estimator as well as asymptotic normality are shown to hold. Using marginal distributions that might violate classical regularity conditions (because the f_j in (3) violates them) poses no real problems from a practical point of view. However showing the necessary theoretical inference results demands a careful study and new techniques. It is precisely this gap in inference results that we bridge in this paper. To achieve such results in the overarching setting of copulas and margins, the innovation is to use a generalized method of moments methodology under non-standard regularity conditions. This provides the necessary framework for a broader class of models where we will focus specifically on IFM.

The outline of the paper is as follows. In Section 2 we discuss the statistical inference methodology for both the margins and the copula parameters obtained through IFM. Section 3 contains the main theoretical results: consistency and asymptotic normality of the IFM estimators. In Section 4 two simulation studies are conducted. Section 5 shows practical applicability on two data examples and Section 6 finishes with a short discussion.

2. Statistical inference for IFM

Papers dealing with statistical inference for IFM include [35], [22], [28] and [26]. The latter paper shows that asymptotic normality results hold even when the model (both margins and copula) is misspecified. In [26] it is also mentioned that IFM can be seen as a special case of a two-step maximum likelihood estimator. Two-step maximum likelihood estimators are in turn a special case of a generalized method of moments (GMM) estimator where the moments are the first order conditions of the likelihood. In this light, results obtained for GMM can also be applied to IFM if the necessary conditions are satisfied. As mentioned in [15], the most important part of two-step estimation is that uncertainty on parameter estimates introduced in the first step should be taken into account in the uncertainty of parameter estimates in the second step. So just viewing the first step parameter estimates as being fixed might lead to underestimated uncertainty. Many results have been published on GMM estimators, yet most of these only consider the case where standard regularity conditions are satisfied. When this is not the case, the main result to fall back on is [30] and therein mentioned related works.

The intuition for statistical inference results to hold in the global picture of both steps, as also mentioned in [15], is that the results should hold in both steps separately. Asymptotically, first step inference results can then be incorporated in the second step. So as a starting point, regularity conditions should hold for both steps, in addition to some further conditions depending on the exact result one is after. However, in our context the classical regularity conditions are violated in the first step. Yet, even then, albeit under a slightly different set of conditions, consistency and asymptotic normality results equivalent to the classical results can be obtained for the margins. Intuitively, it is thus expected that under these surrogate conditions and the classical regularity conditions for the second step, global statistical inference results equivalent to those under standard regularity conditions can be obtained.

In the sequel $c\big(F_{Z_1}(Z_1;\boldsymbol{\eta}_1),\ldots,F_{Z_d}(Z_d;\boldsymbol{\eta}_d);\boldsymbol{\theta}_C\big)$ is sometimes denoted shortly as $c\big(\mathbf{U};\boldsymbol{\theta}_C\big)$ or as $c\big(\boldsymbol{F_Z}(\boldsymbol{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C\big)$, where $\mathbf{U}=(U_1,\ldots,U_d)=\big(F_{Z_1}(Z_1;\boldsymbol{\eta}_1),\ldots,F_{Z_d}(Z_d;\boldsymbol{\eta}_d)\big)=\boldsymbol{F_Z}(\boldsymbol{Z};\boldsymbol{\eta})$. Further, we use the Euclidean norm, $\|\boldsymbol{z}\|=\sqrt{\sum_{j=1}^d z_j^2}$, unless specified differently.

It is implicitly assumed that the model is correctly specified (both the margins and the copula) and that standard regularity conditions (Assumptions (R1)$-$(R3) listed below) hold for $c(\mathbf{U};\boldsymbol{\theta_C})$. Assumption (R4) needs to be added here. We also require that the likelihood with respect to the copula parameters under given margins is uniquely maximized.

(R1) $\boldsymbol{\theta}_C^0\in\mathcal{N}(\boldsymbol{\theta}_C)\subset\boldsymbol{\Theta}_C$, with $\mathcal{N}(\boldsymbol{\theta}_C)$ an open set, $\boldsymbol{\Theta}_C$ compact and $\boldsymbol{\theta}_C^0$ is the true copula parameter.

(R2) $\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln \big(c\big(\mathbf{u}; \boldsymbol{\theta}_C\big)\big)$, $\frac{\partial^2}{\partial \boldsymbol{\theta}_C^2} \ln \big(c\big(\mathbf{u}; \boldsymbol{\theta}_C\big)\big)$ exist for almost all $\mathbf{u} \in [0,1]^d$ and for each $\boldsymbol{\theta}_C$ in a neighborhood of $\boldsymbol{\theta}_C^0$.

(R3) $0 < E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln \big(c\big(\mathbf{U}; \boldsymbol{\theta}_C\big)\big)\right)^2\right] < \infty$, $\forall \boldsymbol{\theta}_C \in \boldsymbol{\Theta}_C$, i.e. the Fisher information matrix is positive definite and bounded.

(R4) There exists a function $H(\mathbf{u})$ such that $\left\|\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln \big(c\big(\mathbf{u}; \boldsymbol{\theta}_C\big)\big)\right\| \leq H(\mathbf{u})$ for almost all $\boldsymbol{\theta}_C \in \boldsymbol{\Theta}_C$ and for all $\mathbf{u} \in [0,1]^d$ and $E[H(\mathbf{U})] < \infty$.

These are mild assumptions. Assumption (R1) is a standard assumption. Note that Assumption (R2) should hold for almost all $\mathbf{u} \in [0,1]$, not for all $\mathbf{u}$ which is important as many copulas derivatives do not exist on the boundaries. Assumption (R3) is needed to assure existence and finiteness of the variance of the estimator. Finally, (R4) is, due to the requirement of the expectation being finite, a rather weak condition necessary for uniform convergence. In Section 3.3 we give some illustrative examples, to show these assumptions are reasonable.

Statistical inference relies on (two-step) maximum likelihood estimation. By (2), a contribution, for an observation $\mathbf{Z}$, in the log-likelihood of a copula $C(\mathbf{u}; \boldsymbol{\theta}_C)$, $\mathbf{u} \in [0,1]^d$ with marginal cumulative distribution functions F_{Z_j}, $j = 1, \ldots, d$ is

$$\ell(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d, \boldsymbol{\theta}_C; \mathbf{Z}) = \ln \big(c\big(F_{Z_1}(Z_1; \boldsymbol{\eta}_1), \ldots, F_{Z_d}(Z_d; \boldsymbol{\eta}_d); \boldsymbol{\theta}_C\big)\big) + \sum_{j=1}^{d} \ln \big(f_{Z_j}(Z_j; \boldsymbol{\eta}_j)\big) .$$

2.1. Statistical inference for the margins

We assume that the margins, originating from the QBA-distributions, share no common parameters. This reduces the problem of fitting the d-dimensional problem to d univariate problems. For statistical inference, the results obtained in [12] can be used. For these results to hold, the following assumptions need to be met for each margin density function $f_{Z_j}(z_j; \boldsymbol{\eta}_j)$.

(M1) Let $\mathcal{H}_{R,j}$ be a compact subset of $\mathcal{H}_j$ and assume that the true parameter $\boldsymbol{\eta}_j^0 \in \overset{\circ}{\mathcal{H}}_{R,j}$ (with $\overset{\circ}{\mathcal{H}}_{R,j}$ the interior of $\mathcal{H}_{R,j}$).

(M2) $\int_0^\infty |\ln(f_j(s))| f_j(s) ds < \infty$, where $f_j(s)$ is the underlying reference density.

(M3) $\gamma_{j,r} = \int_0^\infty s^{r-1} \frac{\left(f_j'(s)\right)^2}{f_j(s)} ds < \infty, \qquad \text{for } r = 1, 2, 3.$

(M4) $\lim_{s \to \infty} s f_j(s) = 0$ or $\int_0^\infty s f_j'(s) ds = -\frac{1}{2}$.

In case other parameters are present (such as degrees of freedom in case of an asymmetric Student's t component Z_j), two other assumptions are also required.

(M5) The MLE $\widehat{\boldsymbol{\eta}}_{j,n}^{\text{MLE}}$ is a weakly consistent estimator of $\boldsymbol{\eta}_j^0$.

(M6) The following holds

$$\begin{aligned}
& E\left[\frac{1}{2}\left(\frac{\partial^-}{\partial \boldsymbol{\eta}_j} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right) + \frac{\partial^+}{\partial \boldsymbol{\eta}_j} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right)\right)\right]_{\boldsymbol{\eta}_j=\boldsymbol{\eta}_j^0} \\
:= \; & E\left[\overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right)\right]_{\boldsymbol{\eta}_j=\boldsymbol{\eta}_j^0} \\
= \; & \mathbf{0}.
\end{aligned}$$

In this, $\frac{\partial^-}{\partial \boldsymbol{\eta}_j}$ and $\frac{\partial^+}{\partial \boldsymbol{\eta}_j}$ denote the left- respectively right-hand derivative and $\overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}}$ the average of both. This notation is necessary because f_{Z_j} is not necessarily differentiable with respect to μ_j. It then holds that (see [12])

- When (M1) and (M2) hold (or (M5)), the MLE $\widehat{\boldsymbol{\eta}}_{j,n}^{\text{MLE}}$ is (weakly) consistent, i.e. $\widehat{\boldsymbol{\eta}}_{j,n}^{\text{MLE}} \xrightarrow{P} \boldsymbol{\eta}_j^0$ for $n \to \infty$.

- When (M4) holds (or (M6))

$$E\left[\overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right)\right]_{\boldsymbol{\eta}_j=\boldsymbol{\eta}_j^0} = \mathbf{0}. \tag{4}$$

- When (M1)−(M4) hold (or (M1), (M5) and (M6))

$$\sqrt{n}(\widehat{\boldsymbol{\eta}}_{j,n}^{\text{MLE}} - \boldsymbol{\eta}_j^0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \mathcal{I}_j(\boldsymbol{\eta}_j^0)^{-1}\right), \qquad \text{for } n \to \infty, \tag{5}$$

$$\text{with} \qquad \mathcal{I}_j(\boldsymbol{\eta}_j^0) = \begin{bmatrix} \frac{2\left((\alpha_j^0)^3+(1-\alpha_j^0)^3\right)\gamma_{j,3}-(1-2\alpha_j^0)^2}{(\alpha_j^0)^2(1-\alpha_j^0)^2} & -\frac{2\gamma_{j,2}}{\phi_j^0} & -\frac{(1-2\alpha_j^0)(2\gamma_{j,3}-1)}{\alpha_j^0(1-\alpha_j^0)\phi_j^0} \\ -\frac{2\gamma_{j,2}}{\phi_j^0} & \frac{2\alpha_j^0(1-\alpha_j^0)\gamma_{j,1}}{(\phi_j^0)^2} & 0 \\ -\frac{(1-2\alpha_j^0)(2\gamma_{j,3}-1)}{\alpha_j^0(1-\alpha_j^0)\phi_j^0} & 0 & \frac{2\gamma_{j,3}-1}{(\phi_j^0)^2}, \end{bmatrix},$$

for the setting when there is no extra parameter vector $\boldsymbol{\kappa}_j$. For more general results see [12].

If the above assumptions hold for all margins, using the previously mentioned remark that no parameters are shared between margins, the full vector of marginal parameters $\boldsymbol{\eta}$ is consistent in the sense that $\widehat{\boldsymbol{\eta}}_n^{\mathrm{MLE}} \xrightarrow{P} \boldsymbol{\eta}^0$, as $n \to \infty$ and

$$\sqrt{n}(\widehat{\boldsymbol{\eta}}_n^{\mathrm{MLE}} - \boldsymbol{\eta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \mathcal{I}(\boldsymbol{\eta}^0)^{-1}\right), \qquad \text{as}\, n \to \infty,$$

where $\mathcal{I}(\boldsymbol{\eta}^0) = \operatorname{diag}\left(\mathcal{I}_j(\boldsymbol{\eta}_j^0)\right)$ is a block-diagonal matrix of dimension $q \times q$, with $q = \sum_{j=1}^{d} q_j$, with $\mathcal{I}_j(\boldsymbol{\eta}_j^0)$ as the j-th element $(j = 1, \ldots, d)$.

Remark 1. When the domain of one or more margins, say Z_j, is not $\mathbb{R}$, but say $\mathbb{M}_j$, a link function $h_j : \mathbb{M}_j \to \mathbb{R}$ can be used to transform the QBA-margin of a real-valued random variable Y_j to $Z_j = h_j^{-1}(Y_j)$ on the appropriate domain. Under the above assumptions together with the added assumption that the link function h_j is a differentiable function with $h_j'(.) > 0$, similar asymptotic results can be obtained. For a detailed exposition on this link function transformations issue, see [13].

2.2. Generalised method of moments

Treating IFM as a GMM estimator instead of an MLE has the key advantage that it provides a framework in which we can work without the cumbersome burden of already having estimated the margin parameter. Yet, it also has its downside as the conditions under which general results are valid, are stronger than those required for specific cases. Fortunately there are also different results, albeit few, which weaken the general requirements needed for statistical inference for GMM sufficiently such that the moment functions even need not be continuous. In particular we use results of [31] and [30] (Chapter 7) when establishing consistency and asymptotic normality of the IFM estimator in Section 3.

GMM is an example of a minimal distance estimator. Let $\underset{\sim}{\mathbf{Z}} = (\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(n)})$ be n i.i.d. realizations from the random variable $\mathbf{Z}$ and let $\mathbf{g}(\mathbf{Z}; \boldsymbol{\theta}) \in \mathbb{R}^m$ (with $m \geq 1$) represent the vector of (m) moment functions, consisting of the elements $\mathbf{g}_{\boldsymbol{\eta}_j}(Z_j; \boldsymbol{\eta}_j)$, $j = 1, \ldots, d$, followed by $\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\boldsymbol{\eta},\boldsymbol{\theta}_C)$, defined in respectively (9) and (10). Further recall we denote $\boldsymbol{\theta} = (\boldsymbol{\eta}_1^\top,\dots,\boldsymbol{\eta}_d^\top,\boldsymbol{\theta}_C^\top)^\top$. Suppose we want to find $\boldsymbol{\theta}^0$ for which $E[\mathbf{g}(\mathbf{Z};\boldsymbol{\theta}^0)] = \mathbf{0}$. GMM approaches this problem by replacing the expectation with the empirical mean. So by finding $\widehat{\boldsymbol{\theta}}_n$ for which

$$\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\widehat{\boldsymbol{\theta}}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{Z}^{(i)};\widehat{\boldsymbol{\theta}}_n) = \mathbf{0}, \tag{6}$$

or at least as close to zero as possible. This corresponds to minimizing the norm of $\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. The choice of norm is important, so the GMM estimator considers a set of norms

$$\left\|\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})\right\|_{\mathbf{W}}^2 = \widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})^\top\mathbf{W}\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}).$$

In this, the positive (semi-)definite weighting matrix $\mathbf{W}\in\mathbb{R}^{m\times m}$ depends on model parameters. In practice it is replaced with its empirical estimate $\widehat{\mathbf{W}}$. The choice of $\widehat{\mathbf{W}}$ is generally user dependent, but certain choices might be beneficial as is made clear below. The reason of the viability of GMM is that, by the law of large number ([7]), $\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) \xrightarrow{P} E[\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})]$ for $n\to\infty$, so the solution of (6) is expected to solve the original problem.

Using the above, in its general form (see [30]), a GMM estimator is the solution to the maximization problem

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \quad \widehat{\mathbf{Q}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}), \tag{7}$$

with

$$\widehat{\mathbf{Q}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) = -\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{Z}^{(i)};\boldsymbol{\theta})\right]^\top \widehat{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mathbf{Z}^{(i)};\boldsymbol{\theta})\right]. \tag{8}$$

Using the notation of (6), the GMM estimator (7) can be rewritten as the solution to the optimization problem

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left\{-\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})^\top\widehat{\mathbf{W}}\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})\right\}.$$

As a consequence of $\widehat{\mathbf{W}}$ being positive (semi-)definite, for n large enough, $\widehat{\mathbf{Q}}_n$ is bounded from above by zero. In essence a GMM estimator is nothing more than the solution to an optimization problem that provides us with the optimal set of parameters maximizing a given criterion function.

Other estimators such as least squares estimators or extremum estimators such as the MLE can also be expressed in the GMM framework. A MLE, for example, is defined as the parameter set which solves the score functions for zero. By taking $\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})$ in (8) to be the score functions, the MLE becomes the GMM estimator as it yields $\widehat{\mathbf{Q}}_n = 0$ hence providing a maximizer for (7).

Contrary to maximum likelihood estimators, very few results are available for IFM when classical regularity conditions are not met. This void is filled by the GMM estimator as there exist asymptotic results which hold under a less restrictive set of regularity conditions which are easier met by our non-regular (in the classical sense) IFM problem. The main advantage of treating IFM as a GMM estimator thus lies in the availability of asymptotic theory under less restrictive conditions.

We now introduce the GMM estimator specifically for a two-step MLE such as IFM. In the first step, the margins are estimated by MLE. As already explained in the previous paragraph, this can be solved by taking as moment functions the score functions. So for the first step, we have d sets of moment functions given by

$$\mathbf{g}_{\boldsymbol{\eta}_j}(Z_j;\boldsymbol{\eta}_j) = \overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right). \tag{9}$$

For the second step, $\boldsymbol{\eta}$ is replaced by its MLE $\widehat{\boldsymbol{\eta}}_n^{\text{MLE}}$ from the first step. So the only parameters left to estimate are the copula parameters $\boldsymbol{\theta}_C$. This is again done using maximum likelihood, but by using the scores of the log-copula density with respect to the copula parameters as moment functions. In doing so, it provides a second GMM estimator with

$$\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\widehat{\boldsymbol{\eta}}_n^{\text{MLE}},\boldsymbol{\theta}_C) = \frac{\partial}{\partial \boldsymbol{\theta}_C} \ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{Z};\widehat{\boldsymbol{\eta}}_n^{\text{MLE}});\boldsymbol{\theta}_C)\right). \tag{10}$$

The key property is that for two-step estimation, the different sets of moment functions can be stacked as explained in [30]. This bundles the two steps into a single optimization problem where the (combined) set of moment functions is given by

$$\mathbf{g}(\mathbf{Z};\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{g}_{\boldsymbol{\eta}_1}(Z_1;\boldsymbol{\eta}_1) \\ \vdots \\ \mathbf{g}_{\boldsymbol{\eta}_d}(Z_d;\boldsymbol{\eta}_d) \\ \mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\boldsymbol{\eta},\boldsymbol{\theta}_C) \end{bmatrix} = \begin{bmatrix} \overline{\frac{\partial}{\partial \boldsymbol{\eta}_1}} \ln\left(f_{Z_1}(Z_1;\boldsymbol{\eta}_1)\right) \\ \vdots \\ \overline{\frac{\partial}{\partial \boldsymbol{\eta}_d}} \ln\left(f_{Z_d}(Z_d;\boldsymbol{\eta}_d)\right) \\ \frac{\partial}{\partial \boldsymbol{\theta}_C} \ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C)\right) \end{bmatrix}, \tag{11}$$

of dimension $m := q + p$. Note that the MLE of $\boldsymbol{\eta}$ in $\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z}; \boldsymbol{\eta}, \boldsymbol{\theta}_C)$ was dropped in (11) as they are treated as constants in those moment functions.

