

# Approximate Bayesian Inference in Semiparametric Copula Models

Clara Grazian\* and Brunero Liseo†

**Abstract.** We describe a simple method for making inference on a functional of a multivariate distribution, based on its copula representation. We make use of an approximate Bayesian Monte Carlo algorithm, where the proposed values of the functional of interest are weighted in terms of their Bayesian exponentially tilted empirical likelihood. This method is particularly useful when the “true” likelihood function associated with the working model is too costly to evaluate or when the working model is only partially specified.

**Keywords:** multivariate dependence, Bayesian exponentially tilted empirical likelihood, Spearman’s  $\rho$ , tail dependence coefficients, partially specified models.

## 1 Introduction

Copula models are widely used in multivariate data analysis. Major areas of application include econometrics (Huynh et al., 2014), geophysics (Schölzel and Friederichs, 2008), climate prediction (Scheffzik et al., 2013), actuarial science and finance (Cherubini et al., 2004), among the others. A copula allows a useful representation of the joint distribution of a random vector in two steps: the marginal distributions and a distribution function which captures the dependence among the vector components.

From a statistical perspective, whereas it is generally simple to produce reliable estimates of the parameters of the marginal distributions of the data, the problem of estimating the dependence structure, however it is modelled, is crucial and complex, especially in high dimensional situations. A list of important applications can be found in the recent monograph by Joe (2015).

In a frequentist approach to copula models, there are no broadly satisfactory methods for the joint estimation of marginal and copula parameters. The most popular method is the so called *Inference Functions for Margins* method, where the parameters of the marginal distributions are estimated first, and then pseudo-data are obtained by plugging-in the estimates of the marginal parameters. Then inference on the copula parameters is performed using the pseudo-data: this approach does not account for the uncertainty on the estimation of the marginal parameters. A nonparametric alternative may be found in Kauermann et al. (2013), where a penalized hierarchical B-splines approach is proposed.

---

\*Nuffield Department of Medicine, University of Oxford, Old Road Campus, Roosevelt Dr, Oxford OX3 7FZ, United Kingdom, [clara.grazian@ndm.ox.ac.uk](mailto:clara.grazian@ndm.ox.ac.uk)

†Brunero Liseo, MEMOTEF, Sapienza Università di Roma, via del Castro Laurenziano 9, 00161, Italy, [brunero.liseo@uniroma1.it](mailto:brunero.liseo@uniroma1.it)

The literature on Bayesian alternatives is still limited, although they show a great potential for inference in a number of cases, for example in the modelling of multivariate discrete data (Smith and Khaled, 2012) and of conditional copulae (Craiu and Sabeti, 2012); see Smith (2013) for a review on parametric methods and Wu et al. (2014) for a nonparametric approach. An instrumental use of copulas in Bayesian mixture estimation may be found in Burda and Prokhorov (2014). An alternative and more flexible copula construction is based on the so-called vines, where the joint dependence structure of a multivariate random vector is decomposed into several marginal and conditional bivariate copulas. A Bayesian use of this approach can be found in Min and Czado (2010) and Gruber and Czado (2015).

In this work we consider the general problem of estimating a functional of interest of a generic copula: practical illustrations will include the Spearman's  $\rho$  and tail dependence indices. Our method is based on the simulation of a posterior sample weighted in terms of the Bayesian exponentially tilted empirical likelihood (Schennach, 2005). A similar approach, in a frequentist fashion, has been proposed in Oh and Patton (2013), where a simulated method of moments is discussed for copula estimation. The main difference between the two approaches is that in our method the functional represents the actual quantity of interest and no assumption is made on the copula structure, while in Oh and Patton (2013) the functional is only instrumental, as a moment condition, to estimate the parameter of a given parametric copula.

As already stated, the central tool in our approach is the empirical likelihood (Owen, 2001); we adopt an approximate Bayesian approach based on the use of a pseudo-likelihood, along the lines of Mengersen et al. (2013). We use a partially specified model where the prior distribution is explicitly elicited only on the quantity of interest. Its approximate posterior distribution is obtained via the use of the Bayesian exponentially tilted empirical likelihood approximation of the marginal likelihood of the quantity of interest, illustrated in Schennach (2005). This approximation of the true "unknown" likelihood function hopefully reduces the potential bias for incorrect distributional assumptions, very hard to check in complex dependence modeling. Our approach can be adapted both to parametric and nonparametric modeling of the marginal distributions.

A brief review on copula models and empirical likelihood methods is given in Section 2. The method used to approximate the posterior distribution for a functional of the copula is presented in Section 3 and its asymptotic justification is discussed in Section 4. The rest of the paper is devoted to the illustration of the empirical behavior of the proposal on simulated and real data sets; in particular, Section 5 will apply the method for the Bayesian estimation of the Spearman's  $\rho$  in a bivariate setting, while Section 6 is focused on tail dependence coefficients; multivariate extensions are considered in Section 7. Finally, Section 8 contains a real financial application. A discussion concludes the work.

## 2 Preliminaries: Copulae and Empirical Likelihood

A copula represents an alternative way of writing the joint distribution of a random vector  $X = (X_1, \dots, X_d)$ . Given a  $d$ -variate cumulative distribution function  $F$  which

depends on some parameter  $\theta$ , it is possible to show (Sklar, 2010) that there always exists a  $d$ -variate function  $C_\theta : [0, 1]^d \rightarrow [0, 1]$ , such that

$$F(x_1, \dots, x_d; \lambda, \theta) = C_\theta(F_1(x_1; \lambda_1), \dots, F_d(x_d; \lambda_d)),$$

where  $\lambda = (\lambda_1, \dots, \lambda_d)$  is a vector of parameters,  $F_j$  is the marginal distribution of  $X_j$  depending on a parameter vector  $\lambda_j$ . In other terms, the copula  $C$  is a distribution function with uniform margins on  $[0, 1]$ : it binds together the univariate  $F_1, F_2, \dots, F_d$  in order to produce the  $d$ -variate distribution  $F$ . The copula  $C$  does not depend on the marginal distributions, and it accounts for potential dependence among the components of the random vector  $X$ . In the continuous case, the density of a random vector  $(X_1, \dots, X_d)$  has a unique copula representation given by

$$f(x|\lambda, \theta) = c(u; \theta) \prod_{j=1}^d f_j(x_j|\lambda_j), \quad (1)$$

where  $u = (u_1, \dots, u_d) = (F_1(x_1; \lambda_1), \dots, F_d(x_d; \lambda_d))$ ,  $c(u; \theta)$  is the derivative of  $C_\theta$  and  $\theta$  and  $(\lambda_1, \dots, \lambda_d)$  are the parameter of the copula and of the marginal density functions respectively. Given a prior  $\pi(\theta, \lambda_1, \dots, \lambda_d)$  and a sample of size  $n$  of independent multivariate observations  $(x_{i1}, \dots, x_{id})$  for  $i = 1, \dots, n$ , the resulting posterior distribution for the parameter vector is

$$\pi(\theta, \lambda|x) \propto \pi(\theta, \lambda) \prod_{i=1}^n \left[ c(u_i; \theta) \prod_{j=1}^d f(x_{ij}; \lambda_j) \right].$$

Notice that the likelihood function is not separable in  $\lambda_1, \dots, \lambda_d$  and  $\theta$  because the  $u_i$ 's depend on the marginal parameter  $\lambda$ .

In the parametric case, frequentist methods of estimation are generally based either on the simultaneous maximization of the likelihood function in  $\theta$  and  $\lambda$  or on the so-called method of inference functions for margins (IFM) (Joe, 2015): here a maximum likelihood estimate of  $\lambda$  is obtained using only the second factor of (1); then the estimate is plugged into the first factor and an estimate of  $\theta$  is based on pseudo-data  $\hat{u}_{ij} = F(x_{ij}; \hat{\lambda}_j)$ . The two methods are not equivalent in general (Choroś et al., 2010).

The first-step estimation may be performed both parametrically and nonparametrically: Genest et al. (1995) propose a semiparametric approach where nonparametric estimates are contemplated for the marginals and a specific copula function is used. The proposed estimator is shown to be consistent and asymptotically normal.

It is possible to modify the two-step procedure of Joe (2005) within a Bayesian framework, where the joint posterior distribution  $\pi(\theta, \lambda_1, \dots, \lambda_d|x)$  is evaluated through a Monte Carlo Markov chain (MCMC) algorithm, with  $\theta$  and  $(\lambda_1, \dots, \lambda_d)$  generated separately in a Gibbs sampling scheme; see Pitt et al. (2006) for a discussion. Smith (2013) provides a review on sampling schemes and possible prior distributions, in particular in the case of Gaussian copula models, both in a continuous and in a discrete setting.

A Bayesian nonparametric approach is followed by Wu et al. (2014), who model and estimate only the copula density function by using infinite mixture models and treat the

marginals as given. In particular, they focus on a mixture of multivariate skew-normal copulas, in order to circumvent the symmetry limitation of the Gaussian copula and to preserve the simplicity of the estimation, which is not the case with the skewed  $t$  copula. The MCMC implementation follows Kalli et al. (2011).

In all these cases, the goal of the analysis is about the complete dependence structure. The aim of this work is different. First, we allow the marginal distributions  $F_j$ 's to follow either a parametric or a non parametric model. Secondly, we do not make any parametric assumption for the copula function  $C$ . Rather, we limit our goal to the estimation of a given functional of interest of  $C$ , say  $\phi(C)$ . In this respect, we adopt a semiparametric Bayesian strategy for estimating  $\phi(C)$  where the parameter of interest is the particular functional  $\phi$  for which we derive an approximated posterior distribution

$$\pi(\phi|x) \propto \pi(\phi)\hat{L}(\phi;x),$$

where  $\hat{L}(\phi;x)$  is a nonparametric approximation of the likelihood function for  $\phi$ . In particular here we use the Bayesian exponentially tilted empirical likelihood of Schennach (2005).

