Bayesian Mixture Copula Estimation and Selection with Applications

Yujian Liu 1,2, Dejun Xie 1,* and Siyi Yu 2

1 School of Mathematics and Physics, Xi’an Jiaotong-Liverpool University, Suzhou 215123, China; liuyujian1995@163.com
2 School of Economics and Management, Shanghai University of Sport, Shanghai 200438, China; yusiyi@sus.edu.cn
* Correspondence: dejun.xie@xjtlu.edu.cn

Abstract: Mixture copulas are popular and essential tools for studying complex dependencies among variables. However, selecting the correct mixture models often involves repeated testing and estimations using criteria such as AIC, which could require effort and time. In this paper, we propose a method that would enable us to select and estimate the correct mixture copulas simultaneously. This is accomplished by first overfitting the model and then conducting the Bayesian estimations. We verify the correctness of our approach by numerical simulations. Finally, the real data analysis is performed by studying the dependencies among three major financial markets.

Keywords: mixture copulas; copula selection; dependence modeling; Bayesian estimations

1. Introduction

Copula functions are important tools for detecting and modeling statistical relations. While the Pearson correlation remains popular among the applied sciences because of its simplicity, its shortcomings become undeniable in the task of constructing nonlinear behavior; see Murphy [1] (Section 2.5.1) for the undesirable results of using the Pearson correlation in the case of non-linearity. On the other hand, copula functions fully describe the connection between random variables by forming their joint distribution from their univariate margins. Mathematically speaking, following from McNeill et al. [2] (p. 221), a d-dimensional copula $C(u_1, u_2, \ldots, u_d): [0, 1]^d \rightarrow [0, 1]$ is a distribution function with uniform margins. That is, there exists $U_1, U_2, \ldots, U_d$ uniformly distributed such that $C(u_1, u_2, \ldots, u_d) = P(U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d)$. The important Sklar theorem [3] guarantees the validity and usefulness of the copula functions in modeling the co-movement among random variables. It states that for the random variables $X_1, X_2, \ldots, X_d$ with the distribution function $F_d(x_1, x_2, \ldots, x_d)$, there exists a copula function, uniquely defined up to Range $(F_1) \times \text{Range} (F_2) \times \ldots \times \text{Range} (F_d)$ such that $F_d(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d))$, where $C(\cdot)$ refers to a $d$-dimensional copula function. See Embrechts [4] for a careful introduction to the concepts and theorems related to the Sklar theorem with applications in financial risk management.

When applying the copula methods to real life, such as modeling the dependence between global trading markets, it is often insufficient to rely on a single parametric copula family because of the data complexity and heterogeneity. A possible remedy is to use mixture modeling; McLachlan et al. [5] give a recent comprehensive review of this classic statistical topic. In terms of the copula functions, we are therefore motivated to write a copula function as the mixture of others:

$$C_{mix}(u) = \sum_{j=1}^{K} w_j C_j(u; \alpha_j) \quad u \in \mathbb{R}^d, \alpha_j \in \Theta_j, \sum_{j} w_j = 1$$ (1)
where we denote $\Theta_j$ to be the parameter space of the copula $j$ and it is straightforward to verify that (1) is also a copula. In most of the research papers regarding the topics of finite mixture models, it is common to assume that the mixed distributions come from the same parametric family. This is also mentioned in [5,6]. However, in the literature on copula methods, mixture copulas consisting of several parametric copula families are also common. Hu [7] is one of a few pioneering works of the mixture copula application in the field of finance. The author used mixed normal Gumbel/survival Gumbel copula with empirical marginal distributions to model the stock dependence between the FTSE 100 of the UK, Nikkei 225 of Japan, S&P 500 of the USA, and Hang Seng of Hong Kong, S.A.R., and quasi-likelihood was used for parameter estimation, and the chi-square test was applied for the goodness of fit. Arakelian and Karlis [8] use the expectation maximization (E.M.) approach to estimate the mixture copula with two components and use them to detect the changing dependence between financial markets, the combination of the Gaussian, Clayton, and Gumbel copulas is mainly considered there, and the model selection criteria is log-likelihood. Vrac et al. [9] combined the dynamic clustering with the gradient ascent to solve the mixture copula; Frank copula family with nonparametric margins are used in their geographical application, and the best model is selected by minimizing the approximate weight of evidence (A.W.E.), which is defined to be

$$A.W.E. = -\text{MLE} + \sum_j \dim(\Theta_j)(\log N + 3/2).$$

Given the number of mixture components is correct, the asymptotic convergence of their methods is finally obtained. More recently, Liu et al. [10] proposed constructing the semi-parametric conditional mixture copulas to assess the global currency market; their best models are selected by comparing the Bayes information criterion (B.I.C.), and asymptotic consistency is obtained.

The topics discussed in this essay are mixture copula estimation and selection using the Bayesian approach. Some previous works regarding Bayesian copula selections include [11], where the author treats the copula parameters as a nuisance and selects the copula with the highest posterior. That is, for the model $M_l$ and the data $D$

$$M_{\text{best}} = \arg\max_{M_l} P(M_l | D) = \arg\max_{M_l} \int P(D | \theta, M_l) dF(\theta) p(M_l).$$

Their method is free from estimating the copula parameters. However, on the other hand, this selection approach may experience high variance in the case of small data sets with high dimensional copulas, and the computational round-off error becomes significant when we multiply the probability if the data set is large. Silva and Lopes [12] proposed to select the model using deviance information criteria (D.I.C.), expected Bayes information criteria (EBIC), and expected Akaike information criteria (EAIC). They also pointed out the importance of the joint estimation of copula parameters using the Bayesian approach from the perspective of considering parameter dependence. Their work can be viewed as the Bayesian version of the popular frequentist A.I.C. (B.I.C.) approaches. One potential concern here is the trade-off between the computational load and the selection efficiency compared with the classic approach. Wu et al. [13,14] proposed to use the Dirichlet process and select the correct mixture copula from the infinite model. Their methods unify the parameter estimation and the best model selection, and we consider this feature to be convenient and important.

