MDPI

Article

Bayesian Hierarchical Copula Models with a Dirichlet–Laplace Prior

Paolo Onorati [†] and Brunero Liseo *,[†]

The Department MEMOTEF, Sapienza University of Rome, 00161 Roma, Italy
* Correspondence: brunero.liseo@uniroma1.it
† These authors contributed equally to this work.

Abstract: We discuss a Bayesian hierarchical copula model for clusters of financial time series. A similar approach has been developed in recent paper. 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 structures among different clusters. The performance of the proposed model is presented via simulations and a real data analysis.

Keywords: global–local shrinkage prior; MCMC; model-based clustering; GARCH

1. Introduction

There is a large body of literature with respect to hierarchical model settings. The concept to pull the mean of a single group towards the mean across different groups can be found at least in Kelley [1]. Tiao and Tan [2] and Hill [3] 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 [4] discuss and resolve a paradox that arises with the use of Jeffreys' prior. The foundation for the Bayesian hierarchical linear model is established in Lindley and Smith [5]. More recently, Gelman [6] discuss a review on prior distributions for variance parameters in the hierarchical model.

More recently, Zhuang et al. [7] introduced a hierarchical model in a copula framework; they suggest using, for the variance parameters of two different priors, (i) the standard improper prior for scale parameters, which is proportional to σ^{-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 A); in the second case, the use of small parameters of the inverse Gamma priors simply hides the problem without actually solving it; see for example Berger [8].

Hobert and Casella [9] also provide another review on the effect of improper priors in the 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 for the vector of coefficients. Several different global–local shrinkage families of priors have been proposed: Park and Casella [10] and Hans [11] discuss the Bayesian LASSO; Carvalho et al. [12] introduce the *Horseshoe prior*, Armagan et al. [13] propose a Generalized Double Pareto prior. Here, we will use a Dirichlet–Laplace prior, proposed in Bhattacharya et al. [14], with a slight modification; while in a regression framework, it is natural to adopt a prior that 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 follows: The next section is devoted to illustrating the statistical model and the prior distribution, highlighting the differences with the approach described in Zhuang et al. [7]; 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 dataset discussed in Zhuang et al. [7] and compare the results of the two approaches. We conclude with another illustration of the model in the problem of clustering financial time series.

2. Materials and Methods

2.1. The Statistical Model

2.1.1. 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 applications in statistical inferences and a review on the most popular approaches can be found in Hofert et al. [15]. 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 dimensions, we will concentrate on the use of the most popular elliptical versions, namely the Gaussian and Student's t copulae. Since the main objective of the paper is the clusterization of the dependence structure, for the sake of simplicity and without a loss of generality , we will assume that all marginal distributions are known or, equivalently, their parameters have been previously estimated. In this way, we can directly work with the transformed variables: $U_j = F_{X_j}(x_j)$, $j \in \{1, \ldots, n\}$.

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:

$$(U_{1i}, U_{2i}, \ldots, U_{d_i i})|\psi_i \sim c_i(\cdot|\psi_i) \qquad i \in \{1, \ldots, m\}$$

where m denotes the number of groups or clusters. Set the following:

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

and assume the following.

$$\gamma_i|\xi, \tau, \alpha_i \overset{ind}{\sim} Laplace(\xi, \tau\alpha_i) \qquad i \in \{1, \ldots, m\},$$

$$\tau \sim Gamma\left(ma, \frac{1}{2}\right),$$

$$(\alpha_1, \alpha_2, \ldots, \alpha_m) \sim Dirichlet(a, a, \ldots, a),$$

$$\xi \sim Logistic(0, 1).$$

In the previous expressions, b_i and B_i, respectively, denote the lower and the upper bound of the parameter space of the corresponding ψ_i, and γ_i is the mapping of ψ_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 Kendall's Tau, for which its range of values has been restricted to $(0, 1)$ for the Clayton, Joe, and Gumbel copulae, while it is set to $(-1, 1)$ for the Frank copula. In the elliptical case, the Gaussian copula is parametrized in terms of the correlation coefficient ρ, which ranges in $(-1, 1)$; finally, Student's t copula has the additional parameter ν, and that is the number of degrees of freedom: A discrete uniform

prior on $\{1, 2, \ldots, 35\}$ has been used here. When 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 $\boldsymbol{U}$ 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 $(\gamma, \xi, \boldsymbol{\alpha}, \tau)$ is then described as follows:

$$p(\gamma, \xi, \boldsymbol{\alpha}, \tau | U) \propto \prod_{i=1}^{m} \left[\prod_{j=1}^{n_i} \left[c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_i ij} | \gamma_i) \right] p(\gamma_i | \xi, \tau, \alpha_i) \right] p(\xi) p(\tau) p(\boldsymbol{\alpha}),$$

