BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 1251-1262
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp1251-1262

ANALYSIS OF THE EXISTENCE OF THE AGRICULTURAL SECTOR IN MODELING POVERTY IN BENGKULU PROVINCE USING GAUSSIAN COPULA MARGINAL REGRESSION

Sigit Nugroho[1], Dyah Setyo Rini[2*], Pepi Novianti[3], Riki Crisdianto[4], Elisabeth Evelin Karuna[5], Athaya Fairuzindah[6]

[1,2,3,4,5,6]*Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Bengkulu*
Jln. Wr. Supratman, Kandang Limun, Bengkulu, 38371, Indonesia

*Corresponding author's e-mail: ** ***dyah.setyorini@unib.ac.id***

Article History:

Received: 9th October 2024
Revised: 2rd February 2025
Accepted: 4th March 2025
Published: 1st April 2025

Keywords:

Copula;
Gaussian Copula Marginal Regression (GCMR);
Poverty.

ABSTRACT

Bengkulu Province ranks second in the category of the highest percentage of poor people in the Sumatra region, at 14.62% in March 2022, and sixth in Indonesia, which is undoubtedly one of the fundamental problems that requires mutual attention. The phenomenon of high poverty in Bengkulu Province is inseparable from the lives of people whose main livelihood is in the agricultural sector, especially tenant farmers. Therefore, in this study, the Copula and Gaussian Copula Marginal Regression (GCMR) methods are applied to determine how the agricultural sector affects poverty in Bengkulu Province using secondary data obtained from the Bengkulu Provincial Statistics Agency (Susenas 2022). The results show that the Copula model can identify various types of dependency between the number of poor households in each district/city in Bengkulu Province in 2022 (Y) and each of the X variables, namely the Number of Agricultural Business Households (X_1), the Growth Rate of the Agricultural Sector (X_2), the Human Development Index (X_3), and the Open Unemployment Rate (X_4) by considering the different characteristics of dependency such as top-tail, bottom-tail, or negative dependency. Meanwhile, the GCMR model can provide the direction of influence of the independent variables on the dependent variable Y, where it can be seen that the variables X_1, X_2, and X_3 have a negative influence on the variable Y, whie the variable X_4 has a positive impact on the variable Y. Therefore, in general, it can be concluded that either positive or negative dependencies identified by the Copula model can influence the resulting GCMR model by providing more profound complexity regarding the relationship between the variables analyzed.

How to cite this article:

S. Nugroho, D. S. Rini, P. Novianti, R. Crisdianto, E. E. Karuna and A. Fairuzindah., "ANALYSIS OF THE EXISTENCE OF THE AGRICULTURAL SECTOR IN MODELING POVERTY IN BENGKULU PROVINCE USING GAUSSIAN COPULA MARGINAL REGRESSION," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 1251-1262, June, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Poverty is one of the most complex and multidimensional underlying issues and is a central concern for governments in any country. Poverty often hinders access to education, health, and economic opportunities. As a crucial first step for sustainable development, the SDGs seek to address poverty to improve overall quality of life. The Sustainable Development Goals (SDGs) are one of the global and national commitments to improve society's welfare, including 17 goals, of which the first goal is No Poverty. Therefore, poverty reduction based on the national poverty line is an MDGs indicator that must be implemented as SDGs in Indonesia [1].

One of the keys to achieving the MDGs and SDGs in overcoming poverty is the performance of the agricultural sector. Based on several empirical studies, the role of the agricultural sector in poverty alleviation, which is the epicenter of the SDGs, is enormous, covers quantitative aspects, and involves qualitative dimensions. Therefore, poverty alleviation is urgently needed to reduce poverty in Indonesia, especially in Bengkulu Province.

Bengkulu Province ranks second in the category of the highest percentage of poor people in the Sumatra region, at 14.62% in March 2022, and sixth in Indonesia. This percentage is also higher than that of poor people in Indonesia, at 9.54%. The number of poor people in rural areas is also higher than the number of poor people in urban areas [2]. On the other hand, the dominant sector that absorbs labor in Bengkulu Province is the agricultural sector (the impact of the COVID-19 pandemic is a shift of labor from cities to villages), primarily located in rural areas. This condition explains that the agricultural sector is the livelihood of most people with low incomes in Bengkulu Province due to the existence of the agricultural sector as the primary employment of most people living in rural areas.

The phenomenon of high poverty in Bengkulu Province certainly requires more understanding of what is inseparable from the discussion of the lives of people with their main livelihood, namely in agriculture, especially sharecroppers; of course, it is very necessary to reduce the poverty rate in Bengkulu Province through the agricultural sector. Therefore, this study will examine how the agricultural sector affects poverty in Bengkulu Province using the Copula and Gaussian Copula Marginal Regression (GCMR) methods.

