Approximate Bayesian Computation for Factor Copula Models

Konstantinos Giannakopoulos

Supervisor : Dr Samuel Livingstone

Department of Statistical Science
University College London

Word Count: 9614

September 18, 2021

Contents

Abstract

Modelling dependence and estimating parameters in high dimensions has always been a challenge due to the scarcity of models and the complexity of the task. Factor Copulas are a class of multivariate distributions that allow to flexibly model the dependence between random variables, especially in high dimensions and complex scenarios. The Approximate Bayesian Computation (ABC) algorithm is a computational method which obtains the posterior distributions of model parameters through simulations. This method is appealing in cases where it is easy to simulate from the model but the likelihood is either not known in closed-form or is computationally intractable. The likelihood of Factor Copulas can not, typically, be obtained in closed-form which incentivises the implementation of the ABC algorithm. This report investigates the parameter estimation of Factor Copulas through the application of the ABC algorithm. Intuitively, the ABC algorithm approximates the posterior distribution by simulating a large number of parameter values from their prior and accepting those that generated data, using the Factor Copula model, that are "similar" to the observed. Major advantages of estimating Factor Copulas using the ABC algorithm are, first, the fact that it provides the full posterior distribution and second that it allows the use of ABC-related techniques, such as regression adjustment methods, which improve the estimation in simple and often intuitive ways. The objective of this report is to assess the parameter estimation results with respect to sample size and other free parameters and to inspect ways to improve the estimation.

Acknowledgements

"I would like to thank my supervisor Dr Samuel Livingstone for his guidance, support and patience for the duration of this research project. I would also like to thank Dr Rodrigo Targino for his contribution. It has been a pleasure working with both on this project. Lastly, I would like to thank my family and my partner for their constant support and motivation throughout this project".

1 Introduction

In Bayesian Inference, the focus is often set on the posterior distribution $p(\theta|D) \propto p(D|\theta)p(\theta)$ for a certain parameter $\theta \in \Theta \subseteq \mathbb{R}^n$, $n \geqslant 1$, given some observed data D. The posterior distribution is obtained via the likelihood function of the observed data, $p(D|\theta)$, and the prior distribution of the parameter, $p(\theta)$. The Approximate Bayesian Computation (ABC) algorithm is a class of computational methods that is embedded in the field of Bayesian statistics. It allows to obtain the posterior distribution of a parameter while it avoids evaluating the likelihood function of the data. This is particularly useful in cases where the evaluation of the likelihood is extremely hard, time consuming or simply impossible, but it is possible and straightforward to simulate data, $\hat{D}$, from the model, $\hat{D} \sim p(\cdot|\theta)$. The ABC algorithm extends the set of models that can be considered by a researcher, as it adds to the models whose likelihood can be derived analytically, all those models that one can easily simulate data from. Even though ABC methods have strong mathematical foundation, they also make assumptions and approximations that need to be carefully taken into account.

Two main approximations are made when using an ABC algorithm. First, the posterior of the observed data, $p(\theta|D)$ is approximated by $p(\theta|S) \propto p(S|\theta)p(\theta)$, where S is a vector of summary statistics from D. Typically, S is of a much lower dimension compared to D which increases the probability to generate data from the model that are "similar" to the observed. This is a good approximation of the true posterior when S carries all the information from D and that is when S is a sufficient statistic for θ. Second, the ABC algorithm and especially the ABC rejection algorithm described below, approximates the posterior $p(\theta|S)$ through simulations and not by deriving it analytically, which would require evaluating the likelihood as well as the normalising constant. In short, parameter values are simulated from their prior and with each one of them, a data set is generated using the model. If the generated data are "similar" to the observed, the parameter value is accepted and is considered to be a value from a sample from the approximation of the posterior distribution. The "similarity" is assessed using a distance function ρ, that evaluates the distance between the summary statistics of the observed and generated data. If this is lower than a pre-specified threshold ϵ, then the parameter value is accepted. Therefore, the importance of the choice of the summary statistics becomes obvious since they need to be both low dimensional and highly informative. There exists a number of ways to reduce the dimension of the summary statistics within the ABC framework and a comparison of the different methods can be found in Blum et al. (2013).

The ABC rejection algorithm can be extended by adding some further steps in order to potentially improve the quality of the approximation of the posterior. These further steps are the Regression Adjustment methods that were introduced by Beaumont et al. (2002). The main idea is to first perform the rejection algorithm and, as before, accept a portion of the simulated parameter values that produced summary statistics that are similar to those of the observed data. Then, the accepted values are adjusted using a regression model to account for the difference between the observed and the simulated summary statistics. The regression adjustment methods are superior to sampling techniques that are usually used to account for differences between simulations and observations. According to Blum (2017), one of the advantages of the regression adjustment methods is that they are not part of the simulation process and hence can be implemented at a subsequent time. This means that they do not require any ad hoc algorithm development. Another advantage is that they have been tested in a variety of scenarios and have been found to produce smaller errors compared to rejection algorithms. Finally, they can readily be implemented through various software such as the `abc` (Csillery et al., 2012) package in `R` (R Core Team, 2020) and `DIYABC` (Cornuet et al., 2014).

In recent years, ABC methods have been used to analyse complex problems and have become

particularly popular in research fields such as population genetics, ecology and epidemiology (Bertorelle et al., 2010; Lopes and Beaumont, 2010; Beaumont, 2010; Csillery et al., 2010). However, the use of ABC in research fields such as finance or economics has not been extensively investigated yet, with some exceptions such as using ABC to extend the particle filtering methodology (Calvet and Czellar, 2014).

Copula models are used to model multivariate distributions and have been used in applications in a wide variety of fields, such as in insurance (Embrechts et al., 2002), economics (Paarsch and Brendstrup, 2007), finance (Patton, 2006b), epidemiology (Fine, 2000) and hydrology (Genest and Favre, 2007). Most of the copula-based models' likelihood functions are either not available in analytic form or are computationally expensive to obtain. Therefore, estimation methods other than the typical, such as maximum-likelihood estimation (MLE), need to be considered to estimate the copula models. A plausible alternative estimation method for the copula parameters was proposed by Oh and Patton (2013) and that is a modified version of the Simulated Method of Moments (SMM). Other estimation methods that have been proposed are multi-stage maximum-likelihood estimation (Joe, 2005; Patton, 2006a), iterative multi-stage maximum-likelihood (Song et al., 2005), semi-parametric MLE (Chen and Fan, 2006; Chen et al., 2006; Chan et al., 2009) and the MM estimator (Genest and Rivest, 1993; Nasri and Remillard, 2019). However, the first three methods are only applicable in cases where the marginal distributions are parametric or the likelihood is available.

The models used in this report represent a class of copula models, i.e. multivariate distributions, called Factor Copulas, proposed by Patton and Oh (2012). Their likelihood function does not typically have an analytical form so for an estimation method like MLE to be implemented, numerical integration of the likelihood would be required. The SMM estimator as well as the ABC algorithm mentioned above are good estimator candidates for the Factor Copula's parameters, since they do not require evaluating the likelihood. The Factor Copulas model the observed data through latent variables. The copula completely describes the dependence between the observable variables. The structure of the model's latent variables is used only for the copula, thus, the factor-structure is only used to estimate the joint distribution and not the marginals, which can be estimated in many flexible ways. An advantage of this class of models is that the marginal distributions of the copula do not need to be the same as the marginals of the observable variables.

The use of ABC is motivated mainly by the fact that it allows to quantify the uncertainty in the parameters and also allows using ad hoc methods that can improve the estimation. First, the ABC algorithm provides the full posterior distribution of each parameter unlike the SMM estimator which only provides a point estimate. Therefore, when using the ABC algorithm, the estimation uncertainty is available, quantifiable and can be accounted for. Second, there are plenty ABC-related techniques, such as the linear regression adjustment, that can be implemented and can improve the posterior approximations and parameters estimation in simple and often intuitive ways. Adjustment methods range from using a simple linear regression model to more complex, non-linear models, such as Neural Networks (Blum and François, 2009).

In this research project, the application of the ABC algorithm to estimate the parameters of Factor Copula models is examined for different model specifications which represent different scenarios. The examples include modeling the dependence between two, three and nine observable random variables, whose joint distributions may exhibit different characteristics such as asymmetry and or tail dependence. In general, the models considered fall in one of the following categories: (a) "equidependence " models, that are more restrictive since they assume that one common factor exists for all variables and also impose the pair-wise dependence between variables to always be the same, (b) "block-dependence" models, that are more flexible since

they allow for different levels of dependence between variables that belong to different groups. Ways to improve the parameter estimation are considered such as using weights to account for summary statistics of different scale and the quality of the approximations is investigated for different sample sizes and different thresholds in the ABC algorithm. Finally, applications of regression adjustment are presented for adjusting one and multiple parameters simultaneously, through linear and local-linear regression.

1.1 Structure of the Report

The report has the following structure. In Section 2 we introduce the concept of Approximate Bayesian Computation through a short history, the theory behind it and the algorithm is presented as well as a simple example of the implementation of the algorithm. In Section 3, a short definition of copulas is presented and a classification of the main classes of copulas. In Section 4 we introduce and discuss the Factor Copulas and its components. The distributions that are used throughout the report are also presented in detail as well as the SMM estimator in more detail. In Section 5 we first specify the summary statistics for the ABC algorithm as well as some other important details of the estimation process. Then, we obtain approximations of the posteriors of the parameters and assess the accuracy of the parameter estimation for different Factor Copulas using ABC. In Section 6, a summary of the report is provided as well as discussion for further research.

2 Approximate Bayesian Computation

The Approximate Bayesian Computation (ABC) algorithm first appeared as a theoretical idea in the early 1980's (Rubin, 1984). Since then, ABC has been used to approximate likelihood functions through simulations (Diggle and Gratton, 1984) and to make inference about posterior distributions (Tavaré et al., 1997), in cases when they were analytically intractable.

The ABC algorithm makes use of Bayes' Theorem in the case where the conditional probability of a parameter, θ, given data, D, can be obtained using the conditional probability of the data given θ:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}. \tag{2.1}$$

In this case, $p(\theta|D)$ is the posterior distribution of θ, $p(D|\theta)$ is the likelihood, $p(\theta)$ is the prior and $p(D)$ is the marginal likelihood or the prior predictive probability of the data. The normalisation constant $p(D)$ can be ignored if comparing posterior probabilities for different values of θ is of interest. However, the likelihood and prior are necessary and it could be infeasible or computationally intractable to evaluate the likelihood. This is where the use of the ABC algorithm might prove to be useful to overcome this problem.

2.1 ABC Rejection Algorithm

The ABC algorithms avoid computing the likelihood function by approximating it through simulations. The results of the simulations are then compared to the observed data. One of the most common forms of the ABC algorithm is the ABC rejection algorithm. A prior is assigned to the parameter and a 'new' parameter value is simulated from the prior, assuming that there is only one parameter but this can also be extended for multiple parameters. Then a data set is produced $\hat{D}$ using the sampled parameter and a pre specified model. The observed data D and the generated data $\hat{D}$ are compared and if they are found to be very "different", the sampled value of the parameter is rejected. This comparison can be facilitated with a distance function ρ and for a specified threshold ϵ. The algorithm terminates when a certain number of parameters is accepted. The output of the ABC rejection algorithm is a sample from the approximation of the posterior distribution.

Approximate Bayesian Computation (ABC) Rejection Algorithm

We consider a case where there is only one parameter of interest θ:

1. compute the chosen summary statistic S for the observed data D,
2. choose an acceptance threshold ϵ,
3. simulate $\hat{\theta}$ from its prior distribution $p(\theta)$,
4. simulate a 'new' data set $\hat{D}$ using the parameter $\hat{\theta}$,
5. compute $\hat{S}$ from the simulated data set $\hat{D}$,
6. compute the distance $\hat{\rho} = \rho(S, \hat{S}) = ||S - \hat{S}||$,
7. if $\hat{\rho} \leqslant \epsilon$ then accept $\hat{\theta}$; otherwise reject it,
8. repeat steps 3 - 7 until K acceptances occur.

Table 1: Description of the ABC Rejection algorithm.

As the dimension of the data increases the probability of generating data that are "close" to the observed typically decreases, which increases the computational cost of the algorithm. To

overcome this problem, a set of summary statistics from the data $\hat{S}(\hat{D})$ is chosen in order to reduce the dimensionality. This way, we obtain the partial posterior distribution (2.2)

$$p(\theta|\hat{S}) = \frac{p(\hat{S})p(\theta)}{p(\hat{S})}, \tag{2.2}$$

which is the same as the posterior only in the case when the summary statistics are sufficient for the parameter θ. In this case, the efficiency of the ABC rejection algorithm increases and no error is introduced, as sufficient statistics capture all information in the data. A summary of the ABC algorithm can be found in Table 1. In some cases and especially in practice, insufficient but highly informative summary statistics can produce adequate results (Sunnåker et al., 2013).

2.2 Advantages of ABC

As previously mentioned, ABC can be implemented in cases where it is easy to simulate from the model. Furthermore, it is preferred in cases where the model specification does not allow us to obtain a closed-form likelihood, since the ABC algorithm avoids evaluating the likelihood when obtaining the posterior distribution of a parameter. There are three significant advantages to using the ABC algorithm. First, unlike other estimation methods, such as the SMM, it obtains an approximate posterior distribution of the parameter in addition to a point estimate. This allows to quantify the uncertainty in the parameter, using Bayesian inference methods such as credible intervals. Second, it allows to use ABC-related techniques such as the regression adjustment methods which are described in detail in Section 5.9. Regression adjustment has been shown to lead to a posterior that, asymptotically, correctly quantifies uncertainty (Li and Fearnhead, 2017). Third, the posterior mean obtained by the ABC algorithm as the acceptance threshold asymptotically tends to zero, gives estimates that are at least as accurate as any other estimators based on the same summary statistics. This is more formally described by the following theorem, found in Fearnhead and Prangle (2011).

Theorem 2.1. *Let $\tilde{\theta} = E[\theta|S]$. Then, for any function g,*

$$E[L(\theta, \tilde{\theta}; W)|S] \leqslant E[L(\theta, g\{S\}; W)|S],$$

where S is a vector of summary statistics from the observed data, W is a positive definite weight matrix and L is a loss function defined as: $L(\theta, \tilde{\theta}; W) = (\theta - \tilde{\theta})^T W(\theta - \tilde{\theta})$. Furthermore, the ABC posterior mean estimate of θ is, asymptotically as $\epsilon \to 0$, optimal among estimates based on the same summary statistics S.

2.3 Example of ABC

In the following example, from Sunnåker et al. (2013), we consider a system that can be characterized by a Hidden Markov Model (HMM) that consists of two states A and B. There exists a probability of transition, θ, from one state to the other which is the same in both directions and the probability to remain in the same state $q = 1 - \theta$. The calculation of the likelihood of a time series generated by this model is tedious and, therefore, is a good opportunity to employ the ABC algorithm. The summary statistic S used is the number of transitions between the two states in one time series. The distance function used, ρ, is the absolute difference between the summary statistics of the observed and the simulated data accompanied by a acceptance threshold ϵ. In order to obtain the posterior distribution of θ we go through the steps described in Table 2.

Step-by-step ABC Rejection algorithm example
A data set is generated using the previous model with the parameter θ set to be 1/4. The observed data set D: AAAABAABBAAAAAABAAAA.

1. The summary statistic for the observed data is computed, $S = 6$,
2. the acceptance threshold is chosen to be $\epsilon = 2$,
3. $\hat{\theta}$ is simulated from the prior distribution of $\theta \sim f(\theta) = Unif(0, 1)$,
4. a 'new' data set $\hat{D}$ is simulated using the model with the parameter $\hat{\theta}$,
5. the summary statistic $\hat{S}$ is computed for the simulated data set,
6. the distance $\rho(S, \hat{S}) = ||S - \hat{S}||$ is computed,
7. the distance $\rho(S, \hat{S})$ is compared to the threshold ϵ to either accept or reject $\hat{\theta}$,
8. steps 3 to 7 are repeated five times.

Table 2: Implementation of the ABC rejection algorithm.

For a threshold $\epsilon = 2$, the results are presented in Table 3. The fourth sample is accepted as it generated a sequence with 6 transitions between the two states of the model, just as many as in the observed data. The value of θ that generated this sequence is 0.65 which is not close to the true value of θ.

It becomes obvious that we need a much larger number of samples from the prior in order to obtain a reasonable approximation of the posterior distribution. We also need to reduce the threshold to zero to remove any bias introduced by the rejection algorithm. Therefore, the number of samples is increased from 5 to 10^2, 10^3, 10^4 and finally to 10^5 while keeping the length of the sequences fixed to $m = 20$. We also set the threshold to be $\epsilon = \{0, 1, 2\}$ and we obtain the approximations of the posterior distribution as in Figure 1. The posterior concentrates around the true value of θ as the number of the observed data points increases. Since the summary statistic used in this example is sufficient, proven below, we expect the posterior to quickly concentrate around the true value of θ as we increase both the number of samples and the length of the sequences.

