The world of vines

Claudia Czado
Technische Universität München
cczado@ma.tum.de

Lehrstuhl für
Mathematische Statistik

Wien, May 2012

Motivations for vine based models

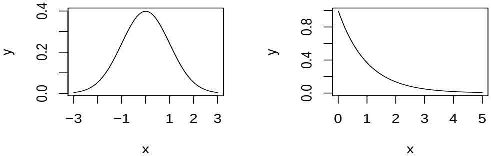
- Many data structures exhibit
  - different marginal distributions
  - nonsymmetric dependencies between some pairs of variables
  - heavy tail dependencies between some pairs of variables
- Cannot be modeled by a Gaussian or multivariate t distribution
- The copula approach allows to model dependencies and marginal distributions separately.
- Marginal time dependencies can be captured by appropriate univariate time series models.
- Elliptical and Archimedean copulas do not allow for different dependency patterns between pairs of variables.

Vine based models can overcome all these shortcomings.

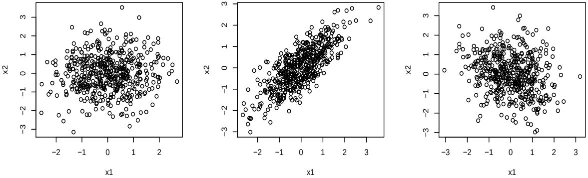
Overview

Multivariate distributions

- Multivariate distributions describe stochastic behavior of several variables jointly.
- Marginal distributions describe stochastic behavior of a single variable (examples: univariate normal, exponential)



- Often used: multivariate normal (left: $\rho = 0$, middle: $\rho = .8$, right:$\rho = -.5$)



How to construct multivariate distributions with different margins?

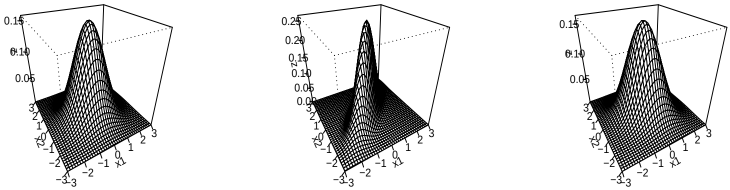
Dependency measures

- Most well known dependency measure is the correlation ρ between two random variables.
- It only measures linear dependencies.
- Non linear dependencies can be detected by Kendall's τ which measures the difference between the concordance and discordance probability.
- Upper (lower) tail dependence measures the probability of joint large (small) occurrences as one moves to the extremes.
- multivariate normal has no tail dependence, while the multivariate t distribution has tail dependence.
- When upper and lower tail dependence are not the same we speak of asymmetric tail dependence.

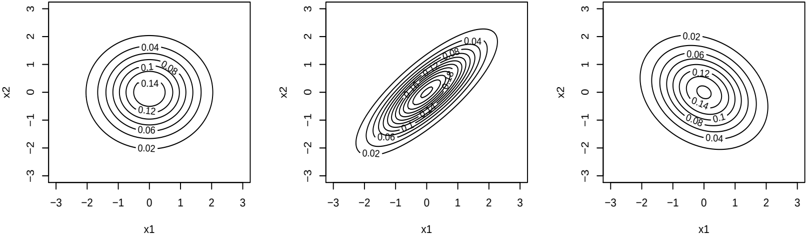
How to separate dependency patterns from the marginal behavior?

Joint density and contour plots

joint density plot (right: $\rho = 0$, middle: $\rho = .8$, left: $\rho = -.25$)



contour plot



Conditional distributions

- vine distributions are defined using conditional distributions
- conditional distributions describe the stochastic behaviour of variables under the condition that other variables are fixed.
- conditional = unconditional distributions if variables are independent

Conditional density of (X_i, X_j) given that $X_k = x_k$

$$f_{i,j|k}(x_i, x_j|x_k) := \frac{f_{ijk}(x_i, x_j, x_k)}{f_k(x_k)}$$

Copula approach

Consider n random variables $\mathbf{X} = (X_1, \ldots, X_n)$ with

	pdf	**cdf**
marginal	$f_i(x_i), i = 1, \ldots, n$	$F_i(x_i), i = 1, \ldots, n$
joint	$f(x_1, \ldots, x_n)$	$F(x_1, \ldots, x_n)$
conditional	$f(\cdot\|\cdot)$	$F(\cdot\|\cdot)$

Copula

A copula with $C(u_1, \ldots, u_n)$ and copula density $c(u_1, \ldots, u_n)$ is a multivariate distribution on $[0, 1]^n$ with uniformly distributed marginals.

