

Research Article

Detection of Heterogeneous Structures on the Gaussian Copula Model Using Projective Power Entropy

Akifumi Notsu,¹ Yoshinori Kawasaki,² and Shinto Eguchi²

¹ Department of Statistical Science, The Graduate University for Advanced Studies, Tachikawa, Tokyo 190-8562, Japan

² The Institute of Statistical Mathematics and The Graduate University for Advanced Studies, Tachikawa, Tokyo 190-8562, Japan

Correspondence should be addressed to Akifumi Notsu; notsu@ism.ac.jp

Received 22 July 2013; Accepted 19 September 2013

Academic Editors: E. Miranda and O. Pons

Copyright © 2013 Akifumi Notsu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We discuss a parameter estimation problem for a Gaussian copula model under misspecification. Conventional estimators such as the maximum likelihood estimator (MLE) do not work well if the model is misspecified. We propose the estimator that minimizes the projective power entropy. We call it the γ -estimator, where γ denotes the power index. A feasible form of the projective power entropy is given that suites the Gaussian copula model. It is shown that the γ -estimator is robust against outliers. In addition the γ -estimator can appropriately detect a heterogeneous structure of the underlying distribution, even if the underlying distribution consists of some different copula components while a single Gaussian copula is used as a statistical model. We explore such an ability of the γ -estimator to detect the local structures in the comparison with the MLE. We also propose a fixed point algorithm to obtain the γ -estimator. The usefulness of the proposed methodology is demonstrated in numerical experiments.

1. Introduction

Applications of copula models have been increasing in number in recent years. There are a variety of applications on finance, risk management [1] and multivariate time series analysis [2]. With copula models, the specification of the marginal distributions is parameterized separately from the dependence structure of the joint distribution. Hence, it gives a convenient way of the construction of flexible and more general multivariate distributions. As far as we know, there exist only a few works that are tackled with the identification and the statistical estimation of the mixture of copula models and most of them rely on MCMC algorithm. In this paper we focus on a misspecified Gaussian copula model. In other words, a sample follows a distribution mixed with different sources but a statistical model we fit is just a single Gaussian copula. It is very hard to construct multivariate copulas for three or more random variables [3], while the Gaussian is an exception. So we start with the Gaussian copula model, but later in Section 5 we will show that our method is closely related to t -copula. As an example of misspecification, we consider that the underlying distribution is

$$\tau c_G(\mathbf{u}; P_1) + (1 - \tau) c_G(\mathbf{u}; P_2), \quad (1)$$

where τ is a mixing proportion and $c_G(\mathbf{u}; P)$ denotes the probability density function of the Gaussian copula with the correlation matrix parameter P . We see that the MLE for P almost surely converges to $\tau P_1 + (1 - \tau) P_2$ under the assumption (1), which means that the MLE fails to detect the structure of the underlying distribution.

We make use of the γ -estimator [4, 5] that can be obtained via minimization of the projective power entropy. Here γ denotes the power index, and if $\gamma \rightarrow 0$, the γ -estimator reduces to the MLE. So the γ -estimator can be regarded as an extension of the MLE. In [5], the robustness of the γ -estimator was investigated in a general setting of parametric model. In [6], the minimum density power divergence estimator was proposed, which also uses power of density, for the covariance matrix of multivariate time series, and the robustness was shown. Our research shows that even if a single Gaussian copula model is incorrectly fitted to the data from the mixture distribution (1), the γ -estimator can detect both P_1 and P_2 separately if P_1 and P_2 are “distinct” enough and τ is close to 0.5.

The γ -estimation for the Gaussian copula model relies on the projective power cross entropy between the underlying distribution and the Gaussian copula model $c_G(\mathbf{u}; P)$.

The projective power cross entropy, which is a function of P , has only one local minimum or some local minima depending on the underlying distribution. We show that if P_1 and P_2 are “distinct” enough and τ is near 0.5, then the projective power cross entropy between the underlying mixture distribution (1) and the Gaussian copula $c_G(\mathbf{u}, P)$ has two local minimizers near P_1 and P_2 , respectively, so we propose to use these local minimizers to detect P_1 and P_2 .

This paper is organized as follows. The γ -estimator and the MLE for the Gaussian copula model and a fixed point algorithm to obtain the γ -estimator and the MLE are given in Section 2. Section 3 states the relationship between the projective power entropy and the γ -estimator. We introduce an appropriate measure for the Gaussian copula model since the projective power entropy is defined with respect to some carrier measure. Section 4 reveals the property of the γ -estimator to detect heterogeneous structures. Section 5 elucidates the relationship between maximum entropy distributions and the γ -estimation. The robustness of the γ -estimator is discussed based on its influence function in Section 6. A simulation study is given in Section 7, and discussions are given in the last section. The proofs for all the theoretical results are provided in the appendix.

2. Estimation of the Gaussian Copula Model

In Section 2.1, the γ -estimator for the Gaussian copula model is discussed, and in the followed subsection the MLE for the Gaussian copula model is given. The last subsection lays out a fixed point algorithm to obtain the γ -estimator and the MLE.

2.1. The γ -Estimator for the Gaussian Copula Model. The density function of the Gaussian copula is given by

$$c_G(\mathbf{u}; P) = \det P^{-1/2} \exp\left(-\frac{1}{2} \mathbf{x}_G(\mathbf{u})^\top (P^{-1} - I_m) \mathbf{x}_G(\mathbf{u})\right),$$

$$\mathbf{u} \in [0, 1]^m, \quad (2)$$

where $\mathbf{x}_G(\mathbf{u}) = (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_m))^\top$, $\Phi(x)$ denotes the cumulative distribution function of the standard normal distribution, P is a correlation matrix, and I_m is the identity matrix of size m . Let $v(P)$ be the $m(m-1)/2$ -dimensional vector which consists of the column-wise stacked lower diagonal elements of P . For example, $v(P) = (p_{21}, p_{31}, p_{32})^\top$ if $m = 3$. The set

$$\{\nu(P) : P \text{ is a correlation matrix of size } m\} \quad (3)$$

is a parameter space of the Gaussian copula model.

Let $\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(n)}$ be a random sample from a copula with the probability density function $c(\mathbf{u})$ while $c_G(\mathbf{u}; P)$ is our statistical model. The loss function associated with the projective power entropy introduced in Section 3.1 is given by

$$L_\gamma(P) = -\det(P)^{-\gamma/2(1+\gamma)} \frac{1}{n} \sum_{i=1}^n \exp\left(-\frac{\gamma}{2} \mathbf{x}^{(i)\top} P^{-1} \mathbf{x}^{(i)}\right), \quad (4)$$

up to constant, where $\mathbf{x}^{(i)} = \mathbf{x}_G(\mathbf{u}^{(i)})$ for $i = 1, \dots, n$. The γ -estimator is proposed as the set of local minimizers of $L_\gamma(P)$ and interpreted as follows. If $L_\gamma(P)$ has a local minimum, the underlying distribution is estimated by $c_G(\mathbf{u}; \hat{P}_\gamma)$ using the minimizer \hat{P}_γ . If $L_\gamma(P)$ has ℓ local minima ($\ell \geq 2$), the underlying distribution is estimated by a mixture of ℓ Gaussian copulas. Each Gaussian copula's parameter is estimated by the corresponding local minimizer.

