

Copula and Vine Modeling for Finance (stat08013)

Luciana Dalla Valle, University of Plymouth, UK *

Abstract

This article gives an introduction to copulas, pair copula constructions and vines and presents some examples of their applications to financial problems.

Keywords: *Dependence, Copula, Credit Risk, Finance, Operational Risk, Pair Copula Constructions, Probability of Default, Sklar's Theorem, Small and Medium-sized Enterprises, Vines.*

*email: luciana.dallavalle@plymouth.ac.uk

1 Introduction to Copulas

Copulas are popular modeling approaches in multivariate statistics allowing the separation of the marginal components of a joint distribution from its dependence structure (see **stat00943**). Copulas allow the calculation of the joint multivariate distribution from the marginal distributions, incorporating their dependence structure. The main advantage of copulas is their flexibility in describing departures from normality, with potentially any type of distribution modeling the marginals and with various classes of copulas accommodating complex dependence patterns.

Sklar [15] proved that a d -dimensional distribution F of the random variables X_1, \dots, X_d can be fully described by its marginal distributions and a function C , called copula, through the relation

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

2 Copula Applications to Finance

In the literature, copulas have been applied to model the dependence between variables in a wide variety of fields. The financial field has benefited from the use of copulas with significant advances especially in the area of risk management (see **stat00014**, **stat04580**, **stat04704**). Motivated by the requirements of the Basel Committee on Banking Supervision (see **stat03550**), several contributions used copulas to efficiently model operational risk (see **stat03758**). The main aim of operational risk models is to evaluate the amount of capital to be set aside to cover unexpected operational losses. This is typically achieved estimating a loss distribution and deriving functions of interest from it, such as the Value at Risk (VaR) (see **stat03746**). One of the main challenges related to operational risk modeling is the

large number of loss distributions to estimate. While traditional approaches treat these distributions as perfectly dependent, thus overestimating VaR, copula distributions are able to model high dimensional operational risks in a more flexible way, taking into account their dependence structure [11], [10].

Another issue related to operational risk modeling is the lack of loss data, which may affect the parameter estimates of the risk model. The use of Bayesian methods allows the integration of the scarce and, sometimes, inaccurate data collected by the bank with prior information provided by experts. Bayesian inference for copulas permits the calculation of the capital required to cover the estimated operational risks and shows advantages in terms of a reduction of capital charge according to different choices of the marginal loss distributions. Following the approach of Bayesian copulas, operational risks can be evaluated in a more adequate way and the uncertainty of model parameters can be accounted for, which implies also the possibility of reporting credible intervals for risk capital charges based on VaR and other risk measures [8], [7].

In addition to operational risk measurement, copulas have been successfully applied in other areas, such as, for example, the evaluation of the productivity of small and medium-sized enterprises (SMEs). When productivity data is collected through a survey, one of the most common issues is the high proportion of nonresponses that leads to self-selection bias (see **stat03338**, **stat03397**, **stat05290.pub2**). Since typically responders have common characteristics, the sample is biased, hence it does not represent the population it is related to, and the sample distribution of the variables differs from the same variables in the population (see **stat05107**). The Heckman two-step procedure [12] considers two equations tied together by a latent factor that allows the missing data associated with the nonresponding subjects to be estimated. Copulas generalize Heckmans two-step method allowing us to assume different distributions for the marginals and to express various dependence structures. In addition,

Bayesian inference allows the inclusion of prior information into the copula model, facilitating the estimation of small samples. With the use of copulas, self-selection bias can be corrected and SMEs productivity can be estimated properly [5].

3 Pair Copula Constructions and Vines

Whilst for bivariate copulas (named *pair copulas*) there exist a long and varied list of distribution families, the selection of copulas in higher dimensions is greatly reduced with the choice of families limited to the elliptical copulas <stat01214> [13], [14]. Recently pair copula constructions (PCCs) and their graphical representation, called *vines*, have been introduced to overcome the lack of flexibility of copulas in high-dimensional cases [3], [4], [2]. PCCs are a flexible class of multivariate copulas based on their decomposition using bivariate (conditional) copulas as building blocks (see **stat08012**).