Sklar's Theorem (1959) for n=2

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \quad (1)$$

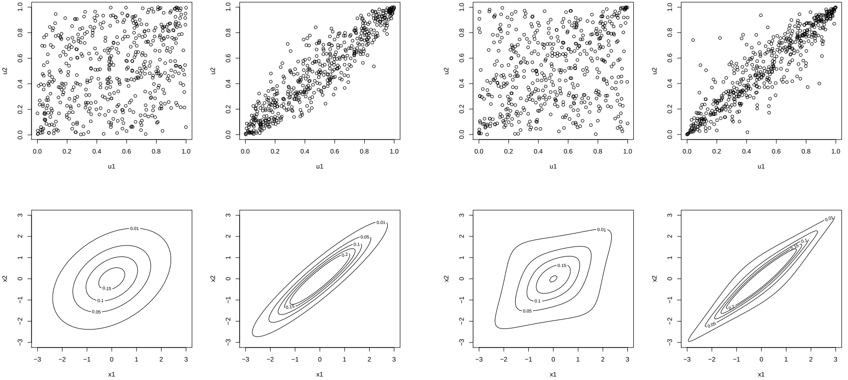
$$f(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2)$$

for some bivariate copula density $c_{12}(\cdot)$ such as normal, t-, Clayton and Gumbel .

Bivariate elliptical copula families

Gaussian copula
(left $\tau = .25$, right: $\tau = .75$)

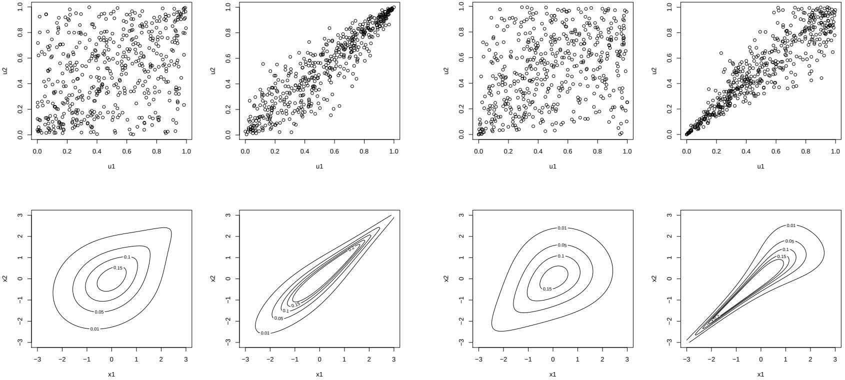
t-copula with $df = 3$
(left $\tau = .25$, right: $\tau = .75$)



Bivariate Archimedian copula families

Gumbel copula
(left $\tau = .25$, right: $\tau = .75$)

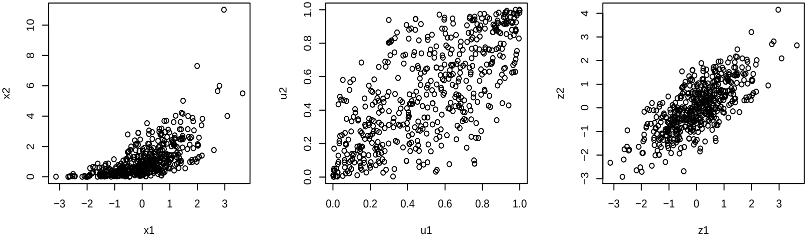
Clayton copula
(left $\tau = .25$, right: $\tau = .75$)



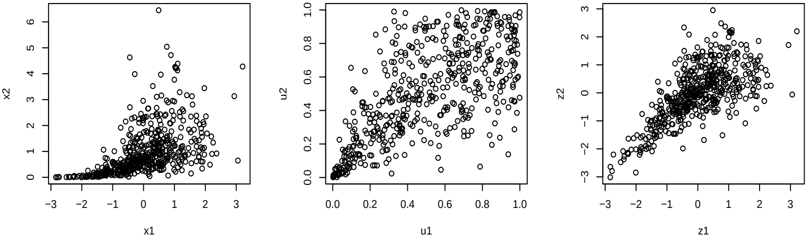
Meta distributions

are build using a copula (u_1, u_2) and different margins (normal/exponential (x_1, x_2) or normal/normal (z_1, z_2))