2.2. The MLE for the Gaussian Copula Model. We consider the MLE for the Gaussian copula model on the same setting as in Section 2.1. The log-likelihood multiplied by $-1/n$ is given by

$$L_0(P) = -\frac{1}{2} \log \det(P)^{-1} + \frac{1}{2n} \sum_{i=1}^n \mathbf{x}^{(i)\top} (P^{-1} - I_m) \mathbf{x}^{(i)}. \quad (5)$$

It is easy to see that $L_0(P)$ and $L_\gamma(P)$ satisfy

$$\lim_{\gamma \rightarrow 0} \frac{L_\gamma(P) + 1}{\gamma} = L_0(P), \quad (6)$$

up to constant, so the MLE will be deemed to be the 0-estimator in terms of the γ -estimator. Generally the γ -estimator can be regarded as an extension of the MLE. It is well known that the MLE does not work well under model misspecification. For example, in the case of (1) the MLE for the Gaussian copula model almost surely converges to $\tau P_1 + (1 - \tau) P_2$, but we cannot detect neither P_1 nor P_2 . If $\tau = 0.5$ and

$$P_1 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}, \quad (7)$$

then $\tau P_1 + (1 - \tau) P_2$ is equal to the identity matrix, which has no meaning in this situation. We cannot use the MLE in the case of misspecification.

2.3. A Fixed Point Algorithm to Obtain the γ -Estimator for the Gaussian Copula Model. We give a fixed point algorithm to obtain the γ -estimator for the Gaussian copula model using the Lagrange-multiplier method. The appendix provides the details of the derivation of the algorithm. We can still make use of this algorithm to obtain the MLE just by setting $\gamma = 0$.

Algorithm

(1) Set an appropriate correlation matrix P_0 .

(2) Given P_t , calculate P_{t+1} by the following update formula:

$$P_{t+1} = \Sigma_t + P_t \operatorname{diag}\left((P_t \odot P_t)^{-1} \operatorname{Diag}(I_m - \Sigma_t)\right) P_t, \quad (8)$$

where \odot denotes the Hadamard product. Σ_t is defined by

$$\Sigma_t = (1 + \gamma) \sum_{i=1}^n w_\gamma(\mathbf{x}^{(i)}, P_t) \mathbf{x}^{(i)\top} \mathbf{x}^{(i)}, \quad (9)$$

where

$$w_\gamma(\mathbf{x}, P) = \frac{\exp\left(-(\gamma/2) \mathbf{x}^\top P^{-1} \mathbf{x}\right)}{\sum_{j=1}^n \exp\left(-(\gamma/2) \mathbf{x}^{(j)\top} P^{-1} \mathbf{x}^{(j)}\right)}. \quad (10)$$

Here $\text{Diag}(M)$ for a square matrix M denotes the column vector which consists of the diagonal elements of M and $\text{diag}(a)$ for a vector a denotes the diagonal matrix whose diagonal elements are the components of a .

(3) For sufficient small given number ε , repeat procedure 2 while

$$\|P_{t+1} - P_t\| > \varepsilon, \quad (11)$$

where $\|A\|$ for a square matrix A denotes the matrix norm defined by $\sqrt{\text{tr}(A^T A)}$.

(4) For all local minimizers, repeat procedures 1-3 for different initial values P_0 .

If we consider the estimation problem on Gaussian distributions with mean $\mathbf{0}$, the update formula for an iteration algorithm to obtain the γ -estimator of the covariance matrix Σ is given by

$$\Sigma_{t+1} = (1 + \gamma) \sum_{i=1}^n w_{\gamma}(\mathbf{x}^{(i)}, \Sigma_t) \mathbf{x}^{(i)\top} \mathbf{x}^{(i)}. \quad (12)$$

See [5] for details. If we consider the optimization problem with the objective function $L_{\gamma}(P)$ without the constraint that the diagonal elements of P are 1, the same iteration algorithm (12) can be deduced. So the second term of the right hand side of (8) appears because of the existence of the constraint.

We make a remark on the algorithm to obtain the MLE, or γ -estimator with $\gamma = 0$. We find rather complicated solution of the MLE if we consider a simpler case of $m = 2$. In [1], an approximate MLE for the Gaussian copula model is shown because it takes quite a while to solve the constrained optimization problem in order to obtain the MLE in high dimensions. The approximate MLE is given by

$$\text{diag}(S)^{-1/2} S \text{diag}(S)^{-1/2}, \quad (13)$$

where $\text{diag}(S)$ is the diagonal matrix whose diagonal elements are equal to those of S . We can easily consider an iteration algorithm to obtain an approximate γ -estimator to combine (12) and (13). The update formula of the algorithm is given by

$$P_{t+1}^* = \text{diag}(\Sigma_{t+1}^*)^{-1/2} \Sigma_{t+1}^* \text{diag}(\Sigma_{t+1}^*)^{-1/2}, \quad (14)$$

where

$$\Sigma_{t+1}^* = (1 + \gamma) \sum_{i=1}^n w(\mathbf{x}^{(i)}, P_t^*) \mathbf{x}^{(i)\top} \mathbf{x}^{(i)}. \quad (15)$$

If n is infinity, P_t and P_t^* converge to the same correlation matrix when t tends to ∞ . However P_t and P_t^* are different in general. P_t is preferred to P_t^* in terms of accuracy.

3. Projective Power Entropy and γ -Estimator

In Section 3.1 the projective power entropy and the γ -estimator are given. In the next subsection we discuss an appropriate measure for the Gaussian copula model.

3.1. Projective Power Entropy and the γ -Estimator. The projective power entropy of $f(\mathbf{x})$ with the index γ and the measure Q is defined as

$$H_{\gamma}(f | Q) = - \left[\int f(\mathbf{x})^{1+\gamma} Q(d\mathbf{x}) \right]^{1/(1+\gamma)}. \quad (16)$$

If Q is the Lebesgue measure denoted by Q_L and $f(\mathbf{x})$ is a probability density function, then we have

$$\lim_{\gamma \rightarrow 0} \frac{H_{\gamma}(f | Q_L) + 1}{\gamma} = H_0(f | Q_L), \quad (17)$$

where $H_0(f | Q_L)$ is defined by $-\int f(\mathbf{x}) \log f(\mathbf{x}) d\mathbf{x}$, which is equivalent to the Boltzmann-Shannon entropy. The projective power cross entropy between $g(\mathbf{x})$ and $f(\mathbf{x})$ with the index γ and the measure Q is defined as

$$C_{\gamma}(g, f | Q) = - \frac{\int g(\mathbf{x}) f(\mathbf{x})^{\gamma} Q(d\mathbf{x})}{\left[\int f(\mathbf{x})^{1+\gamma} Q(d\mathbf{x}) \right]^{\gamma/(1+\gamma)}}. \quad (18)$$

The projective power divergence is given by

$$D_{\gamma}(g, f | Q) = C_{\gamma}(g, f | Q) - H_{\gamma}(g | Q). \quad (19)$$

$D_{\gamma}(g, f | Q)$ satisfies $D_{\gamma}(g, f | Q) \geq 0$, and $D_{\gamma}(g, f | Q) = 0$ if and only if $g(\mathbf{x}) = f(\mathbf{x})$, so $D_{\gamma}(g, f | Q)$ can be seen as a kind of distance between g and f .

