Bayesian Hierarchical Copula Model for Clusters of Financial Time Series

Paolo Onorati
MEMOTEF, Sapienza University of Rome
Brunero Liseo
MEMOTEF, Sapienza University of Rome

Abstract

We discuss a Bayesian hierarchical copula model for clusters of financial time series. A similar approach has been developed in Zhuang et al. (2020). However, the prior distributions proposed there do not always provide a proper posterior. In order to circumvent the problem, we adopt a proper global-local shrinkage prior, which is also able to account for potential dependence structure among different clusters. Performance of the proposed model is presented through simulations and a real data analysis.

1 Introduction

There is a large literature in hierarchical model setting. The concept to pull the mean of a single group towards the mean across group can be found at least in Kelley (1927). Tiao and Tan (1965) and Hill (1965) consider the one-way random effects model and they discuss a Bayesian approach for the analysis of variance because the frequentist unbiased estimator of the variance of random effects could be negative. For the same model Stone and Springer (1965) discuss and resolves a paradox that arises with the use of the Jeffreys’ prior. The foundation for Bayesian hierarchical linear model is established in Lindley and Smith (1972). Gelman (2006) discuss a review on prior distributions for variance parameters in hierarchical model.

More recently, Zhuang et al. (2020) have introduced a hierarchical model in a copula framework: they suggest to use, for the variance parameters of two different priors: i) the standard improper prior for scale parameters, which is proportional to $\sigma^{-2}$ or ii) a vaguely informative prior, say an inverse gamma density with both parameters equal to a small value.

However, both the above proposals might be impractical: in the first case the posterior is simply not proper (as we show in the Appendix); in the second case, the use of small parameters of the inverse Gamma priors simply hides the problem without actually solving it Berger (2006).

Hobert and Casella (1996) is another review on the effect of improper priors in Gibbs Sampling algorithm.

In this paper we propose a Bayesian hierarchical copula model using a different prior. In particular we adopt a global-local shrinkage prior. These prior
distributions naturally arise in a linear regression framework with high dimensional data and where a sparsity constraint is necessary to put for the vector of coefficients. Several different global-local shrinkage families of priors have been proposed: Park and Casella (2008) and Hans (2009) discuss the Bayesian LASSO; Carvalho et al. (2010) introduce the Horseshoe prior, Armagan et al. (2013) propose a Generalized Double Pareto prior. Here we will use a Dirichlet-Laplace prior, proposed in Bhattacharya et al. (2016), with a slight modification; while in a regression framework, it is natural to adopt a prior which shrinks the parameters towards zero, this is not the case for our hierarchical copula model, where the zero value does not have a particular interpretation in the model. For this reason we need to introduce a further level of hierarchy, assuming a prior distribution on the location of the shrinkage point.

The rest of this paper is organized as follow: next section is devoted to illustrate the statistical model, the prior distribution, highlighting the differences with the approach described in Zhuang et al. (2020); we conclude the section with a description of the sampling algorithm. In the third section we perform a simulation study in order to compare the Mean Square Error of the estimates produced by our model and compare them with a standard maximum likelihood approach. Then we reconsider a data set discussed in Zhuang et al. (2020) and compare the results of the two approaches. We conclude with another illustration of the model in a problem of clustering of financial time series.

2 The Statistical Model

2.1 Likelihood and Priors distributions

Copula representation is a way to recast a multivariate distribution in such a way that the dependence structure is not influenced by the shape, the parametrization, and the unit of measurement of the marginal distributions. Their use in statistical inference and a review on the most popular approaches can be found in Hofert et al. (2018). In this paper we will consider several different parametric forms of copula functions: in particular, in the bivariate case, we will use the standard Archimedean families, namely the Joe, Clayton, Gumbel and Frank copulae. For more than two dimension, we will concentrate on the use of the most popular elliptical versions, namely the Gaussian and the Student $t$ copulae. Since the main objective of the paper is the clusterization of the dependence structure, for the sake of simplicity and without loss of generality, we will assume that all the marginal distribution are known, or equivalently, their parameters have been previously estimated. This way, we can directly work with the transformed variables $U_j = F_{X_j}(x_j), j = 1, \ldots, n$.

Let $c_i(\cdot|\psi_i)$ be the generic copula density function associated with the $i$-th group. The statistical model can be stated as follows:

$$(U_{1i}, U_{2i}, \ldots, U_{di})|\psi_i \sim c_i(\cdot|\psi_i) \quad \forall i = 1, 2, \ldots, m,$$

where $m$ denotes the number of groups or clusters. Set

$$\gamma_i = \log \left( \frac{\psi_i - b_i}{B_i - \psi_i} \right),$$
and assume
\[ \gamma_i | \xi, \tau, \alpha_i \sim \text{Laplace}(\xi, \tau \alpha_i) \quad \forall i = 1, 2, \ldots, m, \]
\[ \tau \sim \text{Gamma}(ma, \frac{1}{2}), \]
\[ (\alpha_1, \alpha_2, \ldots, \alpha_m) \sim \text{Dirichlet}(a, a, \ldots, a), \]
\[ \xi \sim \text{Logis}(0, 1). \] 

In the previous expressions, \( b_i \) and \( B_i \) respectively denotes the lower and the upper bound of the parameter space of the corresponding \( \psi_i \), and \( \gamma_i \) is the mapping of \( \psi_i \) into the real axis; \( d_i \) is the dimension of \( i \)-th group and \( a \) is a hyperparameter which we typically set to 1, although different values can be used. In general, the Archimedean copulae are parametrized in terms of the Kendall’s Tau whose range of values has been restricted to \((0, 1)\) for the for Clayton, Joe and Gumbel copulae, while is is set to \((-1, 1)\) for the Frank copula. In the elliptical case, the Gaussian copula is parametrized in terms of the correlation coefficient \( \rho \) which ranges in \((-1, 1)\); finally, the Student-\( t \) copula has the additional parameter \( \nu \), that is the number of degrees of freedom: a discrete uniform prior on \( \{1, 2, \ldots, 35\} \) has been used here. When the dimension \( d \) of the specific group is larger than two, we restrict the analysis to elliptical copulae with an equi-correlation matrix: in that case, it is well known that the range of the correlation parameter is \((-1/(d - 1), 1)\).