2.3. General asymptotic normality result of GMM estimator

In showing asymptotic normality of the IFM estimator with QBA-distributed margins, the most important result is Theorem 7.2 of [30]. The theorem states that asymptotic normality of GMM estimators for which the classical regularity conditions are not satisfied can still be obtained under a milder set of conditions. In this, the moment functions (11) are used. The theorem goes as follows.

Theorem 1 ([30], Theorem 7.2 p2186)**.** *Suppose that*

(I) $\widehat{\mathbf{g}}_n\left(\underset{\sim}{\mathbf{Z}}; \widehat{\boldsymbol{\theta}}_n\right)^\top \widehat{\mathbf{W}}\widehat{\mathbf{g}}_n\left(\underset{\sim}{\mathbf{Z}}; \widehat{\boldsymbol{\theta}}_n\right) \leq \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta})^\top \widehat{\mathbf{W}}\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta}) + o_p(n^{-1})$,

(II) $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}^0$ *as* $n \to \infty$,

(III) $\widehat{\mathbf{W}} \xrightarrow{P} \mathbf{W}$ *as* $n \to \infty$ *with* $\mathbf{W}$ *positive (semi-)definite.*

Suppose further that

$$\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta}) \xrightarrow{P} E[g(\mathbf{Z}; \boldsymbol{\theta})] = \mathbf{g}_0(\boldsymbol{\theta}) \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{12}$$

such that (i) $\mathbf{g}_0(\boldsymbol{\theta}^0) = \mathbf{0}$,

(ii) $\mathbf{g}_0(\boldsymbol{\theta})$ *is differentiable in* $\boldsymbol{\theta}^0$ *with derivative* $\mathbf{G} \in \mathbb{R}^{m\times m}$ *such that* $\mathbf{G}^\top\mathbf{W}\mathbf{G}$ *is non-singular,*

(iii) $\boldsymbol{\theta}^0 \in \overset{\circ}{\Theta}$,

(iv) $\sqrt{n}\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$*, for* $n \to \infty$,

(v) for any $\delta_n > 0$ *such that* $\delta_n \to 0$ *for* $n \to \infty$ *it holds that*

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^0\|\leq\delta_n} \frac{\sqrt{n}\left\|\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta}) - \widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}}; \boldsymbol{\theta}^0) - \mathbf{g}_0(\boldsymbol{\theta})\right\|}{1 + \sqrt{n}\,\|\boldsymbol{\theta} - \boldsymbol{\theta}^0\|} \xrightarrow{P} 0, \qquad \textit{for} \quad n \to \infty.$$

Then $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{G}^\top\mathbf{W}\mathbf{G}\right)^{-1}\mathbf{G}^\top\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}\mathbf{G}\left(\mathbf{G}^\top\mathbf{W}\mathbf{G}\right)^{-1}\right)$ *for* $n \to \infty$.

This is where the choice of $\mathbf{W}$ is of importance as it can heavily influence the variance of the estimator. Certain choices simplify the expression for the asymptotic variance of the estimator. Two notable candidates are $\mathbf{W} = \mathbf{I}_m$, the identity matrix, and $\mathbf{W} = \mathbf{W}_{\boldsymbol{\theta}^0} = \boldsymbol{\Sigma}^{-1}$. Both yield $\left(\mathbf{G}^\top \boldsymbol{\Sigma} \mathbf{G}\right)^{-1}$ as asymptotic variance matrix. Here the impact of using a two-step procedure can be observed as the matrix $\mathbf{G}$ incorporates the added uncertainty. Though the two options for $\mathbf{W}$ mentioned above are asymptotically equivalent, in finite samples the choice might impact the parameter estimates as different moment equations have a different impact on the criterion function of the GMM estimator. In Section S.4 of the Supplementary Material we provide a small simulation study to illustrate the impact of the choice of matrix $\mathbf{W}$ on finite sample level.

Most conditions of Theorem 1 are easily verified with the exception of condition (v). This states that $\widehat{\mathbf{g}}_n$ as in (6), obeys some form of stochastic equicontinuity. Stochastic equicontinuity is defined as follows ([2]). $\forall \varepsilon > 0$, $\exists \delta > 0$ it holds that

$$\lim_{n\to\infty} P\left(\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_{\left\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\right\|<\delta} \left\|\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) - \widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\tilde{\boldsymbol{\theta}})\right\| > \varepsilon\right) < \varepsilon.$$

Primitive conditions for stochastic equicontinuity are given in [32], [31], [2] or [30]. However, what we need to show is a slightly stronger form with the added square-root-n behavior. This can for example be proven by imposing a Lipschitz condition on $\widehat{g}_n(;\boldsymbol{\theta})$ with respect to the parameter vector. A second way to fulfill the assumption is by a remark made in [30] p2186-2187 which states that (v) holds in case $\widehat{\mathbf{g}}_n(\theta) \overset{unif.P}{\to} g_0(\theta)$, i.e. uniform in probability convergence holds for $\widehat{\mathbf{g}}_n$. Lemma S.1 allows us to meet this stochastic equicontinuity condition of $\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})$ by achieving a uniform convergence result. The proof that our model satisfies the conditions of Lemma S.1 is provided in Section S.2 of the Supplementary Material. This result on uniform convergence in turn allows us to establish asymptotic normality of the IFM estimator in Section 3.2.

3. Asymptotics for the IFM parameter estimators

3.1. Consistency

Since we consider the IFM estimator as a special case of a GMM estimator, conditions under which consistency of GMM estimators hold can also be used to show consistency of the IFM estimator. Theorem 2.6 of [30] gives sufficient conditions under which a GMM estimator is consistent. For convenience of the reader, we recall this result in Theorem S.1 of Section S.1 in the Supplementary Material. Theorem 2 states the consistency result for our estimators $\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}}$.

Theorem 2. *Suppose an estimated matrix $\widehat{\mathbf{W}}$ is such that $\widehat{\mathbf{W}} \xrightarrow{P} \mathbf{W}$, where $\mathbf{W}$ is positive semi-definite and invertible. Under Assumptions (M1), (M3), [(M4) or (M6)], (M7) and (M8), and supposing that for the copula Assumptions (R1) to (R3) hold, the IFM estimator for the copula with quantile-based asymmetric margins is consistent, i.e. $\widehat{\boldsymbol{\theta}}_n^{IFM} \xrightarrow{P} \boldsymbol{\theta}_0$ for $n \to \infty$.*

Proof. The proof is based on Theorem S.1, which gives sufficient conditions under which a GMM estimator is consistent. We thus need to check whether the conditions of Theorem S.1 are satisfied with $\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})$ as in (11).

Checking Condition $(C1)$. Weak consistency of $\widehat{\mathbf{W}}$ holds by assumption.

Checking Condition $(C2)$. Per assumption $\mathbf{W}$ is an invertible positive (semi-)definite matrix. Since we have the same number of moment equations as there are parameters, the linear system of equations $\mathbf{W}\mathbf{g}_0(\boldsymbol{\theta}) = \mathbf{0}$ has single solution, which is $\mathbf{g}_0(\boldsymbol{\theta}) = \mathbf{0}$. For $E[\mathbf{g}_{\boldsymbol{\eta}_j}(\mathbf{Z};\boldsymbol{\eta}_j)]$,under Assumption (M4) (or (M6)), this is the case if and only if $\boldsymbol{\eta}_j = \boldsymbol{\eta}_j^0$ (see [12]). By Assumption (R2), $\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\boldsymbol{\eta}^0,\boldsymbol{\theta}_C)$ is continuous and hence, by the Leibniz rule, derivative and integral can be switched. A classical likelihood argument combined with the assumed identifiability then states that $E[\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\boldsymbol{\eta}^0,\boldsymbol{\theta}_C)] = \mathbf{0}$ if and only if $\boldsymbol{\theta}_C = \boldsymbol{\theta}_C^0$. Hence $\mathbf{W}\mathbf{g}_0(\boldsymbol{\theta}) = 0$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^0$.

Checking Condition $(C3)$. This holds by Assumptions (M1) and (R1).

Checking Condition $(C4)$. For $\mathbf{g}_{\boldsymbol{\theta}_C}(\mathbf{Z};\boldsymbol{\eta},\boldsymbol{\theta}_C)$ this condition is satisfied since it is continuously differentiable with respect to $\boldsymbol{\theta}$. The problem lies in the $\mathbf{g}_{\boldsymbol{\eta}_j}(\mathbf{Z};\boldsymbol{\eta}_j)$ as $\ln\left(f_{Z_j}(Z_j,\boldsymbol{\eta}_j)\right)$ is may not be differentiable with respect to μ_j in the point $z_j = \mu_j$. Hence, for the d hyperplanes formed by $z_j = \mu_j$, $j = 1,\ldots,d$, $\mathbf{g}_{\boldsymbol{\eta}_j}(\mathbf{Z};\boldsymbol{\eta}_j)$ may not be continuous. The condition however states that $\mathbf{g}_{\boldsymbol{\eta}_j}(\mathbf{Z};\boldsymbol{\eta}_j)$ must be continuous almost everywhere. This is satisfied as a hyperplane has measure 0 and the finite union of objects with measure 0 also has measure 0. Thus $\mathbf{g}_{\boldsymbol{\eta}_j}(\mathbf{Z};\boldsymbol{\eta}_j)$ satisfies the condition.

Checking Condition ($C5$). This condition is shown to hold for the $\mathcal{L}_1$-norm. This is allowed as we are working in $\mathbb{R}^m$, and it holds that for any $\mathbf{x} \in \mathbb{R}^m$, that $\|\mathbf{x}\| \leq \|\mathbf{x}\|_1$. Furthermore, it suffices to show the condition holds for each component of $\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})$, since $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})\|_1 \leq \left\|\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbf{g}(\mathbf{Z};\boldsymbol{\theta})\right\|_1$. It can be shown component-wise that each element of $\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})$ can be dominated by a function with finite expectation. Some examples are given in Section S.2 in the Supplementary material. Thus condition ($C5$) is satisfied.

In conclusion, consistency of $\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}}$ thus follows from Theorem S.1. □

Remark 2. Note that considering $\widehat{\mathbf{W}}$ a consistent estimator for the variance-covariance matrix $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$, or $\widehat{\mathbf{W}} = \mathbf{I}_m$, would satisfy the assumptions imposed on $\widehat{\mathbf{W}}$ in Theorems 1 and 2. Since the variance-covariance matrix depends on $\boldsymbol{\theta}$ we in fact have $\widehat{\mathbf{W}}_{\boldsymbol{\theta}}$ in this case (see also Theorem S.1). The above choices would lead to the same vastly simplified variance-covariance matrix of the asymptotic normality result in Theorem 1.

Within the proof of consistency, it is implicitly assumed that identifiability of the model holds. This however is guaranteed since identifiability of the model parameters underlying the estimation problem in each of the two estimation steps is ensured. Indeed, for the marginal parameters, this is proven in [12]; and for the copula parameter, this follows from the uniqueness of the copula given that the margins are continuous ([36]).

3.2. Asymptotic normality

With the consistency results available we now tackle the asymptotic normality of the IFM estimator. This is done by adapting Theorem 1 to our framework and checking whether the conditions are satisfied. Before stating the asymptotic normality result, we introduce some notations. Denote

$$\mathbf{G} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\eta}^0} & \mathbf{0} \\ \mathcal{I}^{\top}_{\boldsymbol{\eta}^0\boldsymbol{\theta}^0_C} & \mathcal{I}_{\boldsymbol{\theta}^0_C} \end{bmatrix}, \tag{13}$$

with

$$\mathcal{I}_{\boldsymbol{\eta}^0} = -E\left[\frac{\partial}{\partial\boldsymbol{\eta}}\left(\sum_{j=1}^{d}\overline{\frac{\partial}{\partial\boldsymbol{\eta}}}\ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right)\right)\right]\Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}^0}, \mathcal{I}_{\boldsymbol{\theta}_C^0} = -E\left[\frac{\partial^2}{\partial\boldsymbol{\theta}_C^2}\ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C)\right)\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0},$$

$$\mathcal{I}_{\boldsymbol{\eta}^0\boldsymbol{\theta}_C^0} = -E\left[\frac{\partial}{\partial\boldsymbol{\theta}_C^\top}\left(\overline{\frac{\partial}{\partial\boldsymbol{\eta}}}\ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\mathbf{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C)\right)\right)\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}, \tag{14}$$

with $\mathcal{I}_{\boldsymbol{\eta}^0} \in \mathbb{R}^{q\times q}$ $\mathcal{I}_{\boldsymbol{\theta}_C^0} \in \mathbb{R}^{p\times p}$ and $\mathcal{I}_{\boldsymbol{\eta}^0} \in \mathbb{R}^{p\times q}$.

Theorem 3. *Assume Assumptions (M1)–(M4) (or (M5)–(M6)) and (M7)–(M8) hold for Z_j, $\forall j = 1,\dots,d$. Further suppose Assumptions (R1)– (R4) hold as well as*

(N1) $\widehat{\mathbf{g}}_n\left(\underset{\sim}{\mathbf{Z}};\widehat{\boldsymbol{\theta}}_n^{IFM}\right)^\top \widehat{\mathbf{g}}_n\left(\underset{\sim}{\mathbf{Z}};\widehat{\boldsymbol{\theta}}_n^{IFM}\right) \leq \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})^\top\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) + o_p(n^{-1})$,

(N2) $\mathbf{W}_{\boldsymbol{\theta}^0} = \mathbf{I}_p$.

Then the IFM estimator is asymptotically normally distributed, i.e.

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{IFM} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{G}^\top\boldsymbol{\Sigma}\mathbf{G}\right)^{-1}\right), \qquad \textit{for } n \to \infty,$$

with $\mathbf{G}$ *as in* (13) *and with* $\boldsymbol{\Sigma}^{-1}$ *as in Condition (iv) of Theorem 1.*

Proof. First, Conditions $(I)-(III)$ from Theorem 1 need to be checked. Condition (I) holds by Assumption $(N1)$. By Assumptions (M1), (M3), [(M4) or (M6)], (M7), (M8) and (R1) and (R3), Theorem 2 holds true which implies that the estimator is consistent. Hence Condition (II) is satisfied. Condition (III) was shown to hold in Theorem 2 by Assumption $(N2)$ and if Condition (iv) of Theorem 1 is satisfied.

We next investigate the remaining Conditions $(i)-(v)$ of Theorem 1.

Checking Condition (i). From (12) and the strong law of large numbers applied to $\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta})$ (since $E[\left|\mathbf{g}(\mathbf{Z}^{(i)},\boldsymbol{\theta})\right|] < \infty$),

$$\mathbf{g}_0(\boldsymbol{\theta}) = E\begin{bmatrix}\overline{\frac{\partial}{\partial\boldsymbol{\eta}_1}}\ln\left(f_{Z_1}(Z_1;\boldsymbol{\eta}_1)\right)\\ \vdots \\ \overline{\frac{\partial}{\partial\boldsymbol{\eta}_d}}\ln\left(f_{Z_d}(Z_d;\boldsymbol{\eta}_d)\right)\\ \frac{\partial}{\partial\boldsymbol{\theta}_C}\ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\mathbf{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C)\right)\end{bmatrix}.$$

By (4) and arguments in the proof of Theorem 2, it follows that $\mathbf{g}_0(\boldsymbol{\theta}^0) = \mathbf{0}$.

Checking Condition (ii). $\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln\left(c(\boldsymbol{F}_{\boldsymbol{Z}}(\mathbf{Z};\boldsymbol{\eta});\boldsymbol{\theta}_C)\right)$ is differentiable with respect to $\boldsymbol{\eta}_j$. Thus in extension also its expectation. In addition, $E\left[\frac{\partial}{\partial \boldsymbol{\eta}_j} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right)\right]$ is differentiable with respect to $\boldsymbol{\eta}_j$ in $\boldsymbol{\eta}^0$ (see Theorem 3.4 in [12]). This makes that $\mathbf{g}_0(\boldsymbol{\theta}^0)$ is differentiable with respect to $\boldsymbol{\eta}_j$. As the copula parameter does not appear in the margins, by Assumption (R2), $\mathbf{g}_0(\boldsymbol{\theta}^0)$ is differentiable with respect to the copula parameters as well. This concludes the differentiability part.

For the second part of Condition (*ii*). By the formulation of Condition (*ii*), it follows that

$$\mathbf{G} = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}_0(\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0},$$

leading to the expression in (13). Invertibility of $\mathbf{G}$ is guaranteed by invertibility of $\mathcal{I}_{\boldsymbol{\eta}}$ and $\mathcal{I}_{\boldsymbol{\theta}_C}$ with

$$\mathbf{G}^{-1} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\eta}^0}^{-1} & \mathbf{0} \\ \mathcal{I}_{\boldsymbol{\theta}_C^0}^{-1} \mathcal{I}_{\boldsymbol{\eta}^0 \boldsymbol{\theta}_C^0}^{\top} \mathcal{I}_{\boldsymbol{\eta}^0}^{-1} & \mathcal{I}_{\boldsymbol{\theta}_C^0}^{-1} \end{bmatrix}.$$

Hence $\left(\mathbf{G}^{\top}\mathbf{W}_{\boldsymbol{\theta}^0}\mathbf{G}\right)^{-1} = \mathbf{G}^{-1}(\mathbf{G}^{\top})^{-1}$ exists, and $\mathbf{G}^{\top}\mathbf{W}_{\boldsymbol{\theta}^0}\mathbf{G}$ is thus non-singular.

Checking Condition (iii). This is satisfied by (M1) and (R1).

Checking Condition (iv). The quantity $\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}^0)$ is an empirical mean with $E[\mathbf{g}(\mathbf{Z}^{(i)};\boldsymbol{\theta}^0)] = \mathbf{g}_0(\boldsymbol{\theta}^0) = \mathbf{0}$ and $E[\mathbf{g}(\mathbf{Z}^{(i)};\boldsymbol{\theta}^0)\mathbf{g}(\mathbf{Z}^{(i)};\boldsymbol{\theta}^0)^{\top}] < \infty$. By the multivariate central limit theorem, it follows that $\sqrt{n}\widehat{\mathbf{g}}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$, for $n \to \infty$.

Checking Condition (v). The last condition is satisfied by Proposition 1.

