Analyzing dependent data with vine copulas (Lecture 2)

Claudia Czado <cczado@ma.tum.de>
TU München

École d'été 2018 Sept 2-5,2018

Overview

1. Regular vine distributions
2. Data preparation and exploration or vine based modeling
3. Model selection and estimation in regular vine based models
4. Summary

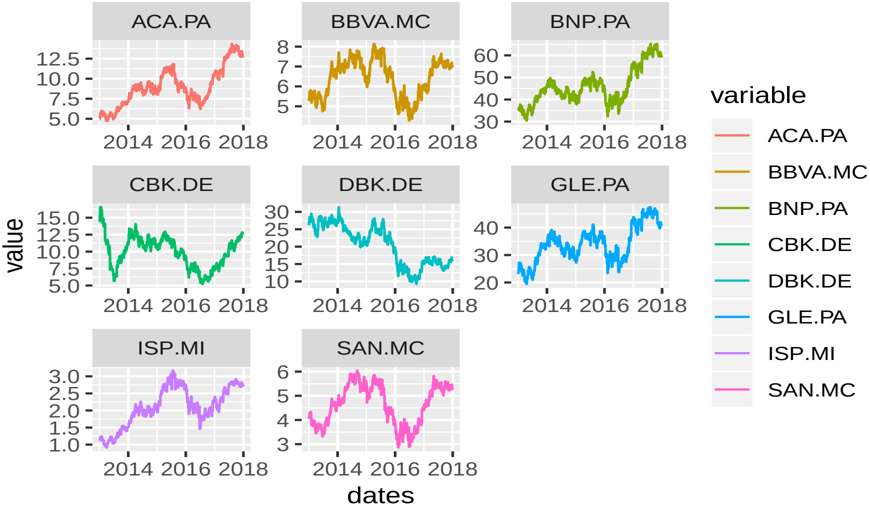
History of regular vine models

- Regular vines started with Joe (1996) who constructed them using mixtures of conditional distribution functions.
- There are many choices of conditioning variables, in $d = 3$ we have 3 possibilities.
- Bedford and Cooke (2002) introduced a graphical structure to organize the sequence of conditioning variables.
- In contrast to Joe (1996) the construction of Bedford and Cooke (2002) is based on densities.
- Gaussian vines were analyzed in Kurowicka and Cooke (2006), while estimation for non Gaussian vines started with Aas et al. (2009).
- Joe (2014) is the up to date reference. An introductory book on vines will appear in 2019.
- Web resources are `vine-copula.org` and `en.wikipedia.org/wiki/Vine_copula`

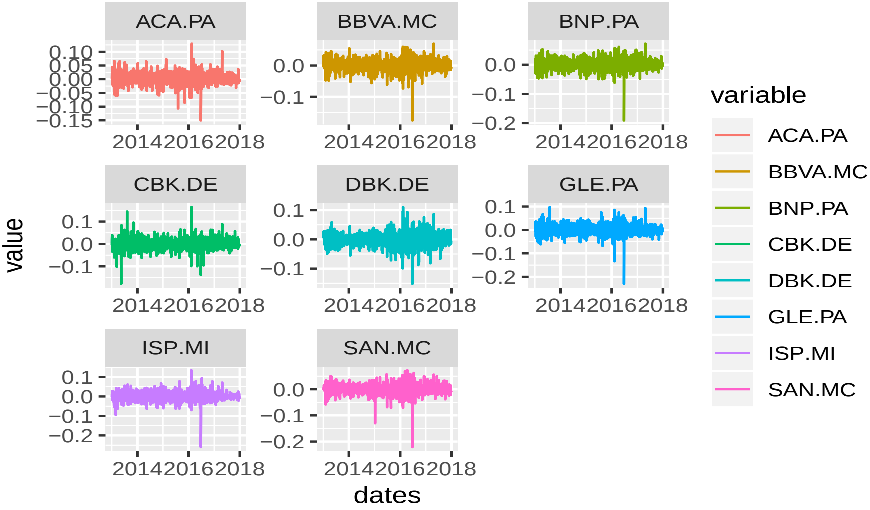
Eight European banks

- Consider daily values of eight European banks between 2013 and 2017 from `Yahoo`
- The following banks were included:
  - **ACA.PA**: Crédit Agricole S.A. (France)
  - **BBVA.MC**: Banco Bilbao Vizcaya Argentaria, S.A. (Spain)
  - **BNP.PA**: BNP Paribas SA (France)
  - **CBK.DE**: Commerzbank AG (Germany)
  - **DBK.DE**: Deutsche Bank AG (Germany)
  - **GLE.PA**: Société Générale Société anonyme (France)
  - **ISP.MI**: Intesa Sanpaolo S.p.A. (Italy)
  - **SAN.MC**: Banco Santander, S.A (Spain)

