IAJ
Indonesian Actuarial Journal
Persatuan Aktuaris Indonesia (PAI)

Vol. 2, No. 1, July (2026)
e-ISSN 3110-6463
https://10.65689/iajvol2no1pp096-111

Portfolio Optimization and Risk Estimation of LQ45 Mining Stock Using The ARIMA-GARCH-EVT-COPULA Hybrid Model as Alternative Pension Fund Investment

Dyah Ayu Mahanani [1*], Galuh Oktavia Siswono [1], Ulil Azmi [1]

[1]*Department of Actuarial, Faculty of Science and Data Analytics, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia*

*Corresponding email: *dayumahanani2866@gmail.com*

Abstract
Economic uncertainty and high commodity market volatility demand more accurate pension fund investment risk management. This study aims to measure risk and determine the optimal portfolio of LQ45 mining stocks as an alternative pension fund investment instrument using a hybrid GARCH-EVT-Copula approach. The data used are daily log returns of ITMG, ADRO, and PTBA stocks for the period June 1, 2020, to June 30, 2025. The ARIMA-GARCH model is used to capture volatility dynamics, while extreme risks are modeled using the EVT approach. Dependencies between stocks are analyzed using the best copula model and Monte Carlo simulations to produce VaR and ES estimates which are then validated using Backtesting. The results show that the best model for ITMG stocks is ARMA(0,0)-GARCH(1,1), for ADRO stocks it has an ARMA(1,1)-GARCH(1,1) model, and for PTBA stocks it has an ARMA(0,1)-GARCH(0,1) model. Based on the EVT estimation, ITMG and ADRO stock have positive GPD shape parameters, indicating a sfat-tailed distribution and a higher potential for extreme risk. In contrast, PTBA stock have negative shape parameters, indicating a limited tail distribution so that extreme risk is relatively more controlled. The scale parameter value also confirms that PTBA's extreme fluctuations are more stable than other stock. The portfolio dependency structure can be modeled using the best copula model, namely the Gumbel Copula, the optimization results give portfolio A a weight of 62.11456% for ITMG stock and 37.88544% for ADRO stock, portfolio B is 19.93126% for ITMG stock and 80.06874% for PTBA stock, while portfolio C is 8.668356% for ADRO stock and 91.33164% for PTBA stock. The results also show that the estimated VaR values obtained are valid at all confidence levels, and the optimal portfolio tends to place PTBA and ITMG stock as dominant stocks to reduce the total portfolio risk. Thus, the ARIMA-GARCH-EVT-Copula approach is proven to be effective in modeling extreme risks and dependency structures between stocks, and can provide a strong basis for making long-term investment decisions for pension fund management.

Article History:
Submited/Recevied: 16 February 2026
First Revised: 18 July 2026
Accepted: 30 July 2026
Published:31 July 2026

Keywords:
ARIMA-GARCH, Copula, Extreme Value Theory, Pension Fund, Risk.

1. INTRODUCTION

Pension fund institutions, as regulated under Law No. 11 of 1992 and POJK 3/POJK.05/2015, are entities that manage long-term savings and are permitted to allocate a portion of their funds into investment instruments such as stocks [1]. To preserve and increase pension fund value, investment decisions require not only fundamental analysis but also robust risk management, particularly in highly volatile sectors such as mining. The mining sector offers attractive return potential but is characterized by substantial price fluctuations driven by global commodity prices, geopolitical uncertainty, and government downstreaming policies. Consequently, constructing an optimal portfolio and accurately measuring downside risk are essential for supporting long-term pension fund investment decisions. Previous studies have widely applied hybrid ARIMA-GARCH models to capture volatility, Extreme Value Theory (EVT) to model tail risk, and copula models to represent dependence among asset returns. However, most studies examine these methods separately or focus on general

stock markets without specifically integrating them into a unified framework for mining stocks intended as pension fund investment alternatives. Furthermore, limited attention has been given to evaluating both portfolio optimization and extreme downside risk simultaneously using a hybrid ARIMA-GARCH-EVT-Copula approach. Therefore, this study aims to address this gap by integrating ARIMA-GARCH, EVT, and Copula models to optimize mining stock portfolios and estimate portfolio risk, providing a more comprehensive framework for long term pension fund investment.