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

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. [14] with a minor modification for the updates of γ and the shrinkage location ξ. Following, Bhattacharya et al. [14], we introduce a vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_m) \in \mathbb{R}^m$ in order to have a latent variable representation of the γ prior; then, the following is obtained.

$$\gamma_i | \xi, \tau, \alpha_i, \beta_i \overset{ind}{\sim} Normal(\xi, \beta_i \tau^2 \alpha_i^2) \ \forall i \in \{1, \ldots, m\},$$

$$\beta_i \overset{iid}{\sim} Exp\left(\frac{1}{2}\right) \quad i \in \{1, \ldots, m\}.$$

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

1. Update $\gamma | \xi, \tau, \boldsymbol{\alpha}, \boldsymbol{\beta}$:

   (a) Sample $\tilde{\gamma}_i$ from a proposal $Cauchy(\gamma_{it}, \delta_\gamma)$ $\quad i \in \{1, \ldots, m\}$;

   (b) Set $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, \ldots, \tilde{\gamma}_m)$ and compute the following.

   $$q = \frac{\prod_{i=1}^{m} \left[\prod_{j=1}^{n_i} \left[c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_i ij} | \tilde{\gamma}_i) \right] p(\tilde{\gamma}_i | \xi_t, \tau_t, \alpha_{it}, \beta_{it}) \right]}{\prod_{i=1}^{m} \left[\prod_{j=1}^{n_i} \left[c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_i ij} | \gamma_{it}) \right] p(\gamma_{it} | \xi_t, \tau_t, \alpha_{it}, \beta_{it}) \right]}$$

   (c) Sample $u \sim U(0, 1)$,

   (d) Set $\gamma_{t+1} = \tilde{\gamma}$ if $u \leq q$; otherwise, $\gamma_{t+1} = \gamma_t$.

2. Update $\xi | \gamma, \tau, \boldsymbol{\alpha}, \boldsymbol{\beta}$:

   (a) Sample $\tilde{\xi}$ from a proposal $Cauchy(\xi_t, \delta_\xi)$;

   (b) Compute the following.

   $$q = \frac{\prod_{i=1}^{m} \left[p(\gamma_{i\,t+1} | \tilde{\xi}, \tau_t, \alpha_{it}, \beta_{it}) \right] p(\tilde{\xi})}{\prod_{i=1}^{m} \left[p(\gamma_{i\,t+1} | \xi_t, \tau_t, \alpha_{it}, \beta_{it}) \right] p(\xi_t)}$$

   (c) Sample $u \sim U(0, 1)$;

   (d) Set $\xi_{t+1} = \tilde{\xi}$ if $u \leq q$; otherwise, $\xi_{t+1} = \xi_t$.

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

4. Update $\boldsymbol{\alpha} | \gamma, \xi, \tau, \boldsymbol{\beta}$: sample $\tilde{\alpha}_i \sim GIG(0, 1, 2|\gamma_{i\,t+1} - \xi_{t+1}|)$ $\quad i \in \{1, \ldots, m\}$, and set the following.

   $$\alpha_{i\,t+1} = \frac{\tilde{\alpha}_i}{\sum_{j=1}^{m} \tilde{\alpha}_j} \quad i \in \{1, \ldots, m\}$$

5. Update $\beta_i | \gamma, \xi, \tau, \alpha \quad i \in \{1, \ldots, m\}$: sample $\tilde{\beta}_i \sim IG(\frac{\tau_{t+1}\alpha_{i\,t+1}}{|\gamma_{i\,t+1} - \tilde{\xi}_{t+1}|}, 1)$ and set the following.

$$\beta_{i\,t+1} = \frac{1}{\tilde{\beta}_i} \qquad i \in \{1, \ldots, m\}.$$

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

$$f(x) \propto x^{p-1} \exp\left(-\frac{1}{2}ax - \frac{1}{2}\frac{b}{x}\right).$$

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

In the case of the Student's 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 $\nu_i | \gamma, \xi, \tau, \alpha, \beta \ \forall i \in \{1, \ldots, m\}$:

    (a)   Sample $\tilde{\nu}$ from discrete uniform distribution in $\{1, 2, \ldots, 35\}$;
    (b)   Compute the following.

$$q = \frac{\prod_{j=1}^{n_i} \left[c(U_{1ij}, U_{2ij}, \ldots, U_{d_i ij} | \gamma_{i\,t+1}, \tilde{\nu}) \right]}{\prod_{j=1}^{n_i} \left[c(U_{1ij}, U_{2ij}, \ldots, U_{d_i ij} | \gamma_{i\,t+1}, \nu_{it}) \right]}$$

    (c)   Sample $u \sim U(0, 1)$;
    (d)   Set $\nu_{i,t+1} = \tilde{\nu}$ if $u \leq q$; otherwise, $\nu_{i,t+1} = \nu_{it}$.

