UDC 519.7, 338.67
DOI: 10.22363/2658-4670-2019-27-4-343-354

Vine copulas structures modeling on Russian stock market

Eugeny Yu. Shchetinin

Department of Data Analysis, Decision Making and Financial Technologies
Financial University under the Government of Russian Federation
49, Leningradsky Prospect, Moscow 125993, Russian Federation

(received: March 5, 2019; accepted: December 30, 2019)

Pair-copula constructions have proven to be a useful tool in statistical modeling, particularly in the field of finance. The copula-based approach can be used to choose a model that describes the dependence structure and marginal behaviour of the data in efficient way, but is usually applied to pairs of securities. In contrast, vine copulas provide greater flexibility and permit the modeling of complex dependency patterns using the rich variety of bivariate copulas which may be arranged and analysed in a tree structure. However, the number of possible configurations of a vine copula grows exponentially as the number of variables increases, making model selection a major challenge in development. So, to learn the best possible model, one has to identify the best possible structure, which necessitates identifying the connections between the variables and selecting between the multiple bivariate copulas for each pair in the structure.

This paper features the use of regular vine copulas in analysis of the co-dependencies of four major Russian Stock Market securities such as Gazprom, Sberbank, Rosneft and FGC UES, represented by the RTS index. For these stocks the D-vine structures of bivariate copulas were constructed, which models are described by Gumbel, Student, BB1and BB7 copulas, and estimates of their parameters were obtained. Computer simulations showed a high accuracy of the approximation of the explored data by D-vine structure of bivariate copulas and the effectiveness of our approach in general.

Key words and phrases: copula, multivariate models, dependence structure, vines, securities

1. Introduction

In the field of financial analysis, finding new useful models and improving the existing ones is a constant struggle. Finding an appropriate multivariate model that efficiently describes the dependence structure as well as marginal behavior of the data being analyzed can be a very challenging task, especially in the case of higher dimensions. The approach that relies on copulas tends to outperform other methods when it comes to financial analysis, for example modeling financial returns.

Usually the Student n-dimensional copula is a good choice for financial data of various kinds [1], and as such deserves special attention. Of course, generally speaking, thorough analysis is needed for the best results – especially if the data being analysed has different behaviour in the tails, in which case the Student copula might not capture the dependence structure very well.

We will be discussing pairwise model into bivariate copulas as laid out by Aas [2]. This approach will let us easily track the parameters relevant to the tail dependence. In order to find the most appropriate approach for our specific case, we will rely on the detailed comparison and overview of different approaches by Berg [3].

The relatively recent concept of vines, introduced by T. Bedford and R. M. Cooke [4], is very relevant to pairwise decomposition of multivariate distributions. Vines essentially a subclass of trees that can be used to efficiently represent a pairwise decomposition. We will focus primarily on D-vines and canonical vines [5], [6]. Our main source for the elements of copula theory is R. B. Nelsen, H. Joe [7]–[9].

2. Basics of copula theory

Deflnition (pair-copula)

A *pair-copula* or simply *copula* is a function $C : [0,1]^2 \to [0,1]$ that satisfies the following properties:

For any $u, v \in [0,1]$

1) $C(u,0) = C(0,v) = 0$;

2) $C(u,1) = u$, $C(1,v) = 1$.

For any $u_1, u_2, v_1, v_2 \in [0,1]$ such that $u_1 \leqslant u_2$ and $v_1 \leqslant v_2$

3) $C(u_1,v_1) - C(u_1,v_2) - C(u_2,v_1) + C(u_2,v_2) \geqslant 0$.

One of the most important theorems of copula theory is Sklar's Theorem. In terms of probability theory, it states that any joint distribution function can be can be written in terms of marginal (univariate) distribution functions and a copula function that can describe the dependence structure between the random variables.

Sklar's Theorem.

