# Bayesian Model Selection of Regular Vine Copulas

Lutz Gruber[1], Claudia Czado[1]

*Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013*
*Paper no. 17*

[1] Technische Universität München, Zentrum Mathematik, Lehrstuhl für Mathematische Statistik, München, Germany
lutz.gruber@ma.tum.de
cczado@ma.tum.de

**Abstract**

Regular vine copulas can describe a wider array of dependency patterns than the multivariate Gaussian copula or the multivariate Student's t copula. We present two contributions related to model selection of regular vine copulas. First is a reversible jump Markov chain Monte Carlo algorithm to estimate the joint posterior distribution of the density factorization, pair copula families and parameters of a regular vine copula. In a second step, we reduce the algorithm to a tree-by-tree stepwise Bayesian procedure that allows for faster computation. A simulation study shows that our algorithm outperforms the model selection methods suggested in current literature and succeeds in recovering the true model when other methods fail. Furthermore, we present an application study that shows how a vine copula-based approach can improve the pricing of exotic financial derivatives using real-life data.

**Keywords**: Dependence Modeling; Vine Copula; Model Selection; Markov Chain Monte Carlo

## 1 Introduction

Multivariate data with rich patterns of dependence are found in many fields in business and science. The tool of choice to model these patterns of dependence are multivariate distributions with uniform margins; these are called copulas [11]. The Gaussian copula, possibly the most widely known copula, even made it into mainstream media as "The Formula That Killed Wall Street" [10]. The signature feature of copulas is that they allow dependence characteristics to be modeled separately from the marginal distributions. This provides the added benefit

that copulas can be introduced to existing models that do not yet incorporate measures of dependence, but feature established models for the margins.

While many classes of bivariate copulas, also known as pair copulas, are well known [6], there is only a very limited number of multivariate copulas available with a closed-form analytical expression. Additionally, these cover only limited patterns of dependence. Regular vine copulas provide a solution to this problem by using nested conditioning with arbitrary pair copulas to describe multivariate dependencies.

More specifically, regular vine copulas are set up in two steps: first is the construction of an $n$-dimensional copula density from (conditional) bivariate copula densities. These are organized in a sequence of linked trees $\mathcal{V} = (T_1, \dots, T_{n-1})$ called the regular vine. Each of the $n - j$ edges of tree $T_j$, $1 \leq j \leq (n-1)$ corresponds to a bivariate copula density that is conditional on $j - 1$ variables. Secondly, a copula family is selected for each of these (conditional) bivariate building blocks from a set of bivariate (parametric) candidate families $\mathbf{B}$. We denote the mapping of the pair copulas to the regular vine by $\mathcal{B}_\mathcal{V}(\boldsymbol{\theta}_\mathcal{V})$, where we write $\boldsymbol{\theta}_\mathcal{V}$ for the parameters of the pair copulas.

We discuss two Bayesian model selection procedures of regular vine copulas [4, 5] that are based on a reversible jump Markov chain Monte Carlo algorithm [3]. Previous work on Bayesian as well as frequentist model selection of regular vine copulas includes [7, 8, 13, 12, 1]. We add to that list a fully Bayesian model selection algorithm that estimates the joint posterior distribution of the vine density factorization $\mathcal{V}$ and the pair copulas $\mathcal{B}_\mathcal{V}(\boldsymbol{\theta}_\mathcal{V})$. Furthermore, we propose an accelerated algorithm that proceeds tree-by-tree in a stepwise Bayesian fashion for faster computation.

Our reversible jump MCMC sampler performs a within-model move and a between-models move at each iteration. The within-model move updates the parameters $\boldsymbol{\theta}_\mathcal{V}$ of the current regular vine copula. The between-models move attempts to change the model to a different regular vine copula that is different in the regular vine $\mathcal{V}$ and/or the pair copula families $\mathcal{B}_\mathcal{V}$. We use random walk proposals for the trees $T_1, \dots, T_{n-1} \in \mathcal{V}$ of the regular vine and for the parameters $\boldsymbol{\theta}_\mathcal{V}$. The proposal weights for the candidate pair copula families $B \in \mathbf{B}$ are proportional to the likelihood of each candidate family, when its parameter is set to its estimated value.