On the other hand, Wu et al.’s [13,14] methods are more suitable for the mixture models where each component is from the same parametric family. For example, $C_{\text{mix}} = \sum_j w_j C(u, \theta_j)$ with $C(\cdot, \theta_j)$ belongs to the normal copula for any $j$. In copula research, it is interesting and meaningful to consider the mixed model with heterogeneous mixture components because the dependence on real data sets is complex and ever-changing. Therefore, we propose to use the Bayesian Monte Carlo sampling approaches so that the copula parameters and the correct heterogeneous components can be determined at once. This is simply performed by writing out the saturated mixture models with all possible components included and estimating it with the Bayesian approaches, which is expected to have more stable behavior than the maximum-likelihood methods [15].
Following the introduction, we proceed to outline some classic parametric copula families, and in Section 3, we present our sampling methods here, including how we overfit the model first and estimate the parameters by using the Bayesian modeling. Connection of our approach with the penalized likelihood method from Wang [16], Cai, and Wang [17] has also been made through the E.M. method. The last two sections are for numerical simulations and real data analysis.

2. Parametric Copula Families

2.1. Elliptical Copulas

The elliptical copulas are one of the most common choices for modeling the dependence structures among variables, especially in high dimensional settings [18]. From the Sklar theorem, copulas are of the form

\[ C(u_1, u_2, \ldots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)). \]  

(2)

Since the elliptical distribution is closed under the marginalization, we can therefore get the corresponding parametric copula implicitly defined by (2). For example, by inverting the marginal of the standard multivariate normal distribution, we obtain the normal copula, which is

\[ C_C(u_1, u_2, \ldots, u_d) = \int_{-\infty}^{\phi^{-1}(u_d)} \cdots \int_{-\infty}^{\phi^{-1}(u_1)} ((2\pi)^d |C|)^{-1/2} \exp\left(-\frac{1}{2}x'C^{-1}x\right) dx, \]  

(3)

where \( C \) is the positive definite correlation matrix, and \( x = (\phi^{-1}(u_1), \phi^{-1}(u_2), \ldots, \phi^{-1}(u_d))' \) with \( \phi(\cdot) \) being the quantile function. On the other hand, taking the same action to the multivariate \( t \) distribution yields the \( t \) copula,

\[ C_{t,C}(u_1, u_2, \ldots, u_d) = \int_{-\infty}^{l_{v}^{-1}(u_d)} \cdots \int_{-\infty}^{l_{v}^{-1}(u_1)} \frac{\Gamma((\frac{v}{2})d)/\Gamma(\frac{v}{2})}{\sqrt{\pi v}d|C|} (1 + \frac{y'C^{-1}y}{v})^{d/2} dy, \]  

(4)

\( l_{v}^{-1}(u) \) is the quantile function of the univariate standard \( t \) distribution with \( v \) degree of freedom and \( y = (l_{v}^{-1}(u_1), l_{v}^{-1}(u_2), l_{v}^{-1}(u_3), \ldots, l_{v}^{-1}(u_d))' \), \( C \) is a correlation matrix. The respective copula density \( c(\cdot) \) can be obtained due to the differentiation

\[ f(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_d^{-1}(u_d)) = c(u_1, u_2, \ldots, u_d) \prod_{j=1}^{d} f_j(F_j^{-1}(u_j)). \]  

(5)

One potential advantage of using the \( t \) copula (4) over the normal copula is its ability to model the tail dependence. That is, we wish to measure the degree of dependence on the upper tail \( \rho_u \) and on the lower tail \( \rho_l \). Taking two dimensional copulas as examples, we have for the corresponding random vector \((X_1, X_2)\)

\[ \rho_l = \lim_{u \to 0} P(X_2 < F_2^{-1}(u) \mid X_1 < F_1^{-1}(u)) = \lim_{u \to 0} \frac{C(u, u)}{u}, \]  

\[ \rho_u = \lim_{u \to 1} P(X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)) = \lim_{u \to 1} \frac{1 - 2u + C(u, u)}{u}. \]  

Calculations lead to \( \rho_l = \rho_u = 0 \) for the normal copula but for the \( t \) copula with \( v \) degree of freedom we have,

\[ \rho_l = \rho_u = 2F_{v+1,2}(-\sqrt{(v+1)(1-c)/(1-c)}), \]  

where \( F_{v+1,2}(\cdot) \) is the \( t \) distribution function with \( v \) degree of freedom. One criticism of the elliptical copula families is their symmetric property \( c(u) = c(1-u) \), which might be unrealistic for modeling the asymmetrical correlation that often occurs in the finan-
cial market [19]. Therefore, many authors have proposed the skewed elliptical copula. Smith et al. [20] proposed the skew \( t \) copula and the estimation of the parameters is performed by MCMC. Wu et al. [13] uses a nonparametric Bayesian approach to construct infinite mixture skew normal copula. Wei et al. [21] explored some theoretical properties of the skew-normal copula. Alternatively, Archimedean families of copulas are another solution for the issue.

### 2.2. Archimedean Copulas

Archimedean copulas have been widely researched and applied in the field of credit risk modeling [2]. They can be constructed by satisfying the following linear additive property, that is,

\[
\varphi^{-1}(C(u_1, u_2, \ldots, u_d)) = \sum_{i=1}^{d} \varphi^{-1}(u_i),
\]

where \( \varphi(\cdot) : [0, +\infty) \rightarrow [0, 1] \) is usually called the Archimedean copula generator satisfying convexity, continuity, and completely monotonicity with \( \varphi(0) = 1 \) and \( \lim_{t \to \infty} \varphi(t) = 0 \). The generator with such properties can be derived from the Laplace transform of the positive random variable \( X \) with its distribution function having \( F_X(0) = 0 \),

\[
\varphi(t) = \int_{0}^{\infty} \exp(-tx)dF_X(x).
\]

Taking different forms of \( \varphi(t) \) yields different Archimedean families of copulas, a comprehensive table can be found in the textbooks [22] (Table 4.1). We give several copulas that we will use with their generators in Table 1, and the corresponding distribution function is \( C(u_1, \ldots, u_d) = \varphi(\sum_{i=1}^{d} \varphi^{-1}(u_i)) \).

<table>
<thead>
<tr>
<th>Copula Type</th>
<th>( \varphi(t) )</th>
<th>( \theta ) Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frank</td>
<td>( \theta^{-1} \ln \left(\frac{1}{1 + \exp(-\theta t)} + \exp(-\theta t)\right) )</td>
<td>( \mathbb{R} \setminus {0} )</td>
</tr>
<tr>
<td>Gumbel</td>
<td>( \exp(-\theta t) )</td>
<td>( [1, +\infty) )</td>
</tr>
<tr>
<td>Clayton†</td>
<td>( (1 + \theta t)^{-\theta^{-1}} )</td>
<td>( (0, +\infty) )</td>
</tr>
</tbody>
</table>

† We denote \((\cdot)_{+} := \max(\cdot, 0)\).