Copula is one of the statistical methods that shows the relationship between variables, and this method is not too strict on distribution assumptions, especially the assumption of normality. In addition, Copula also has the advantage of clearly describing dependencies at extreme points. [3] showed that the Copula method performs better when the assumptions of normality are violated. [4] showed that the Copula model assumes the response variable is more general, not only independent. [5] showed that copulas can determine multivariate correlations. However, the research is still limited to correlation and does not identify the causal relationship. A method that can be used to model causal relationships is Gaussian Copula Marginal Regression (GCMR). This method can overcome the problem of violation of assumptions in classical regression methods, overcome outliers (extreme values) in the data, and has good precision for small samples [6].

One of the studies on GCMR was conducted by [7] to see the relationship between the rice harvest area as a response variable and the rainfall predictor variable that allows extreme observations. These extreme observations often violate the assumption of normality. Therefore, the Gaussian Copula Marginal Regression (GCMR) method is one of the appropriate methods for overcoming this problem. The study results show that the GCMR model is very well used to model response variables that are not normally distributed and have extreme observations. Therefore, this study will apply the Copula and Gaussian Copula Marginal Regression (GCMR) methods to analyze the factors that influence poverty, especially in the agricultural sector in Bengkulu Province. This research is expected to be the basis of government policy to support poverty alleviation in Bengkulu Province through the agricultural sector.

2. RESEARCH METHODS

2.1 Data and Data Sources

The data used in this study is secondary data obtained from the Bengkulu Provincial Statistics Agency (Susenas 2022). The variable used in this study is the number of poor households in each district/city in

Bengkulu Province in 2022 as the response variable (Y). Meanwhile, the predictor variables used are the Number of Agricultural Business Households (X_1), Growth Rate of the Agricultural Sector (X_2), Human Development Index (X_3), and Open Unemployment Rate (X_4). This data will be used to model, identify, and analyze indicators that affect poverty in Bengkulu Province, especially in the agricultural sector, using the Copula and Gaussian Copula Marginal Regression (GCMR) methods.

2.2 Definition of Copula

According to **[8]**, the copula is a method that combines several marginal distributions into a joint distribution that aims to detect dependencies between linear and non-linear random variables. Copula function is a concept that is used as a tool to study the non-linear dependency between events in multivariate cases. Copula is increasingly used in multivariate distribution modeling because it does not require the assumption of normality in the marginals so it is flexible enough to be used **[9]**.

Some advantages of copula include not being strict on distribution assumptions, exceptionally normal distribution, explaining non-linear relationships, and making it easy to build a joint distribution because the marginal distribution of random variables can be different or even the marginal distribution is unknown **[10]**. Copula comes from the Latin copula which means bond or binding. The copula concept was first popularized by a mathematician named Abe Sklar in 1959 whose theorem is known as the Sklar Theorem **[11]**.

2.3 Sklar's Theorem

Sklar's theorem is a theory of copulas and is the basis of many applications of statistical theory. Sklar's theorem explains the role of copulas on multivariate distribution functions with univariate marginal distributions **[12]**. Suppose F and G are the marginal distributions for random variables X and Y respectively then the function H is the joint distribution function for (X, Y). There exists a copula C such that for every x, $y \in R$ holds **[11]**:

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

Based on Sklar's Theorem 1959, the bivariate distribution of $H_{X,Y}(x,\ y)$ can be defined by:

$$\begin{aligned} H_{X,Y}(x,\ y) &= P(X \leq x, Y \leq y) \\ &= C(F_X(x), F_Y(y); \theta), x, y \in R \end{aligned} \tag{2}$$

where θ is the estimated value of the dependency parameter, which can be estimated using the following **Equation (3)** **[13]**:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{T} \log C(F_X(x_i), F_Y(y_i); \theta) \tag{3}$$

Based on the Probability Integral Transform (PIT) theorem, the marginal distribution of variables (X, Y) follows a uniform distribution ranging from 0 to 1. Therefore, by denoting $F_X(x_i)$ and $F_Y(y_i)$ by u_1 and u_2, e can write it as follows:

$$C(F_X, F_Y; \theta) = C(u_1, u_2; \theta), u_1, u_2 \in [0,1] \tag{4}$$

2.4 Copula Family

Copula consists of several Copula families, namely elliptical copula and archimedian copula. Copulas included in the elliptical copula are Gaussian copula and Student's t copula. While the copulas included in the archimedian are Joe copula, Frank copula, Gumble copula and Clayton copula **[14]**.