In conclusion, by Theorem 1, it follows that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \left(\mathbf{G}^{\top}\boldsymbol{\Sigma}\mathbf{G}\right)^{-1}\right)$, for $n \to \infty$. □

3.3. Assumptions (R1)—(R4): illustrative examples

As a first illustration of a copula where all assumptions (R1)—(R4) are met, we consider the bivariate Gaussian copula with copula density

$$c(\rho;u_1,u_2) = \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\rho^2\Phi^{-1}(u_1)^2+\rho^2\Phi^{-1}(u_2)^2-2\rho\Phi^{-1}(u_1)\Phi^{-1}(u_2)\right)},$$

in which $\rho \in \mathcal{N}(\rho^0) \subset (-1,1)$ and Φ^{-1} is the inverse of the cumulative distribution function of a standard normal distribution. Assumption (R1) is thus met. For Assumption (R2), the following expressions are required, in which $x_j = \Phi^{-1}(u_j)$, $j = 1,2$

$$\frac{\partial}{\partial \rho} \ln \left(c(\rho; u_1, u_2)\right) = \frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}(x_1^2 + x_2^2) + \frac{1+\rho^2}{(1-\rho^2)^2} x_1 x_2,$$

and $$\frac{\partial^2}{\partial \rho^2} \ln \left(c(\rho; u_1, u_2)\right) = -\frac{\rho^4 - 1 + (3\rho^2+1)(x_1^2+x_2^2) - 2\rho(\rho^2+3)x_1 x_2}{(1-\rho^2)^3}.$$

Thus, the first two derivatives exist and are finite under Assumption (R1) and $u_j \neq 0, 1$. For Assumptions (R3) and (R4) moments of the x_j are required. Consider as an illustration $E[\Phi^{-1}(U_1)^k]$. By definition and by using the substitution $\Phi^{-1}(u_j) = x_j$ this equals

$$\begin{aligned} E[\Phi^{-1}(U_1)^k] &= \int_{[0,1]^2} \Phi^{-1}(u_1)^k c(\rho; u_1, u_2) du_1 du_2 \\ &= \int_{\mathbb{R}^2} x_1^k \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(-1\rho^2)}(x_1^2 + x_2^2 - 2\rho x_1 x_2)} dx_1 dx_2, \end{aligned}$$

which is the k-th moment of a standard bivariate normal distribution with correlation ρ. To check Assumption (R3), we can thus evaluate the resulting expression using the moments of $(X_1, X_2) \sim \mathcal{N}_2(\mathbf{0}, \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right])$. We need $E[X_1^4] = E[X_2^4] = 3$, $E[X_1^2] = E[X_2^2] = 1$, $E[X_1^3 X_2] = E[X_1 X_2^3] = 3\rho$, $E[X_1^2 X_2^2] = 1 + 2\rho^2$ and $E[X_1 X_2] = \rho$. This gives

$$E\left[\left(\frac{\partial}{\partial \rho} \ln \left(c(\rho; U_1, U_2)\right)\right)^2\right] = \frac{1+\rho^2}{(1-\rho^2)^2},$$

which is strictly positive and finite under (R1).

For Assumption (R4) a similar reasoning can be used. Under Assumption (R1), with ρ not near the boundary of $(-1,1)$, we can set

$$H(\mathbf{u}) = \sup_{\rho \in \mathcal{N}(\rho^0)} \left\| \frac{\partial}{\partial \rho} \ln \left(c(\rho; u_1, u_2)\right) \right\|.$$

Then $E[H(\mathbf{U})]$ is related to some finite combination of moments of X_j which results in (R4) holding true. The last assumption to check is $(N1)$ as $(N2)$ is a fixed choice to be made. For this, is suffices that the margin parameter estimator and that of the copula parameter are root-n consistent. For the QBA-distributed margins, this assumption is satisfied by (5). For ρ, we rely on the result that for a bivariate normal distribution, the MLE of the correlation parameter is indeed also root-n consistent by classical likelihood theory arguments. A similar reasoning can be made for other copulas based on elliptical distributions, as long as the first few moments exist.

A second illustrative example concerning the mildness of Assumptions (R1)—(R4) is provided in Section S.3 of the Supplementary Material.

4. Simulation study

The simulation study consists of two parts. In a first exploratory part, with Setting 1, the aim is to briefly investigate the finite-sample performance of the IFM estimators of the copula and marginal parameters, focusing on the behaviour for increasing sample size. In a second part, with Setting 2, the aim is to study the finite-sample distributions of the estimated parameters more in detail, keeping in mind the asymptotic results obtained in Section 3.2. Throughout the simulation study as well as in Section 5 we consider for the margins the following distributions: the QBA-normal, the QBA-logistic, the QBA-Laplace, and the QBA-Student's t-distributions, with densities as in (3) with as reference density f_j the density of respectively a symmetric standard normal, logistic, Laplace and Student's t.

Throughout this section 1 000 samples of sample size n are simulated from a considered simulation model. From the r-th sample, with $r = 1, \ldots, 1\,000$, we obtain the IFM estimate, denoted $\widehat{\theta}_j^{(r)}$, of a model parameter θ_j. Approximations of the bias, variance and mean squared error (MSE) for each parameter θ_j are then calculated via

$$\text{Abias}(\widehat{\theta}_j) = \frac{1}{1\,000}\sum_{r=1}^{1\,000}\widehat{\theta}_j^{(r)} - \theta_j\,, \qquad \text{AVar}(\widehat{\theta}_j) = \frac{1}{999}\sum_{r=1}^{1\,000}\left(\widehat{\theta}_j^{(r)} - \frac{1}{1\,000}\sum_{l=1}^{1\,000}\widehat{\theta}_j^{(r)}\right)^2$$
$$\text{and} \quad \text{AMSE}(\widehat{\theta}_j) = \text{AVar}(\widehat{\theta}_j) + \text{Abias}(\widehat{\theta}_j)^2. \qquad (15)$$

Setting 1. In this setting we consider three copula models: a ten-dimensional Frank copula, a bivariate Gaussian copula and a six-dimensional Student's t-copula. Margins and their parameters are kept the same in each of the three models: as such, for example, the margins of the Gaussian copula correspond to the first two margins of the Frank Copula.

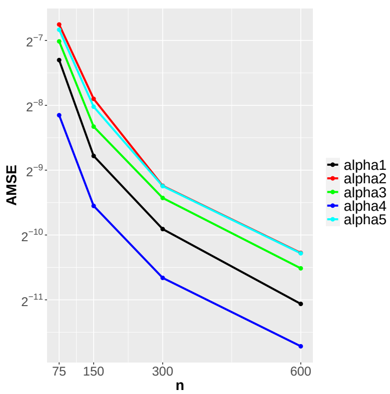
The marginal distributions F_{Z_j} and their parameters are listed in Table 1 (see the columns under the heading 'Setting 1'). Copula parameters are chosen such that they have a pairwise Kendall's tau of approximately 0.2, except for the Student's t-copula, which has pairwise Kendall type dependencies of at most 0.35.

Table 1: Margin distributions and parameters for Settings 1 and 2.

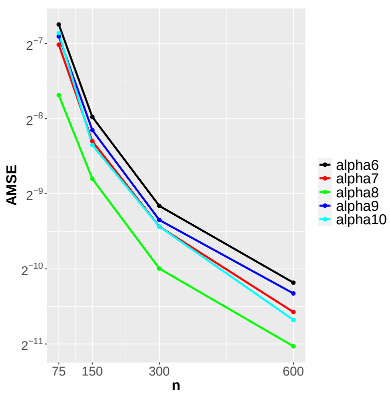
Setting 1						Setting 2				
Margin	Type	α_j	μ_j	ϕ_j	ν_j	Margin	Type	α_j	μ_j	ϕ_j
d_1	QBA-Student's t	0.2	2	0.2	4	d_1	QBA-Laplace	0.4	1	0.5
d_2	QBA-normal	0.7	3	0.4		d_2	QBA-normal	0.7	2	2.9
d_3	QBA-logistic	0.3	-1	0.1		d_3	QBA-logistic	0.3	3	0.1
d_4	QBA-Laplace	0.8	-1	2						
d_5	QBA-Student's t	0.6	4	1	15					
d_6	QBA-normal	0.3	-2	1						
d_7	QBA-logistic	0.3	1	10						
d_8	QBA-Laplace	0.5	-2	0.5						
d_9	QBA-Student's t	0.7	-3	5	25					
d_{10}	QBA-normal	0.2	4	3						

The Archimedean Frank copula is characterized by a single parameter θ_C which governs the dependency. We take $\theta_C = 2$, which corresponds to a weak dependency. For the bivariate Gaussian copula $\theta_C = \rho = 0.2$, whereas for the six-dimensional Student's t-copula $\boldsymbol{\theta}_C = (\text{Vec}(\boldsymbol{\Sigma})^\top, \nu_C)^\top$, with $\nu_C = 3$ and the 6×6 correlation matrix $\boldsymbol{\Sigma}$ given in (S.2) in the Supplementary Material.

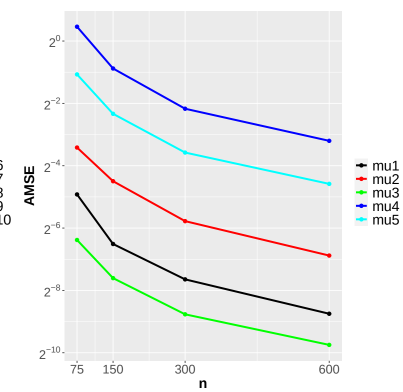
In Setting 1 we consider sample sizes n equal to 75, 150, 300 and 600 (hence doubling each time the sample size). Note that the number of parameters to be estimated in the simulation models is considerable: 34, 8 and 36 parameters for, respectively, the 10-dimensional Frank copula model; the 2-dimensional Gaussian copula model; and the 6-dimensional Student's t-copula model. By employing IFM we estimate each margin separately. Each margin contains at most 4 parameters, and we thus have at least 75 observations to get



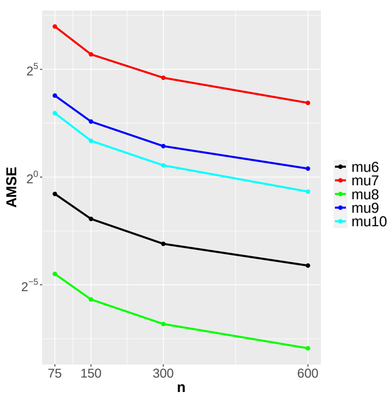
(a) $\boldsymbol{\alpha}$ (part 1).



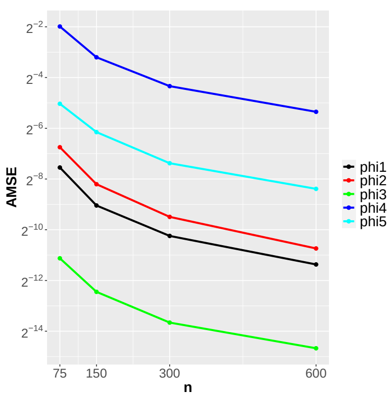
(b) $\boldsymbol{\alpha}$ (part 2).



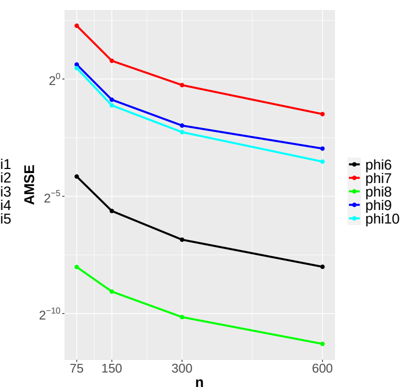
(c) $\boldsymbol{\mu}$ (part 1).



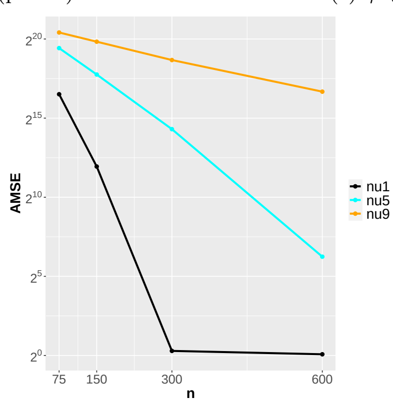
(d) $\boldsymbol{\mu}$ (part 2).



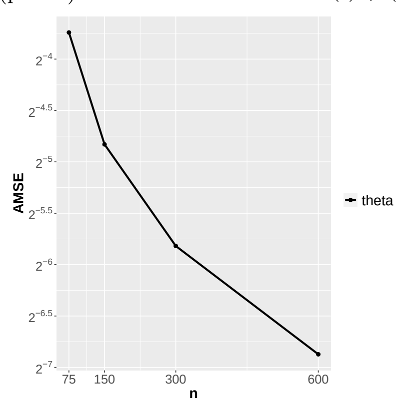
(e) $\boldsymbol{\phi}$ (part 1).



(f) $\boldsymbol{\phi}$ (part 2).



(g) $\boldsymbol{\nu}$.



(h) θ_C.

Figure 1: Setting 1. Approximate Mean Squared Error of the estimated (36) parameters for the ten dimensional Frank-copula with parameters as in Table 1 and $\theta_C = 2$. Results for sample sizes 75, 150, 300 and 600, presented on a vertical $\log_2$-scale.

estimates for them. So in terms of quality of the marginal parameter estimates, we would expect decent results. For all models, margin estimation is performed by the `bobyqa`-

algorithm contained in the R-package nloptr (see [33] and [23]). Twenty different starting values are used per margin and as a convergence criterion, a relative step size of 10^{-5} is used. The number of iterations is capped at 50 000. Regarding estimation of the copula parameters, this is performed under the default setting of the fitCopula function contained in the copula-package (see [19]). Because of numerical stability, the optimization method used differs for the three copula models. For the Frank copula, L-BFGS-B is used, for the Gaussian copula Nelder-Mead and for the Student's t-copula BFGS. In essence this does not make a difference in the quality of the estimation, but it leads to a significant difference in computation time. This is most noticeable for the elliptical copulas as BFGS takes longer to converge (given models with the same dimensions).

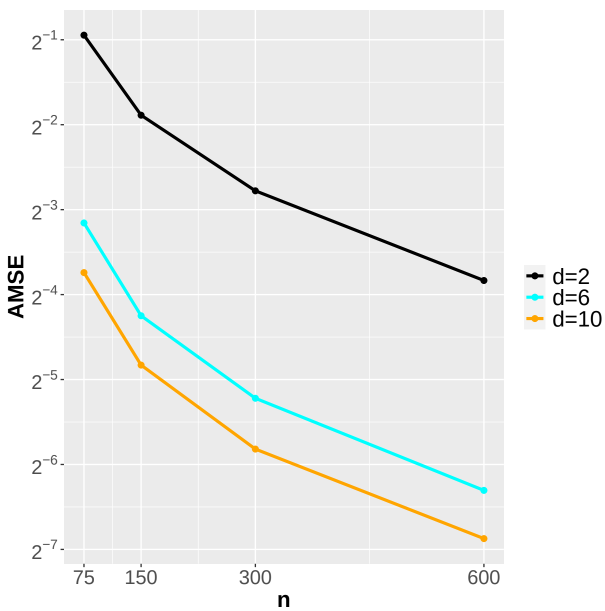


Figure 2: Setting 1. Approximate Mean Squared Error of θ_C for the Frank copula of dimensions 2, 6 and 10. Results for sample sizes 75, 150, 300 and 600, presented on a vertical $\log_2$-scale.

Figure 1 presents the AMSE-values, on a logarithm with base number 2 scale, for the parameter estimates of the Frank copula model. The simulation results regarding the Gaussian and Student's t-copula are presented in the Supplementary Material in Figures S.1 and S.2, respectively. All simulation results show a very good finite-sample performance of the IFM estimation procedure.

Figure 2 summarizes results on AMSE-values (on a $\log_2$-scale) for simulations from a two-, six- and ten-dimensional Frank copula model. This shows a decreasing AMSE of θ_C for increasing dimensions and increasing sample size. The effect of the sample size is as expected; however, the impact of the dimensionality might seem strange at first sight as the AMSE actually improves if we increase the dimensionality. This contradicts the curse of dimensionality. This phenomenon is also reported in [17] and [8]. This is due to the fact that an Archimedean copula is exchangeable, which circumvents the curse of dimensionality (see [17]) and turns it into a blessing of dimensionality.

Table 2: Setting 2: approximate bias, variance and ratio of approximate variance and asymptotic variance. Average $\widehat{\text{As.Var}}$ denotes the average of the estimated asymptotic variance over the different replications.

n	Parameter	α_1	μ_1	ϕ_1	α_2	μ_2	ϕ_2	α_3	μ_3	ϕ_3	θ_C
	True value	0.4	1	0.5	0.4	2	2.9	0.3	3	0.1	1.3
25	ABias	−0.0040	−0.0059	−0.0724	0.0326	0.3187	−0.9689	−0.0218	−0.0113	−0.0286	−0.0098
	AVar	0.0243	0.2926	0.0186	0.0495	22.4019	1.5661	0.0433	0.0636	0.0017	0.0264
	Ratio	2.5302	3.5113	1.5925	2.6702	2.1141	2.1356	2.7408	2.4376	2.0482	1.0335
	Average $\widehat{\text{As.Var}}$	0.0086	0.0709	0.0113	0.0130	6.8762	0.5304	0.0118	0.0183	0.0006	0.0256
50	ABias	−0.0003	0.0034	−0.0254	0.0173	0.2712	−0.4068	−0.0154	−0.0116	−0.0121	0.0069
	AVar	0.0079	0.0832	0.0069	0.0179	8.9105	0.7708	0.0144	0.0226	0.0008	0.0132
	Ratio	1.6452	1.9963	1.1834	1.9342	1.6818	2.1021	1.8228	1.7297	1.9260	1.0345
	Average $\widehat{\text{As.Var}}$	0.0048	0.0384	0.0056	0.0089	4.6942	0.3355	0.0077	0.0113	0.0004	0.0124
100	ABias	−0.0006	−0.0001	−0.0098	0.0084	0.1595	−0.1663	−0.0073	−0.0066	−0.0048	0.0041
	AVar	0.0034	0.0350	0.0032	0.0064	3.5964	0.2780	0.0051	0.0082	0.0003	0.0070
	Ratio	1.4346	1.6806	1.0986	1.3762	1.3576	1.5161	1.2789	1.2570	1.3980	1.0980
	Average $\widehat{\text{As.Var}}$	0.0024	0.0205	0.0030	0.0044	2.4820	0.1825	0.0038	0.0062	0.0002	0.0065
250	ABias	−0.0010	0.0021	−0.0043	0.0036	0.0527	−0.0635	−0.0018	−0.0006	−0.0014	0.0015
	AVar	0.0012	0.0121	0.0012	0.0021	1.1617	0.0830	0.0017	0.0028	0.0001	0.0027
	Ratio	1.2617	1.4473	0.9975	1.1351	1.0963	1.1314	1.0813	1.0849	1.1389	1.0694
	Average $\widehat{\text{As.Var}}$	0.0010	0.0083	0.0012	0.0018	1.0330	0.0738	0.0016	0.0026	0.0001	0.0026
1 000	ABias	−0.0007	−0.0016	−0.0009	0.0008	0.0143	−0.0124	−0.0003	−0.0001	−0.0004	0.0015
	AVar	0.0003	0.0026	0.0003	0.0005	0.2771	0.0183	0.0004	0.0007	0.0000	0.0006
	Ratio	1.1737	1.2295	1.0552	1.0519	1.0461	1.0003	1.0912	1.0913	1.0510	1.0157
	Average $\widehat{\text{As.Var}}$	0.0002	0.0021	0.0003	0.0005	0.2635	0.0185	0.0004	0.0007	0.0000	0.0006

Setting 2. In this setting, we investigate the finite-sample distribution of the IFM estimator. We consider a Gumbel copula with parameter $\theta_C = 1.3$ and margins as indicated in Table 1 (under the heading 'Setting 2'). With this copula parameter we have again a

rather weak dependence structure, with a bivariate Kendall's tau equal to approximately 0.23. The simulation model thus includes 10 parameters. We draw samples of sizes $n = 25$, 50, 100, 250 and 1 000. For each sample size we summarize the simulation results by presenting the approximate bias, variance and MSE, calculated from (15). Furthermore, for each parameter, the approximate (finite-sample) variance is compared to the theoretical (asymptotic) variance established in Theorem 3, and computed via $n^{-1}\left(\mathbf{G}^{\top}\mathbf{\Sigma}\mathbf{G}\right)^{-1}$ with $\mathbf{\Sigma}$ and $\mathbf{G}$ as in *(iv)* of Theorem 1 and (13), respectively. For presentational ease we report the ratio of the approximate (finite-sample) variance and the theoretical (asymptotic) variance. The results can be found in Table 2. To give an idea about the quality of the estimated asymptotic variance, obtained by substituting the estimated parameters into the expression for the asymptotic variance we also report on the average of this estimated (asymptotic) variance over the different replications (denoted as Average $\widehat{\text{As.Var}}$ in the Table 2). Note that the approximate variance approaches the asymptotic (theoretical) variance, at the expected n^{-1}-rate. The ratio gets closer to one with increasing sample size. The approximate bias is smaller for larger sample sizes. Boxplots of the parameter estimates are shown in Figures S.4 and S.5 in the Supplementary Material. These confirm the behaviour summarized in Table 2.