2.1.2. Prior Distribution of ξ

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

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

however $\gamma_i = \log\left(\frac{\psi_i - b_i}{B_i - \psi_i}\right)$, so otherwise is the case.

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

Therefore, given ξ, the median of ψ_i is $Y_i = (B_i e^\xi + b_i)/(1 + e^\xi) \ \forall i \in \{1, \ldots, m\}$. 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 $i \in \{1, \ldots, m\}$ implies a standard logistic density for ξ.

2.1.3. Previous Work

Apart form the prior specification, the model described in previous sections is the one proposed by Zhuang et al. [7]. We restrict our discussion to the case where each copula expression has one parameter only. Their prior can be stated as follows.

$$\gamma_i | \mu_i, \sigma_i^2 \stackrel{ind}{\sim} N(\mu_i, \sigma_i^2) \qquad i \in \{1, \ldots, m\},$$

$$\mu_i | \lambda, \delta^2 \stackrel{iid}{\sim} N(\lambda, \delta^2) \qquad i \in \{1, \ldots, m\},$$

$$\sigma_i^2 \stackrel{iid}{\sim} \pi_{\sigma^2}(\cdot) \qquad i \in \{1, \ldots, m\},$$

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

There is no unique choice for the distributions of $(\sigma^2, \lambda, \delta)$, although the authors suggest using weakly informative priors, for example, inverse gamma densities with small hyperparameters values or, as an alternative, an objective prior: for example, an improper uniform prior. However, one can prove that, in the second case, the posterior distribution cannot be proper no matter what the sample size is. We show this result in Appendix A. When the posterior distribution is improper, the resulting summary statistics are meaningless. In fact, the Markov Chain implied by the MCMC does not have a limiting distribution so the Ergodic theorem does not hold and the posterior is completely useless. Moreover, even the first solution is not feasible. In fact, when an improper prior produces an improper posterior, using a vague proper prior can typically hide—not solve—the problem. In these cases, in fact, as shown in Berger [8] (p. 398), the use of a vague prior approximating an improper prior typically concentrates the posterior mass on some boundary of the parameter space.

3. Results

3.1. 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's t copula with an equi-correlation matrix and set the number of groups m equal to five. We repeat the procedure 100 times; at iteration j for the i-th group, we sample the true value γ_{ij}^T from a standard normal distribution, the degrees of freedom ν_{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 the following:

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

and compute the standard errors.

$$\widehat{SE}_{ij}^{\text{mle}} = \left(\gamma_{ij}^T - \hat{\gamma}_{ij}^{\text{mle}} \right)^2 \quad i \in \{1, \ldots, 5\} \; .$$

In a Bayesian framework, we use the posterior mean as a point estimate, obtained from the use of the MCMC algorithm described above. We ran six independent chains of 2.5×10^5 scans, discarded the first 5×10^4 as a burn-in, and finally computed the $\hat{\gamma}_{ij}^{\text{Bay}}$ via the sample mean of simulation outputs for all $i \in \{1, \ldots, 5\}$. As a tuning parameters, we set $\delta_\gamma = 10^{-3}$ and $\delta_{\tilde{\zeta}} = 10^{-1}$. Then, we compute the following.

$$\widehat{SE}_{ij}^{\text{Bay}} = \left(\gamma_{ij}^T - \hat{\gamma}_{ij}^{\text{Bay}} \right)^2 \quad i \in \{1, \ldots, 5\} \; .$$

Comparison are performed in terms of the corresponding mean square errors.

$$\widehat{MSE}_i^{\text{mle}} = \frac{1}{100} \sum_{j=1}^{100} \widehat{SE}_{ij}^{\text{mle}} \; , \qquad \widehat{MSE}_i^{\text{Bay}} \frac{1}{100} \sum_{j=1}^{100} \widehat{SE}_{ij}^{\text{Bay}} \; ,$$

Table 1 reports values $\widehat{MSE}_i^{\text{mle}}$ against $\widehat{MSE}_i^{\text{Bay}}$ for all groups based on 100 simulations.

Table 1. MSE of the proposed Bayesian Hierarchical Model and of the likelihood-based one.