We also propose a modified version of the algorithm of Mengersen et al. (2013) in a situation where the statistical model is only partially specified and the main goal is the estimation of a finite dimensional quantity of interest. In practice this represents the prototypical semiparametric set-up, where one is mainly interested in some characteristics of the population, although the statistical model may contain nuisance parameters which are introduced in order to produce more flexible models that might better fit the data at hand. In order to make robust inference on the quantity of interest, a reasonable model should account for the uncertainty on the nuisance parameters, in some way. Even if some of these additional parameters are not particularly important in terms of estimation – they often lack of a precise physical meaning – their estimates can dramatically affect inference on the parameter of interest. In these circumstances it might be more reasonable and robust to partially specify the model. Empirical likelihood has been introduced by Owen (2001); it is a way of producing a nonparametric likelihood for a quantity of interest in an otherwise unspecified statistical model. Schennach (2005) proposed an exponentially tilted empirical likelihood which can also be interpreted as a semiparametric Bayesian procedure. Assume that our dataset is composed of  $n$  independent replicates  $(x_1, \dots, x_d)$  of some random vector  $X$  with distribution  $F$  and corresponding density  $f$ . Rather than defining the usual likelihood function in terms of  $f$ , the Bayesian exponentially tilted empirical likelihood is constructed with respect to a given quantity of interest, say  $\phi$ , expressed as a functional of  $F$ , i.e.  $\phi(F)$ , and then a sort of profile likelihood of  $\phi$  is computed in a nonparametric way. More precisely, consider a given set of generalized moment conditions of the form

$$E_F(h(X, \phi)) = 0, \tag{2}$$

where  $h(\cdot)$  is a known function, and  $\phi$  is the quantity of interest. The resulting Bayesian exponentially tilted empirical likelihood  $L_{BEL}(\phi;x)$  is defined as

$$L_{BEL}(\phi, x) = \prod_{i=1}^n p_i^*(\phi), \tag{3}$$

where  $(p_1^*(\phi), \dots, p_n^*(\phi))$  is the solution of

$$\max_{(p_1, \dots, p_n)} \sum_{i=1}^n (-p_i \log p_i),$$

under the constraints  $0 \leq p_i \leq 1$ ,  $\sum_{i=1}^n p_i = 1$ , and  $\sum_{i=1}^n h(x_i, \phi) p_i = 0$ . The third condition induces a profiling of the information towards the quantity of interest, through an unbiasedness condition. Mengersen et al. (2013) proposed a sort of “sampling importance re-sampling” (Rubin et al., 1988) method for dealing with situations where the “true likelihood” evaluation is out of reach and parameter values are “weighted” by the empirical likelihood proposed by Owen (2001). Here we replace the empirical likelihood with the exponentially tilted empirical likelihood proposed by Schennach (2005), in order to guarantee a solid Bayesian justification of the procedure.

### 3 The Bayesian Use of Exponentially Tilted Empirical Likelihood

We describe in detail our method. First we illustrate the role played by the two steps, then we present the general algorithm in a pseudo-code style and we finally comment on the main issues.

#### 3.1 Step 1: Marginal Estimation

We assume that a data set is available in the form of a  $n \times d$  matrix  $x = (x_1, x_2, \dots, x_d)$ , where  $n$  is the sample size and  $d$  is the number of variables. Given that the object of interest is a functional of the copula structure, inference for the marginals is not central in this description: one can either use a parametric or a nonparametric model for the marginals. In the first case, for each  $j = 1, \dots, d$ , we use data available for  $X_j$  to derive an estimate of  $\pi(\lambda_j | x_j)$ ; for example, we can generate a sample  $\lambda_j = (\lambda_j^{(1)}, \lambda_j^{(2)}, \dots, \lambda_j^{(S_j)})$  which is an approximation of the posterior distribution for  $\lambda_j$ . We allow  $S_j$  to vary for  $j = 1, \dots, d$  in order to take into account particular features of the marginal models or the information available for each variable  $X_j$ . Alternatively, Bayesian nonparametric estimates of the marginal distributions may be obtained, see Hjort et al. (2010) for a general review. In Section 4 we will argue that the nonparametric choice may lead to better convergence results of the estimation procedure for the functional  $\phi$ .

#### 3.2 Step 2: Joint Estimation

We also assume that the main focus of the analysis is the estimation of a specific function  $\phi$  of  $C$ ; because of this, we avoid to choose the complete copula structure, in order to prevent estimation biases due to model misspecification, as we will see in Section 5. This is particularly useful and meaningful in those situations where there is no theoretical or empirical evidence that a given parametric copula should be preferred and we are mainly interested in a synthetic measure of dependence, like for example, the upper tail dependence index between two components of  $X$ , discussed in Section 6. Another

popular quantity, which we will consider in Section 5, is the Spearman's  $\rho$  between two components of  $X$ , and its extensions to the multivariate case (Section 7). From this perspective, the problems analyzed in this paper belong to the class of Bayesian semiparametric problems, where the posterior distribution of a quantity of interest,  $\phi$ , is investigated, whereas the complete form of the model is considered a nuisance parameter, for which a nonparametric approach seems more cautious and reasonable.

Then, after the estimation of the marginal distributions, performed in Step 1, we now use a copula representation of the multivariate distribution in order to approximate the posterior distribution  $\pi(\phi|x)$  of  $\phi$ , a single measure of the multivariate dependence structure of  $X$ . The posterior distribution of  $\phi$  is approximated by combining its prior distribution with the Bayesian Exponentially tilted Empirical Likelihood (3),

$$\pi(\phi|x) \propto \pi(\phi)L_{BEL}(\phi;x).$$

As already stated, the Bayesian exponentially tilted empirically likelihood has been introduced by Schennach (2005), however its use has been limited so far, with the remarkable exceptions of Lancaster and Jae Jun (2010) and Yang and He (2012) for quantile regression. This approach also implies the introduction of a nonparametric prior distribution on the nuisance aspects of the model, as detailed in Schennach (2005) and discussed in Section 4.  $L_{BEL}(\cdot)$  is computed for a given choice of moment conditions of the form (2), based on a nonparametric estimator of the functional of interest, for which there must exist an (at least, asymptotically) unbiased estimator  $\hat{\phi}_n$ , i.e. such that  $E_F(\hat{\phi}_n - \phi) = 0$ . The existence of an unbiased estimator is a caveat on the use of empirical likelihoods: the problem might be partially circumvented through the use of a bootstrap likelihood as in Zhu et al. (2016).

### 3.3 The Algorithm

Here we present the pseudo-code of the proposed algorithm, which we briefly indicate as ABSCop (Approximate Bayesian semiparametric copula). In Algorithm 1 the method is described in the case when parametric models are assumed for the marginals. Nevertheless it could be easily modified to manage the case of nonparametric estimation of the marginal densities, or to a mix of the two cases. The final output is a posterior sample drawn from an approximation of the posterior distribution of the quantity of interest  $\phi$ . There are several critical issues both in the practical implementation of the method and in its theoretical properties, which we will now discuss.

**1. Marginal estimation.** As we will see in the real data application in Section 8, as long as one uses a reasonable parametric model for the marginals, the posterior distribution of  $\phi$  will not be seriously affected by this part of model choice.

**2. Weighting the prior sample.** An interesting point to discuss is that the posterior sample sizes used to approximate the marginal parameter posterior distribution may be different, maybe because estimation of some of them need to be more accurate. Therefore values  $S_j$ 's are allowed to change across  $j = 1, \dots, d$ . If, instead, all the  $S_j = S$  for  $j = 1, \dots, d$ , it would be ideally possible to run the second step (for each

**Algorithm 1** ABSCop algorithm

- 
1. STEP 1: Marginal estimation  
 Given a sample  $X = (X_1, X_2, \dots, X_d)$  where  $X_j = (X_{1j}, \dots, X_{nj})$  for  $j = (1, \dots, d)$  with joint cdf  $F_X(x)$  and marginal cdf's  $F_1(x_1; \lambda_1), \dots, F_d(x_d; \lambda_d)$ 
    - 1: **for**  $j = 1, \dots, d$  **do**
    - 2:     Derive a posterior sample for  $\lambda_j: (\lambda_j^1, \dots, \lambda_j^{S_j})$  approximating the marginal posterior  $\pi(\lambda_j | x_j)$
    - 3: **end for**
  2. STEP 2: Joint estimation
    - 1: **for**  $b = 1, \dots, B$  **do**
    - 2:     Draw  $\phi^{(b)} \sim \pi(\phi)$
    - 3:     Sample one value  $\lambda^{s_j}$  from each marginal posterior sample:  $\lambda' = (\lambda_1^{(s_1)}, \dots, \lambda_d^{(s_d)})$
    - 4:     Derive a matrix of uniformly distributed pseudo-data  $u_{ij} = F_j(x_{ij}; \lambda_j^{(s_j)})$

$$u' = \begin{pmatrix} u_{11}^{(s_1)} & u_{12}^{(s_2)} & \dots & u_{1d}^{(s_d)} \\ u_{21}^{(s_1)} & u_{22}^{(s_2)} & \dots & u_{2d}^{(s_d)} \\ \dots & \dots & u_{ij}^{(s_j)} & \dots \\ u_{n1}^{(s_1)} & u_{n2}^{(s_2)} & \dots & u_{nd}^{(s_d)} \end{pmatrix}.$$

- 5:     Compute  $L_{BEL}(\phi^{(b)}; u') = \omega_b$
- 6: **end for**
- 7: **return** A weighted sample of size  $B$  of values for  $\phi$ , where the weights are defined as the  $L_{BEL}$ , given the nonparametric estimate  $\hat{\phi}_n$ .
- 8: Sample with replacement  $(\phi^{(b)} \omega_b)$ ,  $b = 1, \dots, B$ .

