Indonesian Journal of Physics and Nuclear Applications
Volume 3, Number 3, October 2018, p. 89-94
e-ISSN e-ISSN 2550-0570,

COPULA MODELING IN ANALYSIS OF DEPENDENCY OF OIL PALM PRODUCTION AND RAINFALL

Dadan Kusnandar[1,*], Naomi Nessyana Debataraja[1], Shantika Martha1[1]

[1]Department of Statistics, FMIPA, Universitas Tanjungpura, Pontianak, Indonesia
**Corresponding author: dkusnand@untan.ac.id*

Abstract Copula is a method that examines the relationship pattern between variables. Copula is characterized as a nonparametric method with several benefits, i.e., it is independent of the assumption of the distribution, accommodates nonlinear relationship among variables, and is convenient in building joint distribution. This study investigates the relationship and prediction analysis using the copula approach. The method is applied to the monthly data of oil palm production and the amount of rainfall. The results show that the model of Frank Copula is the best model for rainfall and oil palm production relationship.

Keywords Relationship, Correlation coefficient, Archimedean Copula, Elliptical Copula, Maximum Likelihood Estimation

INTRODUCTION

Oil palms are tropical plants that can grow well in the latitude of 12o N to 12o S. The amount of rainfall in the tropical area is considered important to the plant growth and productivity. Some studies show that the amount of rainfall and rainy days affect the productivity of the plants (Prasetyo 2009; Simanjuntak, Sipayung and Irsal 2014). The ideal rainfall for oil palm is 2000-2500 mm per year evenly distributed throughout the year without long dry months. A long dry season could reduce the production of oil palm since fewer minerals in soil are absorbed by the plants (Risza 2014). On the other hand, too much rainfall can cause erosion of topsoil especially in areas with bad topography.

Copula is one method that can be used to describe the relationship between variables without any assumption of the distribution. It can reveal relation of dependency even at extreme points (Anisa and Sutikno 2015). Copula has an important role when one or both variables have an abnormal marginal distribution or have tail dependencies. The use of the copula in investigating the relationship among variables has been studied by many authors in various fields. Anisa and Sutikno (2015) and Udayani, Sumarjaya, and Susilawati (2016) used copula in the analysis of relationship in agricultural research. Murterio and Lourencio (2007) applied copula in the health care utilization, whereas Zhu, Ghosh and Goodwin (2008) used copula in modeling dependence in the design of the insurance contract. Syahrir (2011) estimated copula parameters in the field of climatology. The Copula approach resulted in better estimates even in the presence of extreme observational data and for conditions that violate normality assumptions. This paper discusses the application of the copula in modeling the relationship between two variables, namely rainfall and palm oil production. The aim of the study is to provide an overview of Copula and its application in the field of agriculture.

METHOD OF ANALYSIS

Suppose two-dimensional random variables X having marginal cumulative distribution function of F_{X1}, F_{X2} with domain **R**, i.e., non-decreasing and $F_{Xi}(-\infty)=0$ and $F_{Xi}(\infty)=1$. The Sklar's theorem (See e.g. Embrechst et. al. 2001; Schoelzel and Friederichs 2008) says that the joint distribution of random variable F_X can be written as follows:

$$F_{X1,X2}(x_1,x_2)=C_{X1,X2}(F_{X1}(x_1),F_{X2}(x_2))$$

Where $C_{X1,X2}:[0,1]\times[0,1]\to[0,1]$ is a joint distribution function of the transformed random variables $Z_j=F_{Xj}(x_j)$ for j = 1, 2. Due to this transformation, Z_j always has uniform marginal distribution. The function C_X is called a copula and c_X is the corresponding copula density.

The definition of the parametric form of the copula function or the copula density allows them to be grouped into families. Two important classes are the Elliptical and Archimedean Copula families. Details about other copula families and their selection can be found in Embrechts et al. (2001). Gaussian Copula or Normal Copula is part of the Elliptical Copula family. The normal copula is derived from the transformation of random variables into standard normal distributions. Normal Copula function is given in the following equation:

$$C_{(X1,X2)}(Z_1,Z_2)=F_{N(0,\Sigma)}\left(F_{N(0,1)}{}^{-1}(Z_1),F_{N(0,1)}{}^{-1}(Z_2)\right)$$

Where $F_{N(0,\Sigma)}$ is the distribution function of the standard normal distribution, $F_{N(0,1)}{}^{-1}$ is the inverse distribution function of the standard normal distribution.