Sample No.	**Simulated data** $\hat{D}$	**Summary Statistic** $\hat{S}$	**Distance** $\rho(S, \hat{S})$	**Result**
1	BAABBBBAABBBBABAABAA	9	3	Rejected
2	AAAAAAAAAAAABBAAAAAA	2	4	Rejected
3	ABABABABABABABBABABA	18	12	Rejected
4	ABBBBBAAAABABBAAAAAA	6	0	Accepted
5	AABBABBBBAABBAABABBB	9	3	Rejected
Observed	AAAABAABBAAAAAABAAAA	6		

Table 3: Results of the ABC algorithm.

The approximations of the posterior distributions of θ for the following combinations are presented in Figure 2: $n = 10^3$ samples and sequences of length $m = 100$, $n = 10^4$ and $m = 200$, $n = 10^6$ and $m = 20$ and $n = 5\text{x}10^4$ and $m = 500$. It is now clear that the posterior converges around the true value of θ much quicker as both the number of samples and the length of the sequence increase.

Sufficiency of statistic

The summary statistic used in this example is sufficient. This becomes more clear if we model the possibility of switching state as a binomial random variable with parameter p. Hence, each sequence of simulated data is a sequence of independent binomial random variables. The summary statistic, S, is just the number of successes which is proven below to be a sufficient

statistic for p. Hence, as the sample increases, we can see the posterior distribution to converge to the true value of θ.

Theorem 2.2 (Fisher-Neyman Factorisation)**.** *Let X be a random variable with density $P(x|\theta)$ for some $\theta \in \Theta$. The statistic $t(X)$ is sufficient for θ if and only if the density can be factorized into a function $a(x)$ and a function $b(t, \theta)$, a function of θ but only depending on x through the $t(x)$; i.e.,*

$$p(x|\theta) = a(x)b(t, \theta).$$

Proof of Sufficiency. Let us assume $X \sim$ Binomial(1, p), where $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Hence $X \sim p(x|p) = p^x(1-p)^{1-x}$, $p \in [0, 1]$. We observe sequences of m independent realizations of $X : \tilde{x} = (x_1, \ldots, x_m)$, with

$$p(\tilde{x}|p) = \prod_{i=1}^{m} p^{x_i}(1-p)^{1-x_i} = p^t(1-p)^{1-t},$$

where $t = \sum_{i=1}^{m} x_i$, $t = \{0, 1, \ldots\}$. We can observe that $p(\tilde{x}|p)$ depends on $\tilde{x}$ only through t so according to Theorem 2.2, t is sufficient for p. □

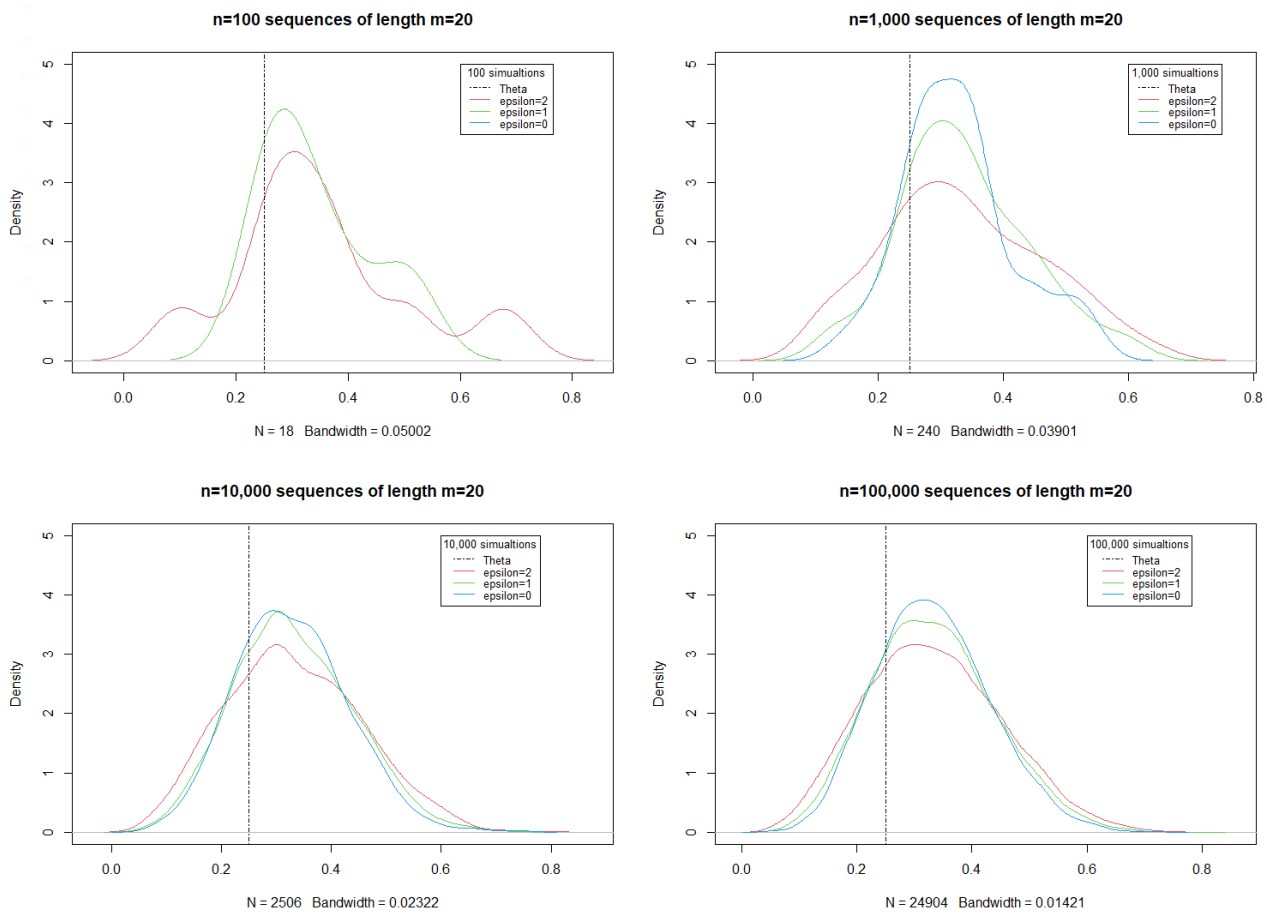

Figure 1: Approximations of the posterior distribution of θ as the number of simulations increases.

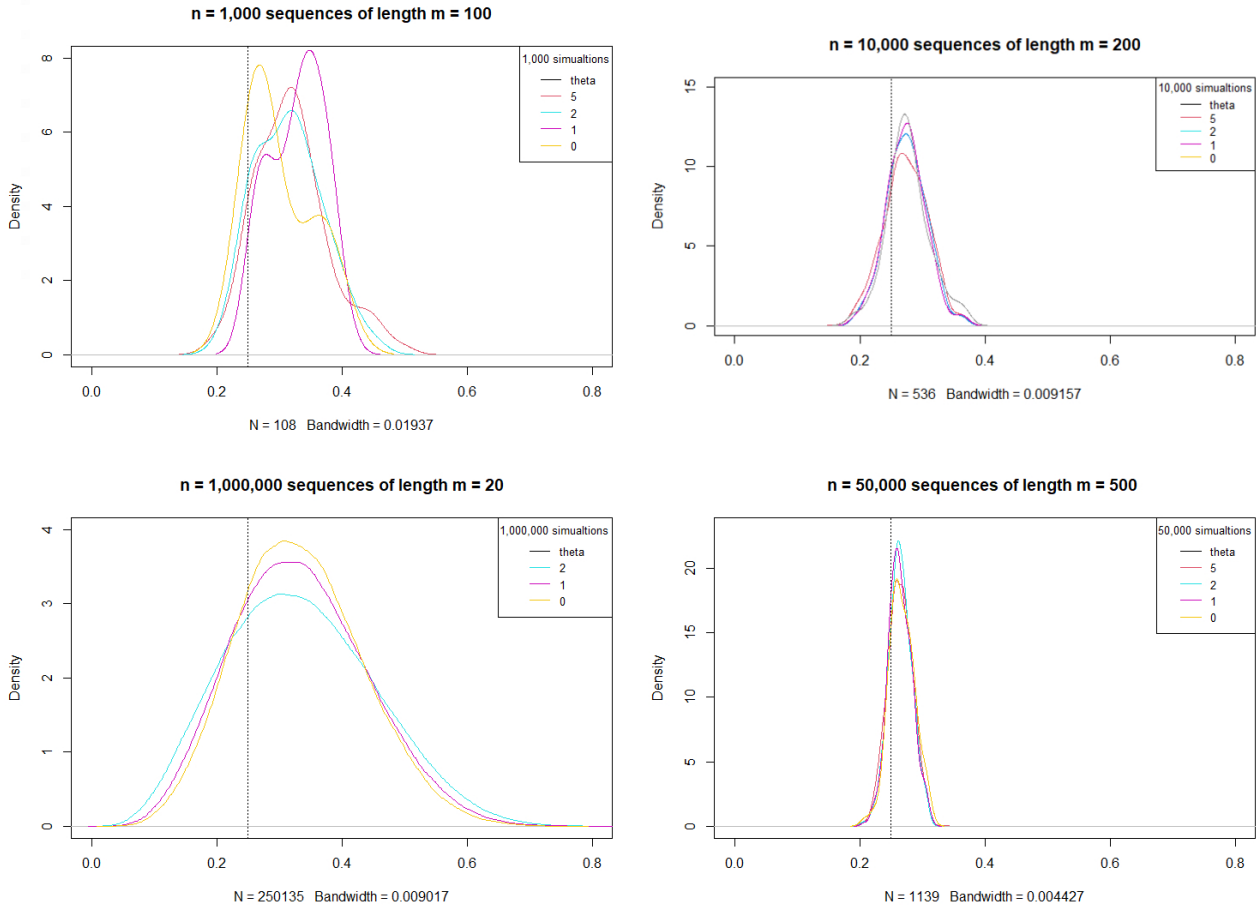

Figure 2: Approximations of the posterior distribution of θ as the number of simulations and the length of the simulated sequences increase.

3 Copulas

Copulas are a great tool to investigate and model the behaviour of not only single variables but also their joint distribution. Copulas are a form of multivariate functions and can elegantly express multivariate distributions (Größer and Okhrin, 2021). According to Sklar's theorem, there exists a copula for any multivariate distribution and the joint distribution is equal to the copula of the marginal distributions (Sklar, 1959).

Copulas allow us to isolate the dependence structure in a multivariate distribution. In practice, they are used in a variety of modelling applications. One of the earliest applications of copula modelling was to model economic risk (Embrechts et al., 2002). The rise in popularity reached its peak around the time of the 2008 financial crisis when they were (naively) used in the valuation of complex financial products such as Collaterized Debt Obligations (CDOs) and Credit Default Swaps (CDSs) (Li, 2000; MacKenzie and Spears, 2014). The fields of application vary from risk modelling to weather forecasting (Schefzik, 2015).

Definition 3.1 (Multivariate Copula)**.** *Let $C : [0,1]^d \to [0,1]$ be a joint cumulative distribution function of a d-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals. Then C is a d-dimensional copula (Nelsen, 1999). For any multivariate distribution function F with margins $F_1, ..., F_d$ there exists a copula C such that:*

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)),\ x_1, \ldots, x_d \in \bar{\mathbb{R}} = \mathbb{R} \cup (-\infty, +\infty).$$

If F is continuous, then the copula C is unique. Otherwise, it is only uniquely determined on $F_1(\bar{\mathbb{R}})$ x $\cdots$ x $F_d(\bar{\mathbb{R}})$. The converse is also true: for any copula C and univariate distribution functions $F_1, \ldots, F_d$, the function $C(F_1(x_1), \ldots, F_d(x_d))$ is a multivariate distribution function with marginals $F_1, \ldots, F_d$.

Some examples of different classes of copulas are the following:

- **Elliptical Copulas:** These are copulas of elliptical distributions such as the Normal or Student's - t and do not have a closed-form,
- **Archimedian Copulas (AC):** These copulas do not make use of Sklar's theorem and they generally have a closed-form expression. The most important Archimedian copulas include the Independence, the Frank, the Clayton and the Gumbel copula. In the multivariate case, these genearalise to the Hierarchical Archimedian copulas (HAC) (Okhrin et al., 2013),
- **Extreme Value Copulas (EV):** These copulas model the dependence of the maximum of some independent and identically distributed random variables and are often used in hydrology, flood and financial models,
- **Archimax Copula:** The Archimax copula (Capéraà et al., 2000) is a bivariate copula which is a combination of AC and EV copulas and is used to model asymptotic dependence between variables. In higher dimensions, these generalise to the Multivariate Archimax copulas (Mesiar and JáGr, 2013),
- **Vine Copulas:** Vine copulas are multivariate copulas introduced by Joe (1996) which divide the dependency structure of d variables into $d(d-1)/2$ bivariate copulas, both between pairs of variables and between pairs of variables conditioned on other variables. There exist three main types of Vine copulas; the canonical vine (C-vine) which assumes that at each level each node connects with all other nodes, the drawable vine (D-vine)

which assumes that each node connects at most with two other nodes which results in a "line" of copulas and the regular vine (R-vine) in which two edges in the tree representation are joined by an edge (in the following tree) only if these edges share a common node,

- **Factor Copulas:** This family of copulas makes use of the idea of factor analysis that random variables can be modelled by latent variables. In Factor Copulas, the variables X_i are represented by a function of the latent variable(s) and an error term. In the simplest case, we have the linear additive model $X_i = \sum_{k=1}^{K} \beta_{ik} Z_k + \epsilon_i,\ i = 1, \ldots, N$, where Z_k are the latent varaibles and ϵ_i the error terms (Joe, 2015; Krupskii and Joe, 2013, 2015).

4 Factor Copulas

Let us consider a set of N random variables $Y_1, \ldots, Y_N$ with joint distribution function F_y and marginal distributions F_i, for $i = 1, \ldots, N$, and copula C. Then:

$$\boldsymbol{Y} \equiv (Y_1, \ldots, Y_N)' \sim F_y = C(F_1, \ldots, F_N) \tag{4.1}$$

The copula $C(F_1, \ldots, F_N)$ describes the dependence among the variables $Y_1, \ldots, Y_N$. The family of models presented in the next subsection will be used to model this dependence. The marginal distributions F_i, for $i = 1, \ldots, N$, are estimated using the empirical distribution function (EDF):

$$\hat{F}_i(z) = \frac{1}{n} \sum_{t=1}^{n} 1\{x_{it} < z\}, \tag{4.2}$$

where n is the number of samples and $\hat{F}_i$ and x_{it} correspond to the variable X_i. Decomposing the joint distribution and working with the marginals and the copula instead of working with the joint distribution F_y directly dissects the modelling task into smaller pieces and allows the use of a variety of models for the univariate distributions F_i.

4.1 The Copula of a Latent Factor Structure

The factor copula models consist of K latent variables Z_i for $i = 1, \ldots, K$ and the terms ϵ_i for $i = 1, \ldots, N$. The model can be described as the following:

$$\begin{aligned} \text{Let } X_i &= \sum_{k=1}^{K} \beta_{ik} Z_k + \epsilon_i, \ i = 1, \ldots, N, \\ \text{so that } (X_1, \ldots, X_N)' &\equiv \boldsymbol{X} = \boldsymbol{BZ} + \boldsymbol{\epsilon}, \end{aligned} \tag{4.3}$$

where $\epsilon_i \overset{iid}{\sim} F_\epsilon$, $Z_k \sim F_{Z_k}$ and $Z_k \perp \epsilon_i \ \forall k, i$. Then:

$$\boldsymbol{X} \sim F_X = C(G_1, \ldots, G_N), \tag{4.4}$$

where $G_1, \ldots, G_N$ are the marginal distributions of $X_1, \ldots, X_N$ and are different to those of the original variables $Y_1, \ldots, Y_N$, $F_i \neq G_i$ in general. The copula of the latent variables X, $C(G_1, \ldots, G_N)$, is used as the model for the copula of the observable variables Y.

The copula implied in (4.3) is generally not known. However, in some specific cases it might be as in the case where $F_\epsilon, F_{Z_1}, \ldots, F_{Z_K}$ are all Normal distributions. In this case the distribution of $\boldsymbol{X}$ is a multivariate Normal which implies a Gaussian copula. For a different choice of distributions for $F_\epsilon, F_{Z_1}, \ldots, F_{Z_K}$ both the joint distribution and the copula of $\boldsymbol{X}$ are not known in closed form. However, we can extract properties of the implied copula by simulating from a model with these distributions for ϵ_i and Z_i. The properties of interest in our case are dependence measures such as Rank Correlation and Quantile Dependence, described in Section 5.1.

