CAUCHY – Jurnal Matematika Murni dan Aplikasi
Volume 11 (1) (2026), Pages 1171-1186
p-ISSN: 2086-0382; e-ISSN: 2477-3344

Risk Analysis of Shallot Farm Income Using D-vine Copula-Based Monte Carlo Simulation

Fatimah Fuzzaroh, Berlian Setiawaty*, and I Gusti Putu Purnaba

School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia

Abstract

Shallot farm income is highly uncertain due to fluctuations in yields, prices, and production costs, which are interdependent and significantly correlated. This study evaluates income risk by modeling the dependence structure among the variables that constitute income, while addressing data limitations. Two approaches are employed. First, a parametric approach models income as a univariate variable under the assumption of a normal distribution, ignoring dependence among its components. Second, a multivariate simulation approach utilizes a D-vine copula, combined with Monte Carlo simulation, to capture the dependence among income components and generate synthetic observations that better represent tail behavior. Risk is measured using Value-at-Risk (VaR) and Expected Shortfall (ES) based on 32 observations of average shallot farm income per harvest season over the period 2014–2024, and the results are compared with empirical estimates. Due to limited data, the empirical approach produces relatively coarse estimates, particularly in the tail region. The normal distribution approach yields higher and smoother estimates, indicating a higher level of risk. In contrast, the D-vine copula approach provides lower estimates than the normal distribution. These differences indicate that each method offers a distinct perspective on income risk.

Keywords: D-vine copula; Expected shortfall; Monte Carlo simulation; Value at Risk.

1. Introduction

Shallots are among the horticultural commodities in high demand for household consumption and the food industry in Indonesia. This makes shallots a strategic commodity for farming. Shallots drive regional economic growth, create community jobs, and provide income for farmers and agricultural business actors [1]. Cirebon Regency contributes 20.70% to total shallot production in West Java Province and ranks second among shallot-producing regions in the province [2].

Despite its economic importance, the income from shallot farming is unstable and subject to various sources of uncertainty. The income of shallot farmers is inherently uncertain and fluctuates due to factors such as price fluctuations, seasonal differences, and production variability. Fluctuations in shallot commodity prices directly lead to uncertain incomes [3]. In addition, differences in planting seasons contribute to variations in farmers' income across periods [4]. Furthermore, production risk constitutes another important source of income uncertainty, as it is influenced by factors such as production costs and planting seasons [5]. Empirical studies further confirm that shallot farming is associated with high income risk and variability, indicating substantial uncertainty in farmers' income [6, 7]

*Corresponding author. E-mail: berlianse@apps.ipb.ac.id

Submitted: March 13, 2026 Reviewed: April 26, 2026 Accepted: May 08, 2026
DOI: https://doi.org/10.18860/cauchy.v11i1.41556

Income in agriculture is often represented as a single random variable, masking its actual underlying structure and dependencies. Accurate modeling and assessing farm income risk requires understanding its component variables:

$$Y = (X_1 \times X_2) - (X_3 + X_4 + C_1) \tag{1}$$

where Y represents income, X_1 denotes the crop yield , X_2 denotes the selling price, X_3 represents the total labor cost incurred over one production season, X_4 represents the total seed cost incurred over one production season, and C_1 denotes all other production costs. Each component of income has dependencies among variables that cannot be ignored. Previous studies have shown that ignoring dependence among risk factors may lead to inaccurate risk estimation, particularly in extreme conditions. Models that fail to capture co-movement dynamics tend to underestimate joint risk, whereas more structured approaches can provide better risk assessment [8]. This issue is also relevant in modeling farm income, where yield, price, and cost are interdependent. Therefore, copula methods are used to capture the dependence structure among these variables.

Most Indonesian farmers do not keep systematic farming records, as a result, data on productivity, selling prices, and production costs are often incomplete and inconsistent [9]. These data limitations create challenges for agricultural risk analysis. High income risk and limited historical data require a careful analysis of shallot farming income risk.

To address both the dependence structure among variables and the limitation of available data, this study employs a copula-based simulation approach. Specifically, a D-vine copula is selected to model the dependence structure among income components, as it provides a flexible framework for capturing complex pairwise dependencies [10]. In the context of agricultural risk, [11] demonstrates that copula modeling is effective in capturing yield–climate dependence, highlighting its ability to represent joint variability driven by environmental factors. In addition, recent studies have extended this framework by using vine copulas to model more flexible, higher-dimensional dependence structures in agricultural applications [12]. A Monte Carlo simulation is then applied to the selected D-vine copula model to generate synthetic observations and enrich the tail behavior of the distribution under data limitations [13].

This study analyzes income risk using two modeling approaches. First, a parametric approach is employed, in which income is modeled as a univariate variable assumed to follow a normal distribution, selected based on its goodness-of-fit to the observed data. While this approach adequately captures the marginal behavior of income, it ignores the dependence among the underlying variables and may fail to represent joint extreme events, such as low prices and high production costs. To address this limitation, a second approach is adopted in which income is reconstructed using a Monte Carlo simulation based on a D-vine copula model. This approach preserves the dependence structure among the components, allowing for a more comprehensive representation of income risk. Both methods are compared with empirical data to assess how well they represent income behavior.

The comparison focuses on risk measures, particularly Value-at-Risk (VaR) and Expected Shortfall (ES), to evaluate the extent of income risk under each modeling approach. By incorporating both parametric and dependence-based simulation methods, this study aims to analyze and compare the income risk of shallot farming in Pabedilan and Babakan villages, Cirebon Regency with a particular emphasis on VaR and ES as risk indicators.

2. Methods

This section describes the data source and the methodological framework employed in this study. It begins with an overview of the data, followed by the univariate and multivariate approaches used to model the income components. In the multivariate framework, copulas are applied to capture the dependence structure among the variables constituting income, with particular emphasis on the D-vine copula. Furthermore, a Monte Carlo simulation based on D-vine copula is conducted to generate synthetic observations and better represent tail behavior for risk estimation purposes. Subsequently, VaR and ES are estimated under different approaches.

2.1. Data Sources

The data used in this study consisted of primary and secondary data. Primary data were collected through surveys and direct interviews with 10 advanced shallot farmers who had at least 10 years of planting experience and cultivated a minimum land area of 1 hectare in Babakan and Pabedilan villages, Cirebon Regency. Primary data included: crop yields (kg/ha), selling price (IDR/kg), labor cost (IDR/ha), and income (IDR/ha) during the 2014-2024 per harvest season.

From the data collection, 62 observations were obtained. Several observations correspond to the same harvest period but come from different farmers. To ensure comparability and reduce individual-level variability, the data were aggregated at the harvest-period level. Specifically, for each harvest period, observations belonging to the same period were combined using the arithmetic mean. This aggregation process resulted in 32 distinct harvest periods. This aggregation implies that increasing the number of sampled farmers within the same harvest period would primarily refine the estimated mean values rather than substantially increase the number of independent observations.