One noticeable property of the Archimedean copula is its exchangeability. That is, for any permutation \( \sigma(i) \) of \( \{1, 2, \ldots, d\} \), we have \( C(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(d)}) = C(u_{1}, u_{2}, \ldots, u_{d}) \). This characteristic would be attractive for some applications, such as portfolio default modeling in the credit market. However, for the more general purpose, it might be undesirable when we have the copula dimension \( d \geq 3 \) since this implies that the connection between variables is assumed to be homogeneous. Some improvement has been made on this problem, including nonexchangeable copulas named asymmetric Archimedean copulas [2,23]

\[
C^\gamma(u_1, u_2, \ldots, u_d) = \left(\prod_{j=1}^{d} u_j^{\gamma_j}\right)C(u_1^{\gamma_1}, u_2^{\gamma_2}, \ldots, u_d^{\gamma_d}).
\]

Otherwise, some amendment in (6) yields so-called nonexchangeable nested Archimedean copula [24].

Different copulas can model different kinds of dependence, and Figure 1 gives us a plot of four different types of copulas. In particular, normal and Frank copulas are symmetric in the sense of \( c(1 - u_1, 1 - u_2) = c(u_1, u_2) \) with 0 tail dependence, but the Frank copula was, in addition, proved to be radial symmetric [25]. The Gumbel copula is able to depict the extreme upper tail dependence with \( p_u = -2^\gamma_{Gumbel} + 2 \) but \( p_l = 0 \).
Oppositely, the Clayton copula is able to describe the extreme lower tail dependence with 
\( p_l = 2^{-\theta_{\text{Clayton}}} \) for \( \theta_{\text{Clayton}} > 0 \), but \( \rho_u = 0 \). Hence by mixing those four copulas, setting

\[
C_{\text{mix}}(u_1, u_2) = w_1 C_{\text{fr}}(u_1, u_2) + w_2 C_{\text{No}}(u_1, u_2) + w_3 C_{\text{Cl}}(u_1, u_2) + w_4 C_{\text{Gu}}(u_1, u_2),
\]

(7)

We would be able to reconstruct vast amounts of nontrivial dependence from here \[17\].

![Figure 1](image)

**Figure 1.** scatter plots of different families of copula with 2000 points, \( \theta = 0.6, 5, 3, \text{and} 3 \) for normal, Clayton, Frank, and Gumbel copulas, respectively.

3. Estimation and Selection

The main model we use to conduct simulations and real data analysis is (7). We also work with three-dimensional mixture Gaussian copulas in Section 4.3 to demonstrate the application of our approach in the higher dimensional situation. The general starting point is to construct the model by writing out

\[
C_{\text{mix}}(u) = \sum_{j=1}^{K} w_j C_j(u; \theta_j),
\]

(8)

with the knowledge that a true model has the form

\[
C^0(u) = \sum_{j=1}^{K'} w_j C_j(u; \theta^0_j), \text{ for } K' \leq K.
\]

We then proceed to directly estimate (8) by the Bayesian approach, where \( C_j(\cdot), C_k(\cdot) \) for any \( j, k \leq K \) can either from the same parametric family or not, although for the former case, one needs to take extra measures for the label switching problems \[6\] (Section 22.3). Rousseau and Mengersen \[15\] showed that by applying this approach to the standard finite mixture distribution, it would clear out the redundant components asymptotically. In particular, they showed for \( w = (w_1, w_2, \ldots, w_K) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \ldots, \alpha_K) \) with \( \text{dim} \left( \theta_j \right) / 2 \)
plus some other regularity conditions, the posterior estimation of weights has the property \( \sum_{j=K'}^K E[w_j | D] = O_p(1/\sqrt{n}) \). This result has shown us the extra stability of the Bayesian estimation due to its shrinkage property compared with the maximum likelihood approach (MLE) since the MLE of an over-fitted model only guarantees the convergence to an unidentifiable set with the limiting distribution \( C_\infty(\cdot) = C^0(\cdot) \) in the domain as \( n \to \infty \) [26]. However, the asymptotic results do not guarantee sparsity. This would cause a failure to identify the correct number of components if, for example, \( i,j,k \leq K, w_i C_i(u) + w_j C_j(u) = w_k C_k(u) \) is achievable in the model setting.

On the other hand, Cai and Wang [17] approached the mixed copula estimation and selection problem using penalized MLE approach. In terms of its nature, this approach is quite similar to Bayesian estimation. However, the authors only applied penalties to the weighting parameters, whereas the Bayesian counterpart typically applies penalties to all parameters. The connection between these two approaches is established using the expectation maximization (EM) approach of the posterior, as outlined at the end of Section 3.2, where we compare the maximization form of the posterior mode and the penalized MLE.

### 3.1. Markov Chain Monte Carlo

We show the sampling algorithm of the model (7), but the spirit of the estimation remains the same for all forms of (8). It is especially straightforward to extend the work to the high dimensional implicit copulas [27], including some skew elliptical copulas. The pseudo-likelihood for the i.i.d data is

\[
p(D_n | \theta_0) = \prod_{i=1}^n c_{\text{mix}}(F_1(X_{i1}; \theta_0), F_2(X_{i2}; \theta_0), \ldots, F_d(X_{id}; \theta_0); \theta_0) \prod_{j=1}^d f_j(X_{ij}; \theta_0).
\]

Or, to avoid misspecification of the marginal models, we use the semiparametric approach. The pseudo-likelihood for the i.i.d data is

\[
p(D_n; \theta_0) \propto \prod_{i=1}^n c_{\text{mix}}(\hat{F}_{i1}(X_{i1}), \hat{F}_{i2}(X_{i2}), \ldots, \hat{F}_{id}(X_{id}); \theta_0) \prod_{j=1}^d f_j(X_{ij}; \theta_0)
\]

where we have

\[
\hat{F}_{ij}(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbb{I}(X_{ij} < x).
\]

Other alternatives of the margins \( \hat{F}_n = (\hat{F}_{n1}, \hat{F}_{n2}, \ldots, \hat{F}_{nd}) \) such as kernel density estimations are also available [28]. Therefore, only \( \theta_0 \) is estimated here. In this paper, we focus on the discussion of semiparametric cases.

