The world of vines

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

Lehrstuhl für Mathematische Statistik

4th Vine Workshop, Munich, May 2011

Overview

Motivations for vines

- Many multivariate data structures exhibit
  - different marginal distributions
  - nonsymmetric dependencies between some pairs of variables
  - heavy tail dependencies between some pairs of variables
- These cannot be modeled with standard parametric distributions such as the Gaussian ormultivariate t distribution
- The copula approach allows to model dependencies and marginal distributions separately.
- However standard multivariate copula models such as the elliptical and Archimedean copulas do not allow for different dependency models between pairs of variables.

Vine models can overcome all these shortcomings.

Some history of vine models

- Joe (1996) gave a probabilistic construction of multivariate distributions functions based on simple building blocks called pair-copulas.
- Bedford and Cooke (2001) and Bedford and Cooke (2002) organized these constructions in a graphical way called regular vines and gave expression for the joint density.
- Estimation for the Gaussian case was considered in the book by Kurowicka and Cooke (2006).
- Aas et al. (2009) used the PCC construction to construct flexible multivariate copulas based on pair-copulas such as bivariate Gaussian, t-, Gumbel and Clayton copulas and provided likelihood expressions.
- First and second vine workshops took place in Delft in Nov. 2007 and Dec. 2008, a third one took place in Oslo in Dec. 2009. Workshop results are published in Kurowicka and Joe (2011).
- A recent survey about PCC models is Czado (2010).

Copula approach

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

- joint pdf $f(x_1, \ldots, x_n)$ and marginal pdf's $f_i(x_i), i = 1, \ldots, n$
- joint cdf $F(x_1, \ldots, x_n)$ and marginal cdf's $F_i(x_i), i = 1, \ldots, n$
- $f(\cdot|\cdot)$ denote corresponding conditional pdf's.
- $F(\cdot|\cdot)$ denote corresponding conditional cdf's.

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

$$\begin{aligned} f(x_1, x_2) &= c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2) \qquad (1) \\ f(x_2|x_1) &= c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \end{aligned}$$

for some bivariate copula density $c_{12}(\cdot)$.

Common bivariate copula distributions

- **Elliptical copulas**
  According to Sklar copulas can be created using multivariate distributions F , i.e.

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2)), u_1, u_2 \in (0, 1)$$

  - **Normal copulas** (derived from bivariate normal with zero means, unit variances and correlation ρ)
  - **t-copulas** (derived from bivariate t-distribution with zero mean, degree of freedom ν and association ρ)
- **Archimedean copulas**
  - **Clayton** $C(u_1, u_2) = (u_1^{-\delta} + u_2^{-\delta} - 1)^{-1/\delta}, \delta > 0$
  - **Gumbel** $C(u_1, u_2) = \exp\left[-\left\{(-\log u_1)^{\delta} + (-\log u_1)^{\delta})\right\}^{-1/\delta}\right], \delta > 1$
- **Reference books:** Joe (1997) and Nelsen (2006)

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

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

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

Pair-copula constructions (PCC) in n dimensions

Factorization

$$f(x_1, \dots, x_n) = [\prod_{t=2}^{n} f(x_t|x_1, \dots, x_{t-1})] \cdot f_1(x_1) \quad (2)$$

For distinct $i, j, i_1, \cdots, i_k$ with $i < j$ and $i_1 < \cdots < i_k$ let

$$c_{i,j|i_1,\cdots,i_k} := c_{i,j|i_1,\cdots,i_k}(F(x_i|x_{i_1}, \cdots, x_{i_k}), (F(x_j|x_{i_1}, \cdots, x_{i_k}))$$

Reexpress $f(x_t|x_1, \cdots, x_{t-1})$ as

$$\begin{aligned} f(x_t|x_1, \cdots, x_{t-1}) &= c_{1,t|2,\cdots,t-1} \times f(x_t|x_2, \cdots, x_{t-1}) \\ &= [\prod_{s=1}^{t-2} c_{s,t|s+1,\cdots,t-1}] \times c_{(t-1),t} \times f_t(x_t) \end{aligned}$$

PCC decomposition

Using (2) and $s = i, t = i + j$ it follows that

$$f(x_1, \dots, x_n) = [\prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,(i+j)|(i+1),\cdots,(i+j-1)}] \cdot [\prod_{k=1}^{n} f_k(x_k)] \quad (3)$$

- A decomposition such as (3) is called a pair copula decomposition (PCC). There are many.
- Bedford and Cooke (2001) introduced a graphical structure called regular vine tree structure to help organize them.