To address these challenges, the ARIMA-GARCH approach is employed to handle autocorrelation and heteroskedasticity in log-return data, thereby producing i.i.d. residuals [2]. These residuals are required for extreme risk modeling using Extreme Value Theory (EVT) with the Peaks Over Threshold (POT) method, which provides more accurate results compared to the Block Maxima approach [3]. In addition, the dependence structure between stocks is modeled using Copula methods from the Archimedean family—namely Clayton, Gumbel, and Frank Copulas—which are capable of capturing dependence structures, particularly in the tail distribution [4].

This study aims to estimate extreme risk and construct an optimal portfolio of mining sector stocks using the ARIMA-GARCH-EVT-Copula approach. Risk measurement is conducted through Value at Risk (VaR) and Expected Shortfall (ES), whose accuracy is evaluated using backtesting procedures. This hybrid approach is able to produce more accurate estimation results, providing a sound basis for decision-making [5].

This research is expected to provide theoretical and statistical contributions to risk management in mining stock portfolios amid systemic challenges and risks, as well as offer insights into their potential as one of the investment instruments for pension funds.

2. METHODS

2.1 Return

Return is the compensation value expected by investors for the opportunity cost and the risk of a decline in purchasing power due to the effects of inflation [6]. Return calculation in this study uses the log return method because it has more stable and accurate performance, allowing risk analysis and long-term investment strategy formation to be more reliable [7]. Therefore, this research employs log return calculations as they perform better than simple returns. The log return method is particularly suitable when the data exhibit indications of extreme values, as it is able to provide more robust and reliable analytical results..

$$\boldsymbol{R_{i,t}^{ln} = \ln(P_t) - \ln(P_{t-1}) = \ln\left(\frac{P_t}{P_{t-1}}\right)} \tag{1}$$

This formula represents the ratio of the stock's closing price at time t to the stock's closing price in the previous period.

2.2 Time Series

Time series is a collection of data observed sequentially over time, indicating that a variable changes as time progresses. Time series data can be classified into discrete time series and continuous time series [8]. The main objective of time series analysis is to understand the historical patterns of the data, forecast future values, and optimize systems based on the existing level of risk, as visualized in **Figure 1**.

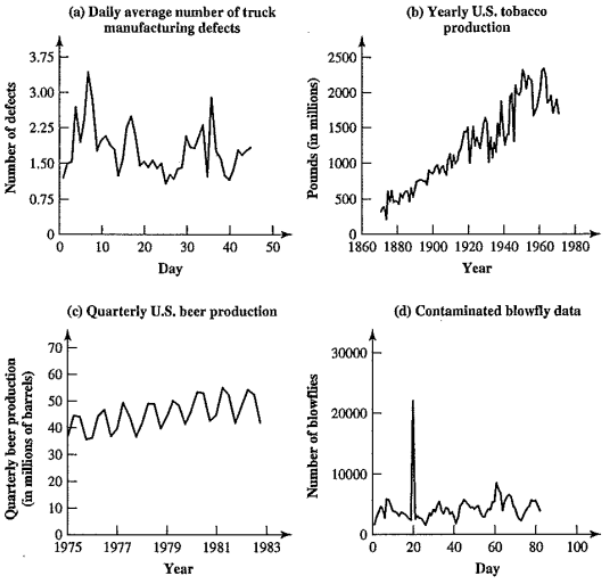


Figure 1. Time Series Plot
Source : Wei, 2016

The fundamental characteristic of time series data is that the observations are interdependent or correlated with one another; therefore, the assumption of independence no longer holds, and more advanced methods are required for proper analysis.

