Chapter 1

Bayesian Inference for D-vines: Estimation and Model Selection

Claudia Czado and Aleksey Min

Technische Universität München, Zentrum Mathematik,
Boltzmannstr. 3, 85747 Garching, Germany
*cczado@ma.tum.de**

During the last two decades the advent of fast computers has made Bayesian inference based on Markov Chain Monte Carlo (MCMC) methods very popular in many fields of science. These Bayesian methods are good alternatives to traditional maximum likelihood (ML) methods since they often can estimate complicated statistical models for which a ML approach fails. In this paper we review available MCMC estimation and model selection algorithms as well as their possible extensions for a D-vine pair copula constructions (PCC) based on bivariate $t-$copulas. However the discussed methods can easily be extended for an arbitrary regular vine PCC based on any bivariate copulas. A Bayesian inference for Australian electricity loads demonstrates the addressed algorithms at work.

1.1. Introduction

Pair copula constructions (PCC) for multivariate copulas have been successful in extending the class of available multivariate copulas (see Fischer *et al.* (2007) and Berg and Aas (2008)). Estimation of the corresponding copula parameters has been done so far using maximum likelihood (ML). However the number of parameters of a PCC model to be estimated can be considerable. So far it is facilitated by numerical optimization of the log likelihood to obtain ML estimates.

For inference purposes one needs to have reliable standard error estimates for the estimated parameters. The standard approach for this is to impose regularity conditions such that asymptotic normality of the param-

*Technische Universität München, Zentrum Mathematik, Boltzmannstr. 3, 85747 Garching, Germany

eter estimates holds and to approximate the estimated standard errors by evaluating numerically the Hessian matrix. However for large parameter vectors this evaluation is time consuming and its reliability is unclear. In addition numerical estimates of the Hessian matrix might result in non positive definite matrices yielding negative variance estimates. Out of these reasons Min and Czado (2008) started to investigate Bayesian inference for PCC models based on Markov Chain Monte Carlo (MCMC) methods (see Metropolis *et al.* (1953) and Hastings (1970)). Bayesian inference has the advantage of providing natural interval estimates based on the posterior distribution and does not rely on asymptotic normality. In addition the Bayesian approach is able to incorporate prior information which might be available from the data context or previous data analyses.

Min and Czado (2008) developed and implemented Bayesian MCMC algorithms for D-vines based on pair $t-$copulas. While this solved the problem of obtaining reliable interval estimates for parameters needed for inference purposes, the problem of model selection needed to be approached. The gain of flexibility using PCC is huge, however the problem which PCC model to choose becomes important. In particular Morales-Napoles *et al.* (2008) have shown that the number of PCC models even in small dimensions can be enormous, so it is impossible to fit all models and compare them. Therefore efficient model selection strategies are needed. While Heinen and Valdesogo (2008) approached this problem by using truncated PCC constructions, Min and Czado (2009) approached the problem of reducing a chosen PCC by using reversible jump (RJ) MCMC methods suggested by Green (1995) and successfully applied to search large model spaces as is needed in PCC's. The purpose of this paper is to give an overview of these Bayesian estimation and model selection procedures and illustrate their usefulness in a data set involving Australian electricity loads from Czado *et al.* (2009).

The proposed methodology is developed for data transformed to the copula level, i.e. for data living on the multivariate unit cube. We will use a parametric and nonparametric approach to create the copula data used for illustration. We have chosen this data set to facilitate comparison of two step estimation procedures to the joint estimation procedure of Czado *et al.* (2009) and Gärtner (2008). For this data set it turns out that two step estimation procedures are nearly as efficient as the joint estimation procedure, thus making the extra effort required for the joint estimation less needed.

The paper is organized as follows. In Section 1.2 we briefly consider a general D-vine decomposition for a multivariate density. Section 1.3 presents the likelihood of D-vine PCC's based on $t-$copulas. In Section 1.4 we survey MCMC estimation and model selection algorithms as well as their possible extensions for a D-vine PCC based on $t-$copulas. Sections 1.5 and 1.6 illustrates the discussed MCMC methods in the Bayesian analysis of Australian electricity loads from Czado *et al.* (2009). In Section 1.7 we summarize our findings and discuss further open problems as well as future research directions.

1.2. D-vine

Using Sklar's theorem in d dimensions multivariate distributions on $\mathbb{R}^d$ with given margins can be easily constructed. However this general approach does not provide a solution for the construction of flexible multivariate distributions. In this section we give such a construction proposed first by Joe (1996), organized by Bedford and Cooke (2002) and applied to Gaussian copulas only. Later Aas *et al.* (2007) used bivariate Gaussian, t, Gumbel and Clayton copulas as building blocks to increase model flexibility.