**Output:** a sample of size  $B$  of values approximately from the posterior distribution of  $\phi$ .

---

$b = 1, \dots, B$ ) for vector  $\lambda^T = (\lambda_1^{(s)}, \dots, \lambda_d^{(s)})$  without sampling for each  $b = 1, \dots, B$  only a single value from each sample of the marginal posterior distributions. In this case *Point* 4 of Step 2 in Algorithm 1 will consist of  $S$  matrices  $u^s$ ,  $s = 1, \dots, S$ ; consequently,  $L_{BEL}$  will provide a set of weights  $\omega_{bs}$ ,  $b = 1, \dots, B$  and  $s = 1, \dots, S$ . In this case, the posterior distribution of  $\phi$  may be approximated by simply combining the sample from the prior distribution, say  $(\phi_1, \dots, \phi_B)$  with an average of the weights, say  $\bar{\omega}_b = n^{-1} \sum_{i=1}^S \omega_{bs}$  for  $b = 1, \dots, B$ . This version of the algorithm will be used in Section 8 and the relative Algorithm is available in the Appendix C of the Supplementary Material (Grazian and Liseo, 2017). This last version of the algorithm is of course more accurate since it considers, at each iteration, the global uncertainty in the marginal distributions. However, its computational burden may be heavy. Algorithm 1, therefore, is presented in a more manageable version, where, at each simulation  $b = 1, \dots, B$  the marginal posterior distributions are considered as approximated with a sample of size one, randomly selected among the entire marginal posterior sample.

**3. Choice of priors.** Prior elicitation is necessary for the marginal parameters  $\lambda$  and for the quantity of interest  $\phi$ . While the marginal estimation does not present peculiar issues, the elicitation of the prior on  $\phi$  could be potentially important. However, the most common functional of interest are, in general, defined on a compact space; as a consequence, a default objective choice, in the absence of specific information, is the uniform distribution. Other choices are clearly possible; our simulation studies, not reported here, suggests that the resulting posterior distribution seems to be robust in terms of prior choices. The prior on the nonparametric component of the copula is implicitly provided by the use of the exponentially tilted empirical likelihood, as proved in Schennach (2005), and discussed in Section 4: this aspect, of course, has pros and cons. The main advantage is the ease of elicitation: one need not to elicit about complex aspects of the multivariate dependence structure. This is mainly in the spirit of the so called partially specified models, quite popular in the econometric literature. Another obvious advantage is the implied robustness of the method, with respect to different prior opinions about non-essential aspects of the dependence structure. The most important disadvantage is its inefficiency when compared to a parametric copula, under the assumption that the parametric copula is the true model. On the other hand, we will see in Section 5 that the parametric approach may lead to completely wrong results in case of misspecification. Another aspect to consider is that model selection procedures are not yet fully developed in the copula literature: this is mainly due to the fact that most of the differences among the various copula models refer to the tail behavior, and it is rare to have enough data on the tails to perform reliable model comparison.

**4. Existence of an unbiased estimator.** The Bayesian exponentially tilted empirical likelihood is based on moment conditions of the form (2). As already sketched in § 3.2, this kind of conditions implies, at least implicitly, the existence of an unbiased estimator for the quantity of interest. In practical applications, there are often available only asymptotically unbiased estimators. This is the case, for example, of the Spearman's  $\rho$ : its sample counterpart  $\hat{\rho}_n$  is only asymptotically unbiased so the moment condition  $E[\hat{\rho}_n - \rho] = 0$  is only valid for large samples. Finally, a note on computational issues: the most demanding step of Algorithm 1 is the evaluation of  $L_{BEL}$ . This entails an optimization procedure over the hypercube  $[0, 1]^{n-1}$ , based on Lagrange multipliers, however this may be easily and fastly implemented in R using the generic function `optim`.

## 4 Theoretical Background

The method described in the previous section is based on several different theoretical results. In this section we collect some more theoretical considerations, in order to better clarify advantages and limitations of the proposal.

**Two-Step Estimation** The inferential step has been split into two parts: first, the marginal distributions of the multivariate random variable are estimated; then, pseudo-data are created in order to provide a semiparametric estimate of the quantity of interest.



The “two-step” issue is at the core of the often unsatisfactory behaviour of estimation procedures based on the *Inference From the Margins* method, see for a review Joe (2015), Section 10.1: the main drawback of that approach is that it fails to properly account for the uncertainty on the parameter estimates of the marginal distributions.

However this problem is much less serious in our setting; in fact, we produce, for each coordinate of the multivariate distribution, a sample from the joint posterior distribution of the parameters which appear in that marginal. So the actual level of information on those parameters is completely transferred to the *second step* of the procedure, which creates, for each run of the posterior simulation, a different set of pseudo-data and then *takes averages* on them. Provided that the estimation procedure for the marginals is consistent, we are consistently creating multiple “pseudo-data”.

**Bayesian Semi-Parametric Interpretation** In Schennach (2005) it is argued and proved that the Bayesian exponentially tilted empirical likelihood has a precise Bayesian interpretation, which we now describe in our context. The infinite dimensional parameter space for a copula model can be written as  $(C, F_1, \dots, F_d)$ ; however, the interest of the analysis is in a low-dimensional function  $\phi(C)$ . Then the copula  $C$  can be represented as  $C = (\phi, C^*)$ , where  $C^*$  belongs to an infinite dimensional metric space  $(H, d_H)$  and represents all those aspects of the dependence structure not related to  $\phi$ . The global nuisance parameter for the model is  $\xi = (C^*, F_1, \dots, F_d)$ . At each iteration of Algorithm 1, i.e. for fixed values of the marginal parameters  $\lambda$ , the computation of the Bayesian exponentially tilted empirical likelihood may be read as the evaluation of the integrated likelihood of  $\phi$ , say

$$L_{BEL}^{(\lambda)}(\phi; u) = \int_{\Xi} L(\phi, \xi; u) d\Pi(\xi),$$

where  $u = \{[u_{ij}]_{ij}, i = 1, \dots, n, j = 1, \dots, d\}$ ,  $u_{ij} = F_j(x_{ij}; \lambda_j)$  and  $\Pi(\xi)$  is the non-parametric prior process implicitly induced by the use of the Bayesian exponentially tilted empirical likelihood and specified by Theorem 1 in Schennach (2005). In brief,  $\Pi(\xi)$  is a prior process which tends to favor distributions with a high level of entropy. Algorithm 1 takes an average of  $L_{BEL}^{(\lambda)}(\phi; u)$  with respect to the posterior distributions of the parameters of the marginals  $F_1, \dots, F_d$ . Consequently, it produces an approximation of the integrated likelihood  $L(\phi; x)$  which is combined with the genuine prior for  $\phi$  in order to obtain  $\pi(\phi | x)$ .

**Partial Specification of the Model** Our model is, in some sense, only partially specified, since we are interested in a specific aspect of the copula. In order to make Bayes’ theorem applicable, we again invoke Theorem 1 in Schennach (2005), which gives a fully Bayesian interpretation of the model. It is true however, that the Bayesian exponentially tilted empirical likelihood is a valid approximation of the integrated likelihood for the parameter of interest  $\phi$  only when the moment conditions (2) are valid. It may happen, as already noticed, that available estimators of  $\phi$  are only asymptotically unbiased; as a consequence, the moment conditions (and the entire method as well) are valid only for large samples. Moreover, the quantities of interest considered in this paper (Spearman’s  $\rho$  and tail dependence indices  $\lambda_L$  and  $\lambda_U$ ) are defined in terms of the copula and the corresponding estimators are based on the empirical copula

$$\hat{C}_n(u) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbb{I}_{\{\hat{U}_{ij} \leq u_j\}}, \quad u = (u_1, u_2, \dots, u_d) \in [0, 1]^d, \quad (4)$$

where  $\hat{U}_{ij}$  are the pseudo-data obtained after the first step of the procedure (the estimation of the marginals). It is then clear that, in order to use the Bayesian exponentially tilted empirical likelihood, the empirical copula must be a consistent estimator of the copula. In order to check this condition, suppose we have obtained some estimates of  $F_1, \dots, F_d$ . For the moment we assume that they come from a Bayesian nonparametric procedure, which is asymptotically equivalent to a procedure based on the empirical marginal cumulative distribution functions. Pseudo-data are obtained from this first step of the procedure  $u_{ij} = \hat{F}_j(x_{ij})$  and they are used for deriving nonparametric estimates of  $\phi$ ; the joint use of a nonparametric procedure for the estimation of the marginals and of an asymptotically unbiased estimator in (2) provides a global procedure which is clearly asymptotically equivalent to an estimate based on the empirical copula  $\hat{C}_n$ . On the other hand, it is known (Fermanian et al., 2004) that  $(\hat{C}_n - C)$  is weakly convergent to a Gaussian process in  $\ell^\infty[0, 1]$ ; more precisely, suppose  $(X_{11}, X_{21}, \dots, X_{d1}), \dots, (X_{1n}, X_{2n}, \dots, X_{dn})$  are independent random vectors with distribution function  $F$  and marginals  $F_1, F_2, \dots, F_d$ . The empirical estimator of the copula function  $C(u_1, u_2, \dots, u_d) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_d^{-1}(u_d))$  is

$$\hat{C}_n(u_1, u_2, \dots, u_d) = \hat{F}_n(\hat{F}_{1n}^{-1}(u_1), \hat{F}_{2n}^{-1}(u_2), \dots, \hat{F}_{dn}^{-1}(u_d)),$$

where  $\hat{F}_n, \hat{F}_{1n}, \hat{F}_{2n}, \dots, \hat{F}_{dn}$  are the joint and marginal empirical distribution functions of the observations. The empirical copula process is defined as  $\mathbb{C}_n = \sqrt{n}(\hat{C}_n - C)$  and if the  $j$ -th first order partial derivative exists and is continuous on  $V_{d,j} = \{u \in [0, 1]^d : 0 < u_j < 1\}$ , then  $\mathbb{C}_n$  converges weakly to a Gaussian process  $\{\mathbb{G}_C(u_1, u_2, \dots, u_d), 0 < u_1, u_2, \dots, u_d < 1\}$  in  $\ell^\infty([0, 1]^d)$ ; for details, see Theorem 3 in Fermanian et al. (2004). If a Bayesian nonparametric procedure which is asymptotically equivalent to the empirical distribution function is used, we may still advocate Theorem 3 in Fermanian et al. (2004) and the obtained empirical copula is again consistent. A similar, but less general, argument may be used if the marginals are estimated parametrically. In this case, the Bayesian procedure will be asymptotically equivalent to a maximum likelihood approach and Joe (2005) shows that the two-step procedure based on maximum likelihood estimates is consistent. However, for finite sample sizes, there may still be a problem: if we are using a wrong model on the marginals, the entire posterior sample may be misleading and the subsequent step might be biased. This problem is, of course, common to any parametric statistical procedure for copula estimation.