We specifying the prior of \( w \) and \( \theta_0 \) with

\[
\pi(w) \sim \text{Dir}(\alpha_1, \alpha_2, \ldots, \alpha_K)
\]

\[
\pi(\theta) \sim N_d(0, I_d)
\]

Note that for any copula parameters which do not have the range \(( -\infty, +\infty) \), when convenient, we transfer them from the original parameter space to \( \mathcal{R} \) so that \( \theta = \phi(\theta_0) \in \mathcal{R} \) and

\[
\pi(\theta_0) \sim \text{Dir}(\alpha_0, \alpha_0, \ldots, \alpha_0) \quad \text{and} \quad \pi(\theta) \sim N_d(0, I_d).
\]

\[
\pi(\theta_0) \sim \text{Dir}(\alpha_0, \alpha_0, \ldots, \alpha_0) \quad \text{and} \quad \pi(\theta) \sim N_d(0, I_d).
\]
(-\infty, +\infty). Hence, we will be able to unify the prior to be normal. In case of the model (7), denote \( \theta_{\text{mix}} \in (-\infty, +\infty)^d \), the original parameters can be obtained by

\[
\theta_{\text{ori}}^{\text{clayton}} = \exp (\theta_{\text{mixclayton}})
\]
\[
\theta_{\text{ori}}^{\text{gumbel}} = \exp (\theta_{\text{mixgumbel}}) + 1
\]
\[
\theta_{\text{ori}}^{\text{normal}} = \frac{1 - \exp (-\theta_{\text{mixnormal}})}{1 + \exp (-\theta_{\text{mixnormal}})},
\]
\[
\theta_{\text{ori}}^{\text{frank}} = \theta_{\text{mixfrank}}
\]

where \( \theta_{\text{ori}} \) refer to the parameters in the classical copula settings. We augment our data to \((X_i, Z_i)\), where \( Z_i \) denotes the cluster of the point \( i \), so that \( p(X_i \mid Z_i = k, \theta_{\text{mix}}) \propto c_k(F(X_i); \theta_k) \). The Metropolis–Hasting algorithm of sampling the posterior \( p(\theta_{\text{mix}}, w \mid D_t) \) follows as:

1. Setting initial values \( \theta_{\text{mix}}^{(0)}, w^{(0)} \).
2. Denote the current round to be \( t \), iteratively updating \( Z_i^{(t)} \) such that \( p(Z_i^{(t)} \mid Z_i^{(t)} \setminus i, D_t, w) \propto p(X_i \mid Z_i^{(t)} \setminus i, D_t, w) p(Z_i^{(t)} \mid w) \) for \( i = 1, 2, \ldots, n \) using Gibbs procedure; this can be sampled from the multinomial distribution with \( p_k = \frac{w_k c_k(F(X_i); \theta_k)}{\sum_k w_k c_k(F(X_i); \theta_k)} \) with \( k = 1, 2, \ldots, K \).
3. For all \( i = 1, 2, \ldots, K \), we propose \( f(\theta_i^t \mid \theta_i^{t-1}) \sim N_d(\theta_i^{t-1}, \Sigma) \) where \( \Sigma \) is updated every 50 iterations from the sample variance of previously accepted points. We accept the \( \theta_i^t = \theta_i^t \) with the acceptance rate

\[
a_i = \frac{\prod_{j=1}^{n_i} c_i(X_{ij}; \theta_i^t) \pi(\theta_i^t) f(\theta_i^{t-1} \mid \theta_i^t)}{\prod_{j=1}^{n_i} c_i(X_{ij}; \theta_i^{t-1}) \pi(\theta_i^{t-1}) f(\theta_i^t \mid \theta_i^{t-1})}.
\]

4. Update \( w \sim \text{Dir}(a_1 + \sum_{i=1}^n I(Z_i^{(t)} = 1), \ldots, a_K + \sum_{i=1}^n I(Z_i^{(t)} = K)) \).
5. Repeat steps 2-4 until the stopping criteria are reached, for example, after 10,000 iterations. The MCMC method would be sufficient for our purpose. However, by setting up the EM method for the posterior mode, we can bridge between the Bayesian methods and the penalized likelihood methods discussed in [16,17]. In addition, if the gradient information of the copula is available, it would be faster to work with the EM to get the parameter estimations.

### 3.2. EM Algorithm

Start from the complete data \((X_i, Z_i)\) where \( Z_i \) is the cluster label as previously. Therefore, we denote \( Q(Z) := \log p(w, \theta_{\text{mix}}, Z \mid X) \); our goal is to work iteratively so that

\[
(w^{t+1}, \theta^{t+1}) = \arg\max_{\theta, w} \int Q(Z) p(Z \mid X, \theta^{t+1}, w^{t}) dZ = \arg\max_{\theta, w} E_{p(Z \mid X, \theta^{t}, w^t)} Q
\]

In more detail,

\[
Q(Z) = \log p(w, \theta_{\text{mix}}, Z \mid X)
\]
\[
\propto \log p(X \mid w, \theta_{\text{mix}}, Z) + \log p(Z \mid w) + \log p(w, \theta_{\text{mix}})
\]

\[
\propto \log \prod_{i=1}^{n} \prod_{j=1}^{K} c_i(F_n(X_i); \theta_j)^{1(Z_i = j)} + \log \prod_{i=1}^{n} \prod_{j=1}^{K} w_j^{I(Z_i = j)}
\]

\[
+ \log \prod_{j=1}^{K} w_j^{(a_j-1)} - \sum_{j=1}^{K} \frac{1}{2} ||\theta_j||^2 + C,
\]

where we have denoted the irrelevant constant to be \( C \), and \( p(w, \theta_{\text{mix}}) = p(w)p(\theta_{\text{mix}}) \).
Hence, we take the expectation so that the argmax of (10) would be equivalent as

$$\arg\max_{w, \theta_{\text{mix}}} \sum_{ij} \log c_j(\hat{F}_n(X_i); \theta_j) E(I(Z_i = j))$$

$$+ \sum_{ij} E(I(Z_i = j)) \log w_j + \sum_{j=1}^{K} (\alpha_j - 1) \log w_j - \frac{1}{2} \sum_{j=1}^{K} ||\theta_j||^2$$

$$= \sum_{ij} r_{ij}^t \log c_j(\hat{F}_n(X_i); \theta_j) + n \sum_{ij} r_{ij}^t \log w_j - (1 - \frac{1}{K}) \sum_{j} \log w_j - \frac{1}{2} \sum_{j} ||\theta_j||^2,$$

where we have taken $\alpha_j = 1/K$ to make it less informative while satisfying the regularity condition of [15] and

$$r_{ij}^t = \frac{w_j^t c_j(y_i | \theta_j^t)}{\sum_j w_j^t f_j(y_i | \theta_j^t)}.$$

To achieve the maximum, we differentiate with respect to $w_j$ while adding the Lagrange multiplier $\lambda (1 - \sum w_j)$, we have

$$w_j^{t+1} = \frac{1}{N + 1 - K} \left( \sum_i w_i^t c_j(y_i | \theta_i^t) \right) - (1 - \frac{1}{K}).$$

Differentiate with respect to $\theta_j$, and it can be solved numerically using quasi-Newton methods.