Gaussian copula



Clayton copula



Pair-copula constructions in 3 dimensions

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1)$$

Using Sklar for $f(x_1, x_2)$, $f(x_2, x_3)$ and $f_{13|2}(x_1, x_3|x_2)$ implies

$$\begin{aligned}
f_{2|1}(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2)) f_2(x_2) \\
f_{3|12}(x_3|x_1, x_2) &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) f_{3|2}(x_3|x_2) \\
&= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{23}(F_2(x_2), F_3(x_3)) f_3(x_3)
\end{aligned}$$

$$\begin{aligned}
f(x_1, x_2, x_3) &= c_{13|2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)) c_{23}(F_2(x_2), F_3(x_3)) \\
&\times c_{12}(F_1(x_1), F_2(x_2)) \\
&\times f_3(x_3) f_2(x_2) f_1(x_1)
\end{aligned}$$

Only bivariate copulas and univariate conditional cdf's are used. This can be easily generalized to n dimensions.

Regular vine distributions

- Many PCC's are feasible. Bedford and Cooke (2002) introduced a graphical structure to help organize them.
- Gaussian vines were analyzed in Kurowicka and Cooke (2006) while estimation for Non Gaussian ones started with Aas et al. (2009).
- Pair copulas model (un)conditional dependencies between two variables.
- A parametric regular vine distribution $R(\mathcal{V}, \mathcal{C}, \boldsymbol{\theta})$ with specified margins has three components:

Components of a regular vine distribution

1. tree structure (set of linked trees) $\mathcal{V}$
2. Parametric bivariate copulas $\mathcal{C} = \mathcal{C}(\mathcal{V})$ for each edge in the tree structure
3. Corresponding parameter value $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathcal{C}(\mathcal{V}))$

Regular vine tree structure

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

Vine tree structure (Bedford and Cooke (2002))

- Tree T_1 is a tree on nodes 1 to n.
- Tree T_j has $n+1-j$ nodes and $n-j$ 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 cases:

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

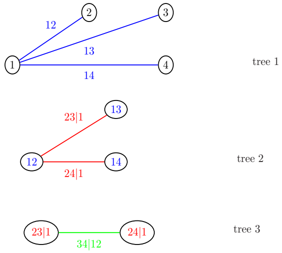
C and D-vines

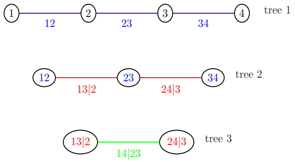
C-vine: each tree has a unique node connected to $n-j$ edges

$$f_{1234} = [\prod_{i=1}^{4} f_i] \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{34|12}$$

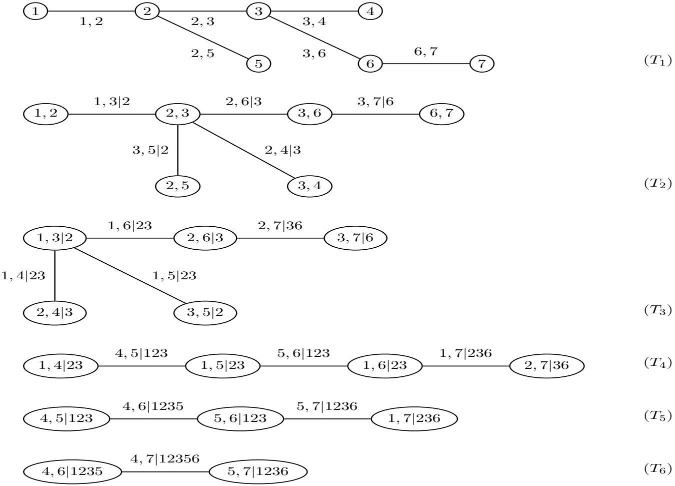
D-vine: no node is connected to more than 2 edges

$$f_{1234} = [\prod_{i=1}^{4} f_i] \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{14|23}$$