The choice of the distributions for the common variables Z_k and the idiosyncratic variables ϵ_i as well as the number K of factors Z_k are of great importance to the flexibility of the model. In most applications, the distributions should allow the model to capture asymmetry and or tail dependency. Additionally, allowing for a larger number of factors K gives a lot more

flexibility to the model at the cost of adding more parameters that will need to be estimated. In Sections 5.4 and 5.6 we set $K = 1$ since we only have to estimate a bivariate copula, but as the dimension increases a larger K will probably have to be chosen. One way to determine the number of factors has been proposed in Oh and Patton (2017), introduced by Cattell (1966). They suggest that, under certain conditions, the number of factors corresponds to the number of "large" eigenvalues of the covariance or correlation matrix.

4.2 Specification of Distributions

The distributions that appear throughout this report in the Factor Copulas are the Normal distribution, Skewed Normal distribution, Student's-t distribution and Skewed-t distribution. All of these appear in the Factor Copulas presented in Section 4.3 or in Section 5. These distributions when used in Factor Copulas, produce copulas with different characteristics. For example, the Normal copula is symmetrical and has zero tail dependence, the t(ν)-t(ν) copula is symmetrical and exhibits tail dependence, the Skewed Normal - Normal is asymmetrical and exhibits zero tail dependence while the Skewed t(ν)-t(ν) is both asymmetrical and exhibits tail dependence. The densities of the distributions as well as their parametrisation in `R` is described below:

- **Normal distribution**: This appears both in Section 4.3 and in Section 5.4. The parametrisation in `R` is:

  rnorm(**n** = Number of samples, **mean** = Mean of the Normal Distribution, **sd** = Standard Deviation of the distribution).

- **Skewed Normal distribution**: This appears in Section 4.3. The parametrisation of this distribution is as defined in `R`'s `sgt` (Davis, 2015) package:

  $$f(x;\mu,\sigma,\lambda) = \frac{1}{v\sigma\sqrt{\pi}} e^{-\left(\frac{|x-\mu+m|}{v\sigma(1+\lambda sign(x-\mu+m))}\right)^2}, \tag{4.5}$$

  where $m = \frac{2v\sigma\lambda}{\sqrt{\pi}}$. This is implemented in `R` as:

  rsgt(**n** = No. of samples, **mu** = Mean, **sigma** = Standard Deviation, **lambda** = Skewness parameter, **p = 2**, **q = Inf**).

- **Student's-t distribution**: This appears in Section 4.3 and in Sections 5.6, 5.7 and 5.8. The parametrisation of this distribution is as defined in Hansen (1994) and implemented with `R`'s `sgt` package. Hansen's parametrisation of Student's - t probability density function with unit variance is:

  $$g(z|\eta) = \frac{\Gamma(\eta+1)}{\sqrt{\pi(\eta-2)}\Gamma(\frac{\eta}{2})}(1+\frac{z^2}{\eta-2})^{-(\eta+1)/2},$$

  where $2 < \eta < \infty$. The equivalent of this density in `R` using the `sgt` package is:

  rsgt(**n** = No. of samples, **mu** = Mean, **sigma** = Standard Deviation, **lambda = 0** (Skewness parameter), **p = 2**, **q** = Free parameter, controls kurtosis),

where $1 < q < \infty$ and $-1 < \lambda < 1$. This distribution belongs to the Skewed Generalized T Distribution which is a 5 parameter distribution. By setting $p = 2$ and $\lambda = 0$ we obtain Student's - t distribution. The correspondence between Hansen's parametrisation and `sgt`'s is setting ***sigma*** $= 1$ to obtain unit variance and $\eta = 2q$.

- **Skewed t distribution**: This appears in Section 4.3 and in Sections 5.6, 5.7 and 5.8. The parametrisation of the Skewed t density was first introduced by Hansen (1994) and served as a more flexible family of probability density functions:

$$g(z|\eta,\lambda) = \begin{cases} bc\left(1+\frac{1}{\eta-2}\left(\frac{bz+\alpha}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & z < -\alpha/b, \\ bc\left(1+\frac{1}{\eta-2}\left(\frac{bz+\alpha}{1-\lambda}\right)^2\right)^{-(\eta+1)/2} & z \geqslant -\alpha/b, \end{cases} \tag{4.6}$$

where $2 < \eta < \infty$ and $-1 < \lambda < 1$. The α, b and c are constants given by:

$$\alpha = 4\lambda c\left(\frac{\eta-2}{\eta-1}\right), \tag{4.7}$$

$$b^2 = 1 + 3\lambda^2 - \alpha^2, \tag{4.8}$$

$$c = \frac{\Gamma(\frac{\eta+1}{2})}{\sqrt{\pi(\eta-2)}\Gamma(\frac{\eta}{2})}. \tag{4.9}$$

This is implemented in `R` using the `sgt` package:

rsgt(**n** = No. of samples, **mu** = Mean, **sigma** = Standard Deviation, **lambda** = Skewness parameter, **p = 2**, **q** = Free parameter).

4.3 Illustration of Factor Copulas

A simple yet illuminating example of the flexibility of the Factor Copula models is the following. A random sample of size $N = 800$ is drawn from bivariate distributions that have been constructed using four different Factor Copulas. The marginal distributions are $F_i \sim N(0,1)$, $i = 1, 2$. The variances of the latent variables Z_1 and ϵ are set to be 1, $\sigma^2_{Z_1} = \sigma^2_{\epsilon_i} = 1$ and the loading weights β_{11} and β_{21} are selected in a way that the common factor Z_1 accounts for 0.7 of the variance of both X_1 and X_2 and the correlation between X_1 and X_2 is approximately 0.7.

The four Factor Copulas are then generated in a way that for each of them the distributions of Z_1 and ϵ_i are:

Distributions in Factor Copulas	
Z_1	ϵ_i
1. Normal($\mu = 0, \sigma^2 = 1$)	Normal($\mu = 0, \sigma^2 = 1$)
2. Student's-t($\nu = 4$)	Student's-t($\nu = 4$)
3. Skewed Normal(λ = -0.9)	Normal($\mu = 0, \sigma^2 = 1$)
4. Skewed t($\nu = 4, \lambda$ = -0.9)	Student's-t($\nu = 4$)

Table 4: Distributions of the latent variables used in the Factor Copulas.

The scatterplots of the Factor Copulas are presented in Figure 3. The first bivariate distribution implies a Gaussian copula which is a symmetric copula with zero tail dependence. In the second case (top - right), the copula is also symmetric but with non-zero tail dependence. In the third

case (bottom - left), the copula generates asymmetric dependence but zero tail dependence and in the fourth case , the copula generates both asymmetric dependence and lower tail dependence. Generally, when the distributions of the latent variables are either Normal or Skewed Normal we observe zero tail dependence between the variables. When a symmetric distribution with heavier tails is used, such as Student's - t with $\nu = 4$ degrees of freedom, the number of samples in both the upper and lower tail (in this case, samples in the $[-\infty, -2]^2$ and $[2, +\infty]^2$) increases. Finally, when the skewness parameter λ of the Skewed distributions is negative, the clustering of the samples in the lower tail is more potent than in the upper tail, especially when distributions with heavy tail are used, and the higher the value of λ, in absolute terms, the more significant the effect.

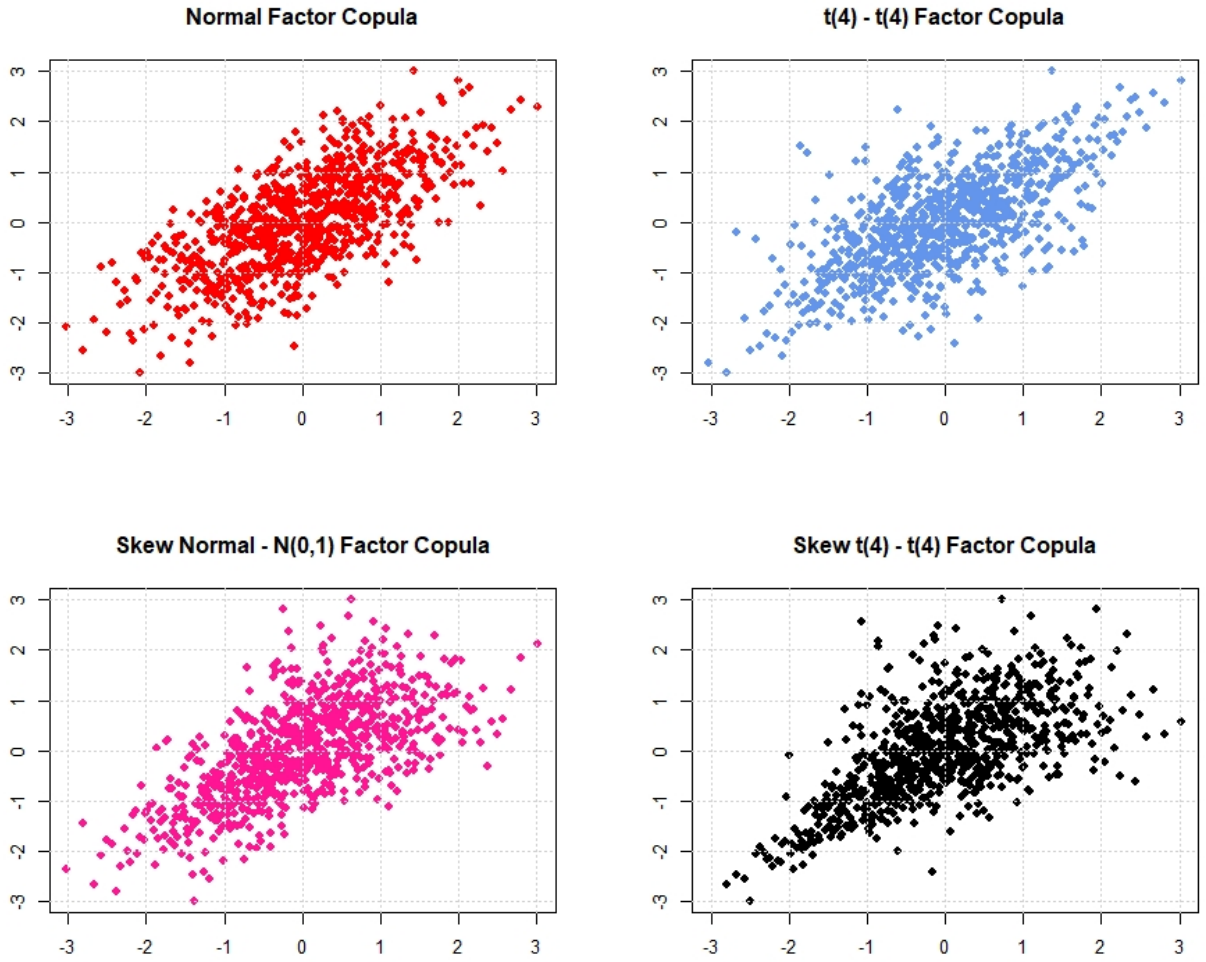

Figure 3: Scatterplots of samples from four different Factor Copulas, with N(0,1) margins and with linear correlation 0.7.

4.4 Non Linear Copulas

The family of Factor Copula models (4.3) can be generalised by including a non-linear and non-additive "link" function. An example could be the model with the following structure:

$$\begin{aligned} &X_i = h(Z_1, \ldots, Z_K, \epsilon_i),\ i = 1, \ldots, N, \\ &Z_k \sim F_{Z_k},\ \epsilon_i \overset{iid}{\sim} F_\epsilon \text{ and } Z_k \perp \epsilon_i\ \forall k, i, \\ &\boldsymbol{X} \sim F_X = \boldsymbol{C(G_1, \ldots, G_N)}, \end{aligned} \tag{4.10}$$

where $h(Z_1, \ldots, Z_K, \epsilon_i) : \mathbb{R}^N \longrightarrow \mathbb{R}$ is an non-linear and non-additive function. These copulas do not typically have closed-from likelihoods and simulation based methods are required to estimate them. However, in some cases, copulas with closed-form densities can be produced through certain combinations of the distributions F_{Z_i} and F_{ϵ_i} with an appropriate "link" function h (see Oh and Patton, 2017, Chapter 2.5).

4.5 Simulated Method of Moments

A popular estimation method for Factor Copula models, proposed by Oh and Patton (2013), is based on another estimation method call Simulated Methods of Moments (SMM) (see McFadden, 1989; Pakes and Pollard, 1989). This is a simulation-based estimation method that shares features with the SMM and can be implemented for Factor Copulas when the marginal distributions are estimated through the Empirical Distribution Function (EDF). The "moments" used in this estimator are functions of rank statistics and they are only affected by changes on the copula and not the marginal distributions.

Let us assume that we have observed data of length T from a multivariate distribution $F_X(\theta)$ with marginal distributions $G_i(\theta)$ and copula $C(\theta)$. Let us also consider m_T to be a vector of dependence measures from the observed data and $\tilde{m}_S(\theta)$ a vector of the same dependence measures computed through S simulations from simulated data from the multivariate distribution $F_X(\theta)$ using the simulated parameter θ from its prior, $p(\theta)$.

Then we define the difference $g_{T,S}(\theta) = m_T - \tilde{m}_S(\theta)$. The SMM estimator of θ is:

$$\hat{\theta}_{T,S} = arg\min_\theta Q_{T,S}(\theta), \tag{4.11}$$

with

$$Q_{T,S}(\theta) = g'_{T,S}(\theta)\, \hat{W}_T\, g_{T,S}(\theta), \tag{4.12}$$

where $\hat{W}_T$ is a weight matrix, which in this case is either the Identity matrix or the Efficient Weight matrix defined in Section 5. The weak consistency and the asymptotic normality of the SMM estimator is also proved by Oh and Patton (2013).

Weak Consistency
For the assumptions mentioned in (Oh and Patton, 2013, Chapter 2.2), the SMM estimator, $\hat{\theta}_{T,S}$, converges in probability to the true value of the parameter, θ_0:

$$\hat{\theta}_{T,S} \xrightarrow{p} \theta_0,\ \text{as } T, S \to \infty. \tag{4.13}$$

Asymptotic Normality

For the assumptions mentioned in (Oh and Patton, 2013, Chapter 2.3), it is proven that the SMM estimator is asymptotically normally distributed:

$$\frac{1}{\sqrt{1/T + 1/S}}(\hat{\theta}_{T,S} - \theta_0) \xrightarrow{d} N(0, \Omega_0), \text{ as } T, S \to \infty, \tag{4.14}$$

where

$$\Omega_0 = (G_0' W_0 G_0)^{-1} G_0' W_0 \Sigma_0 W_0 G_0 (G_0' W_0 G_0)^{-1},$$

with $\Sigma_0 = avar(\hat{m}_T)$, $G_0 = \bigtriangledown_\theta g_0(\theta_0)$, $g_0(\theta) = \lim_{T,S\to\infty} g_{T,S}(\theta)$ and $W_0 = \lim_{T\to\infty} \hat{W}_T$.

5 Estimating Factor Copulas with ABC

5.1 Dependence Measures

The summary statistics used in the ABC algorithm are chosen to be dependence measures that are not affected by different marginal distributions but only from changes in the copula. The chosen dependence measures are the pair-wise Spearman's rank correlation and quantile dependence evaluated at the 5%, 10%, 90% and 95% quantiles, as they are only functions of the multivariate distribution. Pearson's linear correlation is not considered as it is affected by changes in the marginal distributions of the data (Oh and Patton, 2017). This give us five dependence measures for each pair of variables (X_i, X_j).

These measures for the pair (x_i, x_j) are defined as in Nelsen (2010):

$$\rho_{ij} = 12E[F_i(x_i)F_j(x_j)] - 3 = 12 \iint uvC_{ij}(u,v)\,du\,dv - 3, \tag{5.1}$$

$$\lambda_{ij}^q = \begin{cases} P[F_i(x_i) \leqslant q, F_j(x_j) \leqslant q] = \frac{C_{ij}(q,q)}{q}, & \text{if } q \in (0, 0.5], \\ P[F_i(x_i) > q, F_j(x_j) > q] = \frac{1-2q+C_{ij}(q,q)}{1-q}, & \text{if } q \in (0.5, 1), \end{cases} \tag{5.2}$$

where C_{ij} is the copula of (x_i, x_j).

The sample rank correlation and quantile dependence are computed using the following formulas:

$$\hat{\rho}_{ij} = \frac{12}{T}\sum_{t=1}^{T} \hat{F}_i(\hat{x}_i)\hat{F}_j(\hat{x}_j) - 3, \tag{5.3}$$

$$\hat{\lambda}_{ij}^q = \begin{cases} \frac{1}{Tq}\sum_{t=1}^{T} 1\{\hat{F}_i(\hat{x}_i) \leqslant q, \hat{F}_j(\hat{x}_j) \leqslant q\}, & \text{if } q \in (0, 0.5], \\ \frac{1}{T(1-q)}\sum_{t=1}^{T} 1\{\hat{F}_i(\hat{x}_i) > q, \hat{F}_j(\hat{x}_j) > q\}, & \text{if } q \in (0.5, 1), \end{cases} \tag{5.4}$$

where $\hat{F}_i(z) = \frac{1}{T+1}\sum_{t=1}^{T} 1\{\hat{x}_{it} \leqslant z\}$.