Secondary data were obtained from various official sources, including Statistics Indonesia (BPS), the Ministry of Agriculture, the Cirebon Regency Agriculture Office, the Commodity Futures Trading Supervisory Agency, as well as relevant literature and journals. These data included standardized production costs covering labor, seed, and other inputs such as fertilizers and pesticides, based on government recommendations. In this study, production costs reported by farmers were replaced with these standardized values to reduce bias arising from differences in farmers' efficiency and input usage. As a result, income calculations are more comparable across observations.

2.2. Univariate Approach

To model income risk, the analysis is conducted on the loss variable defined as $L = -Y$. Several candidate univariate distributions were fitted using Maximum Likelihood Estimation (MLE), and their goodness-of-fit was evaluated using the Anderson–Darling (AD) statistic, which emphasizes discrepancies in the distribution tails. Based on this criterion, the normal distribution was selected as the most appropriate marginal model. Accordingly, the loss variable L is assumed to follow a normal distribution. The corresponding probability density function (PDF) and cumulative distribution function (CDF) are given by [14]:

$$
\begin{aligned}
f_L(l;\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(l-\mu)^2}{2\sigma^2}\right), \quad l \in \mathbb{R}, \\
F_L(l;\mu,\sigma) &= \Phi\left(\frac{l-\mu}{\sigma}\right),
\end{aligned}
$$

where $\mu \in \mathbb{R}$ is the mean parameter, $\sigma > 0$ is the standard deviation, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

2.3. Multivariate Approach

This subsection describes the multivariate approach used to model the joint behavior of income components. It begins with the specification of marginal distributions, followed by the application of copulas to capture dependence, with particular emphasis on the D-vine copula structure.

2.3.1. Distribution of Income Components

The random variable components of income (X_1,X_2,X_3, and X_4) are modeled with generalized logistic, log-logistic, and Weibull distributions based on their data characteristics and on goodness-of-fit results from the AD test. A random variable is said to follow a generalized logistic distribution with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma > 0$, and shape parameter $k > 0$ if its PDF and CDF are given by [15]:

$$f(x;\mu,\sigma,k) = \frac{k}{\sigma}\frac{e^{-\frac{(x-\mu)}{\sigma}}}{\left[1+e^{-\frac{(x-\mu)}{\sigma}}\right]^{k+1}}, \quad -\infty < x < \infty,$$

$$F(x;\mu,\sigma,k) = \left[1+e^{-\frac{(x-\mu)}{\sigma}}\right]^{-k}, \quad -\infty < x < \infty.$$

A random variable is said to follow a log-logistic distribution with shape parameter $\alpha > 0$ and the scale parameter $\beta > 0$ if its PDF and CDF are given by [15]:

$$f(x;\alpha,\beta) = \frac{\alpha(x/\beta)^{\alpha-1}}{\beta[1+(x/\beta)^{\alpha}]^2}, \quad x > 0,$$

$$F(x;\alpha,\beta) = \frac{1}{1+(x/\beta)^{-\alpha}}, \quad x > 0.$$

A random variable is said to follow a Weibull distribution with the shape parameter $\alpha > 0$ and the scale parameter $\beta > 0$ if its PDF and CDF are given by [15]:

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \quad x \geq 0,$$

$$F(x;\alpha,\beta) = 1-\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \quad x \geq 0.$$

The cost component C_1, which includes fertilizers, pesticides, and other miscellaneous expenses, is treated as a piecewise-constant variable reflecting government-regulated subsidized prices. Over the 2014–2024 period, C_1 takes only two values corresponding to policy changes: $c_{1,1} = 26316$ (10^3IDR/ha) for the 2014–2019 period and $c_{1,2} = 34343$ (10^3IDR/ha) for the 2020–2024 period.

In the simulation, the value of C_1 is assigned according to the production period represented in the data. Since labor cost X_3 exhibits a clear increasing trend over time, it is used as a proxy for the production period. Threshold $= 34534$ (10^3IDR/ha) is defined based on the observed data. For each simulated observation, if $X_3 <$ threshold, then C_1 is assigned $c_{1,1}$; otherwise, if $X_3 \geq$ threshold, then C_1 is assigned $c_{1,2}$.

2.3.2. Copulas

Copulas provide a flexible approach to representing the dependence structure between random variables. The fundamental concept of a copula is to link marginal distributions to a joint distribution, as stated in Sklar's Theorem. Suppose that the function F is the joint cumulative distribution function of two random variables, and F_1 and F_2 represent the marginal cumulative distribution functions of each random variable. Based on Sklar's Theorem, there exist a copula function C of 2-dimension that connects these marginal functions, such that for every $x_1, x_2 \in \mathbb{R}$, the following holds [16]:

$$F(x_1,x_2) = C\big(F_1(x_1), F_2(x_2)\big).$$

Let $U_1 = F_1(X_1)$ and $U_2 = F_2(X_2)$, then U_1 and U_2 are uniform random variables on $[0,1]$. Copula C can be constructed from the joint distribution function F through quantile transformation as [16]:

$$C(u_1,u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right), \quad u_1,u_2 \in [0,1],$$

where

$$u_1 = F_1(x_1) \text{ and } u_2 = F_2(x_2).$$

The parameters of the marginal distributions and the copula are estimated using the Inference Functions for Margins (IFM) method, which is a two-step estimation procedure. In the first step, the parameters of the marginal distributions are estimated by maximizing the log-likelihood function of each marginal distribution. In the second step, the copula parameters are estimated by maximizing the log-likelihood function of the copula using the transformed data $u_i = F_i(x_i)$.

Suppose $C(u_1, u_2)$ is a bivariate copula that is differentiable with $u_1, u_2 \in [0, 1]$. The conditional distribution function is defined as [16]:

$$h_{1|2}(u_1 \mid u_2) = \frac{\partial C(u_1, u_2)}{\partial u_2}.$$

In the pair-copula construction (PCC) framework, these conditional distribution functions, known as h-functions, are used to transform variables into conditional uniform variables. These transformed variables are then used as inputs for higher-level copulas in the vine structure.

Several copula families were considered in this study, including Gaussian, student-t, Clayton, Gumbel, Frank, and Joe copulas. The selection of the best-fitting copula was based on the AIC, where the model with the smallest AIC value was preferred. Based on model selection criteria, the copulas that best fit the data are the Clayton, Joe, and student-t copulas. Therefore, only these selected copulas are described in detail below.