Let $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}$ be a random sample from a probability density function $g(\mathbf{x})$ and $f(\mathbf{x}; \boldsymbol{\theta})$ a statistical model. Since we want to find the closest distribution to $g(\mathbf{x})$ in the model $f(\mathbf{x}; \boldsymbol{\theta})$, we want to find the minimizer of $D_{\gamma}(g, f(\cdot; \boldsymbol{\theta}) | Q)$, which is equal to the minimizer of $C_{\gamma}(g, f(\cdot; \boldsymbol{\theta}) | Q)$. If Q has the Radon-Nikodym derivative q , then $C_{\gamma}(g, f(\cdot; \boldsymbol{\theta}) | Q)$ is equal to

$$- \int g(\mathbf{x}) \kappa_{\gamma}(\boldsymbol{\theta} | Q) f(\mathbf{x}; \boldsymbol{\theta})^{\gamma} q(\mathbf{x}) d\mathbf{x}, \quad (20)$$

where

$$\kappa_{\gamma}(\boldsymbol{\theta} | Q) = \left(\int f(\mathbf{x}, \boldsymbol{\theta})^{1+\gamma} q(\mathbf{x}) d\mathbf{x} \right)^{-\gamma/(1+\gamma)}. \quad (21)$$

Empirically the projective power cross entropy $C_{\gamma}(g, f(\cdot; \boldsymbol{\theta}) | Q)$ can be estimated by

$$L_{\gamma}(\boldsymbol{\theta} | Q) = - \frac{1}{n} \sum_{i=1}^n \kappa_{\gamma}(\boldsymbol{\theta} | Q) f(\mathbf{x}^{(i)}; \boldsymbol{\theta})^{\gamma} q(\mathbf{x}^{(i)}), \quad (22)$$

which is called the loss function. Note that

$$E(L_{\gamma}(\boldsymbol{\theta} | Q)) = C_{\gamma}(g, f(\cdot; \boldsymbol{\theta}) | Q). \quad (23)$$

The original γ -estimator $\hat{\boldsymbol{\theta}}_{\gamma}$ is defined by the minimizer of the loss function $L_{\gamma}(\boldsymbol{\theta} | Q)$:

$$\hat{\boldsymbol{\theta}}_{\gamma} = \underset{\boldsymbol{\theta}}{\text{argmin}} L_{\gamma}(\boldsymbol{\theta} | Q). \quad (24)$$

Note that we are not necessarily seeking to the global minimizer. Rather, we allow the loss function to be multimodal, so we refer the γ -estimator to the set of the local minimizers. See [4, 5] for details of the γ -estimator.

3.2. Choice of the Carrier Measure. In calculating the γ -estimator, the measure Q can be determined by each user. Here we propose, for Gaussian copula models, the use of a measure, denoted by Q_G , of which Radon-Nikodym derivative is given by $J(\mathbf{x}_G)^{-\gamma}$, where $J(\mathbf{x}_G)$ is the Jacobian of the transformation $\mathbf{x}_G(\mathbf{u})$. From now on we refer this choice to Q_G and explain its rationale by virtue of invariance.

We assume that $\mathbf{x} = (x_1, \dots, x_m)^T \sim \phi(\mathbf{x}; P)$, where $\phi(\mathbf{x}; P)$ denotes the probability density function of the m -dimensional Gaussian distribution with mean $\mathbf{0}$ and correlation matrix P . Let $\mathbf{u} = (\Phi(x_1), \dots, \Phi(x_m))^T$; then $\mathbf{u} \sim c_G(\mathbf{u}; P)$. If the underlying distribution of \mathbf{x} is $g(\mathbf{x})$, then $\mathbf{u} \sim c(\mathbf{u})$, where $c(\mathbf{u})$ is given by $g(\mathbf{x}_G(\mathbf{u}))J(\mathbf{x}_G)$. It is noteworthy that the projective power cross entropy between $g(\mathbf{x})$ and $\phi(\mathbf{x}; P)$ based on \mathbf{x} is not always equal to the projective power cross entropy between $c(\mathbf{u})$ and $c_G(\mathbf{u}; P)$ based on \mathbf{u} . So the γ -estimator based on \mathbf{x} does not coincide with the γ -estimator based on \mathbf{u} .

It is natural for us to require the equivalence of the two γ -estimators, and therefore we employ the measure $Q_G(\mathbf{u})$. It is striking that the projective power cross entropy between $c(\mathbf{u})$ and $c_G(\mathbf{u}; P)$ calculated under the measure Q_G is equal to the projective power cross entropy between $g(\mathbf{x})$ and $\phi(\mathbf{x}; P)$ calculated under the Lebesgue measure Q_L ,

$$C_\gamma(c, c_G(\cdot; P) | Q_G) = C_\gamma(g, \phi(\cdot; P) | Q_L), \quad (25)$$

which is proportional to

$$-\det(P)^{-\gamma/2(1+\gamma)} \int g(\mathbf{x}) \exp\left(-\frac{\gamma}{2} \mathbf{x}^T P^{-1} \mathbf{x}\right) d\mathbf{x}. \quad (26)$$

Obviously there is equalization of the two γ -estimators. Note that the loss function associated with cross entropy (26) becomes (4).

The argument above extends to a general statement. For given one to one transformation $\mathbf{y}(\mathbf{x}) : \mathbf{x} \mapsto \mathbf{y}$, $\mathbf{x}(\mathbf{y})$ denotes the inverse function of $\mathbf{y}(\mathbf{x})$ and $J(\mathbf{y} \mapsto \mathbf{x})$ denotes the Jacobian of the transformation $\mathbf{x}(\mathbf{y})$. Any nonnegative functions $g(\mathbf{x})$, $f(\mathbf{x})$ satisfy

$$\begin{aligned} C_\gamma(g(\mathbf{x}(\cdot))J(\mathbf{y} \mapsto \mathbf{x}), f(\mathbf{x}(\cdot))J(\mathbf{y} \mapsto \mathbf{x}) | Q) \\ = C_\gamma(g, f | Q_L), \end{aligned} \quad (27)$$

if and only if the Radon-Nikodym derivative of Q is equal to $J(\mathbf{y} \mapsto \mathbf{x})^{-\gamma}$. When $g(\mathbf{x})$ and $f(\mathbf{x})$ are the probability density functions, to consider the projective power cross entropy on \mathbf{x} under the Lebesgue measure is equal to consider the projective power cross entropy on \mathbf{y} under the measure having $J(\mathbf{y} \mapsto \mathbf{x})^{-\gamma}$ as its Radon-Nikodym derivative.

4. Property of the γ -Estimator

The γ -estimator for the Gaussian copula model under infinite sample size is equal to the set of the local minimizers of $C_\gamma(c, c_G(\cdot; P) | Q_G)$. In this section we leave aside the empirical loss function $L_\gamma(P)$ for the moment and investigate the property of the γ -estimator (at n infinity) through $C_\gamma(c, c_G(\cdot; P) | Q_G)$. First we consider the case where there is no misspecification.

Theorem 1. If $c(\mathbf{u}) = c_G(\mathbf{u}; P_0)$, then $C_\gamma(c, c_G(\cdot; P) | Q_G)$ has the local minimum at P_0 .

In this case we note that the γ -estimator is equal to $\{P_0\}$, which implies Fisher consistency. For asymptotic properties the γ -estimator has asymptotic consistency and normality.