Let $f(x_1, \ldots, x_d)$ be a $d-$dimensional density function and $c(u_1, \ldots, u_d)$ be the corresponding copula density function. For a pair of integers r and s $(1 \leq r \leq s \leq d)$ a set $r:s$ denotes all integers between r and s, namely $r : s \; := \{r, \ldots, s\}$. If $r > s$ then $r : s = \varnothing$. Further let $\mathbf{X}_{r:s}$ denote the set of variables $\{X_r, \ldots, X_s\}$ and $u_{i|r:s}$ denote a conditional cumulative distribution function (cdf) $F_{i|r:s}(u_i|\mathbf{u}_{r:s})$. It is well known that the density $f(x_1, \ldots, x_d)$ can be factorized as

$$f(x_1, \ldots, x_d) = f_d(x_d) \cdot f_{d-1|d}(x_{d-1}|x_d) \cdot f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d) \times \ldots \cdot f_{1|2\cdots d}(x_1|x_2, \ldots, x_d). \quad (1.1)$$

The above factorization is a simple consequence from the definition of conditional densities and is invariant with respect to permutation of the variables.

The second factor $f_{d-1|d}(x_{d-1}|x_d)$ on the right hand side of (1.1) can be represented as a product of a copula density and the marginal density $f_d(x_d)$ in the following way. Consider the bivariate density function $f_{(d-1)d}(x_{d-1}, x_d)$ with marginal densities $f_{d-1}(x_{d-1})$ and $f_d(x_d)$, respectively. Using Sklar's theorem for $d = 2$, we have that the conditional density $f_{d-1|d}(x_{d-1}|x_d)$ is given by

$$\begin{aligned} f_{d-1|d}(x_{d-1}|x_d) &= \frac{f_{(d-1)d}(x_{d-1}, x_d)}{f_d(x_d)} \\ &= c_{(d-1)d}(F_{d-1}(x_{d-1}), F_d(x_d)) \cdot f_{d-1}(x_{d-1}). \end{aligned} \tag{1.2}$$

Similarly, the conditional density $f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d)$ is given by

$$\begin{aligned} f_{d-2|(d-1)d}(x_{d-2}|x_{d-1}, x_d) &= \frac{f_{(d-2)(d-1)|d}(x_{d-2}, x_{d-1}|x_d)}{f_{d-1|d}(x_{d-1}|x_d)} \\ = \;& c_{(d-2)(d-1)|d}(F_{d-2|d}(x_{d-2}|x_d), F_{d-1|d}(x_{d-1}|x_d)) \\ \times \;& f_{d-2|d}(x_{d-2}|x_d) \\ = \;& c_{(d-2)(d-1)|d}(u_{d-2|d}, u_{d-1|d}) \cdot c_{(d-2)d}(x_{d-2}, x_d) \cdot f_{d-2}(x_{d-2}). \end{aligned} \tag{1.3}$$

The copula density $c_{(d-2)(d-1)|d}(\cdot, \cdot)$ is the conditional copula density corresponding to the conditional distribution $F_{(d-2)(d-1)|d}(x_{d-2}, x_{d-1}|x_d)$. Further $F_{d-i|d}(x_{d-i}|x_d)$ is the conditional distribution function of x_{d-i} given x_d for $i = 1, 2$. Note that in general the conditional copula density $c_{(d-2)(d-1)|d}(F_{d-2|d}(x_{d-2}|x_d), F_{d-1|d}(x_{d-1}|x_d))$ depends on the given conditioning value x_d. By induction the jth factor ($j = 4, \ldots, d$) in (1.1) is given by

$$\begin{aligned} & f_{j|1:(j-1)}(x_j|\mathbf{x}_{1:(j-1)}) \\ &\quad = c_{1j|2:(j-1)}(u_{1|2:(j-1)}, u_{j|2:(j-1)}) \cdot f_{j|2:(j-1)}(x_j|\mathbf{x}_{2:(j-1)}) \\ &\quad = \prod_{t=1}^{j-2} c_{tj|(t+1):(j-1)}(u_{t|(t+1):(j-1)}, u_{j|(t+1):(j-1)}) \\ &\quad \times c_{(j-1)j}(u_{j-1}, u_j) \cdot f_j(x_j). \end{aligned} \tag{1.4}$$

Thus we can represent each term on the right hand side of (1.1) as the product of the corresponding marginal density and copula density terms. Combining (1.2)–(1.4), expression (1.1) can be rewritten as

$$f(x_1, \ldots, x_d) = \prod_{t=1}^{d} f(x_t) \times \prod_{j=2}^{d} \prod_{t=1}^{j-1} c_{tj|(t+1):(j-1)}(u_{t|(t+1):(j-1)}, u_{j|(t+1):(j-1)}). \tag{1.5}$$

The density $f(x_1, \ldots, x_d)$ is the product of d marginal densities and $d(d-1)/2$ pair copula density terms. The pair copula density terms are unconditional copulas evaluated at marginal distribution function values or conditional copulas evaluated at univariate conditional distribution function values. The above construction was defined in Aas *et al.* (2007) and was called the D-vine pair copula construction (PCC) for multivariate distributions.

1.3. D-Vines PCC based on $t-$copulas

From now on, we use as building pair copulas of the PCC model (1.5) bivariate $t-$copulas. However the estimation and model selection methodology is generic and applies much more widely. Further we assume that the margins of $\mathbf{X}$ are uniform. This is motivated by the standard two step semi-parametric copula estimation procedure suggested by Genest *et al.* (1995), where approximate uniform margins are obtained by applying the empirical probability integral transformation to standardized fitted residuals based on specified marginal models.