Regular vine distribution

An n-dimensional vine tree structure is a sequence of n-1 trees

- Tree j has $n+1-j$ nodes and $n-j$ edges.
- Edges in tree j become nodes in tree $j+1$.
- **Proximity condition:** Two nodes in tree $j+1$ are joined by an edge if the corresponding edges in tree j share a node.

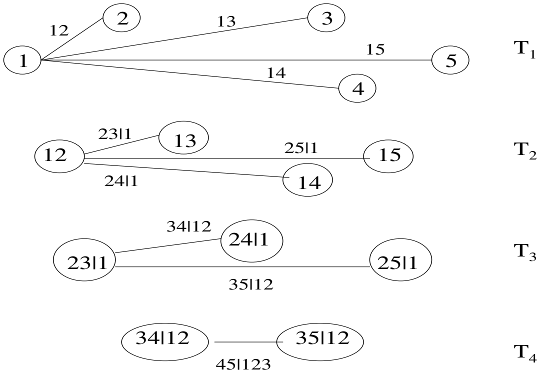
A regular vine distribution is defined by

- A regular vine tree structure
- Each edge corresponds to a pair-copula density.
- The density of a regular vine distribution is defined by the product of pair copula densities over the $\frac{n(n-1)}{2}$ edges identified by the regular vine tree structure and the product of the marginal densities.

Canonical vine distributions

are regular vine distribution for which each tree has a unique node that is connected to $n - j$ edges.

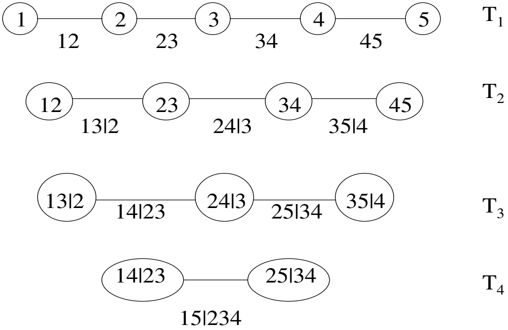
$$f_{12345} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \cdot c_{12} \cdot c_{13} \cdot c_{14} \cdot c_{15} \cdot c_{23|1} \cdot c_{24|1} \cdot c_{25|1} \cdot c_{34|12} \cdot c_{35|12} \cdot c_{45|123}$$



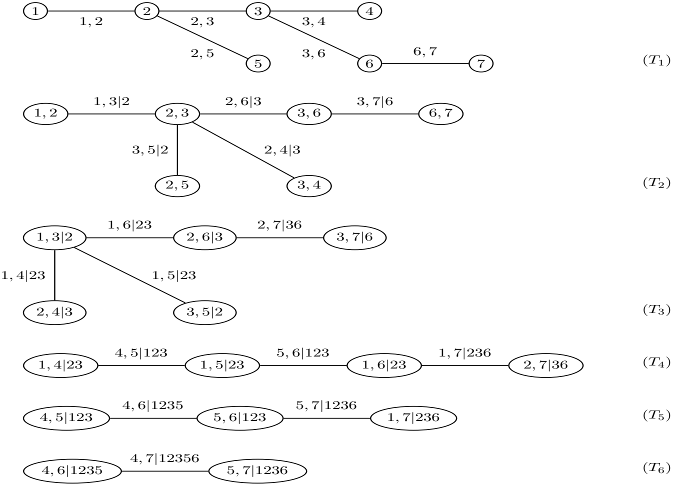
D-vine distributions

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

$$f_{12345} = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \cdot c_{12} \cdot c_{23} \cdot c_{34} \cdot c_{45} \cdot c_{13|2} \cdot c_{24|3} \cdot c_{35|4} \cdot c_{14|23} \cdot c_{25|34} \cdot c_{15|234}$$