In Sections S.5 and S.6 of the Supplementary Material we further provide illustrations, via simulations in the above Setting 2, of the finite-sample distributions of the estimators, as compared to the asymptotic normal distribution.

5. Real data examples

In the two real data examples, we use a parametric copula C with flexible skew asymmetric marginal distributions coming from the family defined in (3). We consider the same four candidate QBA-distributions as in Section 4: the QBA-normal, the QBA-logistic, the QBA-Laplace, and the QBA-Student's t-distributions. Candidates for the copula C are four Archimedean copulas (Gumbel, Joe, Frank and Clayton copula), as well as a Gaussian and Student's t-copula. These six copulas together with the four margins leads to $6 + 4 \times d$ possible parametric models that we consider in our analysis. As a benchmark to compare the fitted distributions to, we use the multivariate skew-normal and the skew-t distribution, available in the literature.

The comparison of different models is based on an information criterion. A commonly-used information criterion is Akaike's information criterion (AIC, [1]):

$$\mathrm{AIC}_n = -2\ell\left(\widehat{\boldsymbol{\theta}}_n^{\mathrm{ML}}; \mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(n)}\right) + 2k, \tag{16}$$

in which k is the number of model parameters to be estimated. Note that since we used IFM parameter estimation instead of MLE, using the AIC formula above is no longer completely correct. In order to give accurate results, the penalty should be slightly modified. This results into the copula information criteria (CIC), which is of the form $\mathrm{CIC}_n = -2\ell\left(\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}}; \mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(n)}\right) + 2\widetilde{k}_n$, where $\widetilde{k}_n$ is a penalty term based on the Fisher information and variance of the score functions of both the margins and the copula. More precisely

$$\widetilde{k}_n = \sum_{j=1}^{d} \widetilde{k}_{\boldsymbol{\eta}_j} + \widetilde{k}_{\boldsymbol{\theta}_C} = \sum_{j=1}^{d} \mathrm{tr}\left(I_{\boldsymbol{\eta}_j}^{-1}\mathcal{K}_{\boldsymbol{\eta}_j}\right) + \mathrm{tr}\left(I_{\boldsymbol{\eta}}^{-1}\mathcal{K}_{\boldsymbol{\eta}}^{*}\right) + \mathrm{tr}\left(-I_{\boldsymbol{\theta}_C}^{-1} I_{\boldsymbol{\eta},\boldsymbol{\theta}_C}^{T} I_{\boldsymbol{\eta}}^{-1}\mathcal{K}_{\boldsymbol{\eta},\boldsymbol{\theta}_C} + I_{\boldsymbol{\theta}_C}^{-1}\mathcal{K}_{\boldsymbol{\theta}_C}\right),$$

in which $I_{\boldsymbol{\eta}_j}$, $I_{\boldsymbol{\theta}_C}$ and $I_{\boldsymbol{\eta},\boldsymbol{\theta}_C}$ are as defined in (14) and

$$\begin{aligned}
\mathcal{K}_{\boldsymbol{\eta}_j} &= E\left[\left(\frac{\partial}{\partial \boldsymbol{\eta}_j} \ln f_{Z_j}(Z_j; \boldsymbol{\eta}_j)\right)\left(\frac{\partial}{\partial \boldsymbol{\eta}_j} \ln f_{Z_j}(Z_j; \boldsymbol{\eta}_j)\right)^{\top}\right], \\
\mathcal{K}_{\boldsymbol{\theta}_C} &= E\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\eta}); \boldsymbol{\theta}_C)\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\eta}); \boldsymbol{\theta}_C)\right)^{\top}\right], \\
\mathcal{K}_{\boldsymbol{\eta},\boldsymbol{\theta}_C} &= E\left[\left(\frac{\partial}{\partial \boldsymbol{\eta}_j} \ln f_{Z_j}(Z_j; \boldsymbol{\eta}_j)\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}_C} \ln c(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\eta}); \boldsymbol{\theta}_C)\right)^{\top}\right], \\
\text{and} \quad \mathcal{K}_{\boldsymbol{\eta}}^{*} &= E\left[\left(\frac{\partial}{\partial \boldsymbol{\eta}_j} \ln f_{Z_j}(Z_j; \boldsymbol{\eta}_j)\right)\left(\frac{\partial}{\partial \boldsymbol{\eta}} \ln c(F_{Z_j}(\boldsymbol{F}_{\boldsymbol{Z}}(\boldsymbol{z}; \boldsymbol{\eta}); \boldsymbol{\theta}_C)\right)^{\top}\right].
\end{aligned}$$

If the margins and copula are correctly specified and regularity conditions fulfilled, then the CIC penalty simplifies to the classical AIC penalty as shown in [2] and the likelihood contribution of (16) is obtained using the IFM parameter estimates. Hence using the IFM two-step maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}}$ in (16) one gets the two-stage AIC criterion, denoted by AIC_{2ML} in the tables, in which the log-likelihood is given by

$$\ell\left(\widehat{\boldsymbol{\theta}}_n^{\mathrm{IFM}}; \underset{\sim}{\mathbf{Z}}\right) = \sum_{i=1}^{n}\left(\ln\left(c\big(F_{Z_1}(Z_1^{(i)}; \widehat{\boldsymbol{\eta}}_{1,n}^{\mathrm{IFM}}), \ldots, F_{Z_d}(Z_d^{(i)}; \widehat{\boldsymbol{\eta}}_{d,n}^{\mathrm{IFM}}); \boldsymbol{\theta}_C\big)\right) + \sum_{j=1}^{d}\ln\left(f_{Z_j}(Z_j^{(i)}; \widehat{\boldsymbol{\eta}}_{j,n}^{\mathrm{IFM}})\right)\right).$$

A full exposition on the above information criteria is given in [25]. As noted in the latter paper, the difference between AIC_{2ML} and CIC is only notable in small samples. In larger samples, this difference is minimal. This justifies the use of the AIC_{2ML} criterion for comparison here. As for the skew-normal and skew-t distributions parameter estimates are obtained through MLE, and hence (16) remains valid. In our analysis we proceed as follows. The best fitted margins are chosen using the AIC criterion. Using the pseudo-observations obtained from the best fitting margins, a copula is chosen (from the 6 considered copulas), and the copula that leads to a minimal CIC is then considered the best copula.

Since we only consider 4 candidate margins, we carry out a test to see whether or not the chosen margin (denoted F_{Z_j}) is appropriate for the corresponding variable. For this purpose a Kolmogorov-Smirnov (KS) goodness-of-fit test (see [27] and [37]) is conducted. This test compares the empirical cumulative distribution function $F_{Z_j,n}$ based on the observations of a variable Z_j (say) with the fit $F_{Z_j}(\cdot; \widehat{\boldsymbol{\eta}}_{j,n}^{\mathrm{IFM}})$. For a margin F_{Z_j}, the test statistic is $K_n = \sup_{z\in\mathbb{R}}\left|F_{Z_j,n}(z) - F_{Z_j}(z; \widehat{\boldsymbol{\eta}}_{j,n}^{\mathrm{IFM}})\right|$. As the test statistic K_n involves estimated parameters, it is a known that the critical values of the KS-test are no longer applicable. In order to correct for this behaviour, the P-value of the test is obtained from a bootstrap adaptation proposed in [3]. Under mild conditions, fulfilled by Assumptions (M1)−(M6) and uniform continuity of the reference density $f_j(\cdot)$, this bootstrap test is consistent. The smooth bootstrap procedure is as follows, first the test statistic from the sample against the fitted distribution is calculated. Next, B independent samples of the same size as the data are drawn from the fitted distribution. For each of the B samples, the chosen distribution is then refitted and the test statistic from the refitted distribution against the original fitted distribution is computed. The P-value is then obtained by taking the number of bootstrap samples which yield a larger test statistic than the data. In our analysis 1 000 bootstrap samples are used.

5.1. *Pokémon data*

This data set contains information on the characteristics (statistics) of 800 existing Pokémon characters. The dataset is freely available at `https://www.kaggle.com/mlomuscio/pokemon`. Variables in this data set are the inherent statistics each Pokémon has. These are: Hitpoints (HP), Attack (Atk), Defence (Def), Special attack (Sp. Atk), Special Defence (Sp. Def) and Speed (Spd). The best fitting margin distributions with their corresponding parameters and AIC value are given in Table 3. In Figure S.6 in Section S.7.1 of the Supplementary Material, the four fitted margins for the variable Defence are depicted. All four fits clearly show that the data on Defence is right skewed. This is also reflected in the estimate $\widehat{\alpha}_{\text{Def}}^{\text{MLE}} = 0.2458$ in Table 3 for the selected QBA-logistic margin. Thus, using skewed margins is necessary for this data set.

Table 3: Pokémon data: best fitting margin (with parameters) based on AIC for each of the variables and corresponding AIC and p-value of Smooth bootstrapped KS-test.

Variable	Best fitted margin	$\widehat{\alpha}^{\text{MLE}}$	$\widehat{\mu}^{\text{MLE}}$	$\widehat{\phi}^{\text{MLE}}$	$\widehat{\nu}^{\text{MLE}}$	AIC	Smooth Bootstrap KS-test
HP	QBA-Student's t	0.3001	54.4601	8.4933	7.3854	7312.223	0.010
Atk	QBA-normal	0.3172	60.6742	13.6339		7795.042	0.517
Def	QBA-logistic	0.2458	51.0036	5.9594		7643.270	0.097
Sp. Atk	QBA-normal	0.2177	45.3211	10.4044		7747.145	0.293
Sp. Def	QBA-Student's t	0.2670	52.2203	9.9730	28.4425	7515.755	0.036
Spd	QBA-normal	0.3255	51.9137	12.5273		7637.763	0.005

Some further caution is needed when applying the bootstrap KS-test discussed above. Indeed, despite the data being regarded as continuous, there are accumulations of mass at multiples of 5. This creates quite large jumps in the empirical cumulative distribution function $F_{Z_j,n}$. To accommodate this near discrete behaviour, a smooth goodness-of-fit test is applied. This is obtained by replacing $F_{Z_j,n}$ in the classical tests with a smoothed version. As shown in [34], these type of tests have better power and are overall better when data show a more discrete behaviour. The smooth tests are close to the classical $F_{Z_j,n}$-based test as shown in [41] and asymptotically equivalent in terms of efficiency (see [4]). We thus opted for a smoothed Kolmogorov-Smirnov test (available in the `snpar`-package) with a uniform kernel. Bandwidth is taken as proposed in [41], i.e. $h = 2.0362\,((8\pi)/3)^{1/5}\,(IQR(Z_{j,1},\ldots,Z_{j,n}))/(1.349))^{2/3}\,n^{-1/3}$, with $IQR(Z_{j,1},\ldots,Z_{j,n})$ the interquartile range of the observations on the variable Z_j. As the parameters are estimated, we apply the smoothed KS-test combined with the bootstrapping principle. The results for the smooth bootstrap KS-test can be found in Table 3. These indicate that the fits for Attack, Defence and Special Attack are appropriate at a 5% significance level whereas at a 1% significance level all variables except Speed are modelled appropriately.

Table 4: Pokémon data: log-likelihood, AIC_{2ML} and CIC contributions of copula fits using pseudo-observations obtained through the margin of Table 3.

Copula	Log-likelihood	AIC_{2ML} contribution	CIC contribution
Gumbel	542.9811	−1083.9622	−1082.445
Clayton	625.5084	−1249.0168	−1244.927
Frank	613.4586	−1224.9172	−1223.599
Joe	362.8394	−723.6788	−720.091
Gaussian	800.7414	−1571.4827	−1552.521
Student's t	945.1691	−1858.3381	−1856.663

With the margins fitted, the pseudo-observations can be obtained. We fit all six copulas to them, which resulted into the copula contribution to the log-likelihood, AIC- and CIC-values listed in Table 4. From this, it is immediately clear that a Student's t-copula provides the best fit. The estimated copula parameters are $\widehat{\boldsymbol{\Sigma}}^{\mathrm{IFM}}_{\text{Pokémon}}$ given in (S.3) (in the Supplementary Material) and $\widehat{\nu}_C^{IFM} = 7$. All estimated correlations are positive and rather large, except for the variable Speed for which the estimates are slightly lower. These positive correlations are to be expected as overall, characteristic features increase all together (evolution to more competitive Pokémons). The lower correlation between Speed and other characteristics can be explained by balancing reasons as Pokémon usually having lower Speed because one or more of their other characteristics are rather large, preventing them from becoming overpowered as they can be outspeeded.

For comparison of our Student's t-copula-based model with the two benchmarks (multivariate skew-normal and skew-t distributions), we give in Table 5 the total AIC_{2ML} value of our model and those of the two benchmark models. Clearly the copula-based model provides a better fit. This follows from the AIC_{2ML} value always being larger or equal to the AIC of the model, as IFM parameters are sub-optimal when considered in the full MLE context. Since we already have a lower criterion value, estimating the model with MLE would only improve this thereby further distancing the copula-based model from the skew-elliptical ones.

Table 5: Pokémon data: AIC_{2ML} for the selected copula-based and the skew-elliptical models.

Model	Student's t-copula	Skew-normal	Skew-t
AIC_{2ML}	43792.86	44758.60	44397.44

5.2. White wine quality

The second data set is a subset of the quality of wine data used in [6]. For a first analysis here, three variables are selected for white wines only: Volatile Acidity (amount of tartaric acid in g per dm^3), pH and Sulphates (amount of potassium sulphate in g per dm^3). According to [16] these characteristics are important when it comes to predicting the quality of wines. Figure 3 depicts histograms for the data on each variable. Pairwise scatter plots are in Figure S.7 in the Supplementary Material. As the data consists of wines of different quality, which has an impact on the variables considered here, we only take into account the 880 wines which are rated at quality 7. All three variables are continuous, but they are rounded to 0.01 (0.005 for Volatile Acidity), and hence we also here apply the smoothed bootstrap KS-test. We consider our $6 + 4 \times 3 = 18$ parametric models and the two benchmark models (for comparison purpose). Tables 6 and S.2 report on the results.

From Table 6 we can conclude that Volatile Acidity and Sulphates are well described by our selected margin model, whereas this is to a lesser extent the case for the variable pH. Concerning the margins, the histograms in Figure 3 already show clear right-skewness. See

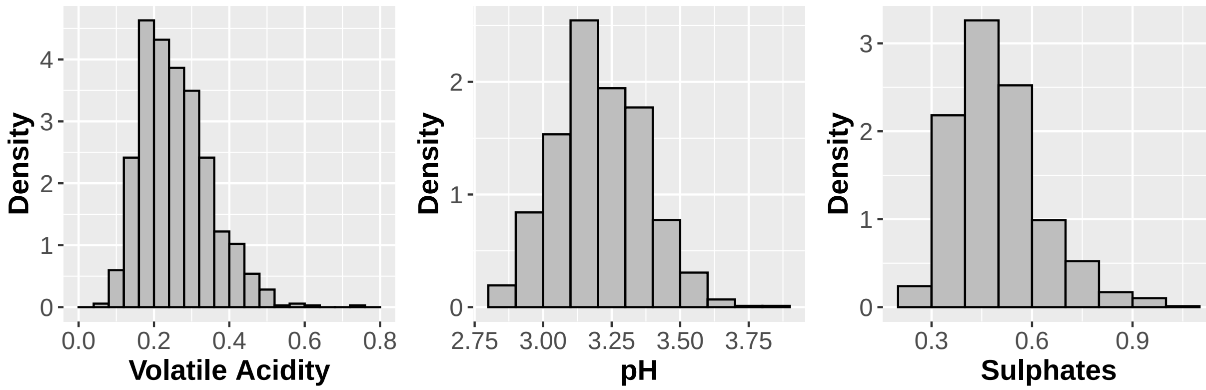


Figure 3: White wine data: histograms of the variables.

Table 6: Wine data: best fitting margin (with parameters) based on AIC, for each of the variables, and corresponding AIC and p-value of the smooth bootstrapped KS-test.

Variable	Best fitted margin	$\widehat{\alpha}^{\mathrm{MLE}}$	$\widehat{\mu}^{\mathrm{MLE}}$	$\widehat{\phi}^{\mathrm{MLE}}$	AIC	Smooth Bootstrap KS-test
Volatile acidity	QBA-normal	0.2333	0.1905	0.0305	−1827.3514	0.190
pH	QBA-normal	0.4012	3.1636	0.0755	−753.7036	0.011
Sulphates	QBA-logistic	0.2441	0.4068	0.0250	−1219.6643	0.383

also the estimates $\widehat{\alpha}_j^{\mathrm{MLE}}$ in Table 6 which are all smaller than 0.5. With $\widehat{\alpha}_{pH}^{\mathrm{MLE}} = 0.4012$, the distribution for the variable pH seems closer to symmetry than for the other two variables.

The copula fits using the pseudo-observations from these margins are given in Table S.2 in the Supplementary Material. Contrary to the first data example, a Gaussian copula is selected here. The estimated correlation matrix, given in (S.4), indicates that the pseudo-observations for pH are positively correlated with the pseudo-observations of both Sulphates and Volatile acidity, whereas the latter two are slightly negatively correlated. The selected copula model clearly outperforms both benchmark skew-elliptical models, as can be concluded from the AIC_{2ML} values for the selected copula-based model (−3828.686), the skew-normal distribution (−3701.502) and the skew-t-distribution (−3709.944).

5.3. Further investigation of wine data

In a further analysis we were interested in finding differences, based on the same three variables, between red and white wines of different quality (wines of quality categories 4 and 7). Section S.7.3 contains the detailed analysis. Figure 4 shows densities of the fitted univariate margins and contour plots of the fitted bivariate margins. This figure clearly reveals the main differences between the four groups of wines, described in Section S.7.3.