	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

3.2. 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. [7] and we include it for comparative purposes; to this end, we quantify the goodness of fit of the model using a predictive approach based on the conditional version of the Widely Applicable Information Criterion, WAIC, in a hierarchical setting, as discussed in Millar [16]. The second one deals with clustering financial time series.

3.2.1. Column Vertebral Data

We apply our model to the Column Vertebral Data, available at the 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 the degree of spondylolisthesis (DS). As in Zhuang et al. [7], 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. Computations were performed 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. [7], 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 three groups of subjects. We run six independent chains of 2.5×10^6 simulations and discard the first 5×10^5. We also set $\delta_\gamma = 10^{-3}$ and $\delta_\xi = 10^{-1}$. We did not report any convergence issues, and the multiple Gelman–Rubin test scores for each of the six implemented models Gelman [17] were very close to the optimal value 1. In terms of the goodness of fit, we have computed the WAIC index for all six models. Our findings is that the most significant relation is the one between PI and PT. Table 3 compares

the results of Zhuang et al. [7] (model A) with our ones (model B). The main difference between the results obtained with the two methods is related to the posterior uncertainty quantification. Credible intervals obtained with model B are systemically larger than those obtianed with model A. Our feeling is that it depends on the fact that results in model A are obtained by running a chain where some hyperparameters are fixed to some estimated values, as explained in Zhuang et al. [7]. Fixing values of the hyperparameters eliminates a critical source of variation, inducing shrinkage in credible intervals size.

For the ease of comparisons, we follow Zhuang et al. [7] and report the results not in terms of parameter γ but rather according the natural parameter of each copula, that is, ρ for the Gaussian copula and θ for the Archimedean ones.

3.2.2. Financial Data Application

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

As a risk measure, we consider the so-called tail index, which measures the strength of dependence between two variables when one of them takes extremely low values. Following De Luca and Zuccolotto [18], 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 λ_L of (Y_1, Y_2) is defined as follow:

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

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 λ_L, we use the empirical estimator discussed in [19]:

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

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

$$d(Y_1, Y_2) = 1 - \lambda_L(Y_1, Y_2) \ ,$$

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. [20] proposed a conditioned correlation coefficient estimated using a nonparametric approach; Fuchs et al. [21] analyzed dissimilarity measure applicable to a multivariate lower tail coefficient.

Table 3. Fitted parameters of copulae.

Group	Features	Copula	Posterior Mean	Model A Posterior s.d.	Posterior CI (95%)	Posterior Mean	Model B Posterior s.d.	Posterior CI (95%)
	PI vs. PT	Gaussian	0.696	0.046	(0.599, 0.775)	0.632	0.073	(0.469, 0.751)
	PI vs. SS	Gaussian	0.726	0.040	(0.633, 0.793)	0.680	0.076	(0.506, 0.789)
Disk Hernia	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)
	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)
Spondylolisthesis	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)
	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)
Healthy	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	(−0.000, 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)

We consider the "S&P 500 Full Dataset" available at Kaggle: It contains more relevant information for the components of S&P 500. We take the daily closing prices from 5 June 2000 to 5 June 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 log-differences and filter data by fitting; for each time series, an ARMA(1,1)GJR-GARCH(1,1) model with Student's t innovations was used; then, we extracted residuals and transformed them via the fitted cumulative distribution function in order to obtain pseudo-data. Computations were performed using the CRAN package `rugarch`. Hence, we compute the empirical estimator of the lower tail coefficient for any possible pair and the dissimilarity measure associated and use them to feed the clustering algorithm. Due to computational complexities, we used the coarsest partition under the constraint that the largest group must have at most 10 components. We obtained 30 groups with dimensions of more than one and discarded instruments that belong to groups with only one component. The final number of instruments was thus reduced to 93.

We ran the MCMC algorithm described above for the 30 clusters, performing 12 independent chains of 10^5 scans and discarding the first 1.5×10^4 as they burned in. Tuning parameters were set to $\delta_\gamma = 10^{-6}$, $\delta_\xi = 10^{-3}$. Moreover, in this example, we did not report any convergence issues, and the Gelman–Rubin test score was 1.02. For each scan and for any group, we compute the lower tail coefficient via the following 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's t random variable with ν degrees of freedom. The copula used in this example was a Student's t copula with an equi-correlation matrix: As a consequence, we obtained a single value for the lower tail coefficient for each cluster. Table 4 reports the results for each pair that belongs to the same group. Finally, we report the estimation results.

Table 4. Posterior distributions for lower tail coefficients.