Daily asset values for eight banks



Daily return values for eight banks



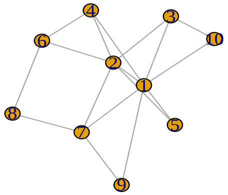
Outline

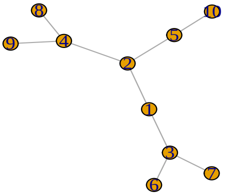
How do vines work in higher dimensions?

- Which pairs of variables are needed?
- What are the conditioning variables?

Some graph theoretic background

- A graph is a pair $G = (N, E)$ with node set N and edge set E.
- A path is a graph $P = (N_p, E_p)$ with node set $N_p = \{\nu_0, \nu_1, \dots, \nu_k\}$ and edge set $E_p = \{\{\nu_0, \nu_1\}, \{\nu_1, \nu_2\}, \dots, \{\nu_{k-1}, \nu_k\}$.
- A graph T is a tree if any two nodes of T are connected by a unique path in T.





Regular vine distributions

A parametric regular vine distribution $R(\mathcal{V}, \mathcal{C}, \boldsymbol{\theta})$ for the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with marginal distributions $F_j, j = 1, \ldots, d$ has three components:

Components of a regular vine distribution

1. **Tree structure:** set of linked trees $\mathcal{V}$
2. **Parametric bivariate copulas:** Set $\mathcal{C} = \mathcal{C}(\mathcal{V})$ for each edge in tree structure. Members of $\mathcal{C}$ are called pair copulas.
3. **Corresponding parameters:** $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathcal{C}(\mathcal{V}))$

Abbreviation: R-vine $R(\mathcal{V}, \mathcal{C}, \boldsymbol{\theta})$

Regular vine tree structure

An n-dimensional vine tree structure $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$ is a sequence of linked $d-1$ trees with

Vine tree structure (Bedford and Cooke (2002))

- Tree T_1 is a tree on nodes 1 to d.
- Tree T_j has $d+1-j$ nodes and $d-j$ edges.
- Edges in tree T_j become nodes in tree T_{j+1}.
- **Proximity condition:** Two nodes in tree T_{j+1} can be joined by an edge only if the corresponding edges in tree T_j share a node.

Special vine tree sequences:

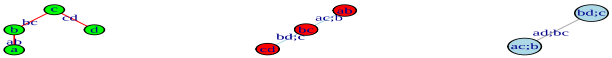
- D-vines use only path like trees
- canonical C-vines use only star like tree

Edges in the vine tree sequence

Consider the tree sequence $\mathcal{V} = \{T_1, \ldots, T_{d-1}\}$, where tree T_k has edge set E_k and node set N_k.

Example:

- Let a, b, c, d nodes in T_1 with edges $e_1 = ab$, $e_2 = bc$, $e_3 = cd$ three connected edges in E_1 of Tree T_1.
- Then e_1, e_2 and e_3 are nodes in T_2.
  - The edge between e_1 and e_2 in Tree T_2 we denote by $e_{a,c;b}$.
  - The edge between e_2 and e_3 in Tree T_2 we denote by $e_{b,d;c}$.
- The edges $e_{a,c;b}$ and $e_{b,d;c}$ are nodes in tree T_3 and the edge between these two nodes is denoted by $e_{a,d;b,c}$.



General: Any edge in tree T_k can be characterized by

$$e_{i,j;D}, \text{ where } D \text{ is a set of } k-1 \text{ indices.}$$

Pair copulas and edges

- Let (i, j, D) be chosen such that $e_{ij;D}$ is edge in tree T_k.
- Let $C_{ij;D}$ be the pair copula associated with edge $e_{ij;D}$.
- Then $C_{ij;D}$ $(c_{ij;D})$ is the copula (density) associated with the bivariate conditional distribution (X_i, X_j) given $\mathbf{X}_D = \mathbf{x}_D$.
- Since we assume the simplifying assumption this copula is independent of the specific value $\mathbf{x}_D$.