We note that the goal of the EM method is to find the mode of the log posterior

$$\log p(w, \theta_{\text{mix}} | X) \propto \sum_i \sum_j w_j c_j(\hat{F}_n(X); \theta_j) - (1 - \frac{1}{K}) \sum_j \log w_j - \frac{1}{2} \sum ||\theta_j||^2$$

$$= \sum_i \sum_j w_j c_j(\hat{F}_n(X); \theta_j) - n \sum_j \Omega_{(1-1/K)}(w_j) - n \sum_j \Omega_{(1/2)}(\theta_j).$$

(13)

This form shares a similar structure as (3.2) in [16] or (3) in [17] despite the fact that they do not penalize the copula parameters. Intuitively, it would be beneficial to penalize them in order to regularize the parameters of $C_j(\cdot)$ when this copula has 0 weightings. Wang [16] proved the $\sqrt{n}$—asymptotic consistency and sparsity of their semiparametric SCAD-penalized likelihood approaches. The consistency of our Bayesian methods will be tested empirically in the next part. However, the theoretical demonstrations are more challenging to consider with Dirichlet distribution priors due to the singularity of log posterior distribution priors due to the singularity of $w_j = 0$ [29].

One shortcoming of using the EM method is the difficulty in obtaining the confidence interval of estimators. Bootstrap could be a very computationally intensive solution. On the other hand, one may consider the fisher information matrix $-\nabla^2 \log p(X | \hat{w}, \hat{\theta}_{\text{mix}})$ as an asymptotic approximation of the precision matrix. Gelman et al. [6] (p. 324) provide an approach to iteratively calculate the asymptotic variance matrix along with the parameter estimations.

4. Numerical Simulations

4.1. Markov Chain Monte Carlo

We perform two types of numerical simulations. Firstly, we assume that the marginal distributions of the data are perfectly known. Therefore, we focus on the estimation of the copula using the data $(U_{i1}, U_{i2}, \ldots, U_{id}) \sim (F_{i1}(X_1), F_{i2}(X_2), \ldots, F_{id}(X_d))$ for $i = 1, 2, \ldots, n$. the dimension $d$ is set to be 2 for our simulation purpose. Our working model is (7). That is,

$$C_{\text{mix}}(u_1, u_2) = w_1 C_{Fr}(u_1, u_2) + w_2 C_{No}(u_1, u_2) + w_3 C_{Cl}(u_1, u_2) + w_4 C_{Ga}(u_1, u_2).$$
We sample the data from different true models, which are submodels of (7), and we estimate them using the MCMC method of Section 3.1. Secondly, we assume that the marginal distributions are unknown, we hence estimate the margins empirically using \( \hat{F}_{np}(x) = \frac{1}{n+1} \sum_{i=1}^{n} I(X_{ip} \leq x) \). Thus, we have \( (\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_d) = (\hat{F}_1(X_1), \hat{F}_2(X_2), \ldots, \hat{F}_d(X_d)) \) for \( i = 1, 2, \ldots, n \) and the copula parameters can be estimated thereafter.

We simulate 3000 iterations for all models, with the first 2500 points discarded as the burning stage. The number of the sample points is \( n = 400, 800, 2000 \). Tables 2 and 3 display the simulation results. In general, the weighting parameters as well as the copula parameters of non-zero weighting components approach the truth with decreasing Monte Carlo standard deviation. The mean and error estimations of the copula parameters with zero weightings remain close to its priors, which might be considered as an advantage over the penalized method used in [16,17] as they proved that the zero weighting copula parameters would end up randomly in their parameter spaces by using their penalized likelihood approach. Three major misidentification cases were found in tables, that is, \( n = 400, 800 \) of Frank copulas simulations in Table 2 and \( n = 800 \) of Frank copulas in the Table 3. All cases mentioned seem to be misidentified as normal copulas, which are understandable as the normal copula and Frank copula share very similar structures with zero tail dependence.

4.2. Expectation Maximization

In this part, we investigate the performances of the EM algorithm introduced in Section 3.2. The approach is computationally demanding. Therefore, we only show the results with the sample size of \( n = 200, 400, 800 \) for one-component copulas. Data are generated directly from the true copula models. More specifically, for each sample size of \( n = 200, 400, 800 \), we generate 10 batches of data from the true distribution. Every batch is learned by the EM method, and the stopping criteria are 1000 full iterations or the absolute sum of the parameters increase less than 0.001 for an iteration. We calculate the mean and variance estimators for each sample size. Table 4 displays the results of the EM approach. It shows comparable outcomes with the MCMC. Although all algorithms fail to distinguish the Frank copulas from the normal ones due to their similarities, other copulas are selected with satisfactory accuracy. One clear advantage of using the EM is its convenience in introducing an exit mechanism for unlikely copulas during the training process. That is, due to the shrinkage term of the weight in (12), we can eliminate components when their corresponding weights fall down to non-positive during the training. By adding this procedure, we can automatically consider fewer mixture components at later stages. As we can see from Table 4, there are many components with deterministic 0 weightings. However, the shortcomings of the EM approach are also very clear. It is more computationally demanding especially when we seek to obtain some estimation errors or work with high dimensional copulas. On the other hand, the EM seeks to find the posterior mode which is less favorable than the posterior mean in statistical decision theories, while the MCMC approach gives full posterior distributions, and it is well acknowledged that the performance of the EM could be affected by starting points.
Table 2. MCMC estimations of the copula with the marginal distributions fully known. The numbers inside parentheses indicate standard errors, and estimations of the true components are denoted in bold font.