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)

Table 4. *Cont.*

Group	Components	Posterior Mean	Posterior s.d.	Posterior CI (95%)
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 NSC UNP	0.3395	0.0674	(0.2672, 0.4535)
14	T VZ	0.3338	0.0699	(0.2368, 0.4509)
15	CAH MCK	0.3337	0.0691	(0.2414, 0.4401)
16	BAC C 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 MTB PNC RF TFC USB	0.2294	0.0666	(0.1526, 0.3273)
26	ATO EVRG	0.2201	0.0692	(0.1256, 0.3412)

Table 4. *Cont.*

Group	Components	Posterior Mean	Posterior s.d.	Posterior CI (95%)
27	ETR NEE PEG	0.1923	0.0652	(0.1175, 0.2953)
28	AEE AEP DTE DUK ED ES LNT WEC XEL	0.1768	0.0633	(0.1174, 0.2855)
29	ARE BXP DRE FRT KIM MAA PLD REG SPG	0.1522	0.0605	(0.0874, 0.2439)
30	EW SYK	0.0008	0.0011	(0.0000, 0.0028)

4. Conclusions

We discussed and improved a fully Bayesian analysis for a hierarchical copula model proposed in Zhuang et al. [7]. We 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 information coming from the data. Our prior belongs to the large family of globa–local shrinkage densities, with an extra stage in the hierarchy, due to the absence of a significant shrinkage value; we 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 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 the Student's *t* copula.

Abbreviations

The following abbreviations are used in this manuscript:

S&P   Standard and Poor's 500 stock exchange index;
mle or MLE Maximum likelihood estimator;
MSE   Mean squared error;
MCMC   Markov chain Monte Carlo.

Appendix A

Here, we show that the prior proposed in Zhuang et al. [7] leads to an improper posterior.

The statistical model consists of m d-dimensional copulae governing different sets of observations.

$$\left(U_{1i}, U_{2i}, \ldots, U_{d_i i}\right)|\theta_i \sim c_i(\cdot|\theta_i) \qquad i \in \{1, \ldots, m\}.$$

Let $\gamma_i = \eta_i g_i(\theta_i)$; here, η_i is a scaling parameter that can be considered known. One-to-one mapping functions $g_i(\cdot)$ are needed to put all dependence parameters on the real line. Zhuang et al. [7] made the following assumptions.

$$\gamma_i|\mu_i, \sigma_i^2 \overset{ind}{\sim} N(\mu_i, \sigma_i^2) \qquad i \in \{1, \ldots, m\};$$

$$\mu_i|\lambda, \delta^2 \overset{iid}{\sim} N(\lambda, \delta^2) \qquad i \in \{1, \ldots, m\}.$$

Hyper-parameters σ_i's, λ, and δ^2 are given a suitable prior distribution. For the moment, we do not specify the priors and set the following.

$$\sigma_i^2 \overset{iid}{\sim} \pi_{\sigma^2}(\cdot) \qquad i \in \{1, \ldots, m\}.$$

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

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 ω.

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 A1. *If $\pi_{\sigma_i^2}(\sigma_i^2) \propto \sigma_i^{-2}$, for $i \in \{1, \ldots, m\}$, 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 γ is not integrable:

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

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

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

$$\pi(\gamma|U) \propto \int_{S(\boldsymbol{\mu})} \int_{s(\sigma^2)} \int_{S(\delta^2)} \int_{S(\lambda)} \prod_{i=1}^{m} \Big[\prod_{j=1}^{n_i} \Big(c_i(U_{1ij}, U_{2ij}, \ldots U_{d_iij}|\gamma_i) \Big) \Big] \times$$

$$\pi(\gamma|\boldsymbol{\mu}, \sigma^2)\pi(\boldsymbol{\mu}|\lambda, \delta^2)\pi(\sigma^2)\pi(\lambda)\pi(\delta^2)d\lambda d\delta^2 d\sigma^2 d\boldsymbol{\mu}$$

$$\propto \prod_{i=1}^{m} \Big[\prod_{j=1}^{n_i} \Big(c_i(U_{1ij}, U_{2ij}, \ldots U_{d_iij}|\gamma_i) \Big) \Big] \int_{S(\boldsymbol{\mu})} \int_{s(\sigma^2)} \pi(\gamma|\boldsymbol{\mu}, \sigma^2)\pi(\sigma^2) \times$$

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

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

with

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

and

$$\pi(\boldsymbol{\mu}) = \int_{S(\delta^2)} \int_{S(\lambda)} \pi(\boldsymbol{\mu}|\lambda, \delta^2)\pi(\lambda)\pi(\delta^2)d\lambda d\delta^2 \ .$$