Next we consider the misspecification case where the true data generating process is given by (1). We see that

$$\begin{aligned} C_\gamma(c, c_G(\cdot; P) | Q_G) &= \tau C_\gamma(c_G(\cdot; P_1), c_G(\cdot; P) | Q_G) \\ &\quad + (1 - \tau) C_\gamma(c_G(\cdot; P_2), c_G(\cdot; P) | Q_G), \end{aligned} \quad (28)$$

which is proportional to

$$\begin{aligned} -\det P^{-\gamma/2(1+\gamma)} &\left[\tau \det P_1^{-1/2} \det(P_1^{-1} + \gamma P^{-1})^{-1/2} \right. \\ &\quad \left. + (1 - \tau) \det P_2^{-1/2} \det(P_2^{-1} + \gamma P^{-1})^{-1/2} \right]. \end{aligned} \quad (29)$$

$C_\gamma(c, c_G(\cdot; P) | Q_G)$ is a weighted mean of the two projective power cross entropy. Each component is a unimodal function, bounded above by 0 and has one local minimum at P_1 and P_2 , respectively. So we expect that $C_\gamma(c, c_G(\cdot; P) | Q_G)$ has two local minima and these local minimizers are near P_1 and P_2 , respectively, if P_1 and P_2 are sufficiently "distinct." However it is hard to formulate such a phenomenon mathematically so we show through easy examples and a graph that such a phenomenon occurs. To obtain numerical solutions, we use the expected (or population) version of the algorithm in Section 2.3.

Example 2. In the case with dimension 2, $C_\gamma(c, c_G(\cdot; P) | Q_G)$ is a univariate function of ρ , which is the nondiagonal element of P . Let P_1 and P_2 be

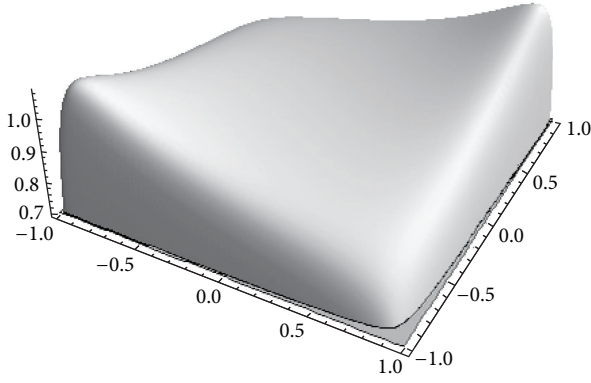
$$P_1 = \begin{pmatrix} 1 & \rho_* \\ \rho_* & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & -\rho_* \\ -\rho_* & 1 \end{pmatrix}, \quad (30)$$

$\gamma = 1$, and $\tau = 0.5$. If $\rho_* > \sqrt{6 - \sqrt{28}} \doteq 0.842$, then $C_\gamma(c, c_G(\cdot; P) | Q_G)$ has two local minima in the interval $(-1, 0)$ and $(0, 1)$, respectively.

Example 3. Suppose the true correlation matrices P_1 and P_2 are given as follows, and P stands for the parameterization of the statistical model we fit:

$$\begin{aligned} P_1 &= \begin{pmatrix} 1 & 0.9 & 0.9^2 \\ 0.9 & 1 & 0.9 \\ 0.9^2 & 0.9 & 1 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 1 & -0.9 & 0.9^2 \\ -0.9 & 1 & -0.9 \\ 0.9^2 & -0.9 & 1 \end{pmatrix}, \\ P &= \begin{pmatrix} 1 & \rho_1 & \rho_1 \rho_2 \\ \rho_1 & 1 & \rho_2 \\ \rho_1 \rho_2 & \rho_2 & 1 \end{pmatrix}. \end{aligned} \quad (31)$$

We also set $\tau = 0.5$ and $\gamma = 1$. Note that $-C_\gamma(c, c_G(\cdot; P) | Q_G)$ is a function of ρ_1 and ρ_2 . Figure 1 shows the graph of $-C_\gamma(c, c_G(\cdot; P) | Q_G)$. We can see there exist two local maxima at $(0.86, 0.86)$ and $(-0.86, -0.86)$.

FIGURE 1: Illustration of $-C_\gamma(c, c_G(\cdot; P) | Q_G)$.

Example 4. Suppose $\tau = 0.4$ and $\gamma = 1$. When P_1, P_2 , and P are given by

$$\begin{aligned} P_1 &= \begin{pmatrix} 1 & 0.9 & 0.7 & 0.7 \\ 0.9 & 1 & 0.9 & 0.7 \\ 0.7 & 0.9 & 1 & 0.7 \\ 0.7 & 0.7 & 0.7 & 1 \end{pmatrix}, \\ P_2 &= \begin{pmatrix} 1 & -0.9 & 0.7 & 0.7 \\ -0.9 & 1 & -0.9 & -0.7 \\ 0.7 & -0.9 & 1 & 0.7 \\ 0.7 & -0.7 & 0.7 & 1 \end{pmatrix}, \\ P &= \begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_4 & \rho_5 \\ \rho_2 & \rho_4 & 1 & \rho_6 \\ \rho_3 & \rho_5 & \rho_6 & 1 \end{pmatrix}. \end{aligned} \quad (32)$$

Then $C_\gamma(c, c_G(\cdot; P) | Q_G)$ has two local minima at

$$\begin{aligned} P &= \begin{pmatrix} 1 & 0.871 & 0.686 & 0.699 \\ 0.871 & 1 & 0.871 & 0.683 \\ 0.686 & 0.871 & 1 & 0.699 \\ 0.699 & 0.683 & 0.699 & 1 \end{pmatrix}, \\ P &= \begin{pmatrix} 1 & -0.892 & 0.695 & 0.700 \\ -0.892 & 1 & -0.892 & -0.696 \\ 0.695 & -0.892 & 1 & 0.700 \\ 0.700 & -0.696 & 0.700 & 1 \end{pmatrix}. \end{aligned} \quad (33)$$

Like these examples, $C_\gamma(c, c_G(\cdot; P) | Q_G)$ has some local minima depending on the underlying distribution. Owing to this property we can detect the heterogeneous structures of the underlying distribution under misspecification.

5. Maximum Entropy Distribution

So far we have considered the γ -estimation of the Gaussian copula model. In this section we uncover that the choice of copula model can be characterized in terms of the maximum entropy distribution. In this regard, the most closely related work is in [2], in which the MLE on meta- t -distribution is addressed. A t -copula is deduced from a multivariate t -distribution while the meta- t -distribution is constructed by

linking a t -copula to univariate t -distributions as its marginal distributions. In our framework, [2]'s work can be interpreted as the maximum likelihood estimation of t -copulas with the marginals estimated simultaneously. Actually the γ -estimation of Gaussian copulas and the maximum likelihood estimation of t -copulas look very similar and share a common idea.

