Technische Universität München

Zentrum Mathematik

Parent orderings in pair-copula Bayesian networks

Diplomarbeit

von

Philippe Bretan

Themensteller/in:	Prof. Claudia Czado, Ph.D.
Betreuer/in:	Prof. Claudia Czado, Ph.D.; Alexander Bauer
Abgabetermin:	03.12.2012

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 03.12.2012

Acknowledgments

First I would like to express my gratitude to my thesis supervisor Prof. Claudia Czado. She adopted the supervision after my original supervisor Prof. Dr. Thomas Klein left the university. All along she was up to date with my current work, gave me directions and supported me with her deep knowledge in the field of copulas and her experience in writing scientific reports.

Further I would like to thank my second supervisor, Alexander Bauer, who was also constantly involved in my work. His research provided the foundation for my thesis and he could support me with his knowledge in this specific field. Some of the scripts for the statistical program R that I used were written by him. He introduced me to work with them and helped me out with other programming issues.

My thanks also goes to Prof. Dr. Thomas Klein who was my original supervisor and proposed the subject which turned out to be very interesting because it belongs to a rather new field of research with a great potential of practical application.

Moreover, I appreciate the use of the Linux cluster of the Leibniz-Rechenzentrum (LRZ). It provides a good environment for the statistical program R as one can submit larger tasks to the cluster and also access the cluster from home.

At last I would like to thank the developers of the most important R libraries that I used: Peter Hepperger for the PccOptimizeR-package, which is currently not published, Ulf Schepsmeier and Eike Christian Brechmann for the CDVine-package and together with Jakob Stoeber for the VineCopula-package.

Zusammenfassung

Die Diplomarbeit beschäftigt sich mit Paar-Copula Bayesschen Netzwerken (PCBN) als multivariates statistisches Modell. PCBNs nutzen direkte azyklische Graphen (DAGs) als graphisches Modell für Paar-Copula Konstruktionen. Es wird gezeigt, welche DAGs und PCBN Dekompositionen basierend auf Annahmen zu bedingten Unabhängigkeiten konstruiert werden können. Dabei erhält man unterschiedliche PCBN Dekompositionen durch verschiedene Ordnungen der Eltern in DAGs. Anhand von Simulationen und Anwendungen auf natürliche Daten wird die Qualität und Rechenzeit von PCBNs mit verschiedenen Elternordnungen untersucht und auch mit der Qualität und Rechenzeit von R-Vines verglichen.

Contents

1 Introduction **1**
- 1.1 Pair-copula constructions . . . 2
- 1.2 Pair-copula constructions and directed acyclic graphs . . . 4
- 1.3 Decompositions and immoralities . . . 8

2 Possible models **10**
- 2.1 Markov equivalent DAGs . . . 10
- 2.2 Decomposition of DAGs as pair-copula Bayesian networks (PCBN) . . . 14
- 2.3 Conditional cdfs corresponding to pair-copula decomposition . . . 17

3 Simulation studies **20**
- 3.1 Case study 1 . . . 22
- 3.2 Case study 2 . . . 27
- 3.3 Case study 3 . . . 32
- 3.4 A second look: More copula families . . . 37
- 3.5 The parent ordering algorithm . . . 38

4 Application **39**
- 4.1 Stock market indices . . . 39
- 4.2 Uranium dataset . . . 48
- 4.3 Wine dataset . . . 58

5 Outlook: Optimization of computation time **67**
- 5.1 PCBN decompositions of the uranium dataset . . . 69

6 Conclusion **72**

Appendices **75**

A Uranium dataset: all parent orderings **75**

B Uranium dataset: PCBN decompositions **82**

C Uranium dataset: RVine tree plots **90**

D Wine dataset: all parent orderings **92**

E Wine dataset: PCBN decompositions **97**

F Wine dataset: RVine tree plots **104**

Chapter 1

Introduction

The use of copulas in statistical models has become more popular in recent years, especially for financial data. Copulas have the ability to describe the dependence structure of multiple variables, seperated from their univariate margins. Unfortunately there is not much literature on multivariate copulas and they have shown to be difficult to work with. Bivariate copulas are much more flexible, so Joe (1996) proposed pair-copula constructions (PCC). A PCC utilizes bivariate copulas as building blocks for a multivariate model. Using Sklar‘s theorem, the multivariate distribution is decomposed into bivariate, possibly conditional distributions. The question of structuring a PCC opens up many different classes of models.

Bedford and Cooke (2001, 2002) developed regular vines to define the structure of a PCC. In application to financial data, vine-PCC models outperformed other multivariate copula models, see Czado et al. (2012). However with an increase in number of variables, the number of pair-copulas and parameters in a vinePCC-model grows exponentially. Therefore Hanea et al. (2006) proposed to use conditional independence properties to lower the number of copulas in the model.

This thought lead to the application graphical models. A graphical model is a multivariate probabilistic model that describes conditional independence assumptions between random variables. Lauritzen (1996) provides an introduction to graphical models. If the graph is a directed acyclic graph (DAG), then the factorization of the joint probability distribution can be deduced from the graph. This is also called a Bayesian network.

Bauer et al. (2012) introduced DAG-PCC models. If the conditional independences of a dataset can be expressed in a DAG, then this provides a factorization of the joint probability distribution. In the next step, the conditional distributions functions in the factorization can be decomposed into bivariate, possibly conditional copulas. One problem is that the conditional copulas have conditional distribution functions as arguments which in some cases have to be computed by integration via other margins.

If we assume certain conditional independences neither the corresponding DAG nor its decomposition are generally unique. Different decompositions can be described by differ-

ent parent orderings for vertices with multiple parents. The goal of my thesis is to analyze the quality and computational complexity of models based on different parent orderings and different DAGs which represent the same conditional independences.

In the sections of this first chapter, we will give an introduction to pair-copula constructions and PCC-DAG models. Chapter 2 shows how to find all possible DAGs and their decompositions based on conditional independence assumptions. Following we will compare different parent orderings in the application on simulated datasets in Chapter 3 and in the application on natural datasets in Chapter 4. Chapter 5 analyzes when we have to compute more complex integrals. Finally the results are concluded in Chapter 6.

1.1 Pair-copula constructions

According to its definition, a copula can be described as a multivariate cumulative distribution function (cdf) $C : [0,1]^d \to [0,1], d \in \mathbb{N}$, such that all univariate marginals are uniform on the interval $[0,1]$. Sklar's theorem Sklar (1959) states that every multivariate distribution $F : \mathbb{R}^d \to [0,1]$ can be specified by the marginal distributions and an appropriate copula which characterizes the depedence structure:

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)) \tag{1.1}$$

Assuming that F is absolutely continuos and that the marginals are strictly increasing, we can write for the probability density function (pdf):

$$f(x_1, ..., x_d) = c(F_1(x_1), ..., F_d(x_d)) \prod_{i=1}^{d} f_i(x_i), \tag{1.2}$$

The copula density c is uniquely determined here. Using marginal quantile functions F_i^{-1}, the above equations can be transformed to

$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))$$

$$c(u_1, ..., u_d) = \frac{f(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))}{\prod_{i=1}^{d} f_i(F_i^{-1}(u_i))}$$

with $(u_1, ..., u_d) \in [0,1]^d$.

There does not exist a lot of literature on multivariate copula families and many important bivariate copulas cannot be easily extended to higher-variate copulas. Therefore Bedford and Cooke (2001, 2002) introduced pair-copulas constructions (PCCs): Only (conditional) bivariate copulas, called pair-copulas, are used as building blocks. These constructions are not unique. R-Vines and D-Vines are two of the most known types.

Let F be an absolutely continuos five-dimensional cdf with stricly increasing margins $F_1, ..., F_5$. Then the corresponding cdf can be factorized as

$$f(\mathbf{x}) = f_1(x_1)\ f_{2|1}(x_2|x_1)\ f_{3|12}(x_3|\mathbf{x}_{12})\ f_{4|123}(x_4|\mathbf{x}_{123})\ f_{5|1234}(x_4|\mathbf{x}_{1234})$$

where f_i is the pdf of F_i. Using (1.2) $f_{2|1}$ can be expressed as

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2))\ f_2(x_2)$$

To obtain pair-copulas we have to decompose pdfs that are conditional on multiple variables. For $f_{3|12}$ we write

$$\begin{aligned} f_{3|12}(x_3|\mathbf{x}_{12}) &= \frac{f_{13|2}(\mathbf{x}_{13}|x_2)}{f_{1|2}(x_1|x_2)} \\ &= \frac{f_{13|2}(\mathbf{x}_{13}|x_2)}{f_{1|2}(x_1|x_2)\ f_{3|2}(x_3|x_2)} f_{3|2}(x_3|x_2) \\ &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)\ c_{23}(F_2(x_2), F_3(x_3))\ f_3(x_3) \end{aligned}$$

Note that this decomposition is not unique. We are free to chose the order of the vertices in the conditioning set.

$$f_{3|12}(x_3|\mathbf{x}_{12}) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)\ c_{13}(F_1(x_1), F_3(x_3))\ f_3(x_3)$$

As an example we apply a D-Vine PCC which determines the order of the vertices in the conditioning sets. The formula can be found in Aas et al. (2009). This leads to

$$\begin{aligned} f(\mathbf{x}) =& c_{12}(F_1(x_1), F_2(x_2))\ c_{23}(F_2(x_2), F_3(x_3))\ c_{34}(F_3(x_3), F_4(x_4))\ c_{45}(F_4(x_4), F_5(x_5)) \\ & c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)\ c_{24|3}(F_{2|3}(x_2|x_3), F_{4|3}(x_4|x_3)|x_3) \\ & c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)|x_4)\ c_{14|23}(F_{1|23}(x_1|\mathbf{x}_{23}), F_{4|23}(x_4|\mathbf{x}_{23})|\mathbf{x}_{23}) \\ & c_{25|34}(F_{2|34}(x_2|\mathbf{x}_{34}), F_{5|34}(x_5|\mathbf{x}_{34})|\mathbf{x}_{34})\ c_{15|234}(F_{1|234}(x_1|\mathbf{x}_{234}), F_{5|234}(x_5|\mathbf{x}_{234})|\mathbf{x}_{234}) \\ & f_1(x_1)\ f_2(x_2)\ f_3(x_4)\ f_4(x_4)\ f_5(x_5) \end{aligned}$$

The conditional cdfs can be computed iteratively using the following formula provided by Joe (1996). One can verify that all needed copulas in this formula are part of our decomposition.

$$F_{j|K}(x_j|\mathbf{x}_K) = \frac{\partial C_{jk|(K\setminus\{k\})}(F_{j|(K\setminus\{k\})}(x_j|\mathbf{x}_{(K\setminus\{k\})}), F_{k|(K\setminus\{k\})}(x_k|\mathbf{x}_{(K\setminus\{k\})})|\mathbf{x}_{(K\setminus\{k\})})}{\partial F_{k|(K\setminus\{k\})}(x_k|\mathbf{x}_{(K\setminus\{k\})})}$$

where $k \in K$.

Especially for larger models, the computation of D-vine PCCs is very expensive, because the number of used copulas is $\binom{d}{2}$ for F on $\mathbb{R}^d$. We could simplify the decomposition if we are given conditional independences. Bauer et al. (2012) proposed using a directed acyclic graph (DAG) as a graphical model to specify these conditional independences. This will be explained in the following section.

Because copulas are invariant under strictly increasing transformations of the underlying random variables, we will assume for the sake of simplicity that all univariate marginals are uniform distributions on $[0, 1]$.

1.2 Pair-copula constructions and directed acyclic graphs

A detailed introduction to DAGs as a graphical model can be found in Lauritzen (1996) or Koski and Noble (2009).

Definition 1.1 (Directed acylic graph) *Let $V \neq \varnothing$ be a finite set and let E be subset of $(v, w) \in V^2 \,|\, v \neq w$ such that $(w, v) \notin E$ whenever $(v, w) \in E$. Then $D = (V, E)$ is a directed graph with vertex set V and edge set E. We denote a pair $(v, w) \in E$ by an arrow $v \to w$. A path in D is a sequence $v_1, ...v_n \in V, n \geqslant 2$, such that D contains all arrows $v_i \to v_{i+1}, i \leqslant n-1$. In the special case $v_1 = v_n$ the path $v_1, ...v_n$ is a cycle. If there are no cycles in D, then D is called a directed acyclic graph (DAG).*

We will call two vertices $v, w \in V$ *connected*, if $v \to w$ or $w \to v$. Then we set

$$\begin{aligned} pa(v) &:= \{w \in V \,|\, D \text{ contains } w \to v\} && \text{(set of } \textit{parents} \text{ of } v\text{)}, \\ de(v) &:= \{w \in V \,|\, D \text{ contains a path from } v \text{ to } w\} && \text{(set of } \textit{descendants} \text{ of } v\text{)}, \\ nd(v) &:= V \backslash (de(v) \cup \{v\}) && \text{(set of } \textit{non-descendants} \text{ of } v\text{)} \end{aligned}$$

for all $v \in V$.

Figure 1.1 shows an example of a DAG. It can be verified that it does not contain a circle.

v	1	2	3	4	5
$pa(v)$	$\varnothing$	$\{1\}$	$\{1,2\}$	$\{2\}$	$\{3,5\}$
$de(v)$	$\{2,3,4,5\}$	$\{3,4,5\}$	$\{5\}$	$\{5\}$	$\varnothing$
$nd(v)$	$\varnothing$	$\{1\}$	$\{1,2,4\}$	$\{1,2,3\}$	$\{1,2,3,4\}$

Let $\mathbf{X} \in \mathbb{R}^{|V|}$ be a random variable and P be a probability measure such that $\mathbf{X} \sim P$. P is called *D-Markovian* or to possess the *local D-Markov property* if

$$X_v \perp \mathbf{X}_{nd(v) \backslash pa(v)} \,|\, \mathbf{X}_{pa(v)} \quad \text{for all } v \in V. \tag{1.3}$$

In the case that P has a pdf f, P is D-Markovian if and only if f allows a *D-recursive factorisation*:

$$f(\mathbf{x}) = \prod_{v \in V} f_{v|pa(v)}(x_v | \mathbf{x}_{pa(v)}) \quad \text{for all } x \in \mathbb{R}^{|V|}, \tag{1.4}$$

We will restrict our studies on probability measures P which are D-Markovian given some DAG D.

Let us examine the 5-dimensional DAG in Figure 1.1 as an example. Like defined in

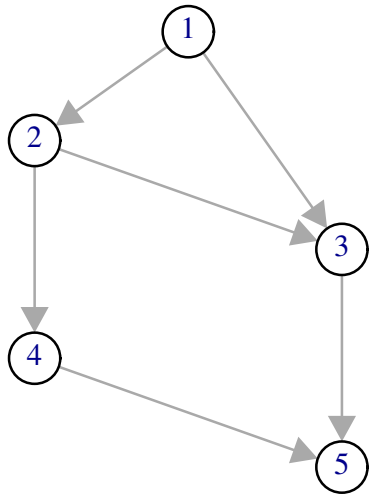


Figure 1.1

(1.3), the local Markov-property requires the following conditional independences:

$$
\begin{aligned}
X_1 &\perp X_\varnothing \mid X_\varnothing \\
X_2 &\perp X_\varnothing \mid X_\varnothing \\
X_3 &\perp X_4 \mid \mathbf{X}_{12} \\
X_4 &\perp \mathbf{X}_{13} \mid X_2 \\
X_5 &\perp \mathbf{X}_{12} \mid \mathbf{X}_{34}
\end{aligned}
$$

Equation (1.4) provides us the factorization

$$f(\mathbf{x}) = f_1(x_1)\, f_{2|1}(x_2|x_1)\, f_{3|12}(x_3|\mathbf{x}_{12})\;\; f_{4|2}(x_4|x_2)\, f_{5|34}(x_5|\mathbf{x}_{34}) \quad \text{for all } x \in \mathbb{R}^5$$

For every vertex v, we choose some strict total order $<_v$ on V to order the elements of $pa(v)$ and set

$$pa(v;w) := \{u \in pa(v) \mid u <_v w, \}, \quad w \in pa(v).$$

Note that for $v, w \in V, v \neq w$, the orders $<_v$ and $<_w$ are not the same in general.

Bauer et al. (2012) proofed that P is uniquely determined by its univariate margins $P_v, v \in V$, and its (conditional) pair-copulas $C_{vw|pa(v;w)}, v \in V, w \in pa(v)$. We can now decompose f into

$$f(\mathbf{x}) = \prod_{v \in V} f_v(x_v) \prod_{w \in pa(v)} c_{vw|pa(v;w)}(F_{v|pa(v;w)}(x_v|\mathbf{x}_{pa(v;w)}), F_{w|pa(v;w)}(x_w|\mathbf{x}_{pa(v;w)})|\mathbf{x}_{pa(v;w)}), \tag{1.5}$$

The gain over regular vines is that with a DAG $D = (V, E), |V| = d$ there are not $\binom{d}{2}$ copulas, but only $\sum_{v \in V} pa(v) = E$. We also have the freedom to choose different parent orderings $<_v$ which leads to different decompositions.

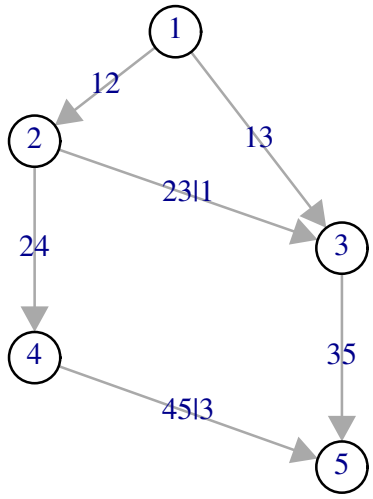


Figure 1.2

For the 5-dimensional DAG in Figure 1.1, we choose $<_3$ and $<_5$, so that $2 <_3 1$ and $3 <_5 4$. This means $pa(3;1) = \{2\}$ and $pa(5;4) = \{3\}$, in all other cases $pa(2;1) = pa(3;2) = pa(4;2) = pa(5;3) = \varnothing$. Then (1.5) can be evaluated to obtain the decomposition

$$\begin{aligned} f(\mathbf{x}) =& f_1(x_1) \cdots f_5(x_5)\, c_{12}(F_1(x_1), F_2(x_2))\, c_{23}(F_2(x_2), F_3(x_3))\, c_{24}(F_2(x_2), F_4(x_4)) \\ & c_{35}(F_3(x_3), F_5(x_5))\, c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)\, c_{45|3}(F_{4|3}(x_4|x_3), F_{5|3}(x_5|x_3)|x_3) \end{aligned} \tag{1.6}$$

It can be noticed that every edge $(v, w) \in D$, $v \to w$, can be associated with exactly one copula $C_{vw|pa(v;w)}$ of the decomposition, and vice versa. This will be called the *corresponding copula* of the edge (v, w). Figure 1.2 shows the indices of the corresponding copulas in our example, according to the decompostion (1.6).

Unfortunately, contrary to regular vines, when applying DAG PCCs, not all conditional cdfs can be obtained through a copula that is available in the decomposition. In some cases, the cdf has to be computed by integrating other margins.

Examining (1.6), the formula (1.1) can be applied to the conditional cdfs

$$\begin{aligned} F_{1|2}(x_1|x_2) &= \frac{\partial C_{12}(F_1(x_1), F_2(x_2))}{\partial F_2(x_2)} \\ F_{3|2}(x_3|x_2) &= \frac{\partial C_{32}(F_3(x_3), F_2(x_2))}{\partial F_2(x_2)} \\ F_{5|3}(x_5|x_3) &= \frac{\partial C_{35}(F_3(x_3), F_5(x_5))}{\partial F_3(x_3)} \end{aligned}$$

For $F_{4|3}$ we have to integrate via other margins because C_{34} is not available. This can be done by using the conditional independency $X_3 \perp X_4 \mid X_2$.

$$
\begin{aligned}
F_{4|3} &= \int_{-\infty}^{x_4} f_{4|3}(y_4|x_3)\ \mathrm{d}y_4 \\
&= f_3^{-1}(x_3) \int_{-\infty}^{x_4} \int_{-\infty}^{\infty} f_{234}(y_2, x_3, y_4)\ \mathrm{d}y_2\ \mathrm{d}y_4 \\
&= f_3^{-1}(x_3) \int_{-\infty}^{\infty} f_2(y_2) f_{3|2}(x_3|y_2) \int_{-\infty}^{x_4} f_{4|2}(y_4|y_2)\ \mathrm{d}y_4\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} f_2(y_2) \frac{f_{23}(y_2, x_3)}{f_2(y_2) f_3(x_3)} \int_{-\infty}^{x_4} \frac{f_{24}(y_2, y_4)}{f_2(y_2)}\ \mathrm{d}y_4\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} f_2(y_2) c_{23}(F_2(y_2), F_3(x_3)) \int_{-\infty}^{x_4} c_{24}(F_2(y_2), F_4(y_4)) f_4(y_4)\ \mathrm{d}y_4\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} f_2(y_2) c_{23}(F_2(y_2), F_3(x_3)) \int_{-\infty}^{x_4} \frac{\partial^2 C_{24}(u_2, u_4)}{\partial u_2 \partial u_4}\Big|_{\substack{u_2=F_2(y_2)\\u_4=F_4(y_4)}} f_4(y_4)\ \mathrm{d}y_4\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} f_2(y_2) c_{23}(F_2(y_2), F_3(x_3)) \frac{\partial}{\partial u_2} \left[\int_{-\infty}^{x_4} \frac{\partial C_{24}(u_2, u_4)}{\partial u_4}\Big|_{\substack{u_2=F_2(y_2)\\u_4=F_4(y_4)}} f_4(y_4)\ \mathrm{d}y_4 \right]\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} f_2(y_2) c_{23}(F_2(y_2), F_3(x_3)) \frac{\partial}{\partial u_2} \left[C_{24}(u_2, F_4(x_4))\Big|_{u_2=F_2(y_2)} \right]\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} c_{23}(u_2, F_3(x_3)) \frac{\partial C_{24}(u_2, F_4(x_4))}{\partial u_2}\Big|_{u_2=F_2(y_2)} f_2(y_2)\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} c_{23}(u_2, F_3(x_3)) h_{2\underline{4}}(u_2, F_4(x_4))\Big|_{u_2=F_2(y_2)} f_2(y_2)\ \mathrm{d}y_2 \\
&= \int_{0}^{1} c_{23}(u_2, F_3(x_3)) h_{2\underline{4}}(u_2, F_4(x_4))\ \mathrm{d}u_2
\end{aligned}
$$

where we use h-functions to express conditional cdfs.

$$
\begin{aligned}
h_{\underline{i}j}(u_i, u_j) &:= F_{i|j}(u_i|u_j) = \frac{\partial C_{ij}(u_i, u_j)}{\partial u_j} \\
h_{i\underline{j}}(u_i, u_j) &:= F_{j|i}(u_j|u_i) = \frac{\partial C_{ij}(u_i, u_j)}{\partial u_i}
\end{aligned}
$$

Note that for $u_i \in [0,1] : u_i = F_i(u_i)$ because we are assuming the univariate marginals X_i to be uniformly distributed on [0,1].

Higher dimensional h-funtions can be defined accordingly

$$
\begin{aligned}
h_{\underline{i}j|K}(v_i, v_j|\mathbf{u}_K) &:= \frac{\partial C_{ij|K}(v_i, v_j|\mathbf{u}_K)}{\partial v_j} \\
h_{i\underline{j}|K}(v_i, v_j|\mathbf{u}_K) &:= \frac{\partial C_{ij|K}(v_i, v_j|\mathbf{u}_K)}{\partial v_i} \\
&\text{where } v_i = F_{i|K}(u_i|\mathbf{u}_K) \text{ and } v_j = F_{j|K}(u_j|\mathbf{u}_K)
\end{aligned}
$$

We examine $F_{2|13}(u_2|u_1, u_3)$ as an example to show that it can be expressed as a h-function of h-functions now:

$$\begin{aligned}
F_{2|13}(u_2|u_1,u_3) &= \int_0^{u_2} f_{2|13}(y_2|u_1,u_3)\ \mathrm{d}y_2 \\
&= \int_0^{u_2} c_{12}(u_1,y_2)c_{23|1}(F_{2|1}(y_2|u_1),F_{3|1}(u_3|u_1)|u_1)\ \mathrm{d}y_2 \\
&= \int_0^{u_2} \frac{\partial F_{2|1}(y_2|u_1)}{\partial y_2}\frac{\partial^2 C_{23|1}(F_{2|1}(y_2|u_1),F_{3|1}(u_3|u_1)|u_1)}{\partial F_{2|1}(y_2|u_1)\ \partial F_{3|1}(u_3|u_1)}\ \mathrm{d}y_2 \\
&= \int_0^{u_2} \frac{\partial^2 C_{23|1}(F_{2|1}(y_2|u_1),F_{3|1}(u_3|u_1)|u_1)}{\partial y_2\ \partial F_{3|1}(u_3|u_1)}\ \mathrm{d}y_2 \\
&= \frac{\partial}{\partial F_{3|1}(u_3|u_1)}\int_0^{u_2}\frac{\partial}{\partial y_2}C_{23|1}(F_{2|1}(y_2|u_1),F_{3|1}(u_3|u_1)|u_1)\ \mathrm{d}y_2 \\
&= \frac{\partial C_{23|1}(F_{2|1}(y_2|u_1),F_{3|1}(u_3|u_1)|u_1)}{\partial F_{3|1}(u_3|u_1)} \\
&= h_{\underline{2}3|1}(h_{1\underline{2}}(u_1,u_2),h_{1\underline{3}}(u_1,u_3)|u_1)
\end{aligned}$$

If we choose the orders $<_v, v \in V$ of the parents differently, than it may be different conditional cdfs that have to been integrated via other margins and the integrations may be of higher or lower dimensions. The main goal of this thesis will be to compare the different decompositions by cost of computation and quality of result. In Chapter 2, we will show that given some conditional independences we cannot only choose the order of parents, but possibly also from different DAGs which all contain exactly the same conditional independences.

1.3 Decompositions and immoralities

As the computation of cdfs via integration of other margins can be very complex, it should be examined when we encounter these. Because the arguments of unconditional copulas are only the marginal cdfs, we can concentrate on conditional copulas.

The decomposition contains conditional copulas for every vertex that has multiple parents. This can be verified in equation (1.5) with:

$$v \in V, |pa(v)| > 1 \ \Rightarrow\ \exists w \in pa(v) : pa(v;w) \neq \varnothing$$

As an example let us examine a vertex k which has two parents i and j (Figure 1.3). This structure can be part of a bigger DAG - with the restriction that k has exactly those 2 parents and the parents are not independent. Equation (1.4) shows that the corresponding factor for the vertex k in the factorization is $f_{k|ij}(x_k|\mathbf{x}_{ij})$. With 2 parents, there exist 2 different parent orderings which leads to 2 different decompositions:

$$\begin{aligned}
j <_k i: &\quad f_{k|ij}(x_k|\mathbf{x}_{ij}) = c_{ik|j}(F_{i|j}(x_i|x_j),F_{k|j}(x_k|x_j),x_j)\ c_{jk}(F_j(x_j),F_k(x_k))\ f_k(x_k) \\
i <_k j: &\quad f_{k|ij}(x_k|\mathbf{x}_{ij}) = c_{jk|i}(F_{j|i}(x_j|x_i),F_{k|i}(x_k|x_i),x_i)\ c_{ik}(F_i(x_i),F_k(x_k))\ f_k(x_k)
\end{aligned}$$

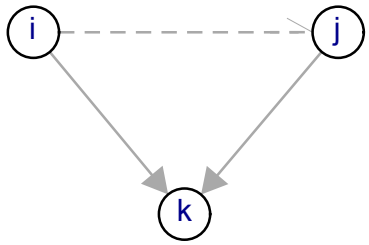


Figure 1.3

In the first decomposition the conditional cdf $F_{k|j}$ can be optained by integrating c_{jk}, and in the second decompositon $F_{k|i}$ can be optained by integrating c_{ik}. For a direct integration of both $F_{i|j}$ or $F_{j|i}$, the copula c_{ij} has to be part of the general decomposition of f. So this condition is seemingly independent from the parent ordering.

If i and j are not connected, than c_{ij} is not available in the decomposition of f, because it only contains copulas of vertices and their respective parents (see (1.5)). So in this case, a computation of $F_{i|j}$ or $F_{j|i}$ by integrating via other margins can not be avoided. We will call this structure in a graph an *immorality*.

Definition 1.2 (Immorality) *Let $D = (V, E)$ be a DAG. An immorality in a graph is a triple (i, k, j) such that $(i, k) \in E$ and $(j, k) \in E$, but $(i, j) \notin E$ and $(j, i) \notin E$.*

If i and j are indeed connected, $c_{ij|K}$ with some, possibly empty, conditioning set K is part of the decomposition. Now let h be another vertex, a parent of i and j. Then there exist both decompositions where the unconditional copula c_{ij} is available and where it is not available, see Figure 1.4 where the copulas of the decomposition are attached to the corresponding edges. In this case, a more complex integration can possibly be avoided by choosing the right decomposition.

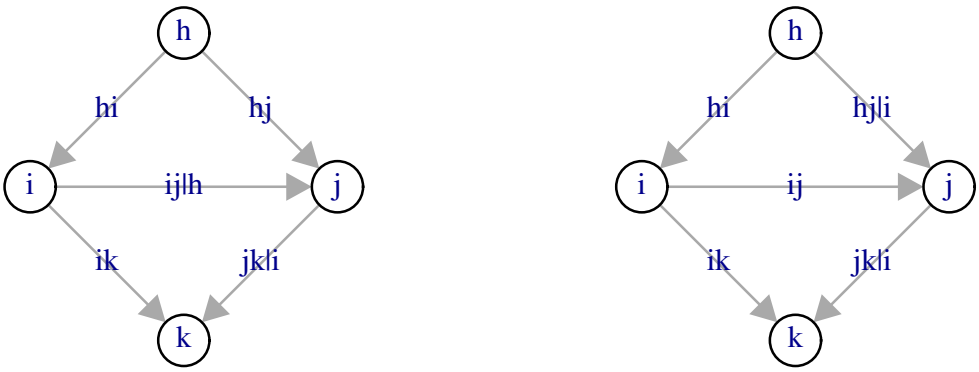


Figure 1.4

Chapter 2

Possible models based on conditional independence assumptions

If we want to apply a pair-copula construction to a dataset with conditional independence assumptions, in most cases there are multiple possible models to choose from (there may be even none). At first we need to find DAGs that match the conditional independences. This is done using an algorithm provided by Koski and Noble (2009, p. 173-176). Then for each DAG there exist different decompositions in the non-trivial case.

2.1 Markov equivalent DAGs

Definition 2.1 (Markov equivalance, I-equivalence) *Let D_1 and D_2 be two DAGs over the same variables. D_1 and D_2 are called Markov equivalent or I-equivalent if the conditional independences induced by the local D-Markov property, see* (1.3)*, are the same for both DAGs.*

Let X_1, X_2, X_3, X_4, X_5 be $\mathbb{R}$-valued random variables with the same conditional independences as our example in Chapter 1:

$$\begin{aligned} X_3 &\perp X_4 \mid \mathbf{X}_{12} \\ X_4 &\perp \mathbf{X}_{13} \mid X_2 \\ X_5 &\perp \mathbf{X}_{12} \mid \mathbf{X}_{34} \end{aligned} \tag{2.1}$$

The first step is to create an undirected graph as a skeleton for the directed graphs. The following Lemma shows how to determine this skeleton and that there exists exactly one.

Lemma 2.2 *Let X be a $\mathbb{R}^d$-valued random variable distributed as P and $D = (V, E)$ a DAG with $|V| = d$ that induces only the conditional independences so that P is D-Markovian regarding D and no additional conditional independences exist.*
Then for $v, w \in V, v \neq e$: v is not connected with w if and only if $X_v \perp X_w \mid \boldsymbol{X}_K$ for some set $K \subset V \backslash \{v, e\}$ which can be empty.

Proof "→" Let v and w be connected.
$\Rightarrow v \notin nd(w)\backslash pa(w)$ and $w \notin nd(v)\backslash pa(v)$
$\Rightarrow$ (1.3) does not induce a conditional independence between X_i and X_j.
"←" Let v and w be not connected and without loss of generality $v \notin de(w)$.
$\Rightarrow v \in nd(w)\backslash pa(w)$
$\Rightarrow$ By (1.3): $X_w \perp X_v \mid \mathbf{X}_{pa(w)}$ □

Using this Lemma, we obtain the skeleton S displayed in Figure 2.1 for the conditional independences given in (2.1).

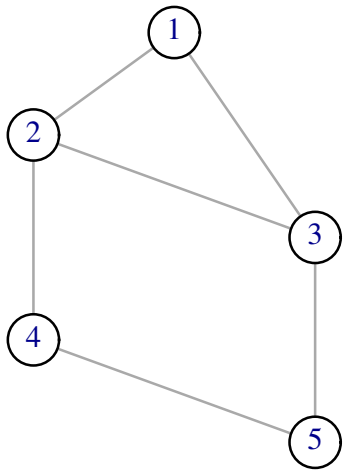


Figure 2.1: skeleton S

In the next step we are determining all immoralities in our DAG. Immoralities are also given with the independence structure, so all possible DAGs will have the same. This is a very important result, because those immoralities (i, k, j), where i and j are not independent, often lead to a decomposition where at least one conditional cdf has to be integrated via other margins, see Chapter 5.

Lemma 2.3 *Let D and P be defined as in Lemma 2.2. Then (i, k, j) is an immorality if and only if*

(i) i is connected with k

(ii) j is connected with k

(iii) i and j are not connected

(iv) $\nexists K \subset V, k \in K : X_i \perp X_j \mid \boldsymbol{X}_K$

Proof *"→" $(i), (ii), (iii)$ apply by definition*
As k is neither parent of i or j, the D-Markov property (1.3) does not induce a conditional

independence as in (iv).
"←" If we direct $k \to i$, then by (1.3) $X_i \perp X_j \mid X_{pa(i)}$ with $k \in pa(i)$. This conflicts with (iv).
If we direct $k \to j$, we find a conditional independence analogously.
$\Rightarrow i \to k$ and $j \to k$. This makes (i, k, j) an immorality. □

In our example vertex 3 is not connected with vertex 4, and both are connected with vertex 5 and there does not exist a $K \subset V, 5 \in K$ such that $X_3 \perp X_4 \mid \mathbf{X}_K$. Thus (3, 5, 4) is an immorality. As these are the only vertices with this structure, there do not exist other immoralities.
The undirected edges (3, 5) and (4, 5) can now be replaced with directed edges and this returns a graph which is partially directed (Figure 2.2). We will call this the corresponding *partially directed graph (pdg).*

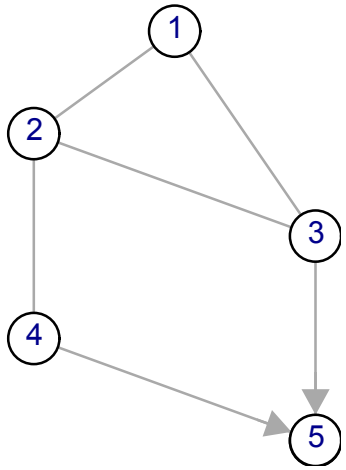


Figure 2.2: Partially directed graph (pdg)

All other edges can now be directed arbitrarily as long as this does not create a circle or additional immoralities. Based on this, Koski and Noble (2009, p. 174-175) provided 4 rules that make it possible to code an algorithm.