A seven dimensional regular vine tree structure



Storing regular vines specifications in matrices

R-vine matrix

(Morales-Napoles (2008),Dissmann (2010)):

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

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

col 1	col 2	col 3	col 4	col 5	col 6
$4,7\|6,5,1,2,3$	$5,7\|6,1,3,2$	$1,7\|6,2,3$	$7,2\|3,6$	$6,2\|3$	$2,3$
$4,6\|5,1,2,3$	$5,6\|1,3,2$	$1,6\|3,2$	$7,3\|6$	$6,3$	
$\color{red}{4,5}\|\color{blue}{1,2,3}$	$5,1\|3,2$	$1,3\|2$	$7,6$		
$4,1\|2,3$	$5,3\|2$	$1,2$			
$4,2\|3$	$5,2$				
$4,3$					

Conditional cdf's

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

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

Univariate v:

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

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

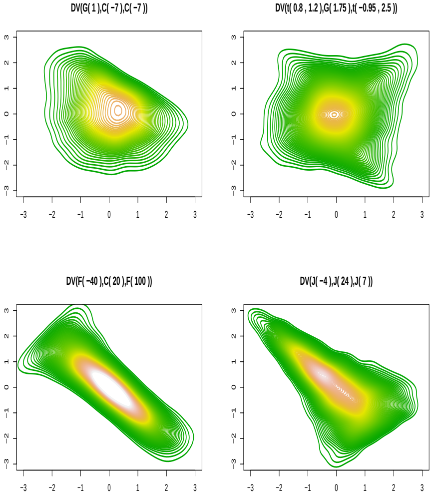
General v:

Under regularity conditions Joe (1996) showed that

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

Illustration of 3 dim. D/C vine

Contours of bivariate 13 margins with standard normal margins after integration (C=Clayton, G=Gumbel, t=Student, F=Frank, J=Joe)



Scope of the vine models

- For simplicity the parameters of the pair-copulas are chosen to be independent of conditioning values. The arguments however depend on the conditioning values.
- Haff et al. (2010) and Stöber and Czado (2011) give some examples where the pair copula parameters depend on the specific conditioning values
- Haff et al. (2010) show that this restriction is not severe in examples.
- The following copula classes are vine copulas
  - multivariate Gaussian copula
  - multivariate t copula
  - multivariate Clayton copula (Takahashi (1965), Stöber and Czado (2011))
- The number of different vine tree structures is huge, see Morales-Nápoles et al. (2010)
- Flexibility is added by allowing for different pair copula families.

Efficient estimation and model selection methods are vital

Estimation methods for PCC's

- **Sequential estimation:**
  - Here the parameters are sequentially estimated starting from the top tree structure.
  - Asymptotic theory is available (Haff (2010)), however analytical standard errors are difficult to compute.
  - Sequential estimates can be used as starting values for maximum likelihood
- **Maximum likelihood estimation:**
  - Asymptotically efficient under regularity conditions.
  - Estimates of standard errors can be based on inverse Hessian matrix
  - Numerical problems for large dimensions, i.e. negative variance estimates might occur
  - Uncertainty in value-at-risk (high quantiles) is difficult to assess
- **Bayesian estimation:**
  - Bayesian estimation is facilitated used Markov Chain Monte Carlo (MCMC) methods based on Metroplis-Hastings steps.
  - Credible interval estimates provide uncertainty assessment for parameter estimates, dependence estimates and value-at-risk estimates.

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

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

Model selection: early approaches

- For data components ordered sequentially (e.g. time) use D-vine with same order
- For $n \leq 4$ and single pair copula type all models are fitted
- Restrict to either D vines or C vines and single pair copula type
- Select D order such that all or most of the strongest pair wise dependencies are contained in the first tree (Aas et al. (2009))
- Brechmann (2010) showed in simulation that AIC works well for pair copula type selection