5.2 Dependence Measures as Summary Statistics

The summary statistics used for the ABC are the 5 dependence measures described above. Let us denote the summary statistics of the sample $\{\hat{x}_t\}_{t=1}^T$ as $\hat{m}_T$ and those obtained by simulations from $F_X(\theta)$ as $\hat{m}(\theta)$. Both of these are vectors of dimension 5. We define the difference between the two vectors of the summary statistics as:

$$g_T(\theta) = \hat{m}_T - \hat{m}(\theta). \tag{5.5}$$

Then we apply the ABC rejection algorithm to obtain a sample from the approximation of the posterior distribution of the parameter θ. More precisely, we set a threshold of acceptance ϵ on $Q_T(\theta)$, which is defined as:

$$Q_T(\theta) = g_T'(\theta)\,\hat{W}_T\,g_T(\theta), \tag{5.6}$$

where $\hat{W}_T$ is either $\hat{W}_T = I$, the identity matrix, or $\hat{W}_T = \Sigma_{T,B}^{-1}$, the efficient weight matrix described below. In practice, we accept only a percentage of the sample from the prior of θ

which is the equivalent of setting ϵ to be equal to a quantile of $Q_T(\theta)$, such as 5% or 1%. For larger values of ϵ there might be a bias towards the prior introduced to the posterior approximation (Beaumont et al., 2002). As ϵ tends zero, the bias introduced also decreases, but the proportion of simulated values also decreases. Therefore, a larger number of parameter values will need to be simulated to accept the desired number of values which increases the computational complexity.

5.3 Bootstrap

The efficient weight matrix, $\hat{\Sigma}_{T,B}^{-1}$, depends on the sample dependence measures (see Section 5.1) and is estimated using independent and identically distributed bootstrap replications with the following steps:

1. Sample with replacement from the observed data, $\{\hat{x}_t\}_{t=1}^T$, to obtain a bootstrap sample, $\{\hat{x}_t^{(b)}\}_{t=1}^T$, $b = 1, \ldots, B$,
2. compute the summary statistics $\hat{m}_T^{(b)}$ using $\{\hat{x}_t^{(b)}\}_{t=1}^T$,
3. compute the sample covariance matrix of $\hat{m}_T^{(b)}$, for $b = 1, \ldots, B$, and scale it by sample size:

$$\hat{\Sigma}_{T,B} = \frac{T}{B} \sum_{b=1}^{B} (\hat{m}_T^{(b)} - \hat{m}_T)(\hat{m}_T^{(b)} - \hat{m}_T)'. \tag{5.7}$$

In cases where the summary statistics are of similar scale, the weight matrix can simply be the identity matrix. In cases where the scale of the summary statistics differs significantly, e.g. by orders of magnitude, some of them might dominate or be dominated by others in the computation of $Q_T(\theta)$. By using $\hat{W}_T = \hat{\Sigma}_{T,B}^{-1}$, weights are assigned to the summary statistics in order for them to contribute similarly to $Q_T(\theta)$.

5.4 Normal Factor Copula for Two Variables

The first and simplest Factor Copula model that is considered is that with two variables X_1 and X_2, both of which depend on a common variable Z that has a standard Normal distribution. Each X_i has a different idiosyncratic term ϵ_i that also follows a Normal distribution with $\sigma_{\epsilon_i} = 1/2$ but the same factor loadings $\beta_{11} = \beta_{21} = \beta$. The "equidependence" assumption is made when using this model which is similar to the equicorrelation assumption of the equicorrelation model of Engle and Kelly (2012). The "equidependence" assumption is that the pair-wise dependence of the variables is the same at all times and arises from the fact that we assume that all variables have one common latent variable. The structure of the model is presented in (5.8):

$$\begin{aligned} &X_i = \beta Z + \epsilon_i, \ i = 1, 2, \\ &Z \sim N(0, 1), \ \epsilon_i \overset{iid}{\sim} N(0, \{1/2\}^2), \ i = 1, 2, \\ &\text{with } Z \perp \epsilon_i, \text{ for } i = 1, 2. \end{aligned} \tag{5.8}$$

The copula of this model is the Gaussian copula and therefore has an analytic form. In this simplest case, we will make use of the ABC rejection algorithm to estimate the parameter β,

which essentially controls the dependence between the variables X_1 and X_2. The true value of β is $1/\sqrt{2}$ so that the marginal distributions of X_i are standard Normal and the correlation between them is approximately $1/2$.

The prior on β is set to be an uninformative, uniform distribution $U(0,1)$, which can produce correlation values, $\rho(X_1, X_2)$, between 0 and 66% but its range can be expanded to allow for different values of the correlation coefficient. It should also be mentioned that this prior is chosen to deal with the fact that ρ is a function of β^2 (5.9) which means that there would be an identifiability issue for β the prior allowed for non-negative values.

$$\rho(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1)}\sqrt{Var(X_2)}} = \frac{\beta^2}{\beta^2 + \frac{1}{2}} \tag{5.9}$$

We explore the cases where $N = \{1000, 10000, 100000\}$ samples are drawn from the prior distribution of β. The threshold of acceptance ϵ for the ABC algorithm is set to be equal to the 10%, 5%, 1% and 0.5% quantiles. For the case of $N = 10000$ samples, both Table 5 and Figure 4 suggest that the identity and the efficient weight matrix produce similar results in approximating the posterior distribution and estimating β. The estimation error is approximately 0.24% when the acceptance threshold is set to be the 1% quantile and the identity matrix is used. When the efficient weight matrix is used, the estimation error of β is 0.03% at the 0.5% level.

Identity matrix				
Accepted Samples	**10%**	**5%**	**1%**	**0.5%**
Posterior mean (Bayes' estimator)	0.7029	0.7014	0.7088	0.7124
True Value	0.7071	0.7071	0.7071	0.7071
Absolute Error	0.0042	0.0057	0.0017	0.0053
Error (%)	0.59	0.81	0.24	0.75

Efficient Weight matrix				
Accepted Samples	**10%**	**5%**	**1%**	**0.5%**
Posterior mean (Bayes' estimator)	0.7017	0.7036	0.7088	0.7069
True Value	0.7071	0.7071	0.7071	0.7071
Absolute Error	0.0054	0.0035	0.0017	0.0002
Error (%)	0.76	0.50	0.24	0.03

Table 5: Estimation results for the Normal Factor Copula model for $N = 10000$.

Similar is the case of $N = 100000$ samples, but the estimation results are more prominent. The posterior distribution concentrates more accurately around the true value of β. The estimation error remains at similar levels as in the previous case but the variability of the distribution is significantly smaller.

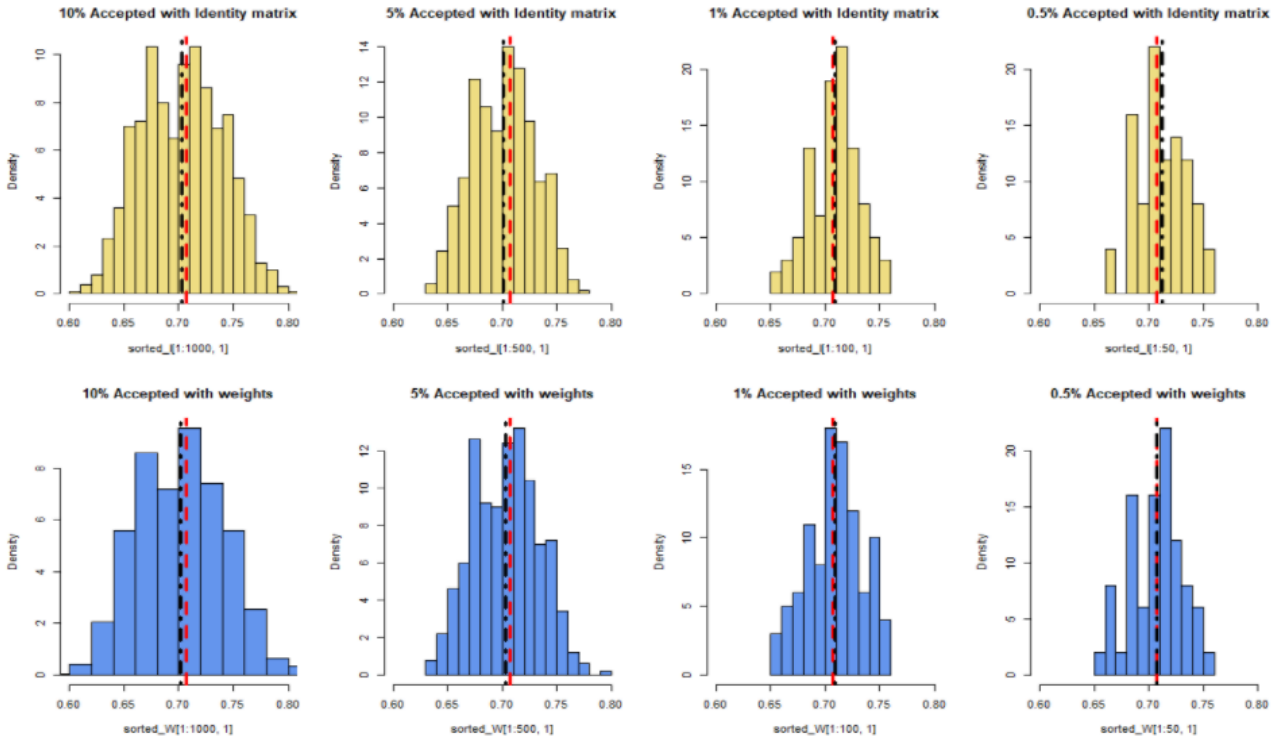

Figure 4: Approximation of the posterior distribution of β using the identity and the efficient weight matrix for $N = 10000$ and with acceptance threshold set to be the 10%, 5%, 1% and 0.5%. (The red line denotes the true value and the black the posterior mean).

5.5 Comparison: Identity vs Efficient Weight matrix

To highlight the effectiveness of the efficient weight matrix we create the following example. In this simulation, the 5% quantile dependence statistic is re-scaled in a way that it is approximately 2 orders of magnitude larger than the rest of the summary statistics. This means that when the identity matrix is used, the difference between the 5% quantile dependence of the sample and the simulations would disproportionately affect $Q_T(\theta)$. This is captured by the efficient weight matrix (Table 6) which assigns a much smaller weight to this summary statistic compared to the rest.

Summary Statistics	**Rank Correlation**	**Re-scaled 5% q.d.**	**10% q.d.**	**90% q.d.**	**95% q.d.**
Rank Correlation	1.6393	-0.0002	-0.1411	-0.2117	-0.0359
Re-scaled 5% q.d.	-0.0002	0.00004	-0.0025	0.0001	-0.0006
10% q.d.	-0.1411	-0.0025	0.7480	0.0268	-0.0082
90% q.d.	-0.2117	0.0001	0.0268	0.8027	-0.2785
95% q.d.	-0.0359	-0.0006	-0.0082	-0.2785	0.4037

Table 6: Efficient weight matrix of the summary statistics. (q.d.: quantile dependence).

The Normal Factor Copula model described in Section 5.4 is used to estimate β using the

Identity matrix and the Efficient Weight matrix. The sample size is $N = 10000$ and the acceptance threshold is set equal to the $10\%, 5\%, 1\%$ and 0.5% quantiles. The results of the approximated posterior distributions of β are presented in Figure 5. It is observed that when using a stricter threshold the posterior distribution concentrates more around the true value of β, as less bias is introduced, and the sample mean approaches the true value of the parameter. The estimation becomes a lot more efficient in the case of the efficient weight matrix, as the approximation of the posterior distribution converges quicker to the true value and the variance of the distribution is smaller for each of the four thresholds used in Table 7. The estimation error for the efficient weight matrix is 27% - 57% smaller than that of the identity matrix.

Identity matrix				
Accepted samples	**10%**	**5%**	**1%**	**0.5%**
Posterior mean (Bayes' estimator)	0.6944	0.6931	0.6948	0.6925
True Value	0.7071	0.7071	0.7071	0.7071
Absolute Error	0.0127	0.0140	0.0123	0.0146
Error (%)	1.80	1.98	1.74	2.07
Standard Deviation*	4.84	4.04	3.05	3.72

Efficient Weight matrix				
Accepted samples	**10%**	**5%**	**1%**	**0.5%**
Sample mean (Bayes' estimator)	0.7017	0.7011	0.6982	0.6996
True Value	0.7071	0.7071	0.7071	0.7071
Absolute Difference	0.0054	0.0060	0.0089	0.0075
Error (%)	0.76	0.85	1.26	1.06
Standard Deviation*	3.10	2.05	1.33	1.41

*Multiplied by 10^2.

Table 7: Comparison of parameter estimation results for the Normal Factor Copula when using the identity matrix and the efficient weight matrix for $N = 10000$ and with summary statistics on different scales.

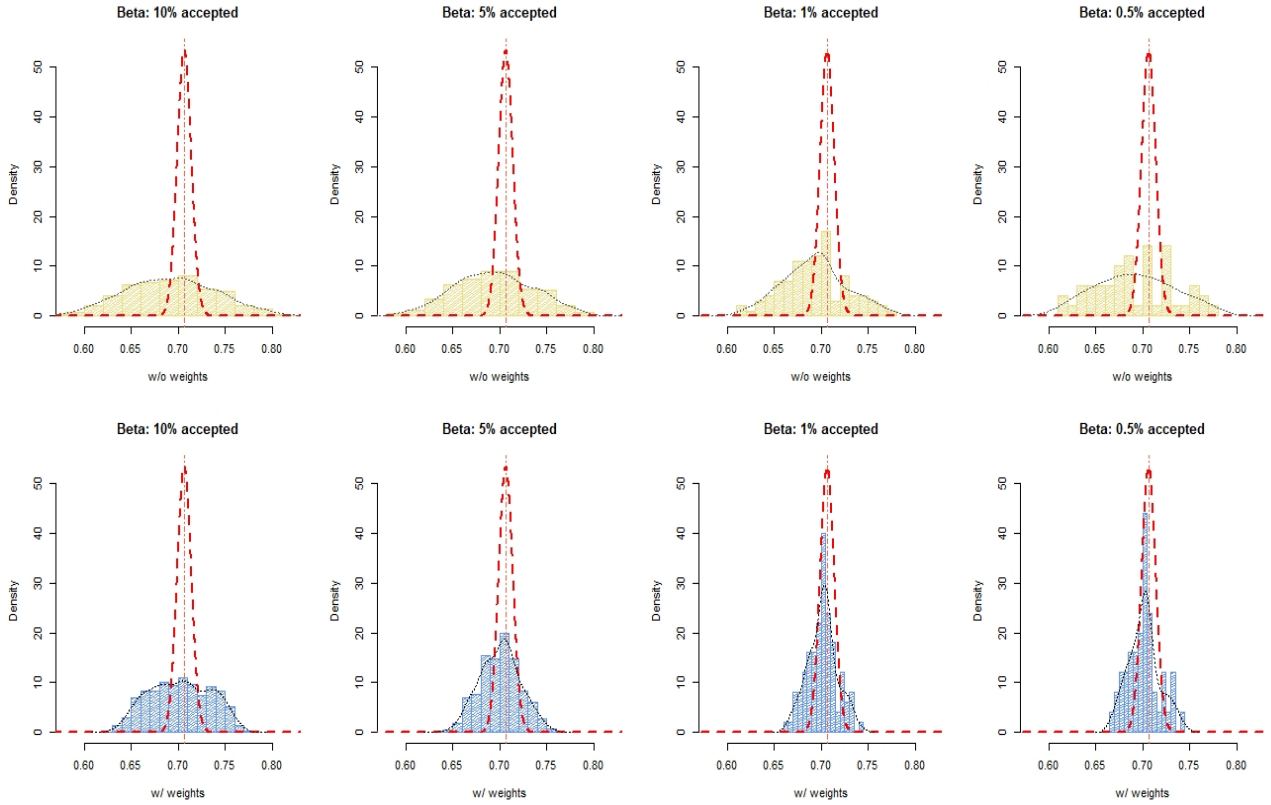

Figure 5: Approximations of the posterior distribution of β using the identity matrix (first row) against using the efficient weight matrix (second row), with $N = 10000$ and with summary statistics on different scales. (The red line is the posterior distribution of β evaluated analytically and the black is the approximation of the ABC algorithm).