Let \( U \) be entire observed sample and let \( U_{ijk} \) be the \( k \)-th observation of \( i \)-th component in the \( j \)-th group, and let \( n_j \) be the number of observation in the \( j \)-th group. The posterior distribution on the parameter vector \( (\gamma, \xi, \alpha, \tau) \) is then:
\[ p(\gamma, \xi, \alpha, \tau | U) \propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[ c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_ij}|\gamma_i) \right] p(\gamma_i | \xi, \tau, \alpha_i) p(\xi) p(\tau) p(\alpha), \]

where \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_m) \) and \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m). \)

The complex form of the posterior distribution requires the use of simulation based methods of inference. In particular, we will adapt the algorithm of Bhattacharya et al. (2016) with a minor modification for the updates of \( \gamma \) and the shrinkage location \( \xi \). Following Bhattacharya et al. (2016), we introduce a vector \( \beta = (\beta_1, \beta_2, \ldots, \beta_m) \in \mathbb{R}^m \) in order to have a latent variable representation of the \( \gamma \) prior; then
\[ \gamma_i | \xi, \tau, \alpha_i, \beta_i \overset{\text{iid}}{\sim} \text{Normal}(\xi, \beta_i \tau^2 \alpha_i^2) \quad \forall i = 1, 2, \ldots, m, \]
\[ \beta_i \overset{\text{iid}}{\sim} \text{Exp}\left(\frac{1}{2}\right) \quad i = 1, 2, \ldots, m. \]

Here we briefly describe the algorithm. Start the chain a time 0 by drawing a sample from the prior. At time \( t \) we use the following updating procedure:

1. update \( \gamma | \xi, \tau, \alpha, \beta \):
   (a) sample \( \tilde{\gamma}_i \) from a proposal Cauchy\((\gamma_i, \delta_i)\) \( i = 1, 2, \ldots, m, \)
(b) set \( \mathbf{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, \ldots, \tilde{\gamma}_m) \), and compute

\[
q = \frac{\prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[ \tilde{c}_i(U_{1ij}, U_{2ij}, \ldots, U_{d,ij}) | \gamma_{it+1} \right] p(\gamma_{it} | \xi_t, \alpha_t, \beta_t)}{\prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[ c_i(U_{1ij}, U_{2ij}, \ldots, U_{d,ij}) | \gamma_{it} \right] p(\gamma_{it} | \xi_t, \alpha_t, \beta_t)}
\]

(c) sample \( u \sim U(0, 1) \)
(d) set \( \gamma_{t+1} = \tilde{\gamma} \) if \( u \leq q \), \( \gamma_{t+1} = \gamma_t \) otherwise.

2. Update \( \xi | \gamma, \tau, \alpha, \beta \):
   (a) sample \( \tilde{\xi} \) from a proposal Cauchy \( \xi_t, \delta \xi \),
   (b) compute
   \[
   q = \frac{\prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[ p(\gamma_{it+1} | \tilde{\xi}, \tau_t, \alpha_t, \beta_t) \right] p(\tilde{\xi})}{\prod_{i=1}^{m} \prod_{j=1}^{n_i} \left[ p(\gamma_{it+1} | \xi_t, \tau_t, \alpha_t, \beta_t) \right] p(\xi_t)}
   \]
   (c) sample \( u \sim U(0,1) \).
   (d) set \( \xi_{t+1} = \tilde{\xi} \) if \( u \leq q \), \( \xi_{t+1} = \xi_t \) otherwise.

3. Update \( \tau | \gamma, \xi, \alpha, \beta \): sample \( \tau_{t+1} \sim GIG(0, 1, 2 \sum_{i=1}^{m} |\gamma_{it+1} - \xi_{it+1}|) \).

4. Update \( \alpha | \gamma, \xi, \tau, \beta \): sample \( \tilde{\alpha}_i \sim GIG(0, 1, 2 |\gamma_{it+1} - \xi_{it+1}|) \) \( \forall i = 1, 2, \ldots, m \), and set
   \[
   \alpha_{it+1} = \frac{\tilde{\alpha}_i}{\sum_{j=1}^{m} \tilde{\alpha}_j} \quad \forall i = 1, 2, \ldots, m
   \]

5. Update \( \beta_i | \gamma, \xi, \tau, \alpha \) \( \forall i = 1, 2, \ldots, m \): sample \( \tilde{\beta}_i \sim IG(\frac{\gamma_{it+1} + \alpha_{it+1}}{\gamma_{it+1} - \xi_{it+1}}, 1) \) and set
   \[
   \beta_{it+1} = \frac{1}{\tilde{\beta}_i} \quad \forall i = 1, 2, \ldots, m
   \]

In the previous statements, \( \text{Cauchy}(a, b) \) denote a one-dimensional Cauchy distribution with location \( a \) and scale \( b \), while \( \text{GIG}(p, a, b) \) is the generalized inverse Gaussian distribution with density function

\[
f(x) \propto x^{p-1} \exp \left( -\frac{1}{2} a x - \frac{1}{2} b \right).
\]

Notice \( \text{IG}(a, b) \) is the inverse Gaussian distribution and it is known that \( X \sim \text{IG}(a, b) \Rightarrow X \sim \text{GIG} \left( -\frac{1}{2}, \frac{a}{2b} \right) \). Finally, \( \delta_\gamma \) and \( \delta_\xi \) are scalar tuning parameters.

In the case of the Student-t copula, we need to add another step between stride 1 and 2 in order to update \( \nu = (\nu_1, \nu_2, \ldots, \nu_m) \):

- update \( \nu_i | \gamma, \xi, \tau, \alpha, \beta \) \( \forall i = 1, 2, \ldots, m \):
  1. sample \( \tilde{\nu} \) from discrete uniform distribution in \( \{1, 2, \ldots, 35\} \)
  2. compute:

\[
q = \frac{\prod_{j=1}^{n} \left[ c(U_{1ij}, U_{2ij}, \ldots, U_{d,ij}) | \gamma_{it+1}, \tilde{\nu} \right]}{\prod_{j=1}^{n} \left[ c(U_{1ij}, U_{2ij}, \ldots, U_{d,ij}) | \gamma_{it+1}, \nu_i \right]}
\]
3. sample $u \sim U(0, 1)$.
4. set $\nu_{i,t+1} = \tilde{\nu}$ if $u \leq q$, $\nu_{i,t+1} = \nu_t$ otherwise.