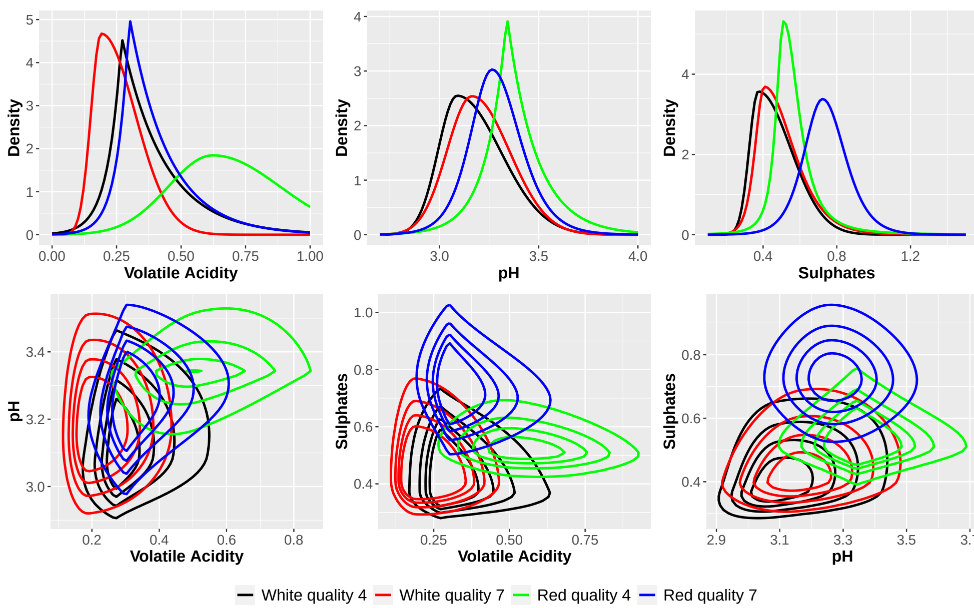


Figure 4: Wine data: univariate and bivariate margins (contour plots) of the fits for white and red wines of quality 4 and 7. Contours represent density levels of 2, 4, 6 and 8.

6. Discussion

In this work, we have shed some light on statistical inference results for parameter estimation for copula-based multivariate distributions with skew margins. More specifically, conditions were provided under which consistency and asymptotic normality hold for a family of margins that are not necessarily continuously differentiable, but still behave "well enough". We obtained these results under the specific setting that margins come from the quantile-based asymmetric family of distributions. However, this is not a requirement as similar conditions for other families of possibly non-smooth distributions can be derived.

In obtaining the results, the more general framework of GMM estimators turns out to be crucial. This allows for the necessary tools in obtaining the statistical inference results.

The main disadvantage of choosing a parametric estimation technique for the margins in IFM is that the model must be correctly specified. By using asymmetric margins, the risk of misspecification is lowered. If clear signs of skewness are present in the marginal data, asymmetric margins do have a significant advantage over classic symmetric margins. Misspecification might still be present. Therefore, a close inspection of whether the chosen margins are suitable should be conducted. If they provide a viable choice, an efficiency argument works in favor of using the parametric margins. Otherwise, one can resort to a semi-parametric setting, for which statistical inference results are also available in the dedicated literature. A mix of both parametric and non-parametric margins might be of interest as it may combine the best of both worlds. This is part of future research.

In the provided examples, we restrict ourselves to classical families of copulas. The attentive reader might note that all of these are in a sense symmetric. The Archimedean copulas are exchangeable and the Gaussian and Student's t-copula elliptically symmetric. If there are clear indications of asymmetric dependence in the data, symmetric copulas no longer suffice. However, the general conditions stated here do not discriminate on the type of copula. As long as a copula meets the conditions, the obtained results apply. These can thus readily be incorporated in other families of copulas as well to allow for even more flexible modelling. It must be stated though, that tractability might be lost in doing so. So one must weigh the advantages against the disadvantages. It might also be of interest to study more specifically how the tail behaviour of the multivariate copula impacts the multivariate distribution. A first step in such a study would be to describe in an adequate way how to measure multivariate tail dependence. There are various approaches to measure multivariate tail dependence. See for example [14] for a recent paper in this respect.

SUPPLEMENTARY MATERIAL

Additional Material. The Supplementary Material contains some preliminaries in Section S.1, the proof of Proposition 1 in Section S.2, and some additional simulation results in Sections S.4–S.6. An additional example for which Assumptions (R1)—(R4) are checked is provided in Section S.3. Finally Section S.7 contains supplements to the real data applications. See PaperCopTwoPSupplement.pdf.

R-codes. R-codes for all numerical examples are provided in PaperCopTwoPRCodes.zip.

Declaration of Competing Interest. The authors declare that there is no conflict of interest for their paper entitled "Estimation in copula models with two-piece skewed margins using the inference for margins method".

Acknowledgements. The authors thank an Associate Editor and two reviewers for their valuable comments, which led to an improvement of the manuscript. The first and second author gratefully acknowledge support from the Research Fund KU Leuven [C16/20/002 project]. The third author was supported by Special Research Fund (Bijzonder Onderzoeksfonds) of Hasselt University [BOF14NI06].

References

[1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

[2] Andrews, D. W. K. (1992). Generic uniform convergence. *Econometric Theory*, 8(2):241–257.

[3] Babu, G. J. and Rao, C. R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhyã: The Indian Journal of Statistics (2003-2007)*, 66(1):63–74.

[4] Butorina, Y. O. and Nikitin, Y. Y. (2011). On large deviations of smoothed kolmogorov-smirnov's statistics. *Vestnik St. Petersburg University: Mathematics*, 44(2):97–102.

[5] Choroś, B., Ibragimov, R., and Permiakova, E. (2010). Copula estimation. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, pages 77–91, Berlin, Heidelberg. Springer.

[6] Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553.

[7] Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[8] Embrechts, P. and Hofert, M. (2013). Statistical inference for copulas in high dimensions: A simulation study. *ASTIN Bulletin*, 43(2):81–95.

[9] Fechner, G. (1897). *Kollektivmasslehre*. Engelmann.

[10] Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

[11] Gibbons, J. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-gaussian distributions. *Applied Physics Letters*, 22(11):568–569.

[12] Gijbels, I., Karim, R., and Verhasselt, A. (2019a). On quantile-based asymmetric family of distributions: Properties and inference. *International Statistical Review*, 87(3):471–504.

[13] Gijbels, I., Karim, R., and Verhasselt, A. (2019b). Quantile estimation in a generalized asymmetric distributional setting. In Steland, A., Rafajłowicz, E., and Okhrin, O., editors, *Stochastic Models, Statistics and Their Applications*, pages 13–40, Cham. Springer International Publishing.

[14] Gijbels, I., Kika, V., and Omelka, M. (2020). Multivariate tail coefficients: properties and estimation. *Entropy*, 22:Article #728.

[15] Greene, W. (2012). *Econometric Analysis*. Pearson, 7 edition. Chapter 14.

[16] Gupta, Y. (2018). Selection of important features and predicting wine quality using machine learning techniques. *Procedia Computer Science*, 125:305 – 312. The 6th International Conference on Smart Computing and Communications.

[17] Hofert, M. (2012). Likelihood inference for archimedean copulas in high dimensions under known margins. *Journal of Multivariate Analysis*, 110:133 – 150.

[18] Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018). *Elements of Copula Modeling with R*. Use R! Springer International Publishing, Cham, Switzerland, 1st ed. edition.

[19] Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J. (2020). copula: Multivariate dependence with copulas. url:
https://CRAN.R-project.org/package=copula.

[20] Hutson, A. D., Wilding, G. E., Mashtare, T. L., and Vexler, A. (2015). Measures of biomarker dependence using a copula-based multivariate epsilon–skew–normal family of distributions. *Journal of Applied Statistics*, 42(12):2734–2753.

[21] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401 – 419.

[22] Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.

[23] Johnson, S. G. (2018). The NLopt nonlinear-optimization package. url: http://ab-initio. mit.edu/nlopt.

[24] Kim, G., Silvapulle, M. J., and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics & Data Analysis*, 51(6):2836 – 2850.

[25] Ko, V. and Hjort, N. L. (2019a). Copula information criterion for model selection with two-stage maximum likelihood estimation. *Econometrics and Statistics*, 12:167 – 180.

[26] Ko, V. and Hjort, N. L. (2019b). Model robust inference with two-stage maximum likelihood estimation for copulas. *Journal of Multivariate Analysis*, 171(C):362–381.

[27] Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, 23:83–92.

[28] Louzada, F. and Ferreira, P. H. (2016). Modified inference function for margins for the bivariate clayton copula-based SUN tobit model. *Journal of Applied Statistics*, 43(16):2956–2976.

[29] Nelsen, R. (2006). *An Introduction to Copulas*, volume 139 of *Springer Series in Statistics*. Springer New York, New York, NY.

[30] Newey, W. K. and McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. In *Handbook of Econometrics*, volume 4, pages 2111 – 2245. Elsevier.

[31] Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057.

[32] Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, 1(3):295–313.

[33] Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report, University of Cambridge, Department of Applied Mathematics and Theoretical Physics.

[34] Racine, J. S. and Van Keilegom, I. (2019). A smooth nonparametric, multivariate, mixed-data location-scale test. *Journal of Business & Economic Statistics*, 38(4):1–12.

[35] Shih, J. and Louis, T. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399.

[36] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

[37] Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279–281.

[38] Smith, M. S., Gan, Q., and Kohn, R. J. (2012). Modelling dependence using skew t copulas: Bayesian inference and applications. *Journal of Applied Econometrics*, 27(3):500–522.

[39] Smith, M. S. and Vahey, S. P. (2016). Asymmetric forecast densities for u.s. macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence. *Journal of Business & Economic Statistics*, 34(3):416–434.

[40] Wallis, K. F. (2014). The two-piece normal, binormal, or double Gaussian distribution: Its origin and rediscoveries. *Statistical Science*, 29(1):106–112.

[41] Wang, J., Cheng, F., and Yang, L. (2013). Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics*, 25(2):395–407.

[42] Wei, Z. and Kim, D. (2018). On multivariate asymmetric dependence using multivariate skew-normal copula-based regression. *International Journal of Approximate Reasoning*, 92:376–391.

[43] Wei, Z., Kim, S., and Kim, D. (2016). Multivariate skew normal copula for nonexchangeable dependence. *Procedia Computer Science*, 91:141–150.

[44] Yoshiba, T. (2018). Maximum likelihood estimation of skew-t copulas with its applications to stock returns. *Journal of Statistical Computation and Simulation*, 88(13):2489–2506.

SUPPLEMENTARY MATERIAL

to the paper

Estimation in copula models with two-piece skewed margins using the inference for margins method

by

Jonas Baillien, Irène Gijbels and Anneleen Verhasselt

S.1. Preliminaries

Lemma S.1 (ULLN, [3], Lemma 2.4 p2129). *Let $Z^{(1)}, \ldots, Z^{(n)}$ be i.i.d. replicates from Z, a univariate random variable with density function $f_Z(z;\boldsymbol{\eta})$, $\boldsymbol{\eta} \in \mathcal{H}$. If*

(U1) $\mathcal{H}$ is compact,

(U2) $\frac{\partial}{\partial \boldsymbol{\eta}} \ln\left(f_Z(z^{(i)};\boldsymbol{\eta})\right)$ is continuous almost everywhere in $\boldsymbol{\eta} \in \mathcal{H}$,

(U3) $\sup\limits_{\boldsymbol{\eta}\in\mathcal{H}} \left\| \frac{\partial}{\partial \boldsymbol{\eta}} \ln\left(f_Z(z;\boldsymbol{\eta})\right) \right\| \leq G(z)$, and $E[G(Z)] < \infty$.

Then

$$\underset{\boldsymbol{\eta}\in\mathcal{H}}{sup} \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\eta}} \ln\left(f_Z(Z;\boldsymbol{\eta})\right) - E\left[\frac{\partial}{\partial \boldsymbol{\eta}} \ln\left(f_Z(Z;\boldsymbol{\eta})\right)\right] \right\| \xrightarrow{P} 0,$$

and $E\left[\frac{\partial}{\partial \boldsymbol{\eta}} \ln\left(f_Z(Z;\boldsymbol{\eta})\right)\right]$ is continuous in $\boldsymbol{\eta}$.

Theorem S.1 ([3], Theorem 2.6 p2132). *With $\underset{\sim}{\mathbf{Z}}$ as previously defined, if*

(C1) $\widehat{\mathbf{W}}_{\boldsymbol{\theta},n} \xrightarrow{P} \mathbf{W}_{\boldsymbol{\theta}}$, as $n \to \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

(C2) $\mathbf{W}_{\boldsymbol{\theta}}$ is positive semi-definite and $\mathbf{W}_{\boldsymbol{\theta}}\mathbf{g}_0(\boldsymbol{\theta}) = \mathbf{0}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^0$,

(C3) $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}$ compact,

(C4) $\mathbf{g}(\mathbf{z};\boldsymbol{\theta})$ is continuous almost everywhere in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$,

(C5) $E\left[\sup\limits_{\boldsymbol{\theta}^0\in\boldsymbol{\Theta}} \|\mathbf{g}(\mathbf{Z};\boldsymbol{\theta})\|\right] < \infty$.

Then $\boldsymbol{\theta} \xrightarrow{P} \boldsymbol{\theta}^0$, for $n \to \infty$.

S.2. Uniform law of large numbers for $\widehat{g}_n$

Proposition 1. *Under Assumption (M1), (R1), (R2) and (R4), the uniform law of large numbers (Lemma S.1) applies to $\widehat{g}_n$ as in* (11), *i.e.*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \widehat{g}_n(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) - g(\underset{\sim}{\mathbf{Z}};\boldsymbol{\theta}) \right\| \xrightarrow{P} 0 \qquad \text{as } n \to \infty.$$

Proof. We want to use Lemma S.1 to show that $\widehat{\mathbf{g}}_n(\boldsymbol{\theta})$ is uniformly convergent to $\mathbf{g}_0(\boldsymbol{\theta})$, and $\mathbf{g}_0(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Hence, for

$$\frac{1}{n}\sum_{i=1}^{n} \overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right),$$

$\forall j = 1,\dots,d$, the conditions of Lemma S.1 need to be fulfilled. Condition $(U1)$ of Lemma S.1 holds by Assumption (M1). For Condition $(U2)$ it suffices to note that

$$\overline{\frac{\partial}{\partial \boldsymbol{\eta}_j}} \ln\left(f_{Z_j}(Z_j;\boldsymbol{\eta}_j)\right),$$

possibly has a single discontinuity for the derivative with respect to μ in the point $\mu = z^{(i)}$. Rest to check Condition $(U3)$. This can be checked on a function-to-function basis. As an example, consider Z to follow a QBA-normal distribution (other distributions can be checked in a similar fashion). We have

$$\frac{\partial}{\partial\alpha}\ln\left(f_Z(z;\boldsymbol{\eta})\right) = \frac{1-2\alpha}{\alpha(1-\alpha)} + \begin{cases} \frac{1-\alpha}{\phi^2}(z^2-2z\mu+\mu^2) & \text{if } z \le \mu \\ \frac{-\alpha}{\phi^2}(z^2-2z\mu+\mu^2) & \text{if } z > \mu \end{cases}$$

$$\overline{\frac{\partial}{\partial\mu}}\ln\left(f_Z(z;\boldsymbol{\eta})\right) = \begin{cases} \left(\frac{1-\alpha}{\phi}\right)^2 (z-\mu) & \text{if } z < \mu \\ \frac{1}{2}\left(\left(\frac{1-\alpha}{\phi}\right)^2 + \left(\frac{\alpha}{\phi}\right)^2\right)(z-\mu) & \text{if } z = \mu \\ \left(\frac{\alpha}{\phi}\right)^2 (z-\mu) & \text{if } z > \mu \end{cases}$$

$$\frac{\partial}{\partial\phi}\ln\left(f_Z(z;\boldsymbol{\eta})\right) = -\frac{1}{\phi} + \begin{cases} \frac{(1-\alpha)^2}{\phi^3}(z^2-2z\mu+\mu^2) & \text{if } z \le \mu \\ \frac{\alpha^2}{\phi^3}(z^2-2z\mu+\mu^2) & \text{if } z > \mu. \end{cases}$$

With the imposed bounds on the parameters α, μ and ϕ, there exist finite δ_i, $i = 1,\dots,8$ depending solely on ε such that

$$\left\| \overline{\frac{\partial}{\partial \boldsymbol{\eta}}} \ln\left(f_Z(z;\boldsymbol{\eta})\right) \right\|^2 \le \left\| \begin{pmatrix} \delta_1 + \delta_2 z + \delta_3 z^2 \\ \delta_4 + \delta_5 z \\ \delta_6 + \delta_7 z + \delta_8 z^2 \end{pmatrix} \right\|^2 = G(z)^2$$

By the existence and finiteness of the moments of Z under the imposed restrictions of the domain of the parameters, $E[|G(Z)|^2] < \infty$ and thus also $E[|G(Z)|] < \infty$. From this, uniform convergence of $\widehat{g}_{n,\boldsymbol{\eta_j}}$ $j = 1, \ldots, d$ holds.

Lemma S.1 can also be used to show uniform convergence of $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}_C} \ln\left(c\left(\mathbf{U}^{(i)}; \boldsymbol{\theta}_C\right)\right)$. Condition $(U1)$ of Lemma S.1 is fulfilled by (R1). Condition $(U2)$ is true $\forall \boldsymbol{\theta}_C \in \boldsymbol{\Theta}_C$ by (R2) and $\mathbf{U} \in]0,1[^d$. As the borders of the domain have measure 0, possible discontinuities there pose no problem. This is necessary due to problems occurring at the boundaries as the copula density itself may fail to be continuous, as noted in [4] and [5]. To round up, Condition $(U3)$ is satisfied by (R4).