Based on our partially directed graph, this leads to 8 different DAGs. (Figure 2.3)
As an example let us direct $3 \to 1$ and $1 \to 2$. Then we can not choose $2 \to 3$ because this would create a circle $(1, 2, 3)$. We also can not choose $4 \to 2$ because this would create a new immorality $(1, 2, 4)$. If we direct $3 \to 2$ and $2 \to 4$, then this causes no circles or new immoralities. This is DAG 5 in the Figure.

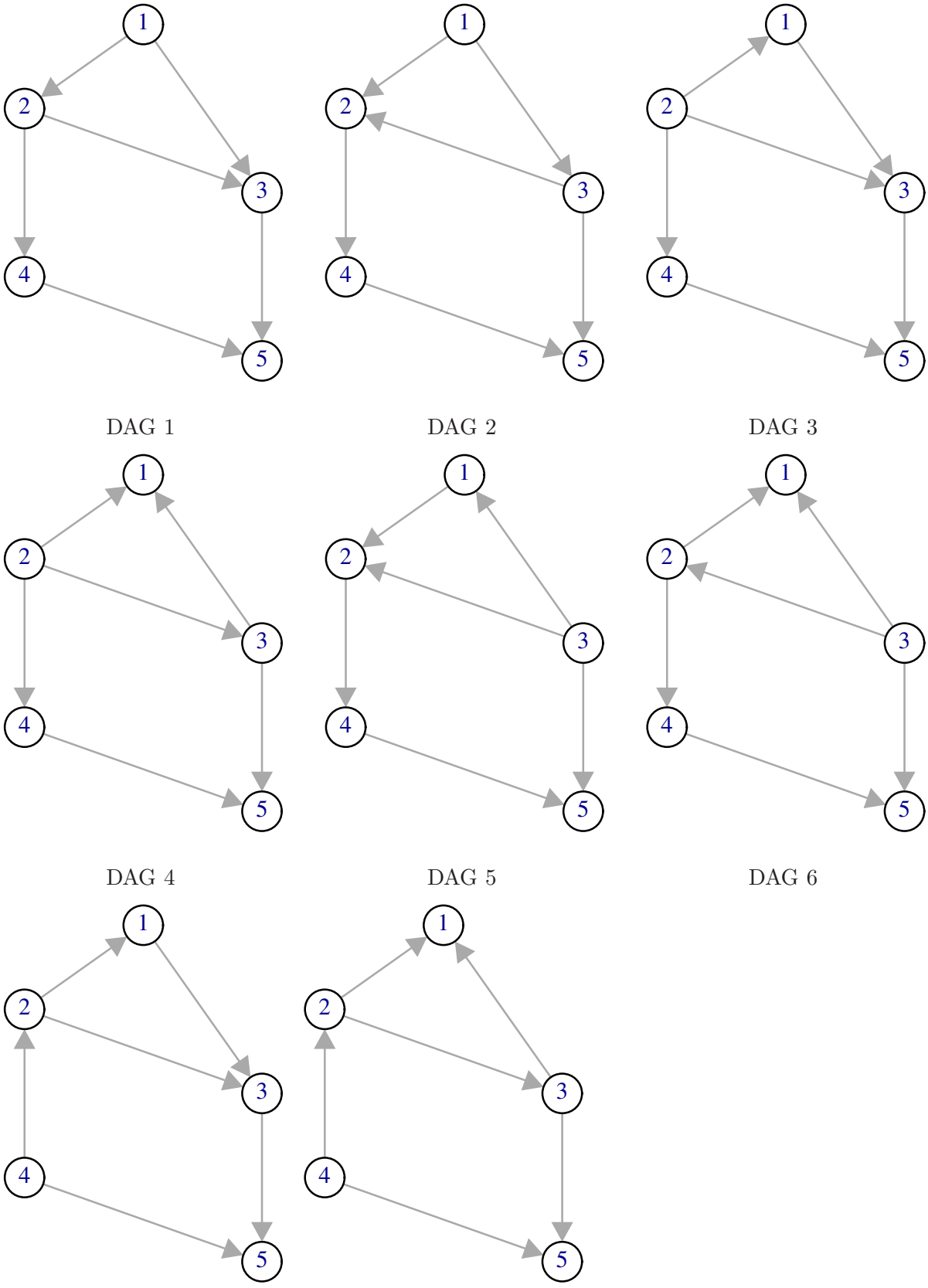


Figure 2.3: All possible DAGs based on the partially directed graph (same conditional independences)

Let P_i be probability measures with $X_i \sim P_i$ and let $f_{i|J}$ be the pdf of $P_{i|J}$. Then we can derive the following density factorizations based on the DAGs by evaluating (1.4)

DAG 1: $f(\mathbf{x}) = f_1(x_1)f_{2|1}(x_2|x_1)f_{3|12}(x_3|\mathbf{x}_{12})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 2: $f(\mathbf{x}) = f_1(x_1)f_{3|1}(x_3|x_1)f_{2|13}(x_2|\mathbf{x}_{13})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 3: $f(\mathbf{x}) = f_2(x_2)f_{1|2}(x_1|x_2)f_{3|12}(x_3|\mathbf{x}_{12})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 4: $f(\mathbf{x}) = f_2(x_2)f_{3|1}(x_3|x_1)f_{1|23}(x_1|\mathbf{x}_{23})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 5: $f(\mathbf{x}) = f_3(x_3)f_{1|3}(x_1|x_3)f_{2|13}(x_2|\mathbf{x}_{13})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 6: $f(\mathbf{x}) = f_3(x_3)f_{2|3}(x_2|x_3)f_{1|23}(x_1|\mathbf{x}_{23})f_{4|2}(x_4|x_2)f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 7: $f(\mathbf{x}) = f_4(x_4)f_{2|4}(x_2|x_4)f_{1|2}(x_1|\mathbf{x}_2)f_{3|12}(x_3|x_{12})f_{5|34}(x_5|\mathbf{x}_{34})$
DAG 8: $f(x) = f_4(x_4)f_{2|4}(x_2|x_4)f_{3|2}(x_3|\mathbf{x}_2)f_{1|23}(x_1|x_{23})f_{5|34}(x_5|\mathbf{x}_{34})$

2.2 Decomposition of DAGs as pair-copula Bayesian networks (PCBN)

If a vertex v has one parent w, then the decomposition of $f_{v|w}(x_v|x_w)$ is unique:

$$f_{v|w}(x_v|x_w) = c_{vw}(F_v(x_v), F_w(x_w))f_v(x_v)$$

If a vertex v has multiple parents $|pa(v)| > 1$, then the decomposition of $f_{v|pa(v)}(x_v|\mathbf{x}_{pa(v)})$ is not unique and has to be determined by the parent orderings $<_v$.

As an example we examine DAG 1: Vertex 3 and vertex 5 both have 2 parents. The possible parent orderings $<_3$ and $<_5$ lead to the following decompositions:

$$\begin{aligned}
2 <_3 1: &\quad f_{3|12}(x_3|x_{12}) = c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)c_{23}(F_2(x_2), F_3(x_3))f_3(x_3)\\
1 <_3 2: &\quad f_{3|12}(x_3|x_{12}) = c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)c_{13}(F_1(x_1), F_3(x_3))f_3(x_3)\\
4 <_5 3: &\quad f_{5|34}(x_5|x_{34}) = c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)|x_4)c_{45}(F_4(x_4), F_5(x_5))f_5(x_5)\\
3 <_5 4: &\quad f_{5|34}(x_5|x_{34}) = c_{45|3}(F_{4|3}(x_4|x_3), F_{5|3}(x_5|x_3)|x_3)c_{35}(F_3(x_3), F_5(x_5))f_5(x_5)
\end{aligned}$$

This leads to 4 different pair-copula decompositions of $f(x)$.

PCBN A) $f(x) = f_1(x_1)\cdots f_5(x_5)c_{12}(F_1(x_1), F_2(x_2))c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)$
$c_{23}(F_2(x_2), F_3(x_3))c_{24}(F_2(x_2), F_4(x_4))c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)|x_4)$
$c_{45}(F_4(x_4), F_5(x_5))$

PCBN B) $f(x) = f_1(x_1)\cdots f_5(x_5)c_{12}(F_1(x_1), F_2(x_2))c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)|x_2)$
$c_{23}(F_2(x_2), F_3(x_3))c_{24}(F_2(x_2), F_4(x_4))c_{45|3}(F_{4|3}(x_4|x_3), F_{5|3}(x_5|x_3)|x_3)$
$c_{35}(F_3(x_3), F_5(x_5))$

PCBN C) $f(x) = f_1(x_1)\cdots f_5(x_5)c_{12}(F_1(x_1), F_2(x_2))c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)$
$c_{13}(F_1(x_1), F_3(x_3))c_{24}(F_2(x_2), F_4(x_4))c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)|x_4)$
$c_{45}(F_4(x_4), F_5(x_5))$

PCBN D) $f(x) = f_1(x_1)\cdots f_5(x_5)c_{12}(F_1(x_1), F_2(x_2))c_{23|1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)|x_1)$
$c_{13}(F_1(x_1), F_3(x_3))c_{24}(F_2(x_2), F_4(x_4))c_{45|3}(F_{4|3}(x_4|x_3), F_{5|3}(x_5|x_3)|x_3)$
$c_{35}(F_3(x_3), F_5(x_5))$

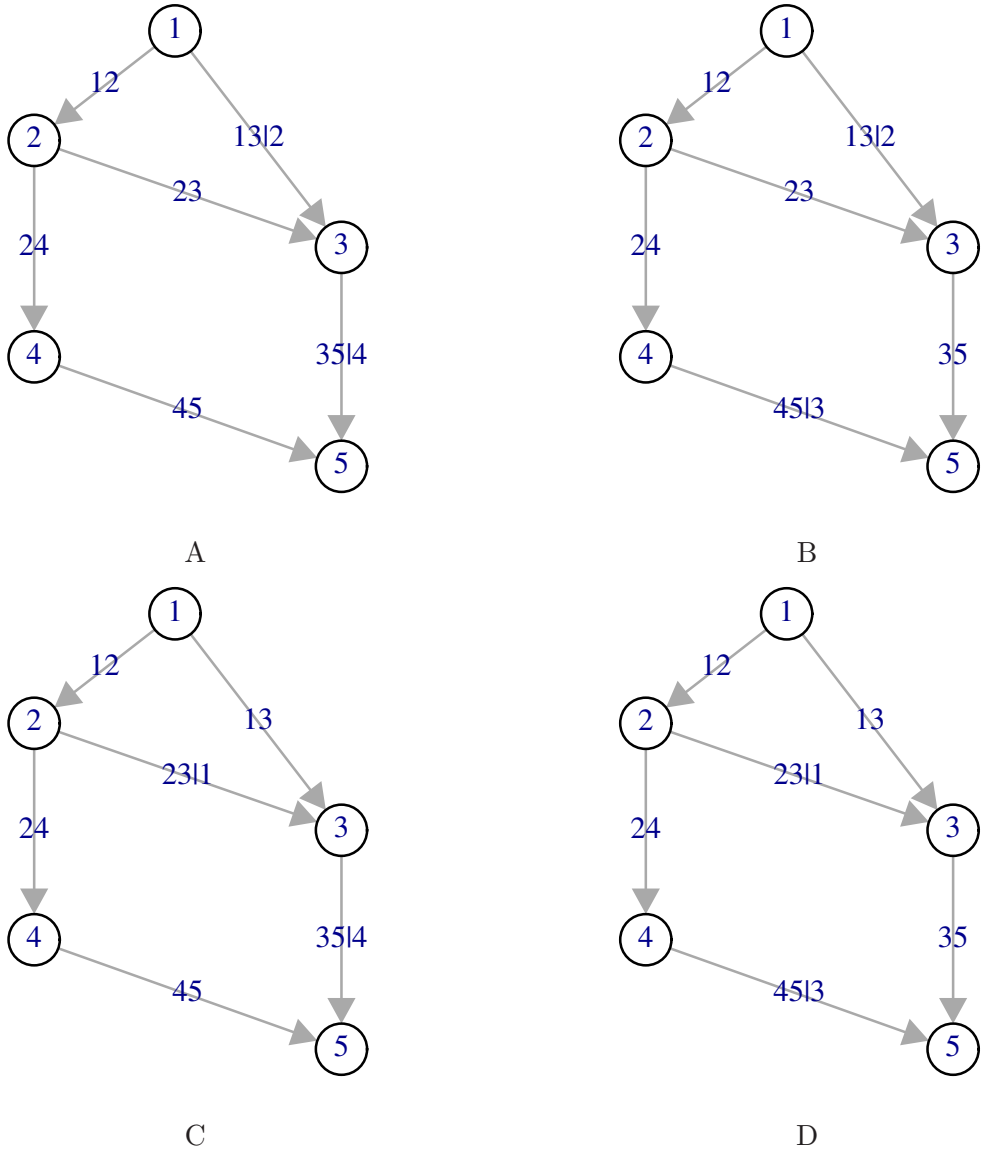


Figure 2.4: PCBN decompositions for DAG 1
(using different parent orderings)

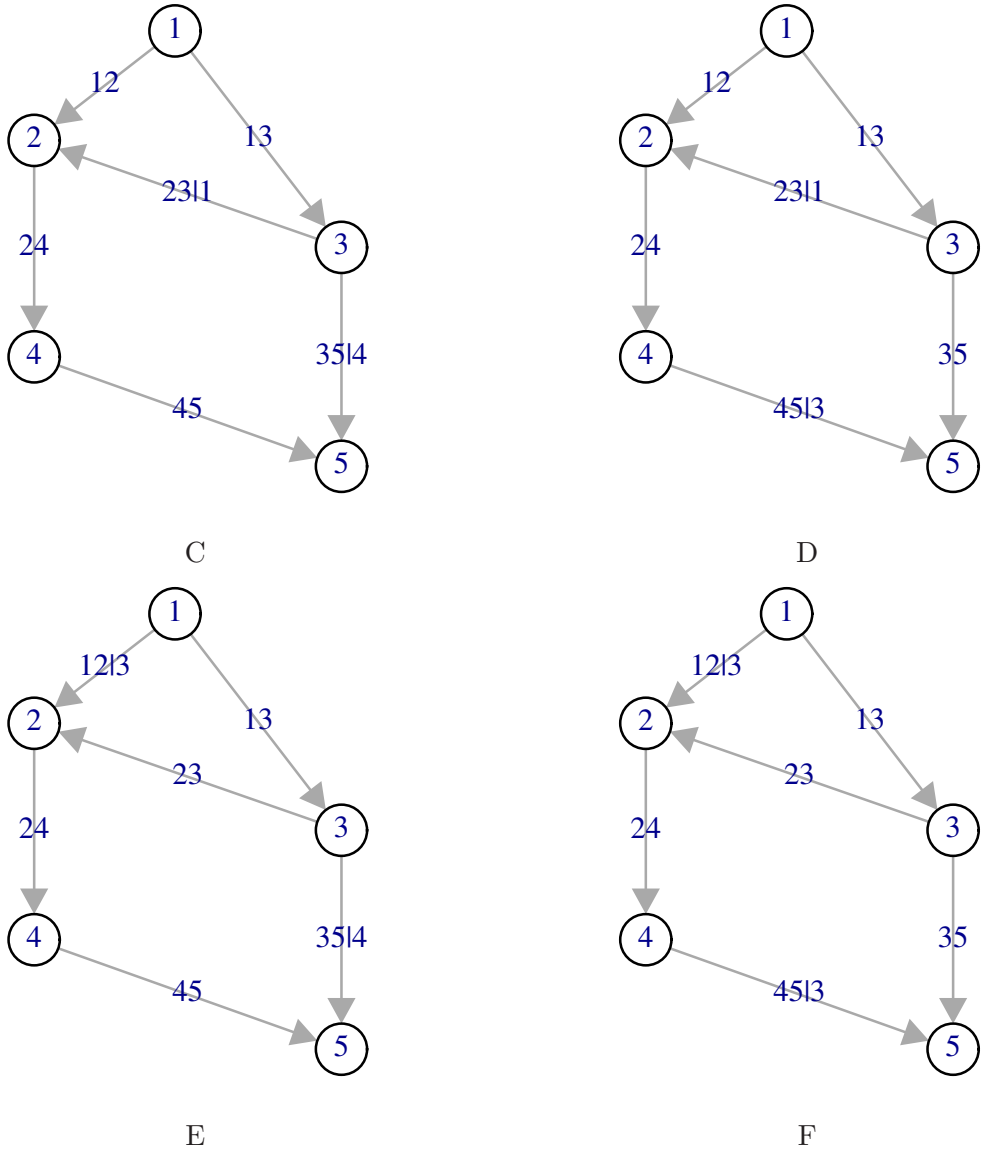


Figure 2.5: PCBN decompositions for DAG 2
(using different parent orderings)

By attaching copulas to corresponding edges, we can also easily determine possible decompositions by simply looking at the DAGs. When a vertex v has only one parent w, the corresponding pdf $f_{v|w}$ has only one decomposition and we can attach c_{vw} to that edge. In the case of multiple parents, we can order them differently. In the DAGs of our example (Figure 2.3) there are always 2 vertices who have each 2 parents (vertex 5 and one of vertex 1, 2 or 3). With 2 parents, the edge to one parent will be unconditional, while the other edge is conditioned on the first parent. It can be verified this way that the DAGs 3 and 7 have the same possible decomposition as DAG 1, in all of them the vertices 3 and 5 are the vertices with 2 parents.

In the DAGs 2, 4, 5, 6 and 8, either vertex 1 or 2 has multiple parents. This yields 2 new decomposition using the copula $c_{12|3}$:

PCBN E) $f(x) = f_1(x_1) \cdots f_5(x_5) c_{13}(F_1(x_1), F_3(x_3)) c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)|x_3)$
$c_{23}(F_2(x_2), F_3(x_3)) c_{24}(F_2(x_2), F_4(x_4)) c_{35|4}(F_{3|4}(x_3|x_4), F_{5|4}(x_5|x_4)|x_4)$
$c_{45}(F_4(x_4), F_5(x_5))$

PCBN F) $f(x) = f_1(x_1) \cdots f_5(x_5) c_{13}(F_1(x_1), F_3(x_3)) c_{12|3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3)|x_3)$
$c_{23}(F_2(x_2), F_3(x_3)) c_{24}(F_2(x_2), F_4(x_4)) c_{45|3}(F_{4|3}(x_4|x_3), F_{5|3}(x_5|x_3)|x_3)$
$c_{35}(F_3(x_3), F_5(x_5))$

Now we can list all possible PCBN decompositions for each DAG:

- DAGs 1, 3, 7: A, B, C, D
- DAGs 2, 5: C, D, E, F
- DAGs 4, 6, 8: A, B, E, F

We call two DAGs *decomposition equivalent* if they have the same possible PCBN decompositions. Note that two DAGs can share some PCBN decompositions even if they are not decomposition equivalent.

2.3 Conditional cdfs corresponding to pair-copula decomposition

Now we can check which conditional cdfs in these PCBN decompositions have to be computed by integrating other margins:

PCBN A) $F_{1|2}$, $F_{3|2}$ and $F_{5|4}$ can be computed using c_{12},c_{23} and c_{45} respectively. $F_{3|4}$ has to be computed by integrating via c_{23} and c_{24}.

PCBN B) $F_{1|2}$, $F_{3|2}$ and $F_{5|3}$ can be computed using c_{12},c_{23} and c_{35} respectively. $F_{4|3}$ has to be computed by integrating via c_{23} and c_{24}.

PCBN C) $F_{2|1}$, $F_{3|1}$ and $F_{5|4}$ can be computed using c_{12},c_{13} and c_{45} respectively. $F_{3|4}$ has to be computed by integrating via c_{12}, c_{13}, $c_{23|1}$ and c_{24}.

PCBN D) $F_{2|1}$, $F_{3|1}$ and $F_{5|3}$ can be computed using c_{12},c_{13} and c_{35} respectively. $F_{4|3}$ has to be computed by integrating via c_{12}, c_{13}, $c_{23|1}$ and c_{24}.

PCBN E) $F_{1|3}$, $F_{2|3}$ and $F_{5|4}$ can be computed using c_{13},c_{23} and c_{45} respectively. $F_{3|4}$ has to be computed by integrating via c_{23} and c_{24}.

PCBN F) $F_{1|3}$, $F_{2|3}$ and $F_{5|3}$ can be computed using c_{13},c_{23} and c_{35} respectively. $F_{4|3}$ has to be computed by integrating via c_{23} and c_{24}.

We can notice that in every PCBN decomposition either $F_{3|4}$ or $F_{4|3}$ has to be computed. And (X_4, X_5, X_3) is the only immorality in these DAGs.

We already determined the integration of $F_{4|3}$ for PCBN B in Section 1.2. This is also valid for PCBN F because both c_{23} and c_{24} are available.

$$\text{PCBN B,F: } F_{4|3} = \int_0^1 c_{23}(u_2, F_3(x_3)) h_{\underline{2}4}(u_2, F_4(x_4)) \, \mathrm{d}u_2$$

The integration of $F_{3|4}$ in PCBN A and PCBN E goes analogue. c_{23} and c_{24} are available as well.

$$\text{PCBN A,E: } F_{3|4} = \int_0^1 c_{24}(u_2, F_4(x_4)) h_{\underline{2}3}(u_2, F_3(x_3)) \, \mathrm{d}u_2$$

The problem for PCBN C and PCBN D is that the copula c_{23} is not available, but $c_{23|1}$. Therefore we have to extend our integration to the margins of vertices 1, 2, 3 and 4.

PCBN D:

$$\begin{aligned}
F_{4|3} &= \int_{-\infty}^{x_4} f_{4|3}(y_4|x_3) \, \mathrm{d}y_4 \\
&= f_3^{-1}(x_3) \int_{-\infty}^{x_4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{1234}(y_1, y_2, x_3, y_4) \, \mathrm{d}y_1 \, \mathrm{d}y_2 \, \mathrm{d}y_4 \\
&= f_3^{-1}(x_3) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(y_1) f_{2|1}(x_2|y_1) f_{3|12}(x_3|(y_1, y_2)) \int_{-\infty}^{x_4} f_{4|2}(y_4|y_2) \, \mathrm{d}y_4 \, \mathrm{d}y_1 \, \mathrm{d}y_2 \\
&= f_3^{-1}(x_3) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{12}(y_1, y_2) c_{23|1}(F_{2|1}(y_2, y_1), F_{3|1}(x_3, y_1)|y_1) f_{3|1}(x_3|y_1) F_{4|2}(x_4|y_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1^{-1}(y_1) f_2^{-1}(y_2) f_{12}(y_1, y_2) c_{23|1}(F_{2|1}(y_2, y_1), F_{3|1}(x_3, y_1)|y_1) \\
&\quad f_1^{-1}(y_1) f_3^{-1}(x_3) f_{13}(y_1, x_3) F_{4|2}(x_4|y_2) \, \mathrm{d}F_1(y_1) \, \mathrm{d}F_2(y_2) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_{12}(F_1(y_1), F_2(y_2)) c_{23|1}(F_{2|1}(y_2, y_1), F_{3|1}(x_3, y_1)|y_1) c_{13}(F_1(y_1), F_3(x_3)) \\
&\quad \frac{\partial C_{24}(F_2(y_2), F_4(x_4))}{\partial F_2(y_2)} \, \mathrm{d}F_1(y_1) \, \mathrm{d}F_2(y_2) \\
&= \int_0^1 \int_0^1 c_{12}(u_1, u_2) c_{23|1}(F_{2|1}(u_2, u_1), F_{3|1}(x_3, u_1)|u_1) c_{13}(u_1, F_3(x_3)) \frac{\partial C_{24}(u_2, F_4(x_4))}{\partial u_2} \, \mathrm{d}u_1 \, \mathrm{d}u_2 \\
&= \int_0^1 \int_0^1 c_{12}(u_1, u_2) c_{23|1}(F_{2|1}(u_2, u_1), F_{3|1}(x_3, u_1)|u_1) c_{13}(u_1, F_3(x_3)) h_{\underline{2}4}(u_2, F_4(x_4)) \, \mathrm{d}u_1 \, \mathrm{d}u_2
\end{aligned}$$

The computation of $F_{3|4}$ in PCBN C works differently because we have to factor out a conditional pdf $f_{3|12}(x_3|(y_1, y_2))$.

PCBN C:

$$
\begin{aligned}
F_{3|4} &= \int_{-\infty}^{x_3} f_{3|4}(y_3|x_4)\ \mathrm{d}y_3 \\
&= f_4^{-1}(x_4) \int_{-\infty}^{x_3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{1234}(y_1, y_2, y_3, x_4)\ \mathrm{d}y_1\ \mathrm{d}y_2\ \mathrm{d}y_3 \\
&= f_4^{-1}(x_4) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1(y_1) f_{2|1}(x_2|y_1) f_{4|2}(x_4|y_2) \int_{-\infty}^{x_3} f_{3|12}(y_3|(y_1, y_2))\ \mathrm{d}y_3\ \mathrm{d}y_1\ \mathrm{d}y_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_1^{-1}(y_1) f_2^{-1}(y_2) f_{12}(y_1, y_2) f_2^{-1}(y_2) f_4^{-1}(x_4) f_{24}(y_2, x_4) \\
&\quad F_{3|12}(x_3|(y_1, y_2))\ \mathrm{d}F_1(y_1)\ \mathrm{d}F_2(y_2) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_{12}(F_1(y_1), F_2(y_2)) c_{24}(F_2(y_2), F_4(x_4)) \\
&\quad \frac{\partial C_{23|1}(F_{2|1}(y_2, y_1), F_{3|1}(x_3, y_1)|y_1)}{\partial F_2(y_2)}\ \mathrm{d}F_1(y_1)\ \mathrm{d}F_2(y_2) \\
&= \int_0^1 \int_0^1 c_{12}(u_1, u_2) c_{24}(u_2, F_4(x_4)) \frac{\partial C_{23|1}(F_{2|1}(u_2, u_1), F_{3|1}(x_3, u_1)|u_1)}{\partial u_2}\ \mathrm{d}u_1\ \mathrm{d}u_2 \\
&= \int_0^1 \int_0^1 c_{12}(u_1, u_2) c_{24}(u_2, F_4(x_4)) h_{2\underline{3}|1}(F_{2|1}(u_2, u_1), F_{3|1}(x_3, u_1)|u_1)\ \mathrm{d}u_1\ \mathrm{d}u_2
\end{aligned}
$$

Chapter 3

Simulation studies

Computation time is not the only reason for a decision. Different decompositions could have highly varying likelihoods for a given dataset. This will be shown in the following 3 case studies.

In each case study we were assigning PCBN A (see Figure 2.4) with certain copula families and parameters. We restricted the use of copula families to Gauss, Gumbel and Clayton. Using the assigned model, 20 simulation runs with a sample size of $N = 5000$ were made. In each run the sampled data was used to fit the following models:

- Attt: PCBN A, true pdg (partially directed graph), true copulas, true parameters
- Atte: PCBN A, true pdg, true copulas, estimated parameters
- Atee: PCBN A, true pdg, estimated copulas, estimated parameters
- Btee: PCBN B, true pdg, estimated copulas, estimated parameters
- Ctee: PCBN B, true pdg, estimated copulas, estimated parameters
- Dtee: PCBN D, true pdg, estimated copulas, estimated parameters
- Etee: PCBN E, true pdg, estimated copulas, estimated parameters
- Ftee: PCBN F, true pdg, estimated copulas, estimated parameters
- Reee: RVine, estimated vine structure, estimated copulas, estimated parameters

As shown in Section 1.3, the PCBNs A-F are the exact 6 PCBNs that are based on the same pdg as A. Note that not all PCBNs can be obtained by the same DAG. PCBNs A-D were applied on DAG 1, PCBNs E and F on DAG 2. In comparison to the PCBNs, we also fitted a RVine to the sampled data.

When estimating copulas, the copula families were chosen sequentially by minimizing the AIC. AIC is the *Akaike information criterion* which is used for determining a well fitting statistical model. For data with sample size N it values not only the maximized

likelihood L of this model, but also the numbers of parameters k as penalty. The *Bayesian information criterion* (BIC) penalizes the number of parameters even stronger.

$$AIC = 2k - 2\ln(L)$$
$$BIC = \ln(N)k - 2\ln(L)$$

After all families were chosen, the parameters were jointly estimated by maximum-likelihood estimation (MLE).

Special focus was put on copula c_{13} which is not available in the PCBNs A and B, but in the PCBNs C, D, E and F. Therefore c_{13} has to be estimated in those models. But c_{13} can also be integrated via the margins c_{12}, c_{23} and $c_{13|2}$.

$$c_{13}(u_1, u_3) = \int_0^1 c_{12}(u_1, u_2)c_{23}(u_2, u_3)c_{13|2}(F_{1|2}(u_1|u_2), F_{3|2}(u_3|u_2)|u_2)\ \mathrm{d}u_2 \tag{3.1}$$

A normalized contour plot of a copula $c_{ij}(u_i, u_j)$ is a contour plot of the joint density $f_{ij}(x_i, x_j) = c_{ij}(F_i(x_i), F_j(x_j))f_i(x_i)f_j(x_j)$ under the assumption that the margins are standard normally distributed. We compared the normalized contour plots of:

(a) bivariate kernel estimation utilizing 5000 sampled (u_1, u_3) values from model *Attt*

(b) parametric bivariate copula family and parameter estimates utilizing 5000 sampled (u_1, u_3) values from model *Attt*

(c) numerically integrated of (3.1) using PCBN model *Attt* sepcification

Especially the comparison of the estimated copula (b) and the integrated copula (c) is a good indicator if the sampled data can be fitted well on another model. In case studies 2 and 3 the copula families and parameters were especially chosen so that c_{13} can not be estimated well by a given copula family.

3.1 Case study 1

In the first study only Gumbel copulas were used. It will be shown that in this case, the copula c_{13} can be estimated well and the likelihoods of the estimated DAG models do not vary much and are on the same level as the true model (Attt) and the RVine.

model	copula	type	par	tau
Attt	c_{12}	G	2.00	0.50
	$c_{13\|2}$	G	1.67	0.40
	c_{23}	G	2.00	0.50
	c_{24}	G	1.43	0.30
	$c_{35\|4}$	G	1.11	0.10
	c_{45}	G	1.67	0.40

Table 3.1: Selected copulas and parameters for PCBN A

The simulated data of the first run is illustrated in Figure 3.1 with pair plots and normalized contour plots.

Figure 3.2 compares the integrated and estimated copulas c_{13}, also in the first run. The estimated copula is a Gumbel copula as well and it can be seen in this figure that it fits the integrated copula almost perfectly.

If all copulas in other models can be well estimated, we might assume that the log-likelihoods of the different models do not vary a lot. Figure 3.3 shows the log-likelihoods of each of the 20 runs in a seperate line. They mostly stay on the same level, only model *Dtee* has a somewhat smaller log-likelihood. Also the estimated RVine does not generally have a higher likelihood than the estimated DAG models.

For the first run, the selected copula families and estimated parameters can be found in Figure 3.4. The selected families of the other DAG-models are almost exclusively Gumbel copulas. Only $c_{12|3}$ is selected as Clayton 180° which is also similar to a Gumbel copula.

