Application of selection and estimation regular vine copula on go public company share

Hasna Afifah R, Lienda Noviyanti, Achmad Bachrudin

1,2,3 Statistics Department, Faculty of Mathematics and Science, State University of Padjadjaran, Indonesia
1 hasnaafifahrusyda@gmail.com
2 lienda@unpad.ac.id
3 achmad.bachrudin@unpad.ac.id

Abstract. The accuracy of financial risk management involving a large number of assets is needed, but information about dependencies among assets cannot be adequately analyzed. To analyze dependencies on a number of assets, several tools have been added to standard multivariate copula. However, these tools have not been adequately used in apps with higher dimensions. The bivariate parametric copula families can be used to solve it. The multivariate copula can be built from the bivariate parametric copula which is connected by a graphical representation to become Pair Copula Constructions (PCCs) or vine copula. The application of C-vine and D-vine copula have been used in some researches, but the use of C-vine and D-vine copula is more limited than R-vine copula. Therefore, this study used R-vine copula to provide flexibility for modeling complex dependencies on a high dimension. Since copula is a static model, while stock values change over time, then copula should be combined with the ARMA-GARCH model for modeling the movement of shares (volatility). The objective of this paper is to select and estimate R-vine copula which is used to analyze PT Jasa Marga (Persero) Tbk (JSMR), PT Waskita Karya (Persero) Tbk (WSKT), and PT Bank Mandiri (Persero) Tbk (BMRI) from august 31, 2014 to august 31, 2017. From the method it is obtained that the selected copulas for 2 edges at the first tree are survival Gumbel and the copula for edge at the second tree is Gaussian.

1. Introduction

The accurate financial risk management involving a large number of assets is needed. Especially the interdependence of assets has not been adequately estimated. Although some advantages were added in multivariate standard, they are inadequate in higher dimensional applications.

Unfortunately, the choice of multivariate copula-copula is quite limited, as opposed to bivariate cases, whereas different types of copula exhibit flexible and complex dependence patterns. Salmon (2009) describes copula Gaussian has a weakness to see the stock dependency structure during a financial crisis. In addition, Student's copula only can capture symmetric dependencies using correlation matrices and Archimedean copula can be used only one or two parameters to model dependencies between possible dozens of variables. For high-dimensional distributions, there is a large number of possible copula pairs. For example, there are 240 constructions for five-dimensional density. To help organize it, Joe (1996) forms a copula model and further explored by Bedford and
The copula is a static model, they must be combined with time series models for modeling movements of stock value (Cherubini et al, 2004). In the financial market, the economic cycle can be observed, where the value of stock prices, indexes, and currencies can sustain value movements, both increase in value (profit), and impairment (loss). The movement of the value of the stock is called volatility. Asset volatility is estimated in order to know how asset movements are. Greater value of volatility return can be concluded that the change of asset value is very large. Historical return data can be used to estimate volatility, i.e. current volatility and future volatility.

The volatility model used can accommodate the characteristics of stock return data. According to Bollerslev et al. 1994, the time series data in finance has two important properties: the probability distribution returns of stock is fat tails and contain of volatility clustering. The volatility model used is Generalized Autoregressive Conditionally Heteroskedasticity (GARCH).

2. Copula

Copula is a Latin which means relationship or bond. The word copula was first used in 1959 for Theorem of Sklar. This theorem is used to describe a function that can combine several distribution functions from several random variables to a joint distribution function. The joint distribution function of some random variables can be determined using copula. These distribution functions can come from several different distributions.

The distribution function of randomized variables that have been combined to form a joint distribution function in the copula form is called the marginal function of the copula. This marginal function spreads uniformly (0,1) (Nelsen, 2006).

For each continuous random variable, the density function of copula is related to the joint density function of random variables in the copula, ie \( x_1, x_2, \ldots, x_d \). The relationship of the two functions can be described canonically as follows:

\[
f(x_1, x_2, \ldots, x_d) = c(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)) \prod_{i=1}^{d} f_i(x_i)
\]  

(1)

2.1 Sklar’s theorem

If \( H \) is a distribution function having \( d \) dimension with a cumulative distribution function \( F_1, F_2, \ldots, F_d \), then will appear a copula for all \( x \) in \( \mathbb{R}^d \), the copula associated with \( F \) is a function distribution \( C: [0,1]^d \to [0,1] \) with margin \( U(0,1) \), as follows:

\[
H(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d))
\]

