ISSN 2229-5518

Identification of Dependent Structure and Prediction of Composite Stock Price Index with C-D Vine Copula Approach

Ira Rosianal Hikmah, Asep Saefuddin, I Wayan Mangku

Abstract— Dependent analysis that is frequently used is Pearson correlation and modelling with regression method. The use of both methods must satisfy the normality assumption. If it is violated then the rank correlation can be used and modelling involve the concept of expected value and density function. If the marginal function of each variable is the same then bivariate or multivariate density function is applied. If it is different then joint density function is applied. With the condition of mutually independent, joint density function can be found by multiplying the marginal functions. However, if the variables are mutually dependent then joint density function can be found by using copula analysis. Analysis of copula is a method to know relationship of two or more variables when each variable can spread normally or not and can be used to the same or different marginal distributions. Vine copula analysis can be used if there are at least two variables. This research used vine copula analysis to analyze composite stock price index and macroeconomic (inflation, exchange rate IDR to USD, and interest rate) data. Vine copula types in this study were C-Vine Copula and D-Vine Copula with ellipse and Archimedean copula family. Dependent structure was obtained from comparing the result of copula family and parameter estimation of each type copula based on AIC and BIC. The best dependent structure in this research was D-Vine copula from Archimedean copula family with MLE method. Furthermore, dependent structure was applied to obtain joint density function and computed expected value as copula regression model. Moreover, this study also applied multiple regression as a comparison with copula regression. MAPE score to copula regression was 4.863% and multiple regression was 16.125%. Because of copula regression MAPE score is smaller than the MAPE of multiple regression then it can be concluded that copula regression model is better than multiple regression.

Index Terms— Archimedean family, C-Vine Copula, composite stock price index, D-Vine Copula.

1 INTRODUCTION

Dependent analysis is important in a daily application, such as in marketing, economics, finance, business, health, and insurance. Dependent analysis that is frequently used is Pearson correlation and modelling with regression method. The use of both methods must satisfy the normality assumption. If the assumption is violated then the validity of the conclusion is reduced. If it is violated then the rank correlation can be used, such as Spearman and Kendall correlation. Modeling can be applied by involving the concept of expected value and density function. If the marginal function of each variable is the same then bivariate or multivariate density function is applied. If it is different then joint density function is applied. In probability concept, with the condition of mutually independent, joint density function can be found by multiplying the marginal functions [5]. However, if the variables are correlated and mutually dependent then joint density function can be found by using copula analysis. Analysis of copula is a method to know relationship of two or more variables when each variable can spread normally or not and can be used to the same or different marginal distributions [12].

- *Ira Rosianal Hikmah is a masters degree student, Departement of Statistics, Bogor Agricultural University, Indonesia. E-mail: ira.hikmah4@gmail.com*
- *Asep Saefuddin is a Lecturer, Departement of Statistics, Bogor Agricultural University, Bogor, Indonesia*
- *I Wayan Mangku is a Lecturer, Departement of Mathematics, Bogor Agricultural University, Bogor, Indonesia*

Copula analysis for two variables is called bivariate copula analysis and for more than two variables is called multivariate copula analysis. Multivariate copula analysis is similar in concept to bivariate cases. In the case of mutually dependent and different marginal distributions, the use of classical copula analysis is very difficult to find the multivariate copula function. Therefore, researchers developed other approach, that is vine copula. The difference is that vine copula can decompose the function of multivariate copula into paired copula functions. These paired copula functions are obtained from bivariate copula families [9] [10].

First step of vine copula analysis is determining the dependent structure by determining the tree and copula families. Determination of the tree and copula families can be obtained by selecting one of several vine copula types. There are several types of vine copula, namely C-Vine (Canonical Vine) and D-Vine (Drawable Vine) Copula [3]. After that, vine copula analysis is continued by estimating the parameters value so the best paired copula function is obtained. Then, by using the paired copula funtions, joint density function can be searched even prediction value can be obtained by calculating the expected value.

Several studies using the vine copula approach have also been done, for example in Insurance [14] and Risk Management [6]. This research used vine copula analysis to analyze Composite Stock Price Index (IHSG) and macroeconomic (inflation, IDR exchange rate to USD, and interest rate) data. In general, economic and financial data do not spread normally even they have some extreme points. In the conventional studies, these macroeconomic factors are assumed to spread

ISSN 2229-5518

normally and mutually independent. However, macroeconomic factors are not always normal and independent because the changes in macroeconomic factors are influenced by the national economy [2]. Therefore, this research aims to examine the dependent structure and to determine the best C-D Vine Copula model between IHSG and macroeconomic factors. This research also predicts IHSG with vine copula approach and classical regression then compares the results of both.