2.3 Stationarity Test

Stationarity testing is an essential preliminary test that must be conducted before ARIMA modeling, as it is used to determine the appropriate model to be applied [9]. If the data are already stationary, this condition will be reflected in the patterns of the ACF and PACF plots. A time series is said to be stationary in variance if it does not exhibit fluctuations over time and tends to be stable; such issues can be addressed using the Box–Cox transformation**.** If the data are not stationary in mean, differencing is required until stationarity is achieved. Stationarity in the mean can also be examined using the Augmented Dickey–Fuller (ADF) test, where the null hypothesis states that the data contain a unit root and are therefore non-stationary.

$$\boldsymbol{\tau_{count}} = \frac{\hat{\boldsymbol{\phi}}_{\boldsymbol{T}}-\mathbf{1}}{\boldsymbol{SE}(\hat{\boldsymbol{\phi}}_{\boldsymbol{T}})} \tag{2}$$

This formula represents the test statistic used to make decisions and draw conclusions from the results. If the value of $\tau_{count} \leq \tau_{n;0,05}$, the null hypothesis is rejected, indicating that the data are stationary. Where $\tau_{count}$is the calculated test statistic for the Augmented Dickey–Fuller (ADF) unit root test, $\hat{\phi}_{T-1}$ is the estimated coefficient of the lagged level variable obtained from the ADF regression, and $SE(\hat{\phi}_T)$ is the standard error of the estimated coefficient. The calculated τ_{count}statistic is then compared with the corresponding critical value to determine whether the null hypothesis of a unit root can be rejected.

2.4 Autoregressive Integrated Moving Average

Autoregressive Integrated Moving Average (ARIMA) can be used to describe stationary time series data using the ARMA (p, q) model, as well as non-stationary time series data using the ARIMA (p, d, q) model [12].

$$\dot{\boldsymbol{Z}}_{\boldsymbol{t}} = \boldsymbol{\phi_1}\dot{\boldsymbol{Z}}_{\boldsymbol{t-1}} + \boldsymbol{\phi_2}\dot{\boldsymbol{Z}}_{\boldsymbol{t-2}} + \cdots + \boldsymbol{\phi_p}\dot{\boldsymbol{Z}}_{\boldsymbol{t-p}} + \boldsymbol{a_t} - \boldsymbol{\theta_1}\boldsymbol{a_{t-1}} - \cdots - \boldsymbol{\theta_q}\boldsymbol{a_{t-q}} \tag{3}$$

Where $\dot{Z}_t$ denotes the stationary time series at time t, $\dot{Z}_{t-j}$ is the observation at lag $j(j = 1,2, \dots, p)$, ϕ_j is the autoregressive (AR) coefficient at lag j, a_t is the white noise error term at time $t$with mean zero and

constant variance, $\theta_k$is the moving average (MA) coefficient at lag $k(k = 1,2, \dots, q)$, a_{t-k} is the white noise error at lag k, p is the order of the autoregressive component, and $q$is the order of the moving average component. The Autoregressive (AR) model is based on the assumption that the current value is influenced by its past values and the information available from previous periods. This model is constructed by establishing a regression relationship among the time series values at the previous p lags. The AR model also determines the order $p$in ARIMA modeling. Equation (3) shows that the current observation is determined by a linear combination of its previous $p$observations and the current as well as past $q$random shocks.

$$\dot{\boldsymbol{Z}}_{\boldsymbol{t}} = \boldsymbol{\phi}_{\boldsymbol{1}}\dot{\boldsymbol{Z}}_{\boldsymbol{t-1}} + \boldsymbol{\phi}_{\boldsymbol{2}}\dot{\boldsymbol{Z}}_{\boldsymbol{t-2}} + \cdots + \boldsymbol{\phi}_{\boldsymbol{p}}\dot{\boldsymbol{Z}}_{\boldsymbol{t-p}} + \boldsymbol{a}_{\boldsymbol{t}} \tag{4}$$