Given that the number of candidate vine tree structures $N = \frac{n!}{2} \times 2^{\binom{n-2}{2}}$ [9] grows super-exponentially in the dimension $n$, model selection is a non-trivial problem. The Metropolis-Hastings proposal mechanisms of our sampling algorithms are designed to achieve rapid convergence of the sampling chain (Figure 1). To enforce model sparsity, we propose to use a shrinkage prior of the form $\pi(\mathcal{V}, \mathcal{B}_\mathcal{V}(\boldsymbol{\theta}_\mathcal{V})) \propto \exp(-\lambda d)$, where $d$ denotes the dimension of the parameter vector $\boldsymbol{\theta}_\mathcal{V}$ and $\lambda$ is a parameter that sets the strength of the shrinkage. Depending on the strength of the shrinkage prior, our algorithm recovers 98% to 100% of the log likelihood of the "true" models used in our simulation studies. This represents a major improvement over existing model selection methods whose

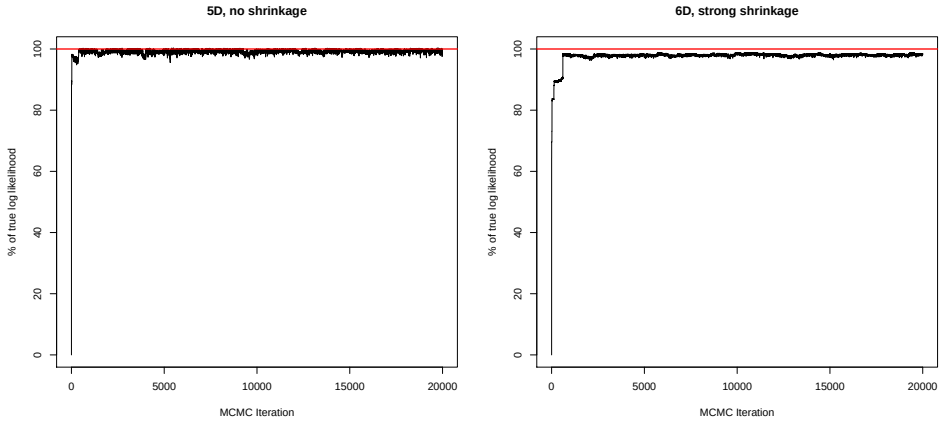


Figure 1: Log likelihood trace plots of a simulation study with known 5-dimensional and 6-dimensional regular vine copulas.

recovery rates hover in the 75% to 85% range.

In a final analysis, we estimate the expected payout of an exotic financial derivative on a basket of securities. First, we strip the return time series from their time-dependencies using GARCH(1,1) models [2]. Then we estimate a regular vine copula to the approximately independent *Uniform(0,1)* data and calculate Monte Carlo estimates of the payouts by drawing from the estimated copula's distribution. Current best practice to model dependencies in financial data is to use the Student's t copula. In recognition of this, we benchmark against it to establish the vine copula's outperformance in a real-life scenario.

# References

[1] Jeffrey F. Dißmann, Eike Christian Brechmann, Claudia Czado, and Dorota Kurowicka. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59(1):52–69, 2013.

[2] Jianqing Fan and Qiwei Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, 2001.

[3] Peter J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[4] Lutz F. Gruber and Claudia Czado. Sequential Bayesian Model Selection of Regular Vine Copulas. Preprint, 2012.

[5] Lutz F. Gruber and Claudia Czado. Bayesian Model Selection of Regular Vine Copulas. Preprint, 2013.

[6] Harry Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 2001.

[7] Aleksey Min and Claudia Czado. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546, 2010.

[8] Aleksey Min and Claudia Czado. Bayesian model selection for D-vine pair-copula constructions. *Canadian Journal of Statistics*, 39:239–258, 2011.

[9] Oswaldo Morales-Napoles. Counting vines. In Dorota Kurowicka and Harry Joe, editors, *Dependence Modeling: Vine Copula Handbook*. World Scientific Publishing Co., 2011.

[10] Felix Salmon. Recipe for Disaster: The Formula That Killed Wall Street. *Wired Magazine*, 17.03, 2009.

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

[12] Michael Smith, Aleksey Min, Carlos Almeida, and Claudia Czado. Modeling Longitudinal Data Using a Pair-Copula Decomposition of Serial Dependence. *Journal of the American Statistical Association*, 105(492):1467–1479, 2010.

[13] Jakob Stöber and Claudia Czado. Detecting regime switches in the dependence structure of high dimensional financial data. Preprint, 2011.