2 Research Method

2.1 Data

The data used in this research is the data obtained from the website of Bank Indonesia (BI) [1] and Indonesia Stock Exchange (BEI) [8]. In this research, they were taken 120 data of IHSG and macroeconomic factors (inflation, IDR exchange rate to USD, and interest rate) from 2006 to 2015. The scope of this research is limited to cross section data. Therefore, the research data are divided into two groups, i.e. 100 data serve as training data to estimate the model and remaining, 20 data serve as testing data to evaluate the model. Training and testing data are selected randomly with the purpose of eliminating the time effects. Description of the variables used in this research are listed in Table 1.

TABLE 1
List of Variables

Variables	Explanation
Y	Composite Stock Price Index / IHSG (points)
X_1	Inflation (%)
X_2	Exchange Rate (IDR/USD)
X_3	Interest Rate (%)

2.2 Data Analysis Procedures

The stages of data analysis in this research are:

1. Data exploration.
   a. Normality test.
   b. Correlation analysis.
2. Perform the classical regression analysis of the training data and calculate mean absolute percentage error (MAPE) value from testing data.

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i-\widehat{Y}_i|}{Y_i}\right)\times 100\%$$

   [13].
3. Test the assumption of the regression model obtained from point 2.
4. Identify the marginal distribution.
5. Analyze vine copula and search for the best copula model from the training data using AIC and BIC [4].
   AIC = -2 log L + 2k
   BIC = -2 log L + k log n.
   a. Transform the data into uniform distribution [0,1].
   b. Perform the vine copula analysis for the four variables (Y, X_1, X_2, X_3).
   c. Search the joint denstiy function $g(y,x_1,x_2,x_3)$.
   d. Perform the vine copula analysis for the three variables (X_1, X_2, X_3).
   e. Search the joint density function $h(x_1,x_2,x_3)$.
6. Perform the copula regression from the training data to determine prediction of IHSG [11].
   a. Search conditional density function $f(y|x_1,x_2,x_3)$.
   b. Calculate expexted value $E(Y|X_1=x_1, X_2=x_2, X_3=x_3)$.
   c. Calculate MAPE value from the testing data.
7. Compare MAPE value of classical and copula regression model to get the best model.

3 Result and Discussion

3.1 Data Exploration

Table 2 shows the results of normality test from training data.

TABLE 2
Normality Test Results

Variables	P-Value
Y	0.000
X1	0.000
X2	0.000
X3	0.000

Based on Table 2, it can be seen that all the variables don't spread normally at the significance level 0.05. Therefore, this research uses spearman and kendall correlation to measure the relationship among variables. The result of correlation analysis with spearman and kendall can be seen in Table 3.

TABLE 3
Correlation Analysis

Variables	Spearman		Kendall	
	Correlation Coefficient	P-Value	Correlation Coefficient	P-Value
Y and X_1	-0.266	0.007	-0.171	0.012
Y and X_2	0.489	0.000	0.307	0.000
Y and X_3	0.489	0.000	-0.348	0.000
X_1 and X_2	0.081	0.423	0.042	0.539
X_1 and X_3	0.703	0.000	0.531	0.000
X_2 and X_3	0.041	0.685	0.03	0.6712

Based on Table 3, it is found that the covariates (macroeconomic factor) used in this research correlate with the response variable that is IHSG, at the significance level 0.05. Therefore, the covariates can be used for the classical regression analysis with the response variable (IHSG).

3.2 Classical Regression Analysis

There are six scenarios used in this research to get the best classical regression equation, namely:

1. $Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$
2. $Y = b_0 + b_1 Ln(X_1) + b_2 X_2 + b_3 Ln(X_3)$
3. $Y = b_0 + b_1 Ln(X_1) + b_2 Ln(X_2) + b_3 Ln(X_3)$
4. $Ln(Y) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$
5. $Ln(Y) = b_0 + b_1 Ln(X_1) + b_2 X_2 + b_3 Ln(X_3)$
6. $Ln(Y) = b_0 + b_1 Ln(X_1) + b_2 Ln(X_2) + b_3 Ln(X_3)$.