A seven dimensional regular vine tree structure



Storing regular vines specifications in matrices

R-vine matrix (Morales-Napoles (2008),Dissmann (2010)):

$$M = \begin{pmatrix} 4 & & & & & & \\ 7 & 5 & & & & & \\ 6 & 7 & 1 & & & & \\ 5 & 6 & 7 & 7 & & & \\ 1 & 1 & 6 & 2 & 6 & & \\ 2 & 3 & 3 & 3 & 2 & 2 & \\ 3 & 2 & 2 & 6 & 3 & 3 & 3 \end{pmatrix}$$

Indices for pair-copulas in corresponding R-vine distribution:

col 1	col 2	col 3	col 4	col 5	col 6
$4,7\|6,5,1,2,3$	$5,7\|6,1,3,2$	$1,7\|6,2,3$	$7,2\|3,6$	$6,2\|3$	$2,3$
$4,6\|5,1,2,3$	$5,6\|1,3,2$	$1,6\|3,2$	$7,3\|6$	$6,3$	
$4,5\|1,2,3$	$5,1\|3,2$	$1,3\|2$	$7,6$		
$4,1\|2,3$	$5,3\|2$	$1,2$			
$4,2\|3$	$5,2$				
$4,3$					

Conditional cdf's

For $\mathbf{v} = (v_1, \ldots, v_n)$ and $\mathbf{v}_{-j} = (v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n)$ $j = 1, \ldots, d$

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

Univariate v:

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

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

General v: Joe (1996)

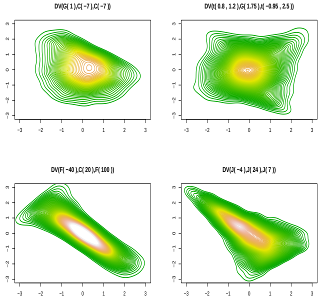
$$F(x|\mathbf{v}) = \frac{\partial\, C_{x,v_j|\mathbf{v}_{-j}}(F(x|\mathbf{v}_{-j}), F(v_j|\mathbf{v}_{-j}))}{\partial F(v_j|\mathbf{v}_{-j})}$$

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

Scope of the vine models

- The following copula classes are vine copulas
  - multivariate Gaussian copula
  - multivariate t copula
  - multivariate Clayton copula (Takahasi (1965))
- The number of different vine tree structures is huge (see Morales-Nápoles et al. (2010)), additional flexibility through choice of copula families.

Contours of bivariate 13 margins with standard normal margins



(C=Clayton, G=Gumbel, t=Student, F=Frank, J=Joe)

Efficient estimation and model selection are vital

Parameter estimation for given 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. (2011)).
  - Asymptotic theory is available (Haff (2010)), 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 et al. (2012) for R-vines)
  - Prior beliefs can be incorporated and credible intervals allow to assess uncertainty for all quantities.

Sequential and ML estimation for PCC's (n=3)

Parameters: $\Theta = (\Theta_{12}, \Theta_{23}, \Theta_{13|2})$
Observations: $\{(x_{1t}, x_{2t}, x_{3t}), t = 1, \cdots, T\}$

Sequential estimates:

Estimate

- Estimate Θ_{12} from $\{(x_{1,t}, x_{2,t}), t = 1, \cdots, T\}$
- Estimate Θ_{23} from $\{(x_{2,t}, x_{3,t}), t = 1, \cdots, T\}$.
- Define pseudo observations

$$\hat{v}_{1|2t} := F(x_{1t}|x_{2t}, \hat{\Theta}_{12}) \text{ and } \hat{v}_{3|2t} := F(x_{2t}|x_{3t}, \hat{\Theta}_{23})$$

Finally estimate $\Theta_{13|2}$ from $\{(\hat{v}_{1|2t}, \hat{v}_{3|2t}), t = 1, \cdots, T\}$.

Maximum likelihood

$$\begin{aligned} L(\Theta|x) &= \sum_{t=1}^{T} [\log c_{12}(x_{1t}, x_{2t}|\Theta_{12}) + \log c_{23}(x_{2t}, x_{3t}|\Theta_{23}) \\ &+ \log c_{13|2}(F(x_{1t}|x_{2t}, \Theta_{12}), F(x_{2t}|x_{3t}, \Theta_{23})|\Theta_{13|2})] \end{aligned}$$

Joint estimation of tree structure, pair copula families and parameters: Sequential approaches

- **Classical approach (Dißmann et al. (2011))**
  - 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 pseudo obs.
- **Bayesian approach**
  - 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 et al. (2012) developed a sequential RJMCMC choosing tree structure, copula families and parameters jointly for T_1 and then fixes the specification for the most sampled T_1 before proceeding to T_2, etc.