In [4], it is analyzed what the maximum projective power entropy distributions would be under the given (population) mean vector and covariance matrix. The answer depends on the power index γ . When $\gamma = 0$, the Gaussian distribution emerges as the maximum projective power entropy distribution. If $\gamma < 0$, the t -distribution comes up. We show that a similar result holds for copulas. Let t_γ be the cumulative distribution function of the t -distribution with degrees of freedom ν , and $\mathbf{x}_{t,\nu}(\mathbf{u}) = (t_\gamma^{-1}(u_1), \dots, t_\gamma^{-1}(u_m))^\top$. We suppose that $\gamma = -2/(\nu + m)$, $Q_{t,\nu}(d\mathbf{u}) = J(\mathbf{x}_{t,\nu})^{-\gamma} d\mathbf{u}$. Let $\mathcal{C}_\gamma(P)$ be the set of probability density functions $c(\mathbf{u})$ on $[0, 1]^m$ which satisfy the following equation:

$$\int_{[0,1]^m} c(\mathbf{u}) \mathbf{x}_{t,\nu}(\mathbf{u}) \mathbf{x}_{t,\nu}(\mathbf{u})^\top d\mathbf{u} = \frac{\nu}{\nu-2} P. \quad (34)$$

Let $f_t(\mathbf{x}; \nu, P)$ denote the probability density function of t -distribution with degrees of freedom ν and correlation matrix P , and let $c_t(\mathbf{u}; \nu, P)$ denote the probability density function of its copula (t -copula):

$$\begin{aligned} c_t(\mathbf{u}; \nu, P) &= f_t(\mathbf{x}_{t,\nu}(\mathbf{u}); \nu, P) J(\mathbf{x}_{t,\nu}), \\ f_t(\mathbf{x}; \nu, P) &= \frac{\Gamma((\nu+m)/2)}{\Gamma(\nu/2)} \det(\nu\pi P)^{-1/2} \\ &\quad \times \left(1 + \frac{\mathbf{x}^\top P^{-1} \mathbf{x}}{\nu} \right)^{-(\nu+m)/2}. \end{aligned} \quad (35)$$

Then we see

$$\operatorname{argmax}_{c \in \mathcal{C}_\gamma(P)} H_\gamma(c | Q_{t,\nu}) = c_t(\mathbf{u}; \nu, P). \quad (36)$$

If $\gamma \rightarrow 0$, then $t_\gamma \rightarrow \Phi$ and $Q_{t,\nu} \rightarrow Q_L$. So we see that

$$\operatorname{argmax}_{c \in \mathcal{C}_0(P)} H_0(c | Q_L) = c_G(\mathbf{u}; P). \quad (37)$$

That is, t -copula can be characterized as the maximum projective power entropy distribution on $[0, 1]^m$. Moreover it has limiting equivalence (by letting $\gamma \rightarrow 0$) with the Gaussian copula which is tagged with the maximum Boltzmann-Shannon entropy distribution. We call these maximum projective power entropy copulas the γ -copulas. Let us consider the relationship between the γ -copula and the γ -estimation. Our method is discussed on the pair of the Gaussian copula (0-copula) and γ -estimator. On the other hand [2] discussed on the pair of γ -copula model ($\gamma < 0$) and the MLE (0-estimator). We see a sort of duality relationship between two choices of the pair.

6. Robustness

In this section we examine robustness of the γ -estimator for the Gaussian copula model through its influence function.

The influence function measures the asymptotic bias caused by contamination at the \mathbf{x} . The boundedness of the influence function means boundedness of the influence from the outlier, hence its robustness. The influence function of the γ -estimator is given in Section 6.1. We show that it is bounded when $\gamma > 0$. In the next subsection, a brief simulation is performed.

6.1. Influence Function. The γ -estimator for the Gaussian copula model can be regarded as a functional $T(g)$ of a distribution g defined by

$$\begin{aligned} \text{argsolve} \int_P \exp\left(-\frac{\gamma}{2} \mathbf{x}^\top P^{-1} \mathbf{x}\right) v\left(P^{-1} - (1 + \gamma) P^{-1} \mathbf{x} \mathbf{x}^\top P^{-1}\right) \\ \times g(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \end{aligned} \quad (38)$$

Let $\psi_\gamma(\mathbf{x}; P)$ be

$$\psi_\gamma(\mathbf{x}; P) = \exp\left(-\frac{\gamma}{2} \mathbf{x}^\top P^{-1} \mathbf{x}\right) v\left(P^{-1} - (1 + \gamma) P^{-1} \mathbf{x} \mathbf{x}^\top P^{-1}\right). \quad (39)$$

Then the influence function $\text{IF}(\mathbf{x}; T, g)$ of the γ -estimator is given by

$$\text{IF}(\mathbf{x}; T, g) = -\left[\int \dot{\psi}_\gamma(\mathbf{x}; T(g)) g(\mathbf{x}) d\mathbf{x}\right]^{-1} \psi_\gamma(\mathbf{x}; T(g)), \quad (40)$$

where $\dot{\psi}_\gamma(\mathbf{x}; P) = (\partial/\partial v(P))\psi_\gamma(\mathbf{x}; P)$. See [7] for details. The boundedness of the influence function is equivalent to the boundedness of $\psi_\gamma(\mathbf{x}; P)$. The following theorem gives a bound of $\psi_\gamma(\mathbf{x}; P)$.

Theorem 5. When $\gamma = 0$, that is, for the MLE, the influence function is not bounded. When $\gamma < 0$, the influence function is not bounded. When $\gamma > 0$, the influence function is bounded and a bound is given by

$$\|\psi_\gamma(\mathbf{x}; P)\| \leq \|v(P^{-1})\| + \frac{2(1 + \gamma)}{e\gamma} \|P^{-1/2} \otimes P^{-1/2}\|, \quad (41)$$

where \otimes denotes the Kronecker product and $\|h\|$ for an m -dimensional vector h denotes the Euclidean norm defined by $\sqrt{h^\top h}$.

For example, if P is equal to I_m , then $\|\psi_\gamma(\mathbf{x}; I_m)\| \leq (2(1 + \gamma)/e\gamma)m$.

6.2. Simulation. This subsection describes the results of Monte Carlo simulations carried out in order to examine the robustness of the γ -estimator for the Gaussian copula model. We generate 500 pseudorandom samples of size 500 from distribution

$$0.9c_G(\mathbf{u}, P) + 0.1c_G(\mathbf{u}, I_{10}), \quad (42)$$

TABLE 1: RMSE of the norm for the γ -estimator and MLE.

	\hat{P}_{GE}	\hat{P}_{MLE}
RMSE	0.155	0.808

where $c_G(\mathbf{u}, I_{10})$ is equal to the independent copula and P is given by

$$P = \begin{pmatrix} 1 & 0.846 & \cdots & \cdots & 0.846 \\ & 1 & 0.846 & \cdots & 0.846 \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}. \quad (43)$$

For each sample, we calculate the γ -estimator \hat{P}_{GE} for the Gaussian copula model with $\gamma = 0.5$ and the MLE \hat{P}_{MLE} for the Gaussian copula model. We use the norm $\|\hat{P} - P\|$ as the accuracy measure. Table 1 shows the root mean squared error (RMSE) of the norm for the γ -estimator and MLE. We can see that the norm for the γ -estimator is less than that for the MLE, so we see that the γ -estimator is more robust than the MLE.

7. Simulation Study

The property of the γ -estimator to detect heterogeneous structures is investigated by a bunch of simulations. A comparison of the γ -estimator with the MLE for a mixture Gaussian copula (1) is also discussed.