By using Adj. R^2, AIC, BIC, and MAPE measurements, scenario 5 is the best equation of the 6 scenarios selected in this research. Thus, the classical regression equation in this research is

$$Ln(Y) = b_0 + b_1 Ln(X_1) + b_2 X_2 + b_3 Ln(X_3).$$

The regression coefficient estimators from 100 training data can be seen in Table 4 and the results of assumption test in Table 5.

TABLE 4
REGRESSION COEFFICIENT ESTIMATION

Estimate	Value	P-Value
b_0	10.25	0.000
b_1	0.2424	0.014
b_2	0.0001217	0.000
b_3	-1.947	0.000

TABLE 5
ASSUMPTION TEST RESULTS

Test	Method	Result
Normality	Shapiro Wilk	Violated
Homoscedasticity	Breusch Pagan	Violated
Not Autocorrelation	Durbin Watson	Not Violated
Not Multicollinearity	VIF	Not Violated

Application of 20 testing data on regression model obtained using the coefficient estimators in Table 4, results MAPE value of 16,125%. It means, there is 16,125% error of classical regression model to predict the IHSG value. Based on assumption test results in Table 5, it is found that there are two assumptions that are violated, namely normality and homoscedasticity assumptions. The violation of homoscedasticity assumption can be ignored without great risk as long as it has the same sample size in each variable [7]. Meanwhile, data transformation can be used to overcome the violation of the normality assumption. Many cases that after the transformation, normality assumption is still violated as in this study, then the classical regression can not be used because the result of regression coefficient estimators are not BLUE (Best Linear Unbiased Estimator). Therefore, another method is used as an alternative to the classical regression model. If the marginal distributions are the same then IHSG can be predicted by calculating the expected value from the existing distribution. If the marginal distributions are different then the expected value is calculated by using the joint density function.

3.3 Marginal Distribution Identification

The results of marginal distribution identification of each variable can be seen in Table 6. The selection of the best distribution is obtained by AIC and BIC measurements. The parameters estimation is performed by maximum likelihood method. Based on Table 6, it is found that the marginal distributions of each variable are different. Covariates approach the lognormal distribution and response variable get near to weibull distribution. Since the distributions of each variable are different, it was searched the joint distribution. According to Bramantya (2014), macroeconomic factors are not entirely mutually independent then the joint distribution can be searched by using the vine copula approach.

TABLE 6
IDENTIFICATION OF MARGINAL DISTRIBUTION

Variables	Distribution	Parameter 1	Parameter 2	AIC	BIC
Y	Normal	3338.81	1332.31	1726	1731
	Lognormal	8.0174	0.4606	1736	1741
	Weibull	2.8172	3763.58	1721	1727
X_1	Normal	6.8724	3.2241	521	527
	Lognormal	1.8327	0.4279	484	489
	Weibull	2.2535	7.7854	506	512
X_2	Normal	10241	1561.84	1758	1763
	Lognormal	9.2235	0.1435	1744	1749
	Weibull	6.3871	10943.8	1776	1781
X_3	Normal	7.6525	1.7327	397	402
	Lognormal	2.0128	0.204	372	377
	Weibull	4.2044	8.3588	412	417

3.4 Vine Copula Analysis

In performing vine copula analysis, first transform the data into a uniform distribution [0,1]. Then, search the dependency structures for the four variables (IHSG, Inflation, Exchange Rate, Interest Rate) and three variables (covariates) using vine copula approach. Tables 7 and 8 present the result of vine copula analysis with AIC and BIC sizes.

TABLE 7
C-D VINE COPULA FOR THE FOUR VARIABLES

Type	Variables	Copula Distribution	MLE Parameter	Kendall Parameter
C-Vine	Y and X_1	Frank	-1.14713	-1.57707
	Y and X_2	Gumbel	1.98255	1.44328
	Y and X_3	Gauss	-0.49466	-0.51986
	X_1 and X_2	Frank	2.31655	1.58746
	X_1 and X_3	Gumbel	2.41288	1.63366
	X_2 and X_3	Frank	1.05111	0.9914
			AIC : -129	AIC : -103
			BIC : -113	BIC : -88
D-Vine	Y and X_1	Frank	-1.14713	-1.57707
	Y and X_2	Gumbel	1.32774	1.04347
	Y and X_3	Gumbel	1.28872	1.0311
	X_1 and X_2	Gumbel	1.86388	1.40945
	X_1 and X_3	Gumbel	2.47448	2.03704
	X_2 and X_3	Frank	-3.10405	-3.32033
			AIC : -147	AIC : -124
			BIC : -131	BIC : -108

Based on Table 7, it is found that the dependent structure for the four variables is D-Vine Copula with Archimedean copula family, namely Frank and Gumbel Copula, and the best parameter estimation is obtained by using MLE method. From Table 8, it is found that the dependent structure for the three variables is D-Vine Copula with Gumbel Copula and the best parameter estimation is also obtained by using MLE method. The dependency structures are used to get the copula regression model with the concept of expected value.