Where $\dot{Z}_t$ denotes the stationary time series at time t, $\dot{Z}_{t-j}$ is the observation at lag j, ϕ_j is the autoregressive coefficient at lag j, a_t is the white noise error term, and $p$is the order of the autoregressive model. The Moving Average (MA) model assumes that the data contain not only information from past values but also the influence of random disturbances or errors. If a high shock occurs in a previous period, its effect may still be felt in the current period. In other words, this model is able to describe how past fluctuations can influence present values. Equation (4) represents the autoregressive (AR) model, in which the current observation depends only on its previous $p$observations and a random error term, without including a moving average component.

$$\dot{\boldsymbol{Z}}_{\boldsymbol{t}} = \boldsymbol{a}_{\boldsymbol{t}} - \boldsymbol{\theta}_{\boldsymbol{1}}\boldsymbol{a}_{\boldsymbol{t-1}} - \cdots - \boldsymbol{\theta}_{\boldsymbol{q}}\boldsymbol{a}_{\boldsymbol{t-q}} \tag{5}$$

Where $\dot{Z}_t$ denotes the stationary time series at time t, a_t is the white noise error term at time $t$with mean zero and constant variance, a_{t-q} is the white noise error at lag $q(q = 1,2, \dots, k)$, $\theta_k$is the moving average (MA) coefficient at lag k, and q is the order of the moving average model. The method used to estimate the parameters of the ARIMA model is Maximum Likelihood Estimation (MLE). Equation (5) represents the Moving Average (MA) model, in which the current observation is expressed as a linear combination of the current random shock and the random shocks from the previous $q$periods. Unlike the autoregressive model, the MA model explains the current value based on past forecast errors rather than past observations. This method is also applied to estimate the ARMA (p, q) components.

2.5 Generalized Autoregressive Conditional Heteroskedasticity

A statistical model used to analyze time series data with non-constant volatility and capable of capturing volatility clustering patterns in the residuals [10]. This model was developed to address higher-order problems, namely the GARCH model.

$$\boldsymbol{\sigma}_{\boldsymbol{re}}^{\boldsymbol{2}} = \boldsymbol{\omega} + \sum_{\boldsymbol{i=1}}^{\boldsymbol{s}} \boldsymbol{\theta}_{\boldsymbol{i}}\boldsymbol{\varepsilon}_{\boldsymbol{t-i}}^{\boldsymbol{2}} + \sum_{\boldsymbol{j=1}}^{\boldsymbol{r}} \boldsymbol{\phi}_{\boldsymbol{j}}\boldsymbol{\sigma}_{\boldsymbol{ret-j}}^{\boldsymbol{2}} \tag{6}$$

Where $\sigma_{re_t}^2$ denotes the conditional variance of the return residuals at time t, ω is a positive constant representing the long-run variance, ε_{t-i}^2 is the squared residual at lag i, θ_i is the ARCH coefficient associated with the lagged squared residuals, $\sigma_{re_{t-j}}^2$ is the conditional variance at lag j, ϕ_j is the GARCH coefficient associated with the lagged conditional variances, $s$is the order of the ARCH component, and $r$is the order of the GARCH component. After the best model has been identified, the next step is to estimate the parameters using the Maximum Likelihood Estimation (MLE) method.

2.6 Extreme Value Theory

The tail behavior of a distribution refers to data characteristics with extreme values that are not reflected by the mean or variance [11]. The Peak Over Threshold (POT) approach considers extreme values that exceed a specified threshold, assuming a Generalized Pareto Distribution (GPD). In determining the threshold, a

method that can provide a more accurate and optimal threshold is required, such as the Mean Residual Life Plot (MRLP) method.

$$\boldsymbol{F_x(x;\xi,u,\sigma)} = \begin{cases} \mathbf{1} - \left(\mathbf{1} - \frac{\xi(x-u)}{\sigma_e}\right)^{-\frac{1}{\xi}}, \xi \neq \mathbf{0},\ \sigma > \mathbf{0} \\ \mathbf{1} - \boldsymbol{exp}\left(-\frac{(x-u)}{\sigma_e}\right), \xi = \mathbf{0}, \sigma > \mathbf{0} \end{cases} \tag{7}$$