### 2.2 Prior distribution of $\xi$

The choice of the prior distribution for the shrinkage location $\xi$ needs some explanation. First of all, notice that, according to our prior specification,

$$P(\gamma_i \leq \xi) = \frac{1}{2} \quad i = 1, \ldots, m,$$

however $\gamma_i = \log \left( \frac{\psi_i - b_i}{\psi_i - B_i} \right)$, so

$$P \left( \psi_i \leq \frac{B_i e^\xi + b_i}{1 + e^\xi} \right) = \frac{1}{2}.$$

Therefore, given $\xi$, the median of $\psi_i$ is $Y_i = (B_i e^\xi + b_i)/(1 + e^\xi)$ for all $i = 1, \ldots, m$. Then it is easy to show that the natural choice of a uniform prior on $Y_i \sim U(b_i, B_i)$ for all $i = 1, \ldots, m$ implies a standard logistic density for $\xi$.

### 2.3 Previous work

Apart from the prior specification, the model described in the previous sections, is the one proposed by Zhuang et al. (2020). We restrict our discussion to the case where each copula expression has one parameter only. Their prior can be stated as follows

$$\gamma_i | \mu_i, \sigma^2_i \overset{ind}{\sim} N(\mu_i, \sigma^2_i) \quad \forall i = 1, 2, \ldots, m,$$

$$\mu_i | \lambda, \delta^2 \overset{ind}{\sim} N(\lambda, \delta^2) \quad \forall i = 1, 2, \ldots, m,$$

$$\sigma^2_i \overset{ind}{\sim} \pi_{\sigma^2}(\cdot) \quad \forall i = 1, 2, \ldots, m,$$

$$\lambda \sim \pi_{\lambda}(\cdot), \quad \delta^2 \sim \pi_{\delta^2}(\cdot).$$

There is no a unique choice for the distributions of $(\sigma^2, \lambda, \delta)$, although the Authors suggest to use weakly informative priors, for example inverse gamma densities with small parameters or, in alternative, an objective prior, for example an improper uniform prior. However one can prove that in the second case the posterior distribution can’t be proper, no matter what the sample size is. We show this result in the Appendix. Moreover even the first solution is not feasible since it tends to hide the problem rather than solving it Berger (2006).

### 3 Simulation Study

We compare the performance of our approach with the results based on a maximum likelihood approach in a simulation study. We will use a Student-$t$ copula with an equi-correlation matrix, and set the number of groups $m$ equal to 5. We repeat the procedure 100 times; at iteration $j$ for the $i$-th group, we sample the true value $\gamma_{ij}^T$ from a standard normal distribution, the degrees of freedom $\nu_{ij}^T$ are sampled from the prior distribution, and the dimensions $d_{ij}$ of the groups
are sampled from the uniform discrete distribution in \{1, 2, \ldots, 5\}. Given the parameters and dimensions of the groups, we sample 20 observations for each group. In the maximum likelihood framework, we estimate

\[
(\hat{\gamma}_{ij}^{\text{mle}}, \hat{\nu}_{ij}^{\text{mle}}) = \arg \max_{\gamma, \nu} \prod_{i=1}^{20} \left[ c(U_{1ij}, U_{2ij}, \ldots, U_{dij} | \gamma_i, \nu_i) \right] \quad i = 1, 2, \ldots, 5,
\]

and compute the standard errors

\[
\hat{SE}_{ij}^{\text{mle}} = (\hat{\gamma}_{ij}^{T} - \hat{\gamma}_{ij}^{\text{mle}})^2 \quad i = 1, 2, \ldots, 5.
\]

In a Bayesian framework, we use the posterior mean as a point estimate, obtained from the use of the MCMC algorithm described above. We have run 6 independent chains of \(2.5 \times 10^5\) scans, discarded the first \(5 \times 10^4\) as a burn-in, and finally compute the \(\hat{\gamma}_{ij}^{\text{bay}}\) through the sample mean of simulation output for all \(i = 1, 2, \ldots, 5\). As a tuning parameters we set \(\delta_\gamma = 10^{-3}\) and \(\delta_\xi = 10^{-1}\). Then we compute

\[
\hat{SE}_{ij}^{\text{Bay}} = (\hat{\gamma}_{ij}^{T} - \hat{\gamma}_{ij}^{\text{bay}})^2 \quad i = 1, 2, \ldots, 5.
\]

Comparison are done in terms of the corresponding mean square errors:

\[
\hat{MSE}_{ij}^{\text{mle}} = \frac{1}{100} \sum_{j=1}^{100} \hat{SE}_{ij}^{\text{mle}}, \quad \hat{MSE}_{ij}^{\text{Bay}} = \frac{1}{100} \sum_{j=1}^{100} \hat{SE}_{ij}^{\text{Bay}}.
\]

Table 1 reports the values \(\hat{MSE}_{i}^{\text{mle}}\) against \(\hat{MSE}_{i}^{\text{Bay}}\) for all groups based on 100 simulations.

|       | 1     | 2     | 3     | 4     | 5     | mean  |
|-------|-------|-------|-------|-------|-------|-------|
| Bayes | 0.1449| 0.1514| 0.1104| 0.1106| 0.1283| 0.1291|
| MLE   | 0.1861| 0.1832| 0.1251| 0.1477| 0.1854| 0.1655|

4 Real Data Applications

This section is devoted to the implementation of the method in two different applications. The first one is the same as in Zhuang et al. (2020) and we include it for comparative purposes. The second one deals with the clustering of financial time series.

4.1 Column Vertebral Data

We apply our model to the Column Vertebral Data, available at UCI Machine Learning Repository. It consists of 60 patients with disk hernia, 150 subjects with spondylolisthesis and 100 healthy individuals; data are available for the following variables: angle of pelvic incidence (PI), angle of pelvic tilt (PT),
lumbar lordosis angle (LL), sacral slope (SS), pelvic radius (PR), and degree of spondylolisthesis (DS). As in Zhuang et al. (2020), we adopt the generalized skew-t distribution for the marginals, use a maximum likelihood estimator in order to calibrate the parameters and then transform data via the fitted cumulative distribution function. Computation were done using the R package sgt available on CRAN. Table 2 reports the values of fitted parameters for the marginals.