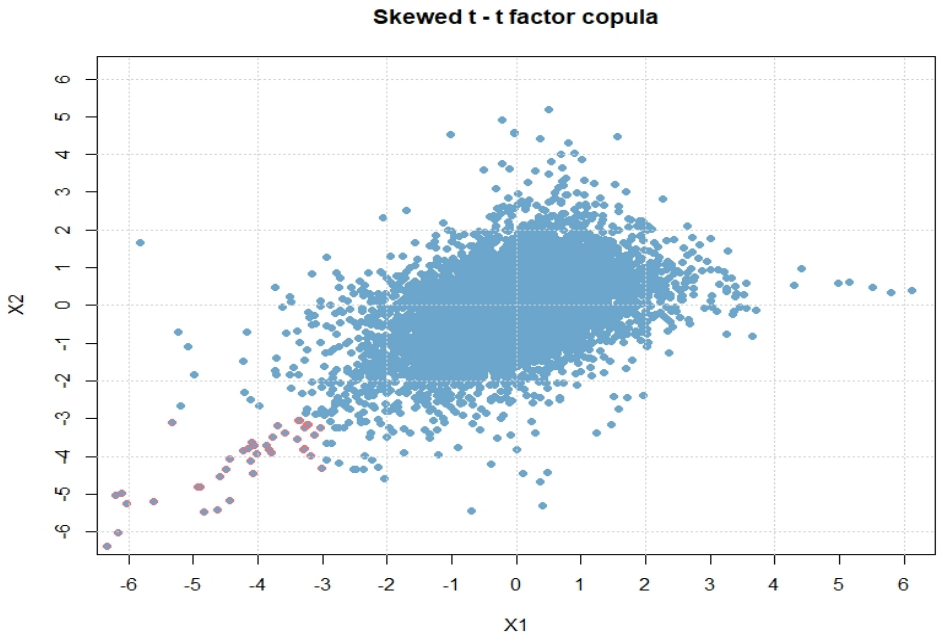

Figure 6: Simulations from the Skewed t-t Factor Copula with correlation 1/2 between the variables X_i, negative skewness $\lambda = -1/2$ and $\nu = 4$ degrees of freedom. The presence of asymmetry and lower tail dependence are obvious compared to the Normal Factor Copula. The simulated data points in the lower tail are highlighted in red.

5.6 Skewed t-t Factor Copula for Two Variables

The next step is to increase the complexity of the Factor Copula model in order to investigate further the parameter estimation using the ABC rejection algorithm. In the case of the Normal Factor copula (see Section 5.4) we have some "nice" properties such as symmetry and zero tail dependence which make the estimation of β simpler. For example, the quantile dependence statistics for $q\%$ and $(1-q)\%$ were theoretically the same due to the symmetry of the model. The Skewed t – t model captures both asymmetry and tail dependence which is desirable when modelling financial data Oh and Patton (2017). The added flexibility of this model comes at the expense of more estimable parameters. Except for the β parameter which controls the dependency between the variables X_i, we also have to estimate the parameters λ and q. The skewness parameter λ allows for asymmetry in the model while q is a function of the degrees of freedom of the distribution, $\nu = 2q$ (see Section 4.2), which impacts the thickness of the tails, thus allowing for tail dependence. The structure of the Factor Copula is the following:

$$\begin{aligned} X_i &= \beta Z + \epsilon_i, \ i = 1, 2, \\ Z &\sim \text{ Skewed t}(\lambda, \nu), \\ \epsilon_i &\overset{iid}{\sim} t(\nu), \\ &\text{with } Z_k \perp \epsilon_i \text{ for } i = 1, 2. \end{aligned} \tag{5.10}$$

The true values of the model parameters that were used to generate the observed data for this example are the following:

Parameters	β	λ	ν
Values	$1/\sqrt{2}$	-1/2	4

The means of the latent variables are both set to zero and their variances are $\sigma_Z^2 = 1$ and $\sigma_{\epsilon_i} = 1/2$. The correlation between the variables X_i is 1/2. The degrees of freedom are $\nu = 4$ and $q = 2$ which appears in the estimation results below, which implies heavy tails in the distributions and strong tail dependency in the copula. The negative value of λ produces lower tail dependence and in financial terms this could be interpreted as: "crashes tend to correlate".

The β prior is once again set to be $U(0, 1)$. The Skewed t density (4.6) is a function of λ^2 which produces an identifiability issue and we can only estimate λ up to sign. To overcome this problem we set the λ prior to be $U(-1, 0)$ instead of $U(-1, 1)$ that would cover the whole range of its values. As for q, three different cases are considered:

Prior distributions of q
1. $q \sim U(1, 10)$,
2. $q \sim U(1, 50)$,
3. $q \sim f(q) = \begin{cases} q = 3, & prob = 1/2, \\ q = 30, & prob = 1/2 \end{cases}$.

Table 8: Prior distributions for the parameter q.

In the first case only a narrow range of values for q are considered. In the second case, a more broad range of values is considered which produce distributions with significantly different tail behavior. The last case is the simplest since one of the two values is selected; $q = 3$ represents

stronger tail dependence while $q = 30$ represents lighter tail dependence. The estimation process will determine whether the model favors "heavy" against "light" tails in the Skewed t and Student's t distributions which can be interpreted as if the model prefers strong or weak tail dependence between variables.

Accepted samples: 0.5%			
Parameters	β	λ	q
Posterior mean (Bayes' estimator)	0.6825	-0.7316	6.2851
True Values	0.7071	-0.50	2
Absolute Error	0.0246	0.2316	4.2851
Error (%)	3.48	46.32	214.26

Accepted samples: 0.1%			
Parameters	β	λ	q
Posterior mean (Bayes' estimator)	0.6691	-0.5516	2.4327
True Values	0.7071	-0.50	2
Absolute Error	0.0310	0.0516	0.4327
Error (%)	5.37	10.32	21.64

Table 9: Estimation results for $N = 50000$ samples of the Skewed t-t Factor copula (Section 5.6) using the Identity matrix.

Three different scenarios are explored for a sample of $N = \{1000, 10000, 50000\}$. In the case of $N = 10000$, the approximation of the posterior distribution of β seems to concentrate around the true value but at a slower pace compared to the Normal factor copula (see Section 5.4). The algorithm rejects the higher values of λ i.e. the smaller values in absolute terms. However, it only meaningfully converges around the true value in the case of $N = 50000$ samples. That means that the algorithm suggests that asymmetry is present. As for q, when either of the first two priors is used, higher values are rejected by the algorithm as the posterior slowly concentrates to the lower values which include its true value and suggest heavier tails in the distributions or stronger dependence between the variables. This becomes more clear in the case of $N = 50000$ samples. When the third prior is used, the "small" value is accepted more frequently at the $N = 10000$ samples and almost exclusively in the case of $N = 50000$ samples. The estimation results for the case where the prior for q is $U(1, 50)$ are presented in Figure 7 and in Table 9.

Less bias is introduced in the approximation of the posterior distributions of the parameters as the acceptance threshold becomes stricter, i.e. less simulated values are accepted. Therefore the posteriors concentrate around the true value of the parameter without much variation. In the case of $N = 50000$ samples and for the acceptance threshold set to be the 0.1% quantile, the estimation error is almost 5.4% for β. This implies correlation 47.2% which is close to the true correlation of 50% between the variables. As for λ the estimation error is 10.3% and based on the accepted values the algorithm suggests strong asymmetry. As for q, the estimation error is 21.6%, $\hat{\nu} = 4.87$ compared to the true value $\nu = 4$, which suggests the presence of fat tails in the distributions and strong tail dependence in the copula. It should also be noted that the results presented in this Section are for the Skewed t-t Factor Copula model using the identity matrix. The results are similar in the case when the Efficient Weight matrix is used.

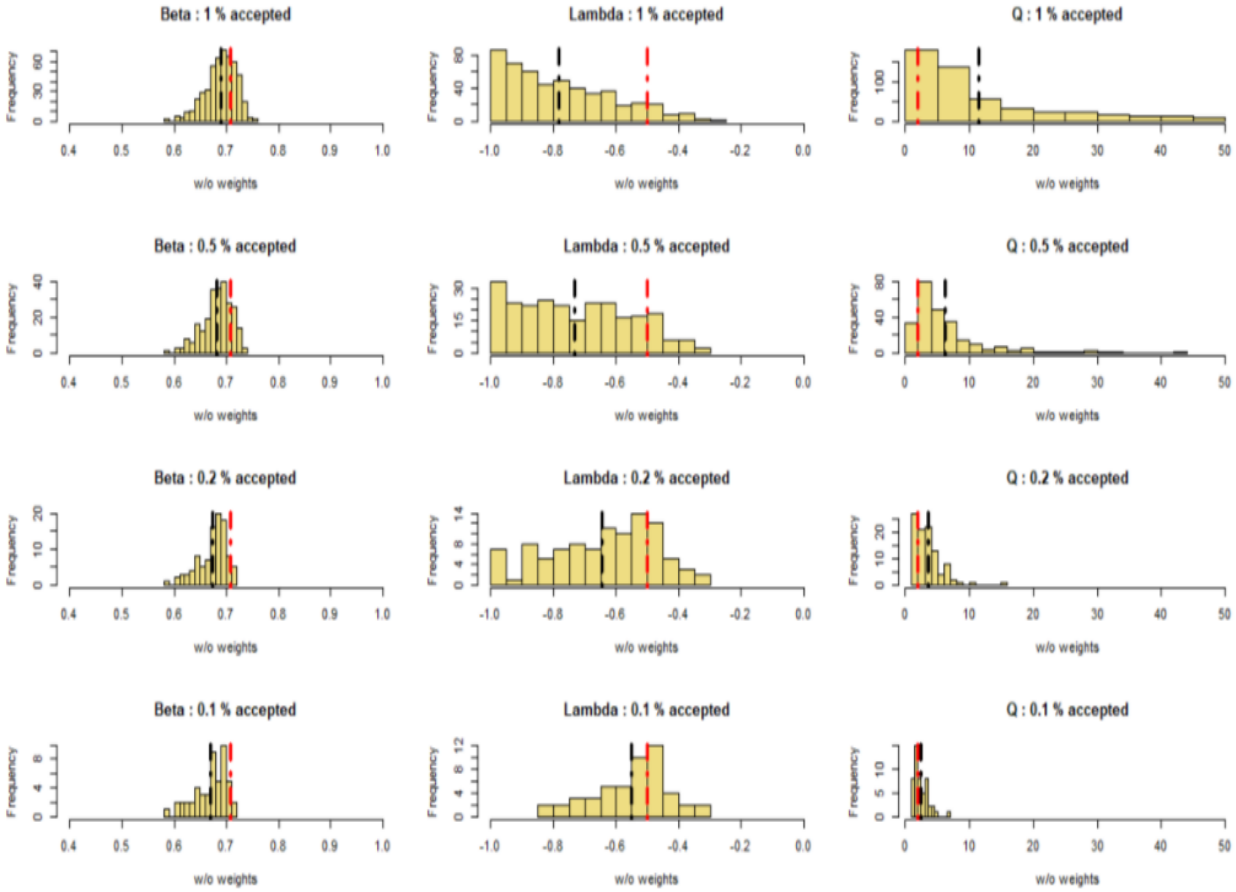

Figure 7: Approximations of the posteriors of β, λ and q of the Skewed t-t model (5.10), using the identity matrix for $N = 50000$ and for acceptance thresholds set to be the $1\%, 0.5\%, 0.2\%$ and 0.1% quantiles. (The red line denotes the true value and the black the posterior mean).

5.7 Skewed t-t Factor Copula for Three Variables

The next step is to repeat the estimation process of Section 5.6 but with three variables. The number of parameters and the parameters themselves remain the same as in the previous Section. The model's structure is the same as in (5.10) but with $i = 1, 2, 3$. A new challenge that arises is how to treat the summary statistics, since we have 3 pairs of variables for each of the 5 statistics which gives us a total of 15 summary statistics. Two different approaches were investigated:

1. average across each of the five summary statistics for all three pairs and end up with a vector $m^*(\theta)$ of dimension 5,

2. maintain all the pair-wise statistics in a vector $\hat{m}(\theta)$ of dimension 15.

The averaging is performed in the following way: Let us assume s_{ij} represents one of the summary statistics, i.e. Rank Correlation, 5%, 10%, 90% or 95% Quantile Dependence, between the pair of variables (i, j), $i, j = 1, \ldots, N$. The following is the pair-wise matrix for a specific

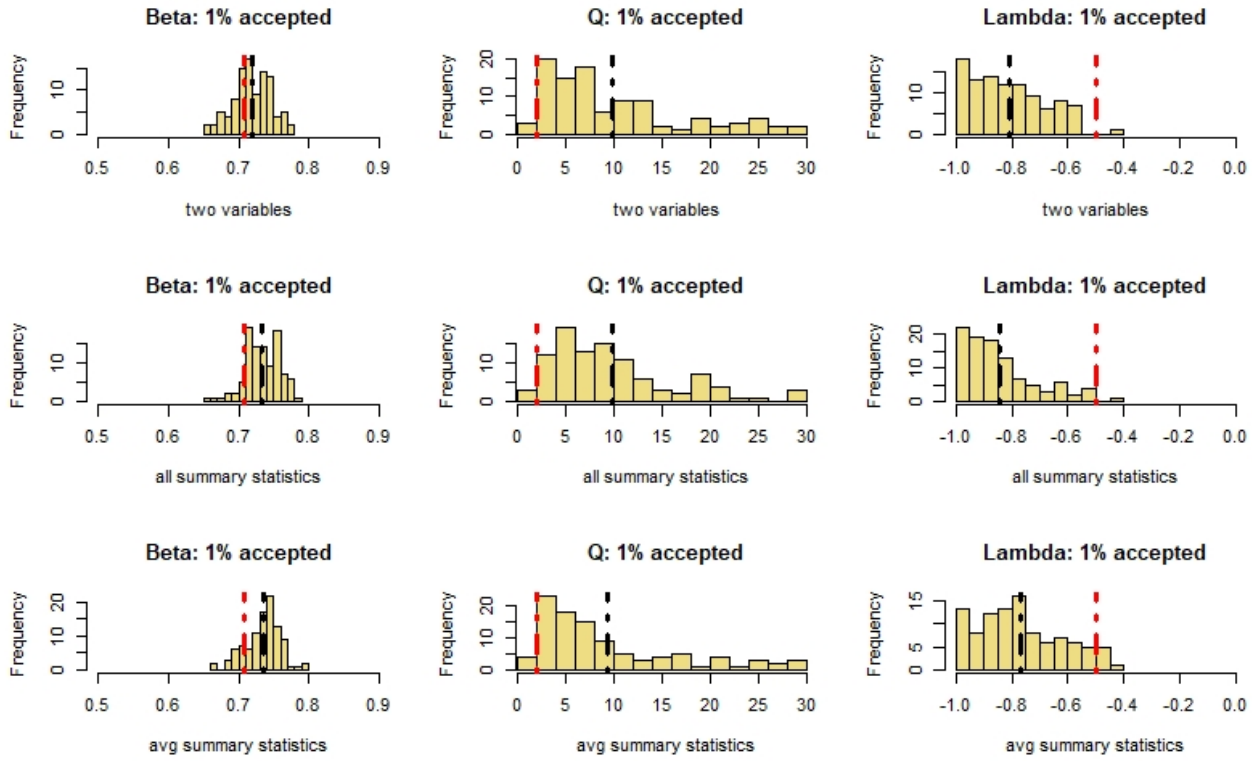

Figure 8: Approximations of the posteriors of β, q and λ using the identity matrix for acceptance threshold the 1% quantile and $N = 10000$ samples. First row: two variable model, Second row: Three variables and all 15 summary statistics, Third row: Three variables and averaged summary statistics. (The red line denotes the true value and the black the posterior mean).

summary statistic:

$$S = \begin{pmatrix} 1 & s_{12} & \cdots & s_{1N} \\ s_{12} & 1 & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1N} & s_{2N} & \cdots & 1 \end{pmatrix}. \tag{5.11}$$

We make use of the "equidependence" property of this model and average the summary statistics across all pairs of variables. In this case, we have $N = 3$ so we reduce the number of summary statistics from 15 to 5:

$$S = \begin{pmatrix} 1 & s_{12} & s_{13} \\ s_{12} & 1 & s_{23} \\ s_{13} & s_{23} & 1 \end{pmatrix}, \tag{5.12}$$

and

$$\bar{s} = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} s_{ij} = \frac{1}{3}(s_{12} + s_{13} + s_{23}). \tag{5.13}$$

So we end up with $m^*(\theta) = [\bar{s}_1, \ldots, \bar{s}_5]$ as we repeat the above process for each of the five summary statistics.

For the case where the summary statistics are averaged and for $N = 10000$ samples, some bias is introduced in the approximate posterior of β. However, the approximations of the posterior

distributions of q and λ concentrate better around the true values of the parameters and the estimation error is reduced. In the second case where all the pair-wise summary statistics are used the approximations of the posteriors resemble those in Section 5.6 for the same number of samples. A comparison of the posterior distributions of the three parameters for $N = 10000$ samples and for the acceptance threshold set equal to the 1% quantile is presented in Figure 8. The first row is from the model used in Section 5.6, the second row is for the Skewed t-t model with three parameters when all 15 summary statistics are used (approach 2 of this Section) and the third row is the same model but with averaging the summary statistics (approach 1 of this Section).