Let X and Y be random variables with distribution functions F and G, respectively, and let H be their joint distibution function. Then there exists a copula $C : [0,1]^2 \to [0,1]$ such that for any $x, y \in \overline{\mathbb{R}}$ the following equation is true:

$$H(x,y) = C(F(x), G(y)). \tag{1}$$

If F and G are continuous, then C is unique. If not, then C is unique only on RanFxRanG (here RanF is the range of F and RanG is the range of G). Conversely, if C is a copula and F and G are distribution functions of X and Y, respectively, then H, defined by (1), is a joint distribution function for the random variables X and Y, and F and G are marginal distribution functions for X and Y, respectively.

It is not hard to describe the n-dimensional case, as well. But first, we have to define the notions of n-Box and the H-volume of an n-Box and discuss notation.

Let us use the following notation:

$$\mathbf{a} = (a_1, a_2, \dots, a_n) \in \overline{\mathbb{R}}^n, \quad \mathbf{b} = (b_1, b_2, \dots, b_n) \in \overline{\mathbb{R}}^n,$$

$\mathbf{a} \leqslant \mathbf{b}$ means $a_k \leqslant b_k$ for all k from 1 to n.

When $\mathbf{a} \leqslant \mathbf{b}$ we will use the following notation:

$$\mathbf{[a,b]} = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n].$$

The construction above is called the n-box. The vectors of the type $\mathbf{c} = (c_1, c_2, \dots, c_n)$ where c_k equals a_k or b_k for all k are called the vertices of the n-box.

The notion of the C-volume of the n-box, $V_C\mathbf{[a,b]}$, is discussed in [10], [11].

Definition (n-copula)

An *n-copula* is a function $C : [0,1]^n \to [0,1]$ that satisfies the following properties:

For any $\mathbf{u} = (u_1, u_2, \dots, u_n)$ in $[0,1]^n$

1) $C(\mathbf{u}) = 0$ if any $u_k = 0$.

2) $C(\mathbf{u}) = u_k$ if all the coordinates except u_k are equal to 0.

For any $\mathbf{a}, \mathbf{b} \in [0,1]^n$ such that $\mathbf{a} \leqslant \mathbf{b}$

3) $V_C\mathbf{[a,b]} \geqslant 0$.

The n-dimensional version of Sklar's theorem is discussed in Nelsen [7], and conditional copulas are discussed in Patton [12].

3. Decomposition of a multivariate distribution function using pair-copula constructions

The general product rule (also called the chain rule of probability) allows us to decompose a multivariate density function in the following, non-unique way:

$$f_{12..n} = f_1 f_{2|1} f_{3|12} \cdots f_{n|12\dots n-1}. \tag{2}$$

If we assume that F is strictly continuous and use the definition of a copula and Sklar's Theorem, we get

$$f_{12..n} = c_{12\dots n} f_1 f_2 \cdots f_n. \tag{3}$$

To get to the pair-copula decomposition we will also have to use the useful factorizations of this type:

$$f_{2|1} = \frac{f_1 f_2 c_{12}}{f_1} = c_{12} f_2. \tag{4}$$

$$f_{3|12} = \frac{f_{123}}{f_{12}} = \frac{f_1 f_{23|1}}{f_{2|1} f_1} = \frac{f_{23|1}}{f_{2|1}} = \frac{f_{2|1} f_{3|1} c_{23|1}}{f_{2|1}} = c_{23|1} f_{3|1} = c_{23|1} c_{13} f_3. \tag{5}$$

Now let's apply (2), (3), (4) and (5) to a 3-dimensional density function to get a pair-copula decomposition:

$$f_{123} = f_1 f_{2|1} f_{3|12} = f_1 c_{12} f_2 c_{23|1} c_{13} f_3 = c_{12} c_{13} c_{23|1} \prod_{i=1}^{3} f_i. \tag{6}$$

If we pick another conditioning variable we get another decomposition, for example

$$f_{123} = f_1 f_{2|1} f_{3|12} = f_1 c_{12} f_2 c_{13|2} c_{23} f_3 = c_{12} c_{13|2} c_{23} \prod_{i=1}^{3} f_i. \tag{7}$$

The number of possible pair-copula decompositions for a 3-variable density function is 24 [13], [14] and this number rises rapidly with the number of dimensions, which makes it very complicated to find the decomposition that best preserves the known information about the dependence structure. The concept of *vines* is very useful in this regard.