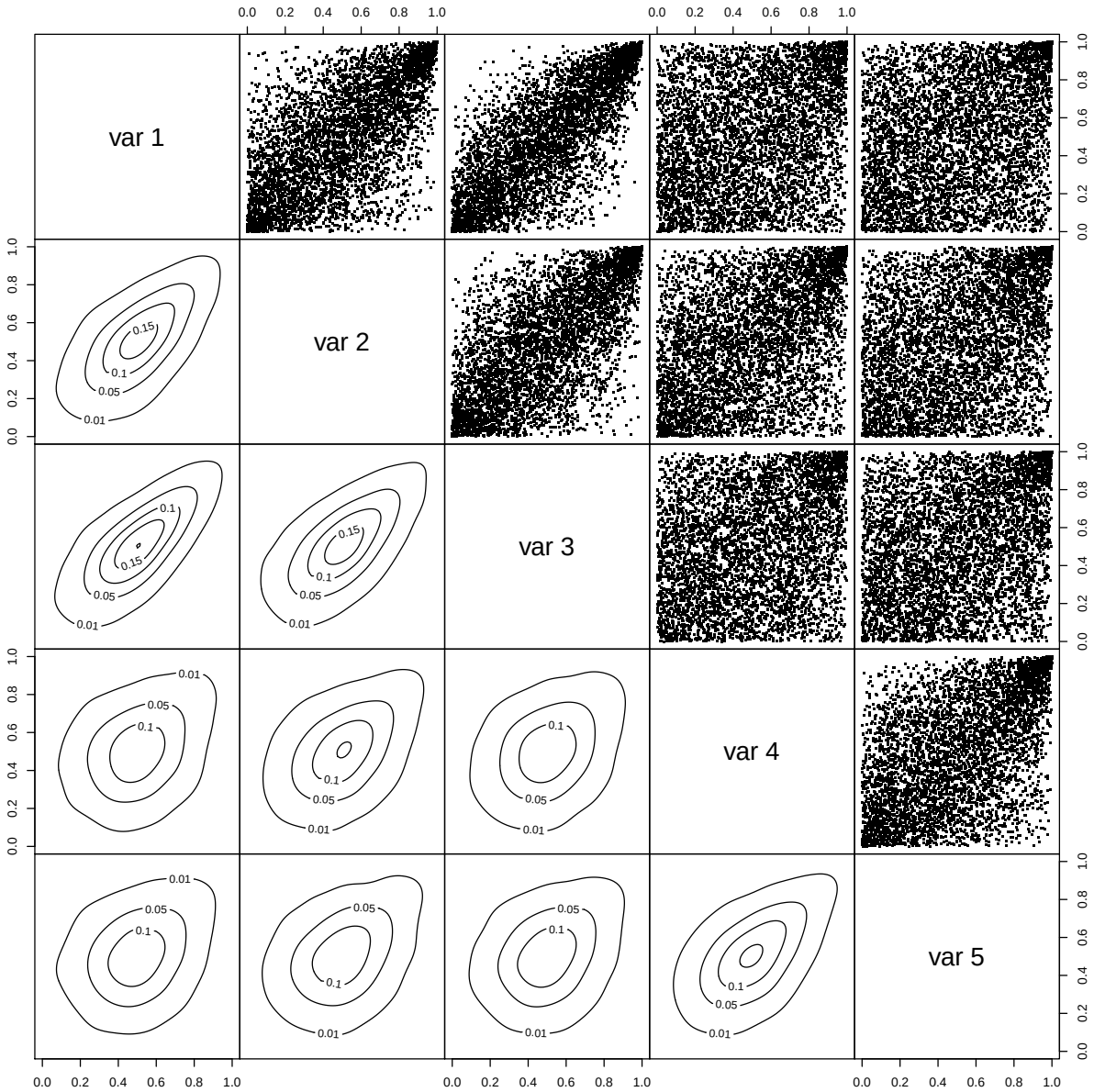
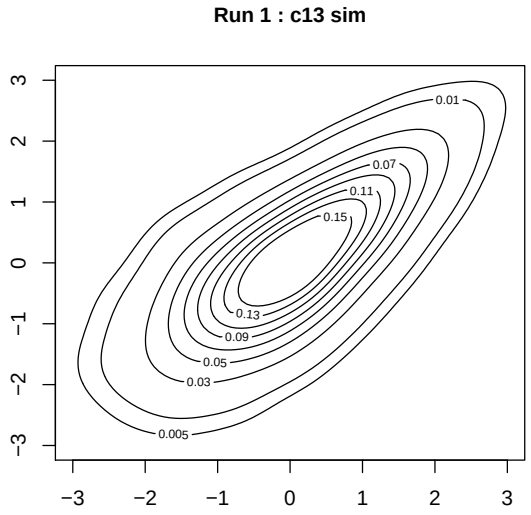
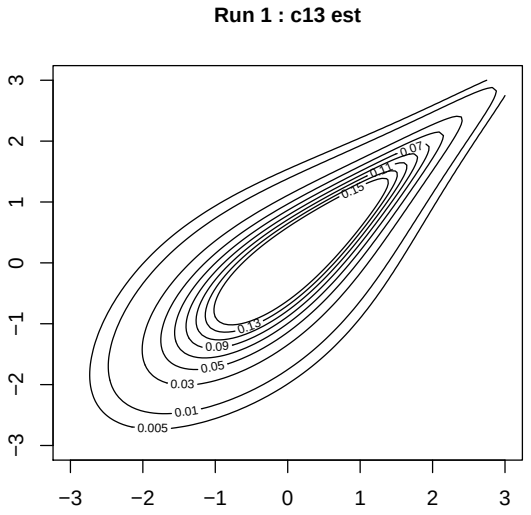


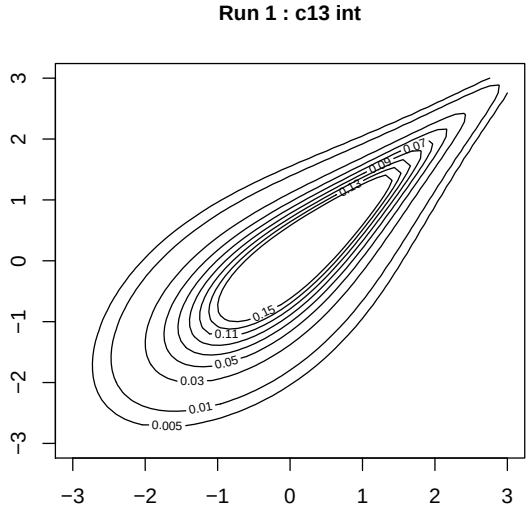
Figure 3.1: upper right: pairs plots of sampled values from model *Attt* (N=5000) bottom left: normalized contour plots based on bivariate kernel estimation utilizing the sampled values from model *Attt*



(a)



(b)



(c)

Figure 3.2: normalized contour plots based on
(a) bivariate kernel estimation utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(b) parametric bivariate copula family and parameter estimates utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(c) numerically integrated of (3.1) using model *Attt* sepcification

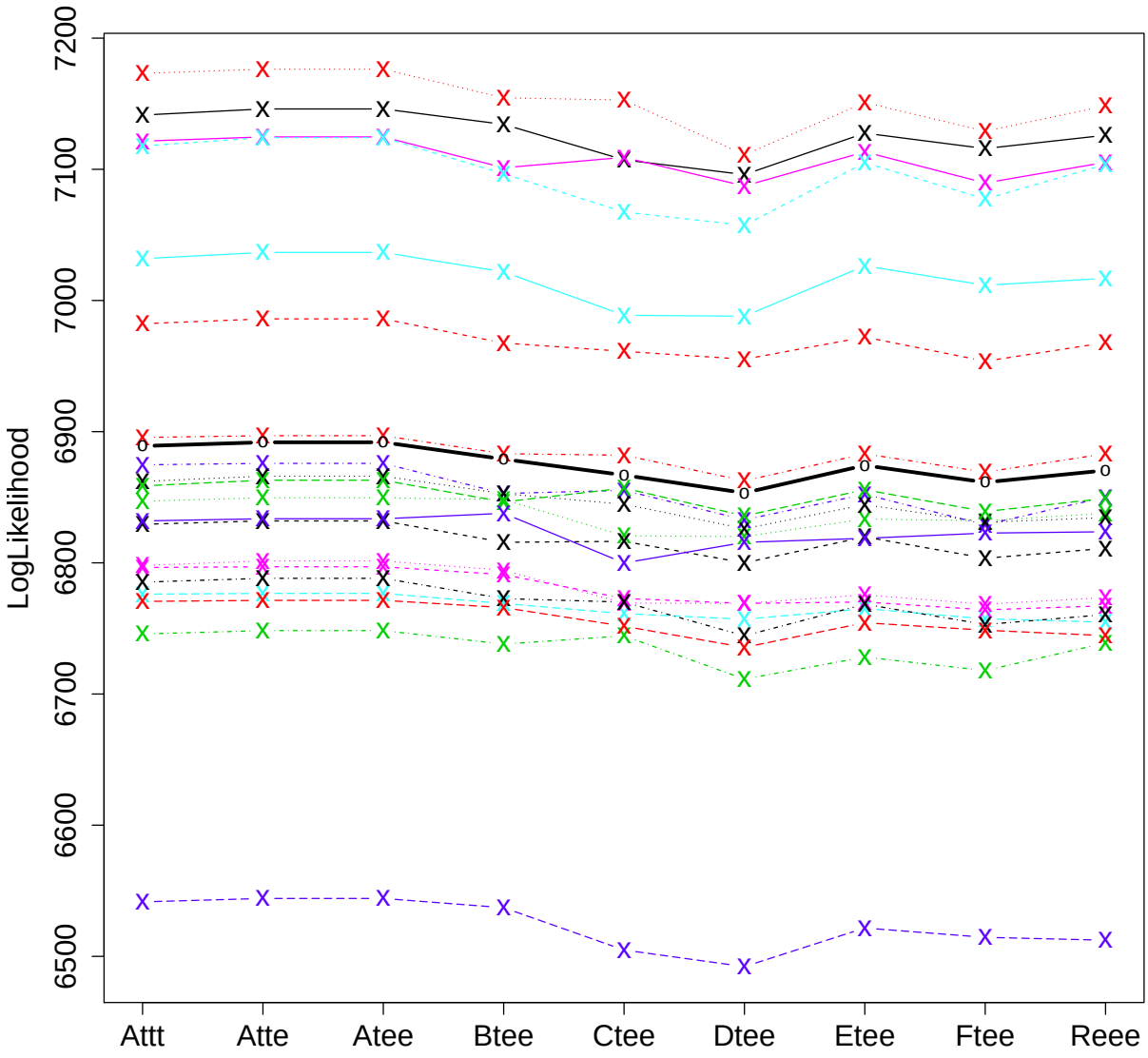


Figure 3.3: LogLikelihood values of 20 simulation runs with N=5000.
One line represents one run. The thick black line represents the means of each model.

model	copula	type	par	tau
Atte	c_{12}	G	1.99	0.50
	$c_{13\|2}$	G	1.68	0.41
	c_{23}	G	2.02	0.50
	c_{24}	G	1.46	0.32
	$c_{35\|4}$	G	1.10	0.09
	c_{45}	G	1.65	0.40
Atee	c_{12}	G	1.99	0.50
	$c_{13\|2}$	G	1.68	0.41
	c_{23}	G	2.02	0.50
	c_{24}	G	1.46	0.32
	$c_{35\|4}$	G	1.10	0.09
	c_{45}	G	1.65	0.40
Btee	c_{12}	G	1.99	0.50
	$c_{13\|2}$	G	1.68	0.41
	c_{23}	G	2.01	0.50
	c_{24}	G	1.45	0.31
	c_{35}	G	1.25	0.20
	$c_{45\|3}$	G	1.53	0.35
Ctee	c_{12}	G	2.02	0.51
	c_{13}	G	2.46	0.59
	$c_{23\|1}$	G	1.28	0.22
	c_{24}	G	1.47	0.32
	$c_{35\|4}$	G	1.10	0.09
	c_{45}	G	1.66	0.40

model	copula	type	par	tau
Dtee	c_{12}	G	2.00	0.50
	c_{13}	G	2.44	0.59
	$c_{23\|1}$	G	1.28	0.22
	c_{24}	G	1.46	0.32
	c_{35}	G	1.25	0.20
	$c_{45\|3}$	G	1.53	0.35
Etee	$c_{12\|3}$	C180°	0.50	0.20
	c_{13}	G	2.45	0.59
	c_{23}	G	2.02	0.50
	c_{24}	G	1.46	0.32
	$c_{35\|4}$	G	1.10	0.09
	c_{45}	G	1.65	0.40
Ftee	$c_{12\|3}$	C180°	0.50	0.20
	c_{13}	G	2.45	0.59
	c_{23}	G	2.01	0.50
	c_{24}	G	1.45	0.31
	c_{35}	G	1.25	0.20
	$c_{45\|3}$	G	1.53	0.35
Reee	$c_{12\|3}$	C180°	0.50	0.20
	c_{13}	G	2.45	0.59
	$c_{14\|23}$	G270°	-1.01	-0.01
	$c_{15\|234}$	C270°	-0.01	0.00
	c_{23}	G	2.02	0.50
	c_{24}	G	1.46	0.32
	$c_{25\|4}$	G	1.07	0.07
	$c_{34\|2}$	G270°	-1.01	-0.01
	$c_{35\|24}$	G	1.06	0.06
	c_{45}	G	1.65	0.39

Figure 3.4: selected copulas for the remaining models
(copulas selected by AIC, parameters determined by joint estimation)

3.2 Case study 2

In the second study we applied Gumbel and Clayton copulas with different rotations. In this case c_{13} can not be estimated well by a given copula family (Gauss, Gumbel, Clayton). This results in high varying likelihoods of the estimated models.

model	copula	type	par	tau
Attt	c_{12}	G	2.5	0.60
	$c_{13\|2}$	C	4.67	0.70
	c_{23}	C180°	4.67	0.70
	c_{24}	G	1.43	0.30
	$c_{35\|4}$	G	3.33	0.70
	c_{45}	G90°	-2	-0.50

Table 3.2: Selected copulas and parameters for PCBN A

It can be seen in Figure 3.6 that the numerically integrated copula c_{13} does not fit the estimated copula, in this case a rotated Gumbel 180°.

This indicates that the log-likelihood of the estimated models with c_{13} in the decomposition is not as high as the log-likelihood of the true model. Figure 3.7 illustrates the log-likelihoods of all models and confirms this expectation.

- Model *Btee* does not have c_{13} in the decomposition. The log-likelihood is on the same level as for Attt, Atte and Atee.
- Models *Ctee* and *Dtee* have c_{13}, as well as c_{12} and $c_{23|1}$ in their decompositions. The log-likelihoods are on the level ~ 13500 while the true model is on the level ~ 20500. This means a very significant decrease of $\sim 34\%$.
- Models *Etee*, *Ftee* and the RVine *Reee* have c_{13}, $c_{12|3}$ and c_{23} in their decompositions. The log-likelihoods are on the level ~ 18000 which means a decrease of $\sim 10\%$.

It is interesting to assert that in this case the quality of the log-likelihood depends very highly on the used copulas in the triangle $(1, 2, 3)$. This is even true for the RVine which contains a higher number of total copulas.

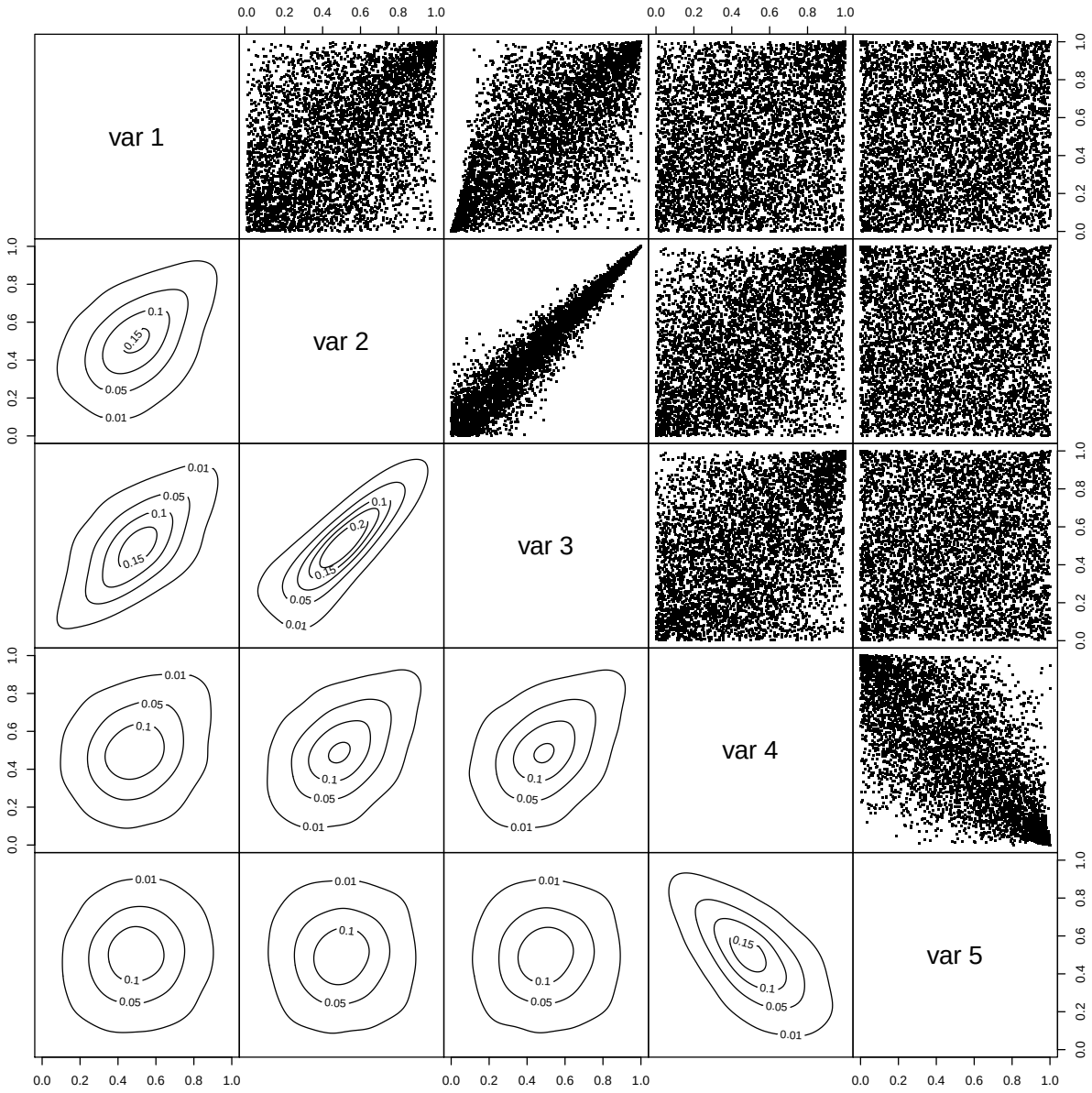
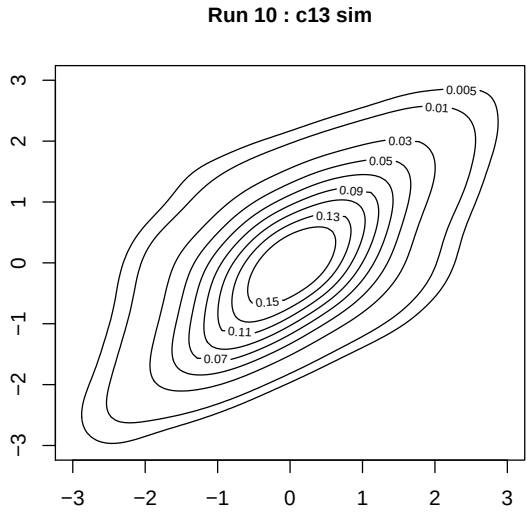
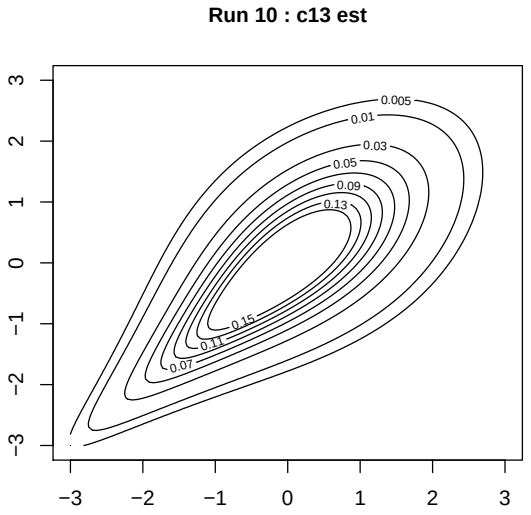


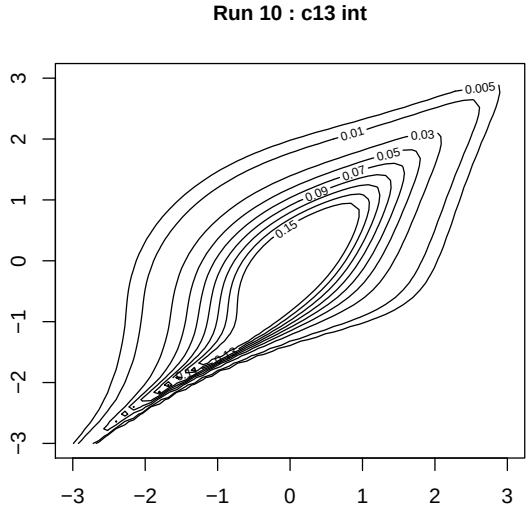
Figure 3.5: upper right: pairs plots of sampled values from model *Attt* (N=5000) bottom left: normalized contour plots based on bivariate kernel estimation utilizing the sampled values from model *Attt*



(a)



(b)



(c)

Figure 3.6: normalized contour plots based on
(a) bivariate kernel estimation utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(b) parametric bivariate copula family and parameter estimates utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(c) numerically integrated of (3.1) using model *Attt* sepcification

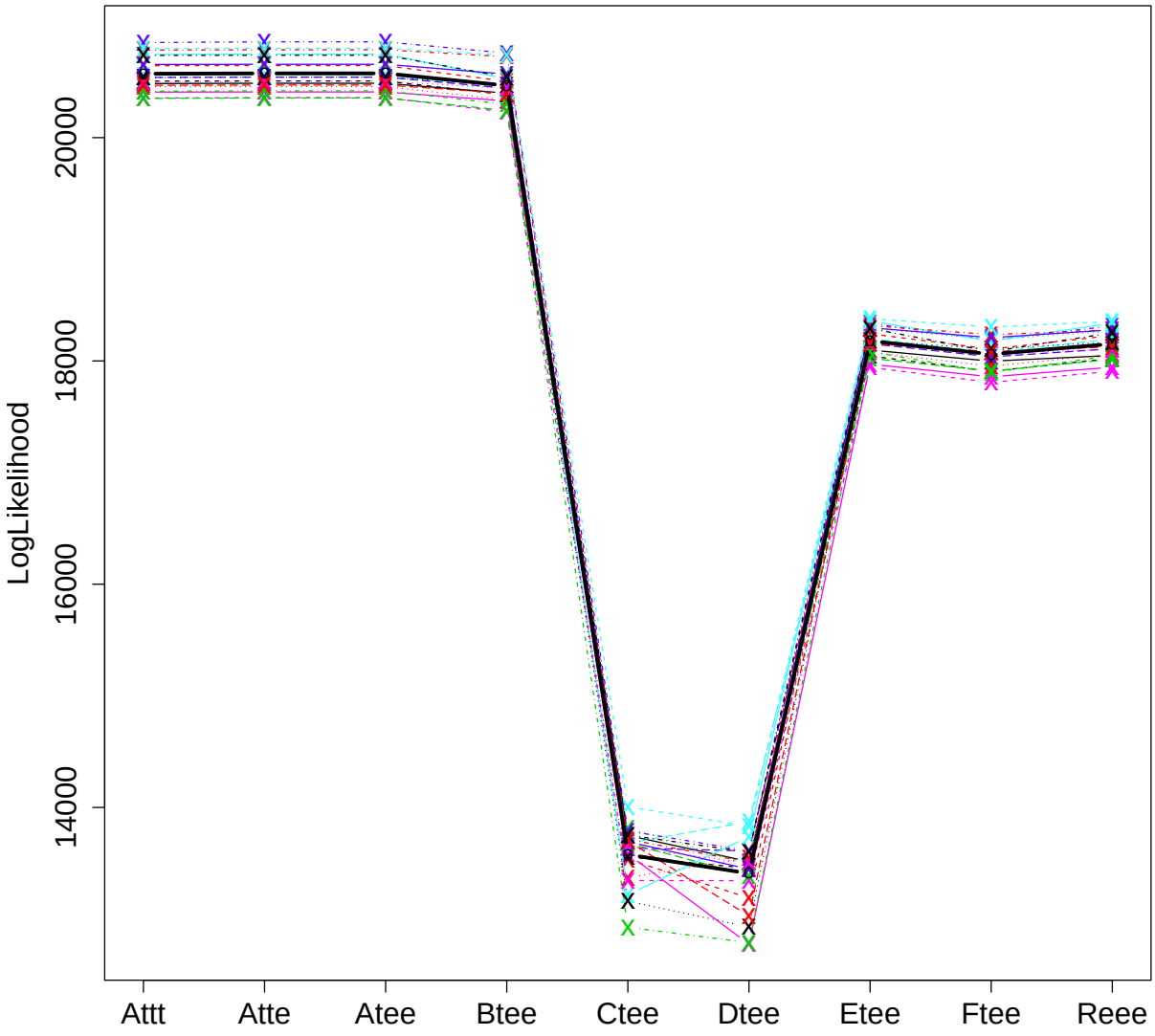


Figure 3.7: LogLikelihood values of 20 simulation runs with N=5000.
One line represents one run. The thick black line represents the means of each model.

model	copula	type	par	tau
Atee	c_{12}	G	1.54	0.35
	$c_{13\|2}$	C	11.33	0.85
	c_{23}	C180°	11.29	0.85
	c_{24}	G	1.44	0.31
	$c_{35\|4}$	G	1.66	0.40
	c_{45}	G90°	-1.99	-0.50
Atee	c_{12}	G	1.54	0.35
	$c_{13\|2}$	C	11.33	0.85
	c_{23}	C180°	11.30	0.85
	c_{24}	G	1.43	0.30
	$c_{35\|4}$	G	1.66	0.40
	c_{45}	G90°	-1.98	-0.50
Btee	c_{12}	G	1.54	0.35
	$c_{13\|2}$	C	11.33	0.85
	c_{23}	C180°	11.30	0.85
	c_{24}	G	1.43	0.30
	c_{35}	G	1.03	0.03
	$c_{45\|3}$	G90°	-2.54	-0.61
Ctee	c_{12}	G	1.54	0.35
	c_{13}	G180°	1.96	0.49
	$c_{23\|1}$	G	5.66	0.82
	c_{24}	G	1.43	0.30
	$c_{35\|4}$	G	1.66	0.40
	c_{45}	G90°	-1.98	-0.50

model	copula	type	par	tau
Dtee	c_{12}	G	1.54	0.35
	c_{13}	G180°	1.96	0.49
	$c_{23\|1}$	G	5.66	0.82
	c_{24}	G	1.43	0.30
	c_{35}	G	1.03	0.03
	$c_{45\|3}$	G90°	-2.56	-0.61
Etee	$c_{12\|3}$	G90°	-3.66	-0.73
	c_{13}	G180°	1.96	0.49
	c_{23}	C180°	11.30	0.85
	c_{24}	G	1.43	0.30
	$c_{35\|4}$	G	1.66	0.40
	c_{45}	G90°	-1.98	-0.50
Ftee	$c_{12\|3}$	G90°	-3.66	-0.73
	c_{13}	G180°	1.96	0.49
	c_{23}	C180°	11.30	0.85
	c_{24}	G	1.43	0.30
	c_{35}	G	1.03	0.03
	$c_{45\|3}$	G90°	-2.54	-0.61
Reee	$c_{12\|3}$	G270°	-3.92	-0.72
	c_{13}	G180°	1.81	0.49
	$c_{14\|23}$	C90°	-0.01	0.00
	$c_{15\|234}$	C270°	-0.02	-0.01
	c_{23}	C180°	11.65	0.85
	c_{24}	G	1.44	0.30
	$c_{25\|4}$	G	1.60	0.38
	$c_{34\|2}$	C180°	0.00	0.00
	$c_{35\|24}$	N	0.18	0.11
	c_{45}	G90°	-1.98	-0.49

Figure 3.8: selected copulas for the remaining models
(copulas selected by AIC, parameters determined by joint estimation)

3.3 Case study 3

In the third case study we applied Gumbel and Gauss copulas. Some are rotated, also in the (1,2,3)-triangle.

model	copula	type	par	tau
Attt	c_{12}	G270°	-2.50	-0.60
	$c_{13\|2}$	G	3.33	0.70
	c_{23}	G270°	-3.33	-0.70
	c_{24}	N	0.45	0.30
	$c_{35\|4}$	G	3.33	0.70
	c_{45}	G270°	-2.00	-0.50

Table 3.3: Selected copulas and parameters for PCBN A

Figure 3.10 shows like in the last study that the numerically integrated copula c_{13} has a form which cannot be described by a classical copula family. In this case the estimated copula is Gumbel copula.

The log-likelihoods are illustrated in Figure 3.11 and are varying similar in a way similar to the last study:

- Model *Btee* does not have c_{13} in the decomposition. The log-likelihood is only on a slightly lower level as for Attt, Atte and Atee.
- Models *Ctee* and *Dtee* have c_{13}, as well as c_{12} and $c_{23|1}$ in their decompositions. The log-likelihoods are on the level $\sim$ 15500 while the true model is on the level $\sim$ 17500. This means a significant decrease of $\sim$ 11%.
- Models *Etee*, *Ftee* and the RVine *Reee* have c_{13}, $c_{12|3}$ and c_{23} in their decompositions. The log-likelihoods are on the level $\sim$ 16000 which means a decrease of $\sim$ 9%.

It is also noticable in this case that the models *Btee*, *Dtee* and *Ftee* have slightly lower log-likelihoods than their counterparts *Atee*, *Ctee* and *Etee* of which the only difference respectively is the use of $c_{45|3}$ instead of c_{45} and c_{35} instead of $c_{35|4}$. The amount of the decrease is also very similar in each run.

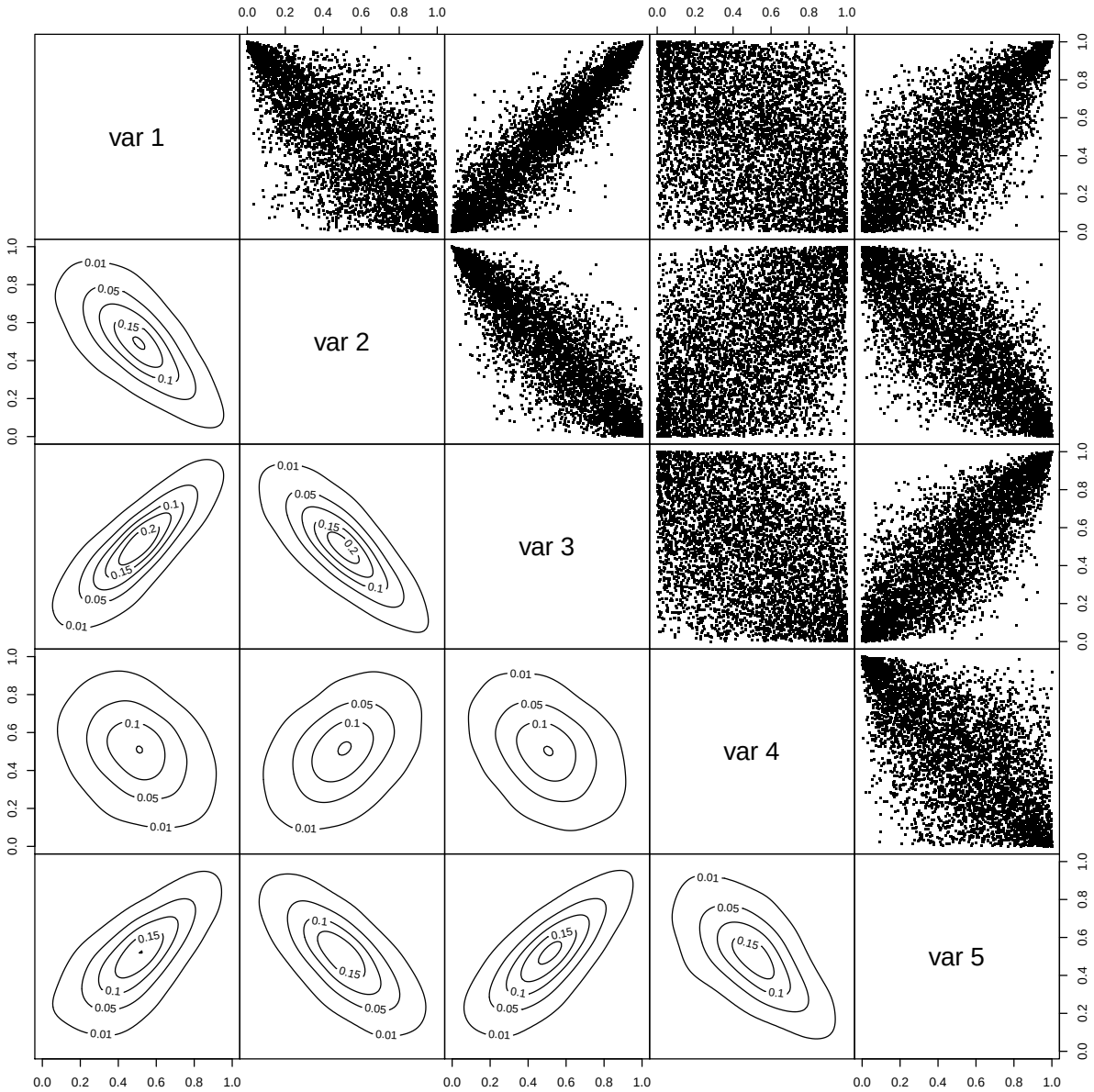
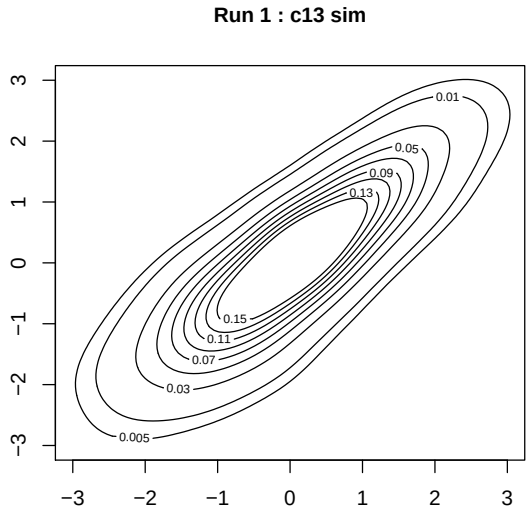
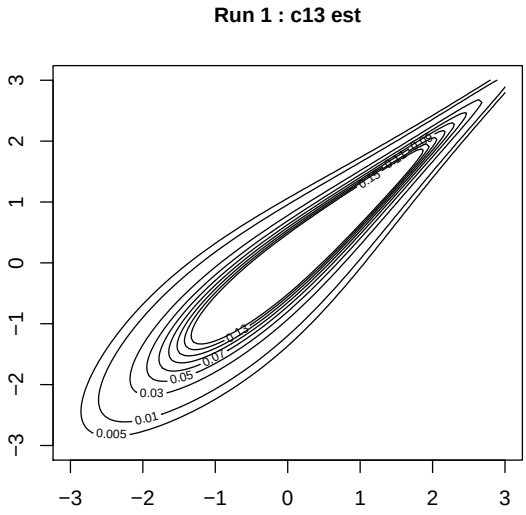


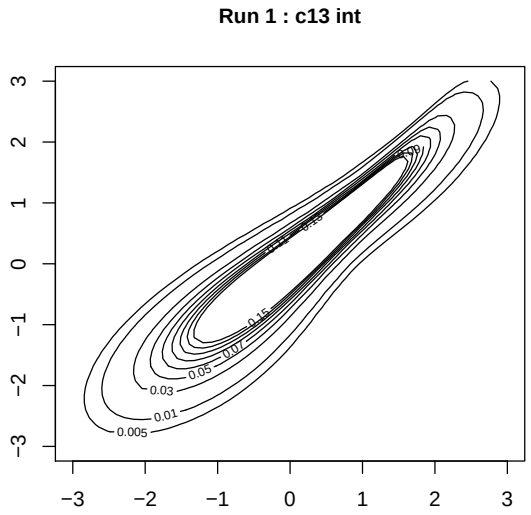
Figure 3.9: upper right: pairs plots of sampled values from model *Attt* (N=5000) bottom left: normalized contour plots based on bivariate kernel estimation utilizing the sampled values from model *Attt*



(a)



(b)



(c)

Figure 3.10: normalized contour plots based on
(a) bivariate kernel estimation utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(b) parametric bivariate copula family and parameter estimates utilizing 5000 sampled (u_1, u_3) values from model *Attt*
(c) numerically integrated of (3.1) using model *Attt* sepcification

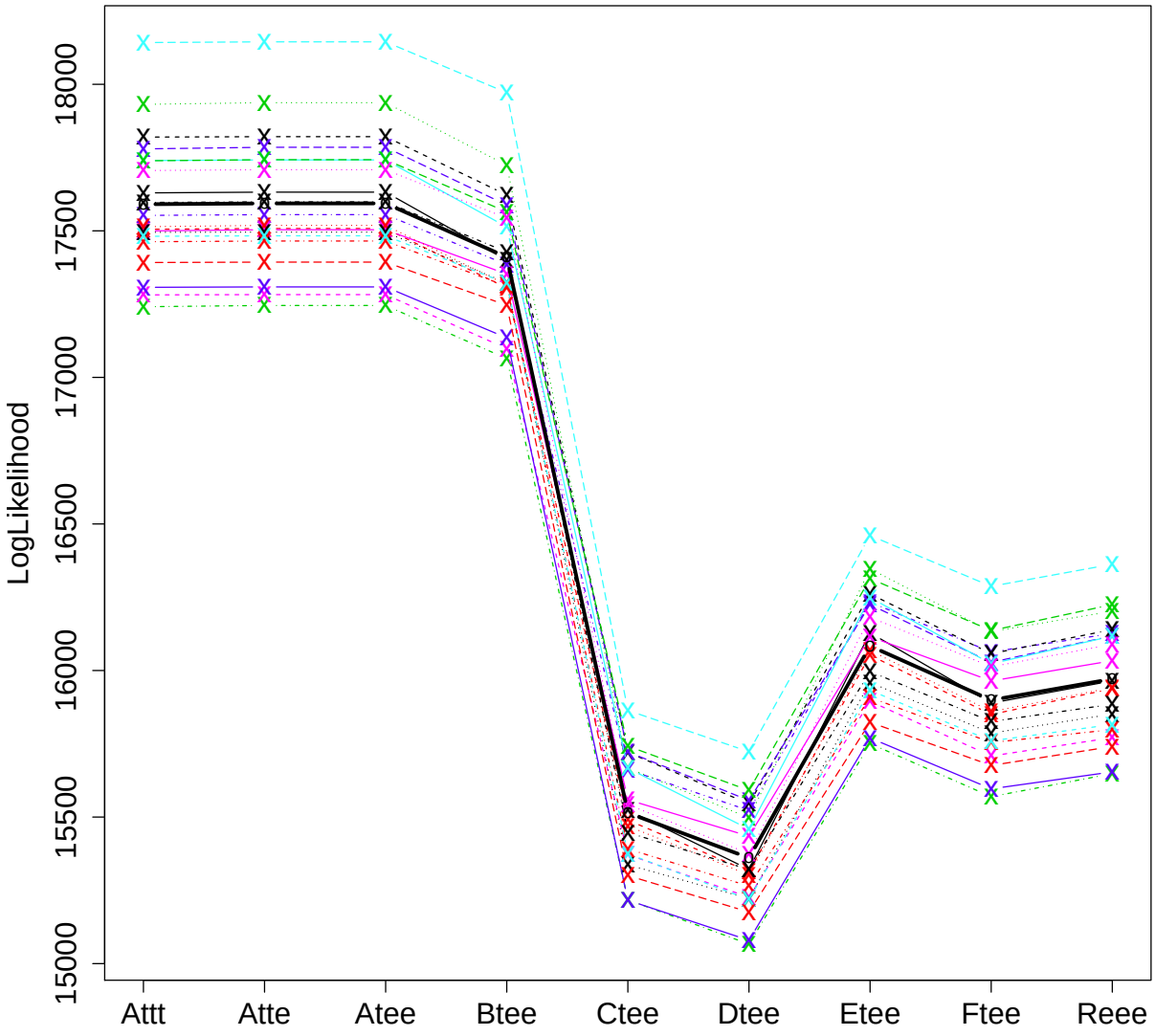


Figure 3.11: LogLikelihood values of 20 simulation runs with N=5000.
One line represents one run. The thick black line represents the means of each model.