When the value of $\xi \leq 0$, the GPD has a long tail, implying that the variable $x$has no finite upper bound, i.e., $u \leq x < \infty$. Conversely, when $\xi > 0$, the GPD has a short (bounded) tail, so that $x$is restricted to the interval $u \leq x \leq c$, where $c = \sigma_\theta/\xi$denotes the finite upper endpoint of the distribution. This endpoint represents the maximum possible value that the exceedance variable can attain under the assumed GPD model; therefore, any observation exceeding $c$has zero probability of occurrence. In other words, the distribution has a finite support, indicating that the tail is bounded rather than extending indefinitely. Subsequently, the parameters of the assumed distribution are estimated by maximizing the log-likelihood function, which is obtained by taking the first derivatives with respect to the distribution parameters (σ_θ, ξ).

2.7 Probability Integral Transform

In copula modeling, each dataset is first transformed into a domain that follows a Uniform [0,1] distribution, since the copula function operates on the space of marginal probabilities that follow a Uniform [0,1] distribution [12]. The Probability Integral Transform is a mathematical transformation process that converts a given variable's distribution into a Uniform [0,1] distribution without altering the dependence structure among variables.

$$\boldsymbol{U = F_X(X)} \tag{8}$$

Thus, U follows the target (domain) distribution, namely the Uniform [0,1] distribution. Each observation X_i is transformed to follow this domain distribution according to its position in the original distribution.

2.8 Copula Archimedean

A non-parametric statistical method used to measure the direction and strength of the relationship between two variables. This test assesses dependence based on ranks and does not rely on the assumption of data normality. The resulting Kendall's Tau value is then used to determine the parameters of the Archimedean copula family [13]. If the two datasets are not independent, a copula model is required.

A copula is a statistical method used to link the marginal distributions of multiple variables in order to capture and understand the dependence among them. Copula models offer the advantage of flexibility in representing various forms of dependence, including tail dependence, which is particularly useful in risk analysis. The density function is defined as follows.

$$\boldsymbol{C(u,v) = \varphi_{ac}^{-1}\big(\varphi_{ac}(u) + \varphi_{ac}(v)\big)} \tag{9}$$

$$\boldsymbol{c(u,v)} = -\frac{\boldsymbol{\varphi'_{ac}(u)\varphi'_{ac}(v)\phi''_{ac}(C(u,v))}}{\left(\boldsymbol{\varphi'_{ac}(C(u,v))}\right)^2} \tag{10}$$

Where $C(u,v)$ denotes the Archimedean copula function, $u$and $v$are uniformly distributed random variables on the interval $[0, 1]$ obtained through the Probability Integral Transform (PIT), φ_{ac} is the generator function of the Archimedean copula, and φ_{ac}^{-1} is the inverse of the generator function. While $c(u,v)$ denotes the probability density function of the Archimedean copula and $C(u,v)$is the Archimedean copula function defined in Equation (9). The copula density function is used to evaluate the joint likelihood of the transformed variables and serves as the basis for estimating the copula parameters using the Maximum Likelihood Estimation (MLE) method. The Archimedean copula is employed because it is a simple copula model and

flexible with respect to the dimensionality of the data. Archimedean copulas include the Clayton copula, Frank copula, and Gumbel copula.

2.9 Mean Variance Efficient Portfolio

The Mean Variance Efficient Portfolio (MVEP) method is one approach used to construct an optimal stock portfolio by minimizing risk for a given level of return, thereby maximizing the expected return while maintaining a low level of risk.

$$\begin{aligned} &\boldsymbol{maximize} \quad && \boldsymbol{w^T\mu - \gamma w^T \Sigma w} \\ &\boldsymbol{subject\ to} \quad && \boldsymbol{w^T 1_N = 1} \end{aligned} \tag{11}$$

Where γdenotes the risk aversion parameter, the portfolio weights $w$are constructed to achieve minimum variance subject to two constraints: the specified target of the portfolio mean return μ_p, expressed as $w^T\mu$, and the condition that the sum of the portfolio proportions equals one, which can be written as $w^T\mathbf{1}_N = 1$.