4. The concept of vines

Vines are a concept first introduced by Bedford and Cooke [4]. A vine is a sequence of trees $\{T_i\}$ in which the edges of T_i are the nodes of T_{i+1}. Each vine is a representation of a particular way of decomposing a multivariate distribution. The two kinds of common vines that we will use in our work are canonical vines and D-vines. Different types of vines represent different types of dependency structures. A canonical vine corresponds to the case where one "main" variable "interacts" with all the others, while in the case of a D-vine there is no such "main" variable. This idea is represented in the illustrations provided in Figures 1 and 2.

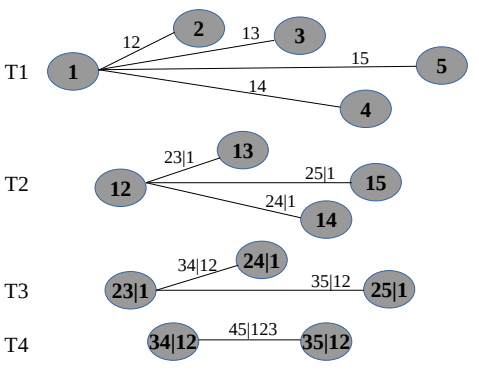


Figure 1. C-vine

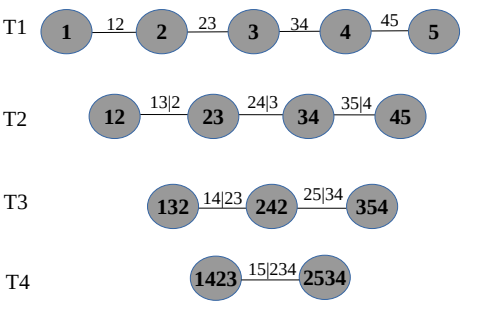


Figure 2. D-vine

The following general formulas give us the expressions for the decomposition of an n-dimensional density function using the D-vine and the canonical vine:

$$\text{D-vine: } f_{12\ldots n} = \prod_{k=1}^{n} f_k \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\ldots,i+j-1}. \tag{8}$$

$$\text{Canonical vine: } f_{12\ldots n} = \prod_{k=1}^{n} f_k \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1,\ldots,j-1}. \tag{9}$$

Each edge in each of the trees corresponds to a pair-copula, the density of which is used as one of the multipliers of the pair-copula construction, as we can see in (8) and (9). The first tree, T_1, should be constructed in a way that best represents the supposed dependence structure of the variables.

Alternative constructions may involve using the copula parameter estimations to get insight into the dependence structure — for example, we could assign a Student-t topula to all the pairs and, knowing that a low number of df indicates strong dependence, could construct a tree that represents that dependence structure.

Algorithm 1. Sequential algorithm

Input: Data $(x_{l1}, \ldots, x_{ln}$, $l = 1, \ldots, N$ (realization of i.i.d. random vectors).

Output: R-vine copula specification, i.e., $\mathcal{V}, B$.

1: Calculate the empirical Kendall's tau $\widehat{\tau}_{j,k}$ for all possible variable pairs $\{j,k\}$, $1 \leqslant j < k \leqslant n$.

2: Select the spanning tree that maximizes the sum of absolute empirical Kendall's taus, i.e.,

$$\max \sum_{e=\{j,k\}\text{in spanning tree}} \left\|\widehat{\tau}_{j,k}\right\|.$$

3: For each edge $\{j,k\}$ in the selected spanning tree, select a copula and estimate the corresponding parameter(s). Then transform $\widehat{F}_{j|k}(x_{lj}|x_{lk})$ and $\widehat{F}_{k|j}(x_{lk}|x_{lj})$, $l = 1, \ldots, N$, using the fitted copula $\widehat{C}_{jk}$ (see (2)).

4: **for** $i = 2, \ldots, n-1$ **do** {Iteration over the trees}

5: Calculate the empirical Kendall's tau $\widehat{\tau}_{j,k|D}$ for all conditional variable pairs $\{j,k|D\}$ that can be part of tree T_i, i.e. all edges fulfilling the proximity condition (see Definition 2.1).