model	copula	type	par	tau
Atte	c_{12}	G270°	-2.51	-0.60
	$c_{13\|2}$	G	3.31	0.70
	c_{23}	G270°	-3.38	-0.70
	c_{24}	N	0.45	0.30
	$c_{35\|4}$	G	3.34	0.70
	c_{45}	G270°	-1.99	-0.50
Atee	c_{12}	G270°	-2.51	-0.60
	$c_{13\|2}$	G	3.31	0.70
	c_{23}	G270°	-3.38	-0.70
	c_{24}	N	0.45	0.30
	$c_{35\|4}$	G	3.34	0.70
	c_{45}	G270°	-1.99	-0.50
Btee	c_{12}	G270°	-2.51	-0.60
	$c_{13\|2}$	G	3.32	0.70
	c_{23}	G270°	-3.37	-0.70
	c_{24}	N	0.43	0.29
	c_{35}	G	2.82	0.65
	$c_{45\|3}$	G270°	-2.26	-0.56
Ctee	c_{12}	G270°	-2.45	-0.59
	c_{13}	G	4.26	0.77
	$c_{23\|1}$	G270°	-1.54	-0.35
	c_{24}	N	0.45	0.29
	$c_{35\|4}$	G	3.34	0.70
	c_{45}	G270°	-2.01	-0.50

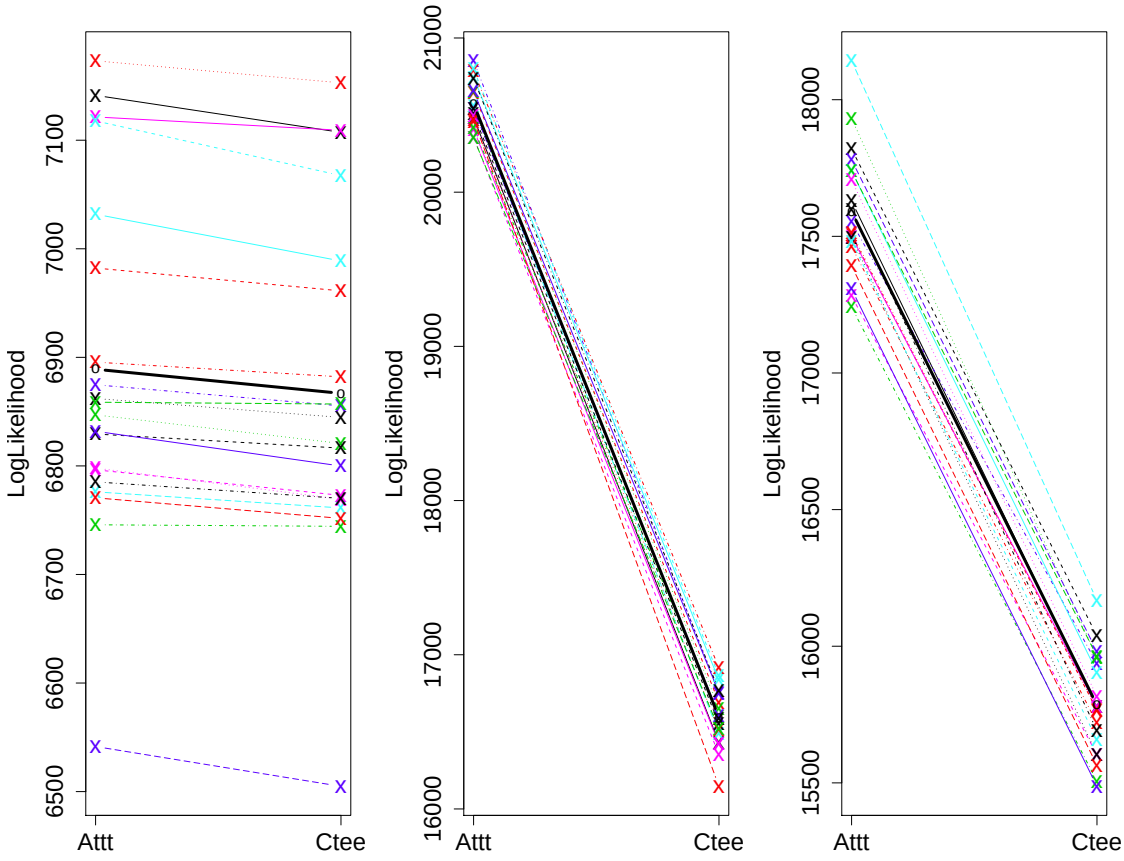
model	copula	type	par	tau
Dtee	c_{12}	G270°	-2.39	-0.58
	c_{13}	G	4.16	0.76
	$c_{23\|1}$	G270°	-1.57	-0.36
	c_{24}	N	0.42	0.28
	c_{35}	G	2.71	0.63
	$c_{45\|2}$	G270°	-2.29	-0.56
Etee	$c_{12\|3}$	G	1.12	0.11
	c_{13}	G	4.38	0.77
	c_{23}	G90°	-3.34	-0.70
	c_{24}	N	0.45	0.30
	$c_{35\|4}$	G	3.34	0.70
	c_{45}	G270°	-1.99	-0.50
Ftee	$c_{12\|3}$	G	1.12	0.11
	c_{13}	G	4.37	0.77
	c_{23}	G90°	-3.33	-0.70
	c_{24}	N	0.44	0.29
	c_{35}	G	2.82	0.64
	$c_{45\|3}$	G270°	-2.26	-0.56
Reee	$c_{12\|3}$	G	-1.12	-0.11
	c_{13}	G	4.28	0.76
	$c_{14\|235}$	G270°	-1.01	-0.01
	$c_{15\|23}$	C	0.03	0.01
	c_{23}	G90°	-3.27	-0.69
	$c_{24\|35}$	N	0.14	0.09
	$c_{25\|3}$	N	-0.19	-0.12
	$c_{34\|5}$	G	1.64	0.39
	c_{35}	G	2.83	0.64
	c_{45}	G270°	-2.02	-0.50

Figure 3.12: selected copulas for the remaining models
(copulas selected by AIC, parameters determined by joint estimation)

3.4 A second look: More copula families

Because the numerically integrated copula in case study 2 seems more fitting to a Student copula (see Figure 3.6), we wanted to test if the likelihoods of the models are still highly varying if we allow all common copula families. These include Gauss, Gumbel, Clayton, Student, Frank and Joe, as well as the rotated versions of Gumbel, Clayton and Joe. Notice that the bivariate Student copula has 2 parameters.

We restricted this second look to the model *Ctee* which had low likelihoods in case study 2 and 3. Figure 3.13 shows plots for the three case studies comparing the likelihoods of the true model *Attt* and *Ctee* with more copula families allowed. It can be noted that the differences are a lot lower, but still significant enough.



(a) Case study 1 (b) Case study 2 (c) Case study 3

Figure 3.13: maximum likelihoods of the 20 simulation runs (N=5000) for models Attt and Ctee with more copula families allowed
One line represents one run. The thick black line represents the means of each model.

3.5 The parent ordering algorithm

Bauer and Czado (2012) developed an algorithm, called *OrderParents*, that tries to determine a set of parent orderings which lead to a high likelihood and low computation time for a given dataset and corresponding DAG.

At first it applies a greedy-type procedure with Kendall's τ edge weight. The next higher parent w_i of v is chosen by optimal Kendall's τ of the pseudo observations $F_{w_i|W_{i-1}}(...)$ and $F_{v|W_{i-1}}(...)$, where W_{i-1} denotes the already chosen lower parents.
The algorithm also permutes some of the orderings to reduce the number of integrals in the corresponding pair-copula decompositions.

Not all of the PCBN decompositions A-F can be assigned to one DAG. We tested the parent ordering algorithm on DAG 1 which has A-D as possible decompositions. The data input were 5000 sampled values of one specific simulation run.
The algorithm chose decomposition A which our simulations where based off in each of the 20 runs for all 3 case studies. This result is especially interesting for case study 1 where the log-likelihoods of all models do not vary a lot.

Chapter 4

Application

In this chapter we will look at three different datasets and calculate the likelihoods and computation times for multiple parent orderings. Because the data is not simulated based on a specific model, it will be interesting to compare the quality of PCBN models with the quality of a RVine.

At first, with empicial data, we have to test conditional independences and determine a partially directed graph corresponding to these. Bauer and Czado (2012) developed another algorithm, called *PCVine*, based on the set of rules presented in Section 2.1. For all distinct vertices $i, j \in V$ and vertex sets $K \subseteq V \backslash \{i, j, \}$, the null hypothesis $H_0 : X_i \perp X_j \mid \mathbf{X}_K$ is tested against the general alternative $H_1 : X_i \not\perp X_j \mid \mathbf{X}_K$. As a suitable test we chose the use of vine copula models and the Rosenblatt transform with a significance level of $\alpha = 0.05$. See Bauer and Czado (2012) for more information. If the immoralities induced by the conditional independences do not allow the creation of a partially directed graph, the algorithm may ignore some of the immoralities.

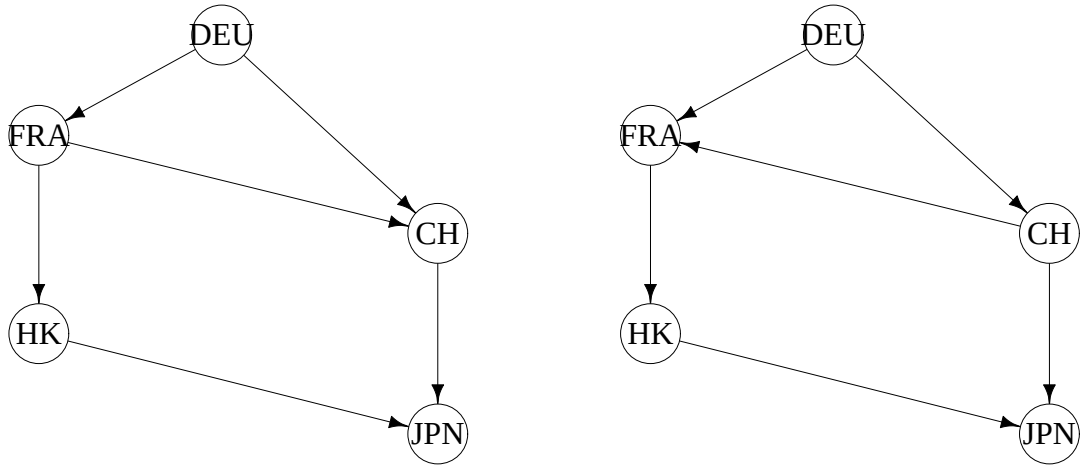
Generally one can find multiple DAGs based on the same partially directed graph, but for the sake of simplicity we restricted the analysis to one DAG for the uranium dataset and for the wine dataset as there are already a lot of possible parent orderings.

As copula families we allowed Gauss, Gumbel, Clayton, Student, Frank and Joe, as well as the rotated versions of Gumbel, Clayton and Joe.

4.1 Stock market indices

This dataset contains the daily log-returns of 10 stock market indices from 7 Januray 2008 to 30 June 2011 ($N = 757$ observations). We restricted our analysis on 5 indices: the German DAX (DEU), the French CAC 40 (FRA), the Swiss Market Index (CH), the Hong Kong Hang Seng Index (HK) and the Japanese Nikkei 225 (JPN).

Interestingly, the partial directed graph determined by the algorithm of Bauer and Czado (2012) is the exact graph we used for our examples in earlier chapters. Figure 4.1 shows two possible DAGs based on the partial directed graph. The structure is the same as for



DAG 1 DAG 2

Figure 4.1: Two DAGs who represent the estimated conditional independence structure of the stock market indices

DAG 1 and DAG 2 in our earlier example, see Figure 2.3.

Figure 4.2 gives an impression of the correlation of the variables, showing pair plots and normalized contour plots based on bivariate kernel estimation utilizing the data values.

As we already know the possible PCBN decompositions A-F for this partial directed graph, we compared the following models:

- PCBN A - PCBN D, the respective PCBN decompositions A-D available for DAG 1
- PCBN E - PCBN F, the respective PCBN decompositions E-F available for DAG 2
- PCBN A_S, a DAG with PCBN decomposition A, only single parameter copulas allowed
- PCBN A_N, a DAG with PCBN decomposition A, only Gauss copulas allowed
- RVine, a RVine selected for the dataset (no independence test)
- RVine_N, a RVine with only Gauss copulas allowed (no independence test)

Table 4.1 shows the log-likelihoods, number of parameters as well the AIC and BIC for the estimated models, both for sequential and joint estimation. Notice that there is not a big difference between the likelihood of models PCBN A - PCBN F. The RVine has a

slightly better likelihood, but also more parameters. The models with restricted copula families had more significant lower likelihoods, with a decrease of $\sim 3\%$.

Tables 4.2 and 4.3 show the chosen copulas and parameters for all models. As experience has shown for financial data, the copulas are mostly Gauss and Student.

To compare the computation time of each model, we measured the time and functions calls of the joint maximum likelihood estimations, see Figure 4.4. It is interesting to note that PCBN C and PCBN D have a much higher time per function call of $17sec$ resp. $45sec$ compared to the $\sim 0.5sec$ of the other PCBN models. In Section 2.3 we analyzed the integration of cdfs via other margins for PCBN decompositions A-F. For A, B, E and F only a single integral was needed while for C and D we needed a double integral. The effect on the computation time can be clearly confirmed here.

We also performed Vuong tests between all models. The Vuong test compares 2 models using the likelihood-ratio. It also allows a correction for the number of parameters using the AIC or BIC. Table 4.5 shows the Vuong test without correction, Table 4.6 shows the Vuong test with AIC correction and Table 4.7 shows the Vuong test with BIC correction. For all tests the significance level α was 0.05.

- The models with unrestricted copula families were always preferred against their restricted counterparts.
- Amongst PCBN A - F no model was preferred.
- Between PCBN A - F and the RVine there was no model was preferred with one single exception: When applying the BIC correction, PCBN A actually was preferred over the RVine.

Additionaly we also ran the parent ordering algorithm of Bauer and Czado (2012), see Section 3.5. When ordering the parents of DAG 1, which has A-D as possible PCBN decompositions, it chose PCBN A, which has a slightly higher likelihood than the other decompositions and is the only one preferred over the RVine in the Vuong test with BIC correction. When comparing the overall computation time of the maximized likelihood PCBN A is also about 9 times faster than the R-Vine.

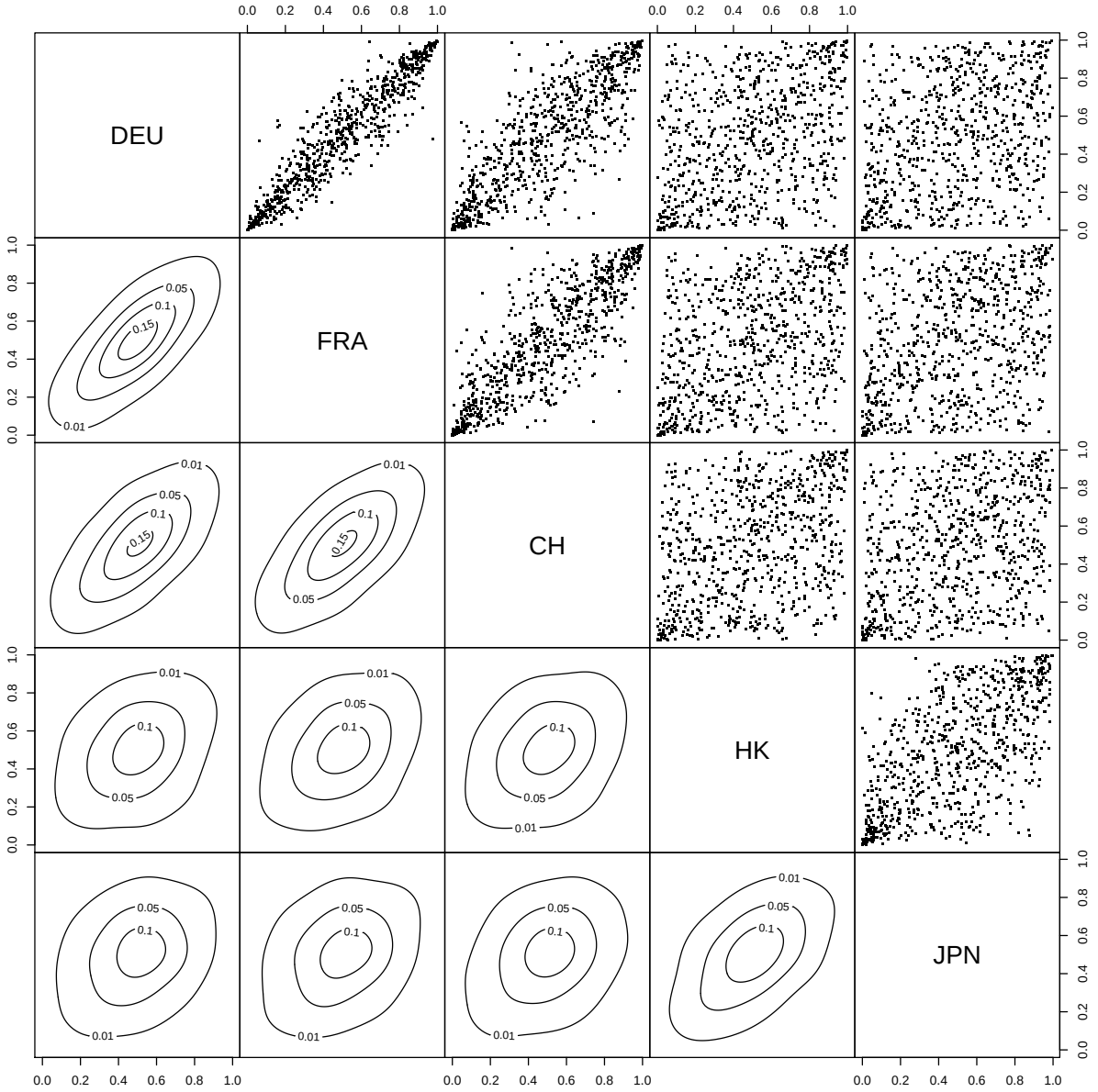


Figure 4.2: stock indices:
upper right: pairs plots of the data values (N=757)
bottom left: normalized contour plots based on bivariate kernel estimation utilizing the data values

model	method	LL	#par	AIC	BIC
PCBN A	seqEst	1746.65	10	-3473.31	**-3427.02**
	jointEst	1747.69	10	-3473.37	**-3427.08**
PCBN B	seqEst	1746.00	10	-3472.00	-3425.70
	jointEst	1746.27	10	-3472.54	-3426.25
PCBN C	seqEst	1745.40	11	-3468.80	-3417.88
	jointEst	1745.40	11	-3468.80	-3417.88
PCBN D	seqEst	1744.54	11	-3467.08	-3416.15
	jointEst	1744.80	11	-3467.61	-3416.69
PCBN E	seqEst	1744.38	11	-3466.76	-3415.84
	jointEst	1745.27	11	-3468.53	-3417.61
PCBN F	seqEst	1743.72	11	-3465.45	-3414.52
	jointEst	1744.79	11	-3467.57	-3416.65
PCBN A_S	seqEst	1692.48	6	-3372.96	-3345.19
	jointEst	1692.50	6	-3373.00	-3345.22
PCBN A_N	seqEst	1691.88	6	-3371.76	-3343.98
	jointEst	1691.88	6	-3371.77	-3343.99
R-Vine	seqEst	**1754.35**	15	**-3478.71**	-3409.26
	jointEst	**1754.48**	15	**-3478.96**	-3409.52
R-Vine$_N$	seqEst	1694.15	10	-3368.31	-3322.01
	jointEst	1694.15	10	-3368.31	-3322.01

Table 4.1: stock indices (N=757):
LogLikelihood, number of total copula parameters, AIC and BIC of the referred models (both for sequential and joint parameter estimation)

model	copula	type	par	par2	tau
A	$c_{DEU,FRA}$	t	0.95	5.11	0.79
	$c_{DEU,CH\|FRA}$	N	0.16		0.11
	$c_{FRA,CH}$	t	0.86	6.19	0.66
	$c_{FRA,HK}$	t	0.45	7.66	0.30
	$c_{CH,JPN\|HK}$	N	0.15		0.10
	$c_{HK,JPN}$	t	0.66	10.90	0.46
B	$c_{DEU,FRA}$	t	0.95	5.11	0.79
	$c_{DEU,CH\|FRA}$	N	0.16		0.11
	$c_{FRA,CH}$	t	0.86	6.04	0.66
	$c_{FRA,HK}$	t	0.44	7.01	0.29
	$c_{CH,JPN}$	G 180°	1.26		0.21
	$c_{HK,JPN\|CH}$	t	0.61	12.63	0.42
C	$c_{DEU,FRA}$	t	0.95	4.97	0.79
	$c_{DEU,CH}$	t	0.84	6.60	0.64
	$c_{FRA,CH\|DEU}$	t	0.38	18.10	0.25
	$c_{FRA,HK}$	t	0.45	7.39	0.30
	$c_{CH,JPN\|HK}$	N	0.15		0.10
	$c_{HK,JPN}$	t	0.66	10.38	0.46
D	$c_{DEU,FRA}$	t	0.95	5.03	0.79
	$c_{DEU,CH}$	t	0.84	6.46	0.64
	$c_{FRA,CH\|DEU}$	t	0.37	18.05	0.24
	$c_{FRA,HK}$	t	0.44	7.01	0.29
	$c_{CH,JPN}$	G 180°	1.26		0.21
	$c_{HK,JPN\|CH}$	t	0.61	12.65	0.41
E	$c_{DEU,FRA\|CH}$	t	0.80	4.66	0.60
	$c_{DEU,CH}$	t	0.85	8.88	0.64
	$c_{FRA,CH}$	t	0.86	7.55	0.66
	$c_{FRA,HK}$	t	0.45	7.65	0.30
	$c_{CH,JPN\|HK}$	N	0.15		0.10
	$c_{HK,JPN}$	t	0.66	10.91	0.46
F	$c_{DEU,FRA\|CH}$	t	0.80	4.65	0.60
	$c_{DEU,CH}$	t	0.85	8.78	0.64
	$c_{FRA,CH}$	t	0.86	7.41	0.66
	$c_{FRA,HK}$	t	0.44	7.01	0.29
	$c_{CH,JPN}$	G 180°	1.26		0.21
	$c_{HK,JPN\|CH}$	t	0.61	12.64	0.41

Table 4.2: stock indices:
selected copulas for models A-F,
parameters determined by joint estimation

model	copula	type	par	par2	tau
A_S	$c_{DEU,FRA}$	N	0.94		0.79
	$c_{DEU,CH\|FRA}$	F	1.11		0.12
	$c_{FRA,CH}$	N	0.86		0.66
	$c_{FRA,HK}$	N	0.47		0.31
	$c_{CH,JPN\|HK}$	N	0.16		0.10
	$c_{HK,JPN}$	N	0.66		0.46
A_N	$c_{DEU,FRA}$	N	0.95		0.79
	$c_{DEU,CH\|FRA}$	N	0.16		0.10
	$c_{FRA,CH}$	N	0.86		0.66
	$c_{FRA,HK}$	N	0.47		0.31
	$c_{CH,JPN\|HK}$	N	0.16		0.10
	$c_{HK,JPN}$	N	0.66		0.46
R-Vine	$c_{DEU,FRA}$	t	0.95	5.23	0.79
	$c_{DEU,CH\|FRA}$	N	0.16		0.10
	$c_{DEU,HK\|FRA,CH}$	t	-0.03	10.81	-0.02
	$c_{DEU,JPN\|FRA,CH,HK}$	F	-0.17		-0.02
	$c_{FRA,CH}$	t	0.86	6.17	0.66
	$c_{FRA,HK}$	t	0.44	7.71	0.30
	$c_{FRA,JPN\|HK}$	G	1.09		0.08
	$c_{CH,HK\|FRA}$	N	0.07		0.04
	$c_{CH,JPN\|FRA,HK}$	C180°	0.14		0.07
	$c_{HK,JPN}$	t	0.66	10.70	0.46
R-Vine$_N$	$c_{DEU,FRA}$	N	0.94		0.79
	$c_{DEU,CH\|FRA}$	N	0.16		0.10
	$c_{DEU,HK\|FRA,CH}$	N	-0.03		-0.02
	$c_{DEU,JPN\|FRA,CH,HK}$	N	0.00		0.00
	$c_{FRA,CH}$	N	0.86		0.66
	$c_{FRA,HK}$	N	0.47		0.31
	$c_{FRA,JPN\|HK}$	N	0.13		0.08
	$c_{CH,HK\|FRA}$	N	0.07		0.04
	$c_{CH,JPN\|FRA,HK}$	N	0.09		0.06
	$c_{HK,JPN}$	N	0.66		0.46

Table 4.3: stock indices:
selected copulas for models A_S, A_N, R-Vine and R-Vine$_N$,
parameters determined by joint estimation

model	overall time	# function calls	time per funtion call
PCBN A	28.40s	46	0.62s
PCBN B	19.24s	46	0.42s
PCBN C	119.13s	7	17.02s
PCBN D	2583.04s	57	45.32s
PCBN E	34.59s	50	0.70s
PCBN F	19.53s	48	0.40s
PCBN A_S	1.38s	34	0.04s
PCBN A_N	1.77s	36	0.05s
R-Vine	263.63s	79	3.34s
R-Vine$_N$	1.50s	3	0.50s

Table 4.4: stock indices:
Computation time of joint parameter estimation

	A	B	C	D	E	F	A_S	A_N	R	R_N
A										
B	0									
C	0	0								
D	0	0	0							
E	0	0	0	0						
F	0	0	0	0	0					
A_S	2	2	2	2	2	2				
A_N	2	2	2	2	2	2	0			
R	0	0	0	0	0	0	1	1		
R_N	2	2	2	2	2	2	0	0	2	

Table 4.5: stock indices:
Vuong tests ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

	A	B	C	D	E	F	A_S	A_N	R	R_N
A										
B	0									
C	0	0								
D	0	0	0							
E	0	0	0	0						
F	0	0	0	0	0					
A_S	2	2	2	2	2	2				
A_N	2	2	2	2	2	2	0			
R	0	0	0	0	0	0	1	1		
R_N	2	2	2	2	2	2	0	0	2	

Table 4.6: stock indices:
Vuong tests with AIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

	A	B	C	D	E	F	A_S	A_N	R	R_N
A										
B	0									
C	0	0								
D	0	0	0							
E	0	0	0	0						
F	0	0	0	0	0					
A_S	2	2	2	2	2	2				
A_N	2	2	2	2	2	2	0			
R	2	0	0	0	0	0	0	0		
R_N	2	2	2	2	2	2	2	2	2	

Table 4.7: stock indices:
Vuong tests with BIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

4.2 Uranium dataset

The uranium dataset consists of log concentrations of 7 chemical elements in water samples (N=655) collected near Grand Junction, Colorado. Concentrations were measured for the following elements: Uranium(U), Lithium (Li), Cobalt (Co), Potassium (K), Cesium (Cs), Scandium (Sc), and Titanium (Ti).

This dataset is interesting because - contrary to most financial data - the best fitting copula families are very diverse. This can be verified with the normalized contour plots based on bivariate kernel estimations in Figure 4.4 and the copula families in Tables 4.11 and 4.12.

We restricted our study on one specific DAG that is based on the partially directed graph determined by the PCVine algorithm of Bauer and Czado (2012). This DAG is illustrated in Figure 4.3.

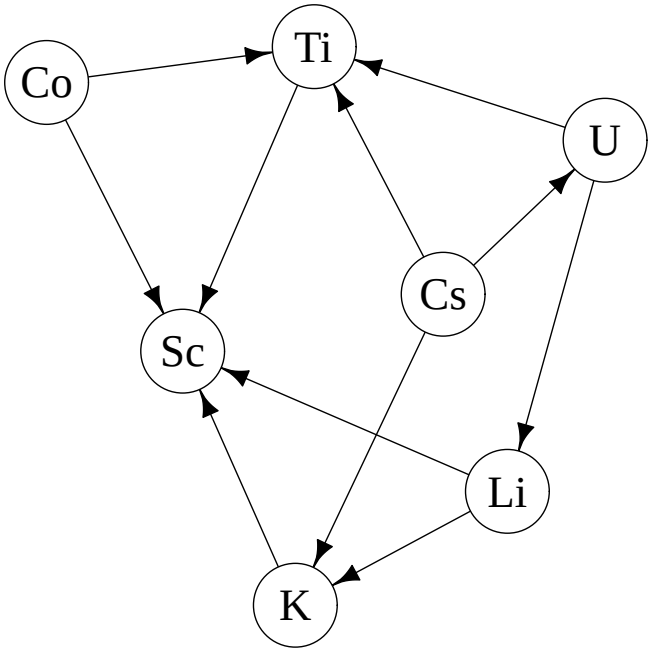


Figure 4.3: uranium data:
DAG based on the partially directed graph determined by the PCVine-algorithm applied on the uranium dataset

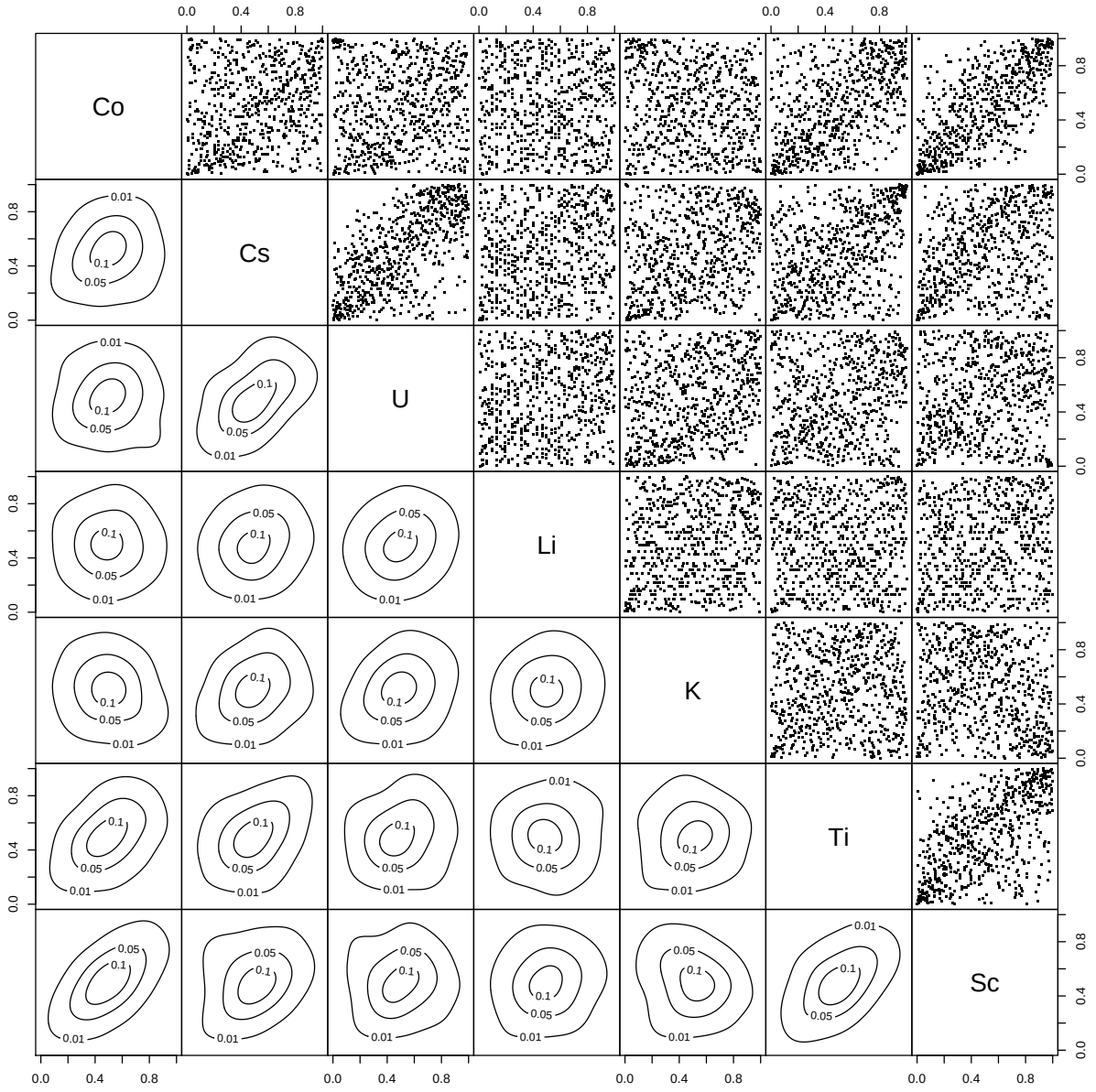


Figure 4.4: uranium data:
upper right: pairs plots of the data values (N=655)
bottom left: normalized contour plots based on bivariate kernel estimation utilizing the data values

The DAG contains 3 vertices that have multiple parents and thus these parents can be ordered differently: K has 2 parents, Ti has 3 parents and Sc has 4 parents. Alltogether that allows $288 = 4!\ 3!\ 2!$ different parent orderings.

At first we tested all 288 orderings. All copulas were chosen by AIC and the sequential log-likelihood was calculated. We also measured the time for the calculation of the sequential log-likelihood as an indicator for the computational complexity. A table of all log-likelihoods and computation times can be found in Appendix A. Unfortunately the computation caused several arithmetic exceptions, so the values of the log-likelihoods cannot be compared. Of the 288 PCBN models we chose 5 for a further analysis and compared them with a RVine.