Regular vine density (Bedford and Cooke 2002)

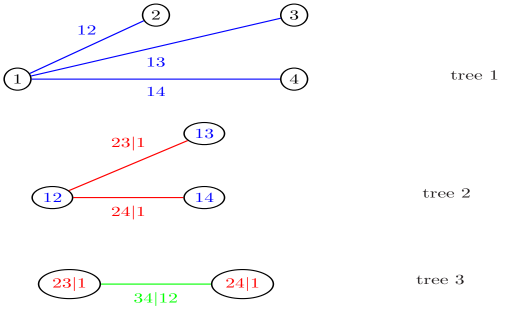
$$f_{1:d}(\mathbf{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{k=1}^{d-1} \prod_{e=e_{ij;D}\in E_k} c_{ij;D}(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D))$$

Canonical C-vine distributions

are regular vine distributions where each tree has a unique node that is connected to $n - j$ edges.

four dimensional C-vine distribution

$$f_{1234} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{34|12}$$

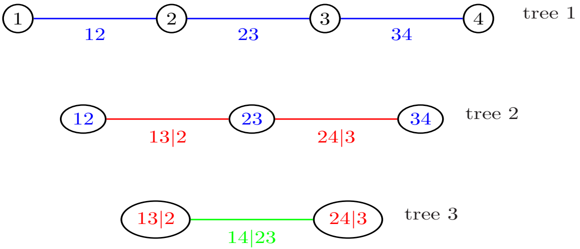


D-vine distributions

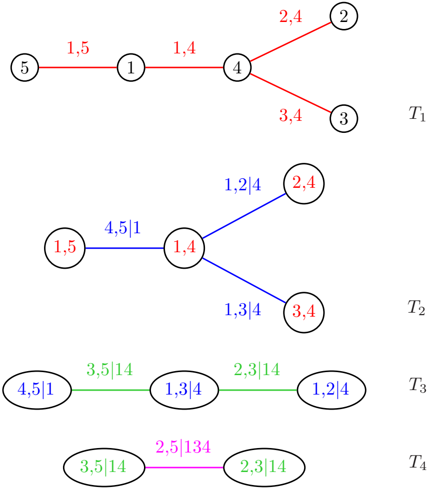
are regular vine distributions where no node in any tree is connected to more than two edges

Four dimensional D-vine distribution

$$f_{1234} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{14|23}$$



Can we see an example of an R-vine?



Density

$$f = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$$

$$\cdot\ c_{14} \cdot c_{15} \cdot c_{24} \cdot c_{34}$$

$$\cdot\ c_{12;4} \cdot c_{13;4} \cdot c_{45;1}$$

$$\cdot\ c_{23;14} \cdot c_{35;14}$$

$$\cdot\ c_{25;134}$$

Conditional distribution functions

For $\boldsymbol{v} = (v_1, \ldots, v_d)$ and $\boldsymbol{v}_{-j} = (v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_d)$

$$f(x|\boldsymbol{v}) = c_{xv_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})) \cdot f(x|\boldsymbol{v}_{-j})$$

Univariate conditioning (v univariate)

Since $f(x|v) = c_{xv}(F_x(x), F_v(v)) f_x(x)$ we have

$$\begin{aligned} F(x|v) &= \int_{-\infty}^{x} \frac{\partial^2 C_{xv}(F_x(u), F_v(v))}{\partial F_x(u)\, \partial F_v(v)} f_x(u) du \\ &= \frac{\partial\, C_{xv}(F_x(x), F_v(v))}{\partial F_v(v)} \end{aligned}$$

Multivariate conditioning (Joe 1996)

$$F(x|\boldsymbol{v}) = \frac{\partial\, C_{x,v_j|\boldsymbol{v}_{-j}}(F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j}))}{\partial F(v_j|\boldsymbol{v}_{-j})}$$

All conditional cdf's in an R-vine can be recursively determined.

R-vine matrices (I)

- Vine tree structure is stored in triangular matrix $M = \{m_{ij}\}$
- Consider the R-vine with edges
  - **Tree 1:** 51,14,42,43
  - **Tree 2:** 45;1, 12;4, 13;4
  - **Tree 3:** 35;14 ,23;14
  - **Tree 4:** 25;134
- Start with edge 25;134 in Tree 4, set $m_{11} = 2$ and $m_{21} = 5$.
- Find partner of 2 in Tree 3:
  - Find edge which has 2 in the conditioned set (23;14)
  - Partner is then 3 and set $m_{31} = 3$
- Partner of 2 in Tree 2 is 1 and therefore set $m_{31} = 1$.
- Partner of 2 in Tree 1 is 4 and therefore set $m_{41} = 4$.
- Column 1 identifies the edges 25;314, 23;14 ;21;4, 24
- Remove all edges containing 2 , giving edges 51,14,43,45;1, 13;4 , 35;14 and do the same with these edges.