(2)

- If \( F \) is a continuous \( d \)-variate distribution function with univariate margins \( F_1, \ldots, F_d \), and quantile functions \( F_1^{-1}, \ldots, F_d^{-1} \), then

\[
C(u) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right), \quad u \in [0,1]^d.
\]

(3)

- If \( F \) is the \( d \)-variate distribution of discrete random variables (in general, some are continuous and partly discrete), then copula is the only unique part of that collection

\[
\text{Range}(F_1) \times \ldots \times \text{Range}(F_d).
\]

In Sklar’s theorem, \( \text{Ran}F_1 \times \text{Ran}F_2 \times \ldots \times \text{Ran}F_d \) are combination of all possibility margin value \( F_i \).
If $F$ is the $d$-variate distribution of some or all of the discrete random variables, then the copula associated with $F$ is not unique. If $H$ is non-continuous or discrete univariate cdf and $Y \sim H$, then $H(Y)$ doesn’t have a $U(0,1)$ distribution. The copula of equation (2) becomes unique. Copula will only be unique in $Range(F_1) \times \cdots \times Range(F_d)$, since $C$ on equation (2) will only be needed to be defined in this set. Such a $C$ must satisfy $C(1,\ldots,1,u_j,1,\ldots,1) = u_j$ where $u_j \in Range(F_j)$ for $j = 1,\ldots,d$; and $C$ can be extended to multivariate distribution with $U(0,1)$ margin.

Based on the Sklar’s theorem, it can be seen that copula has the following properties:

- $Dom \ C = I^n = [0,1]^n = [0,1] \times [0,1] \times \cdots \times [0,1]$ 
- $C(u_1,\ldots,u_n)$ is a non-decreasing function for all component $u_i$
- $C(0,\ldots,0,u_i,0,\ldots,0) = 0$ for each $i \in \{1,\ldots,n\}$, $u_i \in [0,1]$
- $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$ for each $i \in \{1,\ldots,n\}$, $u_i \in [0,1]$
- For all $(a_1,\ldots,a_n),(b_1,\ldots,b_n) \in [0,1]^n$ with $a_i \leq b_i$ be valid with $\sum_{i=1}^{2} \sum_{j=1}^{n} (-1)^{i+j} C(u_{ij_{1}},\ldots,u_{ni_{n}}) \geq 0$ where $u_{ij_{1}} = a_j$ dan $u_{ij_{2}} = b_j$, $j \in \{1,2,\ldots,n\}$.

3. Build the volatility model

Before solving the problem of heteroscedasticity in data through modeling its volatility using GARCH, data must be modelled by Autoregressive Integrated Moving Average (ARIMA). This model is a Box-Jenkins variant of the ARMA model where the model requires analysis of stationarity. In general, regular ARIMA model (without seasonal patterns) have the following mathematical models.

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B) \alpha_t$$

where $\phi_p(B) = (1- \phi_1 B - \cdots - \phi_p B^p)$ and $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ and $(1-B)^d$ is a non seasonal differencing in order $d$. This equation can also be written as ARIMA $(p, d, q)$. When the data is stationary without going through the process of differentation then the used model is ARMA $(p, q)$.

ARMA model have the following mathematical models.

$$X_t - a_1X_{t-1} - \cdots - a_p X_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}$$

Where $a_1,a_2,\ldots,a_p,b_1,b_2,\ldots,b_p \in \mathbb{R}$, $\epsilon_t \sim WN(0, \sigma^2)$

After obtaining a suitable ARIMA or ARMA model, the next step is build a GARCH model from square residual ARIMA or ARMA model. Bollerslev introduces the GARCH model. In the GARCH model, it is assumed that the model mean follows the ARIMA or ARMA model. Example $\epsilon_t = r_t - \mu_t$ dan $\epsilon_t$ follow the GARCH form $(m, s)$ if

$$\epsilon_t = \sigma_t v_t, \ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$

Where $v_t$ is independent random variabel and identic with mean $0$ and varians $1$, $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1,2,\ldots,m$, $\beta_j \geq 0, j = 1,2,\ldots,m$ with $\sum_{i=1}^{m} (\alpha_i + \beta_j) < 1$. Assume that $E(\epsilon_t | F_{t-1}) = 0$, $Var(\epsilon_t | F_{t-1}) = E(\epsilon_t^2 | F_{t-1}) = \sigma_t^2$.