6: Among these edges, select the spanning tree that maximizes the sum of absolute empirical Kendall's taus, i.e.,

$$\max \sum_{e=\{j,k|D\}\text{in spanning tree}} |\widehat{\tau}_{j,k|D}|.$$

7: For each edge $\{j,k|D\}$ in the selected spanning tree, select a conditional copula and estimate the corresponding parameter(s). Then transform $\widehat{F}_{j|k\cup D(x_{lj}|x_{lk},\mathbf{x}_{lD})}$ and $\widehat{F}_{j|k\cup D(x_{lk}|x_{lj},\mathbf{x}_{lD})}$, $l = 1, \ldots, N$, using the fitted copula $\widehat{C}_{jk|d}$ (see (2)).

8: **end for**

5. Numerical experiment: choosing the right vine structure

We will now apply the theory and methods discussed above to the analysis, modeling and visualization of the returns of four major Russian companies. Our data-set consists of the log-returns of Gazprom, Sberbank, Rosneft and FGC UES from 06.06.2014 to 06.06.2018.

We will use the VineCopula package for the R programming language for most of our computational needs [15].

Our main goal is to build a model that best represents core features of our data's dependency structure. We will use the sequential method [13] with Akaike's criterion [11], [16]–[18] (to determinine the most appropriate copula families) and one of the versions of Prim's algorithm (to determine maximum spanning trees [19], [20]) to ultimately determine and specify the most appropriate vine structure. We have provided the results below.

Figure 3 illustrates the D-vine structure of our model.

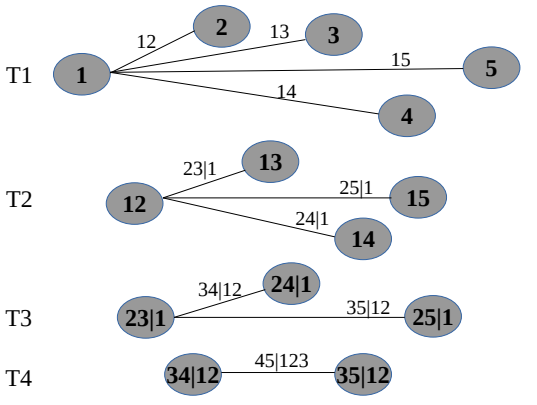


Figure 3. D-vine structure of our model

We also need to verify our model. The verification process involves drawing observations from the vine and comparing the empirical values of Spearman's Rho and some of the plots for the original observations and the sampled observations. In other words, we must observe how well the dependence structure was preserved.

For the sake of brevity, let us denote Rosneft by R, Gazprom by G, FGC UES by F and Sberbank by S.

Using AIC and MLE we have determined that:

1. c_{SF} is a rotated BB1 copula with $\theta = 0.1980236$ and $\delta = 1.421392$.
2. c_{SG} is a rotated BB7 copula with $\theta = 1.920555$ and $\delta = 0.7580773$.
3. c_{GR} is a rotated BB7 copula with $\theta = 2.025809$ and $\delta = 0.9424809$.
4. $c_{GR|S}$ is a rotated Gumbel copula with $\theta = 1.2104850$.
5. $c_{SR|G}$ is a t-copula with $\rho = 0.3746501$ and $\nu = 6.7874375$.
6. $c_{FR|SG}$ is a rotated BB8 copula with $\theta = 1.4269045$ and $\delta = 0.8675492$.

The D-vine tree structure for our model is presented on Fig. 4–6. Corresponding graphs of bivariate copula density models with estimated parameters are shown in Fig. 7.