## 5 Monotonic Dependence

We first illustrate the method in the simple situation  $d = 2$ , and assuming that the two marginal distributions of the data are known: without loss of generality we assume that they are both uniform in  $[0, 1]$ . The Spearman's  $\rho$  between  $X$  and  $Y$  is the correlation coefficient among the transformed variables  $U = F_X(X)$  and  $V = F_Y(Y)$  or, in a copula language,

---

**Algorithm 2** ABSCop algorithm – Spearman’s  $\rho$

---

Given a sample of  $n$  pseudo-observations assumed from an unknown copula function

$$u = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ \dots & \dots \\ u_{n1} & u_{n2} \end{pmatrix},$$

- 1: **for**  $b = 1, \dots, B$  **do**
- 2:     Draw  $\rho^{(b)}$  from its prior distribution, for example  $\rho^{(b)} \sim \text{Unif}(-1, 1)$
- 3:     Compute a nonparametric estimate of the Spearman’s  $\rho$ :

$$\hat{\rho}_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{12}{n^2 - 1} R_i Q_i \right) - 3 \frac{n+1}{n-1},$$

where  $R_i = \sum_{k=1}^n \mathbb{I}(u_{1k} \leq u_{1i}), Q_i = \sum_{k=1}^n \mathbb{I}(u_{2k} \leq u_{2i}), \quad i = 1, \dots, n.$

- 4:     Compute  $L_{BEL}(\rho^{(b)}; u) = \omega_b$
- 5: **end for**
- 6: **return** A weighted sample of size  $B$  of values for  $\rho$ , where the weights are defined as the  $L_{BEL}$ , given the nonparametric estimate  $\hat{\rho}_n$ .
- 7: Sample with replacement  $(\rho^{(b)} \omega_b), b = 1, \dots, B.$

**Output:** a sample of size  $B$  of values approximately from the posterior distribution of  $\rho$ .

---

$$\rho = 12 \int_0^1 \int_0^1 (C(u, v) - uv) dudv = 12 \int_0^1 \int_0^1 C(u, v) dudv - 3. \tag{5}$$

Starting from a sample of size  $n$  from a bivariate distribution, say  $(x_i, y_i), i = 1, \dots, n,$  a possible estimator of  $\rho$ , say  $\hat{\rho}_n$ , obtained by substituting the empirical copula  $\hat{C}_n$  in expression 5, is the correlation among ranks and it can be written as

$$\hat{\rho}_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{12}{n^2 - 1} R_i Q_i \right) - 3 \frac{n+1}{n-1}, \tag{6}$$

where

$$R_i = \text{rank}(x_i) = \sum_{k=1}^n \mathbb{I}(x_k \leq x_i), \quad Q_i = \text{rank}(y_i) = \sum_{k=1}^n \mathbb{I}(y_k \leq y_i), \quad i = 1, \dots, n.$$

We use Algorithm 1 to produce a posterior sample for  $\rho$ , with  $\pi(\rho) = \mathcal{U}(-1, 1)$ : a full description of the application of Algorithm 1 in the specific case of the Spearman’s  $\rho$  in dimension two is described in Algorithm 2; step 1 of Algorithm 1 is avoided in this simulated studies, since the simulation are already in the copula space.

The frequentist properties of estimator (6) have been considered in Borkowf (2002), who shows that the asymptotic variance of  $\rho_n$  is

$$\sigma^2(\hat{\rho}_n) = 144(-9\psi_1^2 + \psi_2 + 2\psi_3 + 2\psi_4 + 2\psi_5), \tag{7}$$

where the  $\psi_i$ 's are terms linked with the moments of the marginal and joint distributions of  $(X_1, Y_1)$  and  $(X_2, Y_2)$ , two independent random vectors with distribution  $F(x, y)$  and marginals  $F_1(X)$  and  $F_2(Y)$  respectively. In particular

$$\begin{aligned} \psi_1 &= \mathbb{E}[F_1(X_1)F_2(Y_1)], \\ \psi_2 &= \mathbb{E}[(1 - F_1(X_1))^2(1 - F_2(Y_1))^2], \\ \psi_3 &= \mathbb{E}[(1 - F(X_1, Y_2))(1 - F(X_2))(1 - F(Y_1))], \\ \psi_4 &= \mathbb{E}[(1 - F_1(\max\{X_1, X_2\}))(1 - F_2(Y_1))(1 - F_2(Y_2))], \\ \psi_5 &= \mathbb{E}[(1 - F_1(X_1))(1 - F_1(X_2))(1 - F_2(\max\{Y_1, Y_2\}))]. \end{aligned}$$

Consistent estimates of the above quantities are available in Genest and Favre (2007). However, in the case of perfect rank agreement, when plugging-in the sample estimates of the  $\psi_j$ 's into expression (7), one gets a negative number. This phenomenon also occurred in our simulations when data were generated from copulae with a value of  $\rho$  close to 1. As an illustration we have simulated 500 samples of size  $n = 1000$  from a bivariate Clayton copula with  $\rho = 0.50$ , a Frank copula with  $\rho = 0.50$ , a Gumbel copula with  $\rho = 0.683$  and a Gaussian copula with  $\rho = 0.8$ . For comparative purposes we have also implemented the nonparametric frequentist procedure described in Genest and Favre (2007), where a confidence interval for the Spearman's  $\rho$  is computed based on the asymptotic distribution of  $\rho_n$ . Table 1 compares the frequentist behavior of the confidence procedure and our proposal, in the case of Clayton, Frank, Gumbel and Gaussian copulae in the case of intermediate and strong positive dependence. One can notice that, for large values of  $\rho$  (i.e. close to 1), the frequentist estimate of the variance is negative in most cases. As a consequence, confidence intervals can not be produced. For a graphical comparison, one may refer to Appendix A in the Supplementary Material.

Table 1 shows that our method produces shorter interval estimates of  $\rho$ , compared to the frequentist approach, while maintaining the correct coverage. Notice that in this simulation study, we have used uniform marginals; this implies that this improvement is only due to the different way of dealing with the dependence structure and not on the accounting for uncertainty in the marginal estimation. The posterior median is always very close to the empirical value  $\hat{\rho}_n$ , however Table 1 shows that the average length

		Ave. Length	Coverage
<b>Clayton</b> ( $\rho = 0.50$ )	<i>Freq.</i>	0.2664	0.998
	<i>Bayes.</i>	0.2597	1.000
<b>Frank</b> ( $\rho = 0.50$ )	<i>Freq.</i>	0.3172	1.000
	<i>Bayes.</i>	0.2735	1.000
<b>Gumbel</b> ( $\rho = 0.68$ )	<i>Freq.</i>	-	-
	<i>Bayes.</i>	0.2966	1.000
<b>Gaussian</b> ( $\rho = 0.80$ )	<i>Freq.</i>	-	-
	<i>Bayes.</i>	0.2931	1.000

Table 1: Simulations from different copulae: average length and empirical coverage of the intervals obtained both via frequentist and Bayesian methods, based on 500 repetitions of the experiment.

of the frequentist intervals is larger than the corresponding Bayesian credible intervals when the frequentist procedure is valid, i.e. the estimated variance is non-negative. As the true value of  $\rho$  increases, the frequentist estimate of the variance tends to be negative (98.4% of the experiments for the Gumbel copula with  $\rho = 0.68$  and 100% of the experiments for the Gaussian copula with  $\rho = 0.80$ ); on the other hand, our procedure performs equally well. The proportion of frequentist intervals with length larger than the corresponding Bayesian interval is 0.564 for the Clayton copula (with  $\rho = 0.5$ ) and 0.892 for the Frank copula (with  $\rho = 0.5$ ); the coverage in the other two cases cannot be evaluated because of the negative frequentist estimate of the variance.

Another advantage of using a semiparametric approach is its robustness with respect to model misspecification. To see this, we have compared our results with a fully parametric approach based on standard MCMC algorithms. In particular, we have re-used the previously simulated data under the following assumptions: Clayton copula and  $\theta \sim \mathcal{TN}(0, 10, -1, \infty)$ , Gumbel copula and  $\theta \sim \mathcal{TN}(1, 10, 1, \infty)$ , Frank copula and  $\theta \sim \mathcal{N}(0, 10)$ , where  $\mathcal{TN}(\mu, \sigma, a, b)$  is a truncated normal distribution with mean  $\mu$ , standard deviation  $\sigma$  and truncation in  $[a, b]$ . Finally, the approximated posterior distributions for the copula parameters are transformed in the corresponding posterior distributions for the Spearman's  $\rho$  relative to that particular copula. Figures 1 and Table 2 show the results of the simulations. It is evident that, although a parametric model produces shorter credible intervals, the choice of the particular parametric copula is crucial. The semiparametric method is clearly the most robust. Graphical comparison of parametric and nonparametric methods for the Clayton and the Gumbel copula are available in the Supplementary Material.