5.8 Multi-Factor Copula

The motivation behind this scenario is a setting where we have N observable random variables, for example stocks, that are divided into M groups, that represent for example different industries. The groups are assumed to be homogeneous in the sense that the variables in the same group exhibit "equidependence", similar to the equicorrelation assumption of the equicorrelation model of Engle and Kelly (2012). The "equidependence" assumption is that in each group the pair-wise dependence is the same at all times.

Therefore we consider the following, more complex, scenario. We want to model nine variables that belong to one of the three groups, hence, $N = 9$ and $M = 3$ with three variables assigned to each group. We now consider a more flexible model that allows for heterogeneous dependence. This is a Factor Copula model with one common factor, Z_0, and a set of group-specific factors Z_S. Each variable also depends on an idiosyncratic variable ϵ_i. The "market-wide" factor Z_0 will allow for asymmetry and kurtosis. The group specific factors Z_S and the idiosyncratic variables ϵ_i are assumed to be symmetric but with flexibility in the tails. More precisely:

$$\begin{aligned} X_i &= \beta_{S(i)} Z_0 + \gamma_{S(i)} Z_{S(i)} + \epsilon_i, \\ &\text{for } i = 1, \ldots, 9 \;\&\; S(i) = \{1, 2, 3\}, \\ Z_0 &\sim \text{ Skewed t}(\lambda, \nu), \\ Z_{S(i)} &\overset{iid}{\sim} t(\nu), \quad S(i) = 1, 2, 3, \\ \epsilon_i &\overset{iid}{\sim} t(\nu), \\ &\text{with } Z_{S(i)} \perp Z_0 \text{ and } \epsilon_i \perp Z_j, \; \forall i, j. \end{aligned} \tag{5.14}$$

The true values of the model parameters that were used to generate the observed data for this example are the following:

Parameters	β_1	β_2	β_3	γ_1	γ_2	γ_3	λ	ν
Values	$1/\sqrt{2}$	$1/3$	0.90	0.40	0.25	0.90	$-1/2$	4

Therefore, there are four latent factors in this model, one common for all variables, Z_0, and three group-specific, Z_S. However, each variable only depends on two of them. The factor loadings are group-specific which means that they are the same for all variables that belong in the same group but can be different for variables from different groups. This results in a "block - equidependence" model (Oh and Patton, 2017). The flexibility of the model to represent different levels of dependence between variables and the "block - equidependence" property is showcased both in Figure 9 and in Table 10 where the scatterplots between some pairs of variables and all the pair-wise correlations are presented.

Correlation	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.00	0.55	0.56	0.28	0.28	0.29	0.47	0.48	0.47
X_2	0.55	1.00	0.55	0.29	0.28	0.29	0.47	0.47	0.48
X_3	0.56	0.55	1.00	0.28	0.28	0.29	0.47	0.48	0.47
X_4	0.28	0.29	0.28	1.00	0.22	0.23	0.28	0.29	0.28
X_5	0.28	0.28	0.28	0.22	1.00	0.22	0.28	0.28	0.28
X_6	0.29	0.29	0.29	0.23	0.22	1.00	0.30	0.30	0.29
X_7	0.47	0.47	0.47	0.28	0.28	0.30	1.00	0.71	0.71
X_8	0.48	0.47	0.48	0.29	0.28	0.30	0.71	1.00	0.72
X_9	0.47	0.48	0.47	0.28	0.28	0.29	0.71	0.72	1.00

Table 10: Pair-wise correlation matrix for the variables X_i. The different values of β_i and γ_i produce different levels of dependence between variables in each of the three groups and between variables from different groups.

As mentioned in Section 2, the larger the number of summary statistics is, the less likely it is to simulate a sample of parameters from their priors that produces summary statistics that are "close" to those of the observed data. This is also known as the "Curse of dimensionality" (Blum, 2017). In this model, the number of pairs of variables is $\binom{9}{2} = \frac{9!}{2!(9-2)!} = 36$ and for each pair we have five summary statistics. This adds up to a vector of summary statistics with dimension 180. In order to obtain a vector of summary statistics of smaller dimension we make use of the "block-equidependence" property. More specifically all variables in the same group m have the "equidependence" property and all pairs of variables for which one of the variables comes from a specific group, r, and the other one from a specific group but different to the first group, t, exhibit the same dependence. This allows us to not only average the intra-group but the inter-group summary statistics as well.

Let us assume that s_{ij} represents one of the summary statistics, i.e. Rank Correlation, 5%, 10%, 90% or 95% Quantile Dependence, between the pair of variables $(i, j),\ i \neq j,\ i, j = 1, \ldots, N$. The following matrix represents the pair-wise matrix for this specific summary statistic:

$$S = \begin{pmatrix} 1 & s_{12} & \cdots & s_{1N} \\ s_{12} & 1 & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1N} & s_{2N} & \cdots & 1 \end{pmatrix}. \tag{5.15}$$

We make use of the "block-equidependence" property. Let us consider the general case where we have N variables, M different groups and k_i variables in each group. The pair-wise matrix $S_{N\text{x}N}$ is then decomposed into $(M\text{x}M)$ sub-matrices S_{ij} of dimensions $(k_i\text{x}k_j)$, according to the M different groups:

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1M} \\ S_{12} & S_{22} & \cdots & S_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ S_{1M} & S_{2M} & \cdots & S_{MM} \end{pmatrix}. \tag{5.16}$$

We create a matrix S^* of the average values:

$$S^* = \begin{pmatrix} s^*_{11} & s^*_{12} & \cdots & s^*_{1M} \\ s^*_{12} & s^*_{22} & \cdots & s^*_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s^*_{1M} & s^*_{2M} & \cdots & s^*_{MM} \end{pmatrix}. \tag{5.17}$$

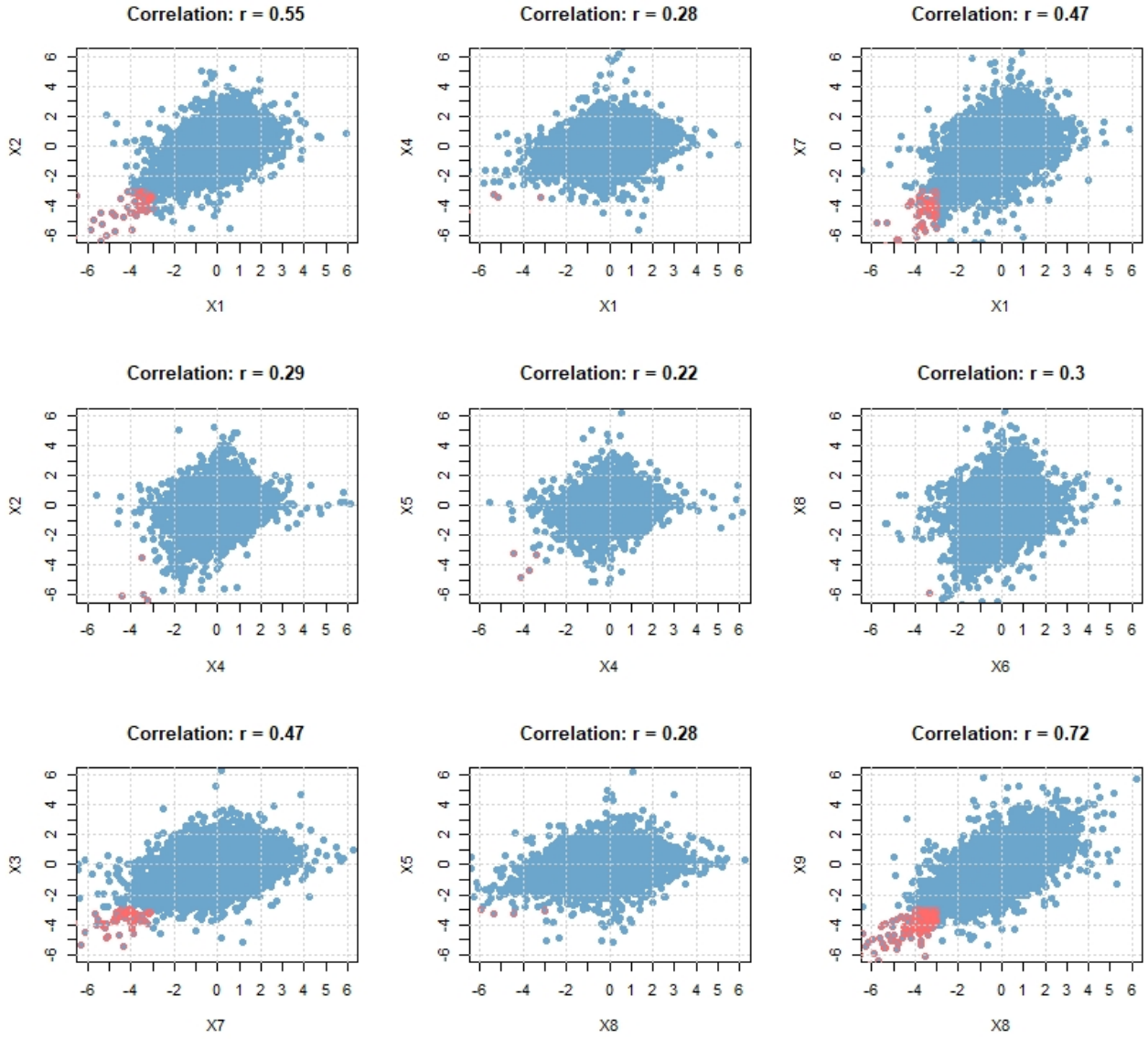

Figure 9: Scatterplots of pairs of variables that belong to the same or different groups and thus exhibit different levels of dependence. The data points in the lower tail are annotated in red to emphasise the asymmetry and the different levels of strength of tail dependence.

The averaging of the values occurs in the following ways: since the diagonal matrices S_{ii} are symmetric, we average the values in the following way:

$$s^*_{ii} = \frac{2}{k_i(k_i-1)} \sum_{i=1}^{k_i-1} \sum_{j=i+1}^{k_i} s_{ij}, \tag{5.18}$$

while for all the non-diagonal matrices S_{rt} with $r \neq t$, we average the values using the following formula:

$$s^*_{rt} = \frac{1}{k_r k_t} \sum_{i=1}^{k_r} \sum_{j=1}^{k_t} s_{ij}. \tag{5.19}$$

Finally, a vector of averaged statistics $[\bar{s}^*_1, \ldots, \bar{s}^*_M]$ is created in the following way:

$$\bar{s}^*_i = \frac{1}{M} \sum_{j=1}^{M} s^*_{ij}, \tag{5.20}$$

which ultimately gives a total of M values for each summary statistic. In our case, we have five summary statistics and three groups, so a vector of 15 summary statistics.

Estimation

This model has eight free parameters that require to be estimated while a model that assumes just one common factor (see Sections 5.6, 5.7) would only have three. The "block - equidependence" assumption, which allows for different levels of dependence between variables, and the chosen distributions of the model's latent variables, which allow for asymmetry and kurtosis, make the model much more flexible while also remaining relatively parsimonious. The model's parameters will be estimated using the ABC Rejection algorithm.

The prior distributions of all β_i and γ_i are set to be $U(0,1)$. The prior for λ is set to be $U(-1,0)$ and for q a discrete uniform distribution with two values, as described in Section 5.6. The estimation results of this model for a sample of $N = 10000$ are presented in Table 11 and the approximations of the posteriors are presented in Figures 10 and 11 with the acceptance threshold set to be the $5\%, 1\%$ and 0.5% quantiles. The posteriors concentrate around the true value for all three β_i, but do not concentrate as quickly for the γ_i's and λ. As for q, it becomes obvious that the smaller value is selected most of the time especially as we decrease the acceptance threshold.

As presented in Table 11, the ABC algorithm suggests "heavy" tails in the distributions of the latent variables 84% of the time, for the 0.5% quantile as the acceptance threshold. Or, alternatively, the observable variables exhibit strong tail dependence and more specifically, lower tail dependence due to the negative asymmetry of the λ parameter. The estimation error for the β_i's varies between 2.9% to 13.1% while that of the γ_i's is between 19% and 56.4%. Similar is the case for λ for which the estimation error is 48.4%.

Accepted samples: 1%

Parameters	β_1	β_2	β_3	γ_1	γ_2	γ_3	λ	q	
Posterior mean	0.791	0.349	0.869	0.498	0.400	0.576	-0.706	Heavy tail	Light tail
True Values	0.707	0.333	0.900	0.400	0.250	0.900	-0.500	74%	26%
Absolute Error	0.084	0.016	0.031	0.098	0.15	0.324	0.206		
Error (%)	11.88	4.80	3.44	24.50	60.00	36.00	41.20		

Accepted samples: 0.5%

Parameters	β_1	β_2	β_3	γ_1	γ_2	γ_3	λ	q	
Posterior mean	0.800	0.355	0.874	0.476	0.391	0.675	-0.742	Heavy tail	Light tail
True Values	0.707	0.333	0.900	0.400	0.250	0.900	-0.500	84%	16%
Absolute Error	0.093	0.0220	0.026	0.076	0.141	0.225	0.242		
Error (%)	13.15	6.61	2.89	19.00	56.40	25.00	48.40		

Table 11: Estimation results of the multi-factor copula model for $N = 10000$ and acceptance thresholds the 1% and 0.5% quantiles.

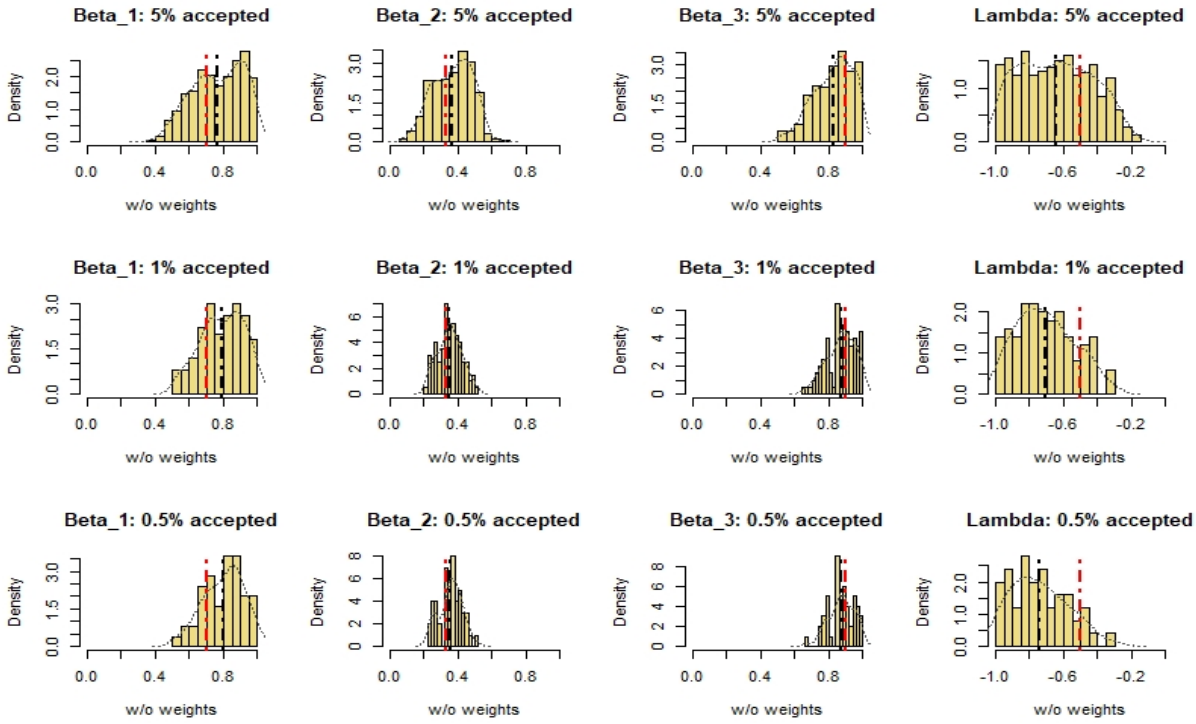

Figure 10: Approximations of the posteriors of β_i and λ using the identity matrix, with the $5\%, 1\%$ and 0.5% quantiles as acceptance thresholds and for $N = 10000$. (The red line denotes the true value and the black the posterior mean).