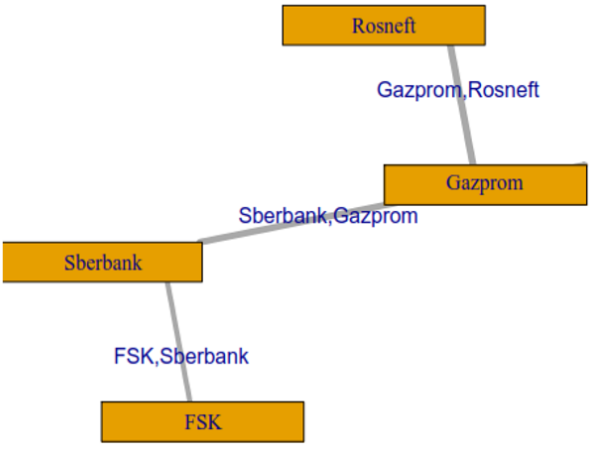


Figure 4. First tree

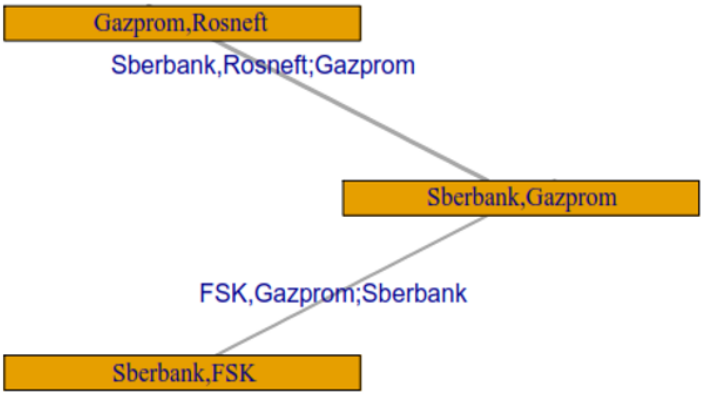


Figure 5. Second tree

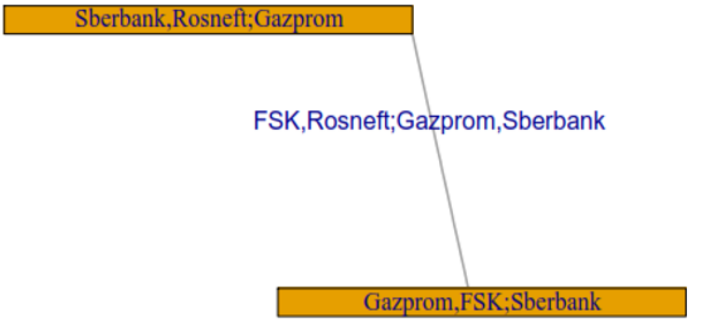
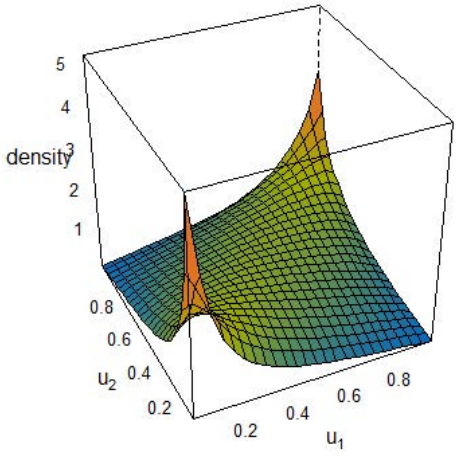
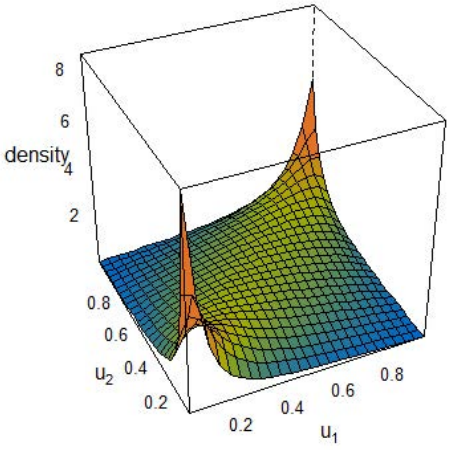