Some Copula models of elliptical copula and archimedean copula can be seen in the following **Table 1** **[15]**:

Table 1. Copula Model

Copula Model	Function $C(u_1, u_2; \theta)$	Generator $\phi(t; \theta)$	Parameter Range (θ)	*Kendall's tau of the copula C*
Independence	u_1, u_2	$-\log(t)$	n/a	0
Clayton	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{\frac{-1}{\theta}}$	$\theta^{-1}(t^{-\theta} - 1)$	$(0, \infty)$	$\frac{\theta}{\theta + 2}$
Gumbel Hougaard	$\exp\left(-(\tilde{u}_1^{\theta} + \tilde{u}_2^{\theta})^{\frac{1}{\theta}}\right)$; $\tilde{u}_j = -\log u_j$	$(-\log(t))^{\theta}$	$[1, \infty)$	$\frac{\theta - 1}{\theta}$
Frank	$-\frac{1}{\theta}\log\left(1 + \frac{u_1^* u_2^*}{\exp(-\theta) - 1}\right)$; $u_j^* = e^{(-\theta u_j - 1)}$	$-\left(\frac{\exp(-\theta t) - 1}{\exp(-\theta) - 1}\right))$	$(-\infty, \infty)$	$1 - \frac{4}{\theta}[1 - D_1(\theta)]$
Joe	$1 - [u_1^* + u_2^* - u_1^* u_2^*]^{\frac{1}{\theta}}$; $u_j^* = (1 - u_j)^{\theta}$	$-\log(1 - (1 - t)^{\theta})$	$[1, \infty)$	$1 - \sum_{k=1}^{\infty} \frac{4}{h(\theta; k)}$
Gaussian	$\Phi_G[\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta]$	n/a	$(-1,1)$	$\frac{2}{\pi}\arcsin(\theta)$
Student's	$t_{2;v}[t_v^{-1}(u_1), t_v^{-1}(u_2); \theta]$; $v \in (2, \infty)$	n/a	$[-1,1]$	$\frac{2}{\pi}\arcsin(\theta)$

2.5 Copula Regression

The regression Copula model formed depends on the Copula distribution function. The predicted Y value is obtained by finding the expected value of the conditional probability function as shown in the following **Equation (5)**:

$$Y = E[y|x] + (y - E[y|x] = E[y|x] + \varepsilon \tag{5}$$

If x_p is the response variable, then finding the predicted x_p value can be written with the following **Equation (6)**:

$$X_p = E[X_p|X_1, X_2, \dots, X_{p-1}] + (x_p - E[X_p|X_1, X_2, \dots, X_{p-1}]) = E[X_p|X_1, X_2, \dots, X_{p-1}] + \varepsilon \tag{6}$$

Suppose $X_1, X_2,$ and Y are continuous random variables with distribution functions $F_1, F_2,$ and G, marginal density functions $f_1, f_2,$ and g, joint distribution function H and joint probability density function h, respectively. By transforming $U = F_1(X_1)$, $V = F_2(X_2)$, and $W = G(Y)$, we get $X_1 = F_1^{-1}(U)$, $X_2 = F_2^{-1}(V)$, and $Y = G^{-1}(W)$, which can be expressed as follows **[16]**:

$$c(u, v, w) = h((F_1^{-1}(u), F_2^{-1}(v), G^{-1}(w))|J| \tag{7}$$

with:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial u} & \frac{\partial x_1}{\partial v} & \frac{\partial x_1}{\partial w} \\ \frac{\partial x_2}{\partial u} & \frac{\partial x_2}{\partial v} & \frac{\partial x_2}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \frac{1}{f_1(x_1)} & 0 & 0 \\ 0 & \frac{1}{f_2(x_2)} & 0 \\ 0 & 0 & \frac{1}{g(y)} \end{vmatrix} \tag{8}$$

and

$$c(u, v, w) = \frac{\partial^3 C(u, v, w)}{\partial u \partial v \partial w} \tag{9}$$

so that **Equation (7)** becomes:

$$c(u,v,w) = h((F_1^{-1}(u), F_2^{-1}(v), G^{-1}(w))\frac{1}{f_1(x_1)f_2(x_2)g(y)} \quad (10)$$

or the function equation h can be written:

$$h(x_1,x_2,y) = (F_1(x_1), F_2(x_2), G(y))f_1(x_1)f_2(x_2)g(y) \quad (11)$$

To determine the Copula regression equation, **Equation (11)** can be used. The probability density function of the conditional random variables Y, $X_1 = x_1$, $X_2 = x_2$ is:

$$m(y|x_1,x_2) = \frac{h(x_1,x_2,y)}{f_{1,2}(x_1,x_2)} \quad (12)$$

assuming X_1 and X_2 are mutually independent random variables, meaning $f_{1,2}(x_1,x_2) = f_1(x_1)f_2(x_2)$, so we get:

$$\begin{aligned} m(y|x_1,x_2) &= \frac{c(F_1(x_1),F_2(x_2),G(y))f_1(x_1)f_2(x_2)g(y)}{f_1(x_1)f_2(x_2)} \\ &= c(F_1(x_1),F_2(x_2),G(y))g(y) \end{aligned} \quad (13)$$