7.1. Simulation Setup. We conducted two kinds of simulation.

Simulation 1. The underlying distribution was constructed based on the one factor Gaussian copula model [8]. Suppose

$$x_i = a_i W + \sqrt{1 - a_i^2} \varepsilon_i, \quad i = 1, \dots, m, \quad (44)$$

where $W, \varepsilon_1, \dots, \varepsilon_m$ have independently the standard normal distribution. Then we see $\mathbf{x} \sim \Phi(\mathbf{x}; P)$, $\mathbf{u} = (\Phi^{-1}(x_1), \dots, \Phi^{-1}(x_m)) \sim c_G(\mathbf{u}; P)$, where $P = (p_{ij})_{ij}$ satisfies

$$p_{ij} = \begin{cases} 1, & i = j, \\ a_i a_j, & i \neq j. \end{cases} \quad (45)$$

Let the underlying distribution be (1), where $c_G(\mathbf{u}; P_1)$ and $c_G(\mathbf{u}; P_2)$ are made from the one factor Gaussian copula model. This model means that the dependence structure is expressed by the mixture of Gaussian copulas. Set $\tau = 0.5$. P_1 is made with

$$\mathbf{a} = (0.92, 0.92, 0.92, 0.92, 0.92, 0.92, 0.92, 0.92, 0.92, 0.92)^\top, \quad (46)$$

and P_2 with

$$\mathbf{a} = (-0.92, 0.92, -0.92, 0.92, -0.92, 0.92, -0.92, 0.92, -0.92, 0.92)^\top. \quad (47)$$

Then we have

$$P_1 = \begin{pmatrix} 1 & 0.846 & \cdots & 0.846 \\ & 1 & \ddots & \vdots \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}, \quad (48)$$

$$P_2 = \begin{pmatrix} 1 & -0.846 & 0.846 & \cdots & -0.846 \\ & 1 & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}.$$

The γ -estimator for the Gaussian copula model with $\gamma = 0.7$ is investigated. Initial values of P which is used in calculating the γ -estimator are $AR(\pm 0.1)$, $AR(\pm 0.3)$, \dots , $AR(\pm 0.9)$, where $AR(\rho)$ is the correlation matrix whose (i, j) component ($i < j$) is equal to ρ^{j-i} . If the γ -estimator has two components G_1 and G_2 such that

$$\|G_1 - P_1\| < \|G_2 - P_1\|, \quad (49)$$

then G_1 is thought of as an estimator of P_1 and denoted by $\hat{P}_{1,GE}$. Similarly G_2 for P_2 and denoted by $\hat{P}_{2,GE}$.

We adopt the MLE for a mixture Gaussian copula model (1). Although P_1 and P_2 are the correlation matrices, we tentatively view them to be the covariance matrices and use EM-Algorithm to obtain an approximate MLE. The obtained estimators \hat{P}_1 and \hat{P}_2 are not necessarily the correlation matrices, so they are transformed into the correlation matrices by

$$\text{diag}(\hat{P}_i)^{-1/2} \hat{P}_i \text{diag}(\hat{P}_i)^{-1/2}, \quad (50)$$

which is denoted by $\hat{P}_{i,MLE}$ for $i = 1, 2$. The initial value of (τ, P_1, P_2) which is used in calculating the MLE is set to $(0.5, AR(0.5), AR(-0.5))$.

A set of data of size n ($n = 500, 1000$) was generated from (1), and the norm of $\hat{P}_{1,GE} - P_1$, $\hat{P}_{2,GE} - P_2$, $\hat{P}_{1,MLE} - P_1$, and $\hat{P}_{2,MLE} - P_2$ were calculated. 500 simulations were carried out, and then, we calculated the RMSE of the norm based on 500 norm values obtained by simulation. The results are shown in Table 3.

Simulation 2. Suppose that the underlying distribution is

$$c(\mathbf{u}) = \tau_1 c_G(\mathbf{u}; P_1) + \tau_2 c_G(\mathbf{u}; P_2) + (1 - \tau_1 - \tau_2) c_G(\mathbf{u}; I_{10}), \quad (51)$$

where $\tau_1 = \tau_2 = 0.45$ and P_1, P_2 are the same in Simulation 1. The other settings are the same as in Simulation 1. The results are shown in Table 5.

7.2. Result

Result of Simulation 1. Table 2 shows the ratio for the γ -estimator to detect two correlation matrices. For $n = 500$ nearly 80 percent was successful, and for $n = 1000$ it worked out almost perfectly. From Table 3, the MLE had

TABLE 2: Ratio of the number of success for the γ -estimator to detect two correlation matrices.

n	500	1000
Ratio	0.768	0.968

TABLE 3: Simulation 1. RMSE of the norm of $\hat{P}_{1,GE} - P_1$, $\hat{P}_{1,MLE} - P_1$, $\hat{P}_{2,GE} - P_2$, and $\hat{P}_{2,MLE} - P_2$.

	n	$\hat{P}_{1,GE}$	$\hat{P}_{1,MLE}$	$\hat{P}_{2,GE}$	$\hat{P}_{2,MLE}$
RMSE	500	0.600	0.184	0.476	0.186
	1000	0.479	0.127	0.431	0.129

TABLE 4: Ratio of the number of success for the γ -estimator to detect two correlation matrices.

n	500	1000
Ratio	0.61	0.966

TABLE 5: Simulation 2. RMSE of the norm of $\hat{P}_{1,GE} - P_1$, $\hat{P}_{1,MLE} - P_1$, $\hat{P}_{2,GE} - P_2$, and $\hat{P}_{2,MLE} - P_2$.

	n	$\hat{P}_{1,GE}$	$\hat{P}_{1,MLE}$	$\hat{P}_{2,GE}$	$\hat{P}_{2,MLE}$
RMSE	500	0.494	0.946	0.563	0.952
	1000	0.468	1.010	0.438	1.032

better performance than the γ -estimator. However this is natural because the MLE is used under no misspecification.

Result of Simulation 2. Table 4 shows the ratio for the γ -estimator to detect two correlation matrices. Compared to the result of Simulation 1 the detection rate at $n = 500$ gets worse while at $n = 1000$ the result is almost alike in Table 2. From Table 5, we find the MLE is considerably underperforming and the γ -estimator is much better.

8. Discussion

We have considered an estimation problem for misspecified Gaussian copula model. By the simulation study our methodology has been found to work well for misspecification. Though we did not consider how to determine the value of γ , this problem was considered in [9] for independent component analysis. It could be possible to follow their method in our problem, but it is currently a future problem.

We choose the measure in terms of invariance. However the γ -estimator obtained is equal to the estimator with normal distribution as a statistical model, so it seems natural. If we use Lebesgue measure in calculating the γ -estimator for Gaussian copula model, we cannot calculate the projective power entropy for all the value of γ and P .

Another issue is to what extent the methodology here works for time series data. Because the basic premise of this paper is that we have data as quantiles, our method would fit, for example, the modeling of unconditional loss distribution [1, page 28]. Such a case is of particular interest when the time horizon over which we measure our losses is relatively large. When we are working on the conditional

modeling, our method should be regarded as a tool for the post analysis. As a typical case, we may want to apply our mixture copula approach to multivariate log-return series which are appropriately standardized and declustered by the multivariate GARCH model fitted to them. See [2] for more details.