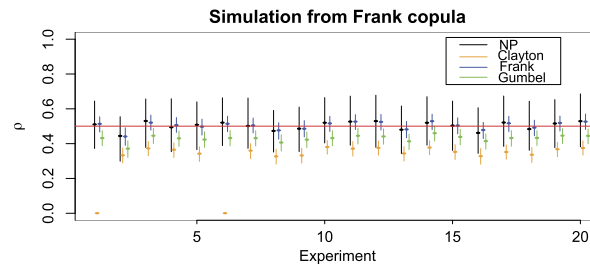


Figure 1: Bayesian point estimates (points) and credible intervals for 20 out of 500 experiments with data from a Frank copula with  $\theta = 3.45$ , obtained by specifying a Clayton model (orange), a Frank model (blue) and a Gumbel model (green) or by using our semiparametric approach (black). The solid red line represent the true value.

## 6 Tail Dependence

Multivariate dependence may be a complicated object. Popular measures like the Spearman's  $\rho$  or the Kendall's  $\tau$  can only capture some aspects of it. For example, dependencies between extreme negative stock returns or large portfolio losses are better explained by tail dependence indices (Sibuya, 1959). Several studies show that, in particular in

Assumption	True Clayton	True Frank	True Gumbel
	$\theta = 1.076$	$\theta = 3.45$	$\theta = 2$
<i>Clayton</i>	0.852	0.000	0.000
<i>Frank</i>	0.920	0.938	0.838
<i>Gumbel</i>	0.052	0.082	0.878
<i>ABSCop</i>	0.999	0.999	0.999

Table 2: Simulations from different copulae: average coverage of the 0.95 equal tails Bayesian credible intervals obtained in 500 repetitions of the experiment.

volatile markets, tail dependence is a useful tool to study the behavior of extremal data in finance. See, for example, Ané and Kharoubi (2003). Unfortunately the tail dependence is delicate to estimate, mostly because of the limited amount of available data in the tails of the distribution.

The concept of tail dependence describes the idea of concordance in the tails of the bivariate distribution, i.e. the amount of dependence in the lower-left quadrant tail or upper-right quadrant tail. The upper and lower tail dependence indices are defined in terms of the survival function:

$$\lambda_U = \lim_{v \rightarrow 1^-} \Pr \{F_X(X) > v | F_Y(Y) > v\}, \lambda_L = \lim_{v \rightarrow 0^+} \Pr \{F_X(X) \leq v | F_Y(Y) \leq v\}, \quad (8)$$

provided the limits exist.  $(X, Y)$  are said to be upper tail dependent if  $\lambda_U > 0$  and upper tail independent if  $\lambda_U = 0$ . Similar definitions apply for  $\lambda_L$ . The above definitions clarify the concept of tail concordance: the upper (lower) tail dependence index is close to one if the probability that the marginal distribution of one variable exceeds a high (low) threshold given that the marginal distribution of the other variable exceeds a high (low) threshold is close to one.

However, the tail dependence indices, as defined in (8), only depend on the copula structure:

$$\lambda_U = \lim_{v \rightarrow 1} \frac{1 - 2v + C(v, v)}{1 - v}, \quad \lambda_L = \lim_{v \rightarrow 0} \frac{C(v, v)}{v}, \quad (9)$$

and, therefore, they may be estimated by using the Bayesian approach proposed in Section 3. It is necessary to choose a nonparametric estimator of  $\lambda_U$  and  $\lambda_L$  in order to apply Algorithm 1. For a review on the parametric and nonparametric estimation of the tail dependence indices, see Frahm et al. (2005). Among the many proposals, here we consider, as a benchmark, the estimator given in Frahm et al. (2005) as a special case of the one proposed in Joe et al. (1992):

$$\hat{\lambda}_L = \frac{\hat{C}_n \left( \frac{k}{n}, \frac{k}{n} \right)}{\frac{k}{n}}, \quad \hat{\lambda}_U = 2 - \frac{1 - \hat{C}_n \left( \frac{n-k}{n}, \frac{n-k}{n} \right)}{1 - \frac{n-k}{n}},$$

where  $\hat{C}_n$  is the empirical copula, and  $0 < k \leq n$  is a parameter tuned by the experimenter. A typical choice, motivated in Joe et al. (1992), is  $k = \sqrt{n}$ . Schmidt and Stadtmüller (2006) prove strong consistency and asymptotic normality for these estimators: the moment conditions for the application of the empirical likelihood approach are, therefore, (asymptotically) valid.

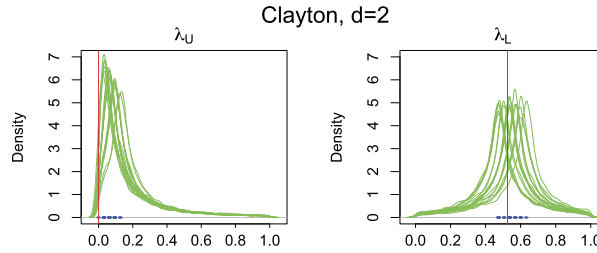


Figure 2: Comparison between frequentist (blue) and Bayesian (green) estimates for  $\lambda_U$  (left) and  $\lambda_L$  (right). 20 out of 500 simulations from a Clayton copula with  $\theta = 1.076$  ( $n = 1000$ ); the circles represent the frequentist point estimates, the lines represent the approximated posterior distributions. The true values are  $\lambda_U^{true} = 0$  and  $\lambda_L^{true} = 2^{-\frac{1}{\theta}}$  (red lines).

Schmidt and Stadtmüller (2006) have also derived the asymptotic variance of  $\hat{\lambda}_L$  and  $\hat{\lambda}_U$ . However, these expressions are of limited use since they depend on unknown quantities. To circumvent this problem, they propose to use the variance of the tail dependence coefficient of a copula for which the same quantities are easy to compute. Nevertheless this method does not provide any quantification of the potential error, which is essential in the particular case of tail dependence coefficients, for which the estimation procedure is, in general, based on a small proportion of the available data. In contrast, with our approach, we are able to provide an approximation of the entire posterior distribution of the index, which can then be summarized in different ways. Figure 2 shows the approximated intervals for the frequentist (obtained via a bootstrap estimation of the variance) and the Bayesian procedure for simulations from a Clayton copula with  $\theta = 1.076$  ( $\lambda_L = 2^{-1/\theta} = 0.525$  and  $\lambda_U = 0$ ). Bayesian intervals are always wider than the corresponding frequentist ones. Nevertheless, the coverage of the frequentist intervals is, on average, around 0.10, far from the nominal 0.95, which is reached by the Bayesian estimates. See also the Supplementary Material for other examples.

## 7 Multivariate Analysis

The extension of the proposed procedure to the multivariate case is straightforward, and no further theoretical issues arise. On the other hand, a broadly satisfactory solution in the frequentist approach has not yet been fully developed. It is important to notice that the way to describe the multivariate dependence with low-dimensional measures is still an open problem, since the number of combinations among variables increases with the dimension; as a consequence, there are several ways to define a multivariate measure of dependence. This partially explains why multivariate functionals of dependence are less used, in practice, than their bivariate counterparts. Formula (5) provides one of the possible ways to express the Spearman’s  $\rho$  and it suggests to interpret it as a measure of expected distance between the actual copula and the *independence* copula  $\Pi(u_1, \dots, u_d) = u_1 \times \dots \times u_d$ . In this sense, the extension to the  $d$ -dimensional setting is:

$$\rho_1 = \frac{\int_{[0,1]^d} (C(u) - \Pi(u)) du}{\int_{[0,1]^d} (M(u) - \Pi(u)) du} = h(d) \left\{ 2^d \int_{[0,1]^d} C(u) du - 1 \right\}, \tag{10}$$

where  $M(u) = \min(u_1, u_2, \dots, u_d)$  is the upper Fréchet–Hoeffding bound, and  $h(d) = (d + 1)/\{2^d - (d + 1)\}$ . Other definitions of the Spearman’s  $\rho$  exist in the literature (Schmid and Schmidt, 2007), for instance:

$$\rho_2 = h(d) \left\{ 2^d \int_{[0,1]^d} \Pi(u) dC(u) - 1 \right\}. \tag{11}$$

Finally, a third generalization of  $\rho$  can be obtained as the average of all the bivariate  $\rho$ ’s. This expression appears in Joe (1990); its rationale is different from those of (10) and (11), and we will not consider it. If  $d = 2$ , then  $\rho_1 = \rho_2$ , but this relation does not necessarily hold in general. Nonparametric estimators of the multivariate  $\rho_k$  for  $k = 1, 2$  to be used in Algorithm 1 are again based on the use of the empirical copula (4) in expressions (10) and (11):

$$\begin{aligned} \hat{\rho}_{1n} &= h(d) \left\{ 2^d \int_{[0,1]^d} \hat{C}_n(u) du - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - \hat{U}_{ij}) - 1 \right\}, \\ \hat{\rho}_{2n} &= h(d) \left\{ 2^d \int_{[0,1]^d} \Pi(u) d\hat{C}_n(u) - 1 \right\} = h(d) \left\{ \frac{2^d}{n} \sum_{i=1}^n \prod_{j=1}^d \hat{U}_{ij} - 1 \right\}. \end{aligned}$$

Asymptotic properties of these estimators are explored and assessed in Schmid and Schmidt (2007). In particular it is known that

$$\sqrt{n}(\hat{\rho}_{kn} - \rho_k) \overset{\sim}{\sim} \mathcal{N}(0, \sigma_k^2), \quad k = 1, 2.$$

The expressions for  $\sigma_k^2$ ,  $k = 1, 2$  are given in Schmid and Schmidt (2007). The variances of the above estimators can be analytically computed only in very few cases. In general, they depend on unknown quantities which must be estimated, for example via bootstrap methods. Bootstrap estimators of  $\rho_1$  and  $\rho_2$  have been proved to be consistent (Schmid and Schmidt, 2006): on the other hand, the bootstrap estimators of the variances tend to dramatically underestimate the variability of  $\hat{\rho}_{kn}$ ,  $k = 1, 2$ . We have performed several simulation experiments and our results always indicate that the coverage of the resulting confidence intervals for both  $\rho_1$  and  $\rho_2$  may be quite far from the nominal value and that the severity of the problem typically depends on the specific copula we sampled from. On our approximate Bayesian side, once an estimator of the multivariate version of  $\rho$  is available, it is easy to apply the procedure presented in Section 3, with no particular modifications.