R-vine matrices (II)

- Resulting R-vine matrix $M = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 5 & 3 & 0 & 0 & 0 \\ 3 & 5 & 5 & 0 & 0 \\ 1 & 1 & 4 & 1 & 0 \\ 4 & 4 & 1 & 4 & 4 \end{bmatrix}$
- **Column 1** identifies 25;314, 23;14, 21;4 , 24.
- **Column 2** identifies 35;14, 31;4; 34.
- **Column 3** identifies 54;1, 51.
- **Column 4** identifies 14.
- Pair copula family matrix $\mathcal{C}$ has following structucе

$$\mathcal{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ c_{25;314} & 0 & 0 & 0 & 0 \\ c_{23;14} & c_{35;14} & 0 & 0 & 0 \\ c_{21;4} & c_{31;4} & c_{54;1} & 0 & 0 \\ c_{24} & c_{34} & c_{51} & c_{14} & 0 \end{bmatrix}$$

- The associated copula parameters are stored similarly as $\mathcal{C}$.

Scope of the vine copula models

- A regular vine cdf with uniform margins is a vine copula.
- **Known vine copula classes**:
  - multivariate Gaussian copula
  - multivariate t copula
  - multivariate Clayton copula (Takahashi (1965)). Stöber et al. (2013) showed this is the only Archimedean copula.
- **Number of $d-1$ vine trees:** (Morales-Nápoles et al. 2010)

$$m(d) = d! \times 2^{\binom{d-2}{2}/2} \quad \textbf{Ex:} \; m(25) \approx 1.1. \times 10^{1}01$$

- **Number of pair copulas:**

$$p(d) = \frac{d \times (d+1)}{2} \quad \textbf{Ex:} \; p(500) \approx 124,000$$

Efficient estimation and model selection are vital

Outline

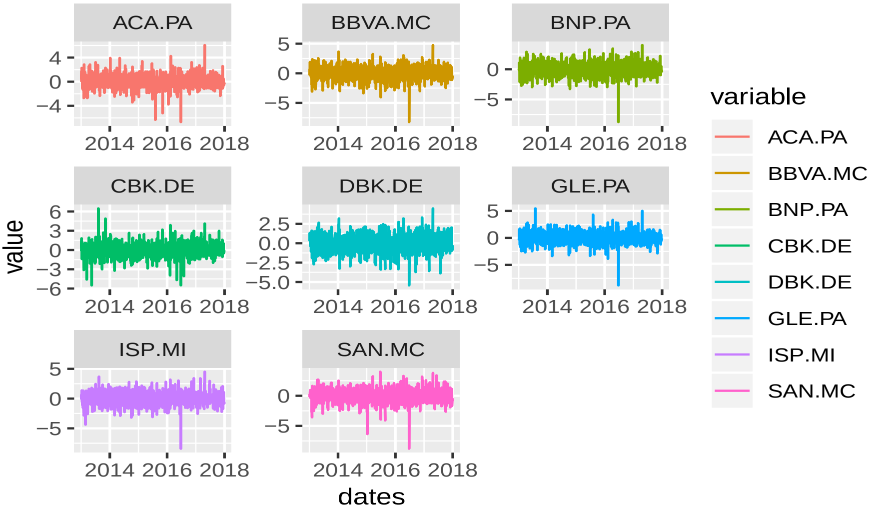
Marginal analysis of bank asset values

- Fit a GARCH(1,1) model with Student t errors to each asset $X_{tj}, t = 1, \ldots, T$ given by