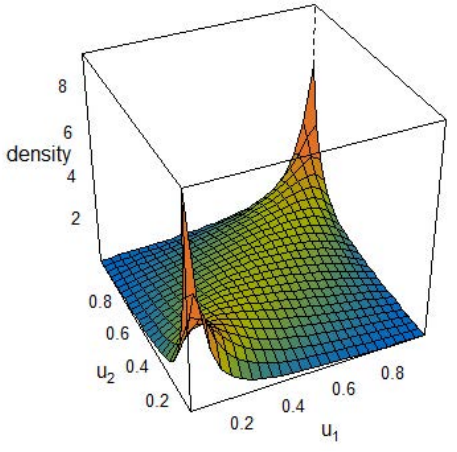
Figure 6. Final tree



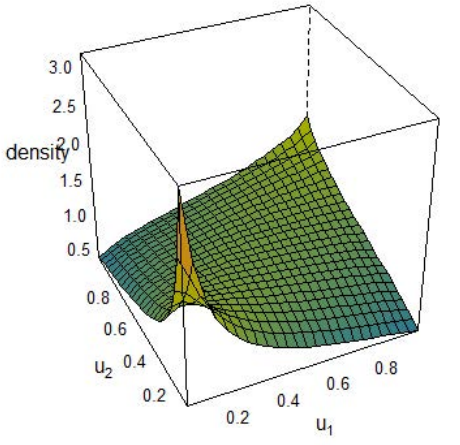
(a) c_{SF}



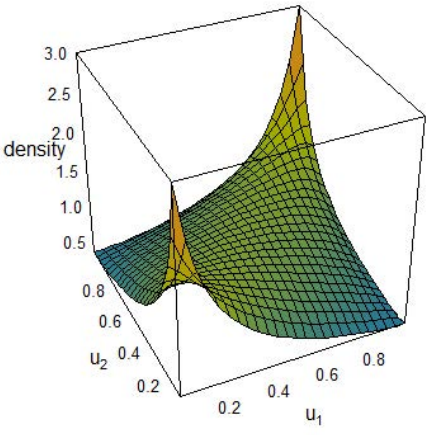
(b) c_{SG}



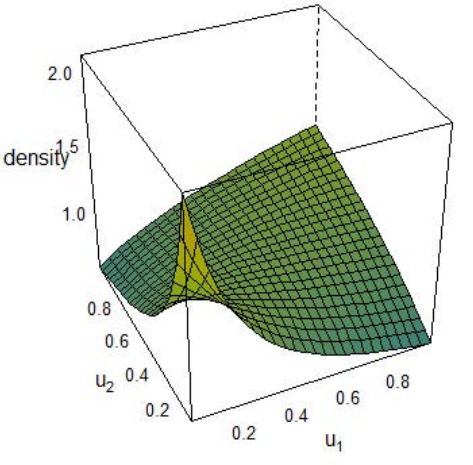
(c) c_{GR}



(d) $c_{GF|S}$



(e) $c_{SR|G}$



(f) $c_{FR|SG}$

Figure 7. Bivariate copula densities for the vine structure of our model

We drew 1003 observations from our D-vine — the same number as in our real-world dataset and calculated Spearman's rho values, shown below in Table 2. Judging from Table 2 and the overlaid plots, the modeled dependencies were captured in a satisfactory way. Graphical comparison of empirical and simulated data with their scatterplots is presented on Fig. 8.