Consider only the following:

$$\pi(\boldsymbol{\mu}) = \int_{0}^{\infty} \int_{-\infty}^{\infty} \pi(\boldsymbol{\mu}|\lambda, \delta^2)\pi(\lambda)\pi(\delta^2)d\lambda d\delta^2$$

$$\propto \int_{0}^{+\infty} \int_{-\infty}^{+\infty} (2\pi\delta^2)^{-\frac{m}{2}} \exp\Big(-\frac{1}{2\delta^2} \sum_{i=1}^{m}(\mu_i - \lambda)^2 \Big) \frac{1}{\delta^2} d\lambda d\delta^2$$

$$\propto \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m}{2}+1} \int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2\delta^2} \sum_{i=1}^{m}(\mu_i^2 - 2\lambda\mu_i + \lambda^2) \Big) d\lambda d\delta^2$$

$$= \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m}{2}+1} \int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2\delta^2} \Big(\sum_{i=1}^{m} \mu_i^2 - 2\lambda \sum_{i=1}^{m} \mu_i + m\lambda^2 \Big) \Big) d\lambda d\delta^2 \ ;$$

and set $\bar{\mu} = \frac{1}{m} \sum_{i=1}^{m} \mu_i$; then, we obtain the following.

$$\pi(\boldsymbol{\mu}) \propto \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m}{2}+1} \exp\Big(-\frac{1}{2\delta^2} \sum_{i=1}^{m} \mu_i^2 \Big) \int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2\frac{\delta^2}{m}} \big(\lambda^2 - 2\lambda\bar{\mu} + \bar{\mu}^2 - \bar{\mu}^2 \big) \Big) d\lambda d\delta^2 =$$

$$= \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m}{2}+1} \exp\Big(-\frac{1}{2\delta^2} \Big(\sum_{i=1}^{m} \mu_i^2 - m\bar{\mu}^2 \Big) \Big) \int_{-\infty}^{+\infty} \exp\Big(-\frac{1}{2\frac{\delta^2}{m}} (\lambda - \bar{\mu}^2) \Big) d\lambda d\delta^2 =$$

$$= \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m}{2}+1} \exp\Big(-\frac{1}{2\delta^2} m \Big(\frac{1}{m} \sum_{i=1}^{m} \mu_i^2 - \bar{\mu}^2 \Big) \Big) \sqrt{2\pi \frac{\delta^2}{m}} d\delta^2 \propto$$

$$\propto \int_{0}^{+\infty} \Big(\frac{1}{\delta^2} \Big)^{\frac{m-1}{2}+1} \exp\Big(-\frac{1}{2\delta^2} \sum_{i=1}^{m}(\mu_i - \bar{\mu})^2 \Big) d\delta^2 \ ,$$

For any choice of $m > 1$, $\pi(\boldsymbol{\mu})$ can be written as follows.

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

Now, we compute the following.

$$\pi(\boldsymbol{\gamma}) = \int_{S(\sigma_1)} \cdots \int_{S(\sigma_m)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \pi(\boldsymbol{\gamma}|\boldsymbol{\mu}, \boldsymbol{\sigma^2})\pi(\boldsymbol{\mu})\pi(\boldsymbol{\sigma^2})d\boldsymbol{\sigma^2}d\boldsymbol{\mu}$$

$$\propto \int_{S(\sigma_1)} \cdots \int_{S(\sigma_m)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \prod_{i=1}^{m} \left[(2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_i^2}(\gamma_i - \mu_i)^2 \right) \right] \times$$

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

$$\propto \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \left(\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} \prod_{i=1}^{m} \left[\int_{S(\sigma_i^2)} \left(\frac{1}{\sigma_i^2} \right)^{\frac{3}{2}} \exp\left(-\frac{1}{\sigma_i^2}\frac{(\gamma_i - \mu_i)^2}{2} \right) d\sigma_i^2 \right] d\boldsymbol{\mu}$$

$$\propto \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \left(\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2 \right)^{-\frac{m-1}{2}} \prod_{i=1}^{m} \left((\gamma_i - \mu_i)^2 \right)^{-\frac{1}{2}} d\boldsymbol{\mu}$$

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

Notice that the following is the case:

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

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

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

and set $K = \sum_{i=1}^{m-1} \mu_i^2$ and $H = \sum_{i=1}^{m-1} \mu_i$: then, we obtain the following.

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

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

So $\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2$ is a convex parabolic function of μ_m, and by the Weierstrass theorem, a global maximum exists for all bounded and closed sets. By integrating μ_m, one obtains the following.