The bivariate $t-$copula (see e.g. Embrechts *et al.* (2003)) has 2 parameters: the association parameter $\rho \in (-1, 1)$ and the df parameter $\nu \in (0, \infty)$ and its density is given by

$$c(u_1, u_2|\nu, \rho) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\sqrt{1-\rho^2}\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^2} \cdot \frac{\left(\left[1+\frac{x_1^2}{\nu}\right]\left[1+\frac{x_2^2}{\nu}\right]\right)^{\frac{\nu+1}{2}}}{\left(1+\frac{x_1^2+x_2^2-2\rho x_1 x_2}{\nu(1-\rho^2)}\right)^{\frac{\nu+2}{2}}},$$

where $x_i := t_\nu^{-1}(u_i)$ for $i = 1, 2$ and $t_\nu^{-1}(\cdot)$ is a quantile function of a $t-$distribution with ν degrees of freedom. Specifying the pair copulas and assuming uniform margins, the conditional distribution function for a bivariate $t-$copula are needed. It is called the $h-$function for the $t-$copula with parameters ρ and ν, and Aas *et al.* (2007) derive it as

$$h(u_1|u_2, \rho, \nu) = t_{\nu+1}\left(\frac{t_\nu^{-1}(u_1) - \rho\, t_\nu^{-1}(u_2)}{\sqrt{\frac{\left(\nu+\left(t_\nu^{-1}(u_2)\right)^2\right)(1-\rho^2)}{\nu+1}}}\right). \tag{1.6}$$

The D-Vine PCC (1.5) with building bivariate $t-$copulas depends now on a $d(d-1)$ dimensional parameter vector $\boldsymbol{\theta}$ given by

$$\boldsymbol{\theta} = (\rho_{12}, \nu_{12}, \rho_{23}, \nu_{23}, \ldots, \rho_{1d|2:(d-1)}, \nu_{1d|2:(d-1)})^t,$$

where $\rho_{tj|(t+1):(j-1)}$ and $\nu_{tj|(t+1):(j-1)}$ are parameters of the $t-$copula density $c_{tj|(t+1):(j-1)}(\cdot, \cdot)$ for $j = 2, \ldots, d$ and $t = 1, \ldots, j-1$. As already noted in Min and Czado (2008) the likelihood $c(\mathbf{u}|\boldsymbol{\theta})$ of the D-vine copula for N $d-$dimensional realizations $\mathbf{u} := (\mathbf{u}_1, \ldots, \mathbf{u}_N)$ of $\mathbf{U} := (U_1, \ldots, U_d)^t$ can be calculated as

$$c(\mathbf{u}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \Big[\prod_{i=1}^{d-1} c\left(u_{i,n}, u_{i+1,n}|\rho_{i(i+1)}, \nu_{i(i+1)}\right) \tag{1.7}$$
$$\times \prod_{j=2}^{d-1} \prod_{i=1}^{d-j} c\left(v_{j-1,2i-1,n}, v_{j-1,2i,n}|\rho_{i(i+j)|(i+1):(i+j-1)}, \nu_{i(i+j)|(i+1):(i+j-1)}\right) \Big],$$

where for $n = 1, \ldots, N$

$$\begin{aligned}
v_{1,1,n} &:= h(u_{1,n}|u_{2,n}, \rho_{12}, \nu_{12}) \\
v_{1,2i,n} &:= h(u_{i+2,n}|u_{i+1,n}, \rho_{(i+1)(i+2)}, \nu_{(i+1)(i+2)}),\ i = 1, \ldots, d-3, \\
v_{1,2i+1,n} &:= h(u_{i+1,n}|u_{i+2,n}, \rho_{(i+1)(i+2)}, \nu_{(i+1)(i+2)}),\ i = 1, \ldots, d-3, \\
v_{1,2d-4,n} &:= h(u_d|u_{d-1}, \rho_{(d-1)d}, \nu_{(d-1)d}), \\
v_{j,1,n} &:= h(v_{j-1,1,n}|v_{j-1,2,n}, \rho_{1(1+j)|2:j}, \nu_{1(1+j)|2:j}),\ j = 2, \ldots, d-2, \\
v_{j,2i,n} & \\
&:= h(v_{j-1,2i+2,n}|v_{j-1,2i+1,n}, \rho_{i(i+j)|(i+1):(i+j-1)}, \nu_{i(i+j)|(i+1):(i+j-1)}) \\
&\qquad \text{for} \quad d > 4,\ j = 2, \ldots, d-3 \quad \text{and} \quad i = 1, \ldots, d-j-2 \\
v_{j,2i+1,n} & \\
&:= h(v_{j-1,2i+1,n}|v_{j-1,2i+2,n}, \rho_{i(i+j)|(i+1):(i+j-1)}, \nu_{i(i+j)|(i+1):(i+j-1)}) \\
&\qquad \text{for} \quad d > 4,\ j = 2, \ldots, d-3 \quad \text{and} \quad i = 1, \ldots, d-j-2 \\
v_{j,2d-2j-2,n} & \\
&:= h(v_{j-1,2d-2j,n}|v_{j-1,2d-2j-1,n}, \rho_{(d-j)d|(d-j+1):(d-1)}, \nu_{(d-j)d|(d-j+1):(d-1)}) \\
&\qquad \text{for} \quad j = 2, \ldots, d-2.
\end{aligned}$$

Note that $v_{j,2i,n}$ and $v_{j,2i+1,n}$ in (1.7) are jth fold superpositions of the h-function (1.6).