Regular vines (or *R-vines*) are a particular type of graphical models, that use a nested set of trees to represent the decomposition of the joint distribution into its bivariate components, incorporating the dependence structure of the variables of interest. The class of regular vines is very general and embraces a large number of possible pair copula decompositions. A vine $\mathcal{V}(d)$ on d variables is a nested set of trees T_1, \dots, T_{d-1} and the edges of tree T_i are the nodes of tree T_{i+1} , $i = 1, \dots, d - 1$. In an R-vine, if two edges of tree T_i share a common node, they are represented in tree T_{i+1} by nodes joined by an edge. Each node corresponds to a variable or a set of variables and a pair copula density is associated to any edge. Two special cases of R-vines are drawable vines (or *D-vines*) and canonical vines (or *C-vines*). A D-vine is an R-vine where each node is connected to no more than two other nodes. A C-vine is an R-vine whose trees are all stars, since each tree T_i has a unique node that is connected to $d - i$ edges.

<Figure 1 near here>

Figure 1 illustrates a vine on $d = 3$ variables, where the trivariate copula is decomposed into the three pair copulas c_{12} , c_{23} and $c_{1,3|2}$, depicted as edges linking the nodes.

4 Vine Applications to Finance

Vines constitute a flexible and very appealing tool for financial analysis [1]. For instance, the successful application of vines to credit risk management contributed to develop accurate models to estimate the probability of default (PD) of firms. Default risk is defined as the risk of a loss when a debtor (e.g. a firm) does not fulfill its commitments in a financial contract, and a default event takes place. The probability of default (PD) is the probability that a default happens (see **stat03755**, **stat03756**, **stat03768**, **stat04695**). The firm value can be modeled via a contingent claim on the underlying securities, given by equity and debt, where the dynamic of the equity is described by a D-vine and calculated using Monte Carlo simulations (see **stat03619**, **stat03876.pub2**). The use of vines allows us to evaluate, at any time t , the firm's ability to service its debts, and consequently to efficiently predict its PD in a flexible way [9].

<Figure 2 near here>

Figure 2 shows an example of 4-dimensional D-vine tree plots used to estimate the PD of a defaulted firm (the Enron company). The left panel shows the first tree, while the right panel shows the second tree of the D-vine. The squares represent nodes that are variables, while the lines represent edges that are dependencies between variables. The names of the variables may be read in the squares (ALT=long term assets, ACT=current term assets, BLT=long term liabilities and BCT=current term liabilities), and the pair copula families

and dependence measures (Kendall's tau) corresponding to pair copula parameters can be read in the edge labels. The thicker the line the higher the dependence between the variables represented by the nodes.

Another noteworthy application of vines is the calibration of companies' sales to integrate different sources of information. When the variables of multidimensional datasets exhibit a high degree of correlation, C-vines can be used to identify the dependence structure and subsequently perform data integration.

<Figure 3 near here>

Figure 3 shows an example of the first two trees of the 9-dimensional C-vine tree plot used for the firm sales data calibration. The variables used to construct the C-vine and corresponding to its nodes, are

- *sales*: firm annual turnover;
- *emp*: average number of employees;
- *rise*: number of managers receiving wage rise;
- *rise2*: number of managers that will receive wage rise in the following year;
- *prom*: number of employees gaining a promotion;
- *horiz*: number of employees involved in horizontal movements;
- *ext*: number of people employed in the external market;
- *grad*: number of newly-graduated employees;
- *qual*: number of newly-qualified employees.

Since in this case *sales* is the target variable dominating the dependencies with all the remaining variables, a C-vine was selected amongst other vine structures, and *sales* was set as the root node. In the first tree of the C-vine, depicted in the left panel of Figure 3, the highest dependencies (thickest arcs) are between the *sales* and the number of employees, and between the *sales* and number of managers receiving wage rise. In the second tree, represented in the right panel of Figure 3, we notice a strong dependence (thick arc) between the central node (*sales, ext*) and the node (*sales, horiz*).

The use of C-vines allows us to model the dependence structure of financial datasets in a flexible way. These results may then be used to calibrate firm sales and perform integration with other financial information [6].