Figure 3 shows the results of a simulation study with a Clayton copula with  $\rho_1 = 0.514$  and  $\rho_2 = 0.346$ . Frequentist intervals obtained via a bootstrap estimate of the variance of  $\hat{\rho}_k$ ,  $k = 1, 2$  are always very narrow; the estimated coverage is about 5.8% and it tends to further decrease as the degree of the dependence increases. It must be said that, at least for reasonably large sample sizes, the frequentist point estimates of  $\rho_1$  and  $\rho_2$  are always very precise: however, the methods for evaluating their standard errors



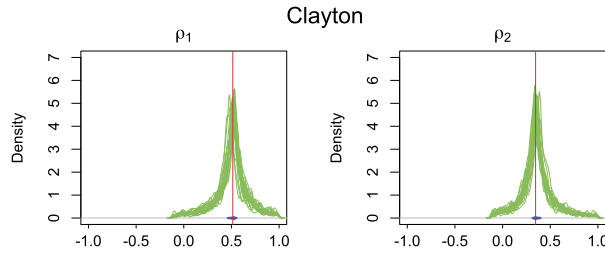


Figure 3: Comparison between frequentist (blue) and Bayesian (green) estimates of  $\rho_1$  (left) and  $\rho_2$  (right) as defined in equations (10) and (11). 20 out of 500 experiments with simulation from a Clayton copula with  $\theta = 1.076$  ( $n = 1000$ ); the true values are the red vertical lines, the blue points represent the frequentist point estimates and the green lines represent the approximated posterior distributions.

	$\hat{\rho}_1^{freq}$	$\hat{\rho}_2^{freq}$	$\hat{\rho}_1^{Bayes}$	$\hat{\rho}_2^{Bayes}$
$d = 2$	0.0032	0.0032	1.1933	1.1801
$d = 3$	0.0026	0.0026	1.0844	1.0853
$d = 4$	0.0026	0.0026	0.9495	0.9594
$d = 5$	0.0027	0.0027	0.8728	0.8914
$d = 6$	0.0027	0.0027	0.8211	0.8224
$d = 7$	0.0030	0.0030	0.8022	0.7882
$d = 8$	0.0031	0.0031	0.7828	0.7541
$d = 9$	0.0032	0.0032	0.7680	0.7492
$d = 10$	0.0035	0.0035	0.7558	0.7439
$d = 25$	0.0047	0.0047	0.7462	0.7480
$d = 50$	0.0073	0.0073	0.7299	0.7634

Table 3: Average lengths of the confidence intervals (based on a bootstrap estimator of the variance of the estimates) and of the corresponding Bayesian 95% credible intervals obtained in 50 repetitions of each experiment of dimension  $d$  by simulating data from a Clayton copula with  $\theta = 1.076$ .

seem to be seriously biased downward. We discuss examples of other copula families in the Supplementary Material, where the coverage can be even worse than in the case studied here.

We notice that the average length of the confidence intervals for  $\rho_1$  and  $\rho_2$  does not change significantly as the dimension  $d$  of the data increases. Table 3 shows the average length of the estimated confidence intervals for  $\rho_1$  and  $\rho_2$  and the average length of the corresponding (approximated) Bayesian equal tailed 95% credible intervals. One can notice that the average length of the Bayesian intervals shows a decreasing pattern as  $d$  increases. Our conjecture is that, for fixed  $n$ , the amount of information on a scalar quantity of interest tends to increase with the dimension of the data; the same phenomenon is less significant on the frequentist side because the length of intervals is always very small.

Multivariate extensions of tail dependence indices are not yet fully developed. An interesting proposal for Archimedean copulae is discussed in Di Bernardino and Rullière (2016); consider a random vector  $X = (X_1, X_2, \dots, X_d)$ , more precisely its version lying in the copula space  $U = (U_1, U_2, \dots, U_d)$  and denote by  $I$  the set  $\{1, 2, \dots, d\}$ . Consider two non-empty subset of  $I$ ,  $I_h \subset I$  and  $\bar{I}_h = I \setminus I_h$  of cardinality  $h \geq 1$  and  $d - h \geq 1$ . Provided that the limits exist, the multivariate tail dependence coefficients are given by

$$\lambda_U^{I_h, \bar{I}_h} = \lim_{u \rightarrow 1^-} \Pr\{U_i \geq u, i \in I_h | U_i \geq u, i \in \bar{I}_h\},$$

$$\lambda_L^{I_h, \bar{I}_h} = \lim_{u \rightarrow 0^+} \Pr\{U_i \leq u, i \in I_h | U_i \leq u, i \in \bar{I}_h\},$$

which describe the relative deviation of upper or lower tail probabilities of a random vector from similar tail probabilities of a subset its component. These coefficients are not uniquely defined, except that in the case of dimension  $d = 2$ , since they depend on the choice of subsets  $I_h$  and  $\bar{I}_h$  (in the case  $d = 2$ ,  $h = d - h = 1$  necessarily). De Luca and Riveccio (2012) make a specific choice of the subsets  $I_h$  and  $\bar{I}_h$  and define the tail dependence coefficients as:

$$\lambda_U = \lim_{u \rightarrow 1^-} \Pr\{F_1(X_1) \geq u | F_2(X_2) \geq u, \dots, F_d(X_d) \geq u\},$$

$$\lambda_L = \lim_{u \rightarrow 0^+} \Pr\{F_1(X_1) \leq u | F_2(X_2) \leq u, \dots, F_d(X_d) \leq u\}$$

by allowing for a simple copula representation

$$\lambda_U = \lim_{u \rightarrow 1^-} \frac{C(1 - u, \dots, 1 - u)}{(1 - u)}; \quad \lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, \dots, u)}{(u)},$$

which is also used in Salazar and Ng (2015). These definitions will be used throughout the paper. We have applied Algorithm 1 to the particular problem of nonparametrically estimating the tail dependence indices for data simulated from a Clayton, a Frank and a Gumbel copula. We have also derived the analytical formulas for  $\lambda_U$  and  $\lambda_L$  for these copulae, which are available in the Supplementary Material.

As an alternative, Di Bernardino and Rullière (2016) propose to estimate the multivariate tail dependence indices through estimation of the copula generator; however, if we assume to have no information about the shape of the copula function, it is difficult to assess the estimation error in this way. On the other hand, our approach may be easily extended to this multivariate setting.

Figure 4 shows the approximated posterior distributions for  $\lambda_U$  and  $\lambda_L$  obtained with Algorithm 1 for 20 out of 500 experiments with simulations from a Clayton copula with  $\theta = 1.076$ . While the frequentist procedure seems very precise in the case of no tail dependence, there is more variability in the estimates when there is tail dependence. In particular, the lack of reliable methods of evaluating the uncertainty linked to the estimates is a crucial problem: in 500 repetitions of the experiment the range of variation of the point estimates for  $\lambda_L$  is [0.000, 0.379]. Simulations from other types of copulas are available in the Supplementary Material.

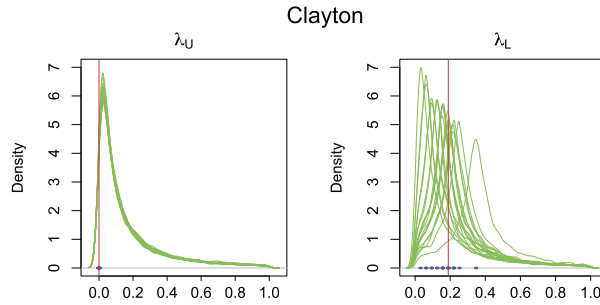


Figure 4: Approximated posterior distributions (green lines) of  $\lambda_U$  (left) and  $\lambda_L$  (right) for simulations from a six-dimension Clayton copula with  $\theta = 1.076$  in 20 out 500 experiments. The blue points are the frequentist point estimates, the red lines are the true values,  $\lambda_U^{true} = 0$  and  $\lambda_L^{true} = 6^{-1/\theta}$ .

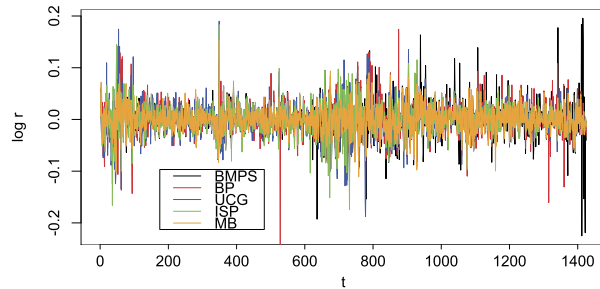


Figure 5: Log-returns of Monte dei Paschi di Siena (BMPS), Banco Popolare (BP), Unicredit (UCG), Intesa-Sanpaolo (ISP), Mediobanca (MB) for weekdays from 01/07/2013 to 30/06/2014.