<table>
<thead>
<tr>
<th>True Copula (Param)</th>
<th>MCMC Estimation</th>
<th>n</th>
<th>Clayon w</th>
<th>θ</th>
<th>Gumbel w</th>
<th>θ</th>
<th>Normal w</th>
<th>θ</th>
<th>Frank w</th>
<th>θ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (0.5)</td>
<td></td>
<td>400</td>
<td>0.089 (0.088)</td>
<td>1.568 (1.670)</td>
<td>0.047 (0.036)</td>
<td>3.801 (2.750)</td>
<td>0.839 (0.105)</td>
<td>0.444 (0.046)</td>
<td>0.025 (0.032)</td>
<td>−0.029 (0.994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.0439 (0.066)</td>
<td>1.173 (1.586)</td>
<td>0.007 (0.011)</td>
<td>2.494 (1.370)</td>
<td>0.940 (0.066)</td>
<td>0.514 (0.026)</td>
<td>0.009 (0.014)</td>
<td>0.113 (0.916)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.039 (0.049)</td>
<td>1.363 (1.424)</td>
<td>0.028 (0.034)</td>
<td>3.087 (1.968)</td>
<td>0.913 (0.063)</td>
<td>0.494 (0.024)</td>
<td>0.020 (0.038)</td>
<td>0.043 (1.120)</td>
</tr>
<tr>
<td>Clayton (5)</td>
<td></td>
<td>400</td>
<td>0.990 (0.011)</td>
<td>4.914 (0.246)</td>
<td>0.005 (0.010)</td>
<td>2.490 (1.612)</td>
<td>0.003 (0.005)</td>
<td>0.526 (0.224)</td>
<td>0.002 (0.003)</td>
<td>0.012 (0.965)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.992 (0.009)</td>
<td>4.876 (0.185)</td>
<td>0.003 (0.006)</td>
<td>2.641 (1.774)</td>
<td>0.003 (0.005)</td>
<td>0.488 (0.207)</td>
<td>0.002 (0.004)</td>
<td>0.037 (0.944)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.996 (0.003)</td>
<td>5.091 (0.133)</td>
<td>0.001 (0.002)</td>
<td>2.411 (1.569)</td>
<td>0.001 (0.002)</td>
<td>0.568 (0.205)</td>
<td>0.001 (0.001)</td>
<td>−0.198 (0.983)</td>
</tr>
<tr>
<td>Gumbel (2.5)</td>
<td></td>
<td>400</td>
<td>0.017 (0.027)</td>
<td>1.676 (1.505)</td>
<td>0.957 (0.038)</td>
<td>2.486 (0.105)</td>
<td>0.022 (0.033)</td>
<td>0.530 (0.210)</td>
<td>0.004 (0.007)</td>
<td>0.080 (1.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.002 (0.004)</td>
<td>1.480 (1.593)</td>
<td>0.991 (0.009)</td>
<td>2.701 (0.071)</td>
<td>0.004 (0.007)</td>
<td>0.545 (0.210)</td>
<td>0.002 (0.005)</td>
<td>0.070 (1.051)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.006 (0.008)</td>
<td>1.442 (1.268)</td>
<td>0.988 (0.014)</td>
<td>2.470 (0.048)</td>
<td>0.005 (0.011)</td>
<td>0.533 (0.194)</td>
<td>0.001 (0.002)</td>
<td>−0.091 (0.968)</td>
</tr>
<tr>
<td>Frank (5)</td>
<td></td>
<td>400</td>
<td>0.061 (0.083)</td>
<td>1.903 (1.512)</td>
<td>0.300 (0.041)</td>
<td>2.386 (2.115)</td>
<td>0.875 (0.087)</td>
<td>0.647 (0.038)</td>
<td>0.033 (0.047)</td>
<td>0.416 (1.037)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.058 (0.041)</td>
<td>3.774 (2.751)</td>
<td>0.019 (0.039)</td>
<td>2.324 (2.043)</td>
<td>0.899 (0.055)</td>
<td>0.603 (0.031)</td>
<td>0.024 (0.031)</td>
<td>0.280 (0.984)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.007 (0.012)</td>
<td>1.358 (1.223)</td>
<td>0.004 (0.007)</td>
<td>2.255 (1.394)</td>
<td>0.205 (0.055)</td>
<td>0.790 (0.041)</td>
<td>0.784 (0.058)</td>
<td>4.408 (0.285)</td>
</tr>
<tr>
<td>0.5 Gumbel (2.5) + 0.5 Clayton (5)</td>
<td></td>
<td>400</td>
<td>0.439 (0.057)</td>
<td>6.079 (0.761)</td>
<td>0.533 (0.059)</td>
<td>2.756 (0.242)</td>
<td>0.024 (0.036)</td>
<td>0.569 (0.207)</td>
<td>0.004 (0.007)</td>
<td>0.176 (1.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.567 (0.034)</td>
<td>5.332 (0.390)</td>
<td>0.429 (0.040)</td>
<td>2.328 (0.143)</td>
<td>0.002 (0.004)</td>
<td>0.514 (0.210)</td>
<td>0.002 (0.004)</td>
<td>0.126 (0.976)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.509 (0.034)</td>
<td>5.111 (0.356)</td>
<td>0.480 (0.032)</td>
<td>2.505 (0.076)</td>
<td>0.005 (0.008)</td>
<td>0.523 (0.200)</td>
<td>0.006 (0.007)</td>
<td>0.182 (1.005)</td>
</tr>
<tr>
<td>0.5 Clayton (5) + 0.5 Normal (0.5)</td>
<td></td>
<td>400</td>
<td>0.513 (0.087)</td>
<td>5.150 (1.054)</td>
<td>0.061 (0.070)</td>
<td>2.606 (2.353)</td>
<td>0.383 (0.095)</td>
<td>0.554 (0.080)</td>
<td>0.044 (0.067)</td>
<td>0.280 (1.036)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>800</td>
<td>0.573 (0.041)</td>
<td>4.107 (0.336)</td>
<td>0.165 (0.079)</td>
<td>1.833 (0.534)</td>
<td>0.191 (0.144)</td>
<td>0.410 (0.126)</td>
<td>0.069 (0.086)</td>
<td>0.365 (1.028)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2000</td>
<td>0.456 (0.035)</td>
<td>5.500 (0.372)</td>
<td>0.069 (0.046)</td>
<td>2.750 (0.788)</td>
<td>0.473 (0.035)</td>
<td>0.466 (0.035)</td>
<td>0.002 (0.003)</td>
<td>−0.105 (0.941)</td>
</tr>
</tbody>
</table>
Table 3. MCMC estimations of the copula with the marginal distributions estimated by empirical distribution. The numbers inside parentheses indicate standard errors, and estimations of the true components are denoted in bold font. The corresponding true marginal distribution is $N(1, 1)$ and $N(0.5, 1)$.