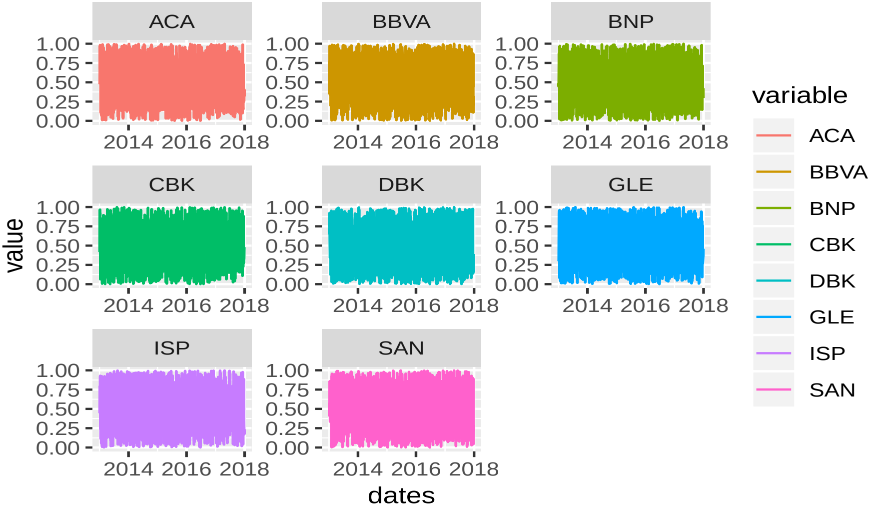
$$\begin{aligned} X_{tj} &= \sigma_{tj}\epsilon_{tj} \\ \sigma_{tj}^2 &= \omega_j + \alpha_j X_{(t-1)j}^2 + \beta_j \sigma_{(t-1)j}^2, \end{aligned}$$

  where ϵ_{tj} are i.i.d standardized Student t with df_j degrees of freedom.
- Form fitted standardized residuals $\hat{\epsilon}_{tj} = \frac{X_{tj}}{\hat{\sigma}_{tj}}$.
- Transform $\hat{\epsilon}_{tj}$ to pseudo copula values using the standard Student t distribution with $\hat{df}_j$.

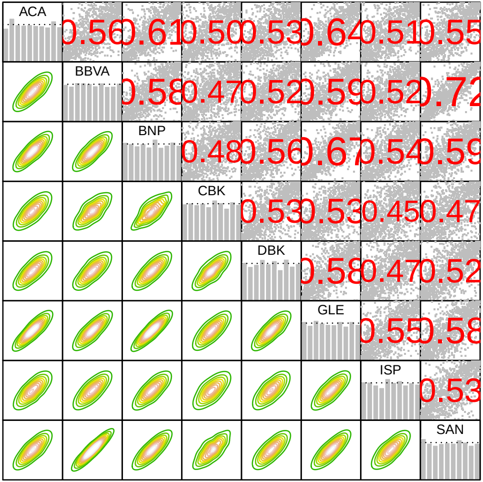
Standardized residuals after GARCH fit



Pseudo copula data for bank data



Pairwise normalized contour plots for banks



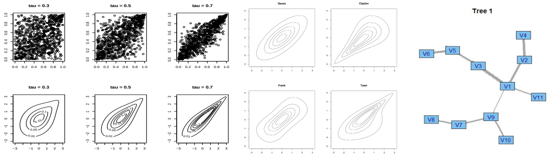
Outline

Model selection and estimation senarios

Need to study three senarios (Czado et al. 2013a):

- for given R-vine tree structure and pair copula families. Parameters are to be estimated.
- for given R-vine tree structure. Copula families and parameters are to be estimated
- all three components are unknown, i.e the tree structure, the pair copula families and their parameters are to be estimated.

Assumption: at least approximate i.i.d. copula sample from an R-vine copula



Known tree structure and copula families

- **Sequential estimation:**
  - Parameters are sequentially estimated starting from the top tree until the last (Aas et al. (2009), Czado et al. (2012)).
  - Asymptotic theory is available (Haff et al. (2013)), however corresponding standard error estimates are difficult to compute.
  - Can be used as starting values for maximum likelihood.
- **Maximum likelihood estimation:**
  - Asymptotically efficient under regularity conditions, again estimated standard errors are numerically challenging.
  - Uncertainty in value-at-risk (high quantiles) is difficult to assess.
- **Bayesian estimation:**
  - Posterior is tractable using Markov Chain Monte Carlo (Min and Czado (2011) for D-vines and Gruber (2011) for R-vines)
  - Prior beliefs can be incorporated and credible intervals allow to assess uncertainty for all quantities.