- PCBN 1: The model with the longest computation time.
- PCBN 2: The model with the shortest computation time.
- PCBN 3: The model with the highest log-likelihood (possibly falsified).
- PCBN 4: The model with the lowest log-likelihood (possibly falsified).
- PCBN 5: The model determined by the OrderParents-algorithm of Bauer and Czado (2012).

where PCBN stands for *pair-copula Bayesian network*. We also compared these models with PCBN 5_N and R-Vine$_N$ where PCBN 5 and respectively the R-Vine use only Gauss pair-copulas.

model	vertex	parent ordering
PCBN 1	K	Li $<_K$ Cs
	Ti	U $<_{Ti}$ Co $<_{Ti}$ Cs
	Sc	Ti $<_{Sc}$ K $<_{Sc}$ Li $<_{Sc}$ Co
PCBN 2	K	Li $<_K$ Cs
	Ti	Co $<_{Ti}$ U $<_{Ti}$ Cs
	Sc	Li $<_{Sc}$ Co $<_{Sc}$ K $<_{Sc}$ Ti
PCBN 3	K	Li $<_K$ Cs
	Ti	Co $<_{Ti}$ Cs $<_{Ti}$ U
	Sc	Ti $<_{Sc}$ K $<_{Sc}$ Co $<_{Sc}$ Li
PCBN 4	K	Li $<_K$ Cs
	Ti	U $<_{Ti}$ Co $<_{Ti}$ Cs
	Sc	K $<_{Sc}$ Ti $<_{Sc}$ Co $<_{Sc}$ Li
PCBN 5	K	Cs $<_K$ Li
	Ti	Co $<_{Ti}$ Cs $<_{Ti}$ U
	Sc	Co $<_{Sc}$ Ti $<_{Sc}$ Li $<_{Sc}$ K

Table 4.8: uranium data:
parent orderings of the models PCBN 1-5

Table 4.9 shows results for the original dataset. Note that the log-likelihood as well as the AIC and the BIC cannot be compared because arithmetic exceptions were caused.

model	LL	#par	AIC	BIC	time	highest integral	#exc.
PCBN 1	749.27	13	-1472.55	-1414.25	1251.20s	4	1
PCBN 2	777.42	12	-1530.84	-1477.03	2.41s	2	0
PCBN 3	791.68	13	-1557.36	-1499.06	359.13s	4	17
PCBN 4	493.09	13	-960.19	-901.89	1061.60s	4	134
PCBN 5	783.55	12	-1543.11	-1489.29	3.28s	2	4
PCBN 5_N	702.47	11	-1382.95	-1333.62	39.25s	2	1
RVine	876.43	26	-1692.86	-1558.32	19.40s	-	0
RVine$_N$	751.33	21	-1460.66	-1366.48	1.60s	-	0

Table 4.9: uranium data: (N=655)
LL = sequential log-likelihood, #par = number of parameters,
time = computation time of the sequential log-likelihood,
highest integral = highest dimension of the integrals in the computation of cdfs
#exc. = number of data values that caused arithmetic exceptions

We then removed values from the dataset repeatedly until no arithmetic exceptions were caused. This reduced the dataset to a samplesize of $N = 496$. Pairs plots and normalized contour plots of the updated dataset are illustrated in Figure 4.5.

Following we performed a new estimation of copula families and parameters and a new calculation of the log-likelihood with the updated dataset. Note that this means that PCBN 1-2 do not necessarily have the longest/shortest computation time for this dataset, and PCBN 3-4 do not necessarily have the highest/lowest log-likelihood.

Table 4.10 shows detailed results for the chosen models.

- There is a significant difference of the log-likelihoods of the different PCBN models: PCBN 5 has a 5% higher log-likelihood than PCBN 4. The RVine is superior to all PCBN models in terms of log-likelihood.
- When we apply a strong penalty for the numbers of parameters (BIC), the RVine is still the best model, but PCBN 5 comes close.
- PCBN 2 and 5 have a computation time that is about 100 times faster than PCBN 1, 3 and 4. It is also about 4-5 times faster than the RVine.
- PCBN 2, 3 and 4 need several minutes for a single evaluation of the sequential log-likelihood. This is caused by 4-dimensional integrals in the computation of the cdfs.
- PCBN 5 which is determined by the OrderParents-algorithm, applied on the orginal dataset with N=655, has a good log-likelihood and computation time compared to the other PCBN models.

model	LL	#par	AIC	BIC	time	highest integral
PCBN 1	553.76	12	-1083.55	-1033.04	235.23s	4
PCBN 2	550.13	13	-1074.27	-1019.58	2.11s	2
PCBN 3	568.67	12	-1113.35	-1062.87	203.19s	4
PCBN 4	542.97	13	-1059.93	-1005.25	234.69s	4
PCBN 5	571.20	12	-1118.39	-1067.92	2.38s	2
PCBN 5_N	509.60	11	-997.19	-950.92	24.15s	2
RVine	619.26	26	-1186.53	-1077.16	9.71s	-
RVine$_N$	533.83	21	-1025.66	-937.32	1.49s	-

Table 4.10: uranium data: (N=496)
LL = sequential log-likelihood, #par = number of parameters,
time = computation time of the sequential log-likelihood,
highest integral = highest dimension of the integrals in the computation of cdfs

- The models PCBN 5_N and R-Vine$_N$ which used only Gauss pair-copulas have a significantly lower log-likelihood than all other models.

The detailed formulas for the decompositions and conditional cdfs of models PCBN 1-5 can be found in Appendix B.

Tables 4.11 and 4.12 show the selected copula families with estimated parameters and corresponding Kendalls τ. The estimated copula families are very diverse, PCBN 1 contains all 6 allowed families. Tree plots of the RVine can be found in Appendix C.

We also performed Vuong tests between all models. Table 4.13 shows the results for the Vuong tests without correction, Table 4.14 for the Vuong tests with AIC correction and Table 4.15 for the Vuong tests with BIC correction. When applying no correction or only AIC correction, the R-Vine is preferred over all other models. With BIC correction, the RVine is not preferred over PCBN 1, 3 and 5.

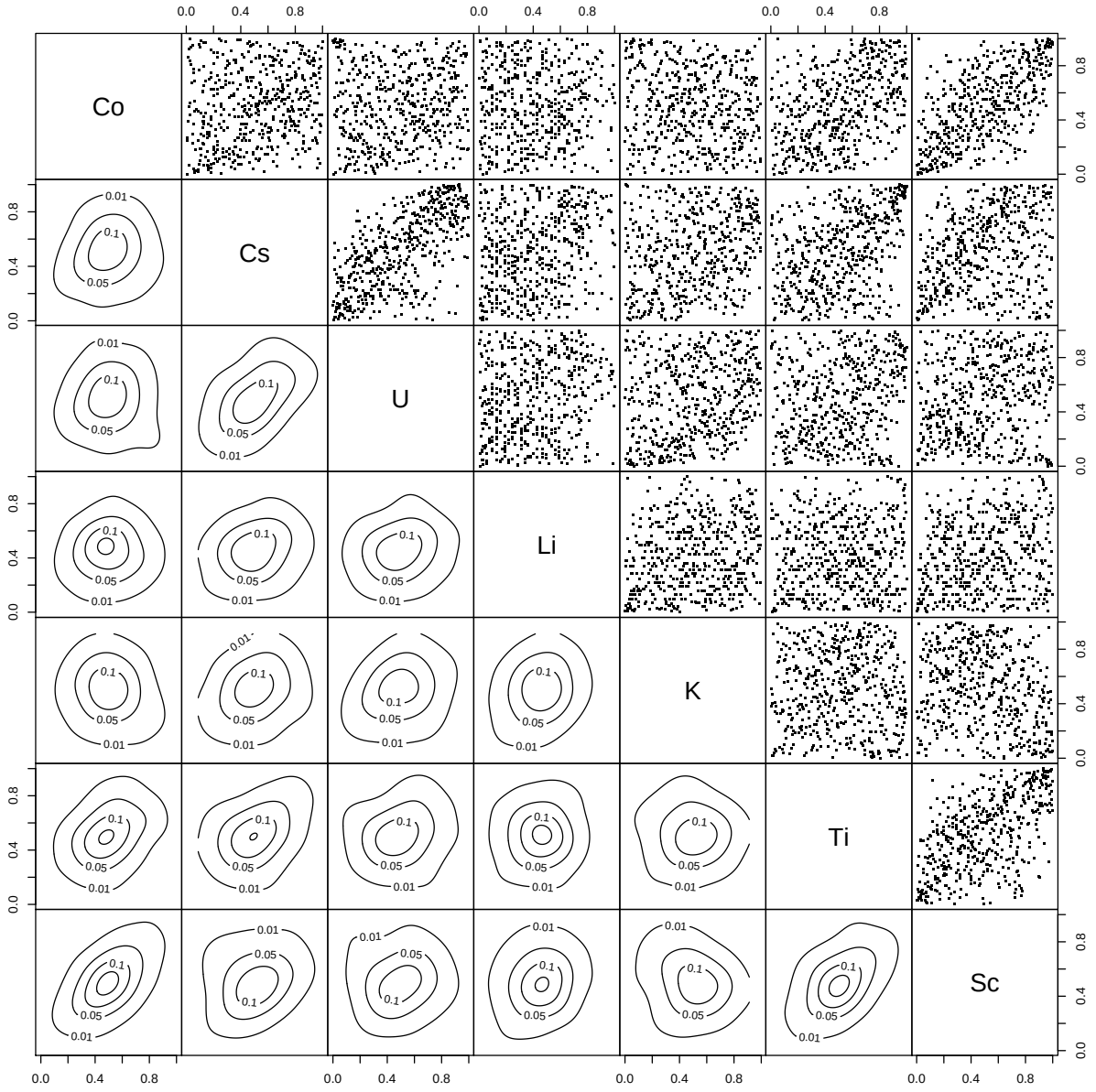


Figure 4.5: uranium data (reduced dataset):
upper right: pairs plots of the data values (N=496)
bottom left: normalized contour plots based on bivariate kernel estimation utilizing the data values

model	copula	type	par	par2	tau
PCBN 1	$c_{16\|3}$	F	3.11		0.32
	$c_{17\|456}$	t	0.60	12.22	0.41
	c_{23}	F	5.50		0.49
	$c_{25\|4}$	F	1.68		0.18
	$c_{26\|13}$	C180°	0.71		0.26
	c_{34}	F	1.30		0.14
	c_{36}	C180°	0.33		0.14
	c_{45}	J180°	1.24		0.12
	$c_{47\|56}$	F	2.10		0.22
	$c_{57\|6}$	G90°	-1.21		-0.17
	c_{67}	N	0.65		0.45
PCBN 2	c_{16}	F	3.22		0.33
	$c_{17\|4}$	t	0.71	10.86	0.5
	c_{23}	F	5.50		0.49
	$c_{25\|4}$	F	1.68		0.18
	$c_{26\|13}$	J	1.56		0.24
	c_{34}	F	1.30		0.14
	$c_{36\|1}$	C180°	0.31		0.13
	c_{45}	J180°	1.24		0.12
	c_{47}	G180°	1.18		0.15
	$c_{57\|14}$	C270°	-0.36		-0.15
	$c_{67\|145}$	t	0.49	5.45	0.33
PCBN 3	c_{16}	F	3.22		0.33
	$c_{17\|56}$	t	0.59	9.62	0.4
	c_{23}	F	5.50		0.49
	$c_{25\|4}$	F	1.68		0.18
	$c_{26\|1}$	C180°	0.82		0.29
	c_{34}	F	1.30		0.14
	$c_{36\|12}$	C90°	-0.15		-0.07
	c_{45}	J180°	1.24		0.12
	$c_{47\|156}$	F	2.29		0.24
	$c_{57\|6}$	G90°	-1.21		-0.17
	c_{67}	N	0.65		0.45

Table 4.11: uranium data:
Estimated copula families, parameters and Kendalls τ.
(1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

model	copula	type	par	par2	tau
PCBN 4	$c_{16\|3}$	F	3.11		0.32
	$c_{17\|56}$	t	0.57	6.98	0.39
	c_{23}	F	5.50		0.49
	$c_{25\|4}$	F	1.68		0.18
	$c_{26\|13}$	C180°	0.71		0.26
	c_{34}	F	1.30		0.14
	c_{36}	C180°	0.33		0.14
	c_{45}	J180°	1.24		0.12
	$c_{47\|156}$	F	1.80		0.19
	c_{57}	C270°	-0.38		-0.16
	$c_{67\|5}$	t	0.64	7.17	0.44
PCBN 5	c_{16}	F	3.22		0.33
	c_{17}	t	0.71	8	0.5
	c_{23}	F	5.50		0.49
	c_{25}	F	1.89		0.2
	$c_{26\|1}$	C180°	0.82		0.29
	c_{34}	F	1.30		0.14
	$c_{36\|12}$	C90°	-0.15		-0.07
	$c_{45\|2}$	J180°	1.17		0.09
	$c_{47\|16}$	G180°	1.18		0.16
	$c_{57\|146}$	C270°	-0.37		-0.16
	$c_{67\|1}$	G	1.49		0.33
PCBN 5_N	c_{16}	N	0.47		0.31
	c_{17}	N	0.70		0.49
	c_{23}	N	0.66		0.46
	c_{25}	N	0.27		0.17
	$c_{26\|1}$	N	0.43		0.28
	c_{34}	N	0.19		0.12
	$c_{36\|12}$	N	-0.10		-0.06
	$c_{45\|2}$	N	0.21		0.13
	$c_{47\|16}$	N	0.26		0.17
	$c_{57\|146}$	N	-0.29		-0.18
	$c_{67\|1}$	N	0.51		0.34

Table 4.12: uranium data:
Estimated copula families, parameters and Kendalls τ.
(1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	1	0						
4	0	0	2					
5	0	1	0	1				
5_N	2	2	2	2	2			
R	1	1	1	1	1	1		
R_N	0	0	2	0	2	0	2	

Table 4.13: uranium data:
Vuong tests ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

vuong	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	1	0						
4	0	0	2					
5	0	1	0	1				
5_N	2	2	2	2	2			
R	1	1	1	1	1	1		
R_N	0	0	2	0	2	0	2	

Table 4.14: uranium data:
Vuong tests with AIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

vuong	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	1	1						
4	0	0	2					
5	0	1	0	1				
5_N	2	2	2	0	2			
R	0	1	0	1	0	1		
R_N	2	2	2	0	2	0	2	

Table 4.15: uranium data:
Vuong tests with BIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

4.3 Wine dataset

Cortez et al. (2009) investigated eleven chemical compenents of red wine samples ($N = 500$). We selected the following 8 components for our analysis: acf = fixed acidity, acv = volatile acidity, acc = citric acid, sf = free sulfur dioxide, st = total sulfur dioxide, den = density, ph = pH and sp = sulphates.

Similar to the uranium dataset, the best fitting copula families are very diverse. Additionaly many values are also negatively correlated. This can be verified with the normalized contour plots based on bivariate kernel estimations in Figure 4.7 and the copula families in Tables 4.18 and 4.19.

We restricted our study on one specific DAG that is based on the partially directed graph determined by the PCVine-algorithm of Bauer and Czado (2012). This DAG is illustrated in Figure 4.6.

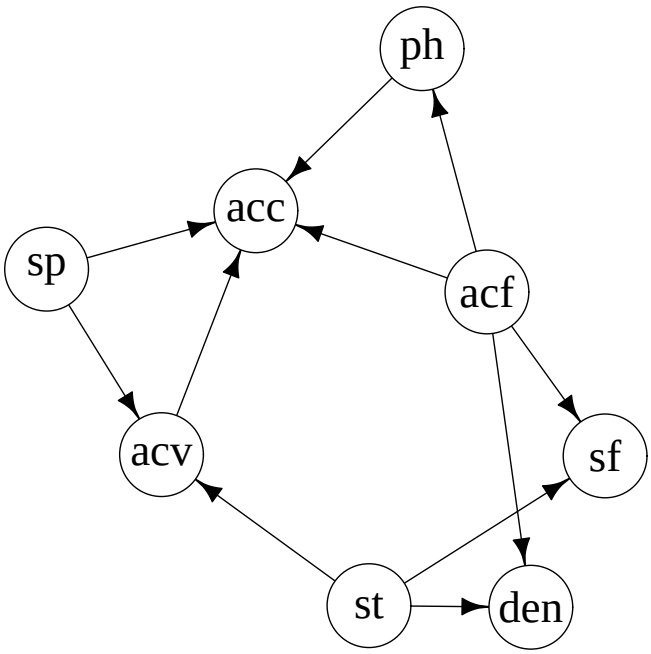


Figure 4.6: wine data:
DAG based on the partially directed graph determined by the PCVine-algorithm applied on the wine dataset

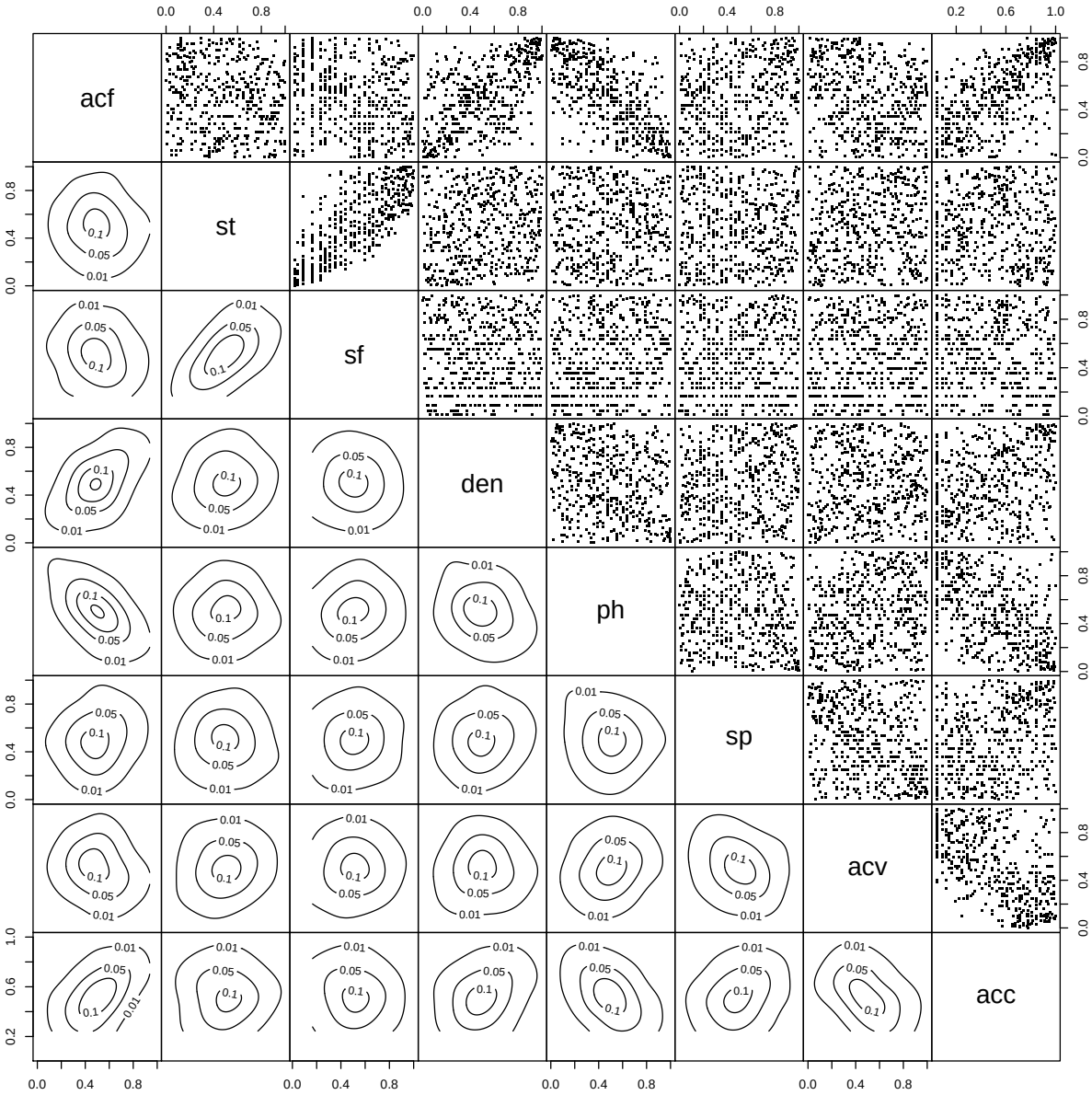


Figure 4.7: wine data:
upper right: pairs plots of the data values (N=500)
bottom left: normalized contour plots based on bivariate kernel estimation utilizing the data values

This DAG contains 4 vertices that have multiple parents, so these can be ordered for different parent orderings: sf, den and acv each have 2 parents, acc has 4 parents. This allows $192 = 4!\ 2!\ 2!\ 2!$ different parent orderings.

At first we tested all 192 orderings. All copulas were chosen by AIC and the sequential log-likelihood was calculated. For this dataset, no exceptions were caused during the computation. We also measured the time for the calculation of the sequential log-likelihood as an indicator for the computational complexity. A table of all log-likelihoods and computation times can be found in Appendix D. Of these 192 models we chose 5 for a further analysis and compared them with a RVine.

model	vertex	parent ordering
PCBN 1	sf	acf $<_{sf}$ st
	den	acf $<_{den}$ st
	acv	st $<_{acv}$ sp
	acc	acv $<_{acc}$ sp $<_{acc}$ acf $<_{acc}$ ph
PCBN 2	sf	acf $<_{sf}$ st
	den	acf $<_{den}$ st
	acv	sp $<_{acv}$ st
	acc	ph $<_{acc}$ acf $<_{acc}$ acv $<_{acc}$ sp
PCBN 3	sf	st $<_{sf}$ acf
	den	acf $<_{den}$ st
	acv	sp $<_{acv}$ st
	acc	acf $<_{acc}$ acv $<_{acc}$ ph $<_{acc}$ sp
PCBN 4	sf	acf $<_{sf}$ st
	den	st $<_{den}$ acf
	acv	st $<_{acv}$ sp
	acc	sp $<_{acc}$ ph $<_{acc}$ acv $<_{acc}$ acf
PCBN 5	sf	acf $<_{sf}$ st
	den	acf $<_{den}$ st
	acv	sp $<_{acv}$ st
	acc	acf $<_{acc}$ acv $<_{acc}$ ph $<_{acc}$ sp

Table 4.16: wine data:
parent orderings of the models PCBN 1-5

- PCBN 1: The model with the longest computation time.
- PCBN 2: The model with the shortest computation time.
- PCBN 3: The model with the highest log-likelihood.
- PCBN 4: The model with the lowest log-likelihood.
- PCBN 5: The model determined by the OrderParents-algorithm of Bauer and Czado (2012).

where PCBN stands for *pair-copula Bayesian network*. We also compared these models with PCBN 5_N and R-Vine$_N$ where PCBN 5 and respectively the R-Vine use only Gauss pair-copulas.

Table 4.17 shows detailed results for the chosen models.

model	LL	#par	AIC	BIC	time	highest integral
PCBN 1	814.08	13	-1602.16	-1547.37	0.1520s	1
PCBN 2	816.56	12	-1609.13	-1558.55	0.0050s	0
PCBN 3	822.31	12	-1620.62	-1570.04	0.0070s	0
PCBN 4	799.60	13	-1573.20	-1518.41	0.0810s	1
PCBN 5	818.09	12	-1612.18	-1561.60	0.0080s	0
PCBN 5_N	749.21	11	-1476.41	-1430.05	0.0020s	0
RVine	990.72	30	-1921.44	-1795.00	10.1000s	-
RVine$_N$	875.32	28	-1694.65	-1576.64	2.8255s	-

Table 4.17: wine data (N=500):
LL = sequential log-likelihood, #par = number of parameters,
time = computation time of the sequential log-likelihood,
highest integral = highest dimension of the integrals in the computation of cdfs

- The PCBN models could not be fitted as well as the RVine. Even with a strong penalty on the higher numbers of parameters of the RVine (BIC), the RVine is significantly better.
- There is also a significant difference in the quality of different PCBN models, especially when comparing the BIC of PCBN 4 and PCBN 5.
- Some models did not need any further computation of a cdf by integration via other margins. Other models needed a 1-dimensional integration. The reason is that those vertices who are parents of the same vertex are mostly not descendants of one common vertex and thus assumed to be independent.
- Because the PCBN models did not need to compute higher-dimensional integrals, they were significantly faster than the RVine. The fastest PCBN was over 1000 times faster than the R-Vine and about 20 times faster than the slowest PCBN.
- PCBN 5 which is determined by the OrderParents-algorithm has a good log-likelihood and computation time compared to the other PCBN models. PCBN 3 is however slightly superior in log-likelihood and computation time, and has also the same number of parameters.

- The models PCBN 5_N and R-Vine$_N$ which use only Gauss pair-copulas have a significantly lower log-likelihood than the corresponding models PCBN 5 and respectively the R-Vine which use more copula families. R-Vine$_N$ perfoms however well in comparison to the PCBN models, but still has a much higher computation time.

The detailed formulas for the decompositions and conditional cdfs of models PCBN 1-5 can be found in Appendix E.

Tables 4.11 and 4.12 show the selected copula families with estimated parameters and corresponding Kendalls τ. The selected copula families are very diverse; PCBN 1 contains all 6 allowed families and different rotations. Also note that the models with lower log-likelihoods have a higher Kendall's τ for the copulas who are conditioned on multiple other vertices: PCBN 3 has $c_{68|157}$ with $\tau = 0.05$, PCBN 4 has $c_{18|567}$ with $\tau = 0.26$. Tree plots of the RVine can be found in Appendix F.

We also performed Vuong tests between all models. Table 4.20 shows the results for the Vuong tests without correction, Table 4.21 for the Vuong tests with AIC correction and Table 4.22 for the Vuong tests with BIC correction. The R-Vine is preferred over the PCBN models even when applying a strong penalty for the number of parameters with the BIC correction. Without correction, R-Vine$_N$ is also preferred over the PCBN models, but not with BIC correction. Between PCBN models only PCBN 2, 3 and 5 are preferred over PCBN 4, else no model is preferred.

model	copula	type	par	par2	tau
PCBN 1	c_{13}	F	-1.46		-0.16
	c_{14}	N	0.66		0.46
	c_{15}	G90°	-1.97		-0.49
	$c_{18\|67}$	t	0.60	7.1	0.41
	$c_{23\|1}$	G180°	2.21		0.55
	$c_{24\|1}$	C	0.35		0.15
	c_{27}	t	0.15	8.81	0.1
	$c_{58\|167}$	C270°	-0.10		-0.05
	$c_{67\|2}$	F	-2.07		-0.22
	$c_{68\|7}$	J	1.16		0.09
	c_{78}	F	-4.27		-0.41
PCBN 2	c_{13}	F	-1.46		-0.16
	c_{14}	N	0.66		0.46
	c_{15}	G90°	-1.97		-0.49
	$c_{18\|5}$	G	1.40		0.28
	$c_{23\|1}$	G180°	2.21		0.55
	$c_{24\|1}$	C	0.35		0.15
	$c_{27\|6}$	t	0.16	8.16	0.1
	c_{58}	F	-3.77		-0.37
	c_{67}	F	-2.05		-0.22
	$c_{68\|157}$	G	1.07		0.06
	$c_{78\|15}$	F	-3.92		-0.38
PCBN 3	$c_{13\|2}$	F	-1.37		-0.15
	c_{14}	N	0.66		0.46
	c_{15}	G90°	-1.97		-0.49
	c_{18}	G	1.75		0.43
	c_{23}	G180°	2.23		0.55
	$c_{24\|1}$	C	0.35		0.15
	$c_{27\|6}$	t	0.16	8.16	0.1
	$c_{58\|17}$	C270°	-0.17		-0.08
	c_{67}	F	-2.05		-0.22
	$c_{68\|157}$	G	1.06		0.05
	$c_{78\|1}$	F	-4.07		-0.39

Table 4.18: wine data:
Estimated copula families, parameters and Kendalls τ.
(1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

model	copula	type	par	par2	tau
PCBN 4	c_{13}	F	-1.46		-0.16
	$c_{14\|2}$	N	0.69		0.48
	c_{15}	G90°	-1.97		-0.49
	$c_{18\|567}$	t	0.39	6.42	0.26
	$c_{23\|1}$	G180°	2.21		0.55
	c_{24}	C	0.20		0.09
	c_{27}	t	0.15	8.81	0.1
	$c_{58\|6}$	F	-3.86		-0.38
	$c_{67\|2}$	F	-2.07		-0.22
	c_{68}	F	2.05		0.22
	$c_{78\|56}$	F	-3.09		-0.31
PCBN 5	c_{13}	F	-1.46		-0.16
	c_{14}	N	0.66		0.46
	c_{15}	G90°	-1.97		-0.49
	c_{18}	G	1.75		0.43
	$c_{23\|1}$	G180°	2.21		0.55
	$c_{24\|1}$	C	0.35		0.15
	$c_{27\|6}$	t	0.16	8.16	0.1
	$c_{58\|17}$	C270°	-0.17		-0.08
	c_{67}	F	-2.05		-0.22
	$c_{68\|157}$	G	1.06		0.05
	$c_{78\|1}$	F	-4.07		-0.39
PCBN 5_N	c_{13}	N	-0.21		-0.13
	c_{14}	N	0.66		0.46
	c_{15}	N	-0.71		-0.51
	c_{18}	N	0.61		0.42
	$c_{23\|1}$	N	0.76		0.55
	$c_{24\|1}$	N	0.26		0.17
	$c_{27\|6}$	N	0.16		0.1
	$c_{58\|17}$	N	-0.15		-0.1
	c_{67}	N	-0.29		-0.19
	$c_{68\|157}$	N	0.14		0.09
	$c_{78\|1}$	N	-0.50		-0.33

Table 4.19: wine data:
Estimated copula families, parameters and Kendalls τ.
(1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

vuong	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	0	0						
4	0	2	0					
5	0	0	0	0				
5_N	2	2	2	2	2			
R	1	1	1	1	1	1		
R_N	1	1	1	1	1	1	2	

Table 4.20: wine data:
Vuong tests ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

vuong	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	0	0						
4	0	2	2					
5	0	0	0	0				
5_N	2	2	2	2	2			
R	1	1	1	1	1	1		
R_N	1	0	0	1	0	1	2	

Table 4.21: wine data:
Vuong tests with AIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

vuong	1	2	3	4	5	5_N	R	R_N
1								
2	0							
3	0	0						
4	0	2	2					
5	0	0	0	1				
5_N	2	2	2	2	2			
R	1	1	1	1	1	1		
R_N	0	0	0	0	0	1	2	

Table 4.22: wine data:
Vuong tests with BIC correction ($\alpha = 0.05$)
0 = no model preferred
1 = left model preferred
2 = upper model preferred

Chapter 5

Outlook: Optimization of computation time

To optimize the computation time of PCBNs, we have to analyze which DAG structures and parent orderings require higher complex integrations to determine the conditional cdfs in the decomposition.

Conditional cdfs appear as arguments of conditional pair-copulas in the decomposition. Equation (1.5) shows that the decomposition only contains conditional pair-copulas if a vertex has multiple parents. We already examined the most simple case, a vertex having 2 parents, in Section 1.3. It showed that if a vertex k has 2 parents i and j which are not independent, we would need to know the conditional cdf $F_{i|j}$ or respectively $F_{j|i}$. This conditional cdf could be determined by integrating the copula c_{ij}. If i and j are not connected, and thus (i, k, j) is an immorality, then c_{ij} is certainly not available in the decomposition and we have to integrate via other margins. One important open question is if every immorality (i, k, j) where i and j are not independent leads to a necessary integration via other margins. Therefor the case that k has more than 2 parents has to be examined.

When we have to determine a conditional cdf $F_{i|j_1,...,j_n}$ by integration via other margins, we have to find a set of vertices $H = \{h_1, ..., h_n\} \subset V \backslash \{i, j_1, ..., j_n\}$ for that we can decompose $f_{ij_1...j_n h_1...h_n}$ with copulas available in the decomposition of $f(\mathbf{x})$.

$$
\begin{aligned}
F_{i|j_1\dots j_n}(x_i|x_{j_1},\dots,x_{j_n}) &= \int_{-\infty}^{x_i} f_{i|j_1\dots j_n}(y_i|x_{j_1},\dots,x_{j_n})\mathrm{d}y_i \\
&= f^{-1}_{j_1\dots j_n}(x_{j_1},\dots,x_{j_n}) \int_{-\infty}^{x_i} f_{ij_1\dots j_n}(y_i,x_{j_1},\dots,x_{j_n})\mathrm{d}y_i \\
&= \frac{\int_{-\infty}^{x_i} \overbrace{\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}}^{n \text{ times}} f_{ij_1\dots j_n h_1\dots h_n}(y_i,x_{j_1},\dots,x_{j_n},y_{h_1},\dots,y_{h_n})\mathrm{d}y_{h_1}\dots\mathrm{d}y_{h_n}\mathrm{d}y_i}{\underbrace{\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}}_{n+1 \text{ times}} f_{ij_1\dots j_n h_1\dots h_n}(y_i,x_{j_1},\dots,x_{j_n},y_{h_1},\dots,y_{h_n})\mathrm{d}y_{h_1}\dots\mathrm{d}y_{h_n}\mathrm{d}y_i}
\end{aligned}
$$

In the last term we can substitute $f_{ij_1\dots j_n h_1\dots h_n}$ with the pair-copula decomposition and compute the conditional cdf. We showed in Sections 1.2 and 2.3 that in some cases, especially when the conditional cdf is only conditioned on 1 variable, we do not need an integration of the denominater. It may also be possible to reduce the dimension of integrals by 1 using a h-function.

The most simple DAG which needs a conditional cdf to be computed via other margins is illustrated in Figure 5.1a, the edge labels are the indices of the corresponding (conditional) pair-copulas. We need to compute

$$
F_{3|2}(x_3|x_2) = \int_0^1 c_{12}(y_1,x_2)h_{1\underline{3}}(y_1,x_3)dy_1 \tag{5.1}
$$

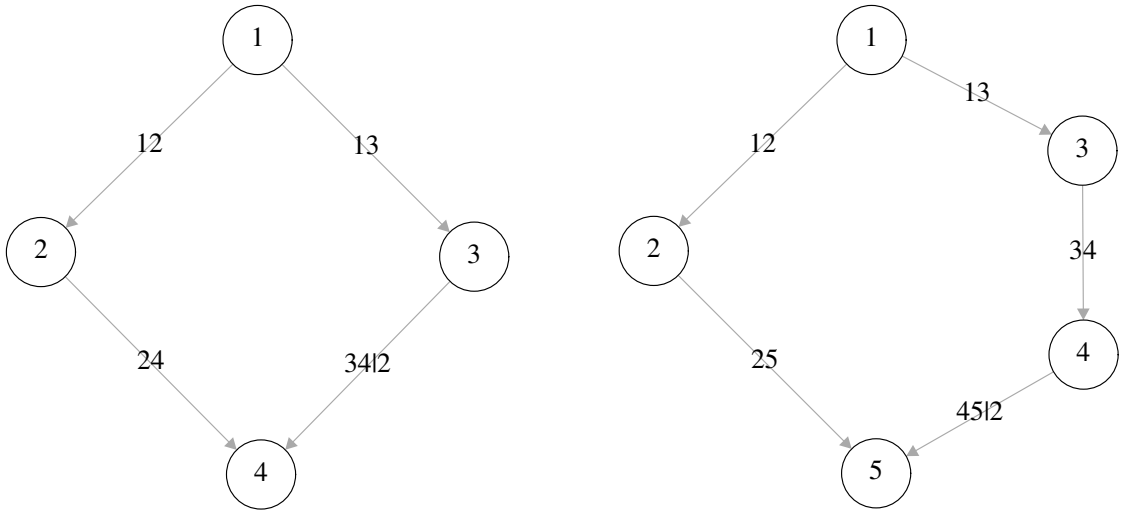


Figure 5.1

We can extend the DAG by one vertex like in Figure 5.1b so that we need to integrate

via one additional margin.

$$F_{4|2}(x_4|x_2) = \int_0^1 \int_0^1 c_{13}(y_1, y_3) c_{12}(y_1, x_2) h_{3\underline{4}}(y_3, x_4) dy_1 dy_3$$

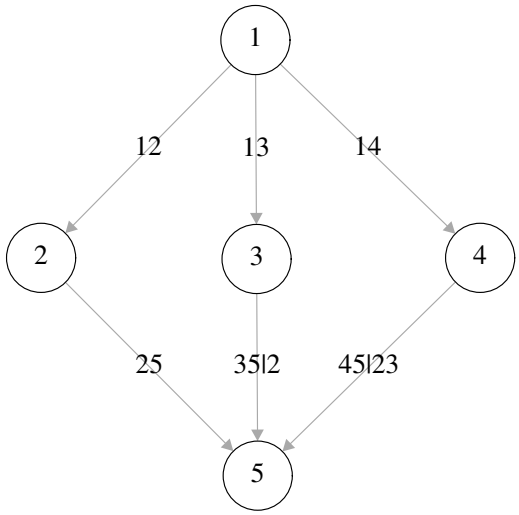


Figure 5.2

Figure 5.2 shows a simple example where we have to compute a conditional cdf conditioned on 2 variables.

$$F_{4|23}(x_4|x_2, x_3) = \frac{\int_0^1 c_{13}(y_1, x_3) c_{12}(y_1, x_2) h_{1\underline{4}}(y_1, x_4) dy_1}{\int_0^1 c_{13}(y_1, x_3) c_{12}(y_1, x_2) dy_1}$$

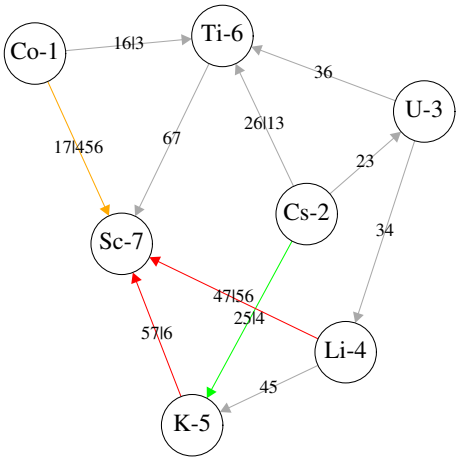
Note that in this DAG we also need to compute $F_{3|2}$ by integration via other margins. The formula is (5.1).