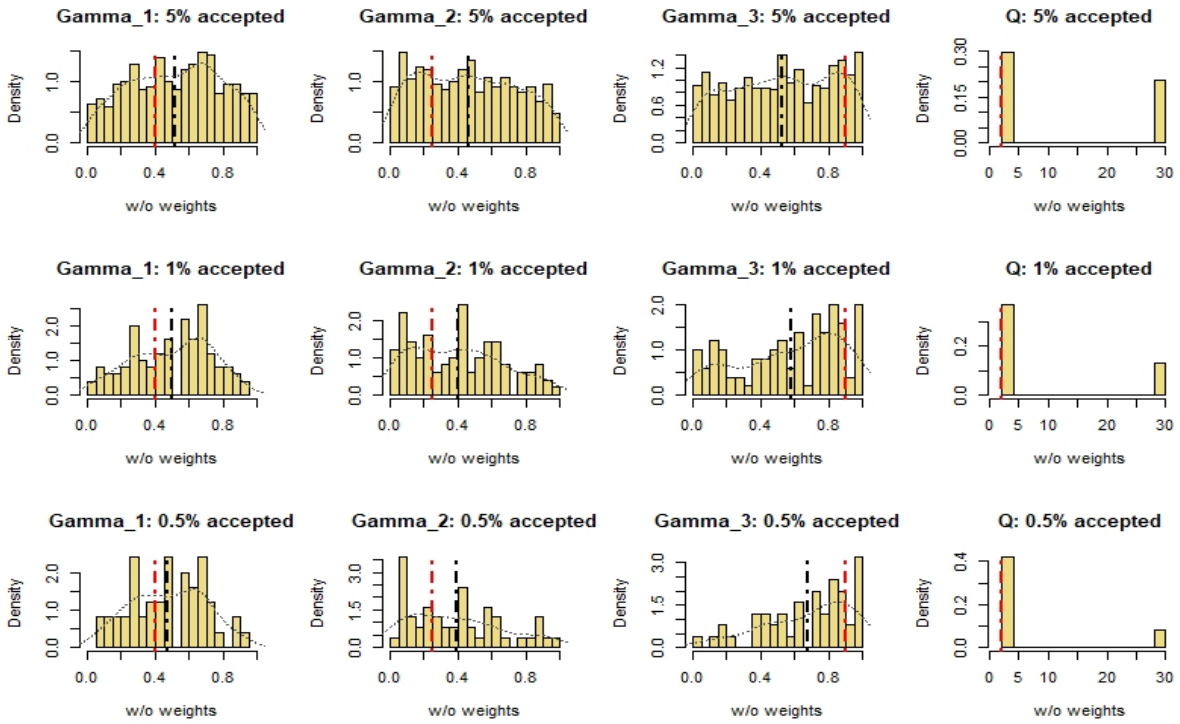

Figure 11: Approximations of the posteriors of γ_i and q using the identity matrix, with the $5\%, 1\%$ and 0.5% quantiles as acceptance thresholds and for $N = 10000$. (The red line denotes the true value and the black the posterior mean).

Accepted samples: 0.2%

Parameters	β_1	β_2	β_3	γ_1	γ_2	γ_3	λ	q	
Posterior mean (Bayes' estimator)	0.745	0.346	0.877	0.394	0.337	0.659	-0.656	Heavy tail	Light tail
True Values	0.707	0.333	0.900	0.400	0.250	0.900	-0.500	85%	15%
Absolute Error	0.038	0.013	0.023	0.006	0.087	0.241	0.156		
Error (%)	5.33	3.93	2.51	1.55	34.73	26.74	31.20		

Table 12: Estimation results of the multi-factor copula model for $N = 50000$ and acceptance thresholds the 0.2% quantile.

Figure 12 shows the approximations of the posterior distributions of the model parameters when the acceptance threshold is set to be the 0.2% quantile which reduces the bias introduced by the algorithm compared to the previous thresholds used. The number of samples is also increased to $N = 50000$. The approximate posterior distributions of the β_i's are very much concentrated around the true parameter values. For the γ_i's, the algorithm accepts values in the right "neighbourhood" but the posterior variance is large compared to the parameters β_i. In the case of λ the algorithm completely rejects values close to zero but a larger number of simulations would be required for the posterior to heavily concentrate around the true value and reduce the variance. Finally, the low value of q is heavily favoured by the algorithm which means that it suggests heavier tails in the distributions and strong tail dependence in the copula.

As shown in Table 12, the estimation error for the β_i's is between 2.51% and 5.33%. The maximum error for the β_i's is reduced by almost 60%. The error for the γ_i's is between 1.55% and 34.73% which means that the maximum error is reduced by more than 38% while it is also worth noting that the minimum estimation error for the γ_i's is reduced by 91%. The error for λ is reduced by 35.5% while the value of q that suggests "heavy" tails is accepted in a slightly higher proportion compared to the case of $N = 10000$ samples.

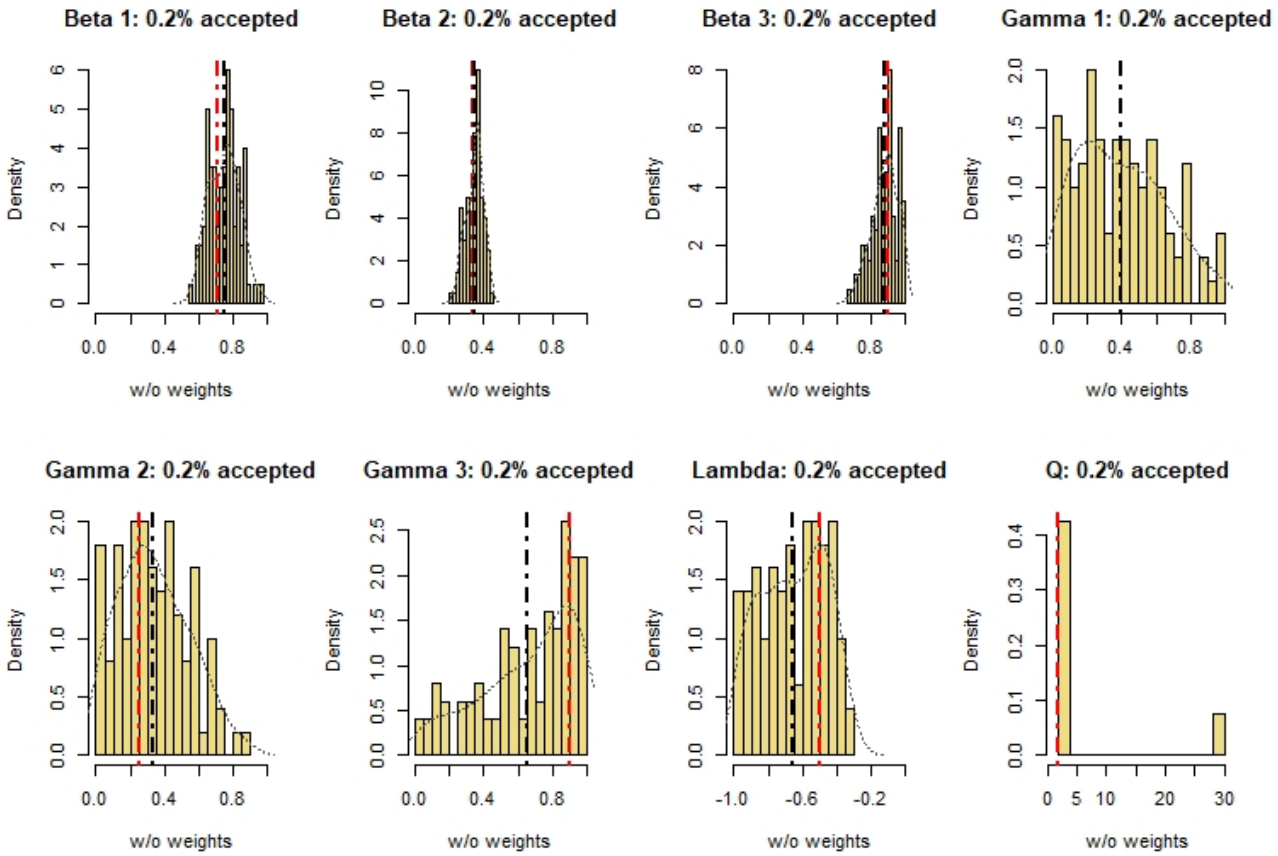

Figure 12: Approximations of the posterior distributions of the model's parameters using the identity matrix with the 0.2% quantile as acceptance threshold and for $N = 50000$. (The red line denotes the true value and the black the posterior mean).

5.9 Regression Adjustment

As previously mentioned, by setting a non-zero acceptance threshold ϵ, the ABC algorithm introduces bias to the approximation of the posterior distribution. That is because instead of sampling from the posterior distribution of the parameter it samples from an approximation of the posterior. However, by setting a relatively small acceptance threshold and a sensible distance function ρ we can expect that the approximation of the posterior distribution will be reasonable. Although the relationship between the choice of ϵ and the accuracy of the approximation has been studied (see Sisson et al. (2007); Dean et al. (2011); Fearnhead and Prangle (2011)), the dependence of the convergence of the approximation to the posterior distribution on the choice of ϵ and the distance measure ρ has not.

One way to reduce the bias introduced by the ABC Rejection algorithm was proposed by Beaumont et al. (2002). In step 7 of the algorithm presented in Table 1, smooth weighting and regression adjustment is proposed. The idea is based on the assumption that the parameter $\hat{\theta}_i$ which is a sample of the prior of θ, for which the difference between the summary statistics of the observed and the simulated data $\rho(S, \hat{S}_i)$ is small is a sample from the approximate posterior distribution and hence the first reasonable step is to weight $\hat{\theta}_i$ based on the value of $\rho(S, \hat{S}_i)$. The next step is to perform (a) linear regression and (b) local-linear regression on $\hat{\theta}_i$ in order to adjust them and weaken the effect of the difference between the sample summary statistic S and those of the simulated sample $\hat{S}_i$.

Linear Regression

Starting with the linear regression we assume that the following model can describe the density that we are estimating for some value α and a vector of coefficients $\boldsymbol{\beta}$:

$$\theta_i = \alpha + (S_i - S)^T \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, N, \tag{5.21}$$

where ϵ_i are uncorrelated errors with zero mean and common variance. In the case where $S_i = S$, i.e. the sample summary statistics match exactly with the observed summary statistics, the mean of the posterior distribution of θ_i is $E[\theta|S] = \alpha$.

The parameters α and $\boldsymbol{\beta}$ are the least squares estimates which means that their values are those that minimize the Sum of Squared Errors (SSE):

$$\sum_{i=1}^{N} [\theta_i - \alpha - (S_i - S)^T \boldsymbol{\beta}]^2. \tag{5.22}$$

The closed-form solution of the above, in matrix form, is:

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = (X^T X)^{-1} X^T \boldsymbol{\theta}, \tag{5.23}$$

where

$$X = \begin{pmatrix} 1 & S_{11}\text{-}S_1 & \cdots & S_{1m}\text{-}S_m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & S_{N1}\text{-}S_N & \ldots & S_{Nm}\text{-}S_m \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_N \end{pmatrix}. \tag{5.24}$$

Therefore, the adjusted values θ_i^* are obtained by:

$$\theta_i^* = \theta_i - (S_i - S)^T \hat{\boldsymbol{\beta}}, \tag{5.25}$$

which are assumed to be closer to a random sample from the posterior distribution $p(\theta|S)$.

Local-Linear Regression

A more flexible assumption is that the linearity and additivity assumed by the linear regression model 5.21 can only apply locally. The local-linear regression model, proposed by Beaumont et al. (2002), minimises:

$$\sum_{i=1}^{N}[\theta_i - \alpha - (S_i - S)^T\boldsymbol{\beta}]^2 K_\delta(||S_i - S||), \tag{5.26}$$

where $K_\delta(t)$ is a kernel such as the Epanechnikov kernel,

$$K_\delta(t) = \begin{cases} c\delta^{-1}(1-(t/\delta)^2), & t \leqslant \delta, \\ 0, & t > \delta, \end{cases} \tag{5.27}$$

where c is a normalising constant, or other kernel functions such as the Gaussian.

The values of $(\hat{\alpha}, \hat{\boldsymbol{\beta}})$ that minimise 5.26, in closed-form, are:

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = (X^T W X)^{-1} X^T W \boldsymbol{\theta}, \tag{5.28}$$

where W is a diagonal matrix whose elements are $K_\delta(||S_i - S||)$. These regression approaches can be extended to adjust more than one parameter at the same time using multivariate multiple regression.

		ABC	**Linear Regression**	**Local-Linear Regression**
Accepted	True value	0.707	0.707	0.707
5%	Posterior Mean	0.702	0.700	0.701
	Error (%)	0.70	1.02	0.92
	St. Dev.	2.131	1.073	1.074
1%	Posterior Mean	0.703	0.705	0.706
	Error (%)	0.63	0.28	0.13
	St. Dev.	1.378	0.875	0.881
0.5%	Posterior Mean	0.703	0.704	0.705
	Error (%)	0.65	0.38	0.27
	St. Dev.*	1.358	0.909	0.919

*Multiplied by 10^2.

Table 13: Comparison of parameter estimation of the Normal Factor Copula model (5.8) using (a) the ABC rejection algorithm without any adjustment, (b) with linear regression adjustment and (c) with local-linear regression adjustment, with acceptance threshold the 1% quantile and for $N = 10000$ samples and with acceptance thresholds the 5%, 1% and 0.5% quantiles.

Other approaches

Other alternatives have been proposed, such as heteroscedastic linear adjustment (Blum, 2017), non-linear adjustment through replacing the local-linear regression by a feed-forward neural network model and also a combination of the two, hence heteroscedastic non-linear adjustment (Blum and François, 2009).

A desirable property of Regression Adjustment is the posterior shrinkage. In the case of homoscedastic linear adjustment, the residuals are cenentered and the their empirical variance is smaller than the total variance of the regression. Thus, it is implied that the empirical variance of the posterior of the adjusted values, θ_{adj}, of the parameter is smaller than the empirical variance of values of the parameter θ obtained by the ABC Rejection algorithm. Therefore,

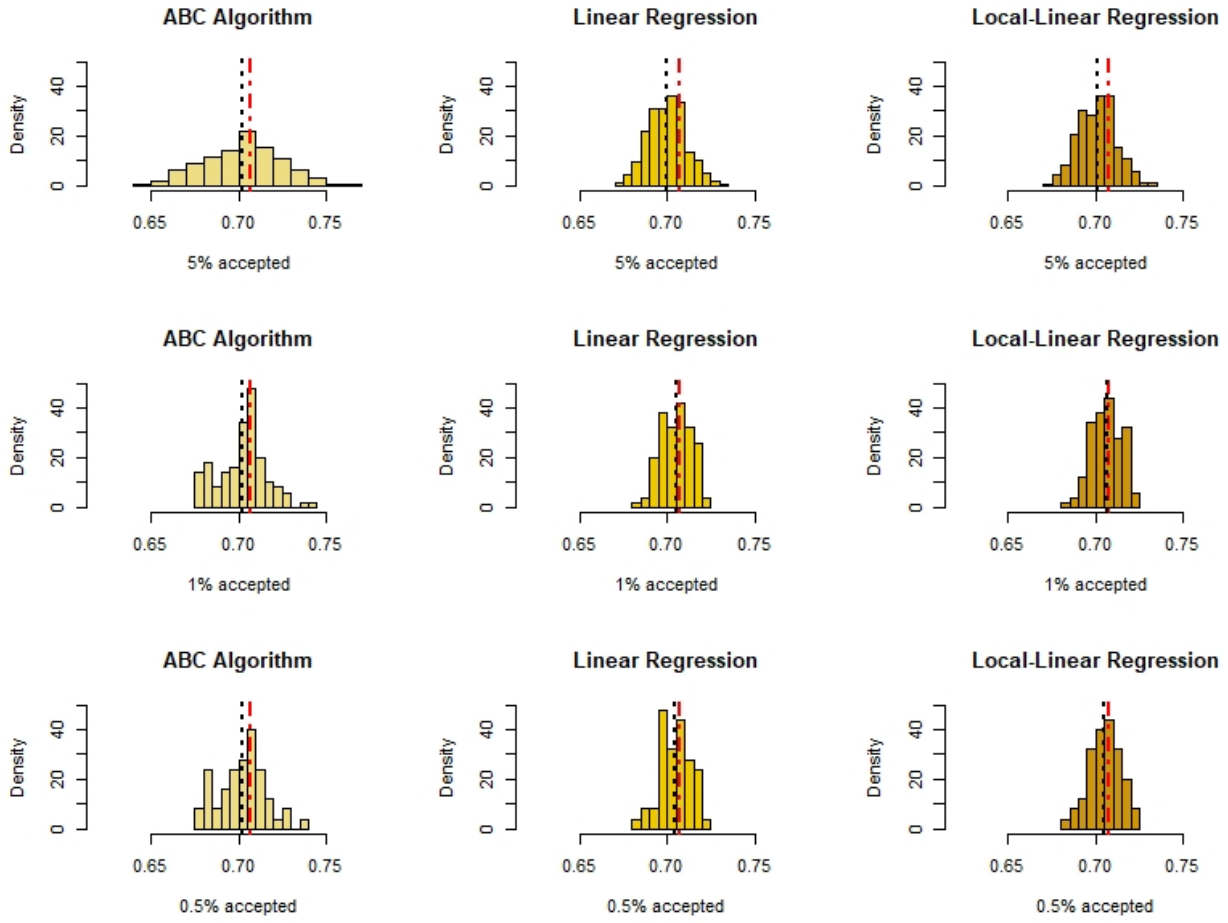

Figure 13: Comparison of the approximations of the posterior distribution of β, using the ABC rejection algorithm (first column), with linear regression adjustment (second column) and with local-linear regression adjustment (third column), while using the Normal Factor Copula model, for $N = 10000$ samples and for acceptance thresholds being the 5%, 1% and 0.5% quantiles. (The red line denotes the true value and the black the posterior mean).

the posterior variance in this case is reduced (Blum, 2017). In the cases of non-linear adjustment and heteroscedastic linear regression adjustment, shrinkage effects have also been reported (Blum, 2010).