For two random variables U_1 and U_3 with parameters $\theta > 0$ that are distributed uniform(0,1), the Clayton copula is defined as [16]:

$$C_\theta(u_1, u_3) = \left(u_1^{-\theta} + u_3^{-\theta} - 1\right)^{-1/\theta}, \quad 0 < \theta < \infty,$$
$$h_{1|3}(u_1 \mid u_3) = \left(u_1^{-\theta} + u_3^{-\theta} - 1\right)^{-1/\theta-1} u_3^{-\theta-1}, \quad u_1, u_3 \in [0, 1].$$

For two random variables U_2 and U_4 with parameters θ that are distributed uniform(0,1), the Joe copula is defined as [16]:

$$C_\theta(u_2, u_4) = 1 - \left[(1-u_2)^\theta + (1-u_4)^\theta - (1-u_2)^\theta(1-u_4)^\theta\right]^{1/\theta}, \quad 1 \le \theta < \infty,$$
$$h_{2|4}(u_2 \mid u_4) = S^{1/\theta-1}(1-u_4)^{\theta-1}\left[1 - (1-u_2)^\theta\right],$$
$$\text{with } S = (1-u_2)^\theta + (1-u_4)^\theta - (1-u_2)^\theta(1-u_4)^\theta, \quad u_2, u_4 \in [0, 1].$$

For two random variables U_3 and U_4 that are distributed uniform(0,1), the student-t copula with parameters $\rho \in (-1, 1)$ and degrees of freedom $\nu > 0$ is defined as [16]:

$$C_{\rho,\nu}(u_3, u_4) = t_{\rho,\nu}\left(t_\nu^{-1}(u_3),\, t_\nu^{-1}(u_4)\right),$$
$$h_{4|3}(u_4 \mid u_3) = t_{\nu+1}\left(\frac{t_\nu^{-1}(u_4) - \rho\, t_\nu^{-1}(u_3)}{\sqrt{\frac{\nu + \left(t_\nu^{-1}(u_3)\right)^2}{\nu+1}(1-\rho^2)}}\right), \quad u_3, u_4 \in [0, 1].$$

where $t_{\rho,\nu}$ denotes the joint cumulative distribution function of the bivariate student-t with correlation ρ and degrees of freedom ν, while t_ν^{-1} is the quantile function of the univariate student-t.

2.3.3. D-vine Copula

The vine structure, copula families, and parameters are selected using the `RVineStructureSelect` function from the VineCopula package in R. This procedure performs a data-driven selection within the class of regular vine (R-vine) copulas, which includes D-vine and C-vine as special cases [17]. The candidate copula families are restricted to a set of commonly used bivariate copulas, including independence, Gaussian, student-t, Clayton, Gumbel, Frank, and Joe copulas.

These families are chosen to capture various types of dependence structures, including symmetric and tail dependence.

For each pair of variables in the vine structure, the optimal copula family and its parameters are selected based on the AIC, where the model with the smallest AIC value is preferred. Based on this data-driven selection procedure, the resulting optimal vine structure corresponds to a D-vine.

The construction of the D-vine structure begins with the measurement of pairwise dependence between random variables. In this study, dependence is quantified using Kendall's tau (τ) coefficient. Kendall's tau is used to measure the strength and direction of dependence between variable pairs and to serve as the basis for determining the copula-based multivariate dependence model.

The D-vine structure is adopted due to its sequential representation of dependence. A D-vine of 4-dimensions consists of 3 linked trees, where the first tree (Tree 1) contains nodes corresponding to the random variables $(X_1, ..., X_4)$ and edges representing bivariate copulas between adjacent variables. The second tree (Tree 2) models conditional dependencies between variables separated by one intermediate variable, with edges corresponding to pair-copulas conditioned on that variable. The third tree (Tree 3) captures higher-order conditional dependence between the remaining pair of variables conditioned on two intermediate variables. Through this recursive construction, the joint density can be expressed as a product of marginal densities and all pair-copula densities specified in the vine decomposition [18].

Suppose that $X_1, \ldots, X_4$ are random variables whose dependency structure will be analyzed. If the absolute values of Kendall's tau correlation coefficients between sequential random variables satisfy the relationship:

$$|\tau(X_1, X_2)| > |\tau(X_2, X_3)| > |\tau(X_3, X_4)| \,.$$

Therefore, the determination of edges or variable pairs in the D-vine copula is based on the order of dependence strength, with variable pairs that have the highest dependence value placed in the first layer, while pairs with weaker dependence are placed in the following layers, with the following structure:

$$\begin{array}{ll} (X_1, X_2),\ (X_2, X_3),\ (X_3, X_4) & \text{Tree 1} \\ (X_1, X_3 \mid X_2),\ (X_2, X_4 \mid X_3) & \text{Tree 2} \\ (X_1, X_4 \mid X_2, X_3) & \text{Tree 3} \end{array}$$

2.3.4. D-vine Copula Simulation

The D-vine copula simulation algorithm described in the following generates a single set of synthetic observations that preserves the dependence structure among income components. To examine the stability and convergence of the Monte Carlo simulation under the D-vine copula framework, this data generation procedure is replicated multiple times.

Data generation is performed using a D-vine copula structure to integrate the income components while preserving their dependence structure. Based on the dependence ordering obtained from Kendall's tau and the selected D-vine copula structure in Section 3.2 and Fig. 3, the sequence of conditional simulation is given by $U_3 \to U_1 \mid U_3 \to U_4 \mid U_3 \to U_2 \mid U_4$. Accordingly, the uniform variables are generated sequentially following this order. The detailed simulation procedure is presented in Algorithm 1.

2.4. Estimation of Value at Risk and Expected Shortfall

Risk measurement in this study is conducted using Value-at-Risk (VaR) and Expected Shortfall (ES) as quantitative indicators of downside income risk. VaR represents the maximum potential loss at a specified confidence level, while ES measures the expected loss given that the loss exceeds the VaR threshold [19, 20]. These two measures are widely used in actuarial science

Algorithm 1 D-vine copula simulation

Require: Number of simulations n, copula parameters
Ensure: Simulated income $y^{(i)}$

1: **for** $i = 1$ to n **do**

2: Generate initial variable $u_3^{(i)}$ from $U_3^{(i)} \sim \text{uniform}(0,1)$.