1.4. Bayesian inference for D-vine PCC based on $t-$copulas

Estimation of D-vine PCC's in a MCMC framework is straightforward and similar to estimation of any multivariate distribution with many parameters. The nature of parameters defined by the specific choice of the copula family should be taken into account. In contrast to multivariate density functions the arguments in a conditional D-vine density term is a complicated function of arguments and the parameters of earlier pair D-vine densities. This makes the evaluation of the log likelihood time consuming. Further the parameter update of PCC's is usually performed by a

Metropolis-Hastings (MH) algorithm (see Metropolis *et al.* (1953) and Hastings (1970)).

Min and Czado (2008) develop and implement one such MCMC algorithm for the estimation of parameters of PCC's. They use noninformative priors for ρ's and ν's. Since estimation of df ν is unstable for large true ν's, its support should be restricted to some finite interval $(1, U)$. A noninformative prior for each ρ results in a uniform distribution on $(-1, 1)$. There are several other possibilities for the choice of priors for ν. Thus Czado *et al.* (2009) use a Cauchy distribution while Dalla Valle (2007) utilizes a truncated Poisson distribution. Joe (2006) developed a uniform prior on the space of positive definite correlation matrices, which imply different beta priors for corresponding partial correlations arising from a D-vine. This alternative prior choice has been used in Czado *et al.* (2009).

There are also several choices of the proposal distributions needed for the MH algorithm. Min and Czado (2008) use a modification of a random normal walk proposal, which is a normal distribution truncated to the support of parameters. Variances of normal distributions are tuned to achieve acceptance rates between 20% and 80% as suggested by Besag *et al.* (1995). Another choice is an independence proposal distribution which is independent of the current value of the sampled parameter. A common independence proposal is a normal distribution with the same mode and inverse curvature at the mode as the target distribution described for example in Gilks *et al.* (1996). This has been used in Czado *et al.* (2009) for the joint MCMC estimation of marginal AR(1) and D-vine copula parameters. Generalizations of the normal independence proposal using $t-$distribution with low degrees of freedom ν, say $\nu = 3$ or $\nu = 5$, are also often used.

The number of pair copulas $n_c = d(d-1)/2$ in (1.7) increases quadratically with dimension d of the data. However if independence or conditional independence is present in data then the number of factors in (1.7), respectively, may reduce drastically. This (conditional) independence is characterized by a unit pair copula density. Therefore the first task on model selection for D-vine PCC's is to determine its non-unit pair copula terms. Min and Czado (2009) derive and implement a reversible jump (RJ) MCMC of Green (1995). The algorithm by Green (1995) allows to explore a huge number of models since only visited models will be fitted. Therefore it is well accepted by the Bayesian community though its derivation and implementation for a particular problem are not simple tasks. Another model selection approach by Congdon (2006) is discussed and utilized in Min and Czado (2008). The recent approach of Congdon (2006) is easy implement but it compares only among prespecified models.

Key points of the RJ MCMC algorithm of Min and Czado (2009) are an introduction of a model indicator vector of dimension n_c and the RJ mechanism for a model change. They associate models with subdecompositions of (1.7) consisting of k $(1 \leq k \leq n)$ pair copula terms. To specify the model indicator pair copulas in the full decomposition (1.7) has to be ordered. Otherwise an identifiability problem occurs since PCC's are invariant with respect to permutation of factors. According to the labeling of Min and Czado (2009), for $d = 4$ the full decomposition for a multivariate copula density is given as follows

$$c(u_1, u_2, u_3, u_4) = c_{12}c_{23}c_{34}c_{13|2}c_{24|3}c_{14|23},$$

where we omit arguments and parameters of pair copulas for brevity. The model indicator $\mathbf{m}_f$ is given by a six dimensional vector $(1, 1, 1, 1, 1, 1)$, where 1 indicates the presence of the corresponding pair copula term. If now some pair copula terms are not present in the decomposition then the corresponding ones are replaced by zeros. For example a model indicator $\mathbf{m} = (1, 1, 1, 1, 1, 0)$ corresponds to the subdecomposition $c(u_1, u_2, u_3, u_4) = c_{12}c_{23}c_{34}c_{13|2}c_{24|3}$ without the last pair copula $c_{14|23}$.

Any RJ MCMC algorithm consists of so-called birth and death moves. For birth moves the dimension of the model parameter increases while for death moves the dimension decreases. Min and Czado (2009) derives acceptance probabilities for both moves in detail. As a proposal distribution for the parameters of the sth pair copula they use a bivariate normal distribution $N_2(\hat{\boldsymbol{\theta}}_s^{\text{MLE}}, \Sigma_s)$ truncated to $(-1, 1) \times (1, U)$. Here $\hat{\boldsymbol{\theta}}_s^{\text{MLE}} = (\rho_s^{\text{MLE}}, \nu_s^{\text{MLE}})'$ denotes the corresponding two dimensional sub-vector of the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_{\mathbf{m}_f}^{\text{MLE}}$ in the full model $\mathbf{m}_f$. Note that there are n_c covariance matrices Σ_s's, which govern the reversible jump mechanism. They are taken of the form $\Sigma_s = diag(\sigma_{s,\rho}^2, \sigma_{s,\nu}^2)$, where $diag(a, b)$ denotes a diagonal matrix with a and b on the main diagonal.