The linear regression adjustment and the local-linear adjustment methods are applied to the estimation of the parameter of the Normal Factor Copula, described in Section 5.4. In Figure 13, the approximation of the posteriors of β are compared and the estimation results are presented in Table 13. The shrinkage effect is obvious in the histograms, both in the case of the linear and the local-linear regression. The posterior standard deviation is reduced by 32.3%-49.6% compared to the one obtained by the ABC rejection algorithm. The estimation error is also reduced by approximately 41.5%-79.3% for each adjustment method respectively, when the acceptance threshold is low, i.e. $0.5, 1\%$.

The regression adjustment techniques are then applied to the estimation results of the Skewed t-t Factor Copula of Section 5.6. The approximations of the posteriors of three parameters, β, λ and q, are obtained in this case. The shrinkage effect is present again for both of the adjustment methods. In Figure 14, the posteriors when the acceptance threshold is set to be the 1% quantile are presented. It seems that for all three parameters both the estimation error and the posterior variance are reduced. From Table 14 we can infer that the effects of the

linear and local-linear adjustment are very similar but at the same time quite significant. The estimation error for β is reduced by at least 94%, for λ by more than 25% and for q by 45%. The posterior standard deviations are reduced by 56.2%, 24% and 22.8% respectively.

Therefore, the regression adjustment techniques can complement the ABC rejection algorithm to reduce the estimation error while also reducing the estimation uncertainty. A combination of a larger sample of observed data and applying the ABC algorithm with a regression adjustment method will lead the posterior distribution to concentrate accurately around the true value of the parameter and will allow to estimate the parameter more accurately and with less uncertainty.

Accepted samples: 1%

Parameters		β	λ	q
	True Values	0.707	-0.50	2
ABC Algorithm	Posterior Mean	0.720	-0.817	10.08
	Error (%)	1.84	63.40	404
	St. Dev.*	2.49	13.89	721
Linear Regression	Posterior Mean	0.707	-0.733	6.44
	Error (%)	$\leqslant 0.01$	46.60	222
	St. Dev.*	1.09	10.43	555
Local-Linear Regression	Posterior Mean	0.706	-0.735	6.42
	Error (%)	0.1	46.98	221
	St. Dev.*	1.09	10.56	556

*Multiplied by 10^2.

Table 14: Comparison of parameter estimation for the Skewed t-t model (5.10) using (a) the ABC rejection algorithm without any adjustment, (b) with linear regression adjustment and (c) with local-linear regression adjustment with acceptance threshold the 1% quantile and for $N = 10000$ samples.

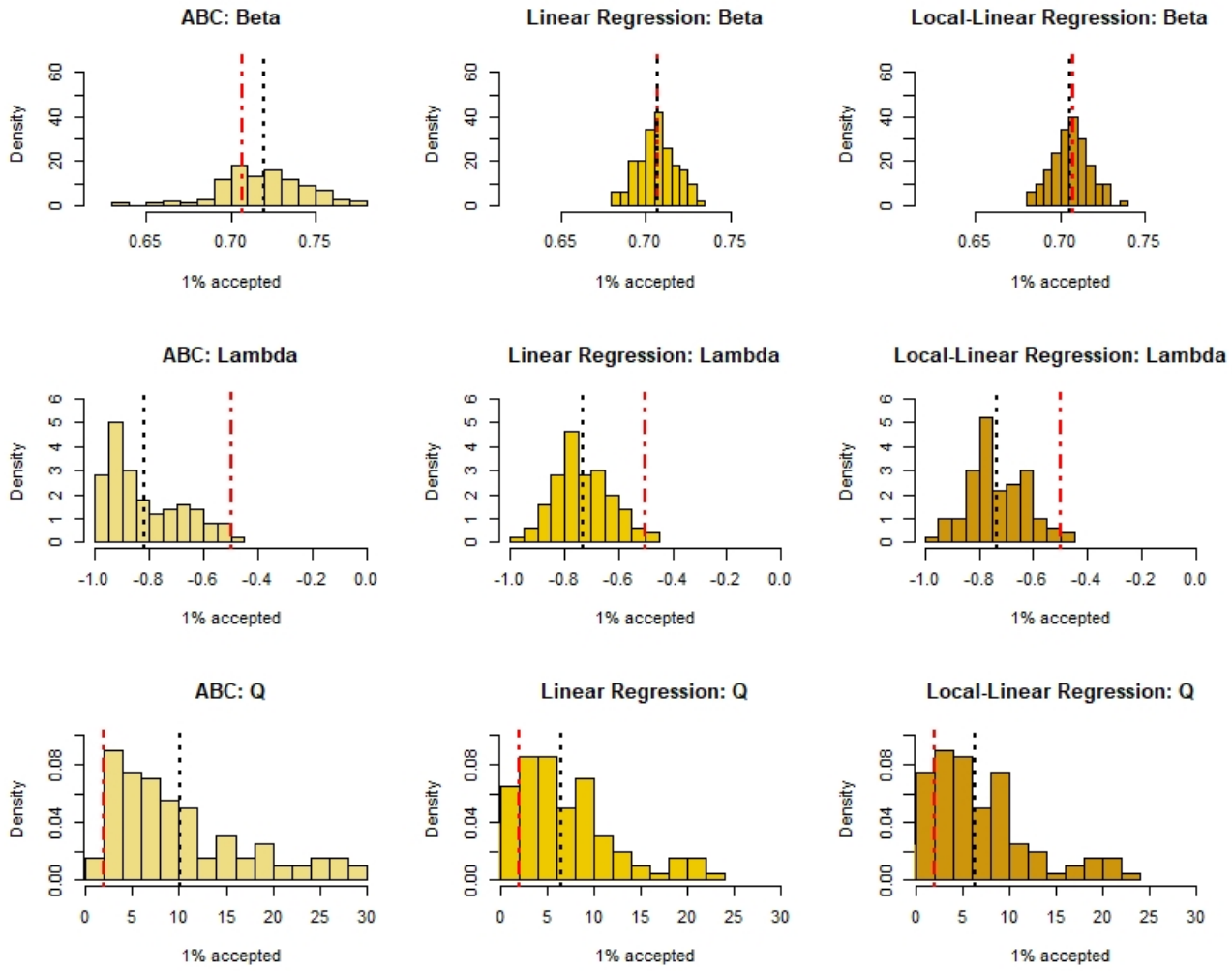

Figure 14: Comparison of the approximations of the posterior distribution of β (first row), λ (second row) and q (third row), using the ABC rejection algorithm (first column), with linear regression adjustment (second column) and with local-linear regression adjustment (third column). The Skewed t-t Factor Copula model was used with $N = 10000$ samples and for acceptance threshold set to be the 1% quantile. (The red line denotes the true value and the black the posterior mean).

6 Conclusion & Discussion

Even though there are plenty of models that can be used to model joint distributions and dependence in lower dimensions, it can be challenging as the dimensionality of the task increases. In the domain of copulas, there are a lot of classes of bivariate models, yet Factor Copulas are just one of the very few that can be used in high-dimensional applications. This class of models is particularly attractive due to their flexibility since by changing the number of latent variables, their distributions or their parameters' values and the factor loadings, one can adjust the overall flexibility of the model depending on the circumstances. For example, for high dimensional applications with large amount of data available one would allow the model to have more flexibility and complexity.

Most Factor Copulas do not typically have a closed-form likelihood. Therefore, simulation methods such as the ABC rejection algorithm can be used to estimate the model parameters. Unlike other estimation methods, the ABC algorithm is a powerful yet elegant way to not only estimate the parameters of a model but to also obtain their full posterior distributions. This allows to quantify the uncertainty in the parameters, using tools and techniques from the field of Bayesian inference. Since the evaluation of the likelihood function is not necessary, ABC allows the use of any model that is easy to simulate data from instead of only those whose likelihood can be evaluated. It also allows the use of techniques directly related to ABC such as the regression adjustment methods that can help reduce both the estimation error and uncertainty. Also, the class of regression models ranges from a simple linear regression to more complex non linear models.

In this report, we used the ABC algorithm to obtain approximations of the posterior distributions of the Factor Copulas' parameters in different settings. The first and simplest model was the bivariate one with Normal latent variables and only one free parameter. Then, we used the ABC algorithm in incrementally more complex scenarios, by using more flexible distributions and thus estimating more parameters. In all cases, the estimation was assessed with respect to the sample size and different acceptance thresholds. The "equidependence" property was assumed by all models by using a common latent variable. Subsequently, the dimensionality of the problem was increased, from two to three to nine variables and a more flexible model was considered that has the "block-equidependence" property. It allowed for different levels of dependence between variables of different groups. Lastly, the application of regression adjustment techniques was investigated for the cases of one and three parameters with the results suggesting that these methods, however simple, can significantly reduce the posterior variance and often the estimation error.

The accuracy of the estimation increases as the sample size increases. This is due to the fact that the true posterior distribution of the parameter concentrates more accurately around the true value as the sample size increases, assuming that the summary statistics used are highly informative. Hence, with an adequate number of simulations, very accurate estimates can be obtained. It seems reasonable to claim that the parameter estimation is rather accurate for sample size $N = 10000$, at least for the size and the complexity of models considered in this report. Applications with these amounts of data are not unreasonable nor unusual since for daily data this sample size corresponds to almost fourty years of historical data. One way to fully utilise the available data is to also use regression adjustment methods instead of just the rejection algorithm.

Although using Factor Copulas for modelling in high dimensions has been researched to a degree, applications in different fields where large amount of data is available will be an interesting field of research. There is a large number of decisions to be made when determining

the flexibility of the model and finding an optimal way to balance the flexibility of the model without over-complicating it is of great interest. Furthermore, there is indication that the ABC algorithm can be implemented in fields such as finance due to the abundance of available data. Further research on the application of the ABC algorithm on the estimation of models other than Factor Copulas would also be interesting as well as a field-specific comparison of the effect of different regression adjustment methods. Finally, another interesting research avenue would be to perform different regression adjustment methods, such as statistical, machine or deep learning regression models, and compare their performance using simple or more complicated models.

References

Beaumont, M. A. (2010). Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics 41*(1), 379–406.

Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate bayesian computation in population genetics. *Genetics 162*(4), 2025–2035.

Bertorelle, G., A. Benazzo, and S. Mona (2010). Abc as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology 19*(13), 2609–2625.

Blum, M. G. (2017). Regression approaches for approximate bayesian computation.

Blum, M. G. B. (2010). Approximate bayesian computation: A nonparametric perspective. *Journal of the American Statistical Association 105*(491), 1178–1187.

Blum, M. G. B. and O. François (2009, Mar). Non-linear regression models for approximate bayesian computation. *Statistics and Computing 20*(1), 63–73.

Blum, M. G. B., M. A. Nunes, D. Prangle, and S. A. Sisson (2013). A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. *Statistical Science 28*(2), 189 – 208.

Calvet, L. E. and V. Czellar (2014, 07). Accurate Methods for Approximate Bayesian Computation Filtering. *Journal of Financial Econometrics 13*(4), 798–838.

Capéraà, P., A.-L. Fougères, and C. Genest (2000, 01). Bivariate distributions with given extreme value attractor. *Journal of Multivariate Analysis 72*, 30–49.

Cattell, R. B. (1966, Apr). The scree test for the number of factors. *Multivariate Behavioral Research 1*(2), 245–276.

Chan, N.-H., J. Chen, X. Chen, Y. Fan, and L. Peng (2009, 01). Statistical inference for multivariate residual copula of garch models. *Statistica Sinica 19*, 53–70.

Chen, X. and Y. Fan (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics 135*(1-2), 125–154.

Chen, X., Y. Fan, and V. Tsyrennikov (2006, 01). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association 101*, 1228–1240.

Cornuet, J.-M., P. Pudlo, J. Veyssier, A. Dehne-Garcia, M. Gautier, R. Leblois, J.-M. Marin, and A. Estoup (2014, 01). DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics 30*(8), 1187–1189.

Csillery, K., M. Blum, O. Gaggiotti, and O. Francois (2010, July). Approximate bayesian computation (abc) in practice. *Trends in Ecology and Evolution 25*(7), 410–418.

Csillery, K., O. Francois, and M. G. B. Blum (2012). abc: an r package for approximate bayesian computation (abc). *Methods in Ecology and Evolution*.

Davis, C. (2015). *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.

Dean, T. A., S. S. Singh, A. Jasra, and G. W. Peters (2011, Mar). Parameter estimation for hidden markov models with intractable likelihoods. *arXiv:1103.5399 [math, stat]*. arXiv: 1103.5399.

Diggle, P. J. and R. J. Gratton (1984). Monte carlo methods of inference for implicit statistical models. *Journal of the Royal Statistical Society. Series B (Methodological) 46*(2), 193–227.

Embrechts, P., A. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond M. Dempster,*, Cambridge: Cambridge University Press.

Engle, R. and B. Kelly (2012, Apr). Dynamic equicorrelation. *Journal of Business Economic Statistics 30*(2), 212–228.

Fearnhead, P. and D. Prangle (2011, Apr). Constructing summary statistics for approximate bayesian computation: Semi-automatic abc. *arXiv:1004.1112 [stat]*. arXiv: 1004.1112.

Fine, J. (2000, Sep). On association in a copula with time transformations. *Biometrika 87*(3), 559–571.

Genest, C. and A. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*, 347–368.

Genest, C. and L.-P. Rivest (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association 88*(423), 1034–1043.

Größer, J. and O. Okhrin (2021, 04). Copulae: An overview and recent developments. *Wiley Interdisciplinary Reviews: Computational Statistics*.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review 35*(3), 705–30.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Lecture Notes-Monograph Series 28*, 120–141.

Joe, H. (2005, Jun). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis 94*(2), 401–419.

Joe, H. (2015). *Dependence modeling with copulas*. Monographs on statistics and applied probability. CRC Press, Taylor Francis Group.

Krupskii, P. and H. Joe (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis 120*, 85–101.

Krupskii, P. and H. Joe (2015, Jun). Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis 138*, 53–73.

Li, D. X. (2000). On default correlation. *The Journal of Fixed Income 9*(4), 43–54.

Li, W. and P. Fearnhead (2017). Convergence of regression adjusted approximate bayesian computation.

Lopes, J. and M. Beaumont (2010, August). Abc: a useful bayesian tool for the analysis of population data. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 10*(6), 826—833.

MacKenzie, D. and T. Spears (2014, 06). 'a device for being able to book pl': The organizational embedding of the gaussian copula. *Social studies of science 44*, 418–40.

McFadden, D. (1989, Sep). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*(5), 995.

Mesiar, R. and V. JáGr (2013, 10). D-dimensional dependence functions and archimax copulas. *Fuzzy Sets and Systems 228*, 78–87.

Nasri, B. R. and B. Remillard (2019, 03). Copula-based dynamic models for multivariate time series. *Journal of Multivariate Analysis 172*.

Nelsen, R. B. (1999). *An Introduction to Copulas*, Volume 139 of *Lecture Notes in Statistics*. Springer New York.

Nelsen, R. B. (2010). *An Introduction to Copulas*. Springer Publishing Company, Incorporated.

Oh, D. H. and A. J. Patton (2013, Jun). Simulated method of moments estimation for copula-based multivariate models. *Journal of the American Statistical Association 108*(502), 689–700.

Oh, D. H. and A. J. Patton (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics 35*(1), 139–154.

Okhrin, O., Y. Okhrin, and W. Schmid (2013). On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics 173*(2), 189–204.

Paarsch, H. and B. Brendstrup (2007, 02). Semiparametric identification and estimation in multi-object, english auctions. *Journal of Econometrics 141*, 84–108.

Pakes, A. and D. Pollard (1989, Sep). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027.

Patton, A. and D. H. Oh (2012). Modelling dependence in high dimensions with factor copulas. *Working Paper, Department of Statistics, Duke University*.

Patton, A. J. (2006a, Mar). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics 21*(2), 147–173.

Patton, A. J. (2006b). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, D. B. (1984, Dec). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics 12*(4).

Schefzik, R. (2015). Multivariate discrete copulas, with applications in probabilistic weather forecasting.

Sisson, S. A., Y. Fan, and M. M. Tanaka (2007, Feb). Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences 104*(6), 1760–1765.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris 8*, 229–231.

Song, P. X.-K., Y. Fan, and J. D. Kalbfleisch (2005, Dec). Maximization by parts in likelihood inference. *Journal of the American Statistical Association 100*(472), 1145–1158.

Sunnåker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz (2013). Approximate bayesian computation. *PLoS Computational Biology 9*(1).

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly (1997, Feb). Inferring coalescence times from dna sequence data. *Genetics 145*(2), 505–518.