3: Generate $r_1^{(i)}$ from $R_1^{(i)} \sim \text{uniform}(0,1)$, then obtain the conditional variable $u_1^{(i)}$ given by $U_3^{(i)} = u_3^{(i)}$ as:

$$u_1^{(i)} = h_{1|3}^{-1}\left(r_1^{(i)} \mid u_3^{(i)}; \theta_{13}\right).$$

4: Generate $r_4^{(i)}$ from $R_4^{(i)} \sim \text{uniform}(0,1)$, then obtain the conditional variable $u_4^{(i)}$ given by $U_3^{(i)} = u_3^{(i)}$ as:

$$u_4^{(i)} = h_{4|3}^{-1}\left(r_4^{(i)} \mid u_3^{(i)}; \rho, \nu\right).$$

5: Generate $r_2^{(i)}$ from $R_2^{(i)} \sim \text{uniform}(0,1)$, then obtain the conditional variable $u_2^{(i)}$ given by $U_4^{(i)} = u_4^{(i)}$ as:

$$u_2^{(i)} = h_{2|4}^{-1}\left(r_2^{(i)} \mid u_4^{(i)}; \theta_{24}\right).$$

6: Return to marginal scale:

$$x_j^{(i)} = F_j^{-1}\left(u_j^{(i)}\right), \quad j = 1, 2, 3, 4.$$

7: Determine the value of $c_1^{(i)}$ based on $x_3^{(i)}$ as:

$$c_1^{(i)} = \begin{cases} c_{1,1}, & \text{if } x_3^{(i)} < 34{,}534{,}000, \\ c_{1,2}, & \text{if } x_3^{(i)} \geq 34{,}534{,}000. \end{cases}$$

8: Compute income:

$$y^{(i)} = (x_1^{(i)} \times x_2^{(i)}) - (x_3^{(i)} + x_4^{(i)} + c_1^{(i)}).$$

9: **end for**

10: **return** $\{y^{(i)}\}_{i=1}^{n}$

and risk management because they capture both threshold risk and tail severity. The simulated income values $y^{(i)}$ in Algorithm 1 are subsequently transformed into losses $l^{(i)} = -y^{(i)}$. All VaR and ES calculations are therefore based on the distribution of $l^{(i)}$, where the focus is placed on the right tail of the loss distribution.

2.4.1. VaR and ES under the Univariate Approach

Under the univariate framework, L is modeled as follows a normal distribution. Suppose that L follows a normal distribution with mean μ as location parameter and standard deviation $\sigma > 0$ as scale parameter, the VaR at level α can be calculated by formula:

$$\text{VaR}_\alpha(L) = \mu + \sigma z_\alpha, \tag{2}$$

where $z_\alpha = \Phi^{-1}(\alpha)$ is the α-quantile of the standard normal distribution [14].

The Expected Shortfall at level α is defined as the expected loss given that the loss exceeds the VaR threshold. For the normal distribution, ES admits a closed-form expression [14]:

$$\text{ES}_\alpha(L) = \mu + \sigma \frac{\phi(z_\alpha)}{1-\alpha}, \tag{3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the PDF and CDF of the standard normal distribution, respectively.

2.4.2. VaR and ES under the Multivariate and Empirical Approach

In the multivariate framework,the distribution is generated through simulation, whereas the empirical approach is based on historical data. Therefore, the estimation of VaR and ES is performed numerically using the ordered values of simulated. Within this framework, multivariate farm income is defined as Eq. (1). Suppose $l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(n)}$ are ordered simulation result of a random variable that has been generated. The percentile used is α, then $m = \lfloor \alpha n \rfloor + 1$, where $\lfloor . \rfloor$ is the floor rounding function, then the VaR estimate for the simulation results is [14]:

$$\widehat{VaR}_\alpha(L) = \ell_{(m)}. \tag{4}$$

Accordingly, the Expected Shortfall is computed as the average of the lowest m simulated values [14]:

$$\widehat{ES}_\alpha(L) = \frac{1}{n-m+1} \sum_{j=m}^{n} \ell_{(j)}. \tag{5}$$

3. Results and Discussion

This section presents the results of the risk analysis under univariate and multivariate approaches. The univariate results are first discussed as a baseline, followed by the multivariate results that account for the dependence among the income components using the D-vine copula. A comparison between these approaches and the empirical estimates is then provided to highlight the differences in VaR and ES and their implications for risk assessment.

3.1. Result under the Univariate Approach

The parametric approach assumes that L follows a normal distribution. Fig. 1 presents the histogram of the empirical losses overlaid with the fitted normal density curve.

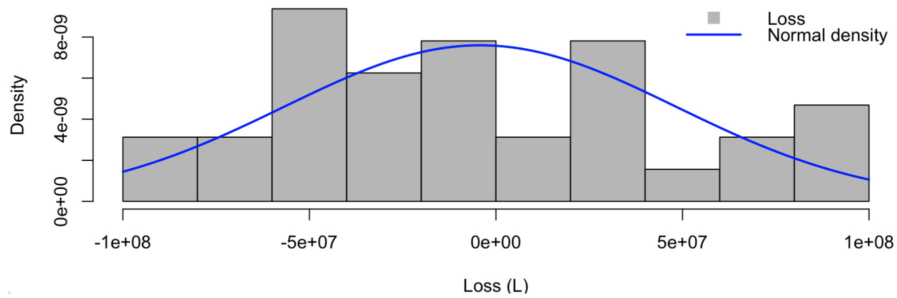

Fig. 1: Histogram of empirical losses with normal distribution

Table 1: Parameter estimation values and goodness-of-fit for losses

Variable	Distribution	Parameter	p-value	AIC
L	Normal	$\widehat{\mu} = -4237254; \widehat{\sigma} = 52509408$	0.268	1232.508

The Shapiro–Wilk test (p-value = 0.2129) indicates that the assumption of normality for L cannot be rejected at the 5% significance level. The estimated parameters are shown in Table 1, consisting of the location parameter $\hat{\mu} = -4{,}237{,}254$, which represents the central tendency of the loss distribution. Since the analysis is conducted on the transformed variable $L = -Y$, a negative mean indicates that, on average, the income remains positive. In other words, the typical outcome corresponds to a gain rather than a loss. The scale parameter $\hat{\sigma} = 52{,}509{,}408$ reflects the variability or dispersion of the loss distribution. The relatively large value of σ suggests substantial variability in income, indicating that farmers may experience significant fluctuations between favorable and unfavorable conditions.

Next, the VaR and ES risk values were estimated using equations Eq. (2) and Eq. (3) using the parameters presented in Table 1. The VaR and ES estimates from the univariate parametric approach are presented in Table 2.

Table 2: VaR and ES results for the univariate approach

Confidence level	VaR (10^3IDR/ha)	ES (10^3IDR/ha)
90%	63056	87915
95%	82133	104074
99%	117917	135711

The VaR estimates indicate that, at the 95% confidence level, the maximum potential loss is 82,133 (10^3IDR/ha), meaning that in most cases, losses are unlikely to exceed this threshold. However, the corresponding ES value of 104,074 (10^3IDR/ha) suggests that if losses exceed this level, the average loss could be substantially higher. This highlights significant tail risk in shallot farming income.

3.2. Result under the Multivariate Approach

The multivariate approach, shallot farm income is determined by several random variables, including crop yield (X_1), selling price (X_2), labor cost (X_3), and seed cost (X_4). Each variable is first modeled individually using an appropriate probability distribution selected based on goodness-of-fit evaluation. The results indicate that the marginal distributions of the income components follow the generalized logistic, log-logistic, and Weibull distributions, according to their respective data characteristics and statistical testing criteria. Fig. 2 presents the histograms of the four income components overlaid with their fitted theoretical distributions.