TABLE 8
C-D VINE COPULA FOR THE FOUR VARIABLES

Type	Variables	Copula	MLE	Kendall

		Distribution	Parameter	Parameter
C-Vine	X_1 and X_2	Gumbel	1.32774	-
	X_1 and X_3	Gumbel	2.72161	-
	X_2 and X_3	Gumbel	1.08479	-
			AIC : -88	AIC : -
			BIC : -81	BIC : -
D-Vine	X_1 and X_2	Gumbel	1.32774	1.04347
	X_1 and X_3	Gumbel	1.28872	1.0311
	X_2 and X_3	Gumbel	2.47448	2.03704
			AIC : -90	AIC : -73
			BIC : -82	BIC : -65

In this research, the prediction stage of IHSG value and evaluation of copula regression model is performed by using 20 testing data and obtain MAPE value of 4,863%. It means that there is 4,863% error of copula regression model to predict IHSG value.

3.5 Comparison of MAPE Values

In this research, it is found that the MAPE value of copula regression model is better than the classical regression model because the MAPE value of copula regression (4,863%) is smaller than the MAPE value of classical regression model (16,125%). It means that the copula regression model yields a smaller error than the classical regression model.

4 Conclusion

This research results show the dependent structures for the four and three variables derive from the D-Vine Copula type, the Archimedean family, with parameter estimation based on the maximum likelihood method (MLE). The dependent structures for the four variables derive from the Frank and Gumbel Copula, while the dependent structures for the three variables derive from the Gumbel Copula. IHSG prediction is obtained by using the classical and copula regression model. The classical regression model yields MAPE value, that is 4,863%, is smaller than the MAPE value of classical regression that is 16,125%. So, this study concludes that the copula regression approach is better than the classical regression.

References

[1] Bank Indonesia (BI). Jakarta[ID]. www.bi.go.id.

[2] Bramantya IS. 2014. Stock Exchange Composite Index Modelling and the Determination of Rank Correlation by Using Copula. Bogor[ID]: Bogor Agricultural University.

[3] Brechmann EC, Schepsmeier U. 2013. Modeling Dependence with C- and D- Vine Copulas: The R Package CD Vine. Journal of Statistical Software. Vol. 52, No. 3, pp. 1-27.

[4] Burnham KP, Anderson DR. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research. Vol. 32. No. 2, pp. 261-304.

[5] Casella, B & Berger, RL. 2002. Statistical Inference, 2nd edition. US: Duxbury, Thomson Learning Inc.

[6] Geidosch M., Fischer M. 2016. Application of Vine Copulas to Credit Portfolio Risk Modeling. Journal of risk and Financial Management. Vol. 9, No. 4, pp. 1-15.

[7] Hays. W.L. 1973. Statistics for the Behavioral Sciences, 2nd edition. New York: Holt, Rinehart and Winston Inc.

[8] Indonesia Stock Exchange (BEI). Jakarta[ID]. www.idx.co.id.

[9] Joe H. 1997. Multivariate Models and Dependence Concept. London: Chapman and Hall.

[10] Nelsen RB. 2006. An Introduction to Copulas 2nd Ed. Oregon: Springer Science and Business Media, Inc.

[11] Parsa RA, Klugman SA. 2011. Copula Regression. Variance: Advancing the Science of Risk. Vol. 5, pp. 45-54.

[12] Sklar, A. 1959. Fonctions de repartition a n dimensions et leurs marges. Paris (France) : Universite Paris.

[13] Sungkawa I, Megasari RT. 2011. Penerapan Ukuran Ketepatan Nilai Ramalan Data Deret Waktu dalam Seleksi Model Peramalan Volume Penjualan PT Satriamandiri Citramulia. *ComTech*. Vol. 2, pp. 636-645.

[14] Zhang B, Wei Y. 2014. Forecasting VaR and ES of Stock Index Portfolio: A Vine Copula Method. Elsevier: Physica A. Vol. 416, pp. 112-124.