Commonly used Archimedean Copula are *Clayton*, Gumbel, and Frank Copula. The general form of Archimedean Copula is as follows:

$$C_{(X1,X2)}(Z_1,Z_2)=\varphi^{-1}(\varphi(Z_1),\varphi(Z_2))$$

Where φ is a generator function. The function $\varphi(Z)$ is a non-decreasing function that maps [0,1] into $\varphi(0)=\infty$ and $\varphi(1)=0$. Generator and bivariate copula for each copula in the family of Archimedean are presented in Table 1. Clayton Copula has lower tail dependencies, Frank Copula has no tail dependencies, while Gumbel Copula has upper tail dependencies (Schoelzel and Friedrichs 2008).

One method for estimating copula parameters is the maximum likelihood estimation (MLE). MLE for this copula is obtained by maximizing the log likelihood function (Udayani et. al. 2016). Estimated Copula parameters are presented in Table 2.

Table 1. Family Copula Archimedian

Copula	Generator $\varphi(Z)$	Bivariate Copula
Clayton	$\frac{z^{-\theta}-1}{\theta},\theta\in(0,\infty)$	$(z_1^{-\theta}+z_2^{-\theta}-1)^{\frac{1}{\theta}}$
Gumbel	$(-\log(z))^{\theta},\theta\in[1,\infty)$	$\exp\left\{-\left[(-\log(z_1))^{\theta}+(-\log((z_2))^{\theta}\right]^{\frac{1}{\theta}}\right\}$

Frank	$\log\left(\dfrac{e^{\theta z}-1}{e^{\theta}-1}\right), \theta \in R\setminus\{0\}$	$\dfrac{1}{\theta}\log\left(1+\dfrac{\left(e^{\theta z_1}-1\right)\left(e^{\theta z_2}-1\right)}{e^{\theta}-1}\right)$

Source: Anisa and Sutikno (2015)

Table 2. Estimates of Copula Parameters

Copula	Estimates of parameters
Gaussian	$\hat{\tau} = \dfrac{2}{\pi}\arcsin\theta$
Clayton	$\hat{\tau} = \dfrac{\theta_c}{\theta_c + 2}$
Gumble	$\hat{\tau} = 1 - \dfrac{1}{\theta_G}$
Frank	where $D_F = \dfrac{1}{\theta}\int_0^{\theta}\dfrac{u}{e^u - 1}du$

Source: Embrechts et. al. (2001)

Test of hypotheses of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is carried out using the following test statistic:

$$z = \frac{\hat{\theta} - \theta_0}{se\left(\hat{\theta}\right)}$$

The null hypothesis is rejected when the z statistic is greater than $z_{\alpha/2}$ of the standard normal distribution (Wang 2010).

RESULT AND DISCUSSION

The data used in this research are secondary data obtained from one estate of PT. Perkebunan Nusantara XIII and Meteorology, Climatology and Geophysics Agency Class II Siantan Station. These secondary data are the monthly production of oil palm plantation in the period of 2010-2015 and monthly rainfall data on the same period. The plot of oil palm production and rainfall are presented in Figure 1.