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

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

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

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

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

Let $A = \max\limits_{\mu_m \in [\gamma_m, \epsilon]} \left(\frac{m-1}{m}\mu_m^2 - \frac{2H}{m}\mu_m + K - \frac{1}{m}H^2 \right)$. The second term of the last expression is as follows:

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

$$\geq \int_{\gamma_m}^{\epsilon} \frac{1}{|\gamma_m - \mu_m|} \frac{1}{A^{\frac{m-1}{2}}} d\mu_m$$

$$= \frac{1}{A^{\frac{m-1}{2}}} \int_{\gamma_m}^{\epsilon} \frac{1}{\mu_m - \gamma_m} d\mu_m$$

$$= \frac{1}{A^{\frac{m-1}{2}}} \Big[\log(\mu_m - \gamma_m) \Big]\Big|_{\gamma_m}^{\epsilon} = +\infty \ ,$$

which also implies the following.

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

For the same argument, one can also see that the following obtains.

$$\pi(\gamma) \propto \int_{S(\mu_1)} \frac{1}{|\gamma_1 - \mu_1|} \int_{S(\mu_2)} \frac{1}{|\gamma_2 - \mu_2|} \cdots \int_{S(\mu_m)} \frac{|\gamma_m - \mu_m|^{-1}}{\left[\sum_{i=1}^m (\mu_i - \bar{\mu})^2 \right]^{\frac{m-1}{2}}} d\boldsymbol{\mu} = +\infty \ ,$$

It follows that

$$\pi(\gamma|U) \propto \prod_{i=1}^m \Big[\prod_{j=1}^{n_i} \Big(c_i(U_{1ij}, U_{2ij}, \ldots U_{d_iij}|\gamma_i) \Big) \Big] \pi(\gamma) = +\infty \ .$$

$\square$

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

Proposition A2. *If $\pi_{\sigma_i^2}(\sigma_i^2) \propto 1$, for $i \in \{1, \ldots, m\}$, 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 is independent of the sample size.*

Proof. As before, one needs to show that the following pseudo-marginal posterior distribution of γ does not have a finite integral.

$$\pi(\gamma|U) = \int_{S(\boldsymbol{\mu})} \int_{S(\boldsymbol{\sigma^2})} \int_{S(\delta^2)} \int_{S(\lambda)} \pi(\gamma, \boldsymbol{\mu}, \boldsymbol{\sigma^2}, \lambda, \delta^2|U) d\lambda d\delta^2 d\sigma^2 d\boldsymbol{\mu}$$

$$\propto \prod_{i=1}^m \Big[\prod_{j=1}^{n_i} \Big(c_i(U_{1ij}, U_{2ij}, \ldots, U_{d_iij}|\gamma_i) \Big) \Big] \pi(\gamma)$$

We use the same notation as in Proposition 1 and assume $m > 3$ (when $m \leq 3$, the theorem is trivially true since $\pi(\boldsymbol{\mu})$ itself is not defined). With a slight modification in the proof of the proposition, we obtain the following:

$$\pi(\boldsymbol{\mu}) = \int_{S(\delta^2)} \int_{S(\lambda)} \pi(\boldsymbol{\mu}|\lambda, \delta^2)\pi(\lambda)\pi(\delta^2)d\lambda d\delta^2 \propto \Big(\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2\Big)^{-\frac{m-3}{2}},$$

and

$$\pi(\boldsymbol{\gamma}) = \int_{S(\sigma_1^2)} \cdots \int_{S(\sigma_m^2)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \pi(\boldsymbol{\gamma}|\boldsymbol{\mu}, \sigma^2)\pi(\boldsymbol{\mu})\pi(\sigma^2)d\boldsymbol{\mu}d\sigma^2$$

$$\propto \int_{S(\sigma_1^2)} \cdots \int_{S(\sigma_m^2)} \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \prod_{i=1}^{m}\left[(2\pi\sigma_i^2)^{-\frac{1}{2}}\exp\Big(-\frac{1}{2\sigma_i^2}(\gamma_i - \mu_i)^2\Big)\right]$$

$$\frac{1}{\Big(\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2\Big)^{\frac{m-3}{2}}} d\boldsymbol{\mu}d\sigma^2$$

$$\propto \int_{S(\mu_1)} \cdots \int_{S(\mu_m)} \Big(\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2\Big)^{-\frac{m-3}{2}} \prod_{i=1}^{m}\left[\int_{S(\sigma_i^2)}\Big(\frac{1}{\sigma_i^2}\Big)^{\frac{1}{2}}\exp\Big(-\frac{1}{\sigma_i^2}\frac{(\gamma_i - \mu_i)^2}{2}\Big)d\sigma_i^2\right]d\boldsymbol{\mu}$$