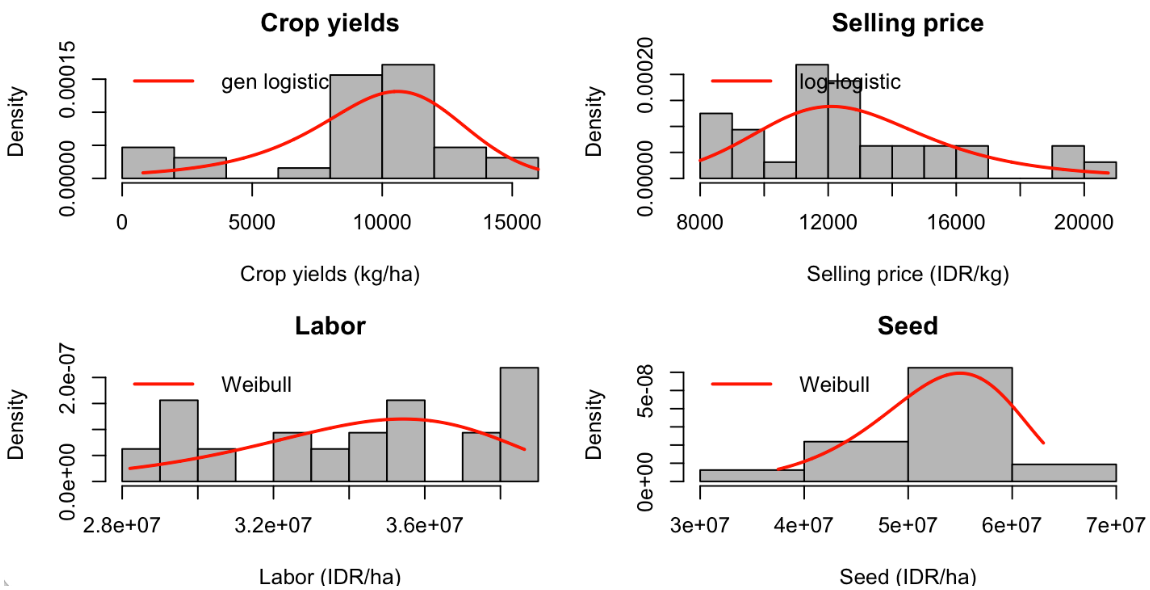

Fig. 2: Histograms of income components with their fitted distributions

The visual agreement between empirical histograms and the corresponding fitted distributions suggests that the selected marginal models adequately capture the variability and distributional patterns of each variable. The parameters for all marginal distribution were estimated using the MLE method and are reported in Table 3, along with the corresponding p-values from the AD goodness-of-fit tests and the AIC values, confirming the suitability of the selected models.

The parameter estimates indicate distinct characteristics for each variable. For X_1 (crop yield), the location parameter $\hat{\mu} = 1255.51$ reflects the average yield level, while the large scale parameter $\hat{\sigma} = 11846.14$ indicates substantial variability among farmers. The positive shape parameter $\hat{k} = 0.399$ suggests a right-skewed distribution.

Table 3: Parameter estimation values and goodness-of-fit for each variable

Variable	Distribution	Parameter	p-value	AIC
X_1	Gen. logistic	$\widehat{k} = 0.399; \widehat{\mu} = 1255.510; \widehat{\sigma} = 11846.140$	0.677	619.599
X_2	Log-logistic	$\widehat{\alpha} = 6.844; \widehat{\beta} = 12628.173$	0.825	611.715
X_3	Weibull	$\widehat{\alpha} = 11.624; \widehat{\beta} = 35712600$	0.419	1058.100
X_4	Weibull	$\widehat{\alpha} = 8.957; \widehat{\beta} = 55713205.767$	0.225	1102.386

For X_2 (selling price), the scale parameter $\hat{\beta} = 12628.17$ represents the median price level, and the shape parameter $\hat{\alpha} = 6.844$ indicates that prices are relatively concentrated around this value, with limited extreme fluctuations.

For X_3 (labor cost) and X_4 (seed cost), both modeled by Weibull distributions, the scale parameters ($\hat{\beta} = 35{,}712{,}600$ and $\hat{\beta} = 55{,}713{,}205$, respectively) reflect the magnitude of accumulated costs per season. The relatively large shape parameters ($\hat{\alpha} = 11.624$ and $\hat{\alpha} = 8.957$) indicate that these costs are fairly concentrated, suggesting similar cost structures among farmers.

After modeling the marginal distributions of the random variables that form income, a preliminary dependence analysis is conducted using several bivariate copula families. This step aims to identify the copula model that best captures the dependence structure between variables. The comparison is based on AIC, with lower values indicating a better fit. The results of this comparison are presented in Table 4.

Table 4: Selection of the best copula model based on AIC

Copula family	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	D-vine
AIC	-11.286	-27.431	-4.683	-8.329	-12.070	-4.683	-36.360

Based on the AIC values, the D-vine copula provides the best fit among the considered models, as indicated by its lowest AIC. Therefore, the dependence structure is further modeled using the D-vine copula approach. After each random variable that forms income is modeled with its distribution, the next step is to model the dependency structure among the random variables to create a D-vine copula. The dependency analysis aims to provide an overview of the strength of the dependency between variables, as measured by Kendall's tau (τ). The variable pairs with the highest τ values are ranked first in order to construct the D-vine copula structure. The results of the Kendall's tau correlation matrix are presented in Table 5.

Table 5: Kendall's tau value between variables

Variable	X_1	X_2	X_3	X_4
X_1	1	-0.071	0.329	0.256
X_2		1	0.087	0.147
X_3			1	0.691
X_4				1

From the information in Table 5, the order of pairs is obtained with $|\tau(X_3, X_4)| > |\tau(X_1, X_3)| > |\tau(X_1, X_4)| > |\tau(X_2, X_4)| > |\tau(X_2, X_3)| > |\tau(X_1, X_2)|$.

The 4-dimensional D-vine copula structure consists of 3 layers of trees. In the first tree (T_1), pairs with the strongest dependence are placed side by side. The pair with the highest dependence is (X_3, X_4), followed by (X_1, X_3), forming the partial sequence $X_1 - X_3 - X_4$. The remaining variable is X_2 is then connected to the variable with the strongest dependence among the endpoints, which is X_4 resulting in the final order $X_1 - X_3 - X_4 - X_2$ that represents the direct dependence between variables.

The second tree (T_2) is formed from conditional variable pairs with one intermediate variable, $(X_1, X_4 \mid X_3)$ and $(X_3, X_2 \mid X_4)$. Furthermore, in the third tree (T_3), second-level conditional variable pairs $(X_1, X_2 \mid X_3, X_4)$ are formed , which represent dependencies after conditioning on two intermediate variables. The D-vine copula structure formed hierarchically from the first to the third tree is shown in the Fig. 3.