2.10 Backtesting

A test conducted to validate previously predicted Value at Risk (VaR) results. This testing is performed to evaluate the VaR model using the quantile regression approach, as it is capable of exploring the conditional distribution under a distribution-free assumption and also allows for serial correlation and conditional heteroskedasticity. In this study, the Kupiec Test is employed; this method works by computing a likelihood ratio and comparing it with the chi-square distribution at a specified significance level [14].

$$\boldsymbol{LR = -2\ ln\left[\frac{(1-p_b)^{n-l}p_b^l}{\left(1-\frac{l}{n}\right)^{n-l}\left(\frac{l}{n}\right)^l}\right]} \tag{12}$$

Where LR denotes the likelihood ratio test statistic of the Kupiec Proportion of Failures (POF) test, $n$is the total number of observations, n is the number of VaR violations (exceptions), p is the expected probability of a VaR violation, corresponding to the significance level of the VaR model. If the LR value is greater than the chi-square value, it can be concluded that the Value at Risk is inaccurate and not properly validated.

3. RESULT AND DISCUSSION

3.1 Stock Selection and Descriptive Statistics

LQ45 index stocks were selected because they exhibit high liquidity, large market capitalization, and strong financial fundamentals, making them suitable for stable and long-term pension fund investments. From the nine mining companies included in the LQ45 index, the sample selection was based on historical data from the period 1 June 2020 to 30 June 2025, an analysis of financial fundamentals including the Debt-to-Equity Ratio (DER) and Return on Equity (ROE), and a review of each company's market capitalization. Companies were screened using the criteria of ROE greater than 12% and DER below 1. Several mining companies were excluded because they did not satisfy these requirements, particularly the ROE threshold of 12%, indicating that they did not meet the desired level of profitability for long-term investment. ROE and DER were employed because they capture the two fundamental dimensions of corporate financial performance: profitability and financial stability. ROE reflects the company's effectiveness in generating returns for shareholders, whereas DER indicates its level of financial leverage and associated risk. The combination of these ratios provides a balanced assessment of a company's ability to generate sustainable profits while maintaining a prudent capital structure. Therefore, ROE and DER are considered sufficient financial indicators for selecting companies suitable for long-term pension fund investments. Based on this screening process, three companies met the specified criteria: PT Indo Tambangraya Megah Tbk (ITMG), PT Bukit Asam Tbk (PTBA), and PT Adaro Energy Indonesia Tbk (ADRO). Furthermore, all three stocks have DER values below 1, indicating strong

capital structures and relatively low dependence on debt, which further supports their suitability for long-term pension fund investments.

This study employs the log returns of daily closing stock prices, as log returns provide a more accurate and stable representation of price changes over time and are widely used in financial time series modeling due to their desirable statistical properties. Using log returns also helps to reduce heteroskedasticity and allows for better comparison across different price levels. The following section presents the time series plots of the stocks analyzed in this study.

Table 1. Descriptive Statistics

Stock	Mean	Varians	Std. Deviation	Skewness	Kurtosis
ITMG	-0,00078	0,00056	0,02383	-0,45735	6,90982
ADRO	-0,00038	0,00008	0,02881	1,15514	20,06161
PTBA	-0,00013	0,00051	0,02264	0,76104	11,04713