## 8 Example: Dependence Among Financial Log>Returns

We now analyze a real dataset containing the log-returns of five Italian financial institutes (Monte dei Paschi di Siena, Banco Popolare, Unicredit, Intesa-Sanpaolo and Mediobanca), as defined by the benchmark stock market index for the Borsa Italiana, the Financial Times Stock Exchange Milano Indice di Borsa (FTSE-MIB). We assume that the log-returns for each bank may be modeled as a generalized autoregressive conditional heteroscedastic model with parameters  $(1, 1)$  and Student- $t$  innovations. Data refers to weekdays from 01/07/2013 to 30/06/2014; they are available on the web-page <https://it.finance.yahoo.com> and are summarized in Figure 5. This example shows the possibility, given by a copula representation of the multivariate distributions of the log-returns, to model the univariate distributions in a flexible and realistic way by considering fat tails, without concern on the existence or easiness of estimation of the multivariate version. GARCH models are known to suffer of the curse of dimensionality (McAleer et al., 2011), however with the copula representation it is possible to sepa-

rately model the marginal distributions in this way and then separately estimate the copula function. We will see in the following that the marginal modeling may have a limited impact on the estimation of the posterior distribution of the functional of interest of the dependence. The GARCH- $t$  model may be expressed via data augmentation (Geweke, 1993) as, for  $t = 1, \dots, T$ ,

$$y_t = \varepsilon_t \sqrt{\frac{\nu - 2}{\nu}} \omega_t h_t; \quad h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta h_{t-1}; \quad (12)$$

$$\varepsilon_t \sim \mathcal{N}(0, 1); \quad \omega_t \sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right),$$

where  $\alpha_0 > 0$ ,  $\alpha_1, \beta \geq 0$ ,  $\nu > 2$  and  $IG(a, b)$  denotes the inverse gamma distribution with shape parameter  $a$  and scale parameter  $b$ . For each institute the posterior distribution of the model parameters  $(\alpha_0, \alpha_1, \beta, \nu)$  may be approximated by using the **R** package `bayesGARCH` (Ardia and Hoogerheide, 2010).

There are several other models which may be used in this setting; for a comparison, we use the quantile distributions, which are defined in terms of their inverse cumulative distribution functions, which are functions of the quantiles of a standard normal distribution (Rayner and MacGillivray, 2002). In particular, we use a  $g$ -and- $k$ , given by

$$Q(z; A, B, g, k) = A + B * \left(1 + \frac{1 - \exp(-gz)}{1 + \exp(-gz)}\right) (1 + z^2)^k z,$$

where  $A$ ,  $B$ ,  $g$  and  $k$  are the parameter of location, scale, skewness and kurtosis respectively and  $z \sim \mathcal{N}(0, 1)$ . Here, we use the extension to time series data proposed by Drovandi and Pettitt (2011), where  $z_i$  follows a  $MA(1)$  model

$$z_i = \eta_i + \alpha \eta_{i-1}, \quad i = 1, \dots, n, \quad (13)$$

where  $n$  is the number of observations and  $\eta_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . Each  $z_i$  is then divided by  $\sqrt{1 + \alpha^2}$  to ensure it is marginally distribution as a standard normal. The most used approach to deal with this model, which is characterized by an intractable likelihood, is through approximate Bayesian computation (Allingham et al., 2009), that is the approach we are using here.

Once the marginal distributions are estimated, it is necessary to derive the pseudo-data to construct the copula. In this particular situation, the density function is analytically unavailable and so the distribution function. It is, therefore, possible to use a nonparametric approach to derive the pseudo-data, by using, for instance, a Pitman–Yor process prior as described in Nieto-Barajas and Contreras-Cristán (2014). This example shows the possibility to perform a separate analysis for the marginal distributions and the joint distribution.

Once the pseudo-data have been derived, it is possible to apply Algorithm 1. Appendix C of the Supplementary Material describes the several steps of implementation for the particular case of marginal GARCH- $t$  models, for the approximation of the posterior distribution of the first version of the multivariate Spearman's  $\rho$ , say  $\rho_1$ ; for the tail dependence indices, it is possible to use a  $Unif(0, 1)$  prior instead of the  $Unif(-1, 1)$ .

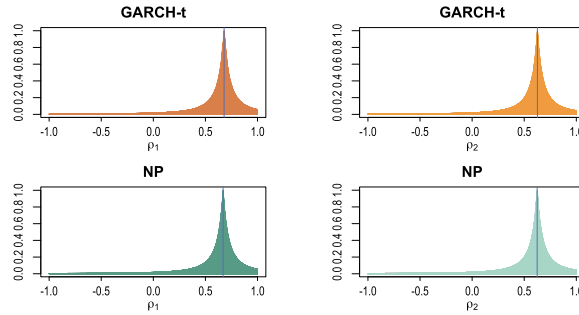


Figure 6: Approximation of the posterior distribution of the multivariate Spearman's  $\rho_1$  (left) and  $\rho_2$  (right) for the log-returns of the investments of five Italian institutes based on 10,000 simulations. On the top, there are the approximation obtained by assuming a GARCH- $t(1,1)$  model, while at the bottom there are the approximation obtained by using a nonparametric procedure in the marginal estimation. The blue vertical lines represent the frequentist estimates.

Without loss of generality, we have decided to describe the case where an equal number of simulations is chosen in the first step, in such a way that the posterior distributions of all the univariate marginals are approximated by samples of equal size.

Figure 6 shows the results of the Bayesian procedure based on Algorithm 1 for  $\rho_1$  and  $\rho_2$  as defined in Section 7. The approximated Bayesian posterior means are 0.604 and 0.600 for  $\rho_1$  with the parametric and the nonparametric procedure respectively and 0.559 and 0.561 for  $\rho_2$ ; the posterior distributions are centered around the frequentist estimates, however they provide a quantification of the uncertainty that is not available with the frequentist procedure (as already shown in Section 7). The impact of the choice of the estimation procedure at the marginal step does not seem to have a great impact on the results. Figure 7 shows the corresponding results for the multivariate tail dependence indices as defined in Section 7. In this case, the posterior distributions are strongly concentrated around small values, the posterior for  $\lambda_U$  is strongly concentrated around 0 and the posterior for  $\lambda_L$  is concentrated around 0.12. The gist of this example is to emphasize the role of the Bayesian approach and the copula representation in the quantification process of tail codependence among different series which would be very hard by simply looking at Figure 5.

## 9 Discussion

This paper describes a novel method for obtaining the posterior distribution of a quantity of interest in cases where either the model is only partially specified or the computation or the evaluation of the complete likelihood is too costly. In particular, we have considered the case of copula models, although extensions to the general semiparametric approach in a Bayesian framework are easy to consider. The proposed method is deliberately approximated, since it avoids a complete specification of the statistical model.

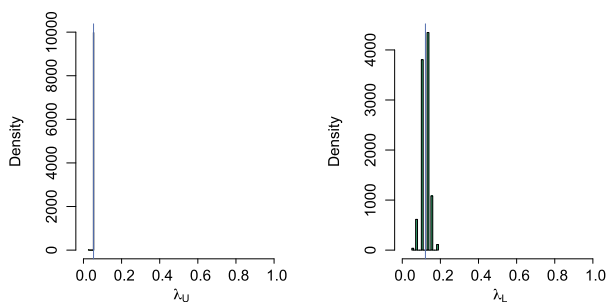


Figure 7: Approximation of the posterior distribution of the multivariate tail dependence indices,  $\lambda_U$  (left) and  $\lambda_L$  (right) for the log-returns of the investments of five Italian institutes based on 10000 simulations. The blue lines are the frequentist estimates.

This may be extremely useful in applications where the user is only interested in a particular aspect of the data, for example in particular summaries of the dependence structure. In these situations, the introduction of any further parametric assumption would generally be difficult to verify and defend and it would presumably introduce a lack of robustness. Finally, our approach provides a natural quantification of the uncertainty of the estimates of common measures of dependence in copula theory, in contrast with standard available methods.

## Supplementary Material

Approximate Bayesian Inference in Semiparametric Copula Models – Supplementary Material (DOI: [10.1214/17-BA1080SUPP](https://doi.org/10.1214/17-BA1080SUPP); .pdf).