5 Related Articles

stat00014

stat00943

stat01214

stat01215

stat07523

stat03338

stat03397

stat03550

stat03619

stat03684

stat03746

stat03758

stat03755

stat03756

stat03768

stat03876.pub2

stat04393

stat04580

stat04695

stat04704

stat05107

stat05290.pub2

stat07457

stat08012

References

- [1] Aas, K. (2016) Pair-Copula Constructions for Financial Applications: A Review. *Econometrics*, **4**, 43.
- [2] Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009) Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**, 182–198.
- [3] Bedford, T. & Cooke, R.M. (2001) Probability density decomposition for conditionally dependent random variables modeled by vines, *Annals of Mathematics and Artificial Intelligence*, **32**, 245–268.
- [4] Bedford, T. & Cooke, R.M. (2002) Vines - a new graphical model for dependent random variables, *Annals of Statistics*, **30**, 1031–1068.

- [5] Dalla Valle, L. (2016) The Use of Official Statistics in Self-Selection Bias Modeling. *Journal of Official Statistics*, **32**, 887–905.
- [6] Dalla Valle, L. (2014) Official Statistics Data Integration Using Copulas. *Quality Technology and Quantitative Management*, **11**, 111–131.
- [7] Dalla Valle, L. (2010) Measuring Operational Risk in a Bayesian Framework, in *Rethinking Risk Measurement and Reporting. Uncertainty, Bayesian Analysis and Expert Judgement*, K. Böcker, ed, Risk Books, London, chap. 14, pp. 395–422.
- [8] Dalla Valle, L. (2009) Bayesian Copulae Distributions, with Application to Operational Risk Management, *Methodology and Computing in Applied Probability*, **11**, 95–115.
- [9] Dalla Valle, L., De Giuli, M.E., Tarantola, C. & Manelli, C. (2016) Default Probability Estimation via Pair Copula Constructions. *European Journal of Operational Research*, **249**, 298–311.
- [10] Dalla Valle, L., Fantazzini, D. & Giudici, P. (2007) Empirical Studies with Operational Loss Data: Dalla Valle, Fantazzini and Giudici Study, in *Operational Risk : a Guide to Basel II Capital Requirements, Models, and Analysis*, A.S. Chernobai, S.T. Rachev & F.J. Fabozzi, eds., John Wiley & Sons, pp. 274–277.
- [11] Fantazzini, D., Dalla Valle, L. & Giudici, P. (2008) Copulae and Operational Risks, *International Journal of Risk Assessment and Management*, **9**, 238–257.
- [12] Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, **47**, 153–161.
- [13] Joe, H. (1997). *Multivariate model and dependence concepts*, Monographs on Statistics and Applied Probability, **73**, Chapman & Hall, London.

[14] Nelsen, R. B. (1999). *An introduction to copulas*, Springer-Verlag, New York.

[15] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231.

6 Figures

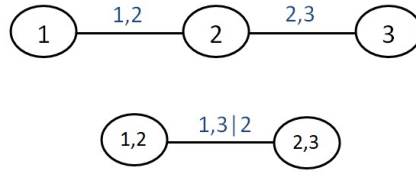


Figure 1: Trivariate vine.

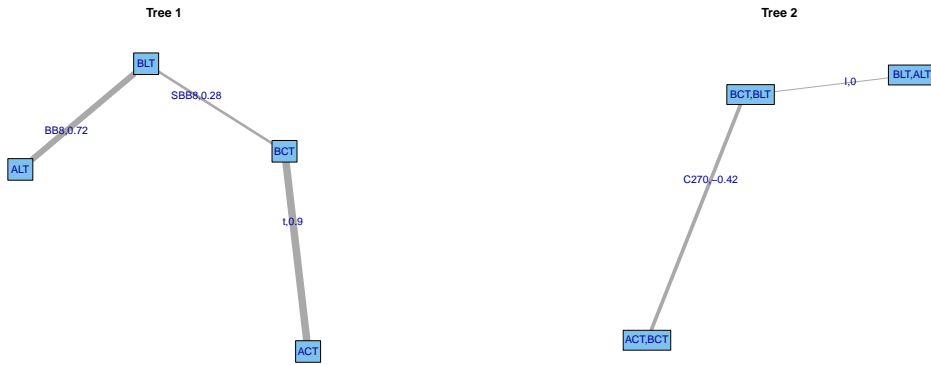


Figure 2: First (left panel) and second (right panel) D-vine trees of the Enron data for the PD estimation.



Figure 3: First (left panel) and second (right panel) C-vine trees for the firm sales data calibration.