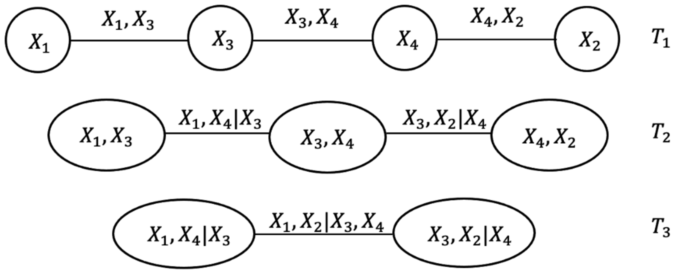

Fig. 3: Illustration of the D-vine copula structure

After the D-vine copula structure is determined, each pair in each tree is modeled using the appropriate bivariate copula. The copula-fitting process is carried out by transforming the data to a uniform scale using the initial marginal distribution from Table 3. The selection of the best copula family for each pair is based on model fit criteria. The copula fitting results and estimated parameters are presented in Table 6.

Table 6: Pairs and parameter values for each copula function

Edge	Copula family	Estimated parameter values
X_1, X_3	Clayton	$\widehat{\theta}_{13} = 0.59$
X_3, X_4	Student-t	$\widehat{\rho} = 0.81; \widehat{\nu} = 2.16$
X_4, X_2	Joe	$\widehat{\theta}_{24} = 1.30$
$X_{1,4\|3}$	Independence	–
$X_{3,2\|4}$	Independence	–
$X_{1,2\|3,4}$	Independence	–

Each pair in tree 2 (T_2) and tree 3 (T_3) is independent, so the density of the copula is 1, therefore:

$$c_{1,4|3} = c_{3,2|4} = c_{1,2|3,4} = 1.$$

To further investigate the dependence structure, two model specifications were compared: one allowing the independence copula (family set 0:6) and one excluding it. When the independence copula is not included, higher-order trees (Tree 2 and Tree 3) still exhibit weak dependencies. However, when independence is allowed, these conditional pairs are selected as independent.

Importantly, the model including the independence copula yields a lower AIC value (AIC $= -36.36$) compared to the model without independence (AIC $= -29.92$). This indicates that the simpler model, in which conditional dependencies vanish in higher trees, provides a better balance between goodness-of-fit and model complexity.

These findings suggest that the remaining dependencies in Tree 2 and Tree 3 are not statistically strong and can be adequately approximated as independent. Consequently, the main dependence structure is effectively captured in the first tree, while higher-order interactions are negligible.

Then the equation of the joint density function (X_1, X_3, X_4, X_2) can be written as:

$$f(x_1, x_3, x_4, x_2) = c_{1,3}^{\text{Clayton}}(F_1(x_1), F_3(x_3)) \times c_{3,4}^{t}(F_3(x_3), F_4(x_4)) \times c_{4,2}^{\text{Joe}}(F_4(x_4), F_2(x_2)) \prod_{i=1}^{4} f_i(x_i).$$

The next step is to perform a Monte Carlo simulation of the selected D-vine structure. The simulation is carried out by generating $n = 10^6$ data points in each iteration to ensure accurate estimation of tail risk measures. To assess the stability of the results, a fixed and moderate number of Monte Carlo replications is performed, and the resulting VaR and ES estimates are averaged across replications. The algorithm for this procedure is described in Algorithm 1. The simulation results are presented in Table 7 and Fig. 4.

Table Table 7 compares the empirical income with the simulated income generated by the D-vine copula model. It should be noted that VaR and ES are computed based on the loss variable $L = -Y$, while the descriptive statistics are presented in terms of income Y for interpretable. The results indicate that the simulated data closely align with the empirical data for mean and standard deviation, suggesting that the model effectively captures the central tendency and variability of income. However, the simulated income spans a broader range, with more pronounced minimum and maximum values, suggesting greater ability to capture extreme outcomes. Additionally, the higher kurtosis observed in the simulated data reflects heavier tails, which is especially pertinent for risk analysis.

Table 7: Simulation and empirical result

Descriptive statistics	Empirical income (10^3IDR/ha)	Simulation with D-vine (10^3IDR/ha)
Min	−92134	−146754
Mean	4237	5002
SD	53349	53019
Varians	2846.181e+9	2811.015e+9
Max	98634	1470684
Skewness	−0.305	0.714
Kurtosis	2.230	3.838

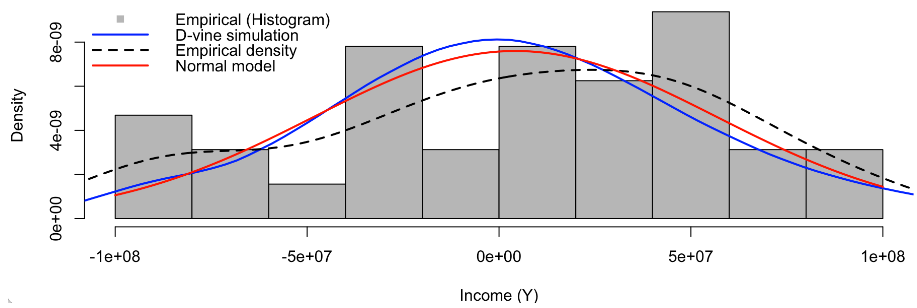

Fig. 4: Comparison of empirical income, D-vine copula simulation, and normal model

A discrepancy is observed in the skewness, the empirical distribution is slightly negatively skewed, whereas the simulated distribution is positively skewed. This indicates that the D-vine copula model does not fully capture the asymmetry of the income distribution, particularly in its representation of the balance between lower and higher income extremes. This difference may be influenced by the limited sample size and the complexity of the dependence structure.

To examine the stability and convergence of the Monte Carlo simulation under the D-vine copula, VaR and ES at the 90%, 95%, and 99% confidence levels are computed in a stepwise manner. The number of simulated observations per iteration (n) is increased progressively, taking value $10^2, 10^3, 10^4, 10^5$, and 10^6. For smaller sample size $n = 10^2, ..., 10^4$, the number of Monte Carlo replications is set to $R = n$. For larger sample sizes $n = 10^5$ and 10^6, a fixed number of replications $R = 10^3$ is used to ensure computational efficiency while maintaining stable estimates.

This stepwise evaluation allows us to assess whether the simulated risk estimates converge toward stable values as n increases. Fig. 5 and Fig. 6 present the convergence behavior of VaR and ES obtained from the simulation procedure.