Full Bayesian approaches

Update mechanism for full RJ MCMC and simulated annealing:

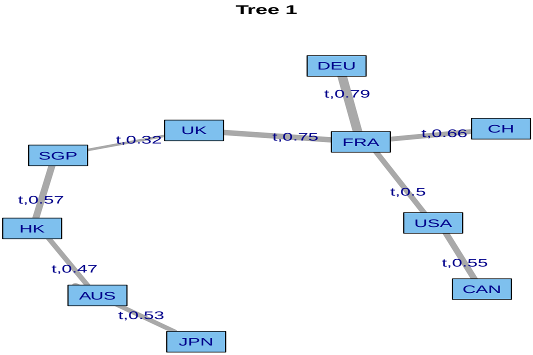
- Choose randomly $k \in \{1, \ldots, n-1\}$ to update trees $T_k, \ldots, T_{n-1}$:
- Generate proposal tree-by-tree: for $T_i \in \{T_k, \ldots, T_{n-1}\}$:
  - Propose tree structure for T_i:
    - random walk: remove randomly one edge and add randomly one edge which is allowed by proximity
    - independent: propose arbitrarily one allowed tree
  - Propose pair copula families for T_i and the corresponding copula parameters using centered parameter proposals at MLE of copula parameter for each pair
- Accept or reject the joint proposal for trees $T_k, \ldots, T_{n-1}$ with acceptance probability α:
  - use MH ratio for RJ MCMC
  - use cooling acceptance probability for simulated annealing

Application 1: National indices

- 10 national indices: AUS, CAN, CH, DEU, FRA, HK, JPN, SGP, UK, USA
- dates: Jan 2008 until June 2011 (757 daily observations)
- marginal time dependencies: AR(1)-GARCH(1,1) with t innovations
- allowed pair-copula families: Clayton, Frank, Gaussian, Gumbel, Joe, independence, and Student's t

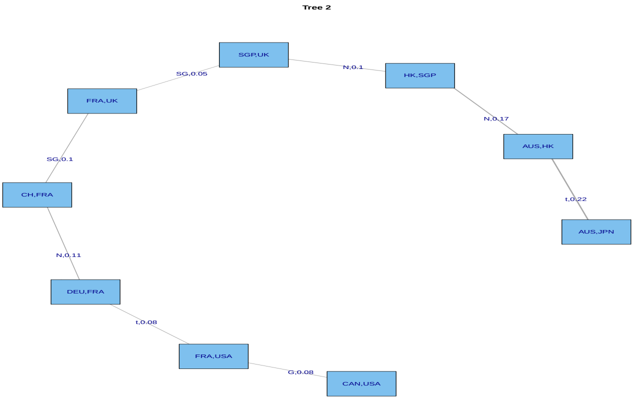
Classical sequential approach with independence tests (I)

First tree: all t-copulas with df between 5 and 14, Kendall's τ estimates between .32 and .79



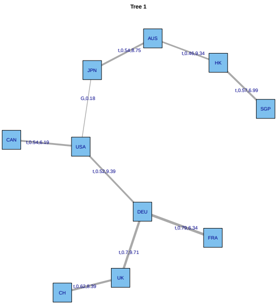
Classical sequential with independence tests (II)

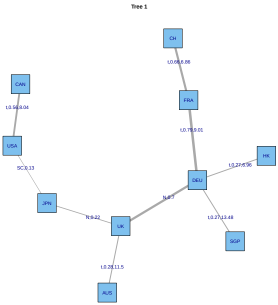
Other trees: few Gumbel, survival Gumbel, Frank, and survival Clayton are used, Kendall's τ estimates vary between .06 and .22, pair copulas on trees 3 and higher can be chosen as independence copula



Bayesian approaches

- Sequential Bayesian approach: same first tree as the classical approach, only one pair-copula family is different, concentrates on 2 first tree structures and about 10 different first tree/copula family combinations using 20000 MCMC iterations per tree
- Simulated annealing (left) and full RJ MCMC (right): different first trees





Model comparison

Log-likelihoods of estimated models:

	# Par.	LL	AIC	BIC
R-vine (sequential, no ind. tests)	61	3,998.0	−7,874.0	−7,591.6
R-vine (sequential, with ind. tests)	35	3,958.9	−7,847.8	−7,685.8
R-vine (full RJMCMC)	55	3,964.0	−7,818.0	−7,563.4
R-vine (simulated annealing)	58	3,996.0	−7,876.0	−7,607.5
Non-Gaussian DAG (part. corr.)	30	3,784.6	−7,509.2	−7,370.3
Non-Gaussian DAG (vine-based)	27	3,772.7	−7,491.4	−7,366.4
Gaussian DAG (part. corr.)	18	3,716.6	−7,397.2	−7,313.9
Gaussian DAG (vine-based)	16	3,708.7	−7,385.4	−7,311.3

For the Bayesian approaches the posterior mode model and estimates are used.

Application 2: S&P select sector indices

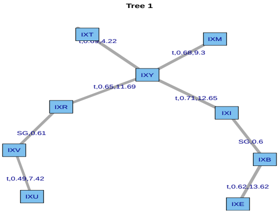
Gruber et al. (2012) use daily log returns from 9 sector indices: 300 trading days before (bear) and after (bull) March 9, 2009 (S&P 500 low)

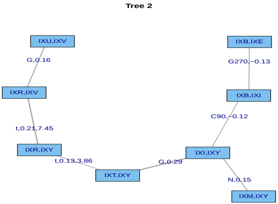
index	S&P code	index name
1	IXB	Materials Select Sector Index
2	IXE	Energy Select Sector Index
3	IXI	Industrial Select Sector Index
4	IXM	Financial Select Sector Index
5	IXR	Consumer Staples Select Sector Index
6	IXT	Technology Select Sector Index
7	IXU	Utilities Select Sector Index
8	IXV	Health Care Select Sector Index
9	IXY	Consumer Discretionary Select Sector Index

Uni. MA(1)-GARCH(1,1) with Gauss innovations remove marginal time dependencies. Copula data formed by using the emp. prob. transform.

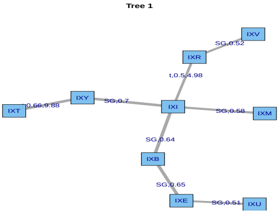
Most sampled models using sequential RJMCMC

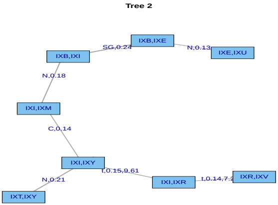
bear market: consumer discretionary (IXY) in center





bull market: industrial (IXI) in center





Further results and comparisons:

- There is evidence of many nonsymmetric dependencies and lower tail dependence for some pairs
- Asymmetry and lower tail dependence occurs more often in the bull market compared to the bear market
- The R-vine obtained by the Dißmann et al. (2011) appraoch agrees in first tree for bear market, while the trees agree for the first two trees for the bull market.
- More independence copulas are chosen by the Dißmann et al. (2011) approach compared to the sequential RJMCMC.
- Strength of dependencies go down as the number of conditioning variables go up.

Special vine models (I)

- vine copulas with time varying parameters:
  - Almeida and Czado (2011) and Almeida et al. (2012) allow an AR(1) driven copula dynamics
  - Almeida and Czado (2011) develops a bivariate Bayesian approach with credible intervals, while Almeida et al. (2012)) use simulated ML and apply it to the stocks of the DAX (29 dim)
  - regime switching vine models were considered by Chollete et al. (2008) and Stöber and Czado (2011)
  - Stöber and Czado (2011) determines crisis and non crisis regime through rolling windows.
- truncated and simplified R-vines:
  - Heinen and Valdesogo (2009) use simplified C-vines in high dimensions
  - Brechmann et al. (2012) derive test to determine truncation level
  - Brechmann and Czado (2011) develops vine sector models
  - Brechmann and Czado (2012b) use a vine based model with VAR backtesting

Special vine models (II)

- Bauer et al. (2012) develop and fit Non Gaussian directed acyclic graphical (DAG) models based on PCC's, first selection methods for building up the DAG graph are developed.
- discrete vine copulas are treated in Panagiotelis et al. (2011)
- Brechmann and Czado (2012a) develop an R-vine model which can capture both between as well as serial dependencies.
- Bernard and Czado (2010) use an R-vine to price multivariate options
- Dependencies between claim numbers and sizes in different insurance risk categories are modeled in Erhardt and Czado (2010)