Appendices

A. Derivation for the Algorithm

We derive the estimation equation for P . Since P is symmetric and positive definite, there exists a matrix R of size m which satisfies $P = RR^\top$. The i th diagonal element of P is expressed by $e_i^\top RR^\top e_i$, where e_i is the m -dimensional column vector whose i th element is 1 and the other elements are 0. Since the diagonal elements of P are equal to 1, Lagrange function becomes

$$\begin{aligned} \Lambda(R, \lambda) = & (\det R^{-1})^{\gamma/(1+\gamma)} \sum_{i=1}^n \exp\left(-\frac{\gamma}{2} \mathbf{x}^{(i)\top} R^{-1\top} R^{-1} \mathbf{x}^{(i)}\right) \\ & + \sum_{i=1}^m \lambda_i (e_i^\top RR^\top e_i - 1), \end{aligned} \quad (\text{A.1})$$

where $\lambda = (\lambda_1, \dots, \lambda_m)^\top$ is Lagrange multiplier. We differentiate (A.1) with respect to R^{-1} with the technique in [10]. The differential of $\sum_{i=1}^m \lambda_i e_i^\top RR^\top e_i$, which is defined in [10, Sections 5.3 and 5.16], is

$$\begin{aligned} d\left(\sum_{i=1}^m \lambda_i e_i^\top RR^\top e_i\right) &= d(\text{tr}(R^\top \text{diag}(\lambda) R)) \\ &= \text{tr}(2R^\top \text{diag}(\lambda) (dR)) \\ &= \text{tr}(-2RR^\top \text{diag}(\lambda) R (dR^{-1})), \end{aligned} \quad (\text{A.2})$$

where $\text{diag}(\lambda)$ is the diagonal matrix whose diagonal elements are $\lambda_1, \dots, \lambda_m$. From Table 2 in [10, Chapter 9] we have

$$\frac{\partial}{\partial R^{-1}} \sum_{i=1}^m \lambda_i e_i^\top RR^\top e_i = -2R^\top \text{diag}(\lambda) RR^\top. \quad (\text{A.3})$$

Set the derivative of (A.1) to 0; then we have

$$\begin{aligned} \frac{\partial \Lambda(R, \lambda)}{\partial R^{-1}} &= \frac{\gamma}{(1+\gamma)} (\det R^{-1})^{\gamma/(1+\gamma)} R^\top \\ &\quad \times \sum_{i=1}^n \exp\left(-\frac{\gamma}{2} \mathbf{x}^{(i)\top} R^{-1\top} R^{-1} \mathbf{x}^{(i)}\right) \\ &\quad - \gamma (\det R^{-1})^{\gamma/(1+\gamma)} R^{-1} \\ &\quad \times \sum_{i=1}^n \exp\left(-\frac{\gamma}{2} \mathbf{x}^{(i)\top} R^{-1\top} R^{-1} \mathbf{x}^{(i)}\right) \mathbf{x}^{(i)} \mathbf{x}^{(i)\top} \\ &\quad - 2R^\top \text{diag}(\lambda) RR^\top \\ &= 0. \end{aligned} \quad (\text{A.4})$$

Multiply R from the left side of (A.4); then (A.4) becomes

$$P = A + aP \text{diag}(\lambda) P, \quad (\text{A.5})$$

where

$$\begin{aligned} A &= (1+\gamma) \sum_{i=1}^n w_\gamma(\mathbf{x}^{(i)}, P) \mathbf{x}^{(i)} \mathbf{x}^{(i)\top}, \\ a &= 2 \left(\frac{\gamma}{1+\gamma} (\det R^{-1})^{\gamma/(1+\gamma)} \right. \\ &\quad \left. \times \sum_{i=1}^n \exp\left(-\frac{\gamma}{2} \mathbf{x}^{(i)\top} R^{-1\top} R^{-1} \mathbf{x}^{(i)}\right) \right)^{-1}. \end{aligned} \quad (\text{A.6})$$

From the constraint about the diagonal elements of P , we have

$$\text{Diag}(I_m - A) = \text{Diag}(aP \text{diag}(\lambda) P). \quad (\text{A.7})$$

In general, for any square matrices X and Y of size m and m -dimensional column vector \mathbf{x} , we have

$$\text{Diag}(X \text{diag}(\mathbf{x}) Y) = (X \odot Y^\top) \mathbf{x}. \quad (\text{A.8})$$

So (A.7) becomes

$$\lambda = \frac{1}{a} (P \odot P)^{-1} \text{Diag}(I_m - A). \quad (\text{A.9})$$

Then we have

$$P = A + P \text{diag}\left((P \odot P)^{-1} \text{Diag}(I_m - A)\right) P, \quad (\text{A.10})$$

and use this estimation equation as an update formula.

B. Proof of Theorem 1

We see that

$$\begin{aligned} C_\gamma(c_G(\cdot; P_0), c_G(\cdot; P) \mid Q_G) \\ \propto -\det P^{1/2(1+\gamma)} \det(P + \gamma P_0)^{-1/2}. \end{aligned} \quad (\text{B.1})$$

Consider a monotone transformation of the right hand side of (B.1) to obtain

$$\begin{aligned} -\log \left[\det P^{1/2(1+\gamma)} \det(P + \gamma P_0)^{-1/2} \right] \\ = -\frac{1}{2(1+\gamma)} \log \det P + \frac{1}{2} \log \det(P + \gamma P_0). \end{aligned} \quad (\text{B.2})$$

For any $P \neq P_0$, let $P_t = (1-t)P_0 + tP$, ($t > 0$) and define $f(t)$ by

$$f(t) = -\frac{1}{2(1+\gamma)} \log \det P_t + \frac{1}{2} \log \det(P_t + \gamma P_0). \quad (\text{B.3})$$

We see

$$\begin{aligned}
 f'(t) &= -\frac{1}{2(1+\gamma)} \operatorname{tr} [P_t^{-1} (P - P_0)] \\
 &\quad + \frac{1}{2} [(P_t + \gamma P_0)^{-1} (P - P_0)] \\
 &= \frac{1}{2(1+\gamma)} \operatorname{tr} \left[\left\{ \left(\frac{1}{1+\gamma} P_t + \frac{\gamma}{1+\gamma} P_0 \right)^{-1} \right. \right. \\
 &\quad \left. \left. - P_t^{-1} \right\} (P - P_0) \right] \quad (\text{B.4}) \\
 &= \frac{1}{2\gamma t} \operatorname{tr} \left[\left\{ \left(\frac{1}{1+\gamma} P_t + \frac{\gamma}{1+\gamma} P_0 \right)^{-1} - P_t^{-1} \right\} \right. \\
 &\quad \left. \times \left\{ P_t - \left(\frac{1}{1+\gamma} P_t + \frac{\gamma}{1+\gamma} P_0 \right) \right\} \right].
 \end{aligned}$$

Let $D_0(c_G(\cdot; P_1), c_G(\cdot; P_2))$ be the Kullback-Leibler divergence between $c_G(\mathbf{u}; P_1)$ and $c_G(\mathbf{u}; P_2)$. It is well known that $D_0(c_G(\cdot; P_1), c_G(\cdot; P_2)) \geq 0$ and equal to 0 if and only if $P_1 = P_2$. So for $P_1 \neq P_2$, we have

$$\begin{aligned}
 D_0(c_G(\cdot; P_1), c_G(\cdot; P_2)) + D_0(c_G(\cdot; P_2), c_G(\cdot; P_1)) \\
 = \frac{1}{2} \operatorname{tr} ((P_1 - P_2)(P_2 - P_1)) > 0. \quad (\text{B.5})
 \end{aligned}$$

If we read as $P_1 = P_t$ and $P_2 = (1/(1+\gamma))P_t + (\gamma/(1+\gamma))P_0$, then we see that $f'(t) > 0$. The proof is complete.