Model selection: more advanced approaches

- Bayesian model selection strategies were used in D-vines with single pair copula type by Min and Czado (2009) and Smith et al. (2010).
- Select tree structure, pair copula types and their parameters for R-vines sequentially using maximal spanning tree algorithms with Kendall's τ as weights and sequential estimation from top tree until the last tree (see later example from Dißmann et al. (2011)).
- For D-vines a Hamiltonian path has be to found, i.e. a traveling salesman problem has to be solved.
- For C-vines a root node in each tree which maximizes the sum of absolute Kendall's τ is found (Czado et al. 2011).
- Kurowicka (2011) starts building trees from last tree to top tree by using empirical partial correlations as approximate measure of pairwise dependence
- Brechmann et al. (2010) searches for truncated vines (independent pair-copulas for later trees) by using Vuong (1989) tests to select truncation level

Special vine models

- vine copulas with time varying parameters
- regime switching vine models
- vine copulas with non parametric pair copulas
- Non Gaussian directed acyclic graphical (DAG) models based on PCC's
- discrete vine copulas
- truncated and simplified R-vines
- spatial vines

Applications

- **Application dimensions:**
  - Gaussian vines in arbitrary dimensions (Kurowicka and Cooke 2006)
  - First non Gaussian D-vine models using joint ML were in 4 dimensions
  - Bayesian D-vine applications in 7 and 12 dimensions with credible intervals
  - Joint ML now feasible in 50 dimensions for R-vines
  - Sequential and joint ML estimation in truncated R-vines and regular vine market sector models in 50 dim. (Brechmann and Czado (2011))
  - Heinen and Valdesogo (2009) propose and sequentially fit a canonical vine autoregressive model in 100 dimensions
- **Application areas:**
  - finance: indices, exchange rates, multivariate options,order books
  - insurance: number of claims, claim size in different categories
  - genetics: gene interactions
  - health: severity of headaches
  - hydrology:
  - images: radar polarimetric data estimates land use

Building R-vine based models for financial Stock Indices

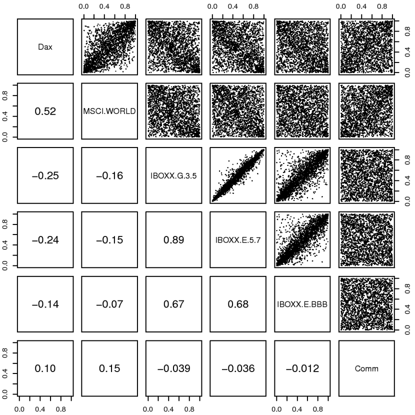
- Investigated 16 indices including 5 equity, 9 fixed income and 2 commodity indices recorded daily between Dec 29, 2001 and Dec 14, 2009
- Data is not an i.i.d. sample, therefore marginal and copula parameters have to be estimated.

Two-step estimation of marginal and copula parameters

- Estimate margins using ARMA(1,1)-GARCH(1,1) model with Gaussian, t and skewed t-innovation for some currencies to form standardized residuals
- We use ranks of standardized residuals to transform to copula data

Reference: Dissmann (2010) and Dißmann et al. (2011)

Pairs plots and Kendall's τ for representatives



$$\text{Kendall's } \tau := 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1$$

Finding Regular Vine Tree Structures

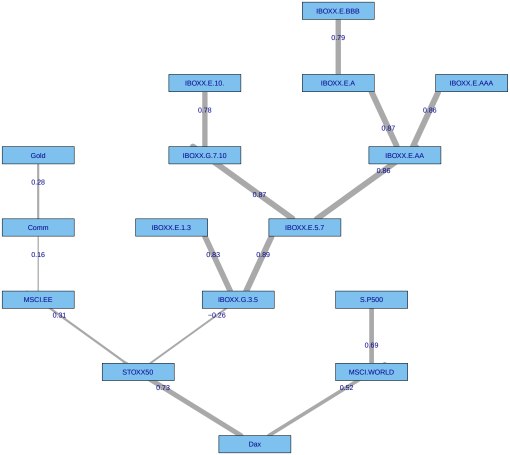
1. Using the maximal spanning tree algorithm of Prim Dißmann et al. (2011) find the first regular vine tree with maximal sum of empirical Kendall's τ values based on pairs.
2. The copula family types (Gauss, t, Gumbel and rotated versions, Frank) for each selected pair in the first tree will be determined by using AIC (see Brechmann (2010))
3. For next tree consider all edges which do not violate the proximity condition. Generate all necessary pseudo observations corresponding to a possible edge and estimate the corresponding Kendall's τ.
4. A test for independence based on Kendall's τ can be used to replace pair copulas by the independence copula.
5. Apply the first step again until all trees and their pair copulas are determined. This will also provide sequential estimates.