However, for all $i \in \{1, \ldots, m\}$, the integral with respect to σ_i^2 is not finite, and this again implies the following.

$$\pi(\boldsymbol{\gamma}|U) \propto \prod_{i=1}^{m}\Big[\prod_{j=1}^{n_i}\Big(c_i(U_{1ij}, U_{2ij}, \ldots U_{d_iij}|\gamma_i)\Big)\Big]\pi(\boldsymbol{\gamma}) = +\infty .$$

$\square$

References

1. Kelley, L.T. *The Interpretation of Educational Measurement*; Measurement and Adjustment Series; WorldBook Company: Yonkers-on-Hudson, NY, USA, 1927.
2. Tiao, G.C.; Tan, W.Y. Bayesian Analysis of Random-Effect Models in the Analysis of Variance. i. Posterior Distribution of Variance-Components. *Biometrika* **1965**, *52*, 37–53. [CrossRef]
3. Hill, B.M. Inference About Variance Components in the One-Way Model. *J. Am. Stat. Assoc.* **1965**, *60*, 806–825. [CrossRef]
4. Stone, M.; Springer, B.G.F. A Paradox Involving Quasi Prior Distributions. *Biometrika* **1965**, *52*, 623–627. [CrossRef]
5. Lindley, D.V.; Smith, A.F.M. Bayes Estimates for the Linear Model. *J. R. Stat. Soc. Ser. B (Methodol.)* **1972**, *34*, 1–41. [CrossRef]
6. Gelman, A. Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper. *Bayesian Anal.* **2006**, *1*, 515–534. [CrossRef]
7. Zhuang, H.; Diao, L.; Yi, G.Y. A Bayesian Hierarchical Copula Model. *Electron. J. Stat.* **2020**, *14*, 4457–4488. [CrossRef]
8. Berger, J. The Case for Objective Bayesian Analysis. *Bayesian Anal.* **2006**, *1*, 385–402. [CrossRef]
9. Hobert, J.P.; Casella, G. The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. *J. Am. Stat. Assoc.* **1996**, *91*, 1461–1473. [CrossRef]
10. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]
11. Hans, C. Bayesian Lasso Regression. *Biometrika* **2009**, *96*, 835–845. [CrossRef]
12. Carvalho, C.M.; Polson, N.G.; Scott, J.G. The Horseshoe Estimator for Sparse Signals. *Biometrika* **2010**, *97*, 465–480. [CrossRef]
13. Armagan, A.; Dunson, D.; Lee, J. Generalized Double Pareto Shrinkage. *Stat. Sin.* **2013**, *23*, 119–143. [CrossRef] [PubMed]
14. Bhattacharya, A.; Pati, D.; Pillai, N.S.; Dunson, D.B. Dirichlet–Laplace Priors for Optimal Shrinkage. *J. Am. Stat. Assoc.* **2016**, *110*, 1479–1490. [CrossRef] [PubMed]
15. Hofert, M.; Kojadinovic, I.; Maechler, M.; Yan, J. *Elements of Copula Modeling with R*; Springer Use R! Series; Springer: New York, NY, USA, 2018.
16. Millar, R. Conditional vs. marginal estimation of the predictive loss of hierarchical models using WAIC and cross-validation. *Stat. Comput.* **2018**, *28*, 375–385. [CrossRef]
17. Gelman, A.; Rubin, D.B. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **1992**, *1*, 457–472.
18. De Luca, G.; Zuccolotto, P. A Tail Dependence-Based Dissimilarity Measure for Financial Time Series Clustering. *Adv. Data Anal. Classif.* **2011**, *5*, 323–340. [CrossRef]

19. Joe, H.; Smith, R.L.; Weissman, I. Bivariate Threshold Methods for Extremes. *J. R. Stat. Soc. Ser. B (Methodol.)* **1992**, *54*, 171–183. [CrossRef]

20. Durante, F.; Pappadà, R.; Torelli, N. Clustering of Financial Time Series in Risky Scenarios. *Adv. Data Anal. Classif.* **2014**, *8*, 359–376. [CrossRef]

21. Fuchs, S.; Di Lascio, F.M.L.; Durante, F. Dissimilarity Functions for Rank-Invariant Hierarchical Clustering of Continuous Variables. *Comput. Stat. Data Anal.* **2021**, *159*, 107201. [CrossRef]