Table 2: Fitted parameters for each margin distribution

| Group            | Feature | µ         | σ         | λ         | p         | q         |
|------------------|---------|-----------|-----------|-----------|-----------|-----------|
| Disk Hernia      | PI      | 50.2874   | 13.9408   | 0.9992    | 104.9370  | 50.7792   |
|                  | PT      | 17.3686   | 6.9609    | 0.3137    | 1.8070    | 68.7768   |
|                  | LL      | 32.8948   | 11.7179   | 1.0000    | 5.2906    | 364.8091  |
|                  | SS      | 30.4401   | 7.8546    | -0.1599   | 3.5617    | 1.4520    |
|                  | PR      | 116.5142  | 12.9605   | -0.1742   | 5.9304    | 0.4001    |
|                  | DS      | 2.4849    | 5.4948    | -0.1557   | 1.7725    | 358.2803  |
| Spondylolisthesis| PI      | 71.6191   | 15.0308   | -0.0261   | 1.6375    | 67.3817   |
|                  | PT      | 20.7980   | 11.4766   | 0.2862    | 1.9411    | 44.5023   |
|                  | LL      | 64.0920   | 16.3405   | 0.2633    | 2.1057    | 73.7317   |
|                  | SS      | 49.5130   | 13.1427   | 0.3057    | 46.4772   | 0.0649    |
|                  | PR      | 114.6216  | 15.5666   | 0.0259    | 1.4962    | 32.5924   |
|                  | DS      | 51.6375   | 52.3930   | 0.5757    | 42.0584   | 0.0520    |
| Healthy          | PI      | 51.5086   | 12.4646   | 0.6837    | 2.5388    | 24.2468   |
|                  | PT      | 12.8140   | 6.7551    | -0.1121   | 1.7036    | 71.8428   |
|                  | LL      | 44.9715   | 187.1274  | 0.3583    | 28.3301   | 0.0707    |
|                  | SS      | 38.8785   | 9.6135    | 0.2867    | 1.9040    | 17.9808   |
|                  | PR      | 124.0712  | 53.4395   | 0.1274    | 55.3812   | 0.0364    |
|                  | DS      | 2.1427    | 6.1430    | 0.3069    | 1.2030    | 7.8901    |

Following Zhuang et al. (2020), we consider the same parametric copulae for the bivariate distributions of the features of interest, and for each of these, we construct our Bayesian hierarchical copula model for the 3 groups of subjects. We run 6 independent chains of $2.5 \times 10^6$ simulations and discard the first $5 \times 10^5$. We also set $δ_{γ} = 10^{-3}$ and $δ_{ξ} = 10^{-1}$. Table 3 compares the results of Zhuang et al. (2020) (model A) with our ones (model B).

For easiness of comparisons, we follow Zhuang et al. (2020) and report the results not in terms of the parameter $γ$, but rather according the natural parameter of each copula, that is $ρ$ for the Gaussian copula, and $θ$ for the Archimedean ones.
| Group        | Features  | Copula       | Model A                              |                      | Model B                              |                      |
|--------------|-----------|--------------|--------------------------------------|----------------------|--------------------------------------|----------------------|
| Disk Hernia  | PI vs PT  | Gaussian     | Posterior mean 0.696 0.046 (0.599,0.775) | Posterior CI (95%)   | Posterior mean 0.632 0.073 (0.469,0.751) | Posterior CI (95%)   |
|              | PI vs SS  | Gaussian     | 0.726 0.040 (0.633,0.793)             | 0.680 0.076 (0.506,0.789) |
|              | DS vs PI  | Gaussian     | 0.161 0.098 (-0.031,0.339)            | 0.229 0.126 (-0.041,0.450) |
|              | DS vs PT  | Frank        | -0.511 0.577 (-1.489,0.522)           | -0.245 0.820 (-1.858,1.340) |
|              | DS vs LL  | Gaussian     | 0.244 0.103 (0.031,0.435)             | 0.265 0.109 (0.037,0.462) |
|              | DS vs PR  | Gaussian     | -0.055 0.113 (-0.263,0.175)           | -0.075 0.126 (-0.315,0.174) |
| Spondylolisthesis | PI vs PT  | Frank        | 5.718 0.505 (0.599,0.775)             | 5.719 0.756 (4.383,7.138) |
|              | PI vs SS  | Gumbel       | 1.729 0.099 (1.554,1.943)             | 1.725 0.128 (1.490,1.984) |
|              | DS vs PI  | Frank        | 3.427 0.431 (2.552,4.245)             | 3.674 0.867 (2.447,4.897) |
|              | DS vs PT  | Survival Clayton | 0.887 0.143 (0.608,1.174) | 1.036 0.193 (0.679,1.422) |
|              | DS vs LL  | Frank        | 3.230 0.426 (2.437,4.104)             | 3.191 0.801 (2.016,4.370) |
|              | DS vs PR  | Joe          | 1.466 0.115 (1.265,1.698)             | 1.421 0.154 (1.121,1.734) |
| Healthy      | PI vs PT  | Gaussian     | 0.633 0.038 (0.555,0.699)             | 0.621 0.057 (0.496,0.717) |
|              | PI vs SS  | Gumbel       | 2.574 0.178 (2.239,2.910)             | 2.552 0.235 (2.115,3.023) |
|              | DS vs PI  | Frank        | 1.822 0.430 (0.936,2.632)             | 1.794 1.100 (0.465,3.139) |
|              | DS vs PT  | Gaussian     | 0.242 0.080 (0.085,0.401)             | 0.210 0.102 (-0.009,0.394) |
|              | DS vs LL  | Frank        | 1.409 0.570 (0.335,2.538)             | 1.661 0.680 (0.362,2.970) |
|              | DS vs PR  | Gaussian     | -0.111 0.093 (-0.289,0.065)           | -0.076 0.123 (-0.310,0.169) |
4.2 Financial Data Application

Grouping financial time series is important for diversification purposes; a portfolio manager should avoid to invest in instruments with high degree of positive dependence and clustering procedures allow to construct groups according to some specific risk measure. This way, financial instruments that belong to the same group will show a certain degree of association, however he strength of dependence within groups may be well different in different groups. It is then important to assess the strength of association for each single cluster and a way to perform it is to use a hierarchical structure, as the one discuss in this paper.