Known vine structure (Brechmann 2010)

- Select a set of candidate familes for the pair copulas.
- For the pair copulas in the first tree use the copula data directly to fit all candidate families for each pair copula C_{ij}.
- Choose as family the one with the lowest AIC. (AIC or BIC not so crucial since only families with 1 or 2 parameters)
- For pair copula $C_{ab;D}$ in tree T_j, $j \geq 2$ with some $a, b \in \{1 : d\}$ and D a subset of $\{1 : d\}$ with $a, b \notin D$ define

Pseudo data in tree T_j

$$u_{i,a|D} := C(u_{ia}|\mathbf{x}_{i,D}) \text{ and } u_{i,b|D} := C(u_{ib}|\mathbf{x}_{i,D}), i = 1, \ldots, n$$

Note these conditional cdf's require the pair copulas and parameters in the trees $T_1, \ldots, T_{j-1}$.

- Fit all candidate families to $\{u_{i,a|D}, u_{i,b|D}\}, i = 1 \ldots, n$ and select the one with the lowest AIC.

Sequential selection (Dißmann et al. 2013)

Idea: Model strong pairwise dependencies first

- For T_1 use a maximal spanning tree (MST) algorithm to find tree which maximizes the sum of absolute empirical pair Kendall's τ.
- Use AIC to choose the pair copula families in T_1.
- Apply MST to the graph of all nodes of T_2 (edges in T_1) with all edges allowed by proximity. Kendall's τ estimates use corresponding pseudo observations
- Continue with the remaining trees.
- Other weight measures such as tail dependence measures can be used (Czado et al. 2013b).

Illustration of the Dissmann algorithm (I)

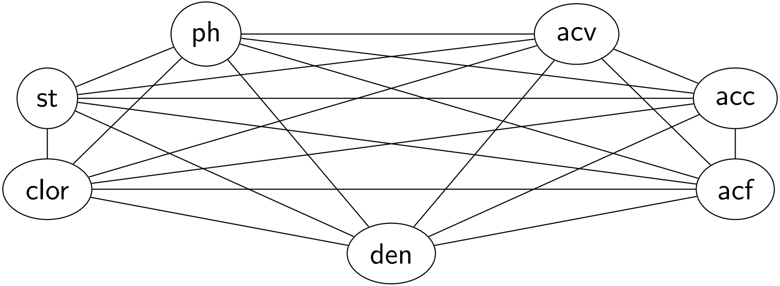


Figure: Complete graph of all pairs from a seven dimensional data set of the first tree.

Illustration of the Dissmann algorithm (II)

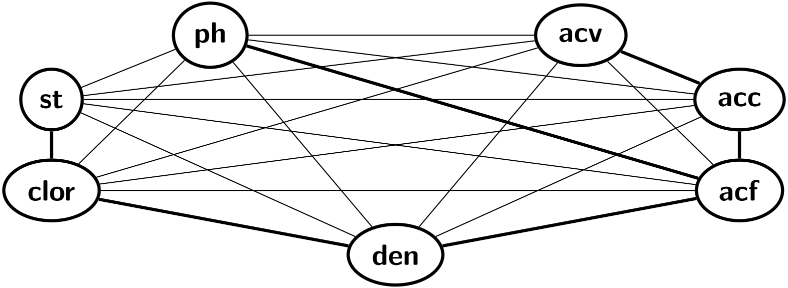


Figure: First tree graph with selected edges highlighted in bold.

Illustration of the Dissmann algorithm (III)

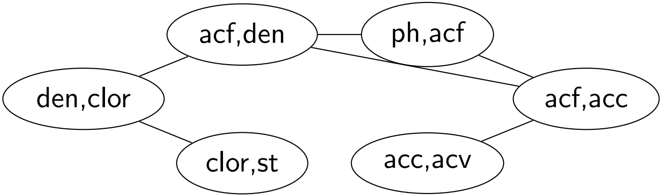


Figure: All pairs of variables of tree T_2 and edges allowed by the proximity condition.

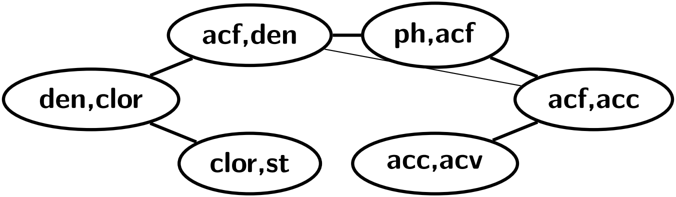


Figure: Tree T_2 with selected edges highlighted in bold.