5.1 PCBN decompositions of the uranium dataset

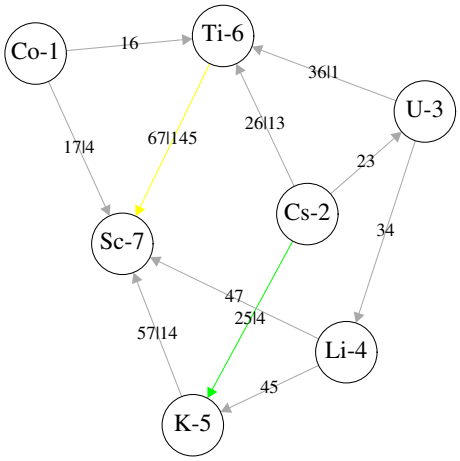
Our analysis of the uranium dataset, see Section 4.2, showed that the computations of PCBN decompositions based on different parent orderings can have very different characteristics.

- PCBN 2 only needed 2 conditional cdfs to be integrated via other margins, the highest integral dimension was 2.
- PCBN 1, PCBN 3 and PCBN 4 needed 4 conditional cdfs to be integrated via other margins, the highest integral dimension was 4.

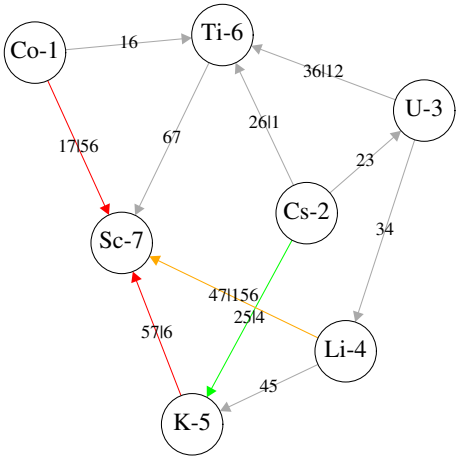
The formulas of the conditional cdfs can be found in Appendix B. We illustrated the necessary integrations of each PCBN model in Figure 5.3.



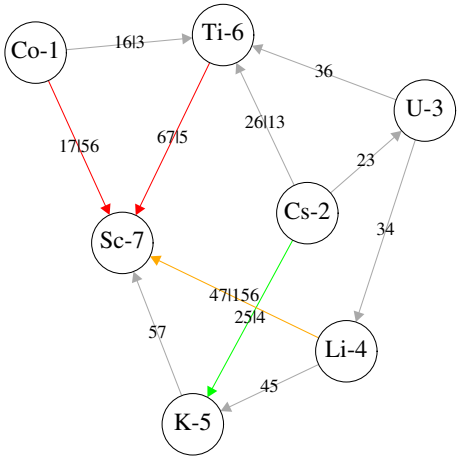
PCBN 1



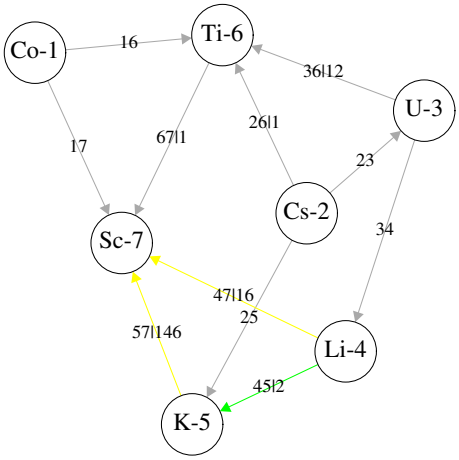
PCBN 2



PCBN 3



PCBN 4



PCBN 5

Figure 5.3: DAGs for PCBN 1-5 of the uranium dataset, described in Section 4.2. The edge labels represent the indices of the corresponding (conditional) pair copulas. The color of the edges illustrate the highest integral dimension in the computation of the conditional cdfs that are arguments of the corresponding copula. (gray = no integral, green = 1, yellow = 2, orange = 3, red = 4)

In PCBNs 1, 3 and 4 we have to determine either $F_{6|5}$ or $F_{5|6}$ because $c_{67|5}$ or respectively $c_{57|6}$ are part of the decomposition. This has to be done by integration via the vertices 1, 2, 3 and 4 which means a 4-dimensional integration. In contrast, PCBN 2 makes use of the conditional independences because $F_{1|4} = F_1$ and $F_{5|14} = F_{5|4}$.

This example shows the big effect of different parent orderings for vertices with more than 2 parents because we have to compute different conditional cdfs. It also gives the idea to order the parents in a way to make use of the (conditional) independences that are present in the data.

Chapter 6

Conclusion

We showed that pair-copula Bayesian networks with different parent orderings can be highly varying in quality and computation time.

Chapter 2 explained how to determine possible DAGs and decompositions for some given conditional independence assumptions. The corresponding DAGs are based on the same skeleton and all immoralities of the graph are also defined by the conditional independence assumptions. This is an important result because immoralities can lead to cdfs that have to be integrated via other margins. So these integrations can not be avoided by the choice of other DAGs or parent orderings. This is shown by the example in this chapter. The example is also an instance where some parent orderings need higher dimensional integrations than other parent orderings.

In Chapter 3 we analyzed three cases with simulated data. The data was sampled by one PCBN with a certain parent ordering and then fitted on PCBNs with other parent orderings. If the decomposition of one fitted PCBN contains a copula that is not available in the decomposition of the PCBN which the data was sampled by, then the results point out that the model has a lower log-likelihood if this copula can not be fitted well by any copula family. So we shouldnt choose our parent ordering solely by optimal computation time but also have to test the quality in comparison with PCBNs with other parent orderings.

Following we applied PCBN models with different parent orderings to natural data in Chapter 4. It was shown that these models can have significantly varying log-likelihoods and very highly varying computation times due to high-dimensional integrations. A fitted RVine had a superior log-likelihood in all cases, but may not be preferred when considering the number of parameters and computation time.

In Chapter 5, we analyzed when a PCBN needs higher-dimensional integrations. A comparison of the PCBN decompositions we used for the uranium dataset suggested that the parents should be ordered in a way to make use of (conditional) independences when computing the conditional cdfs.

Pair-copula Bayesian networks are an interesting alternative to RVines, especially regarding the computation time. The choice of the parent orderings is a very important, but difficult task because one has to optimize both log-likelihood and computation time. The algorithm of Bauer and Czado (2012) which tries to find optimal parent orderings provided good results in my application.

It still remains to analyze the choice of a DAG. The partially directed graph which is induced by the conditional independences allows multiple DAGs in the non-trivial case. Since these DAGs overlap a lot in their possible PCBN decompositions, it would be useful to directly determine all PCBN decompositions based on the conditional independences. It is also an open question how we test natural data for conditional independence. We could adjust our significance niveau to obtain more or less conditional independences which would result in less or respectively more connections in the DAG. Another possiblity is to make special DAG modifications, such as limiting the amount of parents for one vertex.

Bibliography

Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44:182–198, 2009.

Alexander Bauer and Claudia Czado. Pair-copula Bayesian networks. submitted for publication, arXiv:1211.5620[stat.ME], 2012.

Alexander Bauer, Claudia Czado, and Thomas Klein. Pair-copula constructions for non-Gaussian DAG models. *The Canadian Journal of Statistics*, 40(1):86–109, 2012.

Tim Bedford and Roger Marvin Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32:245–268, 2001.

Tim Bedford and Roger Marvin Cooke. Vines—A new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

Claudia Czado, Ulf Schepsmeier, and Aleksey Min. Maximum likelihood estimation of mixed C-vines with application to exchange rates. *Statistical Modelling*, 12(3):229–255, 2012.

Anca Maria Hanea, Dorota Kurowicka, and Roger Marvin Cooke. Hybrid method for quantifying and analyzing Bayesian belief nets. *Quality and Reliability Engineering International*, 22:709–729, 2006.

Harry Joe. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Ludger Rï¿½schendorf, Berthold Schweizer, and Michael Dee Taylor, editors, *Distributions with Fixed Marginals and Related Topics*, volume 28 of *Lecture Notes—Monograph Series*, pages 120–141. Institute of Mathematical Statistics, Hayward, California, 1996.

Timo Koski and John M. Noble. *Bayesian Networks: An Introduction*. John Wiley & Sons, Chichester, 2009.

Steffen Lilholt Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

Abe 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.

Appendix A

Uranium dataset: all parent orderings

This table shows the log-likelihoods and its calculation time of all possible parent orderings for the DAG we used for the uranium dataset. The log-likelihoods have to be interpreted with caution because the calculation caused several arithmetic exceptions.

The columns 'K-1', 'K-2' represent the parents of K in the order that K-1 is the lowest parent. Same applies for the columns of the parents of Ti and Sc.

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
730.49	4.76s	Cs	Li	Co	Cs	U	Co	K	Li	Ti
770.20	66.16s	Cs	Li	Co	Cs	U	Co	K	Ti	Li
728.90	7.11s	Cs	Li	Co	Cs	U	Co	Li	K	Ti
779.49	3.59s	Cs	Li	Co	Cs	U	Co	Li	Ti	K
773.20	63.10s	Cs	Li	Co	Cs	U	Co	Ti	K	Li
783.55	3.28s	Cs	Li	Co	Cs	U	Co	Ti	Li	K
725.01	4.96s	Cs	Li	Co	Cs	U	K	Co	Li	Ti
761.88	75.68s	Cs	Li	Co	Cs	U	K	Co	Ti	Li
723.93	5.11s	Cs	Li	Co	Cs	U	K	Li	Co	Ti
724.83	40.40s	Cs	Li	Co	Cs	U	K	Li	Ti	Co
779.74	280.94s	Cs	Li	Co	Cs	U	K	Ti	Co	Li
720.50	312.91s	Cs	Li	Co	Cs	U	K	Ti	Li	Co
730.88	6.65s	Cs	Li	Co	Cs	U	Li	Co	K	Ti
780.93	3.48s	Cs	Li	Co	Cs	U	Li	Co	Ti	K
720.35	6.40s	Cs	Li	Co	Cs	U	Li	K	Co	Ti
743.11	40.86s	Cs	Li	Co	Cs	U	Li	K	Ti	Co
779.77	8.17s	Cs	Li	Co	Cs	U	Li	Ti	Co	K
758.04	40.77s	Cs	Li	Co	Cs	U	Li	Ti	K	Co
769.51	78.74s	Cs	Li	Co	Cs	U	Ti	Co	K	Li
779.05	3.38s	Cs	Li	Co	Cs	U	Ti	Co	Li	K

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
779.05	3.38s	Cs	Li	Co	Cs	U	Ti	Co	Li	K
790.90	335.89s	Cs	Li	Co	Cs	U	Ti	K	Co	Li
730.71	308.01s	Cs	Li	Co	Cs	U	Ti	K	Li	Co
782.25	10.19s	Cs	Li	Co	Cs	U	Ti	Li	Co	K
760.45	41.30s	Cs	Li	Co	Cs	U	Ti	Li	K	Co
726.83	4.84s	Cs	Li	Co	U	Cs	Co	K	Li	Ti
755.50	62.59s	Cs	Li	Co	U	Cs	Co	K	Ti	Li
724.34	5.95s	Cs	Li	Co	U	Cs	Co	Li	K	Ti
770.34	2.59s	Cs	Li	Co	U	Cs	Co	Li	Ti	K
758.71	62.16s	Cs	Li	Co	U	Cs	Co	Ti	K	Li
772.21	2.92s	Cs	Li	Co	U	Cs	Co	Ti	Li	K
717.28	4.63s	Cs	Li	Co	U	Cs	K	Co	Li	Ti
747.62	64.94s	Cs	Li	Co	U	Cs	K	Co	Ti	Li
716.80	4.95s	Cs	Li	Co	U	Cs	K	Li	Co	Ti
713.86	34.70s	Cs	Li	Co	U	Cs	K	Li	Ti	Co
731.31	255.93s	Cs	Li	Co	U	Cs	K	Ti	Co	Li
701.00	274.62s	Cs	Li	Co	U	Cs	K	Ti	Li	Co
725.60	6.54s	Cs	Li	Co	U	Cs	Li	Co	K	Ti
771.92	2.60s	Cs	Li	Co	U	Cs	Li	Co	Ti	K
713.58	6.67s	Cs	Li	Co	U	Cs	Li	K	Co	Ti
728.54	29.42s	Cs	Li	Co	U	Cs	Li	K	Ti	Co
759.29	6.27s	Cs	Li	Co	U	Cs	Li	Ti	Co	K
750.57	33.87s	Cs	Li	Co	U	Cs	Li	Ti	K	Co
755.88	57.96s	Cs	Li	Co	U	Cs	Ti	Co	K	Li
766.86	2.98s	Cs	Li	Co	U	Cs	Ti	Co	Li	K
779.14	253.50s	Cs	Li	Co	U	Cs	Ti	K	Co	Li
713.99	240.88s	Cs	Li	Co	U	Cs	Ti	K	Li	Co
764.33	7.28s	Cs	Li	Co	U	Cs	Ti	Li	Co	K
756.91	35.76s	Cs	Li	Co	U	Cs	Ti	Li	K	Co
717.03	5.30s	Cs	Li	Cs	Co	U	Co	K	Li	Ti
555.24	73.92s	Cs	Li	Cs	Co	U	Co	K	Ti	Li
714.09	6.75s	Cs	Li	Cs	Co	U	Co	Li	K	Ti
610.23	3.48s	Cs	Li	Cs	Co	U	Co	Li	Ti	K
609.20	62.15s	Cs	Li	Cs	Co	U	Co	Ti	K	Li
562.07	3.05s	Cs	Li	Cs	Co	U	Co	Ti	Li	K
710.79	5.53s	Cs	Li	Cs	Co	U	K	Co	Li	Ti
753.99	70.62s	Cs	Li	Cs	Co	U	K	Co	Ti	Li
709.88	5.57s	Cs	Li	Cs	Co	U	K	Li	Co	Ti
701.39	32.49s	Cs	Li	Cs	Co	U	K	Li	Ti	Co
519.25	283.32s	Cs	Li	Cs	Co	U	K	Ti	Co	Li
685.46	289.72s	Cs	Li	Cs	Co	U	K	Ti	Li	Co
718.29	6.33s	Cs	Li	Cs	Co	U	Li	Co	K	Ti
773.96	3.34s	Cs	Li	Cs	Co	U	Li	Co	Ti	K
707.29	7.09s	Cs	Li	Cs	Co	U	Li	K	Co	Ti
703.13	38.69s	Cs	Li	Cs	Co	U	Li	K	Ti	Co

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
759.12	7.78s	Cs	Li	Cs	Co	U	Li	Ti	Co	K
742.49	36.58s	Cs	Li	Cs	Co	U	Li	Ti	K	Co
608.51	61.31s	Cs	Li	Cs	Co	U	Ti	Co	K	Li
559.30	3.05s	Cs	Li	Cs	Co	U	Ti	Co	Li	K
529.58	297.20s	Cs	Li	Cs	Co	U	Ti	K	Co	Li
699.48	281.81s	Cs	Li	Cs	Co	U	Ti	K	Li	Co
758.45	9.06s	Cs	Li	Cs	Co	U	Ti	Li	Co	K
742.80	36.47s	Cs	Li	Cs	Co	U	Ti	Li	K	Co
717.50	6.94s	Cs	Li	Cs	U	Co	Co	K	Li	Ti
564.79	124.27s	Cs	Li	Cs	U	Co	Co	K	Ti	Li
714.46	8.84s	Cs	Li	Cs	U	Co	Co	Li	K	Ti
611.47	8.64s	Cs	Li	Cs	U	Co	Co	Li	Ti	K
613.03	117.11s	Cs	Li	Cs	U	Co	Co	Ti	K	Li
566.94	11.87s	Cs	Li	Cs	U	Co	Co	Ti	Li	K
711.52	6.35s	Cs	Li	Cs	U	Co	K	Co	Li	Ti
754.16	121.20s	Cs	Li	Cs	U	Co	K	Co	Ti	Li
710.40	6.38s	Cs	Li	Cs	U	Co	K	Li	Co	Ti
701.75	19.89s	Cs	Li	Cs	U	Co	K	Li	Ti	Co
529.90	217.04s	Cs	Li	Cs	U	Co	K	Ti	Co	Li
685.66	143.55s	Cs	Li	Cs	U	Co	K	Ti	Li	Co
718.36	8.20s	Cs	Li	Cs	U	Co	Li	Co	K	Ti
609.52	8.65s	Cs	Li	Cs	U	Co	Li	Co	Ti	K
707.67	8.39s	Cs	Li	Cs	U	Co	Li	K	Co	Ti
731.14	19.32s	Cs	Li	Cs	U	Co	Li	K	Ti	Co
753.66	12.96s	Cs	Li	Cs	U	Co	Li	Ti	Co	K
744.90	19.16s	Cs	Li	Cs	U	Co	Li	Ti	K	Co
612.09	119.11s	Cs	Li	Cs	U	Co	Ti	Co	K	Li
563.89	13.78s	Cs	Li	Cs	U	Co	Ti	Co	Li	K
540.37	215.96s	Cs	Li	Cs	U	Co	Ti	K	Co	Li
699.26	131.36s	Cs	Li	Cs	U	Co	Ti	K	Li	Co
602.38	13.34s	Cs	Li	Cs	U	Co	Ti	Li	Co	K
746.80	21.08s	Cs	Li	Cs	U	Co	Ti	Li	K	Co
710.01	4.68s	Cs	Li	U	Co	Cs	Co	K	Li	Ti
525.80	70.42s	Cs	Li	U	Co	Cs	Co	K	Ti	Li
708.73	6.81s	Cs	Li	U	Co	Cs	Co	Li	K	Ti
771.91	4.28s	Cs	Li	U	Co	Cs	Co	Li	Ti	K
567.88	62.93s	Cs	Li	U	Co	Cs	Co	Ti	K	Li
527.03	3.85s	Cs	Li	U	Co	Cs	Co	Ti	Li	K
704.21	5.67s	Cs	Li	U	Co	Cs	K	Co	Li	Ti
521.48	71.88s	Cs	Li	U	Co	Cs	K	Co	Ti	Li
701.81	5.70s	Cs	Li	U	Co	Cs	K	Li	Co	Ti
687.84	113.80s	Cs	Li	U	Co	Cs	K	Li	Ti	Co
494.49	765.15s	Cs	Li	U	Co	Cs	K	Ti	Co	Li
693.12	978.79s	Cs	Li	U	Co	Cs	K	Ti	Li	Co
710.64	6.72s	Cs	Li	U	Co	Cs	Li	Co	K	Ti

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
772.55	4.24s	Cs	Li	U	Co	Cs	Li	Co	Ti	K
699.02	7.01s	Cs	Li	U	Co	Cs	Li	K	Co	Ti
690.31	113.63s	Cs	Li	U	Co	Cs	Li	K	Ti	Co
771.08	35.73s	Cs	Li	U	Co	Cs	Li	Ti	Co	K
738.18	137.73s	Cs	Li	U	Co	Cs	Li	Ti	K	Co
522.11	74.05s	Cs	Li	U	Co	Cs	Ti	Co	K	Li
524.22	3.96s	Cs	Li	U	Co	Cs	Ti	Co	Li	K
502.43	786.54s	Cs	Li	U	Co	Cs	Ti	K	Co	Li
729.05	973.63s	Cs	Li	U	Co	Cs	Ti	K	Li	Co
770.15	45.05s	Cs	Li	U	Co	Cs	Ti	Li	Co	K
738.16	146.59s	Cs	Li	U	Co	Cs	Ti	Li	K	Co
720.15	5.02s	Cs	Li	U	Cs	Co	Co	K	Li	Ti
554.21	64.59s	Cs	Li	U	Cs	Co	Co	K	Ti	Li
717.19	6.31s	Cs	Li	U	Cs	Co	Co	Li	K	Ti
600.27	3.70s	Cs	Li	U	Cs	Co	Co	Li	Ti	K
607.17	58.16s	Cs	Li	U	Cs	Co	Co	Ti	K	Li
559.73	3.79s	Cs	Li	U	Cs	Co	Co	Ti	Li	K
712.01	4.71s	Cs	Li	U	Cs	Co	K	Co	Li	Ti
746.49	61.68s	Cs	Li	U	Cs	Co	K	Co	Ti	Li
711.65	4.64s	Cs	Li	U	Cs	Co	K	Li	Co	Ti
694.87	9.87s	Cs	Li	U	Cs	Co	K	Li	Ti	Co
734.48	107.83s	Cs	Li	U	Cs	Co	K	Ti	Co	Li
675.88	76.58s	Cs	Li	U	Cs	Co	K	Ti	Li	Co
719.47	6.89s	Cs	Li	U	Cs	Co	Li	Co	K	Ti
598.38	3.09s	Cs	Li	U	Cs	Co	Li	Co	Ti	K
709.11	6.64s	Cs	Li	U	Cs	Co	Li	K	Co	Ti
697.40	11.99s	Cs	Li	U	Cs	Co	Li	K	Ti	Co
593.03	3.27s	Cs	Li	U	Cs	Co	Li	Ti	Co	K
735.40	7.39s	Cs	Li	U	Cs	Co	Li	Ti	K	Co
605.40	58.05s	Cs	Li	U	Cs	Co	Ti	Co	K	Li
556.16	4.29s	Cs	Li	U	Cs	Co	Ti	Co	Li	K
528.00	94.71s	Cs	Li	U	Cs	Co	Ti	K	Co	Li
689.65	67.96s	Cs	Li	U	Cs	Co	Ti	K	Li	Co
595.09	3.28s	Cs	Li	U	Cs	Co	Ti	Li	Co	K
737.20	9.91s	Cs	Li	U	Cs	Co	Ti	Li	K	Co
788.49	2.92s	Li	Cs	Co	Cs	U	Co	K	Li	Ti
768.93	81.94s	Li	Cs	Co	Cs	U	Co	K	Ti	Li
785.75	2.99s	Li	Cs	Co	Cs	U	Co	Li	K	Ti
780.53	2.95s	Li	Cs	Co	Cs	U	Co	Li	Ti	K
772.56	80.47s	Li	Cs	Co	Cs	U	Co	Ti	K	Li
784.31	2.96s	Li	Cs	Co	Cs	U	Co	Ti	Li	K
777.23	2.96s	Li	Cs	Co	Cs	U	K	Co	Li	Ti
761.24	87.00s	Li	Cs	Co	Cs	U	K	Co	Ti	Li
776.51	3.00s	Li	Cs	Co	Cs	U	K	Li	Co	Ti
770.17	34.37s	Li	Cs	Co	Cs	U	K	Li	Ti	Co

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
749.75	377.55s	Li	Cs	Co	Cs	U	K	Ti	Co	Li
748.80	383.55s	Li	Cs	Co	Cs	U	K	Ti	Li	Co
787.24	2.73s	Li	Cs	Co	Cs	U	Li	Co	K	Ti
781.66	3.70s	Li	Cs	Co	Cs	U	Li	Co	Ti	K
771.42	3.21s	Li	Cs	Co	Cs	U	Li	K	Co	Ti
767.38	34.53s	Li	Cs	Co	Cs	U	Li	K	Ti	Co
779.98	7.74s	Li	Cs	Co	Cs	U	Li	Ti	Co	K
756.60	38.86s	Li	Cs	Co	Cs	U	Li	Ti	K	Co
769.36	84.29s	Li	Cs	Co	Cs	U	Ti	Co	K	Li
779.67	3.19s	Li	Cs	Co	Cs	U	Ti	Co	Li	K
791.68	359.13s	Li	Cs	Co	Cs	U	Ti	K	Co	Li
732.44	369.45s	Li	Cs	Co	Cs	U	Ti	K	Li	Co
782.11	9.17s	Li	Cs	Co	Cs	U	Ti	Li	Co	K
758.69	38.79s	Li	Cs	Co	Cs	U	Ti	Li	K	Co
778.86	2.80s	Li	Cs	Co	U	Cs	Co	K	Li	Ti
757.95	77.97s	Li	Cs	Co	U	Cs	Co	K	Ti	Li
776.18	2.59s	Li	Cs	Co	U	Cs	Co	Li	K	Ti
771.31	2.63s	Li	Cs	Co	U	Cs	Co	Li	Ti	K
761.74	68.98s	Li	Cs	Co	U	Cs	Co	Ti	K	Li
772.90	2.54s	Li	Cs	Co	U	Cs	Co	Ti	Li	K
765.42	2.88s	Li	Cs	Co	U	Cs	K	Co	Li	Ti
749.83	84.68s	Li	Cs	Co	U	Cs	K	Co	Ti	Li
763.74	2.50s	Li	Cs	Co	U	Cs	K	Li	Co	Ti
756.78	26.13s	Li	Cs	Co	U	Cs	K	Li	Ti	Co
734.97	317.26s	Li	Cs	Co	U	Cs	K	Ti	Co	Li
706.91	311.14s	Li	Cs	Co	U	Cs	K	Ti	Li	Co
777.42	2.41s	Li	Cs	Co	U	Cs	Li	Co	K	Ti
772.59	3.07s	Li	Cs	Co	U	Cs	Li	Co	Ti	K
759.68	2.71s	Li	Cs	Co	U	Cs	Li	K	Co	Ti
754.12	31.21s	Li	Cs	Co	U	Cs	Li	K	Ti	Co
760.15	6.90s	Li	Cs	Co	U	Cs	Li	Ti	Co	K
744.97	33.08s	Li	Cs	Co	U	Cs	Li	Ti	K	Co
759.26	78.32s	Li	Cs	Co	U	Cs	Ti	Co	K	Li
767.44	2.89s	Li	Cs	Co	U	Cs	Ti	Co	Li	K
780.74	303.73s	Li	Cs	Co	U	Cs	Ti	K	Co	Li
719.26	289.40s	Li	Cs	Co	U	Cs	Ti	K	Li	Co
764.57	7.62s	Li	Cs	Co	U	Cs	Ti	Li	Co	K
750.67	29.95s	Li	Cs	Co	U	Cs	Ti	Li	K	Co
776.90	2.94s	Li	Cs	Cs	Co	U	Co	K	Li	Ti
550.25	91.52s	Li	Cs	Cs	Co	U	Co	K	Ti	Li
773.80	2.81s	Li	Cs	Cs	Co	U	Co	Li	K	Ti
610.47	3.27s	Li	Cs	Cs	Co	U	Co	Li	Ti	K
610.49	77.36s	Li	Cs	Cs	Co	U	Co	Ti	K	Li
562.04	3.06s	Li	Cs	Cs	Co	U	Co	Ti	Li	K
763.72	2.89s	Li	Cs	Cs	Co	U	K	Co	Li	Ti

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
753.93	95.63s	Li	Cs	Cs	Co	U	K	Co	Ti	Li
763.11	2.64s	Li	Cs	Cs	Co	U	K	Li	Co	Ti
731.49	30.66s	Li	Cs	Cs	Co	U	K	Li	Ti	Co
727.03	351.68s	Li	Cs	Cs	Co	U	K	Ti	Co	Li
733.60	333.89s	Li	Cs	Cs	Co	U	K	Ti	Li	Co
775.81	3.10s	Li	Cs	Cs	Co	U	Li	Co	K	Ti
608.57	3.53s	Li	Cs	Cs	Co	U	Li	Co	Ti	K
758.75	3.14s	Li	Cs	Cs	Co	U	Li	K	Co	Ti
731.09	28.95s	Li	Cs	Cs	Co	U	Li	K	Ti	Co
759.46	7.67s	Li	Cs	Cs	Co	U	Li	Ti	Co	K
741.41	32.93s	Li	Cs	Cs	Co	U	Li	Ti	K	Co
610.07	70.87s	Li	Cs	Cs	Co	U	Ti	Co	K	Li
559.06	3.04s	Li	Cs	Cs	Co	U	Ti	Co	Li	K
526.52	343.10s	Li	Cs	Cs	Co	U	Ti	K	Co	Li
701.50	316.08s	Li	Cs	Cs	Co	U	Ti	K	Li	Co
758.54	8.45s	Li	Cs	Cs	Co	U	Ti	Li	Co	K
741.51	32.83s	Li	Cs	Cs	Co	U	Ti	Li	K	Co
776.40	4.32s	Li	Cs	Cs	U	Co	Co	K	Li	Ti
560.74	151.09s	Li	Cs	Cs	U	Co	Co	K	Ti	Li
773.30	4.22s	Li	Cs	Cs	U	Co	Co	Li	K	Ti
613.94	8.57s	Li	Cs	Cs	U	Co	Co	Li	Ti	K
614.61	136.01s	Li	Cs	Cs	U	Co	Co	Ti	K	Li
566.98	11.88s	Li	Cs	Cs	U	Co	Co	Ti	Li	K
763.55	4.65s	Li	Cs	Cs	U	Co	K	Co	Li	Ti
754.25	139.09s	Li	Cs	Cs	U	Co	K	Co	Ti	Li
762.98	4.49s	Li	Cs	Cs	U	Co	K	Li	Co	Ti
730.52	17.50s	Li	Cs	Cs	U	Co	K	Li	Ti	Co
727.07	303.82s	Li	Cs	Cs	U	Co	K	Ti	Co	Li
733.22	153.41s	Li	Cs	Cs	U	Co	K	Ti	Li	Co
775.18	4.73s	Li	Cs	Cs	U	Co	Li	Co	K	Ti
611.80	8.58s	Li	Cs	Cs	U	Co	Li	Co	Ti	K
758.55	4.43s	Li	Cs	Cs	U	Co	Li	K	Co	Ti
730.04	16.60s	Li	Cs	Cs	U	Co	Li	K	Ti	Co
753.68	12.76s	Li	Cs	Cs	U	Co	Li	Ti	Co	K
743.96	19.33s	Li	Cs	Cs	U	Co	Li	Ti	K	Co
613.68	145.20s	Li	Cs	Cs	U	Co	Ti	Co	K	Li
563.72	13.46s	Li	Cs	Cs	U	Co	Ti	Co	Li	K
537.22	266.67s	Li	Cs	Cs	U	Co	Ti	K	Co	Li
703.15	162.91s	Li	Cs	Cs	U	Co	Ti	K	Li	Co
754.62	13.39s	Li	Cs	Cs	U	Co	Ti	Li	Co	K
745.49	20.03s	Li	Cs	Cs	U	Co	Ti	Li	K	Co
776.21	2.87s	Li	Cs	U	Co	Cs	Co	K	Li	Ti
525.95	95.23s	Li	Cs	U	Co	Cs	Co	K	Ti	Li
773.88	2.96s	Li	Cs	U	Co	Cs	Co	Li	K	Ti
554.35	3.89s	Li	Cs	U	Co	Cs	Co	Li	Ti	K

LL	Time	K-1	K-2	Ti-1	Ti-2	Ti-3	Sc-1	Sc-2	Sc-3	Sc-4
568.83	74.01s	Li	Cs	U	Co	Cs	Co	Ti	K	Li
527.45	4.05s	Li	Cs	U	Co	Cs	Co	Ti	Li	K
762.32	3.07s	Li	Cs	U	Co	Cs	K	Co	Li	Ti
520.74	87.68s	Li	Cs	U	Co	Cs	K	Co	Ti	Li
761.85	3.19s	Li	Cs	U	Co	Cs	K	Li	Co	Ti
718.65	114.51s	Li	Cs	U	Co	Cs	K	Li	Ti	Co
493.09	1061.60s	Li	Cs	U	Co	Cs	K	Ti	Co	Li
697.09	1228.30s	Li	Cs	U	Co	Cs	K	Ti	Li	Co
774.59	3.24s	Li	Cs	U	Co	Cs	Li	Co	K	Ti
772.68	4.16s	Li	Cs	U	Co	Cs	Li	Co	Ti	K
756.55	3.15s	Li	Cs	U	Co	Cs	Li	K	Co	Ti
719.00	102.51s	Li	Cs	U	Co	Cs	Li	K	Ti	Co
771.33	35.50s	Li	Cs	U	Co	Cs	Li	Ti	Co	K
737.42	137.29s	Li	Cs	U	Co	Cs	Li	Ti	K	Co
566.55	74.18s	Li	Cs	U	Co	Cs	Ti	Co	K	Li
524.46	3.97s	Li	Cs	U	Co	Cs	Ti	Co	Li	K
501.47	1023.11s	Li	Cs	U	Co	Cs	Ti	K	Co	Li
749.27	1251.20s	Li	Cs	U	Co	Cs	Ti	K	Li	Co
770.16	46.02s	Li	Cs	U	Co	Cs	Ti	Li	Co	K
736.93	139.29s	Li	Cs	U	Co	Cs	Ti	Li	K	Co
769.86	2.67s	Li	Cs	U	Cs	Co	Co	K	Li	Ti
552.88	79.26s	Li	Cs	U	Cs	Co	Co	K	Ti	Li
766.80	2.72s	Li	Cs	U	Cs	Co	Co	Li	K	Ti
603.06	3.18s	Li	Cs	U	Cs	Co	Co	Li	Ti	K
607.10	62.33s	Li	Cs	U	Cs	Co	Co	Ti	K	Li
559.76	3.71s	Li	Cs	U	Cs	Co	Co	Ti	Li	K
756.03	2.84s	Li	Cs	U	Cs	Co	K	Co	Li	Ti
746.68	82.35s	Li	Cs	U	Cs	Co	K	Co	Ti	Li
755.54	2.80s	Li	Cs	U	Cs	Co	K	Li	Co	Ti
721.39	8.35s	Li	Cs	U	Cs	Co	K	Li	Ti	Co
719.05	122.04s	Li	Cs	U	Cs	Co	K	Ti	Co	Li
724.94	77.22s	Li	Cs	U	Cs	Co	K	Ti	Li	Co
768.62	2.58s	Li	Cs	U	Cs	Co	Li	Co	K	Ti
600.89	3.28s	Li	Cs	U	Cs	Co	Li	Co	Ti	K
751.18	2.54s	Li	Cs	U	Cs	Co	Li	K	Co	Ti
721.06	7.79s	Li	Cs	U	Cs	Co	Li	K	Ti	Co
596.42	3.25s	Li	Cs	U	Cs	Co	Li	Ti	Co	K
734.05	6.97s	Li	Cs	U	Cs	Co	Li	Ti	K	Co
605.32	72.81s	Li	Cs	U	Cs	Co	Ti	Co	K	Li
555.96	3.86s	Li	Cs	U	Cs	Co	Ti	Co	Li	K
526.45	114.59s	Li	Cs	U	Cs	Co	Ti	K	Co	Li
693.47	66.68s	Li	Cs	U	Cs	Co	Ti	K	Li	Co
597.94	3.88s	Li	Cs	U	Cs	Co	Ti	Li	Co	K
735.35	7.54s	Li	Cs	U	Cs	Co	Ti	Li	K	Co