Based on **Equation (13)**, the expected value of the conditional random variable Y, $X_1 = x_1$, $X_2 = x_2$ is as follows:

$$\begin{aligned} E(Y|X_1 = x_1, X_2 = x_2) &= \int_{-\infty}^{\infty} y\, m(y|x_1,x_2)\, dy \\ &= \int_{-\infty}^{\infty} y\, c(F_1(x_1),F_2(x_2),G(y))g(y)\ dy \end{aligned} \quad (14)$$

Since $w = G(y)$, hence $G^{-1}(w)$ and $dw = g(y)dy$, the Copula regression equation can be written as follows:

$$E(y|X_1 = x_1, X_2 = x_2) = \int_{-\infty}^{\infty} G^{-1}(w)\, c(u,v,w)\, dw \quad (15)$$

2.6 Gaussian Copula Marginal Regression (GCMR)

The general form of the Gaussian Copula Marginal Regression model is as follows **[6]**:

$$Y_i = g(\boldsymbol{x_i}, \epsilon_i; \boldsymbol{\lambda}), \quad i = 1, \dots, n \quad (16)$$

where $g(.)$ is the corresponding regression function on $\boldsymbol{x_i}$, ϵ_i is the error of the model, and $\boldsymbol{\lambda}$ is the parameter. Among the many possibilities for $g(.)$,, the selection of the appropriate model is as follows:

$$Y_i = F_i^{-1}\{\Phi(\epsilon_i); \boldsymbol{\lambda}\}, \quad i = 1, \dots, n \quad (17)$$

Where ϵ_i is the standard normal variable and $F_i(.;\boldsymbol{\lambda}) = F(.|\boldsymbol{x_i};\boldsymbol{\lambda})$ and $\Phi(.)$ are the cumulative distribution function of Y given by $\boldsymbol{x_i}$ and the standard normal variation, respectively. By the integral transformation theorem, the regression model in Equation (17) can ascertain the desired marginal distribution for the response Y_i and define ϵ_i in terms of the generalized normal error. Equation (17) covers all possible parametric regression models for continuous and noncontinuous responses.

For example, the Gaussian Linear Regression Model $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \sigma\epsilon_i$ corresponds to $F_i(Y_i;\boldsymbol{\lambda}) = \Phi\{(Y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})/\sigma\}$ with $\boldsymbol{\lambda} = (\boldsymbol{\beta}^T,\sigma)^T$, while the log-linear Poisson model can be written as follows:

$$F_i(Y_i;\boldsymbol{\lambda}) = \sum_{j=0}^{Y_i} \frac{e^{-\mu_i}\mu_i^j}{j!} \tag{18}$$

where $\mu_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$ with $\boldsymbol{\lambda} \equiv \boldsymbol{\beta}$.

2.7 Poverty

Poverty is seen as an economic inability to meet basic food and non-food needs measured in expenditure. Based on the poverty measurement conducted by the Central Bureau of Statistics (BPS) using the concept of the ability to meet basic needs (basic needs approach), the poor are defined as people who have an average expenditure per capita per month below the Poverty Line (PL). Technically, the PL is constructed from two components, namely the Food Poverty Line (FPL) and the Non-Food Poverty Line (NFPL). The food poverty line is the minimum food expenditure equivalent to 2,100 kilocalories per capita per day. In contrast, the non-food poverty line is the minimum need for housing, clothing, education, and health [17].

2.8 Research Activity Design

The design or stages in conducting research with the Copula and GCMR methods are as follows:

a. Literature Review and Methodological Exploration

The literature review in this research will be conducted by collecting literature sources relevant to the study. Methodology exploration is intended to review the methods that will be used in the study.

b. Data Collection

This study will use secondary data related to poverty variables in Bengkulu Province based on Susenas data from March 2022. The raw data originates from BPS Bengkulu Province. At this stage, raw data cleaning is also carried out for use in the next stage.

c. Data Exploration

The purpose of data exploration is to provide a complete and accurate description of poverty data in Bengkulu Province. Furthermore, the poverty condition in Bengkulu Province was mapped using district/city poverty percentage data with Arcview GIS software.

d. Copula Modeling

This stage is carried out to determine whether or not there is a correlation between the variables used using the Copula method.

e. GCMR Modeling

After knowing a close relationship exists between variables, this stage models these variables using GCMR. Based on this modeling, it will be known that the agricultural sector will influence poverty in Bengkulu Province.

f. Conclusion

The model obtained in stages 4 and 5 will be interpreted (analyzed) to determine the relationship between the indicators obtained and poverty.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

Descriptive statistics is one of the initial stages needed to determine the characteristics or general description of the data used in further analysis. The descriptive statistics based on several variables used, namely the Number of Poor Households (Y), Number of Agricultural Business Households (X_1), Agricultural Sector Growth Rate (X_2), Human Development Index (X_3), and Open Unemployment Rate (X_4), can be seen in **Table 2** below.