4. Vine copula

Vine copula is another name of the Pair Copula Constructions (PCCs) introduced by Aas, et al (2009). Vine copula is a graphical representation to determine PCC. This is explained with more detail by Kurowicka and Joe (2011). Regular statistical inference vines have been reinforced by Dübmann, et al (2013).
A PCC can be illustrated by three dimensions. Example $\mathbf{x} = (X_1, X_2, X_3)' \sim F$ and assumed that all densities needed exist, then density function of $\mathbf{x}$ is

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{12}(F_1(x_2), F_2(x_3))c_{13}(F_1(x_1), F_3(x_3))c_{23}(F_2(x_1), F_3(x_1))$$

(7)

With Sklar’s theorem, joint density function of three dimensions can be described in bivariat copula as follow

$$f(x_1, x_2, x_3) = f_1(x_1)f_2(x_2)f_3(x_3)c_{12}(F_1(x_2), F_2(x_3))c_{13}(F_1(x_1), F_3(x_3))c_{23}(F_2(x_1), F_3(x_1))$$

(8)

Bivariate $C_{1,2}, C_{1,3}$, and $C_{2,3}$ can be independently selected from each other, so that different dependency structures can be modeled using PCCs, as well as constructing high dimension. For ease of inference, copula $C_{2,3}$ is assumed to depend only on $x_1$ through arguments $F(x_2 \mid x_1)$ and $F(x_3 \mid x_1)$. This is called simplification of the principle which has been investigated by Hobaek Haff, et al (2010).

5. Regular vine copula

The R-vine structure is graphically used for determining the copula required for the construction of a series of copula pairs, where the copula is a multivariate distribution of hypercube units $[0,1]^d$ with a marginal uniform distribution (Nelsen (2006)).

Referring to Kurowicka and Cooke (2006), a regular vine (R-vine) in the $d$ variable consists of a series of first related trees (related to acyclic graphs). $T_1,...,T_{d-1}$ with nodes $N_i$ and edges $E_i$ for $i=1,...,d-1$, where $T_i$ has nodes $N_i = \{1,...,d\}$ and edge $E_i$ and for $i = 2,...,d-1$ tree $T_i$ has nodes $N_i = E_{i-1}$. Besides that, proximity condition needs 2 edges in tree $T_i$. It is combined with tree $T_{i-1}$ if only edge shares the same nodes in tree $T_i$.

Bedford and Cooke (2001) and Kurowicka and Cooke (2006) have shown that an edges of an R-vine tree can be uniquely identified with 2 nodes: a conditioned nodes and series of conditioning nodes. It is denoted by $e=j(e),k(e)\mid D(e)$ where $D(e)$ is a conditioning set.

The multivariate copula connecting $T_1,...,T_{d-1}$ is then constructed by connecting each edge $e=j(e),k(e)\mid D(e)$ to $E_i$ with a bivariate copula density $C_{j(e),k(e)\mid D(e)}$. Referring to the theorem in Kurowicka and Cooke (2006), the density of R-vine copula is:

$$c(F_1(x_1),...,F_d(x_d)) = \prod_{i=1}^{d-1} \prod_{e \in E_i} c_{j(e),k(e)\mid D(e)}(F_j(x_{j(e)} \mid x_{D(e)}), F_k(x_{k(e)} \mid x_{D(e)}))$$

(9)

where $x_{D(e)}$ denotes subvector from $\mathbf{x}=(x_1,...,x_d)'$ showing with contained index in $D(e)$.

Two special types from regular vine (R-vine) copula are canonical vine (C-vine) and drawable vine (D-vine). R-vine is called a C-vine if each tree ($T_i$) have unique nodes with degree $d-i$ (maximum degree) for $i=1,...,d-2$. Therefore, a root node or R-vine is contained in a node with maximum degree in each tree (star structure). And R-vine calls D-vine if all nodes in the first tree $T_1$ have degree no more than two (path structures).

6. Modelling dependency among daily returns of stock

Data used in this research is daily stock return from go public company such as PT Jasa Marga (Persero) Tbk (JSMR), PT Waskita Karya (Persero) Tbk (WSKT), and PT Bank Mandiri (Persero)
Tbk (BMRI) from August 31, 2014 to August 31, 2017 (734 data). The extracted data is the daily closing price of each stock. R software is used in data processing with ‘VineCopula’ package.