As a risk measure, we consider the so called tail index, which measure the strength of dependence between two variables when one of them takes extremely low values. Following De Luca and Zuccolotto (2011) we construct a dissimilarity measure based on the lower tail coefficient. Let \((Y_1, Y_2)\) be a bivariate random vector; the lower tail coefficient \(\lambda_L\) of \((Y_1, Y_2)\) is defined as

\[
\lambda_L = \lim_{u \to 0^+} P(F_{Y_1}(Y_1) \leq u|F_{Y_2}(Y_2) \leq u),
\]

or, equivalently,

\[
\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u},
\]

where \(C(\cdot, \cdot)\) is the cumulative distribution function of the copula associated to \((Y_1, Y_2)\). In order to estimate \(\lambda_L\) we use the empirical estimator discussed in Joe et al. (1992):

\[
\hat{\lambda}_L = \frac{\hat{C}(\sqrt{n}, \sqrt{n})}{\sqrt{n}}
\]

where \(\hat{C}(\cdot, \cdot)\) is the empirical copula and \(n\) is the sample size. The dissimilarity measure is then defined as

\[
d(Y_1, Y_2) = 1 - \lambda_L(Y_1, Y_2)
\]

The preliminary clustering procedure has been implemented using a complete linkage method. Notice that a bivariate lower tail coefficient is not the unique method for modeling dependence on extreme low values: Durante et al. (2014) proposed a conditioned correlation coefficient estimated using a nonparametric approach way; Fuchs et al. (2021) analyze dissimilarity measure applicable to multivariate lower tail coefficient.

We consider the "S&P 500 Full Dataset" available at Kaggle: it contains the more relevant information for the components of S&P 500. We take the daily closing prices from June, 5th, 2000 to June, 5th, 2020, and discard instruments without a complete record for this period. Then we restrict our analysis to 379 components. For all of them, we computed the log-returns by taking the log-differences and filter data by fitting, for each time series, an ARMA(1,1)GJR-GARCH(1,1) model with Student-\(t\) innovations; then we extracted residuals and transformed them through the fitted cumulative distributions functions in order to get the pseudo data. Computation were done using the CRAN package rugarch. Hence we compute the empirical estimator of lower tail coefficient for any possible pair and the dissimilarity measure associated, and use them to feed the clustering algorithm. Due to computational complexity, we have used
the coarsest partition under the constraint that the biggest group must have at most 10 components. We obtained 30 groups with dimension more than 1 and discarded instruments that belong to groups with only 1 component. The final number of instruments was thus reduced to 93.

We run the MCMC algorithm described above for the 30 clusters, doing 12 independent chains of $10^5$ scans, discarding the first $1.5 \times 10^4$ as burn in. Tuning parameters were set to $\delta_\gamma = 10^{-6}$, $\delta_\xi = 10^{-3}$. For each scan and for any group we compute the lower tail coefficient through the formula

$$
\lambda_L = 2T_{\nu+1} \left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right),
$$

where $T_\nu(\cdot)$ is the univariate cumulative distribution function of a Student-$t$ random variable with $\nu$ degree of freedom. The copula used in this example was a Student-$t$ copula with an equi-correlation matrix: as a consequence, we get a single value for the lower tail coefficient for each cluster. The next Table reports the results each pair that belongs in the same group. Finally we report the estimation results.

| Group | Components | Posterior mean | Posterior s.d. | Posterior CI(95%) |
|-------|------------|----------------|----------------|------------------|
| 1     | NTRS STT   | 0.5001         | 0.0592         | (0.4153, 0.5918) |
| 2     | CVX XOM    | 0.4833         | 0.0592         | (0.4061, 0.5715) |
| 3     | AMAT LRCX  | 0.4499         | 0.0633         | (0.3648, 0.5573) |
| 4     | BEN TROW   | 0.4259         | 0.0649         | (0.3457, 0.5359) |
| 5     | CMS PNW    | 0.4256         | 0.0661         | (0.3347, 0.5296) |
| 6     | APD LIN    | 0.4198         | 0.0655         | (0.3389, 0.5274) |
| 7     | PEAK VTR WELL | 0.4170 | 0.0636 | (0.3538, 0.5097) |
| 8     | DHI LEN PHM | 0.3942 | 0.0643 | (0.3137, 0.4895) |
| 9     | MLM VMC    | 0.3827         | 0.0678         | (0.2881, 0.4963) |
| 10    | HD LOW     | 0.3757         | 0.0675         | (0.2828, 0.4851) |
| 11    | COP MRO    | 0.3685         | 0.0681         | (0.2765, 0.4880) |
| 12    | ADP PAYX   | 0.3532         | 0.0692         | (0.2663, 0.4704) |
| 13    | CSX        | 0.3395         | 0.0674         | (0.2672, 0.4535) |
| Group | Components | Posterior mean | Posterior s.d. | Posterior CI(95%) |
|-------|------------|----------------|----------------|-------------------|
| 14    | NSC, T, VZ | 0.3338         | 0.0699         | (0.2368, 0.4509)  |
| 15    | UNP, CAH, MCK | 0.3337       | 0.0691         | (0.2414, 0.4401)  |
| 16    | BAC, JMP, MS | 0.3235       | 0.0671         | (0.2590, 0.4203)  |
| 17    | AIV, AVB, EQR, ESS, UDR | 0.3221 | 0.0668 | (0.2593, 0.4187)  |
| 18    | RSG, WM | 0.3168         | 0.0694         | (0.2275, 0.4255)  |
| 19    | DVN, EOG, NBL | 0.2979       | 0.0682         | (0.2166, 0.4103)  |
| 20    | D, SO | 0.2932         | 0.0708         | (0.1953, 0.4113)  |
| 21    | NI, SRE | 0.2920         | 0.0700         | (0.2022, 0.4032)  |
| 22    | IP, PKG | 0.2914         | 0.0713         | (0.1957, 0.4145)  |
| 23    | CB, TRV | 0.2839         | 0.0715         | (0.1815, 0.4132)  |
| 24    | GL, LNC, MET, UNM | 0.2818 | 0.0677 | (0.2177, 0.3804)  |
| 25    | CMA, FITB, HBAN, KEY | 0.2294 | 0.0666 | (0.1526, 0.3273)  |
| 26    | ATO, EVRG | 0.2201         | 0.0692         | (0.1256, 0.3412)  |
| Group | Components | Posterior mean | Posterior s.d. | Posterior CI(95%) |
|-------|------------|----------------|----------------|-------------------|
| 27    | ETR        | 0.1923         | 0.0652         | (0.1175, 0.2953)  |
|       | NEE        |                |                |                   |
|       | PEG        |                |                |                   |
| 28    | AEE        | 0.1768         | 0.0633         | (0.1174, 0.2855)  |
|       | AEP        |                |                |                   |
|       | DTE        |                |                |                   |
|       | DUK        |                |                |                   |
| 29    | ED         | 0.1522         | 0.0605         | (0.0874, 0.2439)  |
|       | ES         |                |                |                   |
|       | LNT        |                |                |                   |
|       | WEC        |                |                |                   |
|       | XEL        |                |                |                   |
| 30    | EW         | 0.0008         | 0.0011         | (0.0000, 0.0028)  |
|       | SYK        |                |                |                   |

5 Conclusion

We have discussed and improved a fully Bayesian analysis for a hierarchical copula model proposed in Zhuang et al. (2020). We have proposed the use of a proper prior which is able to induce shrinkage and, at the same time, dependence among different clusters of observations. This prior does not mimic the behavior of an improper prior and is better suited for objectively representing the information coming form the data. Our prior belongs to the large family of global-local shrinkage densities, with an extra stage in the hierarchy, due to the absence of a significant shrinkage value; we have experienced that this approach is very effective and useful in the case of parametric copulae depending on a single parameter. In a more general situation, this approach needs to be modified, and this can be easily accommodated.