Bayesian model selection approaches

- Reversible jump (RJ) MCMC (Min and Czado (2011)) and an MCMC with model indicators (Smith et al. (2010)) were used for D-vines choosing between an independence copula and a fixed copula family (nonsequential but tree structure known).
- Gruber and Czado (2015) developed a sequential RJMCMC Bayesian model selection approach, while Gruber and Czado (2018) extends this to a non sequential selection strategy.
- All these approaches are restricted to $d \leq 20$.

R-package `VineCopula`

Schepsmeier et al. (2017)

RVineAIC	AIC of an R-Vine Copula Model
RVineBIC	BIC of an R-Vine Copula Model
RVineCopSelect	Sequential Pair-Copula Selection and Parameter Estimation for R-Vine Copula Models for given Tree Structure
RVineSeqEst	Sequential Parameter Estimation of an R-Vine Copula Model for given Copula families and given Tree Structure
RVineSim	Simulation from an R-Vine Copula Model
RVineStructureSelect	Sequential Specification of R- and C-Vine Copula Models (Dissmann Algorithm)
RVineMLE	Joint Maximum Likelihood Estimation of an R-Vine Copula Model
plot	Tree plots for R-vine matrix object
contour	Contour plots of fitted pair copulas
summary	Summary output for R-vine matrix object

Outline

Summary

- Defined R-vine distributions with
  - arbitrary marginal distributions
  - R-vine tree structure,
  - Set of associated arbitrary pair copulafamilies,
  - Set of associated pair copula parameters.
- There is huge number of R-vine tree structures and a large number of parametric bivariate pair copulas.
- Dependence in multivariate time series is modelled over dependence among standardized residuals of an approriate univariate time series model.
- A sequential estimation approach can be used to estimate parameters.
- The Dissmann algorithm can be used to select an appropriate vine distribution to data.

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009).
Pair-copula constructions of multiple dependence.
Insurance, Mathematics and Economics 44, 182–198.

Bedford, T. and R. M. Cooke (2002).
Vines: A New Graphical Model for Dependent Random Variables.
Annals of Statistics 30(4), 1031–1068.

Brechmann, E. (2010).
Truncated and simplified regular vines and their applications.
Diploma thesis, Technische Universität München.

Czado, C., S. Jeske, and M. Hofmann (2013a).
Selection strategies for regular vine copulae.
Journal de la Société Française de Statistique 154(1), 174–191.

Czado, C., S. Jeske, and M. Hofmann (2013b).
Selection strategies for regular vine copulae.
Journal de la Societe Francaise de Statistique (1), 172–191.

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

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013).
Selecting and estimating regular vine copulae and application to financial returns.
Computational Statistics & Data Analysis 59, 52–69.

References

Gruber, L. and C. Czado (2015).
Sequential Bayesian model selection of regular vine copulas.
Bayesian Analysis 10(4), 937–963.

Gruber, L. and C. Czado (2018).
Bayesian model selection of regular vine copulas.
Bayesian Analysis, in press.

Gruber, L. F. (2011).
Bayesian analysis of R-vine copulas.
Master's thesis, Technische Universität München.

Haff, I. H. et al. (2013).
Parameter estimation for pair-copula constructions.
Bernoulli 19(2), 462–491.

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

Joe, H. (2014).
Dependence modeling with copulas.
CRC Press.

Kurowicka, D. and R. Cooke (2006).
Uncertainty analysis with high dimensional dependence modelling.
Chichester: Wiley.

References

Min, A. and C. Czado (2011).
Bayesian model selection for multivariate copulas using pair-copula constructions.
Canadian Journal of Statistics 39(2), 239–258.

Morales-Nápoles, O., R. Cooke, and D. Kurowicka (2010).
About the number of vines and regular vines on n nodes.
Preprint.

Schepsmeier, U., J. Stöber, E. C. Brechmann, B. Gräler, T. Nagler, and T. Erhardt (2017).
VineCopula: Statistical Inference of Vine Copulas.
Version 2.1.2.

Smith, M., A. Min, C. Almeida, and C. Czado (2010).
Modeling longitudinal data using a pair-copula construction decomposition of serial dependence.
Journal of the American Statistical Association 105, 1467–1479.

Stöber, J., H. Joe, and C. Czado (2013).
Simplified pair copula constructions—limitations and extensions.
Journal of Multivariate Analysis 119, 101–118.

Takahashi, K. (1965).
Note on the multivariate Burr's distribution.
Annals of the Institute of Statistical Mathematics 17, 257–260.