1.5. Application: Australian electricity loads

In this section we illustrate the above discussed estimation algorithms for the Australian electricity loads from Czado *et al.* (2009). We are solely interested in estimating the dependence structure and therefore first marginal AR(1)'s are fitted to extract independent i.i.d. residuals. Now copula data for the Australian electricity loads can be obtained using empirical probability integral transformations or corresponding univariate normal cdf's.

Here we study both the nonparametric and parametric copula data. To facilitate comparison to the models considered in Czado *et al.* (2009) we now investigate the following PCC here and in the sequel:

$$c(u_Q, u_N, u_V, u_S) = c_{QN} \cdot c_{NV} \cdot c_{VS} \cdot c_{QV|N} \cdot c_{NS|V} \cdot c_{QS|NV}, \qquad (1.8)$$

where the parameter dependence of each bivariate $t-$copula and their arguments are dropped to keep the expression short. The subindexes Q, N, V and S correspond to Queensland, New South Wales, Victoria and South Australia, respectively.

For both copula data we run the MH algorithm specified in Min and Czado (2008) for 10000 iterations using Cauchy priors truncated to $(1, 100)$ for the df parameter of each pair as in Gärtner (2008), namely $\pi(\nu) \propto 1/[1 + (v - 1)^2/4]$. The first 500 iterations are considered as burn-in. Proposal variances were determined in pilot runs and resulted in acceptance rates between 23%-77% for all parameters after 10000 iterations. Autocorrelations among the MCMC iterates suggested sub-sampling to reduce these correlations and every 10-th iteration was recorded. Table 1.1 summarizes the estimated posterior distributions for all parameters based on the recorded iterations for both copula data. For comparison we also include the corresponding maximum likelihood estimates (MLE) given in the last column of Table 1.1 as well as results of joint estimation using AR(1) margins and the D-vine PCC model (1.8). In the joint MCMC estimation of Czado *et al.* (2009) a slightly different prior for the ρ parameter has been used.

For both copula data the Bayesian estimates of $\rho_{QV|N}, \rho_{NS|V}$ and $\rho_{QS|NV}$ are not credible at 5% level since the corresponding credible intervals contain 0. They are also not credible at 10% level except for $\rho_{QV|N}$ when the parametric copula data is used. Posterior mode estimates of ν's are larger than 10 only for $\rho_{QV|N}$ and $\rho_{QS|NV}$ while it is smaller than 10 for $\nu_{NS|V}$. At 95% credibility we conclude for both copula data that conditional independence between loads of Queensland and Victoria given loads of New South Wales as well as between loads of New South Wales and South Australia given loads of Victoria are present. Therefore the decomposition (1.8) can be reduced by pair copulas $c_{QV|N}$ and $c_{QS|NV}$. However it is difficult to decide from the above results, whether loads of New South Wales and South Australia are conditionally independent given loads of Victoria. More sophisticated Bayesian model selection procedures discussed in the next section address this problem. For the parametric copula data posterior mode estimates for df's are usually higher then the corresponding

Table 1.1. Estimated posterior mean, mode and quantiles of MCMC as well as MLE of copula parameters for the copula data obtained from the preprocessed Australian load data using Cauchy prior for ν's truncated to $(1, 100)$.