□

S.3. Frank copula and Assumptions (R1)—(R4)

In Section 3.3 it was shown that a bivariate Gaussian copula satisfies Assumptions (R1)—(R4). We here provide a second example for which these assumptions are satsfied: the bivariate Frank-copula. This copula is a member of the Archimedean family, and is governed by a single parameter θ. For the Frank family, the copula parameter can take on any real value, except for zero. Negative values of θ indicate negative dependence and are only allowed in the bivariate case. Any higher dimensional generalization has $\theta > 0$. The example presented here is bivariate, but we limit it to positive dependence. The copula density is given by

$$c(\theta; u_1, u_2) = (1+\theta)(u_1 u_2)^{-1-\theta}(u_1^{-\theta} + u_2^{-\theta} - 1)^{-2-\frac{1}{\theta}}. \tag{S.1}$$

The first two derivatives of the log-copula density are easily obtained as

$$\begin{aligned}\frac{\partial}{\partial\theta} \ln\left(c(\theta; u_1, u_2)\right) = &\frac{1}{1+\theta} - \ln\left(u_1 u_2\right) + \frac{1}{\theta^2} \ln\left(u_1^{-\theta} + u_2^{-\theta} - 1\right) \\ &+ (2 + \frac{1}{\theta})\frac{u_1^{-\theta}\ln\left(u_1\right) + u_2^{-\theta}\ln\left(u_2\right)}{u_1^{-\theta} + u_2^{-\theta} - 1}\end{aligned}$$

$$\frac{\partial^2}{\partial\theta^2}\ln\left(c(\theta;u_1,u_2)\right) = -\frac{1}{(1+\theta)^2} - \frac{2}{\theta^3}\ln\left(u_1^{-\theta}+u_2^{-\theta}-1\right) + \frac{2}{\theta^2}\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}$$
$$-(2+\frac{1}{\theta})\frac{u_1^{-\theta}\ln^2(u_1)+u_2^{-\theta}\ln^2(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}$$
$$+(2+\frac{1}{\theta})\left(\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}\right)^2.$$

Hence, the first two derivatives exist. They are also finite for $u_j \neq 0$, $j = 1,2$. As the denominators have no positive root (for θ), only near zero and at infinity problems can occur. For $\theta \to \infty$, the following two results hold

$$\begin{aligned}
\lim_{\theta\to\infty}\frac{1}{\theta^2}\ln\left(u_1^{-\theta}+u_2^{-\theta}-1\right) &= 0 \\
\lim_{\theta\to\infty}\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1} &= \ln\left(\min\{u_1,u_2\}\right),
\end{aligned}$$

which results in $\frac{\partial}{\partial\theta}\ln\left(c(\theta;u_1,u_2)\right)$ being finite. Similar results hold for the second order derivative. When $\theta \to 0$, Taylor series expansions can be used.

$$\begin{aligned}
\ln\left(u_1^{-\theta}+u_2^{-\theta}-1\right) &= -\theta\ln(u_1u_2) - \theta^2\ln(u_1)\ln(u_2) \\
&\quad -\frac{\theta^3}{2}\ln(u_1)\ln(u_2)\ln(u_1u_2) + \mathcal{O}(\theta^4), \\
\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1} &= \ln(u_1u_2) + 2\theta\ln(u_1)\ln(u_2) \\
&\quad +\frac{3\theta^2}{2}\ln(u_1)\ln(u_2)\ln(u_1u_2) + \mathcal{O}(\theta^3), \\
\frac{u_1^{-\theta}\ln^2(u_1)+u_2^{-\theta}\ln^2(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1} &= \ln^2(u_1)+\ln^2(u_2)+\theta\ln(u_1)\ln(u_2)\ln(u_1u_2)+\mathcal{O}(\theta^2).
\end{aligned}$$

Plugging the above expansions, cut off at the appropriate term in order to get rid of all θ's in the denominators, into the expressions for the first two derivatives of the log-copula density yields

$$\lim_{\theta\to 0}\frac{\partial}{\partial\theta}\ln\left(c(\theta;u_1,u_2)\right) = 1 + \ln(u_1u_2)$$
$$\lim_{\theta\to 0}\frac{\partial^2}{\partial\theta^2}\ln\left(c(\theta;u_1,u_2)\right) = 4\ln(u_1)\ln(u_2) + 2\ln(u_1)\ln(u_2)\ln(u_1u_2) - 1,$$

which are both finite if $u_j \neq 0$. Therefore, Assumption (R2) is satisfied. For (R3), we need to find an expression or a bound for

$$\begin{aligned}
E\left[\left(\frac{\partial}{\partial\theta}\ln\big(c(\mathbf{U};\theta)\big)\right)^2\right] &= -E\left[\frac{\partial^2}{\partial\theta}\ln\big(c(\mathbf{U};\theta)\big)\right] \\
&= \int_0^1\int_0^1\left[\frac{1}{(1+\theta)^2}+\frac{2}{\theta^3}\ln\left(u_1^{-\theta}+u_2^{-\theta}-1\right)\right. \\
&\quad -\frac{2}{\theta^2}\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1} \\
&\quad +(2+\frac{1}{\theta})\frac{u_1^{-\theta}\ln^2(u_1)+u_2^{-\theta}\ln^2(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1} \\
&\quad \left.-(2+\frac{1}{\theta})\left(\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}\right)^2\right] \\
&\cdot c(\theta;u_1,u_2)du_1du_2.
\end{aligned}$$

First note that the latter is strictly positive for all $0<\theta<\infty$. We proceed by splitting up the domain in four regions. The reason for this is that only when u_j approaches zero, the second derivative goes to infinity. let $\varepsilon>0$ such that $\varepsilon\to 0$. These four regions of $[0,1]^2$ are then given by

$$\mathcal{R}_1=[0,\varepsilon)\times[\varepsilon,1],\quad \mathcal{R}_2=[\varepsilon,1]\times[0,\varepsilon),\quad \mathcal{R}_3=[0,\varepsilon)\times[0,\varepsilon)\quad\text{and}\quad \mathcal{R}_4=[\varepsilon,1]\times[\varepsilon,1].$$

For $\mathcal{R}_1$, note that $u_1^{-\theta} >> u_2^{-\theta}-1$. The individual terms of the second derivative can then be approximated by

$$\frac{2}{\theta^3}\ln\left(u_1^{-\theta}+u_2^{-\theta}-1\right)\approx-\frac{2\ln(u_1)}{\theta^2}$$

$$-\frac{2}{\theta^2}\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}\approx-\frac{2\ln(u_1)}{\theta^2}$$

$$\frac{u_1^{-\theta}\ln^2(u_1)+u_2^{-\theta}\ln^2(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}-\left(\frac{u_1^{-\theta}\ln(u_1)+u_2^{-\theta}\ln(u_2)}{u_1^{-\theta}+u_2^{-\theta}-1}\right)^2\approx K_1\ln(u_1)^2,$$

with $K_1>0$ since we can approximate the dominant terms in the latter expression as

$$\frac{u_1^{-\theta}\ln(u_1)^2}{u_1^{-\theta}+u_2^{-\theta}-1}-\frac{u_1^{-2\theta}\ln(u_1)^2}{(u_1^{-\theta}+u_2^{-\theta}-1)^2}.$$

This term is positive since

$$1 \geq \frac{u_1^{-\theta}}{u_1^{-\theta} + u_2^{-\theta} - 1} \geq \left(\frac{u_1^{-\theta}}{u_1^{-\theta} + u_2^{-\theta} - 1} \right)^2 = \frac{u_1^{-2\theta}}{(u_1^{-\theta} + u_2^{-\theta} - 1)^2}.$$

Hence, on $\mathcal{R}_1$

$$0 < -\frac{\partial^2}{\partial \theta} \ln\big(c(\mathbf{u};\theta)\big) \leq \frac{1}{(1+\theta)^2} - D_1 \frac{2\ln(u_1)}{\theta^2} + K_1(2 + \frac{1}{\theta}) \ln(u_1)^2,$$

with $0 < D_1, K_1 < \infty$. In an identical fashion, a bound on $\mathcal{R}_2$ of the same form can be obtained. For $\mathcal{R}_3$, we assume we can write $u_1 = bu_2$, for some positive b and that u_1 and u_2 are small enough such that $u_1^{-\theta} + u_2^{-\theta} \leq u_1^{-\theta} u_2^{-\theta} + 1$. Under these two assumptions, the following bounds are obtained

- $$\frac{2}{\theta^3} \ln\big(u_1^{-\theta} + u_2^{-\theta} - 1\big) \leq \frac{2}{\theta^3} \ln\big(u_1^{-\theta} u_2^{-\theta}\big) = -\frac{2}{\theta^2} \left(\ln(u_1) + \ln(u_2)\right).$$
- $$-\frac{2}{\theta^2} \frac{u_1^{-\theta} \ln(u_1) + u_2^{-\theta} \ln(u_2)}{u_1^{-\theta} + u_2^{-\theta} - 1} \approx -\frac{2}{\theta^2} \left(\frac{\ln(u_1)}{1 + b^{\theta}} + \frac{\ln(u_2)}{1 + b^{-\theta}} \right).$$
- $$\begin{aligned} &\frac{u_1^{-\theta} \ln^2(u_1) + u_2^{-\theta} \ln^2(u_2)}{u_1^{-\theta} + u_2^{-\theta} - 1} - \left(\frac{u_1^{-\theta} \ln(u_1) + u_2^{-\theta} \ln(u_2)}{u_1^{-\theta} + u_2^{-\theta} - 1} \right)^2 \\ &\approx \frac{u_1^{-\theta} \ln(u_1)^2}{u_1^{-\theta}(1 + b^{\theta})} + \frac{u_2^{-\theta} \ln(u_2)^2}{u_2^{-\theta}(1 + b^{-\theta})} - \frac{u_1^{-2\theta} \ln(u_1)^2}{u_1^{-2\theta}(1 + b^{\theta})^2} \\ &\quad - \frac{u_1^{-2\theta} \ln(u_1)^2}{u_1^{-2\theta}(1 + b^{\theta})^2} - \frac{2u_1^{-\theta} \ln(u_1)\, u_2^{-\theta} \ln(u_2)}{u_1^{-2\theta} + 2u_1^{-\theta} u_2^{-\theta} + u_2^{-2\theta}} \\ &\leq \ln(u_1)^2 \left(\frac{1}{1 + b^{\theta}} - \frac{1}{(1 + b^{\theta})^2} \right) + \ln(u_2)^2 \left(\frac{1}{1 + b^{-\theta}} - \frac{1}{(1 + b^{-\theta})^2} \right) \\ &= E_1 \ln(u_1)^2 + E_2 \ln(u_2)^2, \end{aligned}$$

  with $0 < E_1, E_2 < \infty$ since $1 \geq \frac{1}{1+b^{\theta}} \geq \frac{1}{(1+b^{\theta})^2}$ and $1 \geq \frac{1}{1+b^{-\theta}} \geq \frac{1}{(1+b^{-\theta})^2}$.

So on $\mathcal{R}_3$, $\frac{\partial^2}{\partial \theta} \ln\big(c(\mathbf{u};\theta)\big)$ can be bounded above by $\frac{1}{(1+\theta)^2} - D_2 \frac{\ln(u_1)+\ln(u_2)}{\theta^2} + K_2(2+\frac{1}{\theta})(\ln(u_1)^2 + \ln(u_1)^2)$, with $0 < D_2, K_2 < \infty$. Rests there $\mathcal{R}_4$, on which $\frac{\partial^2}{\partial \theta} \ln\big(c(\mathbf{u};\theta)\big)$ is a bounded function, say by a constant M. Now, as all these bounding functions on the four regions of the domain are positive, a bound on the entire domain can be obtained by simply adding them together. Using the obtained bounds, $E[\ln(U_j)] = -1$ and $E[\ln(U_j)^2] = 2$, we arrive at

$$0 < -E\left[\frac{\partial^2}{\partial \theta} \ln\big(c(\mathbf{U};\theta)\big) \right] < \frac{1}{(1+\theta)^2} + \frac{M_1}{\theta^2} + M_2(2 + \frac{1}{\theta}) + M < \infty.$$

Thus, (R3) holds. For (R4) we approach the problem in a similar way as for (R3). By boundedness of the first order derivative, say by a constant $K > 0$, on $\mathcal{R}_4$ the majority of the domain is covered. For the remainder of the domain, a bound can be achieved in the following way. When $u_j \to 0$, $\left|\frac{\partial}{\partial\theta} \ln\left(c(\theta; u_1, u_2)\right)\right| \to \infty$ at a rate proportional to $-\ln(u_j)$ for $u_j \to 0$. The reason for this is, say for $u_1 \to 0$, $u_1^{-\theta} >> u_2^{-\theta} - 1$ and

$$\begin{aligned} \frac{1}{\theta^2} \ln\left(u_1^{-\theta} + u_2^{-\theta} - 1\right) &\approx -\frac{\ln(u_1)}{\theta} \\ (2 + \frac{1}{\theta})\frac{u_1^{-\theta}\ln(u_1) + u_2^{-\theta}\ln(u_2)}{u_1^{-\theta} + u_2^{-\theta} - 1} &\approx 2\ln(u_1) + \frac{\ln(u_1)}{\theta}. \end{aligned}$$

In this light, we suggest a function $H(\mathbf{U})$ of the form

$$H(\mathbf{u}) = K + D_1 \ln(u_1) + D_2 \ln(u_2),$$

with $-\infty < D_j < 0$, $j = 1, 2$. Since $E[\ln(U_1)] = -1$, it can easily be seen that $E[H(\mathbf{U})] < \infty$, thereby fulfilling (R4). The exact choice of K and D_j can be made on some worst case scenario over the subset of interest of the parameter space of θ. On the other hand, for the final assumption, $(N1)$ a similar argument as for the Gaussian copula can be used that, under uniform margins, root-n consistency holds by classical likelihood theory (see e.g. [1]).

S.4. Influence of the choice of W on estimates.

In order to investigate the effect that the weighting matrix $\mathbf{W}$ of the GMM framework has, a small simulation is conducted. From Theorem 3, asymptotically, both the "optimal" and identity matrix result in the same (asymptotic) variance-covariance matrix for the parameters estimators. However, in small samples, this might differ. We consider a bivariate Gaussian copula with two QBA-normal distributed margins. The first margin has parameters $\alpha_1 = 0.3$, $\mu_1 = 0.5$ and $\phi_1 = 0.5$ whereas the second has parameters $\alpha_2 = 0.8$, $\mu_2 = 2$ and $\phi_2 = 2$. The correlation ρ of the copula is set to 0.2 like in the other simulations. We consider sample sizes 25, 50, 100, 250 and 500. It is expected that for the smaller sample sizes, the optimal weighting matrix will outperform the identity, but as the sample size increases, they should perform comparable. In Table S.1, the mean, variance, median and mean absolute deviation (MAD) of the estimated parameters for both weighting matrices are presented for 1 000 Monte Carlo runs (i.e. simulated samples) of the model. For the smaller sample sizes, the optimal weighting matrix clearly outperforms the identity matrix with better, less variable estimates. As the sample size increases, the method using the identity matrix, catches up with the method using the optimal weighting matrix, and eventually even outperforms it. The latter is due to the simpler computation and the lack of the need for matrix inversion, which can cause numerical instabilities.

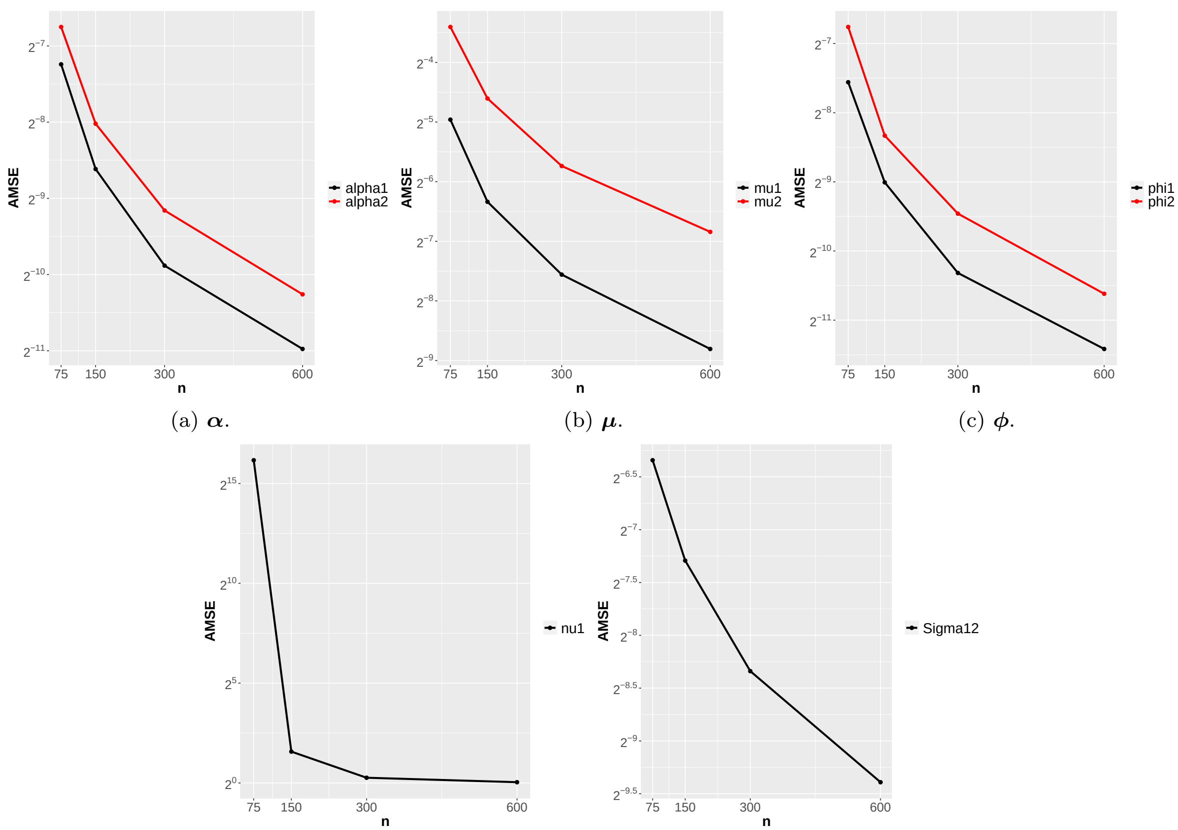
Table S.1: Mean, variance, median and MAD for the above model.