Summary and extensions

- PCC's such as C-, D- and R-vines are very flexible.
- Sequential and MLE parameter estimation of C and D-vines are available in **R package** `CDVine`.
- Sequential and full Bayesian and non Bayesian model selection of vine trees and copula families for regular vines available, but need further testing and development
- Extensions for the future:
  - use of non parametric pair copulas
  - development of spatial vines
  - vines in data mining
- **Vine resource page:** `www-m4.ma.tum.de/forschung/vine-copula-models`
- **Vine workshop book:** Kurowicka and Joe (2011)

- **Thanks to my collaborators** (K. Aas, A. Frigessi , A. Min, E. Brechmann, C. Almeida, M. Smith, A. Panagiotelis, A. Bauer, T. Klein, M. Hofmann, J. Dißmann, H. Joe, J. Stöber, U. Schepsmeier, D. Kurowicka, L. Gruber...)
- Next workshop: **Copulae in Mathematical and Quantitative Finance (Krakow, July 10-11, 2012)**

  `worcotha.mimuw.edu.pl/index.htm`
- **Summerschool for Ph.D. students (Garching 30.7-3.8.2012)** http://www.ma.tum.de/Mathematik/IsamSummerSchool12

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

Almeida, C. and C. Czado (2011).
Efficient Bayesian inference for stochastic time-varying copula model.
to appear in CSDA.

Almeida, C., C. Czado, and H. Manner (2012).
Modeling high dimensional time-varying dependence using d-vine scar models.
preprint.

Bauer, A., C. Czado, and T. Klein (2012).
Pair-copula constructions for non-Gaussian DAG models.
Canadian Journal of Statistics 40, 86–109.

Bedford, T. and R. M. Cooke (2002).
Vines - a new graphical model for dependent random variables.
Annals of Statistics 30(4), 1031–1068.

Bernard, C. and C. Czado (2010).
Multivariate option pricing using copulas.
in revision.

Brechmann, E. and C. Czado (2012a).
Copar - multivariate time series modeling using the copula autoregressive model.
preprint.

Brechmann, E. and C. Czado (2012b).
Risk management with high-dimensional vine copulas: An analysis of the euro stoxx 50.
preprint.

Brechmann, E., C. Czado, and K. Aas (2012).
Truncated regular vines in high dimensions with application to financial data.
Canadian Journal of Statistics 40, 68–85.

Brechmann, E. C. and C. Czado (2011).
Extending the CAPM using pair copulas: The Regular Vine Market Sector model.
Submitted for publication.

Chollete, L., A. Heinen, and A. Valdesogo (2008).
Modeling international financial returns with a multivariate regime switching copula.
Preprint.

Czado, C., U. Schepsmeier, and A. Min (2011).
Maximum likelihood estimation of mixed c-vine pair copula with application to exchange rates.
to appear in Statistical Modeling.

Dißmann, J., E. Brechmann, C. Czado, and D. Kurowicka (2011).
Selecting and estimating regualr vine copulae and application to financial returns.
in revision.

Dissmann, J. F. (2010).
Statistical inference for regular vines and application.
Master's thesis, Technische Universität München.

Erhardt, V. and C. Czado (2010).
Modelling dependent yearly claim totals including zero-claims in private health insurance.
Scandinavian Actuarial Journal, online under DOI: 10.1080/03461238.2010.489762.

Gruber, F., C. Czado, and Stöber (2012).
Bayesian model selection for r-vine copulas using reversible jump mcmc.
preprint.

Haff, I. H. (2010).
Estimating the parameters of a pair copula construction.
preprint.

Heinen, A. and A. Valdesogo (2009).
Asymmetric capm dependence for large dimensions: The canonical vine autoregressive copula model.
Preprint.

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*.

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

Kurowicka, D. and H. Joe (2011).
Dependence Modeling - Handbook on Vine Copulae.
Singapore: World Scientific Publishing Co.

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

Morales-Napoles, O. (2008).
Bayesian belief nets and vines in aviation safety and other applications.
Ph. D. thesis, Technische Universiteit Delft.

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

Panagiotelis, A., C. Czado, and H. Joe (2011).
Pair copula constructions for cultivariate discrete data.
to appear in JASA.

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. and C. Czado (2011).
Detecting regime switches in the dependence stucture of high dimensional financial data.
submitted preprint.

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