Appendix B

Uranium dataset: PCBN decompositions and conditional cdfs

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

PCBN 1

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2,u_3)c_{3,4}(u_3,u_4)c_{4,5}(u_4,u_5)c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))c_{3,6}(u_3,u_6) \\ & c_{1,6|3}(u_1,F_{6|3}(u_6|u_3))c_{2,6|1,3}(F_{2|3}(u_2|u_3),F_{6|1,3}(u_6|u_1,u_3))c_{6,7}(u_6,u_7) \\ & c_{5,7|6}(F_{5|6}(u_5|u_6),F_{7|6}(u_7|u_6))c_{4,7|5,6}(F_{4|5,6}(u_4|u_5,u_6),F_{7|5,6}(u_7|u_5,u_6)) \\ & c_{4,7|5,6}(F_{4|5,6}(u_4|u_5,u_6),F_{7|5,6}(u_7|u_5,u_6))c_{1,7|4,5,6}(F_{1|4,5,6}(u_1|u_4,u_5,u_6),F_{7|4,5,6}(u_7|u_4,u_5,u_6)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3 : c_{2,3}(u_2,u_3)$:
Frank(theta=5.50008)

$3->4 : c_{3,4}(u_3,u_4)$:
Frank(theta=1.29653)

$4->5 : c_{4,5}(u_4,u_5)$:
ReflectUV Joe(theta=1.23974)

$2->5 : c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))$:
$F_{2|4}(u_2|u_4) = \int_0^1 c_{3,4}(v_3,u_4)h1_{2,3}(u_2,v_3)dv_3$
$F_{5|4}(u_5|u_4) = h2_{4,5}(u_4,u_5)$
Frank(theta=1.6798)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

$3->6 : c_{3,6}(u_3, u_6)$:
ReflectUV Clayton(theta=0.328713)

$1->6 : c_{1,6|3}(u_1, F_{6|3}(u_6|u_3))$:
$F_{6|3}(u_6|u_3) = h2_{3,6}(u_3, u_6)$
Frank(theta=3.11038)

$2->6 : c_{2,6|1,3}(F_{2|3}(u_2|u_3), F_{6|1,3}(u_6|u_1, u_3))$:
$F_{2|3}(u_2|u_3) = h1_{2,3}(u_2, u_3)$
$F_{6|1,3}(u_6|u_1, u_3) = h2_{1,6|3}(u_1, F_{6|3}(u_6|u_3))$
ReflectUV Clayton(theta=0.713207)

$6->7 : c_{6,7}(u_6, u_7)$:
Gauss(theta=0.645648)

$5->7 : c_{5,7|6}(F_{5|6}(u_5|u_6), F_{7|6}(u_7|u_6))$:
$F_{5|6}(u_5|u_6) = \int_0^1\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3, u_6)c_{1,6|3}(v_1, F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3), F_{6|1,3}(u_6|v_1, v_3))$
$c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)h2_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))dv_1dv_2dv_3dv_4$
$F_{7|6}(u_7|u_6) = h2_{6,7}(u_6, u_7)$
ReflectU Gumbel(theta=1.20971)

$4->7 : c_{4,7|5,6}(F_{4|5,6}(u_4|u_5, u_6), F_{7|5,6}(u_7|u_5, u_6))$:
$F_{4|5,6}(u_4|u_5, u_6) = (\int_0^{u_4}\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3, u_6)c_{1,6|3}(v_1, F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),$
$F_{6|1,3}(u_6|v_1, v_3))c_{4,5}(v_4, u_5)c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_1dv_2dv_3dv_4)$
$/(\int_0^1\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3, u_6)c_{1,6|3}(v_1, F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3), F_{6|1,3}(u_6|v_1, v_3))c_{4,5}(v_4, u_5)$
$c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_1dv_2dv_3dv_4)$
$F_{7|5,6}(u_7|u_5, u_6) = h2_{5,7|6}(F_{5|6}(u_5|u_6), F_{7|6}(u_7|u_6))$
Frank(theta=2.1038)

$1->7 : c_{1,7|4,5,6}(F_{1|4,5,6}(u_1|u_4, u_5, u_6), F_{7|4,5,6}(u_7|u_4, u_5, u_6))$:
$F_{1|4,5,6}(u_1|u_4, u_5, u_6) = (\int_0^{u_1}\int_0^1\int_0^1 c_{3,6}(v_3, u_6)c_{1,6|3}(v_1, F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),$
$F_{6|1,3}(u_6|v_1, v_3))c_{2,5|4}(F_{2|4}(v_2|u_4), F_{5|4}(u_5|u_4))c_{3,4}(v_3, u_4)c_{2,3}(v_2, v_3)dv_2dv_3dv_1)$
$/(\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3, u_6)c_{1,6|3}(v_1, F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3), F_{6|1,3}(u_6|v_1, v_3))$
$c_{2,5|4}(F_{2|4}(v_2|u_4), F_{5|4}(u_5|u_4))c_{3,4}(v_3, u_4)c_{2,3}(v_2, v_3)dv_1dv_2dv_3)$
$F_{7|4,5,6}(u_7|u_4, u_5, u_6) = h2_{4,7|5,6}(F_{4|5,6}(u_4|u_5, u_6), F_{7|5,6}(u_7|u_5, u_6))$
Student(theta=0.602655,nu=12.2221)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

PCBN 2

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2,u_3)c_{3,4}(u_3,u_4)c_{4,5}(u_4,u_5)c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))c_{1,6}(u_1,u_6) \\ & c_{3,6|1}(u_3,F_{6|1}(u_6|u_1))c_{2,6|1,3}(F_{2|3}(u_2|u_3),F_{6|1,3}(u_6|u_1,u_3))c_{4,7}(u_4,u_7) \\ & c_{1,7|4}(u_1,F_{7|4}(u_7|u_4))c_{5,7|1,4}(F_{5|4}(u_5|u_4),F_{7|1,4}(u_7|u_1,u_4)) \\ & c_{6,7|1,4,5}(F_{6|1,4,5}(u_6|u_1,u_4,u_5),F_{7|1,4,5}(u_7|u_1,u_4,u_5)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3 : c_{2,3}(u_2,u_3)$:
Frank(theta=5.50008)

$3->4 : c_{3,4}(u_3,u_4)$:
Frank(theta=1.29653)

$4->5 : c_{4,5}(u_4,u_5)$:
ReflectUV Joe(theta=1.23974)

$2->5 : c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))$:
$F_{2|4}(u_2|u_4) = \int_0^1 c_{3,4}(v_3,u_4)h1_{2,3}(u_2,v_3)dv_3$
$F_{5|4}(u_5|u_4) = h2_{4,5}(u_4,u_5)$
Frank(theta=1.6798)

$1->6 : c_{1,6}(u_1,u_6)$:
Frank(theta=3.22223)

$3->6 : c_{3,6|1}(u_3,F_{6|1}(u_6|u_1))$:
$F_{6|1}(u_6|u_1) = h2_{1,6}(u_1,u_6)$
ReflectUV Clayton(theta=0.306133)

$2->6 : c_{2,6|1,3}(F_{2|3}(u_2|u_3),F_{6|1,3}(u_6|u_1,u_3))$:
$F_{2|3}(u_2|u_3) = h1_{2,3}(u_2,u_3)$
$F_{6|1,3}(u_6|u_1,u_3) = h2_{3,6|1}(u_3,F_{6|1}(u_6|u_1))$
Joe(theta=1.56066)

$4->7 : c_{4,7}(u_4,u_7)$:
ReflectUV Gumbel(theta=1.17536)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

$1->7 : c_{1,7|4}(u_1, F_{7|4}(u_7|u_4))$:
$F_{7|4}(u_7|u_4) = h2_{4,7}(u_4, u_7)$
Student(theta=0.70732,nu=10.8613)

$5->7 : c_{5,7|1,4}(F_{5|4}(u_5|u_4), F_{7|1,4}(u_7|u_1, u_4))$:
$F_{5|4}(u_5|u_4) = h2_{4,5}(u_4, u_5)$
$F_{7|1,4}(u_7|u_1, u_4) = h2_{1,7|4}(u_1, F_{7|4}(u_7|u_4))$
ReflectV Clayton(theta=0.363444)

$6->7 : c_{6,7|1,4,5}(F_{6|1,4,5}(u_6|u_1, u_4, u_5), F_{7|1,4,5}(u_7|u_1, u_4, u_5))$:
$F_{6|1,4,5}(u_6|u_1, u_4, u_5) = \int_0^1 \int_0^1 c_{2,5|4}(F_{2|4}(v_2|u_4), F_{5|4}(u_5|u_4))c_{3,4}(v_3, u_4)c_{2,3}(v_2, v_3)$
$h2_{2,6|1,3}(F_{2|3}(v_2|v_3), F_{6|1,3}(u_6|u_1, v_3))dv_2dv_3$
$F_{7|1,4,5}(u_7|u_1, u_4, u_5) = h2_{5,7|1,4}(F_{5|4}(u_5|u_4), F_{7|1,4}(u_7|u_1, u_4))$
Student(theta=0.493726,nu=5.44725)

PCBN 3

Note that all univariate marginals are uniform distributions on $[0, 1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0, 1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2, u_3)c_{3,4}(u_3, u_4)c_{4,5}(u_4, u_5)c_{2,5|4}(F_{2|4}(u_2|u_4), F_{5|4}(u_5|u_4))c_{1,6}(u_1, u_6) \\ & c_{2,6|1}(u_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(u_3|u_2), F_{6|1,2}(u_6|u_1, u_2))c_{6,7}(u_6, u_7) \\ & c_{5,7|6}(F_{5|6}(u_5|u_6), F_{7|6}(u_7|u_6))c_{1,7|5,6}(F_{1|5,6}(u_1|u_5, u_6), F_{7|5,6}(u_7|u_5, u_6)) \\ & c_{4,7|1,5,6}(F_{4|1,5,6}(u_4|u_1, u_5, u_6), F_{7|1,5,6}(u_7|u_1, u_5, u_6)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3 : c_{2,3}(u_2, u_3)$:
Frank(theta=5.50008)

$3->4 : c_{3,4}(u_3, u_4)$:
Frank(theta=1.29653)

$4->5 : c_{4,5}(u_4, u_5)$:
ReflectUV Joe(theta=1.23974)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

$2->5 : c_{2,5|4}(F_{2|4}(u_2|u_4), F_{5|4}(u_5|u_4))$:
$F_{2|4}(u_2|u_4) = \int_0^1 c_{3,4}(v_3, u_4)h1_{2,3}(u_2, v_3)dv_3$
$F_{5|4}(u_5|u_4) = h2_{4,5}(u_4, u_5)$
Frank(theta=1.6798)

$1->6 : c_{1,6}(u_1, u_6)$:
Frank(theta=3.22223)

$2->6 : c_{2,6|1}(u_2, F_{6|1}(u_6|u_1))$:
$F_{6|1}(u_6|u_1) = h2_{1,6}(u_1, u_6)$
ReflectUV Clayton(theta=0.818302)

$3->6 : c_{3,6|1,2}(F_{3|2}(u_3|u_2), F_{6|1,2}(u_6|u_1, u_2))$:
$F_{3|2}(u_3|u_2) = h2_{2,3}(u_2, u_3)$
$F_{6|1,2}(u_6|u_1, u_2) = h2_{2,6|1}(u_2, F_{6|1}(u_6|u_1))$
ReflectU Clayton(theta=0.154473)

$6->7 : c_{6,7}(u_6, u_7)$:
Gauss(theta=0.645648)

$5->7 : c_{5,7|6}(F_{5|6}(u_5|u_6), F_{7|6}(u_7|u_6))$:
$F_{5|6}(u_5|u_6) = \int_0^1\int_0^1\int_0^1\int_0^1 c_{1,6}(v_1, u_6)c_{2,6|1}(v_2, F_{6|1}(u_6|v_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|v_1, v_2))$
$c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)h2_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))dv_1dv_2dv_3dv_4$
$F_{7|6}(u_7|u_6) = h2_{6,7}(u_6, u_7)$
ReflectU Gumbel(theta=1.21021)

$1->7 : c_{1,7|5,6}(F_{1|5,6}(u_1|u_5, u_6), F_{7|5,6}(u_7|u_5, u_6))$:
$F_{1|5,6}(u_1|u_5, u_6) = (\int_0^{u_1}\int_0^1\int_0^1\int_0^1 c_{1,6}(v_1, u_6)c_{2,6|1}(v_2, F_{6|1}(u_6|v_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2),$
$F_{6|1,2}(u_6|v_1, v_2))c_{4,5}(v_4, u_5)c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_2dv_3dv_4dv_1)$
$/(\int_0^1\int_0^1\int_0^1\int_0^1 c_{1,6}(v_1, u_6)c_{2,6|1}(v_2, F_{6|1}(u_6|v_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|v_1, v_2))c_{4,5}(v_4, u_5)$
$c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_1dv_2dv_3dv_4)$
$F_{7|5,6}(u_7|u_5, u_6) = h2_{5,7|6}(F_{5|6}(u_5|u_6), F_{7|6}(u_7|u_6))$
Student(theta=0.592164,nu=9.61806)

$4->7 : c_{4,7|1,5,6}(F_{4|1,5,6}(u_4|u_1, u_5, u_6), F_{7|1,5,6}(u_7|u_1, u_5, u_6))$:
$F_{4|1,5,6}(u_4|u_1, u_5, u_6) = (\int_0^{u_4}\int_0^1\int_0^1 c_{2,6|1}(v_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|u_1, v_2))$
$c_{4,5}(v_4, u_5)c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_2dv_3dv_4)$
$/(\int_0^1\int_0^1\int_0^1 c_{2,6|1}(v_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|u_1, v_2))c_{4,5}(v_4, u_5)$
$c_{2,5|4}(F_{2|4}(v_2|v_4), F_{5|4}(u_5|v_4))c_{3,4}(v_3, v_4)c_{2,3}(v_2, v_3)dv_2dv_3dv_4)$
$F_{7|1,5,6}(u_7|u_1, u_5, u_6) = h2_{1,7|5,6}(F_{1|5,6}(u_1|u_5, u_6), F_{7|5,6}(u_7|u_5, u_6))$
Frank(theta=2.28831)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

PCBN 4

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2,u_3)c_{3,4}(u_3,u_4)c_{4,5}(u_4,u_5)c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))c_{3,6}(u_3,u_6) \\ & c_{1,6|3}(u_1,F_{6|3}(u_6|u_3))c_{2,6|1,3}(F_{2|3}(u_2|u_3),F_{6|1,3}(u_6|u_1,u_3))c_{5,7}(u_5,u_7) \\ & c_{6,7|5}(F_{6|5}(u_6|u_5),F_{7|5}(u_7|u_5))c_{1,7|5,6}(F_{1|5,6}(u_1|u_5,u_6),F_{7|5,6}(u_7|u_5,u_6)) \\ & c_{4,7|1,5,6}(F_{4|1,5,6}(u_4|u_1,u_5,u_6),F_{7|1,5,6}(u_7|u_1,u_5,u_6)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3 : c_{2,3}(u_2,u_3)$:
Frank(theta=5.50008)

$3->4 : c_{3,4}(u_3,u_4)$:
Frank(theta=1.29653)

$4->5 : c_{4,5}(u_4,u_5)$:
ReflectUV Joe(theta=1.23974)

$2->5 : c_{2,5|4}(F_{2|4}(u_2|u_4),F_{5|4}(u_5|u_4))$:
$F_{2|4}(u_2|u_4) = \int_0^1 c_{3,4}(v_3,u_4)h1_{2,3}(u_2,v_3)dv_3$
$F_{5|4}(u_5|u_4) = h2_{4,5}(u_4,u_5)$
Frank(theta=1.6798)

$3->6 : c_{3,6}(u_3,u_6)$:
ReflectUV Clayton(theta=0.328713)

$1->6 : c_{1,6|3}(u_1,F_{6|3}(u_6|u_3))$:
$F_{6|3}(u_6|u_3) = h2_{3,6}(u_3,u_6)$
Frank(theta=3.11038)

$2->6 : c_{2,6|1,3}(F_{2|3}(u_2|u_3),F_{6|1,3}(u_6|u_1,u_3))$:
$F_{2|3}(u_2|u_3) = h1_{2,3}(u_2,u_3)$
$F_{6|1,3}(u_6|u_1,u_3) = h2_{1,6|3}(u_1,F_{6|3}(u_6|u_3))$
ReflectUV Clayton(theta=0.713207)

$5->7 : c_{5,7}(u_5,u_7)$:
ReflectV Clayton(theta=0.379217)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

$6->7: c_{6,7|5}(F_{6|5}(u_6|u_5), F_{7|5}(u_7|u_5))$:
$F_{6|5}(u_6|u_5) = \int_0^1\int_0^1\int_0^1\int_0^1 c_{4,5}(v_4,u_5)c_{2,5|4}(F_{2|4}(v_2|v_4),F_{5|4}(u_5|v_4))c_{3,4}(v_3,v_4)c_{2,3}(v_2,v_3)$
$h2_{2,6|1,3}(F_{2|3}(v_2|v_3),F_{6|1,3}(u_6|v_1,v_3))dv_1dv_2dv_3dv_4$
$F_{7|5}(u_7|u_5) = h2_{5,7}(u_5,u_7)$
Student(theta=0.639165,nu=7.16501)

$1->7: c_{1,7|5,6}(F_{1|5,6}(u_1|u_5,u_6), F_{7|5,6}(u_7|u_5,u_6))$:
$F_{1|5,6}(u_1|u_5,u_6) = (\int_0^{u_1}\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3,u_6)c_{1,6|3}(v_1,F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),$
$F_{6|1,3}(u_6|v_1,v_3))c_{4,5}(v_4,u_5)c_{2,5|4}(F_{2|4}(v_2|v_4),F_{5|4}(u_5|v_4))c_{3,4}(v_3,v_4)c_{2,3}(v_2,v_3)dv_2dv_3dv_4dv_1)$
$/(\int_0^1\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3,u_6)c_{1,6|3}(v_1,F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),F_{6|1,3}(u_6|v_1,v_3))c_{4,5}(v_4,u_5)$
$c_{2,5|4}(F_{2|4}(v_2|v_4),F_{5|4}(u_5|v_4))c_{3,4}(v_3,v_4)c_{2,3}(v_2,v_3)dv_1dv_2dv_3dv_4)$
$F_{7|5,6}(u_7|u_5,u_6) = h2_{6,7|5}(F_{6|5}(u_6|u_5),F_{7|5}(u_7|u_5))$
Student(theta=0.572066,nu=6.9782)

$4->7: c_{4,7|1,5,6}(F_{4|1,5,6}(u_4|u_1,u_5,u_6), F_{7|1,5,6}(u_7|u_1,u_5,u_6))$:
$F_{4|1,5,6}(u_4|u_1,u_5,u_6) = (\int_0^{u_4}\int_0^1\int_0^1 c_{3,6}(v_3,u_6)c_{1,6|3}(u_1,F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),$
$F_{6|1,3}(u_6|u_1,v_3))c_{4,5}(v_4,u_5)c_{2,5|4}(F_{2|4}(v_2|v_4),F_{5|4}(u_5|v_4))c_{3,4}(v_3,v_4)c_{2,3}(v_2,v_3)dv_2dv_3dv_4)$
$/(\int_0^1\int_0^1\int_0^1 c_{3,6}(v_3,u_6)c_{1,6|3}(u_1,F_{6|3}(u_6|v_3))c_{2,6|1,3}(F_{2|3}(v_2|v_3),F_{6|1,3}(u_6|u_1,v_3))c_{4,5}(v_4,u_5)$
$c_{2,5|4}(F_{2|4}(v_2|v_4),F_{5|4}(u_5|v_4))c_{3,4}(v_3,v_4)c_{2,3}(v_2,v_3)dv_2dv_3dv_4)$
$F_{7|1,5,6}(u_7|u_1,u_5,u_6) = h2_{1,7|5,6}(F_{1|5,6}(u_1|u_5,u_6),F_{7|5,6}(u_7|u_5,u_6))$
Frank(theta=1.80208)

PCBN 5

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2,u_3)c_{3,4}(u_3,u_4)c_{2,5}(u_2,u_5)c_{4,5|2}(F_{4|2}(u_4|u_2),F_{5|2}(u_5|u_2))c_{1,6}(u_1,u_6)\\ & c_{2,6|1}(u_2,F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(u_3|u_2),F_{6|1,2}(u_6|u_1,u_2))c_{1,7}(u_1,u_7)\\ & c_{6,7|1}(F_{6|1}(u_6|u_1),F_{7|1}(u_7|u_1))c_{4,7|1,6}(F_{4|1,6}(u_4|u_1,u_6),F_{7|1,6}(u_7|u_1,u_6))\\ & c_{5,7|1,4,6}(F_{5|1,4,6}(u_5|u_1,u_4,u_6),F_{7|1,4,6}(u_7|u_1,u_4,u_6)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3: c_{2,3}(u_2,u_3)$:
ReflectUV Joe(theta=1.03637)

$3->4: c_{3,4}(u_3,u_4)$:
Student(theta=-0.156571,nu=5.46424)

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)

$2->5 : c_{2,5}(u_2, u_5)$:
ReflectUV Gumbel(theta=1.1272)

$4->5 : c_{4,5|2}(F_{4|2}(u_4|u_2), F_{5|2}(u_5|u_2))$:
$F_{4|2}(u_4|u_2) = \int_0^1 c_{2,3}(u_2, v_3)h2_{3,4}(v_3, u_4)dv_3$
$F_{5|2}(u_5|u_2) = h2_{2,5}(u_2, u_5)$
ReflectUV Gumbel(theta=1.20495)

$1->6 : c_{1,6}(u_1, u_6)$:
Student(theta=0.139079,nu=4.66412)

$2->6 : c_{2,6|1}(u_2, F_{6|1}(u_6|u_1))$:
$F_{6|1}(u_6|u_1) = h2_{1,6}(u_1, u_6)$
ReflectUV Gumbel(theta=1.11314)

$3->6 : c_{3,6|1,2}(F_{3|2}(u_3|u_2), F_{6|1,2}(u_6|u_1, u_2))$:
$F_{3|2}(u_3|u_2) = h2_{2,3}(u_2, u_3)$
$F_{6|1,2}(u_6|u_1, u_2) = h2_{2,6|1}(u_2, F_{6|1}(u_6|u_1))$
Student(theta=0.719961,nu=12.5461)

$1->7 : c_{1,7}(u_1, u_7)$:
Frank(theta=1.26203)

$6->7 : c_{6,7|1}(F_{6|1}(u_6|u_1), F_{7|1}(u_7|u_1))$:
$F_{6|1}(u_6|u_1) = h2_{1,6}(u_1, u_6)$
$F_{7|1}(u_7|u_1) = h2_{1,7}(u_1, u_7)$
Student(theta=0.606938,nu=6.17406)

$4->7 : c_{4,7|1,6}(F_{4|1,6}(u_4|u_1, u_6), F_{7|1,6}(u_7|u_1, u_6))$:
$F_{4|1,6}(u_4|u_1, u_6) = \int_0^1 \int_0^1 c_{2,6|1}(v_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|u_1, v_2))c_{2,3}(v_2, v_3)$
$h2_{3,4}(v_3, u_4)dv_2dv_3$
$F_{7|1,6}(u_7|u_1, u_6) = h2_{6,7|1}(F_{6|1}(u_6|u_1), F_{7|1}(u_7|u_1))$
$Clayton(theta = 0.184905)$

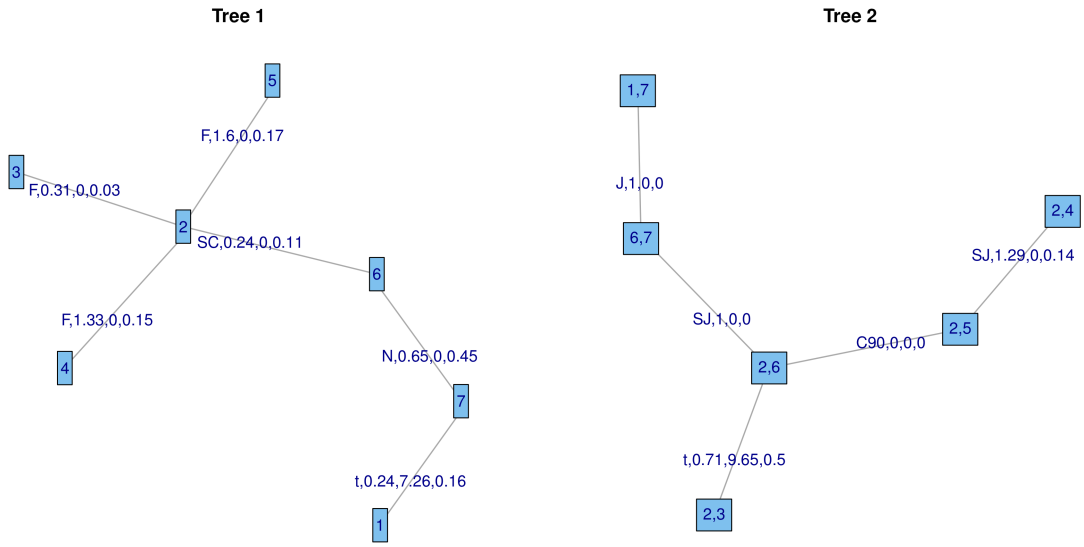
$5->7 : c_{5,7|1,4,6}(F_{5|1,4,6}(u_5|u_1, u_4, u_6), F_{7|1,4,6}(u_7|u_1, u_4, u_6))$:
$F_{5|1,4,6}(u_5|u_1, u_4, u_6) = (\int_0^1 \int_0^1 c_{2,6|1}(v_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|u_1, v_2))$
$c_{3,4}(v_3, u_4)c_{2,3}(v_2, v_3)h2_{4,5|2}(F_{4|2}(u_4|v_2), F_{5|2}(u_5|v_2))dv_2dv_3)$
$/(\int_0^1 \int_0^1 c_{2,6|1}(v_2, F_{6|1}(u_6|u_1))c_{3,6|1,2}(F_{3|2}(v_3|v_2), F_{6|1,2}(u_6|u_1, v_2))c_{3,4}(v_3, u_4)c_{2,3}(v_2, v_3)dv_2dv_3)$
$F_{7|1,4,6}(u_7|u_1, u_4, u_6) = h2_{4,7|1,6}(F_{4|1,6}(u_4|u_1, u_6), F_{7|1,6}(u_7|u_1, u_6))$
Joe(theta=1.1948)

Appendix C

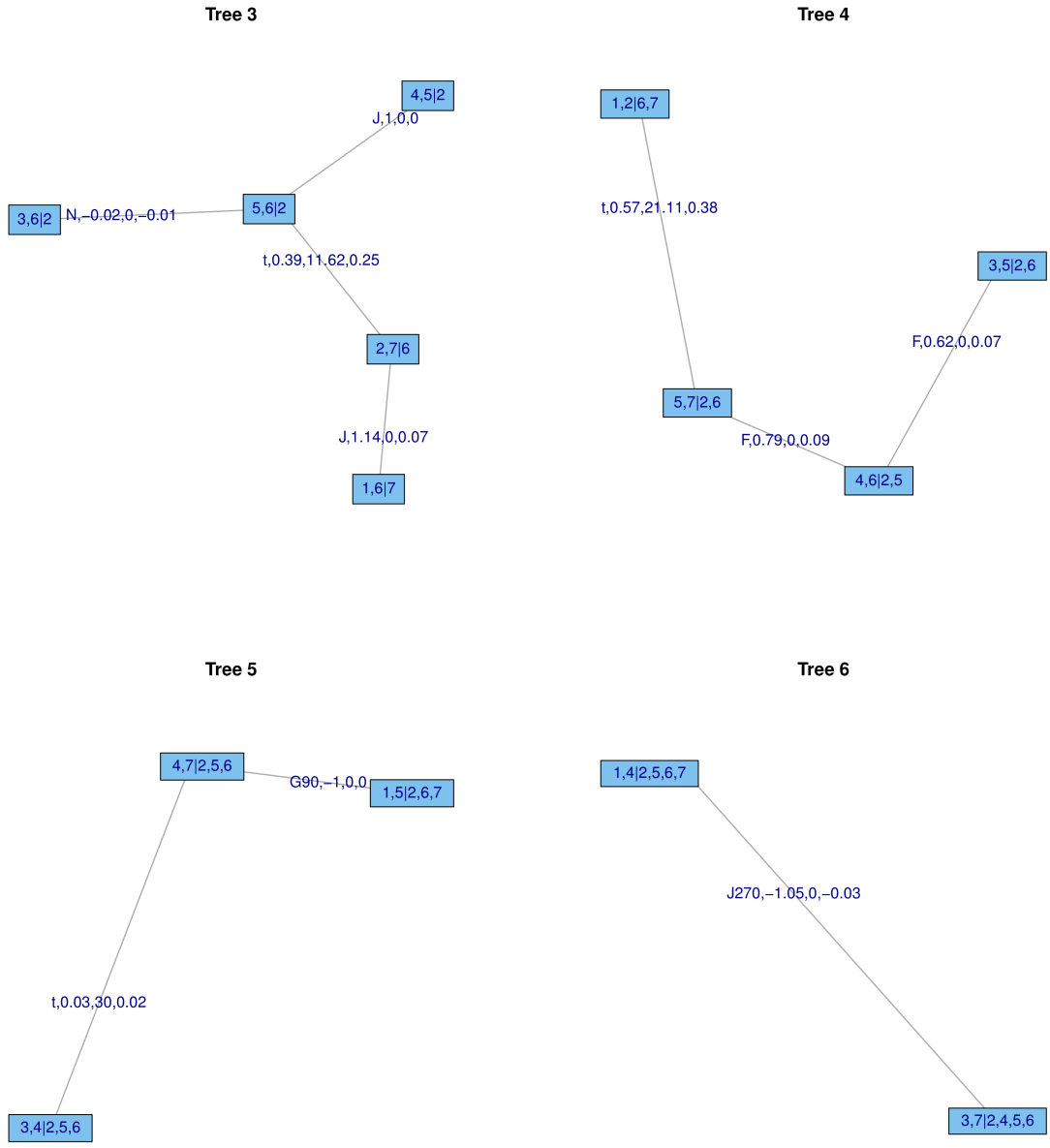
Uranium dataset: RVine tree plots

The following tree plots show the structure as well as the copula families and parameters of the RVine fitted on the uranium dataset. "F,1.6,0,0.17" means a Frank Copula with par1 = 1.6, par2 = 0 (not used) and $\tau = 0.17$.

Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)



Indices: (1 = Co, 2 = Cs, 3 = U, 4 = Li, 5 = K, 6 = Ti, 7 = Sc)



Appendix D

Wine dataset: all parent orderings

This table shows the log-likelihoods and its calculation time of all possible parent orderings for the DAG we used for the wine dataset.

The columns 'sf-1', 'sf-2' represent the parents of sf in the order that sf-1 is the lowest parent. Same applies for the columns of the parents of den, acv and acc.