C. Proof of Theorem 5

If $\gamma = 0$ we see

$$\psi_0(\mathbf{x}; P) = v(P^{-1} - P^{-1} \mathbf{x} \mathbf{x}^\top P^{-1}). \quad (\text{C.1})$$

It is obvious that $\|\psi_0(\mathbf{x}; P)\|$ is not bounded with respect to \mathbf{x} . Next if $\gamma \neq 0$, then let $P^{-1} = (P^{-1/2})^2$, where $P^{-1/2}$ is a symmetric matrix. Set $\mathbf{y} = P^{-1/2} \mathbf{x}$; then

$$\psi_\gamma(\mathbf{x}; P) = \exp\left(-\frac{\gamma}{2} \mathbf{y}^\top \mathbf{y}\right) v(P^{-1} - (1+\gamma) P^{-1/2} \mathbf{y} \mathbf{y}^\top P^{-1/2}). \quad (\text{C.2})$$

Express \mathbf{y} in polar coordinate; then

$$\mathbf{y} = r p(\boldsymbol{\theta}) = r(\cos \theta_1, \sin \theta_1 \cos \theta_2, \dots, \sin \theta_1 \cdots \sin \theta_{m-1})^\top, \quad (\text{C.3})$$

where $0 \leq r, 0 \leq \theta_1, \dots, \theta_{m-2} \leq \pi, 0 \leq \theta_{m-1} \leq 2\pi$. Hence

$$\begin{aligned}
 \psi_\gamma(\mathbf{x}; P) &= \exp\left(-\frac{\gamma}{2} r^2\right) \\
 &\quad \times v(P^{-1} - r^2 (1+\gamma) P^{-1/2} p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top P^{-1/2}). \quad (\text{C.4})
 \end{aligned}$$

If $\gamma < 0$ and $r \rightarrow \infty$, then we see $\|\psi_\gamma(\mathbf{x}; P)\|$ is not bounded. Next if $\gamma > 0$, we see

$$\begin{aligned}
 \|\psi_\gamma(\mathbf{x}; P)\| &\leq \exp\left(-\frac{\gamma}{2} r^2\right) \|v(P^{-1})\| \\
 &\quad + \exp\left(-\frac{\gamma}{2} r^2\right) r^2 (1+\gamma) \\
 &\quad \times \|v(P^{-1/2} p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top P^{-1/2})\|. \quad (\text{C.5})
 \end{aligned}$$

Since $\exp(-(\gamma/2)r^2) \leq 1$, $\exp(-(\gamma/2)r^2)r^2 \leq 2/(e\gamma)$,

$$\begin{aligned}
 \|\psi_\gamma(\mathbf{x}; P)\| &\leq \|v(P^{-1})\| \\
 &\quad + \frac{2(1+\gamma)}{e\gamma} \|\operatorname{vec}(P^{-1/2} p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top P^{-1/2})\|, \quad (\text{C.6})
 \end{aligned}$$

where vec denotes the vec operator. In addition we observe

$$\begin{aligned}
 \|\operatorname{vec}(P^{-1/2} p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top P^{-1/2})\| \\
 = \|(P^{-1/2} \otimes P^{-1/2}) \operatorname{vec}(p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top)\| \quad (\text{C.7}) \\
 \leq \|(P^{-1/2} \otimes P^{-1/2})\| \|\operatorname{vec}(p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top)\|.
 \end{aligned}$$

Since

$$\|\operatorname{vec}(p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top)\| = \operatorname{tr}(p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top p(\boldsymbol{\theta}) p(\boldsymbol{\theta})^\top)^{1/2} = 1, \quad (\text{C.8})$$

we see

$$\|\psi_\gamma(\mathbf{x}; P)\| \leq \|v(P^{-1})\| + \frac{2(1+\gamma)}{e\gamma} \|P^{-1/2} \otimes P^{-1/2}\|. \quad (\text{C.9})$$

D. Proof of the Statement t -Copula Is the Maximum Projective Power Entropy Distribution with $\gamma < 0$

We see

$$\begin{aligned}
 H_\gamma(c_t(\cdot; \nu, P) \mid Q_{t,\nu}) \\
 = -\left[\int_{[0,1]^m} c_t(\mathbf{u}; \nu, P)^{1+\gamma} Q_{t,\nu}(d\mathbf{u}) \right]^{1/(1+\gamma)} \quad (\text{D.1}) \\
 = -\frac{\int_{[0,1]^m} c_t(\mathbf{u}; \nu, P)^{1+\gamma} Q_{t,\nu}(d\mathbf{u})}{\left[\int_{[0,1]^m} c_t(\mathbf{u}; \nu, P)^{1+\gamma} Q_{t,\nu}(d\mathbf{u}) \right]^{\gamma/(1+\gamma)}}.
 \end{aligned}$$

The numerator of (D.1) becomes

$$\begin{aligned}
 \int_{[0,1]^m} c_t(\mathbf{u}; \nu, P)^{1+\gamma} Q_{t,\nu}(d\mathbf{u}) \\
 = \int_{[0,1]^m} c_t(\mathbf{u}; \nu, P) c_t(\mathbf{u}; \nu, P)^\gamma Q_{t,\nu}(d\mathbf{u}) \quad (\text{D.2}) \\
 = \int_{[0,1]^m} c_t(\mathbf{u}; \nu, P) f_t(\mathbf{x}_{t,\nu}(\mathbf{u}); \nu, P)^\gamma d\mathbf{u} \\
 = \int_{[0,1]^m} c(\mathbf{u}) f_t(\mathbf{x}_{t,\nu}(\mathbf{u}); \nu, P)^\gamma d\mathbf{u},
 \end{aligned}$$

for $c \in \mathcal{C}_\gamma(P)$. Hence

$$\begin{aligned}
 H_\gamma(c_t(\cdot; \nu, P) \mid Q_{t,\nu}) &= - \frac{\int_{[0,1]^m} c(\mathbf{u}) c_t(\mathbf{u}; \nu, P)^\gamma Q_{t,\nu}(d\mathbf{u})}{\left[\int_{[0,1]^m} c_t(\mathbf{u}; \nu, P)^{1+\gamma} Q_{t,\nu}(d\mathbf{u}) \right]^{\gamma/(1+\gamma)}} \quad (\text{D.3}) \\
 &= C_\gamma(c, c_t(\cdot; \nu, P) \mid Q_{t,\nu}) \geq H_\gamma(c \mid Q_{t,\nu}).
 \end{aligned}$$

References

- [1] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton Series in Finance, Princeton University Press, Princeton, NJ, USA, 2005.
- [2] R. Zhang, C. Czado, and A. Min, “Efficient maximum likelihood estimation of copula based meta t -distributions,” *Computational Statistics & Data Analysis*, vol. 55, no. 3, pp. 1196–1214, 2011.
- [3] R. B. Nelsen, *An Introduction to Copulas*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2006.
- [4] S. Eguchi, O. Komori, and S. Kato, “Projective power entropy and maximum Tsallis entropy distributions,” *Entropy*, vol. 13, no. 10, pp. 1746–1764, 2011.
- [5] H. Fujisawa and S. Eguchi, “Robust parameter estimation with a small bias against heavy contamination,” *Journal of Multivariate Analysis*, vol. 99, no. 9, pp. 2053–2081, 2008.
- [6] B. Kim and S. Lee, “Robust estimation for the covariance matrix of multivariate time series based on normal mixtures,” *Computational Statistics & Data Analysis*, vol. 57, pp. 125–140, 2013.
- [7] P. J. Huber, *Robust Statistics*, John Wiley & Sons, New York, NY, USA, 1981.
- [8] J. Hull and A. White, “Valuation of a CDO and an n -th to Default CDS Without Monte Carlo Simulation,” *The Journal of Derivatives*, vol. 12, no. 2, pp. 8–23, 2004.
- [9] M. N. H. Mollah, M. Minami, and S. Eguchi, “Exploring latent structure of mixture ICA models by the minimum β -divergence method,” *Neural Computation*, vol. 18, pp. 166–190, 2006.
- [10] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, UK, 1999.