Finally, we have presented an application in a financial context, where the goal was to estimate the lower tail coefficient of several financial time series in a parametric way using a Student-$t$ copula.
References

Armagan, Artin, David Dunson, and Jaeyong Lee (2013). “Generalized Double Pareto Shrinkage”. In: Statistica Sinica 23.1, pp. 119–143. doi:10.5705/ss.2011.048

Berger, James (2006). “The case for objective Bayesian analysis”. In: Bayesian Analysis 1.3, pp. 385–402. doi:10.1214/06-BA115 url:https://doi.org/10.1214/06-BA115

Bhattacharya, Anirban et al. (2016). “Dirichlet–Laplace Priors for Optimal Shrinkage”. In: Journal of the American Statistical Association 110.512, pp. 1479–1490. doi:10.1080/01621459.2014.960967 url:https://doi.org/10.1080/01621459.2014.960967

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (Apr. 2010). “The horseshoe estimator for sparse signals”. In: Biometrika 97.2, pp. 455–480. doi:10.1093/biomet/asq017

Davis, Carter (2015). sgt: Skewed Generalized T Distribution Tree. R package version 2.0. url:https://CRAN.R-project.org/package=sgt

De Luca, Giovanni and Paola Zuccolotto (2011). “A tail dependence-based dissimilarity measure for financial time series clustering”. In: Advances in Data Analysis and Classification 5.4, pp. 323–340. doi:10.1007/s11634-011-0098-3 url:https://doi.org/10.1007/s11634-011-0098-3

Dua, Graff (2017). UCI Machine Learning Repository. url:http://archive.ics.uci.edu/ml

Durante, Fabrizio, Roberta Pappadà, and Nicola Torelli (2014). “Clustering of financial time series in risky scenarios”. In: Advances in Data Analysis and Classification 8.4, pp. 359–376. doi:10.1007/s11634-013-0160-4 url:https://doi.org/10.1007/s11634-013-0160-4

Fuks, Sebastian, F. Marta L. Di Lascio, and Fabrizio Durante (2021). “Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables”. In: Computational Statistics & Data Analysis 159, p. 107201. url:https://www.sciencedirect.com/science/article/pii/S0167947321000359

Gelman, Andrew (2006). “Prior distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper”. In: Bayesian Analysis 1.3, pp. 515–534. doi:10.1214/06-BA117A url:https://doi.org/10.1214/06-BA117A

Ghalanos, Alexios (2020). rugarch: Univariate GARCH models. R package version 1.4-4.

Hans, Chris (2009). “Bayesian Lasso Regression”. In: Biometrika 96.4, pp. 835–845. doi:10.1093/biomet/asp047 url:https://doi.org/10.1093/biomet/asp047

Hill, Bruce M. (1965). “Inference about Variance Components in the One-Way Model”. In: Journal of the American Statistical Association 60.311, pp. 806–825. doi:10.2307/2283247 url:http://www.jstor.org/stable/2283247

Hobert, James P. and George Casella (1996). “The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models”. In: Journal of the American Statistical Association 91.436, pp. 1461–1473. doi:10.2307/2291572 url:http://www.jstor.org/stable/2291572

Hofert, Marius et al. (2018). Elements of Copula Modeling with R. Springer Use R! Series. isbn: 978-3-319-89635-9. url:http://www.springer.com/de/book/9783319896342

Joe, Harry, Richard L. Smith, and Ishay Weissman (1992). “Bivariate Threshold Methods for Extremes”. In: Journal of the Royal Statistical Society. Series B (Methodological) 54.1, pp. 171–183. url:http://www.jstor.org/stable/2345953

Kelley, Lee Truman (1927). The Interpretation of Educational Measurement. Measurement and adjustment series. Yonkers-on-Hudson, N.Y.: WorldBook Company. url:http://hdl.handle.net/2027/mdp.39015001994071
Appendix

Here we show that the prior proposed in Zhuang et al. (2020) leads to an improper posterior.

The statistical model consists of \( m \) \( d \)-dimensional copulae governing different sets of observations

\[
(U_{1k}, U_{2k}, \ldots, U_{dk}) | \theta_i \sim c_i(\cdot | \theta_i) \quad i = 1, 2, \ldots, m.
\]

Let \( \gamma_i = \eta_i g_i(\theta_i) \); here \( \eta_i \) is a scaling parameter which can be considered known. The one-to-one mapping functions \( g_i(\cdot) \) are needed to put all the dependence parameters on the real line. Zhuang et al. (2020) make the following assumptions:

\[
\begin{align*}
\gamma_i | \mu_i, \sigma_i^2 & \sim \text{ind } N(\mu_i, \sigma_i^2) \quad i = 1, 2, \ldots, m; \\
\mu_i | \lambda, \delta^2 & \sim \text{ind } N(\lambda, \delta^2) \quad i = 1, 2, \ldots, m.
\end{align*}
\]

Hyper-parameters \( \sigma_i^2 \)'s, \( \lambda \) and \( \delta^2 \) are given a suitable prior distribution. For the moment we do not specify the priors and set

\[
\begin{align*}
\sigma_i^2 & \sim \text{ind } \pi_{\sigma_2}(\cdot) \quad i = 1, 2, \ldots, m. \\
\lambda & \sim \pi_{\lambda}(\cdot), \quad \delta^2 \sim \pi_{\delta^2}(\cdot)
\end{align*}
\]

Since the \( g_i(\theta_i) \)'s are one-to-one, we write \( c_i(\cdot | \gamma_i) \) instead of \( c_i(\cdot | \theta_i) \). Let \( U \) be the observed sample, and let \( U_{ijk} \) be the \( k \)-th observation of \( i \)-th component in the \( j \)-th group. Let \( n_j \) be the sample size of the \( j \)-th group. Furthermore, let \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_m), \mu = (\mu_1, \mu_2, \ldots, \mu_m) \) and \( \sigma^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2) \). Finally, let \( S(\omega) \) denote the parameter space of the generic parameter \( \omega \).
The next proposition shows that, using standard noninformative priors for scale and location parameters, the resulting posterior will be improper, independently of the sample size.