<table>
<thead>
<tr>
<th>True Copula (Param)</th>
<th>MCMC Estimation</th>
<th>n</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>Normal</th>
<th>Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta$</td>
<td>$\theta$</td>
<td>$\theta$</td>
<td>$\theta$</td>
<td>$\theta$</td>
<td></td>
</tr>
<tr>
<td>Normal (0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.003 (0.005)</td>
<td>1.637 (1.785)</td>
<td>0.114 (0.125)</td>
<td>2.015 (1.073)</td>
<td>0.878 (0.124)</td>
<td>0.590 (0.038)</td>
</tr>
<tr>
<td>800</td>
<td>0.08 (0.08)</td>
<td>1.543 (3.321)</td>
<td>0.008 (0.011)</td>
<td>2.569 (1.758)</td>
<td>0.886 (0.102)</td>
<td>0.568 (0.033)</td>
</tr>
<tr>
<td>2000</td>
<td>0.025 (0.039)</td>
<td>0.845 (1.039)</td>
<td>0.021 (0.021)</td>
<td>1.740 (0.786)</td>
<td>0.952 (0.048)</td>
<td>0.541 (0.022)</td>
</tr>
<tr>
<td>Clayton(5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.987 (0.015)</td>
<td>4.856 (0.240)</td>
<td>0.006 (0.014)</td>
<td>2.648 (1.718)</td>
<td>0.004 (0.007)</td>
<td>0.530 (0.204)</td>
</tr>
<tr>
<td>800</td>
<td>0.994 (0.006)</td>
<td>4.733 (0.185)</td>
<td>0.003 (0.005)</td>
<td>2.957 (1.793)</td>
<td>0.001 (0.002)</td>
<td>0.499 (0.226)</td>
</tr>
<tr>
<td>2000</td>
<td>0.996 (0.004)</td>
<td>5.423 (0.130)</td>
<td>0.002 (0.004)</td>
<td>3.438 (2.236)</td>
<td>0.001 (0.002)</td>
<td>0.539 (0.197)</td>
</tr>
<tr>
<td>Gumbel (2.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.009 (0.018)</td>
<td>1.589 (1.533)</td>
<td>0.971 (0.034)</td>
<td>2.830 (0.122)</td>
<td>0.018 (0.031)</td>
<td>0.554 (0.190)</td>
</tr>
<tr>
<td>800</td>
<td>0.005 (0.007)</td>
<td>2.293 (2.862)</td>
<td>0.981 (0.021)</td>
<td>2.652 (0.084)</td>
<td>0.012 (0.019)</td>
<td>0.484 (0.199)</td>
</tr>
<tr>
<td>2000</td>
<td>0.004 (0.008)</td>
<td>2.295 (2.647)</td>
<td>0.993 (0.008)</td>
<td>2.530 (0.044)</td>
<td>0.001 (0.001)</td>
<td>0.522 (0.216)</td>
</tr>
<tr>
<td>Frank (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.012 (0.016)</td>
<td>2.220 (2.809)</td>
<td>0.963 (0.099)</td>
<td>2.333 (1.019)</td>
<td>0.005 (0.010)</td>
<td>0.532 (0.197)</td>
</tr>
<tr>
<td>800</td>
<td>0.147 (0.044)</td>
<td>4.664 (1.414)</td>
<td>0.006 (0.011)</td>
<td>2.430 (1.634)</td>
<td>0.836 (0.048)</td>
<td>0.599 (0.028)</td>
</tr>
<tr>
<td>2000</td>
<td>0.005 (0.007)</td>
<td>2.334 (3.135)</td>
<td>0.050 (0.028)</td>
<td>2.659 (0.833)</td>
<td>0.016 (0.024)</td>
<td>0.453 (0.208)</td>
</tr>
<tr>
<td>0.5 Gumbel (2.5) + 0.5 Clayton (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.551 (0.085)</td>
<td>4.327 (0.556)</td>
<td>0.363 (0.129)</td>
<td>2.586 (0.276)</td>
<td>0.080 (0.135)</td>
<td>0.603 (0.206)</td>
</tr>
<tr>
<td>800</td>
<td>0.413 (0.046)</td>
<td>5.149 (0.526)</td>
<td>0.538 (0.060)</td>
<td>2.645 (0.167)</td>
<td>0.050 (0.050)</td>
<td>0.558 (0.211)</td>
</tr>
<tr>
<td>2000</td>
<td>0.531 (0.030)</td>
<td>4.792 (0.270)</td>
<td>0.464 (0.031)</td>
<td>2.500 (0.089)</td>
<td>0.004 (0.008)</td>
<td>0.508 (0.197)</td>
</tr>
<tr>
<td>0.5 Clayton (5) + 0.5 Normal (0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>0.502 (0.082)</td>
<td>4.871 (0.992)</td>
<td>0.044 (0.077)</td>
<td>2.376 (1.230)</td>
<td>0.450 (0.109)</td>
<td>0.488 (0.077)</td>
</tr>
<tr>
<td>800</td>
<td>0.526 (0.042)</td>
<td>5.321 (0.461)</td>
<td>0.015 (0.020)</td>
<td>2.841 (1.888)</td>
<td>0.444 (0.054)</td>
<td>0.485 (0.048)</td>
</tr>
<tr>
<td>2000</td>
<td>0.534 (0.034)</td>
<td>4.725 (0.325)</td>
<td>0.106 (0.058)</td>
<td>1.751 (0.353)</td>
<td>0.351 (0.070)</td>
<td>0.512 (0.060)</td>
</tr>
</tbody>
</table>
Table 4. EM estimations of the copula with the marginal distributions fully known. The numbers inside parentheses indicate standard errors, and estimations of the true components are denoted in bold font. The starting value of the EM is $w = (0.25, 0.25, 0.25, 0.25), \theta_{mix} = (1, 1, 1, 1)$.