Copula	2.5%	5%	50%	95%	97.5%	mean	mode	MLE
Nonparametric copula data								
ν_{QN}	3.27	3.39	4.46	6.18	6.75	4.59	4.30	4.46
ν_{NV}	2.61	2.68	3.31	4.32	4.58	3.37	3.27	3.26
ν_{VS}	4.27	4.53	6.04	8.65	9.34	6.24	5.79	6.24
$\nu_{QV\|N}$	10.29	11.50	28.38	79.72	88.49	34.31	22.33	51.80
$\nu_{NS\|V}$	4.96	5.27	7.56	12.63	14.64	8.13	6.92	7.73
$\nu_{QS\|NV}$	8.74	10.01	21.62	73.13	83.50	28.41	17.61	32.18
ρ_{QN}	0.24	0.25	0.30	0.35	0.36	0.30	0.31	0.31
ρ_{NV}	0.28	0.29	0.35	0.40	0.41	0.35	0.35	0.35
ρ_{VS}	0.53	0.53	0.57	0.60	0.61	0.57	0.57	0.57
$\rho_{QV\|N}$	−0.02	−0.01	0.03	0.08	0.09	0.03	0.03	0.03
$\rho_{NS\|V}$	−0.04	−0.03	0.03	0.08	0.09	0.03	0.03	0.03
$\rho_{QS\|NV}$	−0.03	−0.02	0.03	0.08	0.08	0.03	0.03	0.03
Parametric copula data								
ν_{QN}	5.04	5.31	7.24	10.48	11.43	7.48	6.92	7.32
ν_{NV}	4.08	4.24	5.50	7.62	7.96	5.66	5.33	5.52
ν_{VS}	6.79	7.16	10.72	18.03	20.22	11.43	9.99	11.17
$\nu_{QV\|N}$	13.71	15.25	31.13	78.32	86.09	36.77	26.50	38.55
$\nu_{NS\|V}$	5.73	6.15	8.41	13.34	15.55	8.97	7.92	8.34
$\nu_{QS\|NV}$	12.49	14.07	30.11	79.18	87.96	35.91	24.58	58.07
ρ_{QN}	0.27	0.28	0.33	0.38	0.38	0.33	0.33	0.33
ρ_{NV}	0.32	0.33	0.38	0.43	0.44	0.38	0.38	0.39
ρ_{VS}	0.53	0.54	0.57	0.60	0.61	0.57	0.57	0.57
$\rho_{QV\|N}$	−0.00	0.01	0.06	0.12	0.13	0.06	0.07	0.06
$\rho_{NS\|V}$	−0.03	−0.02	0.03	0.09	0.10	0.03	0.03	0.04
$\rho_{QS\|NV}$	−0.03	−0.02	0.03	0.08	0.09	0.03	0.03	0.03
Joint estimation of marginal and copula parameters								
ν_{QN}	5.17	5.45	7.37	11.72	12.71	7.80	6.92	7.32
ν_{NV}	4.11	4.26	5.57	7.65	8.60	5.76	5.36	5.52
ν_{VS}	7.08	7.69	11.45	24.29	29.68	12.89	10.23	11.17
$\nu_{QV\|N}$	15.01	16.28	36.89	84.56	91.77	41.44	29.80	38.55
$\nu_{NS\|V}$	4.20	4.73	14.25	78.22	93.44	22.84	11.60	8.34
$\nu_{QS\|NV}$	12.32	14.58	34.43	78.22	86.81	38.94	29.00	58.07
ρ_{QN}	0.27	0.28	0.34	0.38	0.39	0.34	0.34	0.33
ρ_{NV}	0.33	0.35	0.40	0.45	0.45	0.40	0.40	0.39
ρ_{VS}	0.54	0.55	0.59	0.62	0.63	0.59	0.59	0.57
$\rho_{QV\|N}$	−0.01	−0.00	0.05	0.10	0.11	0.05	0.05	0.06
$\rho_{NS\|V}$	−0.01	0.00	0.05	0.11	0.12	0.06	0.05	0.04
$\rho_{QS\|NV}$	−0.04	−0.03	0.02	0.07	0.08	0.02	0.03	0.03

one for the nonparametric copula data. Difference in estimates of ρ's is here negligible. Further we observe that the joint estimation of AR(1) margins and copula parameters gives results similar to ones for the parametric copula data.

We now compare the two step estimation procedures (estimate margins first, then form standardized residuals and then transform to copula data, either using nonparametric or parametric transformations) to the one step estimation procedure using joint MCMC. For this comparison we see that the credible intervals are similar for the two step parametric and joint estimation procedure except for $\nu_{NS|V}$. For the nonparametric two step estimation procedure the posterior means and modes for the df parameters are lower than for the parametric and joint estimation procedure, thus indicating more heavy tailedness in the data than what is present. Here we consider the joint estimation method as the most appropriate estimation method, since the marginal residuals do not violate the marginal AR(1) model assumption. Overall we see that the loss in efficiency is not huge if one uses a two step estimation procedure compared to a joint estimation procedure for this data set.

1.6. Bayesian model selection for the Australian electricity loads

Based on simulations studies, Min and Czado (2009) have advocated to use $U = 20$ as the upper limit of the prior distribution for ν's. Then the model selection performance of RJ MCMC for PCC's based on $t-$copulas significantly increases. Here we follow their approach.

We run the MH algorithm Min and Czado (2008) to tune proposal variances for the full PCC in (1.8). These tuned variances are used in the stay move to update the corresponding new parameter values. For the birth move we propose new values for $\boldsymbol{\theta}_{\mathbf{m}_s}$, $s = 1, \ldots, 6$ according to the normal $N_2(\hat{\boldsymbol{\theta}}^{\text{MLE}}_{m_s}, \Sigma)$ distribution truncated to $(-1, 1) \times (1, U)$, where $\hat{\boldsymbol{\theta}}^{\text{MLE}}_{m_s^{(\text{new})}} = (\rho_s^{\text{MLE}}, \nu_s^{\text{MLE}})'$ denotes the corresponding two dimensional subvector of the ML estimate $\hat{\boldsymbol{\theta}}^{\text{MLE}}_{\mathbf{m}_f}$ in the full model $\mathbf{m}_f$. The MLE $\hat{\boldsymbol{\theta}}^{\text{MLE}}_{\mathbf{m}_f}$ is determined under the constraints $-1 < \rho < 1$ for $\rho' s$ and $1 < \nu < 20$ for ν's. We consider two birth proposal covariance matrices Σ's, namely $\Sigma_1 = diag(10^2, 100^2)$ and $\Sigma_2 = diag(0.1^2, 3^2)$ to investigate robustness of the procedure. Further we use $\hat{\boldsymbol{\theta}}^{\text{MLE}}_{\mathbf{m}_f}$ and $\mathbf{m}_f$ as initial values for $\boldsymbol{\theta}$ and $\mathbf{m}$, respectively.