Dißmann et al. (2011) provide algorithmic expressions for the joint likelihood of an R-vine and use them for ML estimation.

Investigated vine models

- **mixed R-vine**: R-vine with pair-copula terms chosen individually from 7 bivariate copula types (Gauss, Student-t, Gumbel, survival Gumbel, rotated Gumbel (90 and 270 degrees), Frank).
- **R-vine indep**: Same as mixed R-vine but pair copulas replaced by independence copula, when indicated by independence test based on Kendall's τ.
- **mixed C-vine**: C-vine with pair-copula terms chosen individually from 7 bivariate copula types (see above).
- **mixed D-vine**: D-vine with pair-copula terms chosen individually from 7 bivariate copula types (see above).
- **all t R-vine**: R-vine with each pair-copula term chosen as bivariate Student-t copula (If the df > 30, then replaced with Gaussian.)
- **multivariate Gauss**: R-vine with each pair-copula term chosen as bivariate Gaussian copula,

First regular vine tree for financial indices data



Results

	R-vine mixed	R-vine all t	R-vine all Gauss	R-vine indep.	C-vine mixed	D-vine mixed
Seq. log likelihood	36431.22	36417.35	30445.47	36331.86	36366.89	36300.51
Log likelihood	36514.03	36513.44	31784.07	36396.80	36476.36	36422.53
# Parameters	171	179	120	108	178	176
Indep.	0	0	0	55	0	0
Gauss	16	61	120	8	19	18
Student-t	51	59	0	43	58	56
Gumbel	4	0	0	1	8	7
Surv. Gumbel	7	0	0	1	8	6
Rot. Gumbel	12	0	0	2	11	9
Frank	30	0	0	10	16	24

Using Vuong (1989) tests with Schwartz correction show that mixed R-vine is preferred over all other vine models. A further improvement is visible when independence pair copulas are allowed.

Mixed R-vines are needed

Summary and extensions

- PCC's such as C-, D- and R-vines allow for very flexible class of multivariate distributions
- Efficient parameter estimation methods are available for dimensions up to 50
- Model selection of vine tree structures and pair copula types for regular vines still needs further work
- Efficient distance measures between vine distributions would be useful
- Fast forecasting methods are needed for non Gaussian vine distributions
- Use of Non Gaussian vine models in data mining
- Substantial applications

- **Thank you for your attention and I hope you enjoyed your visit to the world of vines**
- **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 ...)

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 (2001).
Probability density decomposition for conditionally dependent random variables modeled by vines.
Annals of Mathematics and Artificial Intelligence 32, 245–268.

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.

Brechmann, E., C. Czado, and K. Aas (2010).
Truncated regular vines in high dimensions with application to financial data.
preprint.

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

Czado, C. (2010).
Pair-copula constructions of multivariate copulas.
In F. Durante, W. Härdle, P. Jaworki, and T. Rychlik (Eds.), *Workshop on Copula Theory and its Applications*. Springer, Dortrech.

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.
preprint.

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

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

Haff, I. H., K. Aas, and A. Frigessi (2010).
On the simplified pair-copula construction – simply useful or too simplistic?
Journal of Multivariate Analysis 101(5), 1296 – 1310.

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*.

Joe, H. (1997).
Multivariate Models and Dependence Concepts.
London: Chapman & Hall.

Kurowicka, D. (2011).
Optimal truncation of vines.
In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Handbook on Vine Copulae*. Singapore: World Scientific Publishing Co.

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 (2009).
Bayesian model selection for multivariate copulas using pair-copula constructions.
to appear in *Canadian Journal of Statistics*.

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.

Nelsen, R. (2006).
An Introduction to Copulas.
New York: Springer.

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).
Characterization results with respect to vine copulas.
in preparation.

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

Vuong, Q. H. (1989).
Likelihood ratio tests for model selection and non-nested hypotheses.
Econometrica 57, 307–333.