	Parameter:	α_1		μ_1		ϕ_1		α_2		μ_2		ϕ_2		ρ	
	true value:	0.3		0.5		0.5		0.8		2		2		0.2	
	W	Id	Opt	Id	Opt	Id	Opt	Id	Opt	Id	Opt	Id	Opt	Id	Opt
$n = 25$	mean	0.311	0.298	0.499	0.452	0.537	0.538	0.778	0.765	1.478	2.176	2.161	2.569	0.200	0.211
	var	0.014	0.006	1.283	0.293	0.823	0.589	0.004	0.004	2.938	1.904	3.820	1.887	0.038	0.042
	median	0.292	0.289	0.495	0.486	0.456	0.469	0.785	0.774	1.508	2.049	2.055	2.240	0.203	0.221
	MAD	0.120	0.075	0.463	0.269	0.121	0.100	0.065	0.061	1.313	0.866	0.503	0.610	0.189	0.199
$n = 50$	mean	0.299	0.293	0.510	0.402	0.469	0.532	0.783	0.757	1.581	2.116	2.089	2.707	0.200	0.219
	var	0.009	0.004	0.148	0.318	0.010	0.191	0.003	0.004	1.321	1.600	0.414	1.611	0.019	0.022
	median	0.291	0.291	0.466	0.459	0.476	0.481	0.789	0.764	1.664	2.048	2.043	2.316	0.205	0.230
	MAD	0.098	0.060	0.380	0.234	0.105	0.080	0.055	0.063	1.093	0.839	0.436	0.674	0.133	0.138
$n = 100$	mean	0.300	0.295	0.474	0.392	0.504	0.536	0.787	0.748	1.777	2.186	2.114	2.857	0.206	0.230
	var	0.005	0.003	0.251	0.275	0.099	0.099	0.002	0.004	1.357	1.560	0.575	1.322	0.010	0.014
	median	0.297	0.293	0.481	0.452	0.490	0.490	0.792	0.756	1.800	2.092	2.056	2.433	0.211	0.233
	MAD	0.073	0.054	0.292	0.215	0.077	0.064	0.045	0.071	0.871	0.856	0.363	0.829	0.098	0.113
$n = 250$	mean	0.304	0.298	0.473	0.289	0.522	0.595	0.780	0.729	1.807	2.135	2.250	3.168	0.211	0.244
	var	0.002	0.003	0.217	0.494	0.093	0.178	0.002	0.004	1.984	1.600	0.848	1.543	0.007	0.012
	median	0.302	0.294	0.499	0.432	0.502	0.496	0.785	0.733	1.772	2.059	2.094	2.781	0.205	0.238
	MAD	0.048	0.050	0.186	0.194	0.052	0.059	0.037	0.075	0.713	0.795	0.298	1.098	0.063	0.086
$n = 500$	mean	0.306	0.302	0.394	0.224	0.561	0.657	0.769	0.719	2.000	2.111	2.530	3.370	0.225	0.264
	var	0.002	0.003	0.863	0.638	0.188	0.291	0.003	0.004	4.239	1.766	2.313	1.653	0.007	0.013
	median	0.303	0.297	0.498	0.425	0.503	0.504	0.780	0.715	1.736	2.067	2.156	3.148	0.208	0.252
	MAD	0.038	0.051	0.148	0.199	0.044	0.068	0.039	0.073	0.655	0.752	0.307	1.308	0.052	0.079

S.5. Simulations in Setting 1: further details and results

For the Student's t-copula simulation model the correlation matrix equals

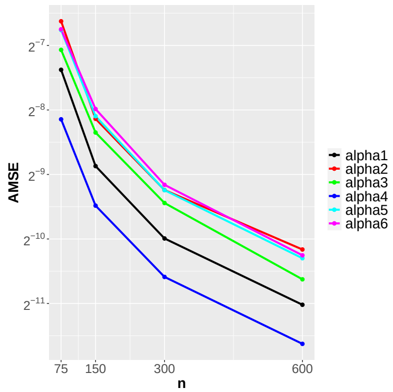
$$\mathbf{\Sigma} = \begin{bmatrix} 1.00 & 0.2 & -0.05 & 0.08 & -0.2 & 0.17 \\ 0.2 & 1.00 & 0.18 & -0.22 & 0.03 & -0.13 \\ -0.05 & 0.18 & 1.00 & -0.08 & 0.3 & -0.34 \\ 0.08 & -0.22 & -0.08 & 1.00 & 0.01 & 0.17 \\ -0.2 & 0.03 & 0.3 & 0.01 & 1.00 & -0.21 \\ 0.17 & -0.13 & -0.34 & 0.17 & -0.21 & 1.00 \end{bmatrix}. \tag{S.2}$$



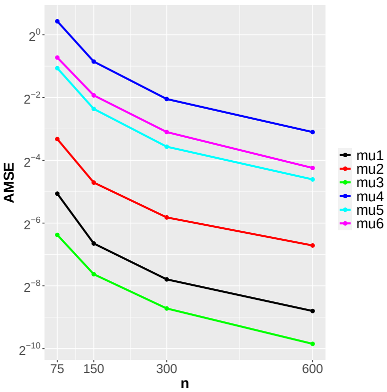
(a) $\boldsymbol{\alpha}$. (b) $\boldsymbol{\mu}$. (c) $\boldsymbol{\phi}$. (d) ν_1. (e) $\Sigma_{1,2}$.

Figure S.1: Setting 1. Approximate Mean Squared Error of the fitted parameters for the bivariate Gaussian copula with parameters as in Table 1 and $\rho = 0.2$. Results for sample sizes 75, 150, 300 and 600 on a vertical $\log_2$-scale.

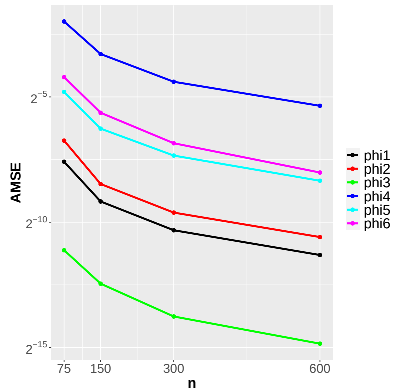
Figure S.1 summarizes simulation results for the second simulation model involving a bivariate Gaussian copula and margins as indicated in Table 1. Note that the finite-sample performance of the IFM estimators is very good.



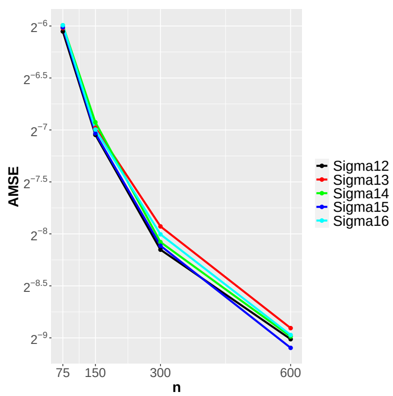
(a) $\boldsymbol{\alpha}$.



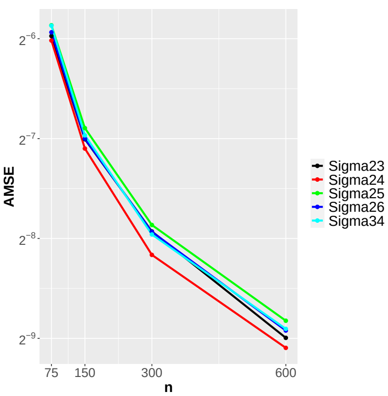
(b) $\boldsymbol{\mu}$.



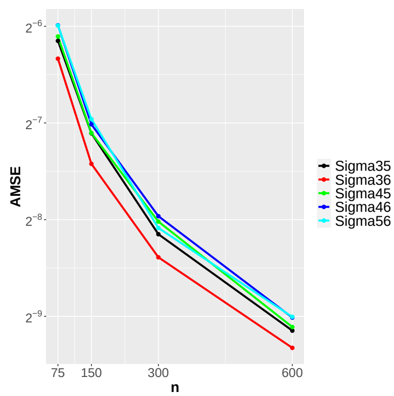
(c) $\boldsymbol{\phi}$.



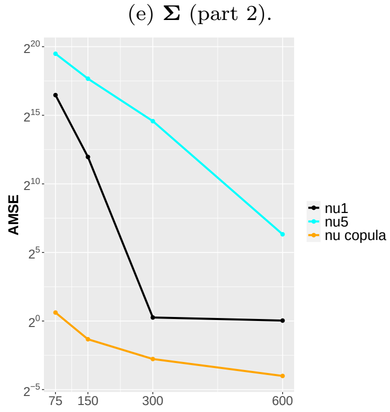
(d) $\boldsymbol{\Sigma}$ (part 1).



(e) $\boldsymbol{\Sigma}$ (part 2).



(f) $\boldsymbol{\Sigma}$ (part 3).



(g) $\boldsymbol{\nu}$.

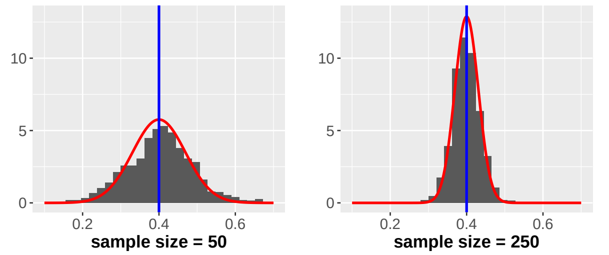
Figure S.2: Setting 1. Approximate Mean Squared Error of the fitted parameters for the six dimensional Student's t-copula with parameters as in Table 1 and (S.2). Results for sample sizes 75, 150, 300 and 600 on a vertical $\log_2$-scale.

Figure S.2 finally presents the simulation results for the six-dimensional Student's t-copula copula, with margins as listed in Table 1. We can draw similar conclusions as for the two other simulation models.

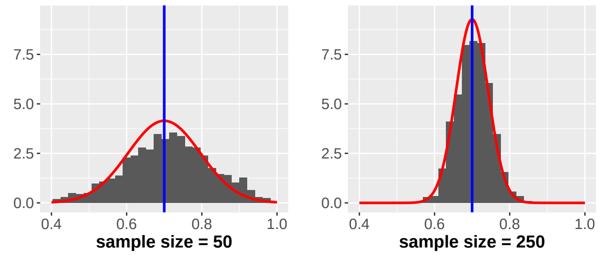
S.6. Simulations in Setting 2: further details and results

S.6.1. Asymptotic normality distributions for the estimated parameters

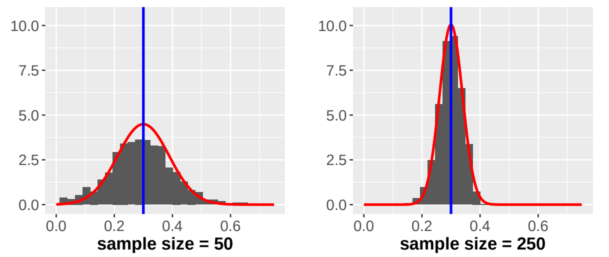
In Figure S.3 we present histograms based on the 1 000 estimated values $\widehat{\theta}_j$ resulting from the simulated samples for Setting 2, for sample sizes $n = 50$ and $n = 250$. For each parameter we also present the approximate (asymptotic) normal distribution, with as mean the true parameter value and as standard deviation the standard deviation established in Theorem 3. As is clearly noted the presented distributions become more and more concentrated around the true parameter value.



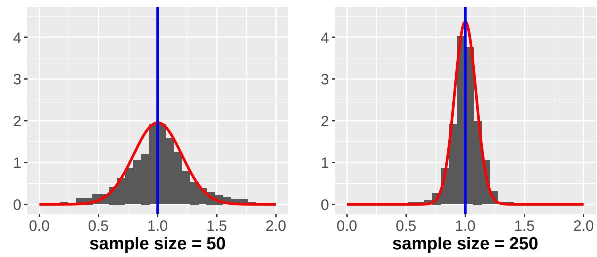
(a) α_1.



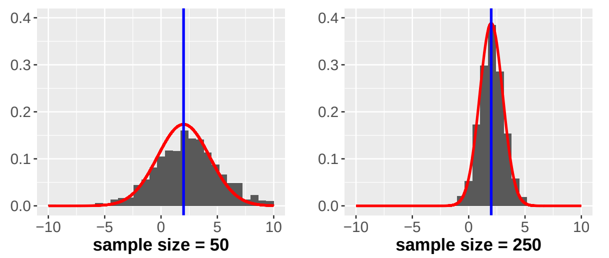
(b) μ_1.



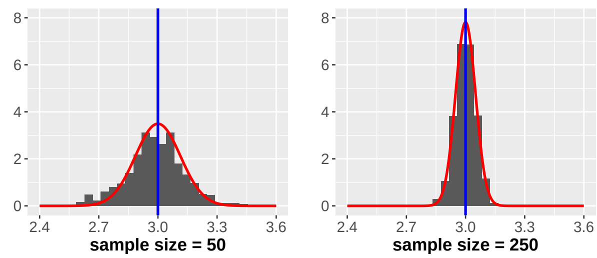
(c) ϕ_1.



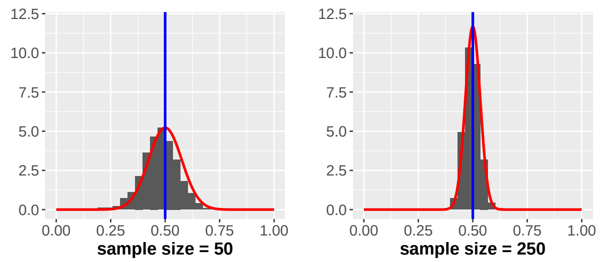
(d) α_2.



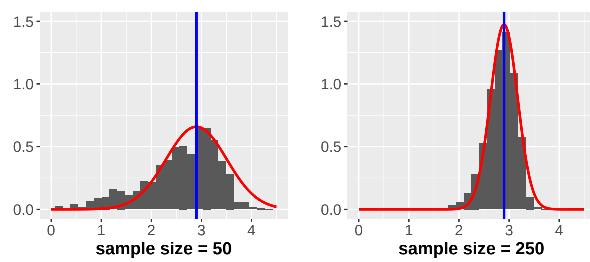
(e) μ_2.



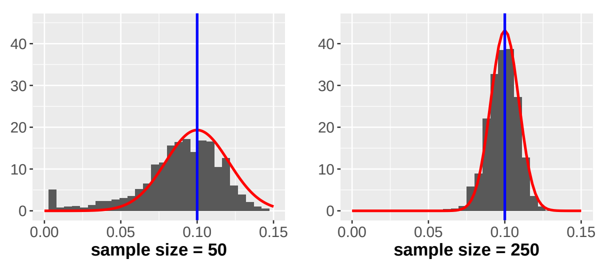
(f) ϕ_2.



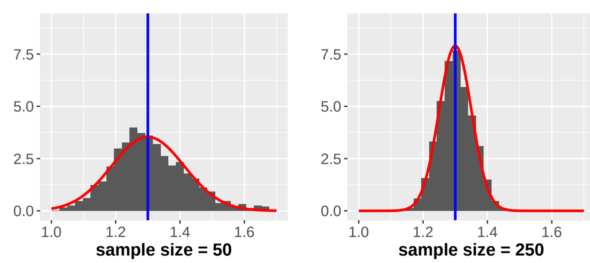
(g) α_3.



(h) μ_3.



(i) ϕ_3.



(j) θ_C.

Figure S.3: Setting 2. Histograms of parameter estimates for sample sizes $n = 50$ and $n = 250$. The red line indicates the theoretical asymptotic distribution for that sample size. The vertical blue line present the true parameter value.

S.6.2. Asymptotic variance-covariance matrix and estimated version

For the simulation model under Setting 2, the asymptotic variance-covariance matrix (up to a factor $1/n$), for which the expression is provided in Theorem 3, relying on expression (13) is equal to

$$\left[\begin{array}{ccccccccc|c}
0.2400 & 0.5000 & 0.1000 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0096 \\
0.5000 & 2.0833 & 0.2083 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0994 \\
0.1000 & 0.2083 & 0.2917 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1513 \\
0 & 0 & 0 & 0.4630 & 10.2040 & -2.5578 & 0 & 0 & 0 & -0.0435 \\
0 & 0 & 0 & 10.2040 & 264.9114 & -56.3650 & 0 & 0 & 0 & 0.6051 \\
0 & 0 & 0 & -2.5578 & -56.3650 & 18.3336 & 0 & 0 & 0 & 0.6645 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3949 & 0.4487 & 0.0752 & -0.1329 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.4487 & 0.6527 & 0.0855 & -0.1539 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.0752 & 0.0855 & 0.0213 & 0.0039 \\
\hline
0.0096 & 0.0994 & 0.1513 & -0.0435 & 0.6051 & 0.6645 & -0.1329 & -0.1539 & 0.0039 & 0.6382
\end{array}\right].$$

For a given sample, of size n, an estimate for this asymptotic variance-covariance matrix is obtained by replacing the true parameter values (in the simulation model) by their estimates, obtained by the two steps IFM method. For each sample we thus get an estimated variance-covariance maxtrix. The next two matrices present the average values, over all Monte Carlo simulations, that we obtained from the samples of sizes $n = 50$ and $n = 1\,000$.

The average of the estimated variance-covariance matrix, based on 1 000 Monte Carlo samples of size $n = 50$ is

$$\left[\begin{array}{ccccccccc|c}
0.2396 & 0.4769 & 0.0935 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0077 \\
0.4769 & 1.9195 & 0.1890 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0790 \\
0.0935 & 0.1890 & 0.2820 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1358 \\
0 & 0 & 0 & 0.4470 & 9.4605 & -2.2751 & 0 & 0 & 0 & -0.0203 \\
0 & 0 & 0 & 9.4605 & 234.7113 & -47.3600 & 0 & 0 & 0 & 0.8625 \\
0 & 0 & 0 & -2.2751 & -47.3600 & 16.7762 & 0 & 0 & 0 & 0.3927 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3851 & 0.4132 & 0.0681 & -0.1077 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.4132 & 0.5636 & 0.0716 & -0.1165 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.0681 & 0.0716 & 0.0192 & 0.0053 \\
\hline
0.0077 & 0.0790 & 0.1358 & -0.0203 & 0.8625 & 0.3927 & -0.1077 & -0.1165 & 0.0053 & 0.6210
\end{array}\right].$$

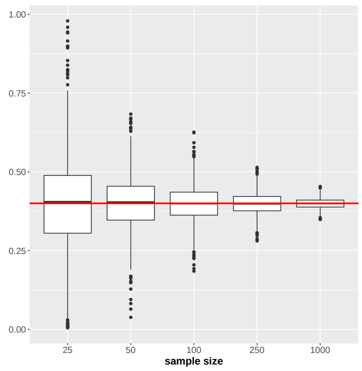
The average of the estimated variance-covariance matrices, based on 1 000 Monte Carlo samples of size $n = 1\,000$ resulted into the matrix

$$\left[\begin{array}{ccccccccc|c}
0.2399 & 0.4992 & 0.1007 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0091 \\
0.4992 & 2.0786 & 0.2098 & 0 & 0 & 0 & 0 & 0 & 0 & 0.0886 \\
0.1007 & 0.2098 & 0.2929 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1489 \\
0 & 0 & 0 & 0.4612 & 10.1515 & -2.5699 & 0 & 0 & 0 & -0.0263 \\
0 & 0 & 0 & 10.1515 & 263.4619 & -56.5172 & 0 & 0 & 0 & 0.9369 \\
0 & 0 & 0 & -2.5699 & -56.5172 & 18.5025 & 0 & 0 & 0 & 0.5365 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3940 & 0.4476 & 0.0752 & -0.1140 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.4476 & 0.6502 & 0.0849 & -0.1266 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.0752 & 0.0849 & 0.0213 & 0.0082 \\
\hline
0.0091 & 0.0886 & 0.1489 & -0.0263 & 0.9369 & 0.5365 & -0.1140 & -0.1266 & 0.0082 & 0.6346
\end{array}\right]$$

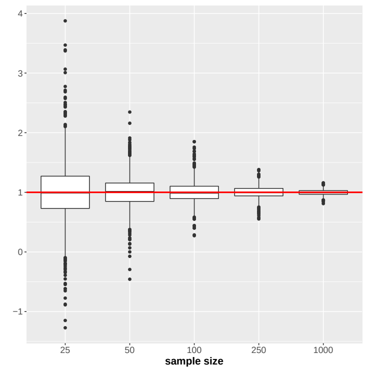
Note that all entries of the average matrix for sample size $n = 1\,000$ are closer to these of the true asymptotic variance-covariance matrix, than when $n = 50$. This is another illustration of the quality of the estimators.