Table 2. Descriptive Statistics

Variable	Min	Mean	Max	Skewness	Kurtosis
Y	5.00	31.30	63.00	0.27	-0.99
X1	10698	31913.1	60351	0.46	-1.00
X2	0.53	3.06	7.05	0.73	-0.72
X3	68.27	70.99	81.45	1.79	2.10
X4	64.96	70.90	77.86	0.25	-1.69

Table 2 generally states that variable Y has the widest range, while variable X_2 has the smallest range. Variable X_1 shows a considerable variation, whereas variables X_3 and X4 have smaller ranges and averages that are quite close to the minimum and maximum values. Variable X_3 and X_4 tend to have a more uniform distribution than the other variables. **Table 2** also provides information about the skewness and kurtosis values, where the skewness value offers an overview of the data distribution whether it is skewed to the left, right, or symmetrical. The kurtosis value indicates whether the data distribution tends to be flat or pointed. Overall, all variables tend to be skewed to the right because they have positive skewness values. For kurtosis, it can be seen that variables Y, X_1, X_2, and X_4 have negative kurtosis values, meaning that this value is far from the ideal condition where the distribution of these data has fewer extreme values.

3.2 Normality Test

Normality test is one of the tests that need to be done to determine whether the data used is normally distributed or not. In this study, the normality test used is the Kolmogorov Smirnov test, where the test results are shown in **Table 3** below:

Table 3. Normality test results

Variable	D-statistic	*p-value*	Decision	Meaning
Y	0.20701	0.7848	Accept H_0	Data is normally distributed
X1	0.18868	0.806	Accept H_0	Data is normally distributed
X2	0.27675	0.3598	Accept H_0	Data is normally distributed
X3	0.27141	0.3828	Accept H_0	Data is normally distributed
X4	0.19475	0.7756	Accept H_0	Data is normally distributed

The normality test results in **Table 3** show the $p_{value} > \alpha$, meaning that H_0 is accepted, giving a conclusion that all variables were normally distributed. After testing the normality of the data, the next step is to test the correlation between each random variable. Based on the normality test conducted, it is known that the data used is normally distributed. Pearson correlation is one of the appropriate correlation tests to use on normally distributed data. However, in this research, correlation testing will be carried out using three correlation tests, namely Pearson Correlation, Kendall Tau Correlation, and Spearman Correlation, as in the next Subchapter.

3.3 Correlation Test

Identifying the relationship between variables is one of the indispensable things in modeling Copula and Gaussian Copula Marginal Regression (GCMR). This study will identify the relationship between variables using three correlation tests: Pearson Correlation, Kendall Tau Correlation, and Spearman Correlation. The test results of the three tests are presented in **Table 4**.

Table 4. Correlation Test Results Between Variables

Pair of Variables	Pearson			Kendall Tau			Spearman		
	Correlation	t-statistic	p-value	Correlation	z-statistic	p-value	Correlation	S-statistic	p-value
Y and X1	-0.0099	-0.0279	0.9784	0.0000	0.0000	1.0000	0.0608	154.9700	0.8675
Y and X2	-0.3517	-1.0628	0.3189	-0.2247	-0.8980	0.3692	-0.3526	223.1800	0.3177
Y and X3	-0.3292	-0.9861	0.3529	-0.3146	-1.2572	0.2087	-0.4316	236.2200	0.2129
Y and X4	0.6114	2.1854	0.0604	0.4045	1.6164	0.1060	0.5836	68.7080	0.0765
X1 and X2	-0.1932	-0.5568	0.5929	-0.1111	20.0000	0.7275	-0.2606	208.0000	0.4697
X1 and X3	-0.3523	-1.0648	0.3180	0.0222	23.0000	1.0000	0.0424	158.0000	0.9186
X1 and X4	0.3152	0.9395	0.3750	0.3778	31.0000	0.1557	0.5152	80.0000	0.1328
X2 and X3	0.0631	0.1790	0.8624	-0.2000	18.0000	0.4843	-0.2485	206.0000	0.4916
X2 and X4	-0.4551	-1.4457	0.1863	-0.5556	10.0000	0.0286	-0.6606	274.0000	0.0440
X3 and X4	-0.1215	-0.3463	0.7380	0.1111	25.0000	0.7275	0.1152	146.0000	0.7588

The correlation test results in Table 4 show that statistically significant correlations are only found between variables X2 and X4 in the Kendall Tau and Spearman methods with a $p_{value} < \alpha$. However, in general, when viewed based on the correlation value, each pair of variables correlates both positively and negatively, but the correlation value is not statistically significant. After identifying the correlation between each pair of variables, the next step is to perform Copula and Gaussian Copula Marginal Regression (GCMR) modeling.