## References

- Allingham, D., King, R. A., and Mengersen, K. L. (2009). “Bayesian estimation of quantile distributions.” *Statistics and Computing*, 19(2): 189–201. [1010](#)
- Ané, T. and Kharoubi, C. (2003). “Dependence structure and risk measure.” *The Journal of Business*, 76(3): 411–438. [1004](#)
- Ardia, D. and Hoogerheide, L. (2010). “Bayesian Estimation of the GARCH(1,1) Model with Student-t Innovations.” *The R Journal*, 2(12): 41–47. [MR2500199](#). [1010](#)
- Borkowf, C. B. (2002). “Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman’s rank correlation.” *Computational Statistics & Data Analysis*, 39(3): 271–286. URL [http://dx.doi.org/10.1016/S0167-9473\(01\)00081-0](http://dx.doi.org/10.1016/S0167-9473(01)00081-0) [1001](#)
- Burda, M. and Prokhorov, A. (2014). “Copula based factorization in Bayesian multivariate infinite mixture models.” *Journal of Multivariate Analysis*, 127: 200–213. [992](#)

- Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula methods in finance*. Wiley Finance Series. John Wiley & Sons, Ltd., Chichester. URL <http://dx.doi.org/10.1002/9781118673331> MR2250804. 991
- Choroś, B., Ibragimov, R., and Permiakova, E. (2010). “Copula estimation.” In *Copula theory and its applications*, 77–91. Springer. MR3051263. 993
- Craiu, V. R. and Sabeti, A. (2012). “In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes.” *Journal of Multivariate Analysis*, 110: 106–120. URL <http://dx.doi.org/10.1016/j.jmva.2012.03.010> 992
- De Luca, G. and Riveccio, G. (2012). “Multivariate tail dependence coefficients for archimedean copulae.” In *Advanced Statistical Methods for the Analysis of Large Data-Sets*, 287–296. Springer. 1008
- Di Bernardino, E. and Rullière, D. (2016). “On tail dependence coefficients of transformed multivariate Archimedean copulas.” *Fuzzy Sets and Systems*, 284: 89–112. URL <http://dx.doi.org/10.1016/j.fss.2015.08.030> 1008
- Drovandi, C. C. and Pettitt, A. N. (2011). “Likelihood-free Bayesian estimation of multivariate quantile distributions.” *Computational Statistics & Data Analysis*, 55(9): 2541–2556. 1010
- Fermanian, J.-D., Radulović, D., and Wegkamp, M. (2004). “Weak convergence of empirical copula processes.” *Bernoulli*, 10(5): 847–860. URL <http://dx.doi.org/10.3150/bj/1099579158> 1000
- Frahm, G., Junker, M., and Schmidt, R. (2005). “Estimating the tail-dependence coefficient: properties and pitfalls.” *Insurance. Mathematics & Economics*, 37(1): 80–100. URL <http://dx.doi.org/10.1016/j.insmatheco.2005.05.008> 1004
- Genest, C. and Favre, A. (2007). “Everything you always wanted to know about copula modeling but were afraid to ask.” *Journal of Hydrologic Engineering*, 347–368. 1002
- Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). “A semiparametric estimation procedure of dependence parameters in multivariate families of distributions.” *Biometrika*, 82(3): 543–552. URL <http://dx.doi.org/10.1093/biomet/82.3.543> 993
- Geweke, J. (1993). “Bayesian Treatment of the Independent Student- t Linear Model.” *Journal of Applied Econometrics*, 8(S): S19–40. 1010
- Grazian, C. and Liseo, B. (2017). “Approximate Bayesian Inference in Semiparametric Copula Models – Supplementary Material” *Bayesian Analysis*. doi: <http://dx.doi.org/10.1214/17-BA1080SUPP>. 997
- Gruber, L. and Czado, C. (2015). “Sequential Bayesian model selection of regular vine copulas.” *Bayesian Analysis*, 10(4): 937–963. 992
- Hjort, N. L., Holmes, C., Mueller, P., and Walker, S. G. (eds.) (2010). *Bayesian Nonparametrics*. Cambridge. URL <https://www.cambridge.org/core/books/bayesian-nonparametrics/884D64B861BEDA1CFA615B219F7086C5> 995

- Huynh, V., Kreinovich, V., and Sriboonchitta, S. (2014). “Modeling dependence in econometrics.” *Advances in Intelligent Systems and Computing*, 251. 991
- Joe, H. (1990). “Multivariate concordance.” *Journal of Multivariate Analysis*, 35(1): 12–30. URL [http://dx.doi.org/10.1016/0047-259X\(90\)90013-8](http://dx.doi.org/10.1016/0047-259X(90)90013-8) 1006
- Joe, H. (2005). “Asymptotic efficiency of the two-stage estimation method for copula-based models.” *Journal of Multivariate Analysis*, 94(2): 401–419. 993, 1000
- Joe, H. (2015). *Dependence modeling with copulas*, volume 134 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL. 991, 993, 999
- Joe, H., Smith, R. L., and Weissman, I. (1992). “Bivariate threshold methods for extremes.” *Journal of the Royal Statistical Society. Series B. Methodological*, 54(1): 171–183. URL [http://links.jstor.org/sici?sici=0035-9246\(1992\)54:1<171:BTMFE>2.0.CO;2-B&origin=MSN](http://links.jstor.org/sici?sici=0035-9246(1992)54:1<171:BTMFE>2.0.CO;2-B&origin=MSN) 1004
- Kalli, M., Griffin, J. E., and Walker, S. G. (2011). “Slice sampling mixture models.” *Statistics and Computing*, 21(1): 93–105. 994
- Kauermann, G., Schellhase, C., and Ruppert, D. (2013). “Flexible Copula Density Estimation with Penalized Hierarchical B-splines.” *Scandinavian Journal of Statistics*, 40(4): 685–705. 991
- Lancaster, T. and Jae Jun, S. (2010). “Bayesian quantile regression methods.” *Journal of Applied Econometrics*, 25(2): 287–307. 996
- McAleer, M., Caporin, M., et al. (2011). “Ranking multivariate GARCH models by problem dimension: An empirical evaluation.” Technical report, Kyoto University, Institute of Economic Research. 1009
- Mengersen, K. L., Pudlo, P., and Robert, C. P. (2013). “Bayesian computation via empirical likelihood.” *Proceedings of the National Academy of Sciences*, 110(4): 1321–1326. 992, 994, 995
- Min, A. and Czado, C. (2010). “Bayesian inference for multivariate copulas using pair-copula constructions.” *Journal of Financial Econometrics*, 8(4): 511–546. 992
- Nieto-Barajas, L. E. and Contreras-Cristán, A. (2014). “A Bayesian nonparametric approach for time series clustering.” *Bayesian Analysis*, 9(1): 147–170. 1010
- Oh, D. H. and Patton, A. J. (2013). “Simulated method of moments estimation for copula-based multivariate models.” *Journal of the American Statistical Association*, 108(502): 689–700. MR3174652. 992
- Owen, A. B. (2001). *Empirical likelihood*. CRC press. 992, 994, 995
- Pitt, M., Chan, D., and Kohn, R. (2006). “Efficient Bayesian inference for Gaussian copula regression models.” *Biometrika*, 93(3): 537–554. 993
- Rayner, G. D. and MacGillivray, H. L. (2002). “Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions.” *Statistics and Computing*, 12(1): 57–75. 1010



- Rubin, D. B. et al. (1988). “Using the SIR algorithm to simulate posterior distributions.” *Bayesian Statistics*, 3(1): 395–402. 995
- Salazar, Y. and Ng, W. L. (2015). “Nonparametric estimation of general multivariate tail dependence and applications to financial time series.” *Statistical Methods & Applications*, 24(1): 121–158. 1008
- Schefzik, R., Thorarinsdottir, T., and Gneiting, T. (2013). “Uncertainty quantification in complex simulation models using ensemble copula coupling.” *Statistical Science*, 28(4): 616–640. URL <http://dx.doi.org/10.1214/13-STS443> 991
- Schennach, S. M. (2005). “Bayesian exponentially tilted empirical likelihood.” *Biometrika*, 92(1): 31–46. URL <http://dx.doi.org/10.1093/biomet/92.1.31> 992, 994, 995, 996, 998, 999
- Schmid, F. and Schmidt, R. (2006). “Bootstrapping Spearman’s multivariate rho.” In *COMPSTAT, Proceedings in Computational Statistics*, 759–766. Springer. 1006
- Schmid, F. and Schmidt, R. (2007). “Multivariate extensions of Spearman’s rho and related statistics.” *Statistics & Probability Letters*, 77(4): 407–416. 1006
- Schmidt, R. and Stadtmüller, U. (2006). “Non-parametric Estimation of Tail Dependence.” *Scandinavian Journal of Statistics*, 33(2): 307–335. 1004
- Schölzel, C. and Friederichs, P. (2008). “Multivariate non-normally distributed random variables in climate research – Introduction to the copula approach.” *Nonlinear Processes in Geophysics*, 15(5): 761–772. URL <http://www.nonlin-processes-geophys.net/15/761/2008/> 991
- Sibuya, M. (1959). “Bivariate extreme statistics, I.” *Annals of the Institute of Statistical Mathematics*, 11(2): 195–210. 1003
- Sklar, M. (2010). “Fonctions de répartition a  $n$  dimensions et leurs marges.” *Annales de L’ISUP*, 54(1–2): 3–6. With an introduction by Denis Bosq. 993
- Smith, M. and Khaled, M. (2012). “Estimation of copula models with discrete margins via Bayesian data augmentation.” *Journal of the American Statistical Association*, 107(497): 290–303. 992
- Smith, M. S. (2013). “Bayesian approaches to copula modelling.” In *Bayesian theory and applications*, 336–358. Oxford Univ. Press, Oxford. URL <http://dx.doi.org/10.1093/acprof:oso/9780199695607.003.0017> 992, 993
- Wu, J., Wang, X., and Walker, S. G. (2014). “Bayesian Nonparametric Inference for a Multivariate Copula Function.” *Methodology and Computing in Applied Probability*, 16(3): 747–763. 992, 993
- Yang, Y. and He, X. (2012). “Bayesian empirical likelihood for quantile regression.” *The Annals of Statistics*, 40(2): 1102–1131. 996
- Zhu, W., Marin, J. M., and Leisen, F. (2016). “A bootstrap likelihood approach to Bayesian computation.” *Australian & New Zealand Journal of Statistics*, 58(2): 227–244. URL <http://dx.doi.org/10.1111/anzs.12156> 996

**Acknowledgments**

The work of the two Authors has been partially supported by the Royal Society Funding for the project “Empirical and Bootstrap Likelihood procedures for Approximate Bayesian Inference” (2015) and the Ministero dell’Istruzione dell’Università e della Ricerca, Italia PRIN 2015. The Authors would like to thank the anonymous referees for many helpful suggestions, which greatly improved the paper. The R codes used in the applications in this paper and the data used in Section 8 are available online on GitHub: <https://github.com/cgrazian/BayesCopule>.