S.6.2.1. Graphical summary of simulation study Setting 2

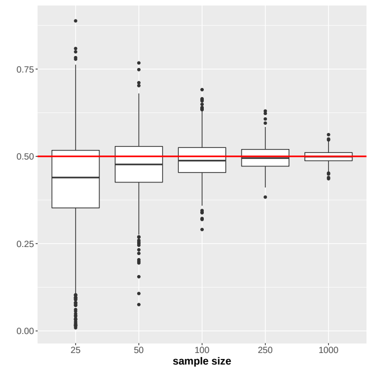
Figures S.4 and S.5 present boxplots of the estimated parameters for the simulation model in Setting 2, for sample sizes n = 25, 50, 100, 250 and 1 000. Figure S.4 gives the results for the parameters of the margins, whereas the results for the copula parameter are in Figure S.5. Note the evolution with increasing sample size: the bias reduces and the variance gets smaller, as expected. The boxplots also reveal more symmetry in the finite-sample distributions of the IFM estimates, for larger sample size.



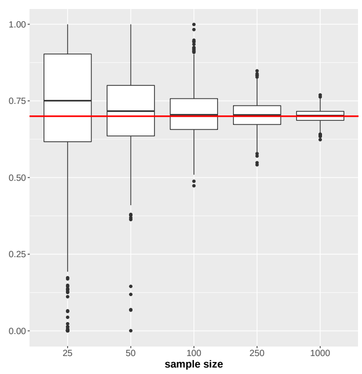
(a) α_1.



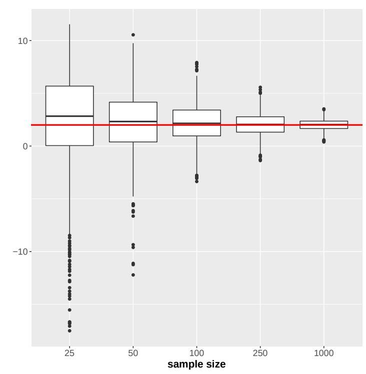
(b) μ_1.



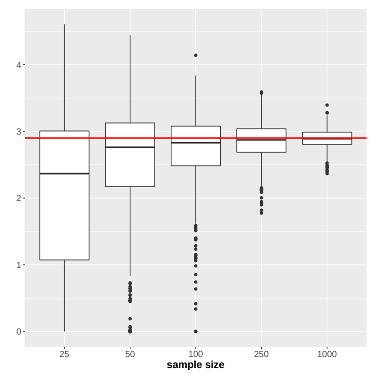
(c) ϕ_1.



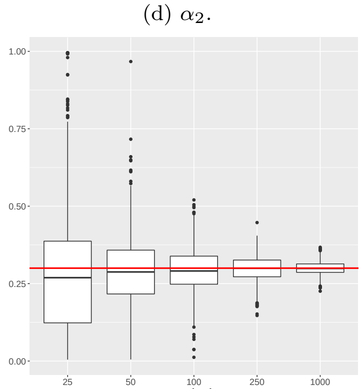
(d) α_2.



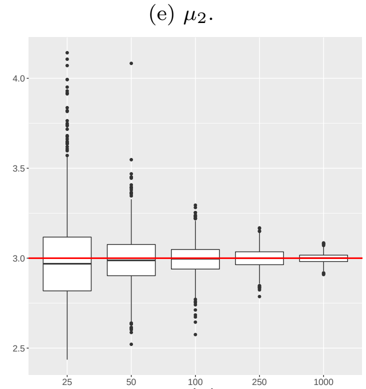
(e) μ_2.



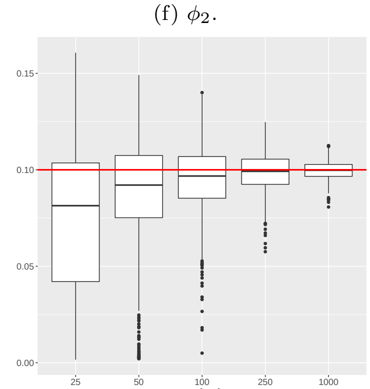
(f) ϕ_2.



(g) α_3.



(h) μ_3.



(i) ϕ_3.

Figure S.4: Setting 2. Boxplots of parameter estimates of the margins for different sample sizes. The red line indicates the true parameter value.

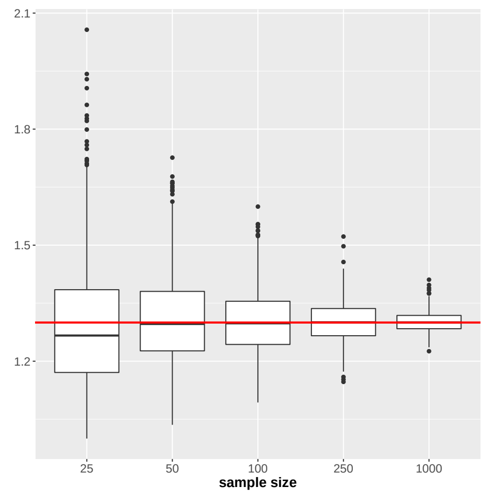


Figure S.5: Setting 2. Boxplots of $\widehat{\theta}_C$ for different sample sizes. The red line indicates the true parameter value.

S.7. Real data examples: supplement

S.7.1. Pokémon data: more details of the analysis

Figure S.6 shows the histogram of the data on the variable Defence, together with the four fitted margins. Due to a rather large tail-weight, the QBA-Student's t and QBA-logistic distributions provide superior fits compared to the QBA-normal and QBA-Laplace distribution.

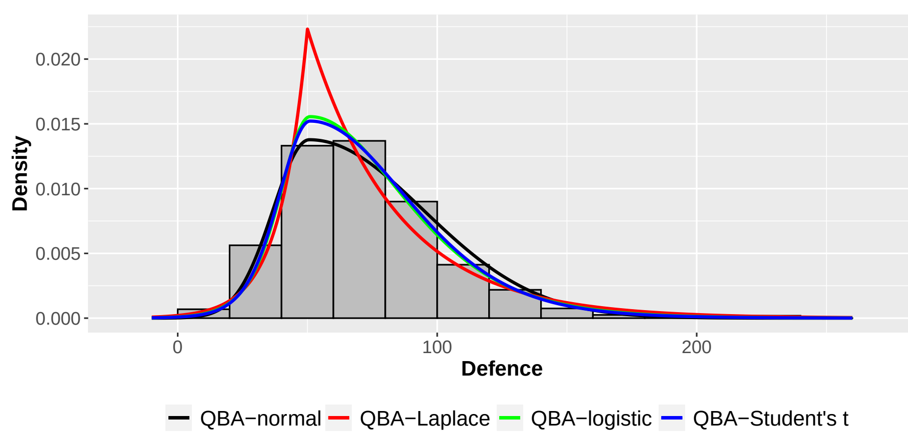


Figure S.6: Pokémon data: four different QBA-distributions fitted to the variable Defence.

The estimated correlation matrix for the (selected) Student's t-copula is

$$\widehat{\mathbf{\Sigma}}^{\text{IFM}}_{\text{Pokémon}} = \begin{bmatrix} 1 & 0.5808 & 0.4847 & 0.5074 & 0.5074 & 0.2653 \\ 0.5808 & 1 & 0.5467 & 0.4155 & 0.3437 & 0.3707 \\ 0.4847 & 0.5467 & 1 & 0.3454 & 0.6050 & 0.1043 \\ 0.4854 & 0.4155 & 0.3454 & 1 & 0.6076 & 0.4658 \\ 0.5074 & 0.3437 & 0.6050 & 0.6076 & 1 & 0.3186 \\ 0.2653 & 0.3707 & 0.1043 & 0.4658 & 0.3186 & 1 \end{bmatrix}. \tag{S.3}$$

S.7.2. White wine data analysis: more details

Figure S.7 depicts pairwise scatter plots for the data on white wines.

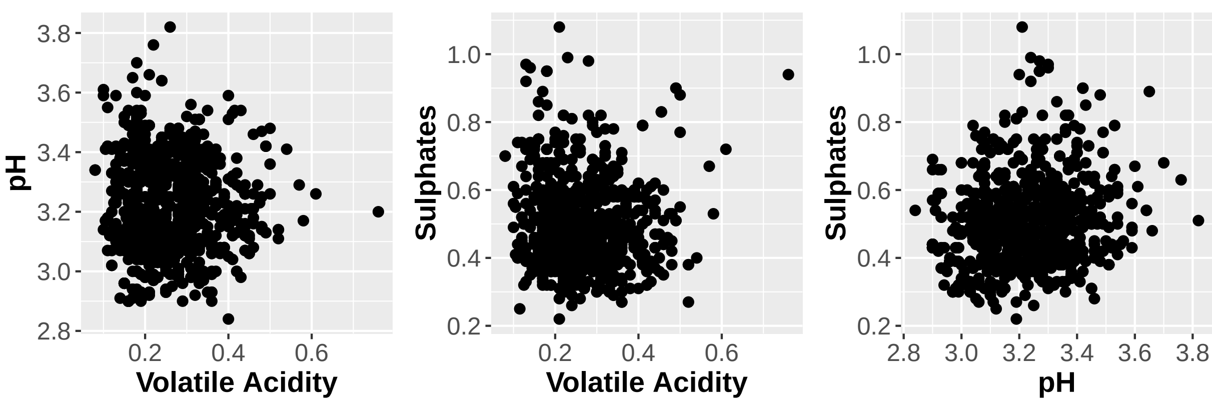


Figure S.7: White wine data. Pairwise scatter plot.

Table S.2 presents the values of the log-likelihood, and of the AIC_{2ML} and CIC contributions of the fitted copula for the white wine data. Among the considered copulas a Gaussian copula is selected, with estimated correlation matrix

$$\widehat{\mathbf{\Sigma}}^{\text{IFM}}_{\text{Wine}} = \begin{bmatrix} 1 & 0.0535 & -0.0427 \\ 0.0535 & 1 & 0.1810 \\ -0.0427 & 0.1810 & 1 \end{bmatrix}. \tag{S.4}$$

Table S.2: Wine data: log-likelihood, AIC_{2ML} and CIC contributions of the fitted copula candidates using the pseudo-observations obtained using the margins of Table 6.

Copula	Log-likelihood	AIC_{2ML} contribution	CIC contribution
Gumbel	5.3958	−8.7916	−8.5984
Clayton	3.8325	−5.6650	−5.5928
Frank	4.9623	−7.9246	−7.8227
Joe	4.0677	−6.1353	−5.9099
Gaussian	16.9834	−27.9667	−27.7091
Student's t	17.5354	−27.0707	−27.2917

S.7.3. Wine data: further analysis

To further investigate the wine data, also lower quality wines and red wines of the same quality are selected in order to see whether there are specific differences in the model to describe the various data appropriately. The same three variables, Volatile Acidity, pH and Sulphates, are considered for both red and white wines of quality 4 and 7. Note that the number of observations differs greatly between the four considered groups. For red wines, we have 53 wines of quality 4 compared to 199 of quality 7 whereas for the white wines, 163 have quality label 4 and 880 have quality 7. This might impact the model as fewer observations may lead to larger uncertainties on both the margins as the copula itself. We employed the same method of analysis as in Sections 5.1 and 5.2. Results are reported in Tables S.3 and S.4.

Table S.3 shows that all four groups have mostly different members of the QBA-family that were selected for the different margins. The most prominent distinction between different qualities of wine can be made for the red wines. As μ indicates the location of the mode, it is immediately clear that for both Volatile Acidity and Sulphates, there is a large discrepancy in the estimated modes for red wines of the two qualities. For red wines of quality 7 the estimated mode of Volatile Acidity is much smaller, combined with a more extreme skewing parameter. For red wines of quality 4, data on Volatile Acidity are more centered around the higher estimated mode value, but due to a larger scale parameter, they are also more spread. Quite the opposite holds true for the amount of Sulphates present in the white wines, as their parameters are quite close. Volatile Acidity of quality 4 white wines is alike to that of quality 7 red wines, but quality 7

Table S.3: Wine data: parameters of best fitting margins for the four considered wine groups based on AIC.

Data	Variable	Best fitting margin	$\widehat{\alpha}^{\text{MLE}}$	$\widehat{\mu}^{\text{MLE}}$	$\widehat{\phi}^{\text{MLE}}$	$\widehat{\nu}^{\text{MLE}}$
Red wine of quality 4	Volatile Acidity	QBA-normal	0.3987	0.6231	0.1037	
	pH	QBA-Laplace	0.4147	3.3400	0.0606	
	Sulphates	QBA-Student's t	0.3300	0.5118	0.0294	2
Red wine of quality 7	Volatile Acidity	QBA-Laplace	0.2360	0.3000	0.0357	
	pH	QBA-logistic	0.4453	3.2653	0.0408	
	Sulphates	QBA-logistic	0.4584	0.7230	0.0367	
White wine of quality 4	Volatile Acidity	QBA-Laplace	0.2416	0.2700	0.0399	
	pH	QBA-normal	0.3070	3.0889	0.0666	
	Sulphates	QBA-normal	0.2186	0.3736	0.0382	
White wine of quality 7	Volatile Acidity	QBA-normal	0.2333	0.1905	0.0305	
	pH	QBA-normal	0.4012	3.1636	0.0755	
	Sulphates	QBA-logistic	0.2441	0.4068	0.0250	

white wines have similar skewness, but a smaller mode. Overall, white wines of quality 7 have the lowest amount of Volatile Acidity. In terms of pH, there is little difference between the four groups of wines, only white wines of quality 4 have a slightly larger right skewed pH combined with a smaller mode to compensate.

Table S.4 shows the chosen copula based on these margins together with the estimated copula parameters. A gaussian copula turned out to be the best option for all groups, except for the white wines of quality 4 where the dependence is modelled through a Student's t-copula. Overall, the dependency structure shows less pronounced dependencies for higher quality wines and also for white wines compared to red wines. Surprisingly, the dependency between pH and Sulphates is negative for red wines, but positive for the white wines. So as pH increases, the amount of Sulphates decreases for red whines, whereas it increases for white wines.

A visual presentation to see the differences was already provided in Figure 4, showing the

Table S.4: Wine data: chosen copula based on AIC for the four groups of wine data.

Data	Chosen copula	Copula parameters	
Red wine of quality 4	Gaussian	$\widehat{\mathbf{\Sigma}}^{\mathrm{IFM}} = \begin{bmatrix} 1 & 0.3414 & -0.2071 \\ 0.3414 & 1 & -0.1850 \\ -0.2071 & -0.1850 & 1 \end{bmatrix}$	
Red wine of quality 7	Gaussian	$\widehat{\mathbf{\Sigma}}^{\mathrm{IFM}} = \begin{bmatrix} 1 & 0.2663 & -0.2356 \\ 0.2663 & 1 & -0.0266 \\ -0.2356 & -0.0266 & 1 \end{bmatrix}$	
White wine of quality 4	Student's t	$\widehat{\mathbf{\Sigma}}^{\mathrm{IFM}} = \begin{bmatrix} 1 & 0.1718 & -0.0996 \\ 0.1718 & 1 & 0.2266 \\ -0.0996 & 0.2266 & 1 \end{bmatrix}$	& $\widehat{\nu}_C^{IFM} = 9.5341$
White wine of quality 7	Gaussian	$\widehat{\mathbf{\Sigma}}^{\mathrm{IFM}} = \begin{bmatrix} 1 & 0.0535 & -0.0427 \\ 0.0535 & 1 & 0.1810 \\ -0.0427 & 0.1810 & 1 \end{bmatrix}$	

densities of the estimated univariate margins and the estimated bivariate margins (contour plots). A 3D-plot of the fitted densities can be found in Figure S.8. Note that all axes are the same for all 4 sub-figures to facilitate an easy and clear visual comparison. For each four the bottom right quarter (starting at the mode) has been removed for visual purposes.

References for Supplementary Material

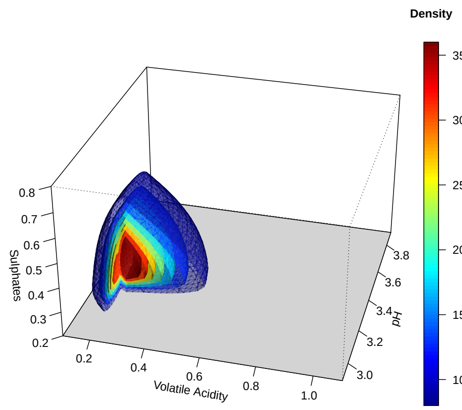
[1] Joe, H. and J. J. Xu (1996, 10). The estimation method of inference functions for margins for multivariate models.

[2] Ko, V. and N. L. Hjort (2019). Copula information criterion for model selection with two-stage maximum likelihood estimation. *Econometrics and Statistics 12*, 167 – 180.

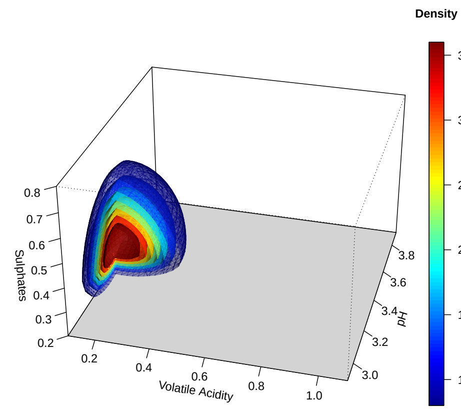
[3] Newey, W. K. and D. McFadden (1994). Chapter 36 large sample estimation and hypothesis testing. In *Handbook of Econometrics*, Volume 4, pp. 2111 – 2245. Elsevier.

[4] Omelka, M., I. Gijbels, and N. Veraverbeke (2009). Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing. *The Annals of Statistics 37*(5B), 3023 – 3058.

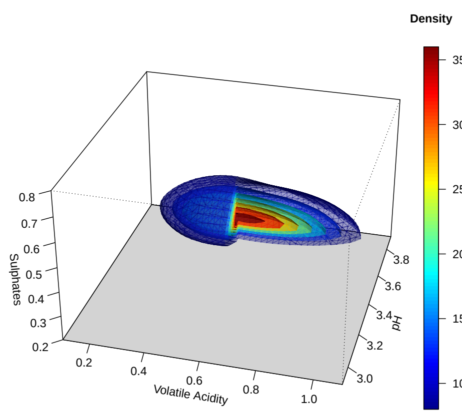
[5] Segers, J. et al. (2012). Asymptotics of empirical copula processes under non-restrictive smoothness assumptions. *Bernoulli 18*(3), 764–782.



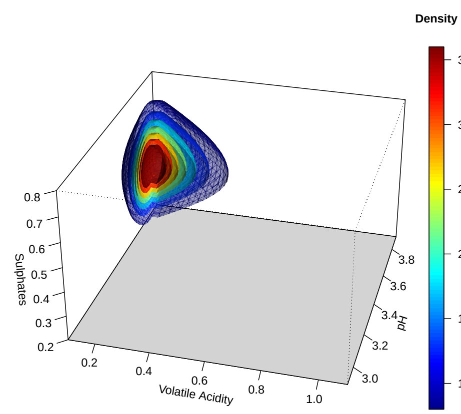
(a) White wine of quality 4.



(b) White wine of quality 7.



(c) Red wine of quality 4.



(d) Red wine of quality 7.

Figure S.8: Wine data: comparison of fitted densities for white and red wines of quality 4 and 7.