**Proposition 1:** If \( \pi_{\sigma^2}(\sigma_i^2) \propto \sigma_i^{-2} \), for \( i = 1, \ldots, m \), and \( \pi_{\delta^2}(\delta^2) \propto \delta^{-2} \), \( \pi_\lambda(\lambda) \propto 1 \), the posterior distribution \( \gamma|U \) is improper for any choice of the copula densities \( c_i(\cdot|\gamma_i) \) and independently of the sample size.

**Proof:** For the sake of clarity set \( d\sigma_2 = d\sigma_1^2 d\sigma_2^2 \ldots d\sigma_m^2 \) and \( d\mu = d\mu_1 d\mu_2 \ldots d\mu_m \).

We need to show that the following pseudo marginal posterior distribution of \( \gamma \) is not integrable:

\[
\pi(\gamma|U) = \int_{S(\mu)} \int_{s(\sigma_2)} \int_{S(\delta)} \int_{S(\lambda)} \pi(\gamma, \mu, \sigma^2, \lambda, \delta^2|U) d\lambda d\delta^2 d\sigma^2 d\mu
\]

\[
\propto \int_{S(\mu)} \int_{s(\sigma_2)} \int_{S(\delta)} \int_{S(\lambda)} \pi(U|\gamma, \mu, \sigma^2, \lambda, \delta^2) \pi(\gamma, \mu, \sigma^2, \lambda, \delta^2) d\lambda d\delta^2 d\sigma^2 d\mu, 
\]

where \( \pi(U|\gamma, \mu, \sigma^2, \lambda, \delta^2) \) represents the likelihood function. Then we obtain

\[
\pi(\gamma|U) \propto \prod_{i=1}^m \left[ \prod_{j=1}^{n_i} c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_{ij}}|\gamma_i) \right] \int_{S(\mu)} \int_{s(\sigma_2)} \pi(\gamma, \mu, \sigma^2) \pi(\sigma^2) \times
\]

\[
\int_{S(\delta)} \int_{S(\lambda)} \pi(\mu|\lambda, \delta^2) \pi(\lambda|\delta^2) d\lambda d\delta^2 d\sigma^2 d\mu 
\]

\[
= \prod_{i=1}^m \left[ \prod_{j=1}^{n_i} c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_{ij}}|\gamma_i) \right] \pi(\gamma), 
\]

with

\[
\pi(\gamma) = \int_{S(\mu)} \int_{s(\sigma_2)} \pi(\gamma|\mu, \sigma^2) \pi(\sigma^2) \pi(\mu) d\sigma^2 d\mu
\]

and

\[
\pi(\mu) = \int_{S(\sigma_2)} \int_{S(\lambda)} \pi(\mu|\lambda, \delta^2) \pi(\lambda|\delta^2) d\lambda d\delta^2. 
\]

Consider only

\[
\pi(\mu) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \pi(\mu|\lambda, \delta^2) \pi(\lambda|\delta^2) d\lambda d\delta^2
\]

\[
\propto \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (2\pi\delta^2)^{-\frac{1}{2}} \exp \left( -\frac{1}{2\delta^2} \sum_{i=1}^m (\mu_i - \lambda)^2 \right) \frac{1}{\delta^2} d\lambda d\delta^2
\]

\[
\propto \int_{-\infty}^{+\infty} \left( \frac{1}{\delta^2} \right)^{\frac{\phi+1}{2}} \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2\delta^2} \sum_{i=1}^m (\mu_i^2 - 2\lambda\mu_i + \lambda^2) \right) d\lambda d\delta^2
\]

\[
= \int_{-\infty}^{+\infty} \left( \frac{1}{\delta^2} \right)^{\frac{\phi+1}{2}} \int_{-\infty}^{+\infty} \exp \left( -\frac{1}{2\delta^2} \left( \sum_{i=1}^m \mu_i^2 - 2\lambda \sum_{i=1}^m \mu_i + m\lambda^2 \right) \right) d\lambda d\delta^2; 
\]
Notice that $F$ or any choice of $m > 1$ can be written as

$$\pi(\mu) \propto \left( \frac{1}{2} \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} \Gamma\left( \frac{m-1}{2} \right) \propto \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}}.$$

Now we compute

$$\pi(\gamma) = \int_{S(\sigma_m)} \cdots \int_{S(\sigma_m)} \pi(\gamma|\mu, \sigma^2) \pi(\mu) \pi(\sigma^2) d\sigma^2 d\mu$$

$$\propto \int_{S(\sigma)} \cdots \int_{S(\sigma) \cdots S(\sigma)} \prod_{i=1}^{m} \left[ (2\pi \sigma_i^2)^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_i^2} (\gamma_i - \mu_i)^2 \right) \right] \times$$

$$\prod_{i=1}^{m} \frac{d\sigma_i^2}{(\sum_{i=1}^{m} (\mu_i - \bar{\mu})^2)^{\frac{1}{2}}} d\mu$$

$$\propto \int_{S(\mu)} \cdots \int_{S(\mu)} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} \prod_{i=1}^{m} \left[ \int_{S(\sigma_i^2)} \left( \frac{1}{\sigma_i^2} \right)^{\frac{3}{2}} \exp \left( -\frac{1}{2\sigma_i^2} (\gamma_i - \mu_i)^2 \right) d\sigma_i^2 \right] d\mu$$

$$\propto \int_{S(\mu)} \cdots \int_{S(\mu)} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} \prod_{i=1}^{m} \left( (\gamma_i - \mu_i)^2 \right)^{-\frac{1}{2}} d\mu$$

$$= \int_{S(\mu)} \frac{1}{\gamma_1 - \mu_1} \int_{S(\mu_2)} \frac{1}{\gamma_2 - \mu_2} \cdots \int_{S(\mu_m)} \frac{1}{\gamma_m - \mu_m} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} d\mu.$$