Note that there are 6 copula terms in (1.8) which can be present or not in a model. Disregarding the model of complete independence this gives that there are $63 = 2^6 - 1$ models to be explored by the RJ MCMC

Table 1.2. Estimated posterior model probabilities $\hat{P}_k = \widehat{P}(M_k|\text{data})$ of all 63 models for the nonparametric copula data obtained from the preprocessed Australian load data using an empirical cdf's. The model probabilities in the third and fourth columns, and in the last fifth column are obtained using RJ MCMC and Congdon's approach, respectively. In parentheses the corresponding model probability estimates for the parametric copula data obtained from the preprocessed Australian load data are given. Further $U = 20$, $\Sigma_1 = diag(10^2, 100^2)$ and $\Sigma_2 = diag(0.1^2, 3^2)$.

Model	Formula	$\hat{P}_k$		
		Σ_1	Σ_2	Cong.
M_{63}: $\mathbf{m} = (1,1,1,1,1,1)$	$c_{QN}c_{NV}c_{VS}c_{QV\|N}$ $\times c_{NS\|V}c_{QS\|NV}$	0.001 (0.000)	0.000 (0.000)	0.002 (0.009)
M_{62}: $\mathbf{m} = (1,1,1,1,1,0)$	$c_{QN}c_{NV}c_{VS}c_{QV\|N}$ $\times c_{NS\|V}$	0.013 (0.054)	0.014 (0.065)	0.062 (0.185)
M_{60}: $\mathbf{m} = (1,1,1,1,0,0)$	$c_{QN}c_{NV}c_{VS}c_{QV\|N}$	0.001 (0)	0.001 (0.000)	0.004 (0.001)
M_{59}: $\mathbf{m} = (1,1,1,0,1,1)$	$c_{QN}c_{NV}c_{VS}$ $\times c_{NS\|V}c_{QS\|NV}$	0.022 (0.012)	0.025 (0.011)	0.090 (0.070)
M_{58}: $\mathbf{m} = (1,1,1,0,1,0)$	$c_{QN}c_{NV}c_{VS}$ $\times c_{NS\|V}$	0.923 (0.933)	0.933 (0.923)	0.710 (0.726)
M_{57}: $\mathbf{m} = (1,1,1,0,0,1)$	$c_{QN}c_{NV}c_{VS}c_{QS\|NV}$	0.001 (0)	0.000 (0.000)	0.003 (0.000)
M_{56}: $\mathbf{m} = (1,1,1,0,0,0)$	$c_{QN}c_{NV}c_{VS}$	0.039 (0.000)	0.026 (0.001)	0.129 (0.009)
M_i: for $i \neq 63, 62, 60, \ldots, 56$		0 (0)	0 (0)	– (–)

algorithm. We enumerate models by the binary representation. Thus the full decomposition $\mathbf{m} = (1,1,1,1,1,1)$ in (1.8) corresponds to 63. If the pair copulas $c_{QV|N}$ and $c_{QS|NV}$ are set equal to 1 then the corresponding model vector is given by $\mathbf{m} = (1,1,1,0,1,0)$. This binary representation corresponds to 58.

Table 1.2 displays estimated posterior probabilities for all possible 63 PCC models and for both copula data obtained from the preprocessed Australian electricity loads based on 100000 iterations with burn-in of 10000. For comparison the last column of Table 1.2 gives approximations to posterior probabilities for the seven models based on the approach by Congdon (2006). The implementation of Congdon's algorithm is similar as presented in Czado *et al.* (2009). Thus our RJ MCMC algorithm with $U = 20$ shows that the PCC model without pairs $c_{MB|T}$ and $c_{ST|M}$ has the highest estimated posterior probability for both choices of Σ independent of the transformation used to obtain copula data. Congdon's approach also supports the above model though with less confidence. However Robert and Marin

(2008) note that there might be considerable bias in Congdon's method.

1.7. Summary and discussion

This paper review methods on Bayesian inference for D-vine PCC's and illustrates their use for a specific data set. The methodology can easily be extended to cover any regular vine PCC model. Since the classical ML approach to PCC's will give only reliable point estimates but not reliable standard error estimates a Bayesian approach is followed here.

To assess the influence of prior distributions we have run the original MH algorithm specified in Min and Czado (2008) for the Australian load data for 10000 iterations using uniform priors on $(1, 100)$ for each df parameter. This means a median value of 50.5 for each df parameter while the truncated Cauchy prior in Czado *et al.* (2009) has its median at 3. This results as expected in a considerable increase of posterior means for ν's which are > 20, while the ρ parameters are not affected. We observe differences in Bayesian estimates only for df's if the corresponding MLE or posterior mode estimates are larger than 20. For small estimated ν's the influence of the prior for ν is negligible. In contrast Bayesian estimates for ρ's are robust with respect to prior distributions for ρ.