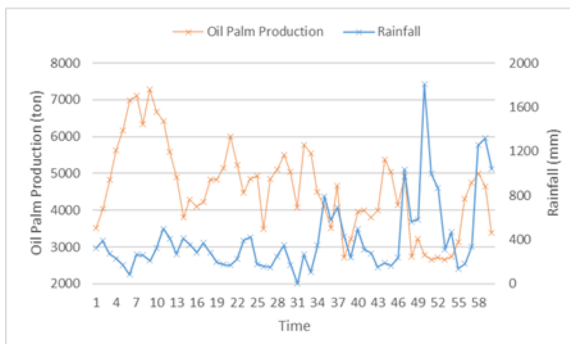

Fig 1. Plot of oil palm production and rainfall over time

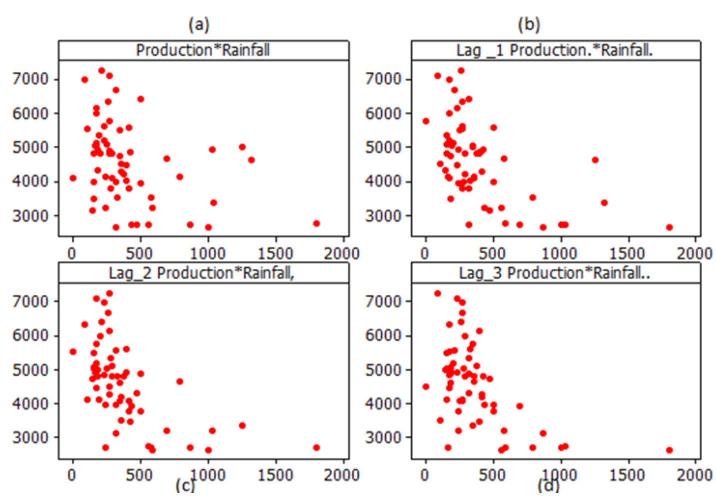

Fig 2. Plots of time lags of oil palm production on rainfall

The plot in Figure 1 shows a tendency of an increasing amount of rainfall in the last two years. It is followed by a decrease in the production of oil palm in the corresponding period of time.

Identification of a pattern in the relationship between the oil palm production and the rainfall was carried out by examining the scatter plot of the two variables. In examining the existence of dependency of the two variables, the oil palm data were plotted on the rainfall data. Different time lags of oil palm production were also plotted on the rainfall data. The plots are presented in Figure 2. Figure 2 (a) representing the data at the same period. In Figure 2 (b), (c) and (d) the different time lags of oil palm production were plotted on the rainfall data. This is carried out to inspect the relationship of rainfall to oil palm production in the next months. The scatter plots show that the pattern of oil palm production and the corresponding time lags versus the rainfall are skewed to the right. The bigger the time lag, the sharper the skewness.

The correlation between the time lags of oil palm production and the rainfall data were also calculated. Table 3 presents the Pearson, Spearman and the Tau Kendall correlation coefficients. They are all significant with negative values. However, normality tests on the data showed that the rainfall data are not normally distributed. This suggests that further investigation is to be carried out.

Table 3. Correlation Coefficients between oil palm production and rainfall data

Variables	**Pearson**		**Spearman**		**Tau-Kendall**	
	Corr Coef	***p-value***	***Corr Coef***	***p-value***	***Corr Coef***	***p-value***
Production vs. Rainfall	-0.350	0.006	-0.399	0.002	-0.278	0.002
Lag 1 Production vs. Rainfall	-0.543	0.000	-0.568	0.000	-0.521	0.000
Lag 2 Production vs. Rainfall	-0.575	0.000	-0.601	0.000	-0.549	0.000
Lag 3 Production vs. Rainfall	-0.537	0.000	-0.487	0.000	-0.512	0.000

With copula approach, the variables need to be transformed into Uniform distribution [0,1]. Three types of copula models were compared with the analysis, i.e., Frank, Clayton, and Gaussian Copula. The Comparison was made by means of parameter estimates and the corresponding log likelihood values. The resulting estimates of the parameters and the corresponding log likelihood values are presented in Table 4.

The value of the log likelihood suggests that of the three models, the Frank Copula fit the data very well. The log likelihood of Frank Copula is always bigger than those of others. This occurs in all pairs of variables. The bigger the estimate indicated the stronger the dependency of the two variables. The Frank Copula model of the time lag 2 oil palm production on rainfall data resulted in the highest value of the log likelihood.