Notice that

$$\sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 = \sum_{i=1}^{m} \mu_i^2 - m\bar{\mu}^2$$

$$= \mu_m^2 + \sum_{i=1}^{m-1} \mu_i^2 - \frac{1}{m} \left( \sum_{i=1}^{m} \mu_i \right)^2$$

$$= \mu_m^2 + \sum_{i=1}^{m-1} \mu_i^2 - \frac{1}{m} \left( \sum_{i=1}^{m-1} \mu_i \right)^2 + 2\mu_m \left( \sum_{i=1}^{m-1} \mu_i + \mu_m \right);$$

16
set \( K = \sum_{i=1}^{m-1} \mu_i^2 \) and \( H = \sum_{i=1}^{m-1} \mu_i \); then

\[
\sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 = \mu_m^2 + K - \frac{1}{m}(H^2 + 2H\mu_m + \mu_m^2)
\]

\[
= \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2.
\]

So \( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \) is a convex parabolic function of \( \mu_m \), and, by Weierstrass theorem global maximum exists for all bounded and closed sets. Integrating out \( \mu_m \), one gets

\[
\int_{S(\mu_m)} \frac{1}{|\gamma_m - \mu_m|} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right) \frac{1}{\mu_m} \, d\mu_m
\]

\[
= \int_{-\infty}^{+\infty} \frac{1}{|\gamma_m - \mu_m|} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \frac{1}{\mu_m} \, d\mu_m
\]

\[
= \int_{-\infty}^{\gamma_m} \frac{1}{|\gamma_m - \mu_m|} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \frac{1}{\mu_m} \, d\mu_m
\]

\[
+ \int_{\gamma_m}^{\epsilon} \frac{1}{|\gamma_m - \mu_m|} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \frac{1}{\mu_m} \, d\mu_m
\]

\[
+ \int_{\epsilon}^{+\infty} \frac{1}{|\gamma_m - \mu_m|} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \frac{1}{\mu_m} \, d\mu_m.
\]

Let \( A = \max_{\mu_m \in [\gamma_m, \epsilon]} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \). The second term of the last expression is such that

\[
\int_{\gamma_m}^{\epsilon} \frac{1}{|\gamma_m - \mu_m|} \left( \frac{m-1}{m} \mu_m^2 - \frac{2H}{m} \mu_m + K - \frac{1}{m}H^2 \right) \frac{1}{\mu_m} \, d\mu_m
\]

\[
\geq \int_{\gamma_m}^{\epsilon} \frac{1}{|\gamma_m - \mu_m|} \frac{1}{A} \, d\mu_m
\]

\[
= \frac{1}{A} \int_{\gamma_m}^{\epsilon} \frac{1}{\mu_m - \gamma_m} \, d\mu_m
\]

\[
= \frac{1}{A} \left[ \log(\mu_m - \gamma_m) \right]_{\gamma_m}^{\epsilon} = +\infty,
\]

which also implies that

\[
\int_{S(\mu_m)} \frac{1}{|\gamma_m - \mu_m|} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right) \frac{1}{\mu_m} \, d\mu_m = +\infty.
\]
For the same argument, one can also see that
\[
\pi(\gamma) \propto \int_{S(\mu_1)} \frac{1}{|\gamma_1 - \mu_1|} \int_{S(\mu_2)} \frac{1}{|\gamma_2 - \mu_2|} \cdots \int_{S(\mu_m)} \frac{1}{|\gamma_m - \mu_m| - 1} \left[ \prod_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right] d\mu = +\infty ,
\]
It follows that
\[
\pi(\gamma|U) \propto \prod_{i=1}^{m} \left( \prod_{j=1}^{n_i} \left( c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_{ij}}|\gamma_i) \right) \right) \pi(\gamma) = +\infty .
\]

A similar argument can be used to prove the following result:

**Proposition 2:** If \( \pi_{\sigma^2}(\sigma_i^2) \propto 1 \), for \( i = 1, \ldots, m \), and \( \pi_{\delta^2}(\delta^2) \propto 1 \), \( \pi_{\lambda}(\lambda) \propto 1 \), the posterior distribution \( \gamma|U \) is improper for any choice of copula densities \( c_i(\cdot|\gamma_i) \) and independently of the sample size.

**Proof:** As before, one needs to show that the following pseudo marginal posterior distribution of \( \gamma \) has not a finite integral:
\[
\pi(\gamma|U) = \int_{S(\mu_1)} \int_{S(\sigma^2)} \int_{S(\delta^2)} \int_{S(\lambda)} \pi(\gamma, \mu, \sigma^2, \lambda, \delta^2|U)d\lambda d\delta^2 d\sigma^2 d\mu
\]
\[
\propto \prod_{i=1}^{m} \left( \prod_{j=1}^{n_i} \left( c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_{ij}}|\gamma_i) \right) \right) \pi(\gamma)
\]
We use the same notation as in Proposition 1, and assume \( m > 3 \). With a slight modification of the proof of Proposition 1, we obtain
\[
\pi(\mu) = \int_{S(\delta^2)} \int_{S(\lambda)} \pi(\mu|\lambda, \delta^2) \pi(\delta^2) d\lambda d\delta^2 \propto \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-3}{2}} ,
\]
and
\[
\pi(\gamma) = \int_{S(\sigma_1^2)} \cdots \int_{S(\sigma_m^2)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \pi(\gamma|\mu, \sigma^2) \pi(\mu) \pi(\sigma^2) d\mu d\sigma^2
\]
\[
\propto \int_{S(\sigma_1^2)} \cdots \int_{S(\sigma_m^2)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \prod_{i=1}^{m} \left[ (2\pi\sigma_i^2)^{\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_i^2} (\gamma_i - \mu_i)^2 \right) \right] \frac{1}{\left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{\frac{m-3}{2}}} d\mu d\sigma^2
\]
\[
\propto \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \left( \sum_{i=1}^{m} (\mu_i - \bar{\mu})^2 \right)^{-\frac{m-3}{2}} \prod_{i=1}^{m} \left[ \int_{S(\sigma_i^2)} \left( \frac{1}{\sigma_i^2} \right)^{\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_i^2} (\gamma_i - \mu_i)^2 \right) \right] d\mu
\]
However, for all \( i = 1, \ldots, m \), the integral with respect to \( \sigma_i^2 \) is not finite and this implies again that
\[
\pi(\gamma|U) \propto \prod_{i=1}^{m} \left( \prod_{j=1}^{n_i} \left( c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_{ij}}|\gamma_i) \right) \right) \pi(\gamma) = +\infty .
\]

\(^1\)When \( m \leq 3 \) the Theorem is trivially true since \( \pi(\mu) \) itself is not defined.