Value of stock return is calculated as follows.

\[
    r_t = \ln \left( \frac{P_t}{P_{t-1}} \right),
\]  

(10)

where:

- \( r_t \) = return of stock,
- \( P_t \) = closing price stock in \( t \) period,
- \( P_{t-1} \) = closing price stock in \( t-1 \) period.

Descriptive of return JSMR, WSKT, and BMRI is as follow.

**Table 1.** Summary statistics of JSMR, WSKT, and BMRI returns.

| Parameter   | JSMR          | WSKT          | BMRI          |
|-------------|---------------|---------------|---------------|
| Min.        | -0.0819166    | -0.060625     | -0.0762274    |
| 1st Qu.     | -0.0093896    | -0.008801     | -0.0082475    |
| Median      | 0.000000      | 0.000000      | 0.000000      |
| Mean        | -0.0001144    | 0.001254      | 0.003014      |
| 3rd Qu.     | 0.0086918     | 0.011116      | 0.0099503     |
| Max.        | 0.0909715     | 0.07509       | 0.0819171     |
| Kurtosis    | 3.664556      | 1.160076      | 2.512133      |
| Skewness    | 0.2680271     | 0.3399456     | 0.08370384    |

Jarque Bera Test

- p-value < 2.2e-16
- p-value = 9.834e-13
- p-value < 2.2e-16

Since the return data contain volatility clustering, first of all, GARCH model is applied on each return before it is modeled by copula. The result is as follow.

**Table 2.** Estimates from univariate ARMA-GARCH models.

| Parameters | JSMR (P-value) | WSKT (P-value) | BMRI (P-value) |
|------------|---------------|---------------|---------------|
| ar1        | 0.5403 (0.008001) | -0.626805 (< 2e-16) | 0.5483 (0.000385) |
| ar2        | - | -0.8653877 (< 2e-16) | - |
| ma1        | -0.5681 (0.004477) | 0.6531046 (8.64e-14) | -0.5255 (0.000674) |
| ma2        | -0.1064 (0.020392) | 0.8371652 (< 2e-16) | -0.1239 (0.002816) |
| omega      | 0.00002097 (0.031505) | 0.0001008 (0.019084) | 0.00007641 (0.007782) |
| alpha1     | 0.2089 (0.000614) | 0.1602536 (0.002706) | 0.1645 (0.000993) |
| beta1      | 0.2049 (0.012611) | 0.5739191 (0.000125) | 0.5946 (0.00000192) |
| beta2      | 0.5461 (0.00000000347) |  |  |
LM Arch & Statistic = 11.73823 & Statistic = 7.944844 & Statistic = 7.852354  \\ AIC & -5.253642 & -5.085729 & -5.288613

Descriptive standardize residual from GARCH model such as histogram, dependency structures, and correlation value are described as follow:

![Figure 1. Pair-plots and Kendall's taus for representatives of each index group.](image)

From GARCH model will be generated standardized residual. The dependencies among these standardized residuals are then modeled with copula after being transformed into marginal uniform data using an integral transformation of empirical or parametric probabilities. To model R-vine copula, the following steps should be taken:

- Plot any possible pairs of variables on the data and estimate values of Kendall's tau.
- Determine the class of R-vine along with the choice of the copula family according to the dataset of each variable pairs. The pair-copula can be individually selected from 7 types of bivariate copulas, those are: Gauss, Student-t, Gumbel, Gumbel survival, rotated gumbel 90 degrees, rotated gumbel 270 degrees, and Frank. To select one possible R-vine for a given dataset, pairs must be chosen to determine used family copula. For the selection process of the R-vine copula specification based on Kendall's tau, the first stage is to define the first tree \( T_1 = (N_1, E_1) \) for R-vine, followed by the second tree, and so on. Trees are chosen in such a way as to produce the most robust dependency model choices. Maximum Spanning Tree (MST) algorithm is used to select the tree. It can maximize the number of absolute \( \text{absolute Kendall's tau} \). A Maximum Spanning Tree (MST) algorithm is clearly described on Dişman (2013).

- Select the suitable pair-copula with the data through AIC criteria.
- Estimated parameters for pair-copula through maximum likelihood estimation.

Since the possibility of R-vine on \( n \) variables increases with increasing \( n \) by the number of \( n \) \((n!/2 \times 2^{(n-2)})\) , algorithms are needed to facilitate that process. Dişman (2013) developed a sequential heuristic method for selecting tree structures from R-vine.