The convergence analysis confirms that the Monte Carlo estimates stabilize as the number of simulations increases, with negligible fluctuations observed for $n > 10^4$. Therefore, $n = 10^6$ replications are adopted in the final estimation to ensure numerical stability. Table 8 reports the final multivariate risk estimates, obtained from Monte Carlo simulations with $n = 10^6$ and a

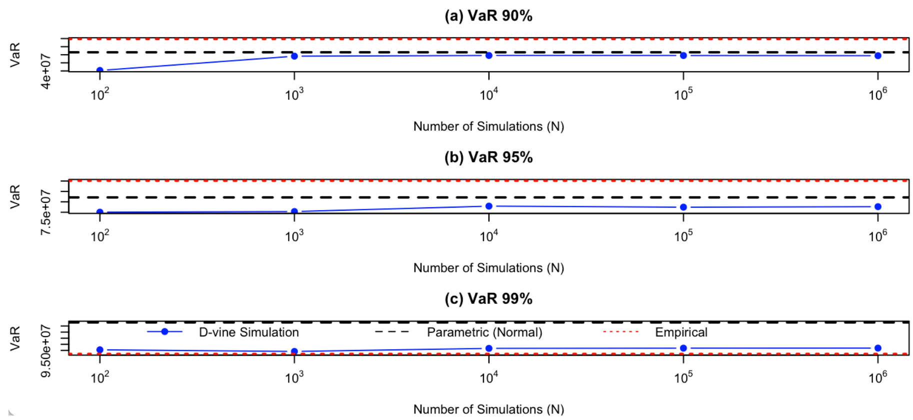

Fig. 5: VaR convergence for D-vine simulation

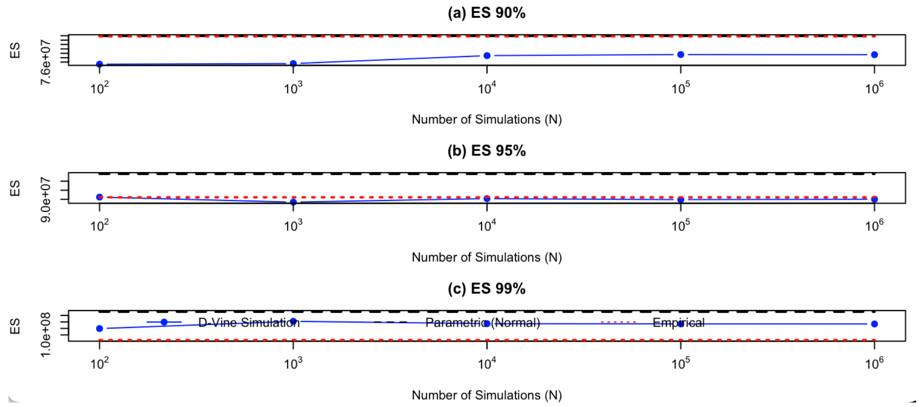

Fig. 6: ES convergence for D-vine simulation

fixed number of replications to ensure the stability of the results. The reported VaR and ES values are computed as the average across replications and represent the multivariate downside income risk under the D-vine copula.

Table 8: VaR and ES results for the multivariate approach

Confidence level	VaR (10^3IDR/ha)	ES (10^3IDR/ha)
90%	58852	79381
95%	77686	90062
99%	96877	115534

The multivariate D-vine copula simulation results, at the 95% confidence level, the VaR is estimated at 77,686 (10^3IDR/ha), indicating that under typical conditions, potential losses are unlikely to exceed this threshold. However, the corresponding ES value of 90,062 (10^3IDR/ha) shows that when losses do exceed the VaR level, the average loss remains substantially higher, reflecting the presence of tail risk.

3.3. Comparison between Univariate and Multivariate Approach

The empirical VaR and ES are computed directly from the historical data using Eq. (4) and Eq. (5), shown in Table 9, without imposing any distributional assumptions. These measures reflect the observed downside risk based solely on the available sample.

From an economic perspective, at the 95% confidence level, the VaR is estimated at 90,134 (10^3IDR/ha), indicating that under most observed conditions, potential losses are unlikely to exceed this threshold. The corresponding ES value of 91,134 (10^3IDR / ha) suggests that, conditional on losses exceeding the VaR level, the average loss remains at a similar high magnitude.

Table 9: VaR and ES results for each approach

Risk estimation	Confidence level	Empirical	Normal	D-vine copula
VaR (10^3IDR/ha)	90%	79465	63056	58852
	95%	90134	82133	77686
	99%	92134	117917	96877
ES (10^3IDR/ha)	90%	87700	87915	79381
	95%	91134	104074	90062
	99%	92134	135711	115534

The empirical approach produces relatively coarse estimates due to limited data, as reflected by the identical VaR and ES values at the confidence level 99%, suggesting limited sensitivity to extreme losses and that the behavior of the tail is not well captured. As a result, this approach is more appropriate for preliminary analysis or for decision-makers with a risk-averse perspective who prefer simpler and less variable estimates, thus motivating the use of parametric and copula-based approaches to obtain smoother and more stable estimates of tail risk.

The normal distribution approach yields the highest VaR and ES values across all confidence levels, reflecting a more conservative assessment of extreme risk. Its smooth parametric form produces a well-defined and continuous tail, resulting in more stable estimates compared to the empirical approach. However, this smoothness is driven by distributional assumptions rather than the underlying data structure, as the model does not account for the dependence among income components. As a result, the estimated tail risk may be overstated or not fully aligned with the actual joint behavior of the variables. Consequently, this approach is more suitable as a conservative benchmark or for decision-makers with a risk-seeking perspective in the sense of accommodating more extreme potential losses.

Conversely, the D-vine copula approach provides moderate risk estimates between the empirical and normal approaches. It captures joint dynamics and tail dependence through simulation, offering smoother tail behavior relevant to correlated income components that can amplify losses. This is particularly relevant for income components, where high input costs may occur simultaneously and amplify total production expenses; when combined with low production or unfavorable prices, this interaction can lead to extreme loss scenarios. However, the resulting risk estimates remain conditional on the specified copula structure and simulation design. Therefore, this approach offers a balanced representation of risk and may be appropriate for decision-makers with a risk-neutral perspective.

Overall, each approach offers distinct advantages and limitations, and the choice of method should be aligned with the decision-maker's risk attitude and the level of model complexity required.

4. Conclusion

This study examines the measurement of shallot farming income risk in Babakan and Pabedilan villages in Cirebon Regency by considering the stochastic nature of income components and their interdependencies. A parametric approach based on the normal distribution and a D-vine copula model was employed, where the D-vine uses Monte Carlo simulation to preserve both the marginal distributions and the dependence structure among variables. The resulting estimates were then compared with those obtained from the empirical approach.

The approaches differ in how they represent income risk, particularly in terms of tail behavior and the treatment of dependence among income components. These differences lead to distinct implications for risk assessment, where the empirical approach may be more suitable for decision-makers with a risk-averse perspective, the normal distribution approach for those willing to risk-seeking perspective, and the D-vine copula approach for decision-makers with a risk-neutral perspective.