Based on the descriptive statistics presented in **Table 1.**, all log return series exhibit small negative mean values, indicating no significant downward trend during the observation period and supporting the assumption of stationarity. This suggests that investment risk is primarily driven by volatility and extreme events rather than by changes in the mean return. PTBA recorded the smallest negative mean (-0.00013764), while ITMG had the largest (-0.00078254). In terms of risk, ADRO showed the highest variance and standard deviation, indicating the greatest price volatility, whereas PTBA exhibited the lowest volatility. Regarding distributional characteristics, ITMG was negatively skewed (-0.4573508), while ADRO (1.155149) and PTBA (0.7610453) were positively skewed. Furthermore, all three stocks had kurtosis values exceeding 3, indicating leptokurtic distributions with a high likelihood of extreme events. ADRO recorded the highest kurtosis (20.06161), suggesting the greatest exposure to extreme risk among the three stocks.

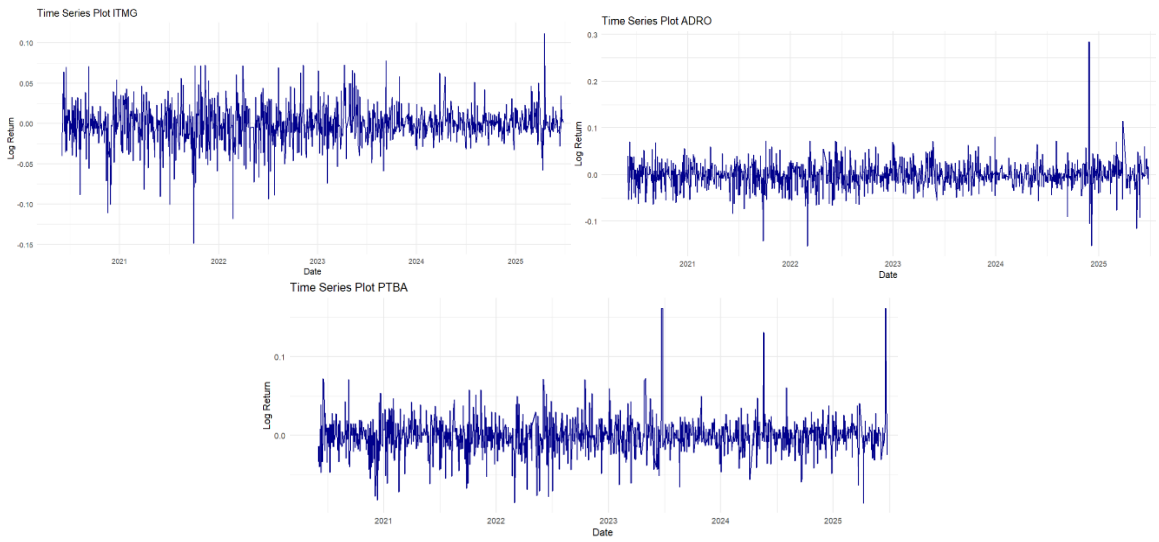


Figure 2. Log Return Time Series Plot

Figure 2. shows that all three log return series exhibit fluctuating movements with clear evidence of volatility clustering. PTBA demonstrates the most stable price movements, although several sharp fluctuations are observed during 2023 and 2025. ITMG exhibits moderate volatility with occasional extreme positive and negative returns, while ADRO displays the highest volatility, characterized by significant price spikes, particularly in 2025, and sharp declines in earlier periods. This finding is consistent with ADRO's high kurtosis value reported in **Table 1.**, indicating a greater likelihood of extreme events. Overall, the non-constant volatility observed in all three series suggests the presence of heteroskedasticity, supporting the use of volatility models such as GARCH. The preliminary analysis using descriptive statistics and time series plots indicates the presence of volatility clustering, heteroskedasticity, leptokurtic distributions, and heavy tails in all three

mining stocks. These conditions require a more comprehensive modeling approach; therefore, a hybrid method is employed to handle extreme risk and the non-normal characteristics of the distribution.