LL	Time	sf-1	sf-2	den-1	den-2	acv-1	acv-2	acc-1	acc-2	acc-3	acc-4
818.09	0.0080s	acf	st	acf	st	sp	st	acf	acv	ph	sp
817.45	0.1120s	acf	st	acf	st	st	sp	acf	acv	ph	sp
817.71	0.0080s	acf	st	acf	st	sp	st	acf	acv	sp	ph
817.07	0.0980s	acf	st	acf	st	st	sp	acf	acv	sp	ph
816.19	0.0070s	acf	st	acf	st	sp	st	acf	ph	acv	sp
815.55	0.1130s	acf	st	acf	st	st	sp	acf	ph	acv	sp
808.41	0.0070s	acf	st	acf	st	sp	st	acf	ph	sp	acv
807.76	0.0690s	acf	st	acf	st	st	sp	acf	ph	sp	acv
809.00	0.0070s	acf	st	acf	st	sp	st	acf	sp	acv	ph
808.36	0.0820s	acf	st	acf	st	st	sp	acf	sp	acv	ph
807.86	0.0060s	acf	st	acf	st	sp	st	acf	sp	ph	acv
807.21	0.1200s	acf	st	acf	st	st	sp	acf	sp	ph	acv
815.03	0.0110s	acf	st	acf	st	sp	st	acv	acf	ph	sp
814.38	0.0840s	acf	st	acf	st	st	sp	acv	acf	ph	sp
814.91	0.0110s	acf	st	acf	st	sp	st	acv	acf	sp	ph
814.25	0.1340s	acf	st	acf	st	st	sp	acv	acf	sp	ph
815.77	0.0150s	acf	st	acf	st	sp	st	acv	ph	acf	sp
815.11	0.1030s	acf	st	acf	st	st	sp	acv	ph	acf	sp
810.55	0.0140s	acf	st	acf	st	sp	st	acv	ph	sp	acf
809.89	0.1320s	acf	st	acf	st	st	sp	acv	ph	sp	acf

LL	Time	sf-1	sf-2	den-1	den-2	acv-1	acv-2	acc-1	acc-2	acc-3	acc-4
814.73	0.0110s	acf	st	acf	st	sp	st	acv	sp	acf	ph
814.08	0.1520s	acf	st	acf	st	st	sp	acv	sp	acf	ph
813.99	0.0100s	acf	st	acf	st	sp	st	acv	sp	ph	acf
813.36	0.1430s	acf	st	acf	st	st	sp	acv	sp	ph	acf
816.56	0.0050s	acf	st	acf	st	sp	st	ph	acf	acv	sp
815.95	0.1320s	acf	st	acf	st	st	sp	ph	acf	acv	sp
807.79	0.0070s	acf	st	acf	st	sp	st	ph	acf	sp	acv
807.15	0.1160s	acf	st	acf	st	st	sp	ph	acf	sp	acv
814.40	0.0110s	acf	st	acf	st	sp	st	ph	acv	acf	sp
813.77	0.0910s	acf	st	acf	st	st	sp	ph	acv	acf	sp
808.18	0.0110s	acf	st	acf	st	sp	st	ph	acv	sp	acf
807.54	0.1260s	acf	st	acf	st	st	sp	ph	acv	sp	acf
806.19	0.0110s	acf	st	acf	st	sp	st	ph	sp	acf	acv
805.52	0.0690s	acf	st	acf	st	st	sp	ph	sp	acf	acv
804.07	0.0100s	acf	st	acf	st	sp	st	ph	sp	acv	acf
803.40	0.1230s	acf	st	acf	st	st	sp	ph	sp	acv	acf
806.60	0.0110s	acf	st	acf	st	sp	st	sp	acf	acv	ph
805.89	0.0730s	acf	st	acf	st	st	sp	sp	acf	acv	ph
805.18	0.0110s	acf	st	acf	st	sp	st	sp	acf	ph	acv
804.47	0.1240s	acf	st	acf	st	st	sp	sp	acf	ph	acv
802.95	0.0110s	acf	st	acf	st	sp	st	sp	acv	acf	ph
802.23	0.1190s	acf	st	acf	st	st	sp	sp	acv	acf	ph
802.13	0.0130s	acf	st	acf	st	sp	st	sp	acv	ph	acf
801.41	0.0870s	acf	st	acf	st	st	sp	sp	acv	ph	acf
802.65	0.0110s	acf	st	acf	st	sp	st	sp	ph	acf	acv
801.94	0.0770s	acf	st	acf	st	st	sp	sp	ph	acf	acv
800.89	0.0100s	acf	st	acf	st	sp	st	sp	ph	acv	acf
800.18	0.1190s	acf	st	acf	st	st	sp	sp	ph	acv	acf
817.51	0.0070s	acf	st	st	acf	sp	st	acf	acv	ph	sp
816.87	0.1070s	acf	st	st	acf	st	sp	acf	acv	ph	sp
817.13	0.0080s	acf	st	st	acf	sp	st	acf	acv	sp	ph
816.49	0.1250s	acf	st	st	acf	st	sp	acf	acv	sp	ph
815.61	0.0070s	acf	st	st	acf	sp	st	acf	ph	acv	sp
814.97	0.1060s	acf	st	st	acf	st	sp	acf	ph	acv	sp
807.83	0.0060s	acf	st	st	acf	sp	st	acf	ph	sp	acv
807.18	0.0900s	acf	st	st	acf	st	sp	acf	ph	sp	acv
808.42	0.0070s	acf	st	st	acf	sp	st	acf	sp	acv	ph
807.77	0.1090s	acf	st	st	acf	st	sp	acf	sp	acv	ph
807.28	0.0070s	acf	st	st	acf	sp	st	acf	sp	ph	acv
806.63	0.1210s	acf	st	st	acf	st	sp	acf	sp	ph	acv
814.45	0.0110s	acf	st	st	acf	sp	st	acv	acf	ph	sp
813.79	0.0860s	acf	st	st	acf	st	sp	acv	acf	ph	sp
814.32	0.0110s	acf	st	st	acf	sp	st	acv	acf	sp	ph
813.67	0.0770s	acf	st	st	acf	st	sp	acv	acf	sp	ph
815.19	0.0140s	acf	st	st	acf	sp	st	acv	ph	acf	sp

LL	Time	sf-1	sf-2	den-1	den-2	acv-1	acv-2	acc-1	acc-2	acc-3	acc-4
814.53	0.0830s	acf	st	st	acf	st	sp	acv	ph	acf	sp
809.97	0.0140s	acf	st	st	acf	sp	st	acv	ph	sp	acf
809.31	0.1070s	acf	st	st	acf	st	sp	acv	ph	sp	acf
814.15	0.0110s	acf	st	st	acf	sp	st	acv	sp	acf	ph
813.49	0.0900s	acf	st	st	acf	st	sp	acv	sp	acf	ph
813.41	0.0100s	acf	st	st	acf	sp	st	acv	sp	ph	acf
812.77	0.1340s	acf	st	st	acf	st	sp	acv	sp	ph	acf
815.98	0.0080s	acf	st	st	acf	sp	st	ph	acf	acv	sp
815.36	0.1360s	acf	st	st	acf	st	sp	ph	acf	acv	sp
807.21	0.0070s	acf	st	st	acf	sp	st	ph	acf	sp	acv
806.57	0.0760s	acf	st	st	acf	st	sp	ph	acf	sp	acv
813.82	0.0110s	acf	st	st	acf	sp	st	ph	acv	acf	sp
813.19	0.0980s	acf	st	st	acf	st	sp	ph	acv	acf	sp
807.60	0.0100s	acf	st	st	acf	sp	st	ph	acv	sp	acf
806.95	0.1430s	acf	st	st	acf	st	sp	ph	acv	sp	acf
805.61	0.0110s	acf	st	st	acf	sp	st	ph	sp	acf	acv
804.93	0.0960s	acf	st	st	acf	st	sp	ph	sp	acf	acv
803.49	0.0100s	acf	st	st	acf	sp	st	ph	sp	acv	acf
802.81	0.0810s	acf	st	st	acf	st	sp	ph	sp	acv	acf
806.02	0.0110s	acf	st	st	acf	sp	st	sp	acf	acv	ph
805.31	0.0670s	acf	st	st	acf	st	sp	sp	acf	acv	ph
804.60	0.0110s	acf	st	st	acf	sp	st	sp	acf	ph	acv
803.89	0.1250s	acf	st	st	acf	st	sp	sp	acf	ph	acv
802.36	0.0100s	acf	st	st	acf	sp	st	sp	acv	acf	ph
801.65	0.1250s	acf	st	st	acf	st	sp	sp	acv	acf	ph
801.54	0.0140s	acf	st	st	acf	sp	st	sp	acv	ph	acf
800.83	0.1190s	acf	st	st	acf	st	sp	sp	acv	ph	acf
802.07	0.0110s	acf	st	st	acf	sp	st	sp	ph	acf	acv
801.36	0.1070s	acf	st	st	acf	st	sp	sp	ph	acf	acv
800.31	0.0090s	acf	st	st	acf	sp	st	sp	ph	acv	acf
799.60	0.0810s	acf	st	st	acf	st	sp	sp	ph	acv	acf
822.31	0.0070s	st	acf	acf	st	sp	st	acf	acv	ph	sp
821.67	0.0940s	st	acf	acf	st	st	sp	acf	acv	ph	sp
821.93	0.0080s	st	acf	acf	st	sp	st	acf	acv	sp	ph
821.29	0.1010s	st	acf	acf	st	st	sp	acf	acv	sp	ph
820.41	0.0080s	st	acf	acf	st	sp	st	acf	ph	acv	sp
819.77	0.1260s	st	acf	acf	st	st	sp	acf	ph	acv	sp
812.63	0.0070s	st	acf	acf	st	sp	st	acf	ph	sp	acv
811.98	0.0680s	st	acf	acf	st	st	sp	acf	ph	sp	acv
813.22	0.0070s	st	acf	acf	st	sp	st	acf	sp	acv	ph
812.57	0.0900s	st	acf	acf	st	st	sp	acf	sp	acv	ph
812.08	0.0070s	st	acf	acf	st	sp	st	acf	sp	ph	acv
811.43	0.1100s	st	acf	acf	st	st	sp	acf	sp	ph	acv
819.25	0.0110s	st	acf	acf	st	sp	st	acv	acf	ph	sp
818.60	0.0800s	st	acf	acf	st	st	sp	acv	acf	ph	sp

LL	Time	sf-1	sf-2	den-1	den-2	acv-1	acv-2	acc-1	acc-2	acc-3	acc-4
819.13	0.0110s	st	acf	acf	st	sp	st	acv	acf	sp	ph
818.47	0.0710s	st	acf	acf	st	st	sp	acv	acf	sp	ph
819.99	0.0140s	st	acf	acf	st	sp	st	acv	ph	acf	sp
819.33	0.1360s	st	acf	acf	st	st	sp	acv	ph	acf	sp
814.77	0.0140s	st	acf	acf	st	sp	st	acv	ph	sp	acf
814.11	0.0860s	st	acf	acf	st	st	sp	acv	ph	sp	acf
818.95	0.0110s	st	acf	acf	st	sp	st	acv	sp	acf	ph
818.30	0.1280s	st	acf	acf	st	st	sp	acv	sp	acf	ph
818.21	0.0100s	st	acf	acf	st	sp	st	acv	sp	ph	acf
817.57	0.1360s	st	acf	acf	st	st	sp	acv	sp	ph	acf
820.78	0.0070s	st	acf	acf	st	sp	st	ph	acf	acv	sp
820.17	0.1210s	st	acf	acf	st	st	sp	ph	acf	acv	sp
812.01	0.0070s	st	acf	acf	st	sp	st	ph	acf	sp	acv
811.37	0.1210s	st	acf	acf	st	st	sp	ph	acf	sp	acv
818.62	0.0110s	st	acf	acf	st	sp	st	ph	acv	acf	sp
817.99	0.1110s	st	acf	acf	st	st	sp	ph	acv	acf	sp
812.40	0.0100s	st	acf	acf	st	sp	st	ph	acv	sp	acf
811.76	0.1320s	st	acf	acf	st	st	sp	ph	acv	sp	acf
810.41	0.0110s	st	acf	acf	st	sp	st	ph	sp	acf	acv
809.74	0.0850s	st	acf	acf	st	st	sp	ph	sp	acf	acv
808.29	0.0100s	st	acf	acf	st	sp	st	ph	sp	acv	acf
807.62	0.0660s	st	acf	acf	st	st	sp	ph	sp	acv	acf
810.82	0.0110s	st	acf	acf	st	sp	st	sp	acf	acv	ph
810.11	0.1160s	st	acf	acf	st	st	sp	sp	acf	acv	ph
809.40	0.0110s	st	acf	acf	st	sp	st	sp	acf	ph	acv
808.69	0.0850s	st	acf	acf	st	st	sp	sp	acf	ph	acv
807.17	0.0090s	st	acf	acf	st	sp	st	sp	acv	acf	ph
806.45	0.1130s	st	acf	acf	st	st	sp	sp	acv	acf	ph
806.35	0.0130s	st	acf	acf	st	sp	st	sp	acv	ph	acf
805.63	0.0820s	st	acf	acf	st	st	sp	sp	acv	ph	acf
806.87	0.0110s	st	acf	acf	st	sp	st	sp	ph	acf	acv
806.16	0.0840s	st	acf	acf	st	st	sp	sp	ph	acf	acv
805.11	0.0100s	st	acf	acf	st	sp	st	sp	ph	acv	acf
804.40	0.0680s	st	acf	acf	st	st	sp	sp	ph	acv	acf
821.72	0.0080s	st	acf	st	acf	sp	st	acf	acv	ph	sp
821.09	0.0910s	st	acf	st	acf	st	sp	acf	acv	ph	sp
821.35	0.0080s	st	acf	st	acf	sp	st	acf	acv	sp	ph
820.71	0.1290s	st	acf	st	acf	st	sp	acf	acv	sp	ph
819.83	0.0070s	st	acf	st	acf	sp	st	acf	ph	acv	sp
819.18	0.1400s	st	acf	st	acf	st	sp	acf	ph	acv	sp
812.05	0.0080s	st	acf	st	acf	sp	st	acf	ph	sp	acv
811.40	0.0910s	st	acf	st	acf	st	sp	acf	ph	sp	acv
812.64	0.0080s	st	acf	st	acf	sp	st	acf	sp	acv	ph
811.99	0.1010s	st	acf	st	acf	st	sp	acf	sp	acv	ph
811.50	0.0080s	st	acf	st	acf	sp	st	acf	sp	ph	acv

LL	Time	sf-1	sf-2	den-1	den-2	acv-1	acv-2	acc-1	acc-2	acc-3	acc-4
810.85	0.1210s	st	acf	st	acf	st	sp	acf	sp	ph	acv
818.67	0.0110s	st	acf	st	acf	sp	st	acv	acf	ph	sp
818.01	0.0930s	st	acf	st	acf	st	sp	acv	acf	ph	sp
818.54	0.0110s	st	acf	st	acf	sp	st	acv	acf	sp	ph
817.89	0.1460s	st	acf	st	acf	st	sp	acv	acf	sp	ph
819.41	0.0150s	st	acf	st	acf	sp	st	acv	ph	acf	sp
818.75	0.1060s	st	acf	st	acf	st	sp	acv	ph	acf	sp
814.19	0.0140s	st	acf	st	acf	sp	st	acv	ph	sp	acf
813.53	0.0860s	st	acf	st	acf	st	sp	acv	ph	sp	acf
818.37	0.0110s	st	acf	st	acf	sp	st	acv	sp	acf	ph
817.71	0.0830s	st	acf	st	acf	st	sp	acv	sp	acf	ph
817.63	0.0100s	st	acf	st	acf	sp	st	acv	sp	ph	acf
816.99	0.1210s	st	acf	st	acf	st	sp	acv	sp	ph	acf
820.20	0.0080s	st	acf	st	acf	sp	st	ph	acf	acv	sp
819.58	0.1380s	st	acf	st	acf	st	sp	ph	acf	acv	sp
811.43	0.0070s	st	acf	st	acf	sp	st	ph	acf	sp	acv
810.79	0.0810s	st	acf	st	acf	st	sp	ph	acf	sp	acv
818.04	0.0110s	st	acf	st	acf	sp	st	ph	acv	acf	sp
817.41	0.0990s	st	acf	st	acf	st	sp	ph	acv	acf	sp
811.82	0.0100s	st	acf	st	acf	sp	st	ph	acv	sp	acf
811.17	0.0800s	st	acf	st	acf	st	sp	ph	acv	sp	acf
809.83	0.0110s	st	acf	st	acf	sp	st	ph	sp	acf	acv
809.15	0.1240s	st	acf	st	acf	st	sp	ph	sp	acf	acv
807.71	0.0100s	st	acf	st	acf	sp	st	ph	sp	acv	acf
807.03	0.1110s	st	acf	st	acf	st	sp	ph	sp	acv	acf
810.24	0.0100s	st	acf	st	acf	sp	st	sp	acf	acv	ph
809.53	0.0800s	st	acf	st	acf	st	sp	sp	acf	acv	ph
808.82	0.0110s	st	acf	st	acf	sp	st	sp	acf	ph	acv
808.11	0.0850s	st	acf	st	acf	st	sp	sp	acf	ph	acv
806.58	0.0110s	st	acf	st	acf	sp	st	sp	acv	acf	ph
805.87	0.1170s	st	acf	st	acf	st	sp	sp	acv	acf	ph
805.76	0.0130s	st	acf	st	acf	sp	st	sp	acv	ph	acf
805.05	0.0900s	st	acf	st	acf	st	sp	sp	acv	ph	acf
806.29	0.0110s	st	acf	st	acf	sp	st	sp	ph	acf	acv
805.58	0.1020s	st	acf	st	acf	st	sp	sp	ph	acf	acv
804.53	0.0100s	st	acf	st	acf	sp	st	sp	ph	acv	acf
803.82	0.0850s	st	acf	st	acf	st	sp	sp	ph	acv	acf

Appendix E

Wine dataset: PCBN decompositions and conditional cdfs

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

PCBN 1

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{1,3}(u_1,u_3)c_{2,3|1}(u_2,F_{3|1}(u_3|u_1))c_{1,4}(u_1,u_4)c_{2,4|1}(u_2,F_{4|1}(u_4|u_1))c_{1,5}(u_1,u_5)c_{2,7}(u_2,u_7) \\ & c_{6,7|2}(u_6,F_{7|2}(u_7|u_2))c_{7,8}(u_7,u_8)c_{6,8|7}(F_{6|7}(u_6|u_7),F_{8|7}(u_8|u_7))c_{1,8|6,7}(u_1,F_{8|6,7}(u_8|u_6,u_7)) \\ & c_{5,8|1,6,7}(F_{5|1}(u_5|u_1),F_{8|1,6,7}(u_8|u_1,u_6,u_7)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$1->3 : c_{1,3}(u_1,u_3)$:
Frank(theta=-1.45943)

$2->3 : c_{2,3|1}(u_2,F_{3|1}(u_3|u_1))$:
$F_{3|1}(u_3|u_1) = h2_{1,3}(u_1,u_3)$
ReflectUV Gumbel(theta=2.21036)

$1->4 : c_{1,4}(u_1,u_4)$:
Gauss(theta=0.66346)

$2->4 : c_{2,4|1}(u_2,F_{4|1}(u_4|u_1))$:
$F_{4|1}(u_4|u_1) = h2_{1,4}(u_1,u_4)$
Clayton(theta=0.34996)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

$1->5 : c_{1,5}(u_1, u_5)$:
ReflectU Gumbel(theta=1.96949)

$2->7 : c_{2,7}(u_2, u_7)$:
Student(theta=0.153808,nu=8.81457)

$6->7 : c_{6,7|2}(u_6, F_{7|2}(u_7|u_2))$:
$F_{7|2}(u_7|u_2) = h2_{2,7}(u_2, u_7)$
Frank(theta=-2.06778)

$7->8 : c_{7,8}(u_7, u_8)$:
Frank(theta=-4.27285)

$6->8 : c_{6,8|7}(F_{6|7}(u_6|u_7), F_{8|7}(u_8|u_7))$:
$F_{6|7}(u_6|u_7) = \int_0^1 c_{2,7}(v_2, u_7) h1_{6,7|2}(u_6, F_{7|2}(u_7|v_2))dv_2$
$F_{8|7}(u_8|u_7) = h2_{7,8}(u_7, u_8)$
Joe(theta=1.16464)

$1->8 : c_{1,8|6,7}(u_1, F_{8|6,7}(u_8|u_6, u_7))$:
$F_{8|6,7}(u_8|u_6, u_7) = h2_{6,8|7}(F_{6|7}(u_6|u_7), F_{8|7}(u_8|u_7))$
Student(theta=0.602789,nu=7.10369)

$5->8 : c_{5,8|1,6,7}(F_{5|1}(u_5|u_1), F_{8|1,6,7}(u_8|u_1, u_6, u_7))$:
$F_{5|1}(u_5|u_1) = h2_{1,5}(u_1, u_5)$
$F_{8|1,6,7}(u_8|u_1, u_6, u_7) = h2_{1,8|6,7}(u_1, F_{8|6,7}(u_8|u_6, u_7))$
ReflectV Clayton(theta=0.104061)

PCBN 2

Note that all univariate marginals are uniform distributions on $[0, 1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0, 1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{1,3}(u_1, u_3) c_{2,3|1}(u_2, F_{3|1}(u_3|u_1)) c_{1,4}(u_1, u_4) c_{2,4|1}(u_2, F_{4|1}(u_4|u_1)) c_{1,5}(u_1, u_5) c_{6,7}(u_6, u_7) \\ & c_{2,7|6}(u_2, F_{7|6}(u_7|u_6)) c_{5,8}(u_5, u_8) c_{1,8|5}(F_{1|5}(u_1|u_5), F_{8|5}(u_8|u_5)) \\ & c_{7,8|1,5}(u_7, F_{8|1,5}(u_8|u_1, u_5)) c_{6,8|1,5,7}(F_{6|7}(u_6|u_7), F_{8|1,5,7}(u_8|u_1, u_5, u_7)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$1->3 : c_{1,3}(u_1, u_3)$:
Frank(theta=-1.45943)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

$2 -> 3 : c_{2,3|1}(u_2, F_{3|1}(u_3|u_1))$:
$F_{3|1}(u_3|u_1) = h2_{1,3}(u_1, u_3)$
ReflectUV Gumbel(theta=2.21036)

$1 -> 4 : c_{1,4}(u_1, u_4)$:
Gauss(theta=0.66346)

$2 -> 4 : c_{2,4|1}(u_2, F_{4|1}(u_4|u_1))$:
$F_{4|1}(u_4|u_1) = h2_{1,4}(u_1, u_4)$
Clayton(theta=0.34996)

$1 -> 5 : c_{1,5}(u_1, u_5)$:
ReflectU Gumbel(theta=1.96949)

$6 -> 7 : c_{6,7}(u_6, u_7)$:
Frank(theta=-2.04587)

$2 -> 7 : c_{2,7|6}(u_2, F_{7|6}(u_7|u_6))$:
$F_{7|6}(u_7|u_6) = h2_{6,7}(u_6, u_7)$
Student(theta=0.157316,nu=8.15814)

$5 -> 8 : c_{5,8}(u_5, u_8)$:
Frank(theta=-3.77097)

$1 -> 8 : c_{1,8|5}(F_{1|5}(u_1|u_5), F_{8|5}(u_8|u_5))$:
$F_{1|5}(u_1|u_5) = h1_{1,5}(u_1, u_5)$
$F_{8|5}(u_8|u_5) = h2_{5,8}(u_5, u_8)$
Gumbel(theta=1.39641)

$7 -> 8 : c_{7,8|1,5}(u_7, F_{8|1,5}(u_8|u_1, u_5))$:
$F_{8|1,5}(u_8|u_1, u_5) = h2_{1,8|5}(F_{1|5}(u_1|u_5), F_{8|5}(u_8|u_5))$
Frank(theta=-3.91674)

$6 -> 8 : c_{6,8|1,5,7}(F_{6|7}(u_6|u_7), F_{8|1,5,7}(u_8|u_1, u_5, u_7))$:
$F_{6|7}(u_6|u_7) = h1_{6,7}(u_6, u_7)$
$F_{8|1,5,7}(u_8|u_1, u_5, u_7) = h2_{7,8|1,5}(u_7, F_{8|1,5}(u_8|u_1, u_5))$
Gumbel(theta=1.06595)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

PCBN 3

Note that all univariate marginals are uniform distributions on $[0,1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0,1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{2,3}(u_2,u_3)c_{1,3|2}(u_1,F_{3|2}(u_3|u_2))c_{1,4}(u_1,u_4)c_{2,4|1}(u_2,F_{4|1}(u_4|u_1))c_{1,5}(u_1,u_5)c_{6,7}(u_6,u_7) \\ & c_{2,7|6}(u_2,F_{7|6}(u_7|u_6))c_{1,8}(u_1,u_8)c_{7,8|1}(u_7,F_{8|1}(u_8|u_1)) \\ & c_{5,8|1,7}(F_{5|1}(u_5|u_1),F_{8|1,7}(u_8|u_1,u_7))c_{6,8|1,5,7}(F_{6|7}(u_6|u_7),F_{8|1,5,7}(u_8|u_1,u_5,u_7)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$2->3 : c_{2,3}(u_2,u_3)$:
ReflectUV Gumbel(theta=2.2281)

$1->3 : c_{1,3|2}(u_1,F_{3|2}(u_3|u_2))$:
$F_{3|2}(u_3|u_2) = h2_{2,3}(u_2,u_3)$
Frank(theta=-1.37383)

$1->4 : c_{1,4}(u_1,u_4)$:
Gauss(theta=0.66346)

$2->4 : c_{2,4|1}(u_2,F_{4|1}(u_4|u_1))$:
$F_{4|1}(u_4|u_1) = h2_{1,4}(u_1,u_4)$
Clayton(theta=0.34996)

$1->5 : c_{1,5}(u_1,u_5)$:
ReflectU Gumbel(theta=1.96949)

$6->7 : c_{6,7}(u_6,u_7)$:
Frank(theta=-2.04587)

$2->7 : c_{2,7|6}(u_2,F_{7|6}(u_7|u_6))$:
$F_{7|6}(u_7|u_6) = h2_{6,7}(u_6,u_7)$
Student(theta=0.157316,nu=8.15814)

$1->8 : c_{1,8}(u_1,u_8)$:
Gumbel(theta=1.74972)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

$7->8: c_{7,8|1}(u_7, F_{8|1}(u_8|u_1))$:
$F_{8|1}(u_8|u_1) = h2_{1,8}(u_1, u_8)$
Frank(theta=-4.07114)

$5->8: c_{5,8|1,7}(F_{5|1}(u_5|u_1), F_{8|1,7}(u_8|u_1, u_7))$:
$F_{5|1}(u_5|u_1) = h2_{1,5}(u_1, u_5)$
$F_{8|1,7}(u_8|u_1, u_7) = h2_{7,8|1}(u_7, F_{8|1}(u_8|u_1))$
ReflectV Clayton(theta=0.168986)

$6->8: c_{6,8|1,5,7}(F_{6|7}(u_6|u_7), F_{8|1,5,7}(u_8|u_1, u_5, u_7))$:
$F_{6|7}(u_6|u_7) = h1_{6,7}(u_6, u_7)$
$F_{8|1,5,7}(u_8|u_1, u_5, u_7) = h2_{5,8|1,7}(F_{5|1}(u_5|u_1), F_{8|1,7}(u_8|u_1, u_7))$
Gumbel(theta=1.05813)

PCBN 4

Note that all univariate marginals are uniform distributions on $[0, 1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0, 1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{1,3}(u_1, u_3)c_{2,3|1}(u_2, F_{3|1}(u_3|u_1))c_{2,4}(u_2, u_4)c_{1,4|2}(u_1, F_{4|2}(u_4|u_2))c_{1,5}(u_1, u_5)c_{2,7}(u_2, u_7) \\ & c_{6,7|2}(u_6, F_{7|2}(u_7|u_2))c_{6,8}(u_6, u_8)c_{5,8|6}(u_5, F_{8|6}(u_8|u_6)) \\ & c_{7,8|5,6}(F_{7|6}(u_7|u_6), F_{8|5,6}(u_8|u_5, u_6))c_{1,8|5,6,7}(F_{1|5}(u_1|u_5), F_{8|5,6,7}(u_8|u_5, u_6, u_7)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$1->3: c_{1,3}(u_1, u_3)$:
Frank(theta=-1.45943)

$2->3: c_{2,3|1}(u_2, F_{3|1}(u_3|u_1))$:
$F_{3|1}(u_3|u_1) = h2_{1,3}(u_1, u_3)$
ReflectUV Gumbel(theta=2.21036)

$2->4: c_{2,4}(u_2, u_4)$:
Clayton(theta=0.19719)

$1->4: c_{1,4|2}(u_1, F_{4|2}(u_4|u_2))$:
$F_{4|2}(u_4|u_2) = h2_{2,4}(u_2, u_4)$
Gauss(theta=0.686996)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

$1->5 : c_{1,5}(u_1, u_5)$:
ReflectU Gumbel(theta=1.96949)

$2->7 : c_{2,7}(u_2, u_7)$:
Student(theta=0.153808,nu=8.81457)

$6->7 : c_{6,7|2}(u_6, F_{7|2}(u_7|u_2))$:
$F_{7|2}(u_7|u_2) = h2_{2,7}(u_2, u_7)$
Frank(theta=-2.06778)

$6->8 : c_{6,8}(u_6, u_8)$:
Frank(theta=2.04603)

$5->8 : c_{5,8|6}(u_5, F_{8|6}(u_8|u_6))$:
$F_{8|6}(u_8|u_6) = h2_{6,8}(u_6, u_8)$
Frank(theta=-3.85949)

$7->8 : c_{7,8|5,6}(F_{7|6}(u_7|u_6), F_{8|5,6}(u_8|u_5, u_6))$:
$F_{7|6}(u_7|u_6) = \int_0^1 h2_{6,7|2}(u_6, F_{7|2}(u_7|v_2))dv_2$
$F_{8|5,6}(u_8|u_5, u_6) = h2_{5,8|6}(u_5, F_{8|6}(u_8|u_6))$
Frank(theta=-3.08501)

$1->8 : c_{1,8|5,6,7}(F_{1|5}(u_1|u_5), F_{8|5,6,7}(u_8|u_5, u_6, u_7))$:
$F_{1|5}(u_1|u_5) = h1_{1,5}(u_1, u_5)$
$F_{8|5,6,7}(u_8|u_5, u_6, u_7) = h2_{7,8|5,6}(F_{7|6}(u_7|u_6), F_{8|5,6}(u_8|u_5, u_6))$
Student(theta=0.390518,nu=6.42285)

PCBN 5

Note that all univariate marginals are uniform distributions on $[0, 1]$, so $f_i(u_i) = 1, F_i(u_i) = u_i$ for $u_i \in [0, 1]$.
Decomposition of this PCBN:

$$\begin{aligned} f(\mathbf{u}) =& c_{1,3}(u_1, u_3)c_{2,3|1}(u_2, F_{3|1}(u_3|u_1))c_{1,4}(u_1, u_4)c_{2,4|1}(u_2, F_{4|1}(u_4|u_1))c_{1,5}(u_1, u_5)c_{6,7}(u_6, u_7) \\ & c_{2,7|6}(u_2, F_{7|6}(u_7|u_6))c_{1,8}(u_1, u_8)c_{7,8|1}(u_7, F_{8|1}(u_8|u_1)) \\ & c_{5,8|1,7}(F_{5|1}(u_5|u_1), F_{8|1,7}(u_8|u_1, u_7))c_{6,8|1,5,7}(F_{6|7}(u_6|u_7), F_{8|1,5,7}(u_8|u_1, u_5, u_7)) \end{aligned}$$

Output for all copulas including the formulas for the conditional cdfs:
($h1_{i,j} = h_{\underline{i}j}$, $h2_{i,j} = h_{i\underline{j}}$)

$1->3 : c_{1,3}(u_1, u_3)$:
Frank(theta=-1.45943)

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)

$2->3 : c_{2,3|1}(u_2, F_{3|1}(u_3|u_1))$:
$F_{3|1}(u_3|u_1) = h2_{1,3}(u_1, u_3)$
ReflectUV Gumbel(theta=2.21036)

$1->4 : c_{1,4}(u_1, u_4)$:
Gauss(theta=0.66346)

$2->4 : c_{2,4|1}(u_2, F_{4|1}(u_4|u_1))$:
$F_{4|1}(u_4|u_1) = h2_{1,4}(u_1, u_4)$
Clayton(theta=0.34996)

$1->5 : c_{1,5}(u_1, u_5)$:
ReflectU Gumbel(theta=1.96949)

$6->7 : c_{6,7}(u_6, u_7)$:
Frank(theta=-2.04587)

$2->7 : c_{2,7|6}(u_2, F_{7|6}(u_7|u_6))$:
$F_{7|6}(u_7|u_6) = h2_{6,7}(u_6, u_7)$
Student(theta=0.157316,nu=8.15814)

$1->8 : c_{1,8}(u_1, u_8)$:
Gumbel(theta=1.74972)

$7->8 : c_{7,8|1}(u_7, F_{8|1}(u_8|u_1))$:
$F_{8|1}(u_8|u_1) = h2_{1,8}(u_1, u_8)$
Frank(theta=-4.07114)

$5->8 : c_{5,8|1,7}(F_{5|1}(u_5|u_1), F_{8|1,7}(u_8|u_1, u_7))$:
$F_{5|1}(u_5|u_1) = h2_{1,5}(u_1, u_5)$
$F_{8|1,7}(u_8|u_1, u_7) = h2_{7,8|1}(u_7, F_{8|1}(u_8|u_1))$
ReflectV Clayton(theta=0.168986)

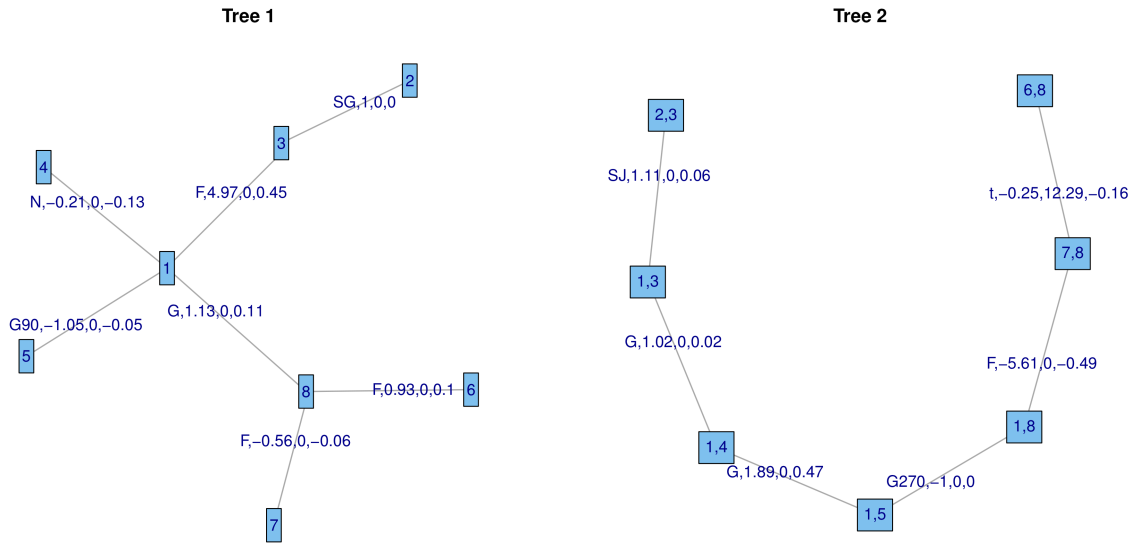
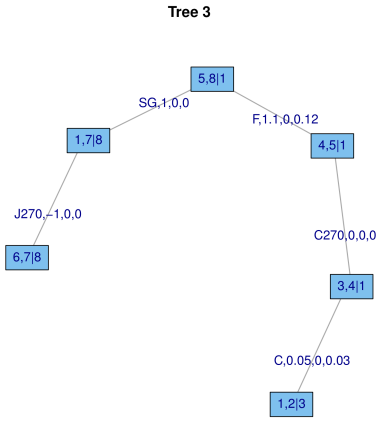
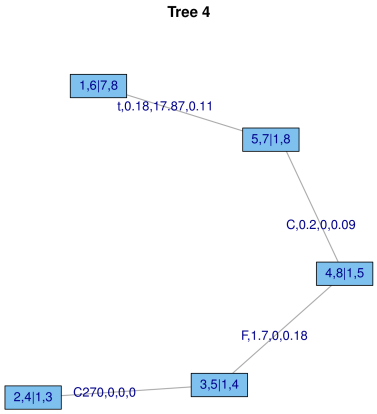
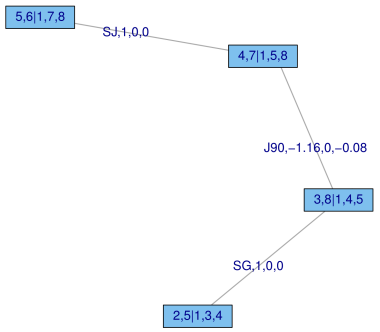
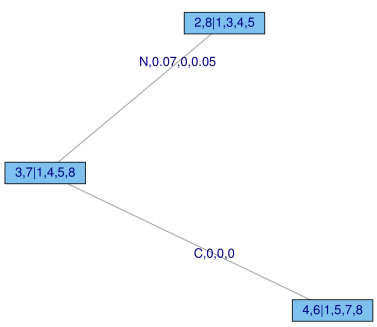
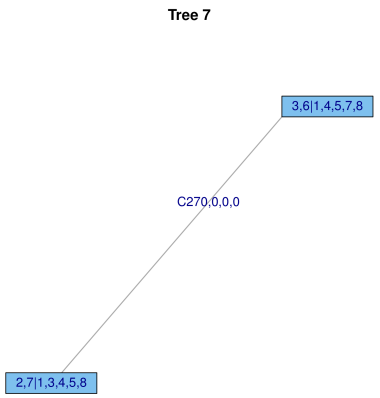
$6->8 : c_{6,8|1,5,7}(F_{6|7}(u_6|u_7), F_{8|1,5,7}(u_8|u_1, u_5, u_7))$:
$F_{6|7}(u_6|u_7) = h1_{6,7}(u_6, u_7)$
$F_{8|1,5,7}(u_8|u_1, u_5, u_7) = h2_{5,8|1,7}(F_{5|1}(u_5|u_1), F_{8|1,7}(u_8|u_1, u_7))$
Gumbel(theta=1.05813)

Appendix F

Wine dataset: RVine tree plots

The following tree plots show the structure as well as the copula families and parameters of the RVine fitted on the wine dataset. "F,1.6,0,0.17" means a Frank Copula with par1 = 1.6, par2 = 0 (not used) and $\tau = 0.17$.

Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)



Indices: (1 = acf, 2 = st, 3 = sf, 4 = den, 5 = ph, 6 = sp, 7 = acv, 8 =acc)