Table 1

Empirical Spearman's Rho values for the original observations

	G	F	R
S	**0.64**	**0.5**	0.63
G	-	0.49	**0.67**
F		-	0.49

Table 2

Empirical Spearman's Rho values for the observations from sampling

	G	F	R
S	0.6	0.54	0.65
G	-	0.5	0.65
F		-	0.53

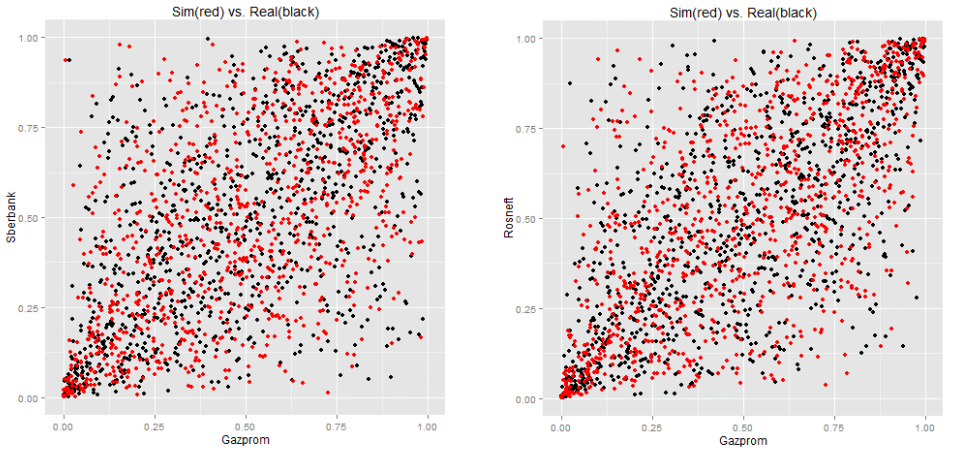


Figure 8. Real and simulated data comparison

6. Conclusions

In this paper we have demonstrated the usefulness of the vine copula-based approach to modeling a real-world dataset with a complex dependence structure. We have successfully specified a model that captures some of the essential dependencies that characterize our dataset. In a sense, by focusing, for the sake of brevity, exclusively on C-vines and D-vines and specific methods of copula selction and parameter estimations, we were forced to neglect other approaches which could provide additional insights. Extensive functionality provided by the VineCopula package for the R programming language let us circumvent many computational dificulties, allowing for faster and more efficient analysis.

References

[1] K. Aas and I. Hobaek Haff, "The generalized hyperbolic skew Student's t-distribution," *Journal of Financial Econometrics*, vol. 4, pp. 275–309, Jan. 2006. DOI: 10.1093/jjfinec/nbj006.

[2] K. Aas, C. Czado, A. Frigessi, and H. Bakken, "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 182–198, 2009.

[3] D. Berg, "Copula goodness-of-fit testing: an overview and power comparison," *European Journal of Finance*, vol. 15, pp. 675–701, 2009. DOI: 10.1080/13518470802697428.

[4] T. Bedford and R. M. Cooke, "Vines–a new graphical model for dependent random variables," *The Annals of Statistics*, vol. 30, no. 4, pp. 1031–1068, 2002. DOI: 10.1214/aos/1031689016.

[5] A. Panagiotelis, C. Czado, H. Joe, and J. Stöber, "Model selection for discrete regular vine copulas," *Comput. Stat. Data Anal.*, vol. 106, pp. 138–152, 2017. DOI: 10.1016/j.csda.2016.09.007.

[6] J.-D. Fermanian, "Recent developments in copula models," *Econometrics*, vol. 5, no. 34, 2017. DOI: 10.3390/econometrics5030034.

[7] R. B. Nelsen, *An introduction to copulas*. New York: Springer, 1999.

[8] H. Joe, H. Li, and A. K. Nikoloulopoulos, "Tail dependence functions and vine copulas," *Journal of Multivariate Analysis*, vol. 101, pp. 252–270, 2010. DOI: 10.1016/j.jmva.2009.08.002.

[9] H. Joe, "Dependence comparisons of vine copulae with four or more variables," in *D. Kurowicka and H. Joe (Eds.), Dependence Modeling*. Singapore: World Scientific, 2010.

[10] A. K. Nikoloulopoulos, H. Joe, and H. Li, "Vine copulas with asymmetric tail dependence and applications to financial return data," *Computational Statistics and Data Analysis*, vol. 56, no. 11, pp. 3659–3673, 2012.

[11] *Modeling dependence in econometrics*. Berlin, Heidelberg: Springer Verlag, 2014. DOI: 10.1007/978-3-319-03395-2.

[12] A. J. Patton, "Modelling asymetric exchange rate dependence," *International Economic Review*, vol. 47, no. 2, pp. 527–556, 2006. DOI: 10.1111/j.1468-2354.2006.00387.x.

[13] E. C. Brechmann, C. Czado, and K. Aas, "Truncated regular vines in high dimensions with application to financial data," *Canadian Journal of Statistics*, vol. 40, no. 1, pp. 68–85, 2012. DOI: 10.1002/cjs.10141.