Table 4. Estimate of copula parameters of lag production on rainfall

Variables	Copula	Estimates	Log-Likelihood
Oil palm Prod.	Frank	**-1.796**	**9.456**
Versus	Clayton	-0.786	2.654
Rainfall	Gaussian	-1.367	6.506

Lag 1 Prod.	Frank	**-1.823**	**10.328**
versus	Clayton	-0.791	3.438
Rainfall	Gaussian	-1.423	6.836
Lag 2 Prod.	Frank	**-1.984**	**12.452**
versus	Clayton	-0.811	5.041
Rainfall	Gaussian	-1.394	6.873
Lag 3 Prod.	Frank	**-1.872**	**11.942**
versus	Clayton	-0.786	4.847
Rainfall	Gaussian	-1.420	6.765

This suggests that the amount of rainfall could affect the oil palm production in the next two months in this estate. A negative value of the estimate indicates a negative relationship between the two variables.

CONCLUSION

The pattern of rainfall and oil palm production relationships at different time lags mostly follows the Frank Copula model. The strongest relationship between rainfall and oil palm production occurs in lag 2 of oil palm production. This indicates that the amount of rainfall could affect oil palm production in the next two months.

A sequence of simulated, variable specific power levels has been selected for ORIGEN2 fuel depletion computation. This simulated power level sequence incorporates reactivity feedback from earlier burnable poison and fuel depletion.

ACKNOWLEDGEMENT

We would like to thank the Faculty of Mathematics and Natural Sciences for providing the research fund. Thanks should also go to PT. Perkebunan Nusantara XIII and Meteorology, Climatology and Geophysics Agency Class II Siantan Station who provided the data for analysis. Travel assistance and support from the Committee of the 3rd International Symposium is highly appreciated.

REFERENCES

Anisa, KN, Sutikno (2015). Analysis of relationship of rainfall and El-Nino Southern Oscillation indicator in center production of rice in East Java with Copula approach. Jurnal Sains dan Seni ITS, 4(1): 49-54. (in Bahasa Indonesia).

Embrechts P, Lindskog F, McNeil A (2001). Modeling dependence with Copulas and applications to risk management. In: Rachev ST (ed), Handbook of heavy tailed distributions in finance (pp. 22-26). San Diego: Elsevier.

Prasetyo, EI (2009). Analysis of relationship of rainfall and oil palm production with transfer function model [thesis]. Bogor: Institut Pertanian Bogor. (In Bahasa Indonesia)

Genest C, Nešlehová J (2010). Copulas: Introduction to the theory and implementation in R with applications in finance and insurance. Course material for the workshop presented at the 38th Annual Meeting of the Statistical Society of Canada, Quebec (QC), May 23, 2010.

Murteira JM, Lourenço ÓD (2007). Health care utilization and self-assessed health: Specification of bivariate models using Copulas. Health, Econometrics and Data Group. The University of York. ISSN 1751-1976.

Risza S (2009). Oil Palm - an effort to increase productivity. Yogyakarta: Kanisius. (in Bahasa Indonesia).

Schoelzel C, Friederichs P (2008). Multivariate non-normally distributed random variables in climate research – Introduction to The Copula Approach. Nonlin. Processes Geophys, 15: 761–772.

Simanjuntak LN, Sipayung R, Irsal (2014). Influence of rainfall and rainy days on the production of oil palm aged 5, 10 and 15 years at the Garden Begerpang Estate PT. PP London Sumatra Indonesia, Tbk. Jurnal Online Agroteknologi, 2: 1141-1151. (in Bahasa Indonesia).

Syahrir I (2011). Estimation of Copula parameter and its application in climatology. Surabaya: TS. (in Bahasa Indonesia).

Udayani LG, Sumarjaya IW, Susilawati M (2016). Analysis of relationship of rice production and ENSO indicator in Tabanan District with Copula approach. E-Jurnal Matematika, 5: 164-169. (in Bahasa Indonesia).

Wang A (2010). Goodness-of-Fit Tests for Archimedean Copula Models. Statistica Sinica, 20: 441-453.

Zhu Y, Ghosh SK, Goodwin BK (2008). Modeling dependence in the design of whole insurance contract - A Copula - based model approach. Selected Paper for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, Florida, July 27-29, 2008.