<table>
<thead>
<tr>
<th>True Copula (Param)</th>
<th>EM Estimations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.3. Higher Dimensional Cases

We proceed to test the effectiveness of our approach in a higher-dimensional case. As the classic Archimedean families of copulas are rarely used in high-dimensional applications due to the restriction of their parameter spaces, we apply the more commonly used Gaussian mixture copulas with 3 components to perform the estimations, while the dimension of the data is set to be 3. That is, we use the MCMC sampler to estimate the model

\[ c_{\text{NormalMix}} = w_\alpha c_\alpha(u; \Sigma_\alpha) + w_\beta c_\beta(u; \Sigma_\beta) + w_\gamma c_\gamma(u; \Sigma_\gamma). \]  

(14)

A major obstacle to performing MCMC of such type is the sampling of the correlation matrices \((\Sigma_\alpha, \Sigma_\beta, \Sigma_\gamma)\). The valid sampler should generate symmetric positive definite matrices every time with every entry from 0 to 1 and 1 in their diagonal. Readers are referred to [30] for a detailed approach. On the other hand, when performing the MCMC sampling with mixture copulas from the same parametric families, it should also be noticed that label-switching problems often occurred. This is because that (14) has 3! equivalent forms by just switching the labels; some engineering efforts should be made to mitigate the circumstances. After every round of iteration, one can post-process the model so that the component with the highest weighting always ranks first. In addition, if the weightings are too close to distinguish, further criteria such as \(\det|\Sigma| + \text{trace}(\Sigma)\) should be used.

In this study, we use the data sampled from

\[
\begin{align*}
(\Sigma_{\alpha}^{12}, & \Sigma_{\alpha}^{23}, \Sigma_{\alpha}^{13}) = (0.7, 0.7, -0.6) \\
(\Sigma_{\beta}^{12}, & \Sigma_{\beta}^{23}, \Sigma_{\beta}^{13}) = (0.6, 0.6, 0.6) \\
(\Sigma_{\gamma}^{12}, & \Sigma_{\gamma}^{23}, \Sigma_{\gamma}^{13}) = (-0.7, 0.7, 0.7). 
\end{align*}
\]

Additionally, we set the true weighting of the experiments to be \((w_\alpha, w_\beta, w_\gamma) = (1, 0, 0)\) and \((w_\alpha, w_\beta, w_\gamma) = (0, 0.7, 0.3)\), respectively. Therefore, the true model lies in the parameter spaces of (14). Table 5 displays the results of experiments. It shows good signs of convergence and too close to distinguish, further criteria such as \(\det|\Sigma| + \text{trace}(\Sigma)\) should be used.

Table 5. MCMC estimations of the 3-dimensional mixture Gaussian copulas with the marginal distributions fully known. The numbers inside parentheses indicate standard errors, and estimations of the true components are denoted in bold font. Comp is the abbreviation for component and the components are ordered by their weightings.
5. Real Data Analysis

In the real data analysis, we use financial trading data from three major indices, that is, Standard & Poors 500 (SP500), Shanghai Composite Index (SSEC), and Hang Seng Index (HSI). Daily close prices from 9 October 2017 to 29 September 2022 were extracted; we aligned three series with the common trading days among them, and other days were omitted. To ease the analysis of the dependence pattern among them, we take the log returns respectively so that

$$R_i = \log P_i - \log P_{i-1}, \quad i = 1, 2, 3.$$  

Table 6 shows the Pearson and Spearman correlation among markets. SSEC and HSI display strong levels of dependence, while their connection with SP500 is relatively weak for those two markets. However, as we argued previously, the single metric of correlation does not give the full picture of the dependence. It is therefore reasonable to apply the mixture copula models for further analysis. In addition, the Ljung–Box tests to the absolute values $|R_i|$ of series indicate all series are correlated to themselves through time. Moreover, the augmented Dickey–Fuller tests show that they are covariance stationary. To apply the copula models to the autocorrelated data, we use the standard method of standardizing. That is, the autocorrelation is removed by rescaling the volatility of the GARCH(1,1) model; assume

$$R_t = \mu + \sigma_t z_t \quad \text{i.i.d} \quad z_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha \sigma_{t-1}^2 + \beta z_{t-1}^2 + \gamma,$$  

(15)

Table 6. Pearson and Spearman correlation among three markets from October 2017 to September 2020.

<table>
<thead>
<tr>
<th></th>
<th>SSEC</th>
<th>HSI</th>
<th>SP500</th>
<th>SSEC</th>
<th>HSI</th>
<th>SP500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>0.699</td>
<td>0.18</td>
<td>1</td>
<td>0.679</td>
<td>0.173</td>
</tr>
<tr>
<td>Spearman Correlation</td>
<td>0.699</td>
<td>1</td>
<td>0.25</td>
<td>0.679</td>
<td>1</td>
<td>0.224</td>
</tr>
</tbody>
</table>

Table 7. Parameters estimation of the stocks data with mean estimator and 90% credible interval.

<table>
<thead>
<tr>
<th></th>
<th>SSEC-HSI</th>
<th>SSEC-SP500</th>
<th>HSI-SP500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>$w$</td>
<td>0.280 (0.144, 0.372)</td>
<td>0.685 (0.508, 0.814)</td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td>2.53 (1.65, 3.65)</td>
<td>0.168 (0.069, 0.247)</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$w$</td>
<td>0.104 (0.015, 0.257)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta$</td>
<td>1.484 (1.130, 2.368)</td>
<td></td>
</tr>
</tbody>
</table>
6. Conclusions

In this paper, we discuss the method of selecting and estimating the mixture copula simultaneously. This is achieved by first overfitting the model with all potential mixture components and then estimating the parameters by Bayesian methods. The MCMC and EM methods are proposed to learn the parameters, and we have performed numerical simulations to validate the correctness. Furthermore, we apply the methodology to the financial markets to detect the asymmetry dependencies among them. For future research, the effectiveness of this method for general mixture models can be thoroughly investigated and tested. In addition, a full and thorough comparison among various model selection approaches can be studied. We also expect this method to be useful in improving other empirical studies, such as value at risk and conditional value at risk calculations in financial risk management.

Author Contributions: Conceptualization, D.X.; Methodology, Y.L.; Resources, S.Y.; Writing—original draft, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Please contact the corresponding author to request the data.

Conflicts of Interest: The authors declare no conflict of interest.

References
25. Frank, M.J. On the simultaneous associativity of F(x,y) and x+y-F(x,y). *Aequationes Math.* 1979, 19, 194–226. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.