After going through the process described previously then obtains the suitable R-vine structure with data as follow.
and the parameters for pair-copula is as follow.

| Edge         | Bivariate Copula       | par | par2 | tau | utd | ltd |
|--------------|------------------------|-----|------|-----|-----|-----|
| **Tree 1**   |                        |     |      |     |     |     |
| WSKT, JSMR   | Survival Gumbel / SG   | 1.33| 0.25 | -   | 0.31|
| BMRI, WSKT   | Survival Gumbel / SG   | 1.29| 0.23 | -   | 0.29|
| **Tree 2**   |                        |     |      |     |     |     |
| BMRI, JSMR; WSKT | Gaussian / N       | 0.25| 0.16 | -   | -   |

Result from the selection process of the R-vine copula specification based on Kendall’s tau is 2 trees. The first tree’s bivariate copula family from edges \( C_{2,1}, C_{3,2} \) are survival Gumbel and the second tree’s bivariate copula family from edge \( C_{1,3|2} \) is Gaussian. Those copulas are chosen in such a way as to produce the most robust dependency model choices.

7. Conclusion

R-vine copula has different choices for individual copula types for each pair-copula, these choices are too many to be investigated. With R-vine copula, multivariate copula modeling become more flexible.

In the formation of R-vine copula, a general selection approach have been used sequentially to select the representation of trees together. Selection is done by selecting the copula type from the bivariate copula family and corresponding parameter prediction. Algorithm for finding the maximum spanning tree is used in the selection of each theoretical graphics. Absolute Kendall empirical values are used as weights, but other weights are possible. In financial data, tail behavior may be useful to be investigated.

The output of this selection procedure gives the best result of R-vine tree, the corresponding pair-copula types, and corresponding parameters. The results of this selection process can be used as a starting option for determining maximum likelihood estimates.

The best result of R-vine trees from the data is that contained edges \( C_{2,1}, C_{3,2} \) at the first tree and \( C_{1,3|2} \) at the second tree, and the most corresponding pair-copula types from edges \( C_{2,1}, C_{3,2} \) is survival Gumbel, and from edge \( C_{1,3|2} \) is Gaussian. Result of R-vine copula formation can be used to measure risk of investment portfolio.
References

[1] Aas K, Czado C, Frigessi A and Bakken H 2009 Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* **44**(2) pp 182–198.

[2] Bedford T and R M Cooke 2002 Vines - a new graphical model for dependent random variables *Annals of Statistics* **30** pp 1031-1068.

[3] Cherubini U, E Luciano and W Vecchiatto 2004 *Copula Methods in Finance* (Chichester: John Wiley & Sons)

[4] Dißmann J, E C Brechmann, C Czado, and D Kurowicka 2013 Selecting and estimating regular vine copulae and application to financial returns *Journal of Computational Statistics & Data Analysis* **59** pp 52-69.

[5] Joe H and J J Xu 1996 The estimation method of inference functions for margins for multivariate models *Technical Report* **166** Department of Statistics, University of British Columbia

[6] Kurowicka D and R M Cooke 2006 *Uncertainty Analysis with High Dimensional Dependence Modelling* (Chichester: John Wiley)

[7] Kurowicka D 2011. Optimal truncation of vines. In: Kurowicka, D., Joe H. (Eds), Dependence Modeling *Handbook on vine copulae* (Singapore : World Scientific Publishing Co)

[8] Nelsen R B 2006 *An Introduction to Copulas 2nd Edition.* (New York : Springer)

[9] Salmon, F 2009 Recipe for disaster: The formula that killed Wall Street *Wired Magazine* **17**(3)

[10] Yahoo! Finance. 1997. *Pt Jasa Marga (Persero) Tbk (JSMR.JK).* https://finance.yahoo.com/quote/JSMR.JK?p=JSMR.JK. September 16th 2017.

[11] Yahoo! Finance. 1997. *PT Wijaya Karya (Persero) Tbk (WIKA.JK).* https://finance.yahoo.com/quote/WIKA.JK?p=WIKA.JK. September 16th 2017.

[12] Yahoo! Finance. 1997. *PT Bank Mandiri (Persero) Tbk (BMRI.JK).* https://finance.yahoo.com/quote/BMRI.JK?p=BMRI.JK. September 16th 2017.