[14] J. Di'smann, E. Brechmann, C. Czado, and D. Kurowicka, "Selecting and estimating regular vine copulae and application to financial returns," *Computational Statistics & Data Analysis*, vol. 59, pp. 52–69, 2013. DOI: 10.1016/j.csda.2012.09.01.

[15] E. C. Brechmann and U. Schepsmeier, "Modeling dependence with C- and D-vine copulas: the R package CDVine," *Journal of Statistical Software*, vol. 52, no. 3, pp. 1–27, 2013. DOI: 10.18637/jss.v052.i03.

[16] S. Konishi and G. Kitagawa, *Information criteria and statistical modeling.* 2007. DOI: 10.1007/978-0-387-71887-3.

[17] H. Manner and O. Reznikova, "A survey on time-varying copulas: specification, simulations and application," *Econometric Reviews*, vol. 31, no. 6, pp. 654–687, 2012. DOI: 10.1080/07474938.2011.608042.

[18] L. Chollete, A. Heinen, and A. Valdesogo, "Modeling international financial returns with a multivariate regime switching copula," *J. Financ. Econ.*, vol. 7, pp. 437–480, 2009. DOI: 10.2139/ssrn.1102632.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms, 2rd Edition.* MIT Press, 2001.

[20] D. E. Allen, M. A. Ashraf, M. McAleer, R. J. Powell, and A. K. Singh, "Financial dependence analysis: applications of vine copulas," *Statistica Neerlandica*, vol. 67, no. 4, pp. 403–435, 2013. DOI: 10.1111/stan.12015.

For citation:

E. Y. Shchetinin, Vine copulas structures modeling on Russian stock market, Discrete and Continuous Models and Applied Computational Science 27 (4) (2019) 343–354. DOI: 10.22363/2658-4670-2019-27-4-343-354.

Information about the authors:

Eugeny Yu. Shchetinin — Doctor of Physical and Mathematical Sciences, lecturer of Department of Data Analysis, Decision Making and Financial Technologies of Financial University under the Government of Russian Federation (e-mail: riviera-molto@mail.ru, phone: +7(917)5390698, ORCID: https://orcid.org/0000-0003-3651-7629, ResearcherID: O-8287-2017, Scopus Author ID: 16408533100)

УДК 519.7, 338.67
DOI: 10.22363/2658-4670-2019-27-4-343-354

Моделирование многомерных структур статистической зависимости на российском фондовом рынке

Е. Ю. Щетинин

Департамент анализа данных, принятия решений и финансовых технологий
Финансовый университет при Правительстве Российской Федерации
Ленинградский проспект, д. 49, Москва, 125993, Россия

Модели копул являются эффективным инструментом в статистическом моделировании, в частности в области финансового анализа. Подход к моделированию многомерных структур с их использованием позволяет описать как структуру статистической зависимости, так и маржинальные свойства данных, но обычно применяется к парам ценных бумаг. Наряду с этим, модели вьющихся копул обеспечивают большую гибкость и позволяют моделировать сложные структуры зависимостей, используя большое разнообразие двумерных копул, которые могут быть организованы в древовидную структуру. Однако число возможных конфигураций вьющихся копул растёт экспоненциально по мере увеличения числа ценных бумаг, что делает выбор модели основной научной проблемой. Таким образом, чтобы построить модель многомерных структур ценных бумаг, нужно определить наилучшую возможную структуру, которая требует выявления связей между её переменными, а также выбора между несколькими двумерными копулами для каждой пары в структуре.

В данной работе продемонстрировано применение регулярных вьющихся копул в финансовом анализе статистических связей крупнейших российских ценных бумаг, таких как Газпром, Сбербанк, Роснефть и ФСК ЕЭС, представленных в индексе РТС. Для этих ценных бумаг были построены D-vine структуры попарных копул, включающих модели Гумбеля, Стьюдента, BB1 и BB7, а также получены оценки их параметров. Компьютерное моделирование показало высокую точность аппроксимации исследуемых данных и эффективность предложенного подхода в целом.

Ключевые слова: финансовый анализ, ценные бумаги, многомерные структуры статистических связей, копулы, вьющиеся копулы