3.4 Copula Modelling

One of the steps that needs to be done before performing Copula modeling is identifying the marginal distribution. The marginal distribution is determined using several appropriate distributions for each variable; then, we select the best distribution based on the smallest AIC and BIC values. The results of the marginal distribution identification can be seen in Table 5 below.

Table 5. Identify the Marginal Distribution

Variable	GOF Criterion	lnorm	Norm	Weibull	Gamma	Unif	Logis
Y	AIC	90.07	88.07	87.46	88.65	**85.21**	88.43
	BIC	90.68	88.68	88.06	89.25	**85.81**	89.04
X_1	AIC	222.79	223.40	222.60	222.43	220.26	**88.43**
	BIC	223.39	224.00	223.20	223.03	220.86	**89.04**
X_2	AIC	43.25	44.77	42.53	42.45	**41.50**	44.83
	BIC	43.86	45.38	43.14	43.06	**42.10**	45.43
X_3	AIC	57.65	58.69	63.82	57.99	**55.57**	55.85
	BIC	58.25	59.29	64.43	58.59	**56.18**	56.46
X_4	AIC	61.29	61.46	62.46	61.34	**55.14**	62.50
	BIC	61.89	62.06	63.07	61.94	**55.75**	63.10

Table 5 shows the results of identifying the marginal distribution of each variable used based on the smallest AIC and BIC values. The results show that variables Y, X_2, X_3, and X_4 follow the Uniform distribution, while variable X_1 follows Logistic distribution. After identifying the marginal distribution, the next step is to select the Copula model from each pair of variables. The results of the Copula model selection are presented in Table 6 below.

Table 6. Copula Model Selection

Variable	Copula	Log-likelihood	AIC	BIC	Selected Copula		
					Tipe	Parameter	Kendall-Tau
Y - X_1	Gaussian	0.04	1.93	2.23	Joe	1.48	0.21
Y~Unif (5,63)	t	0.03	3.94	4.54			

Variable	Copula	Log-likelihood	AIC	BIC	Selected Copula		
					Tipe	Parameter	Kendall-Tau
X_1~log (30.59, 9.34)	Clayton	0.35	1.31	1.61			
	Gumbel	0.19	1.63	1.93			
	Frank	0.02	1.97	2.27			
	Joe	**0.42**	**1.16**	**1.46**			
Y- X_2 Y~Unif (5,63) X_2~ Unif (0.53, 7.05)	Gaussian	0.73	0.54	0.84	Clayton	-0.82	-0.29
	t	0.67	2.67	3.28			
	Clayton	**0.82**	**0.37**	**0.67**			
	Gumbel	0.61	0.77	1.08			
	Frank	0.55	0.9	1.21			
	Joe	0.65	0.69	0.99			
Y- X_3 Y~Unif (5,63) X_3~log (68.27, 81.45)	Gaussian	**1.17**	**-0.34**	**-0.04**	Gaussian	-0.59	-0.4
	t	1.11	1.77	2.38			
	Clayton	1.04	-0.09	0.22			
	Gumbel	0.98	0.03	0.34			
	Frank	0.91	0.18	0.49			
	Joe	0.9	0.21	0.51			
Y- X_4 Y~Unif (5,63) X_4~log (64.96, 77.86)	Gaussian	1.99	-1.97	-1.67	Frank	5.2	0.47
	t	1.98	0.04	0.64			
	Clayton	1.09	-0.18	0.12			
	Gumbel	1.53	-1.05	-0.75			
	Frank	**2.05**	**-2.09**	**-1.79**			
	Joe	0.92	0.17	0.47			

Table 6 shows the results of the Copula model selection for each pair of variables used. For Y and X_1 follow Copula Joe (1.48;0,21), which means that the parameter value θ_{YX_1} of 1.48 indicates that the variables Y and X_1 have an upper (positive) tail dependence, where the data pattern will tend to cluster at the bottom of the tail. For Y and X_2 follow the Clayton Copula (-0.82;-0.29), which means that the parameter value θ_{YX_2} of -0.82 is able to explain that the variables Y and X_2 have a strong lower tail dependence but not on the upper tail. For Y and X_3 follow the Copula Gaussian (-0.59;-0.4), which means that the parameter value θ_{YX_3} of -0.59 indicates that the variables Y and X_3 have a negative dependency. Furthermore, for Y and X_4 follow the Frank Copula (5.2;0.47). In Frank Copula, it should be noted that the dependence between variables can be explained broadly either positively or negatively. That is, the larger the parameter θ_{YX_4}, the stronger the positive dependency captured by the Copula. Therefore, it can be concluded that the parameter value θ_{YX_4} of 5.2 indicates that the variables Y and X_4 have a fairly strong positive dependency. After identifying the marginal distribution and selecting the Copula model, then the Copula modeling can be done as follows:

a. Joe's Copula Model for Y and X_1

$$H_{X_1,Y}(x,y) = 1 - \left[\left(1 - \frac{x-5}{63-5}\right)^{1.48} + \left(1 - \frac{1}{1+\exp\left(-\frac{(x-30.59)}{9.34}\right)}\right)^{1.48}\right.$$

$$-\left(1-\frac{x-5}{63-5}\right)^{1.48}\left(1-\frac{1}{1+\exp\left(-\frac{(x-30.59)}{9.34}\right)}\right)^{1.48}\Bigg]^{1/1.48}$$

b. Clayton Copula Model for Y and X_2

$$H_{X_2,Y}(x,y) \quad = \quad \left(\left(\frac{x-5}{63-5}\right)^{0.82}+\left(\frac{x-0.53}{7.05-0.53}\right)^{0.82}-1\right)^{1/0.82}$$

c. Gaussian Copula Model for Y and X_3

$$H_{X_3,Y}(x,y) \quad = \quad \Phi_{\mathrm{G}}\left[\Phi^{-1}\left(\left(\frac{x-5}{63-5}\right)\right),\Phi^{-1}\left(\frac{x-68.27}{81.45-68.27}\right);-0.59\right]$$

d. Frank Copula Model for Y and X_4

$$H_{X_4,Y}(x,y) \quad = \quad -\frac{1}{5.2}log\left(1+\frac{e^{\left(-(5.2)\left(\frac{x-5}{63-5}\right)-1\right)}e^{\left(-(5.2)\left(\frac{x-64.96}{77.86-64.96}\right)-1\right)}}{\exp(-5.2)-1}\right)$$

3.5 Gaussian Copula Marginal Regression (GCMR) Modeling

The basic concept of GCMR modeling is to combine a marginal regression model with a Gaussian Copula matrix. The correlation matrix of the Copula will make it possible to handle various forms of dependence that arise, both in the analysis of longitudinal data, time series, and geostatistics [6]. In this study, the GCMR model analysis was carried out using the gcmr package in the R program. The results of the GCMR model analysis can be seen in **Table 7**.

Table 7. Estimation of Parameters GCMR

	Estimation	Std. Error	Z-value	P-value
(Intercept)	3.79E+00	8.48E-07	4.47E+06	< 2E-16
X1	-1.31E-05	1.77E-06	-7.41E+00	1.27E-13
X2	-3.47E-02	2.09E-06	-1.66E+04	< 2E-16
X3	-7.69E-02	5.88E-05	-1.31E+03	< 2E-16
X4	7.82E-02	5.77E-05	1.36E+03	< 2E-16
Gaussian Copula (θ)	2.69E-03	4.53E-10	5.93E+06	<2E-16

Table 7 shows the parameter estimation results of the GCMR model. The results show that all independent variables, namely the Number of Agricultural Business Households (X_1), Agricultural Sector Growth Rate (X_2), Human Development Index (X_3), and Open Unemployment Rate (X_4) have a significant effect on the dependent variable Number of Poor Households (Y) with its Gaussian Copula parameter $\theta = 2.69\mathrm{E}-03$. Based on the estimation values obtained, the GCMR model can be made as follows:

$$\hat{Y}_i = g(3.79-1.31E-05X_1-3.47E-02X_2-7.68E-0.2X_3+7.82E-02X_4;2.69E-03)$$

This GCMR model indicates that based on the parameter estimation values obtained for $\beta_0 = 3.79$, $\beta_1 = -1.31E-05$, $\beta_2 = -3.47E-02$, and $\beta_3 = -7.68E-0.2$, it can be concluded that variables X_1, X_2, and X_3 have a negative influence on the dependent variable Y. The higher the number of agricultural business households, the growth rate of the agricultural sector, and the Human Development Index, the lower the number of low-income families. Meanwhile, the parameter $\beta_4 = 7.82E - 02$ shows that the variable X_4 has a positive influence on the dependent variable Y. So, the higher the Open Unemployment Rate, the higher the Number of Poor Households. However, when looking at the relationship or correlation between variables in the Gaussian Copula model, it can be concluded that the dependency in the Gaussian Copula model is very weak with a parameter value of θ=2.69E-03, although this parameter is statistically significant, it can be said that the dependency effect captured by the Copula is almost non-existent. This study also provides information that economic and social variables have a significant influence on poverty, which is in line with several previous studies that show that factors such as the agricultural sector, human development, and unemployment can affect poverty levels, especially in regions with economies that are primarily dependent on agriculture. However, the dependency between the variables modeled with the GCMR shows a reasonably low dependency. Poverty depends not only on linearly interrelated factors but also on other external factors not observed in this model.