This study is subject to several limitations, including the relatively limited sample size and the assumption of specific copula structures. In addition, as the analysis focuses on Babakan and Pabedilan villages, further empirical validation is needed before applying these findings to other regions. Future research may explore alternative copula families and extend the analysis to other commodities and regions.

CRediT Authorship Contribution Statement

Fatimah Fuzzaroh: Conceptualization, Methodology, Formal Analysis, Writing original draft.
Berlian Setiawaty: Supervision, Validation, Writing–Review & Editing.
I Gusti Putu Purnaba: Supervision, Validation, Writing–Review.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (specifically Grammarly) were utilised exclusively for grammar editing during the preparation of this work. The analysis, interpretation, and core content were not generated by AI. All substantive result and discussions were entirely developed by the authors themselves.

Declaration of Competing Interest

The authors declare no competing interests.

Funding and Acknowledgments

This research received funding from the School of IPB University for publishing the article.

Data and Code Availability

The dataset was not publicly available.

References

[1] Ary Bakhtiar, Mulia Salsabila, and Mohd Fauzi bin Kamarudin. "Prospects for Shallots Agribusiness Development in Pamekasan Regency". In: *Jurnal AGRISEP: Kajian Masalah Sosial Ekonomi Pertanian dan Agribisnis* (Mar. 2025), pp. 451–464. DOI: 10.31186/jagrisep.24.01.451-464.

[2] Badan Pusat Statistik. *Komoditas Pertanian Subsektor Hortikultura Bawang Merah*. Tech. rep. Pusat Data dan Sistem Informasi Pertanian Kementerian Pertanian, 2023. https://satudata.pertanian.go.id/assets/docs/publikasi/Outlook_Bawang_Merah_2023.pdf.

[3] Aulia Adetya, Dwi Rachmina, and Lukman Mohammad Baga. "The Influence of Farmer Entrepreneurial Behavior on Shallot Farming Performance". In: *Jurnal Agrisep: Kajian Masalah Sosial Ekonomi Pertanian dan Agribisnis* 24.2 (2025), pp. 579–600. DOI: 10.31186/jagrisep.24.02.579-600.

[4] Yogi Ricardo, Camelya Francisca, Ela Amelia, and Dina Dwirayani. "Analysis of Shallot Farming Income in Pabedilan District, Cirebon Regency". In: *Journal of Agricultural Sciences (Agrosci)* 2.6 (2025), pp. 372–382. DOI: 10.62885/agrosci.v2i6.816.

[5] Dewi Sahara et al. "Sustainability of Shallot Farming System in Lowland Central Java Province, Indonesia: MICMAC Analysis Approach". In: *Environmental Challenges* 20 (2025), p. 101212. DOI: 10.1016/j.envc.2025.101212.

[6] Sriyadi, Zuhud Rozaki, and Wiwi Susanti. "Risk Behavior of Shallot Farmers in Highland and Lowland Regions of Java, Indonesia". In: *Organic Farming* 11.1 (2025), pp. 49–60. DOI: 10.56578/of110104.

[7] Triyono, Lusi Soviyanti, Nur Rahmawati, Sutrisno, and Mhd Fauzi Kamarudin. "Income Risk Analysis and Sustainability of Shallot Farming in Bantul Yogyakarta, Indonesia". In: *BIO Web of Conferences (ICAFS 2025)*. Vol. 208. EDP Sciences, 2026, p. 01005. DOI: 10.1051/bioconf/202620801005.

[8] Yueming Cheng. "Monte Carlo-Based VaR Estimation and Backtesting Under Basel III". In: *Risks* 13 (2025), p. 146. DOI: 10.3390/risks13080146.

[9] N. Palupi, A. Listyaningrum, and G. B. Nugraha. *UGM and PT Pupuk Indonesia Advance Agricultural Efficiency through Precision Farming Technology*. Published by Universitas Gadjah Mada; Accessed: 2026-03-29. Nov. 2025. https://ugm.ac.id/en/news/ugm-and-pt-pupuk-indonesia-advance-agricultural-efficiency-through-precision-farming-technology/.

[10] Yan Sun and Ke Wang. "The Implication of Copula-Based Models for Crop Insurance and Reinsurance Under Systemic Risk". In: *Frontiers in Environmental Science* 10 (2022), p. 916494. DOI: 10.3389/fenvs.2022.916494.

[11] Marwah Soliman, Nathaniel K. Newlands, Vyacheslav Lyubchich, and Yulia R. Gel. "Multivariate Copula Modeling for Improving Agricultural Risk Assessment under Climate Variability". In: *Variance* 16.1 (2023), pp. 1–20. DOI: 10.66573/001c.74221.

[12] Resminawati, Retno Budiarti, and I Wayan Mangku. "Pembandingan Metode Copula dan D-vine Copula dalam Estimasi Tail Value at Risk pada Komoditas Pertanian". Thesis. Institut Pertanian Bogor, 2021. http://repository.ipb.ac.id/handle/123456789/107803.

[13] Xianli Wang, Zhigang Zhao, Feilong Jie, Jingjing Xu, Sheng Li, Kun Hao, and Youliang Peng. "A Copula Function–Monte Carlo Method-Based Assessment of the Risk of Agricultural Water Demand in Xinjiang, China". In: *Agriculture* 14.11 (2024), p. 2000. DOI: 10.3390/agriculture14112000.

[14] Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. *Loss Models: From Data to Decisions*. 4th ed. New York: Wiley, 2012. DOI: 10.1002/9781118787106.

[15] Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume 2*. 2nd ed. Wiley Series in Probability and Statistics. New York: Wiley, 1995. https://www.wiley.com/en-us/Continuous+Univariate+Distributions%2C+Volume+2%2C+2nd+Edition-p-9780471584940.

[16] Roger B. Nelsen. *An Introduction to Copulas*. 2nd ed. Springer, 2006. DOI: 10.1007/0-387-28678-0.

[17] Eike Christian Brechmann and Ulf Schepsmeier. "Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine". In: *Journal of Statistical Software* 52.3 (2013), pp. 1–27. DOI: 10.18637/jss.v052.i03.

[18] Claudia Czado and Thomas Nagler. "Vine Copula Based Modeling". In: *Annual Review of Statistics and Its Application* 9 (2022), pp. 453–477. DOI: 10.1146/annurev-statistics-040220-101153.

[19] Emese Lazar, Jingqi Pan, and Shixuan Wang. "On the Estimation of Value-at-Risk and Expected Shortfall at Extreme Levels". In: *Journal of Commodity Markets* 34 (2024), p. 100391. DOI: 10.1016/j.jcomm.2024.100391.

[20] Turnika Afdatul Rafni and Dina Agustina. "Risk Comparison in Optimal Portfolios: A Study of Value at Risk (VaR) and Tail Value at Risk (TVaR)". In: *Mathematical Journal of Modelling and Forecasting* 3.1 (June 2025), pp. 47–55. DOI: 10.24036/mjmf.v3i1.40.