Min and Czado (2008) report that numerically evaluated Hessian matrix and bootstrap methods are good alternative for getting reliable standard errors if dimension of data $d < 4$. The estimated Hessian matrix can often fail to give reliable standard estimates since negative variance estimates might occur. Further for high dimensional data of $d > 4$ with thousands multivariate observations the bootstrap and Hessian approach becomes much more time consuming in contrast to MCMC methods as Min and Czado (2008) find. In a simulation study Min and Czado (2009) show that for model selection purposes the upper limit for ν should be set to 20. Then the model performance of their RJ MCMC algorithm is significantly improved. The results of the RJ MCMC analysis for the Australian load data are robust with regard to the choice of proposal distribution for the birth move as Min and Czado (2009) notice. In the next step we plan to derive and implement a RJ MCMC algorithm when the copula family of pair-copulas is not fixed anymore and it can vary within a catalogue of bivariate copulas including the independence copula.

In some problems joint estimation of marginal and copula parameters has recently been found to be important. Thus Kim *et al.* (2007) have shown that a separate estimation of the marginal parameters may have an essential influence on the parameter estimation of multivariate copulas. Therefore inference based on joint estimates might be lead to quite different results compared to the inference ignoring estimation errors in the marginal parameters. In financial applications marginal time series usually follow ARMA or GARCH models. Here our future research will concentrate on joint estimation of marginal (ARMA) GARCH and PCC parameters. Finally this all can be generalized to PCC models with time varying parameters since financial data usually shows that the dependence structure changes over time (s. for example Patton (2004)).

s.

Acknowledgement

Claudia Czado acknowledges the support of the Deutsche Forschungsgemeinschaft (Cz 86/1-3).

References

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2007). Pair-copula constructions of multiple dependence, *To appear in Insurance Math. Econom. DOI 10.1016/j.insmatheco.2007.02.001* .

Bedford, T. and Cooke, R. M. (2002). Vines–a new graphical model for dependent random variables, *Ann. Statist.* **30**, 4, pp. 1031–1068.

Berg, D. and Aas, K. (2008). Models for construction of multivariate dependence, *accepted for publication in European Journal of Finance* .

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian computation and stochastic systems, *Statist. Sci.* **10**, 1, pp. 3–66, with comments and a reply by the authors.

Congdon, P. (2006). Bayesian model choice based on Monte Carlo estimates of posterior model probabilities, *Computational Statistics and Data Analysis* **50**, pp. 346 – 3572.

Czado, C., Gärtner, F. and Min, A. (2009). Analysis of Australian electricity loads using joint Bayesian inference of D-vines with autoregressive margins, *Submitted Preprint* available at http://www-m4.ma.tum.de/Papers/index.html.

Dalla Valle, L. (2007). Bayesian copulae distributions with application to operational risk management, *To appear in Methodol. Comput. Appl. Probab. DOI 10.1007/s11009-007-9067-x* .

Embrechts, P., Lindskog, F. and McNeil, A. J. (2003). Modelling dependence with copulas and applications to risk management, in *Handbook of heavy tailed distributions in finance* (Elsevier/North-Holland, Amsterdam).

Fischer, M., Köck, C., Schlüter, S. and Weigert, F. (2007). An empirical analysis of multivariate copula models, *Preprint* available at www.statistik.wiso.uni-erlangen.de/lehrstuhl/mfischer/mfwork.html.

Gärtner, F. (2008). *Bayesian analysis of multivariate time series models based on pair-copula construction* (Technische Universität München, Diploma thesis in the Zentrum Mathematik).

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika* **82**, 3, pp. 543–552.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. Gilks, W. R. (ed.) et al., Markov chain Monte Carlo in practice. London: Chapman & Hall. 1-19 (1996).

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**, 4, pp. 711–732.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**, pp. 97–109.

Heinen, A. and Valdesogo, A. (2008). A canonical vine autoregressive model for large dimensions, Preprint available at www.core.ucl.ac.be/∼valdesogo/papers.html.

Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters, in L. Rüschendorf and B. Schweizer and M. D. Taylor (ed.), *Distributions with Fixed Marginals and Related Topics*.

Joe, H. (2006). Generating random correlation matrices based on partial correlations, *Journal of Multivariate Analysis* **97**, pp. 2177–2189.

Kim, G., Silvapulle, M. J. and Silvapulle, P. (2007). Comparison of semiparametric and parametric methods for estimating copulas, *Comput. Statist. Data Anal.* **51**, 6, pp. 2836–2850.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Phisics* **21**, pp. 1087–1092.

Min, A. and Czado, C. (2008). Bayesian inference for multivariate copulas using pair-copula constructions. *Submitted Preprint* available under http://www-m4.ma.tum.de/Papers/index.html.

Min, A. and Czado, C. (2009). Bayesian model selection for multivariate copulas using pair-copula constructions. *Submitted Preprint* available at http://www-m4.ma.tum.de/Papers/index.html.

Morales-Napoles, O., Cooke, R. and Kurowicka, D. (2008). About the number of vines and regular vines on n nodes, (to be submitted).

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* **2**, pp. 130–168.

Robert, C. and Marin, J.-M. (2008). On some difficulties with a posterior probability approximation technique. Preprint, available under http://arxiv.org/abs/0801.3513.