4. CONCLUSIONS

Based on the results of existing research, it can be concluded that the Copula model can identify various types of dependence between variable Y and each variable X_1, X_2, X_3, and X_4, taking into account differences in dependence characteristics such as up-tail, down-tail, or negative dependence. Meanwhile, the GCMR model can provide the direction of influence of the independent variables on the dependent variable Y, where it can be seen that variables X_1, X_2, and X_3 have a negative influence on variable Y, while variable X_4 has a positive influence on variable Y. Therefore, it can be concluded that positive or negative dependencies identified by the Copula model can influence the resulting GCMR model by providing more profound complexity regarding the relationship between the variables analyzed.

ACKNOWLEDGMENT

The research team would like to thank the Faculty of Mathematics and Natural Sciences University of Bengkulu for funding the activity through the Unggulan 2024 scheme. In addition, this research is also supported by LPPM University of Bengkulu.

REFERENCES

[1] Bappenas, PEDOMAN TEKNIS PENYUSUNAN RENCANA AKSI TUJUAN PEMBANGUNAN BERKELANJUTAN (TPB)/ SUSTAINABLE DEVELOPMENT GOALS (SDGs), Jakarta: Bappenas, 2020.

[2] BPS, BERITA RESMI PROFIL KEMISKINAN DI INDONESIA MARET 2022, Jakarta: BPS, 2022.

[3] P. P. Oktaviana, "PENDEKATAN COPULA UNTUK PENYUSUNAN PETA KERAWANAN PUSO TANAMAN PADI DI JAWA TIMUR DENGAN INDIKATOR EL-NINO SOUTHERN OSCILLATION (ENSO)," 2012.

[4] H. Joe, DEPENDENCE MODELING WITH COPULAS, New York: Chapman and Hall/CRC, 2014.

[5] M. Hofert, I. Kojadinovic, M. Machler and J. Yan, "ELEMENTS OF COPULA MODELING WITH R," pp. 9-79, 2017.

[6] G. Masarotto and C. Varin, "GAUSSIAN COPULA MARGINAL REGRESSION," *Electronic Journal of Statistics*, vol. 6, pp. 1517-1549, 2012.

[7] Sutikno, H. Kuswanto and D. Ratih, "GAUSSIAN COPULA MARGINAL REGRESSION FOR MODELING EXTREME DATA WITH APPLICATION," *Journal of Mathematics and Statistics*, vol. 10, no. 2, pp. 192-200, 2014.

[8] C. Czado, R. Kastenmeier, E. C. Brechmann and A. Min, "A MIXED COPULA MODEL FOR INSURANCE CLAIMS AND CLAIM SIZES," *Scandinavian Actuarial Journal*, vol. 4, pp. 278-305, 2012.

[9] H. O. Pintari and R. Subekti, "PENERAPAN METODE GARCH-VINE COPULA UNTUK ESTIMASI VALUE AT RISK (VAR) PADA PORTOFOLIO," *Jurnal Fourier*, vol. 7, no. 2, pp. 63-77, 2018.

[10] S. W. Agustini, M. Hadijati and N. Fitriyani, "ANALISIS DEPENDENSI FAKTOR MAKROEKONOMI TERHADAP TINGKAT HARGA EMAS," *Eigen Mathematics Journal*, pp. 82-91, 2019.

[11] R. B. Nelsen, An Introduction to Copulas, Portland: Springer, 2006.

[12] A. Khairiati, R. Budiarti and I. G. P. Purnaba, "PERBANDINGAN ANALISIS REGRESI LINEAR DENGAN," *JAMBURA JOURNAL OF MATHEMATICS,* vol. 4, no. 2, pp. 209-219, 2022.

[13] U. Cherubini, E. Luciano and W. Vecchiato, COPULA METHODS IN FINANCE., John Wiley & Sons, 2004.

[14] I. S. Prayoga and A. Ahdika, "PEMODELAN KERUGIAN BENCANA BANJIR AKIBAT CURAH HUJAN EKSTREM MENGGUNAKAN EVT DAN COPULA," *Jurnal Aplikasi Statistika & Komputasi Statistik,* vol. 13, no. 1, pp. 35-46, 2021.

[15] J. Rizal, A. Y. Gunawan, S. Indratno and I. Meilano, "THE APPLICATION OF COPULA CONTINUOUS EXTENSION TECHNIQUE FOR BIVARIATE DISCRETE DATA: A," *Mathematical Modelling of Engineering Problems,* vol. 8, no. 5, pp. 793-804, 2021.

[16] A. Khairiati, R. Budiarti and I. G. P. Purnaba, "PERBANDINGAN ANALISIS REGRESI LINEAR DENGAN ANALISIS REGRESI COPULA PADA DATA KEUANGAN," *Jambura Journal of Mathematics,* vol. 4, no. 2, pp. 209-219, 2022.

[17] BPS, DATA STRATEGIS BPS, Jakarta: BPS, 2012.