3.2 Modelling

The ARIMA model is used to identify and capture autocorrelation patterns present in the data, while the GARCH model is employed to provide more accurate information on volatility. The modeling process consists of stationarity testing, model identification, parameter estimation, residual diagnostic testing, selection of the best ARIMA model, heteroskedasticity testing, GARCH model identification, and selection of the best GARCH model.

Stationarity testing is conducted by examining two aspects: the mean and the variance. Based on the ACF plots, all stock log return series are stationary in the mean, as they do not exhibit significant movements over time. This finding is further supported by the results of the Augmented Dickey–Fuller (ADF) test, which yields P-Values below 0.05, leading to the rejection of the null hypothesis H_0. Therefore, it can be concluded that the log return data of ITMG, ADRO, and PTBA stocks are stationary in the mean and do not require a differencing process. Consequently, the differencing order $d$is set to zero, and the ARIMA model can be specified as ARIMA$(p, 0, q)$ or equivalently ARMA(p, q). This analysis is also supported by the Box–Cox test results, which indicate a lambda value of 1.

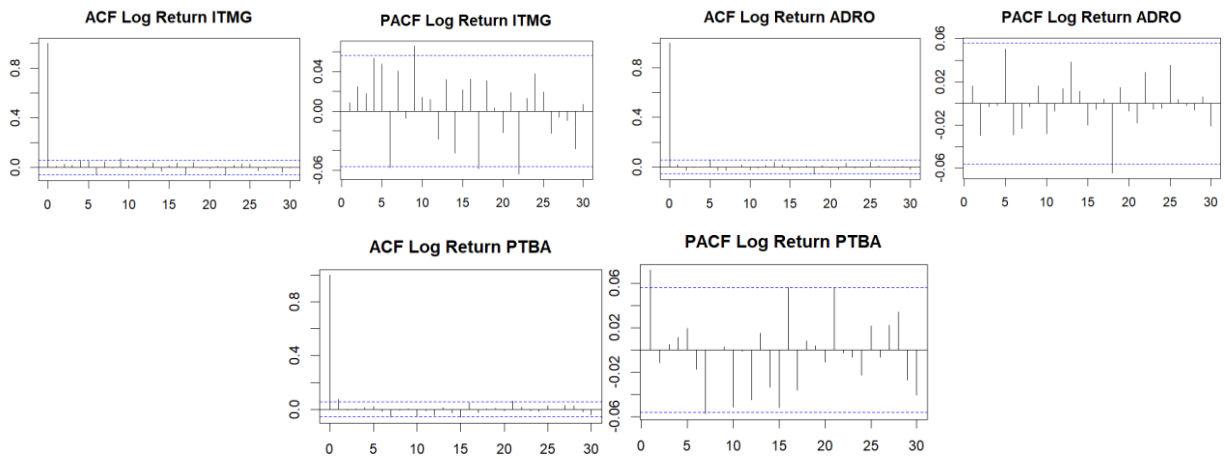


Figure 3. ACF PACF Stock Plot

Next, ARIMA model identification is carried out based on the ACF and PACF plots. Based on these plots, the ITMG and ADRO stocks exhibit a relatively large number of significant lags, making the principle of parsimony necessary. For ITMG stock, three candidate ARIMA models are considered: ARIMA White Noise, ARIMA(0,0,1), and ARIMA(1,0,0). Meanwhile, for ADRO stock, the candidate models are ARIMA(0,0,1), ARIMA(1,0,0), and ARIMA(1,0,1). For PTBA stock, the candidate models are ARIMA(0,0,1), ARIMA(1,0,0), and ARIMA(1,0,1).

Table 2. Parameter Estimation of ARIMA Modelling

Stock	Model	Parameter	Coef	P-Value
ITMG	ARIMA *White Noise*	ϕ_0	-	-
		θ_0	-	-
	ARIMA(0,0,1)	θ_1	0,0079	0,7751
	ARIMA(1,0,0)	ϕ_1	0,0084	0,2556
ADRO	ARIMA(0,0,1)	θ_1	0,0079	0,7751
	ARIMA(1,0,0)	ϕ_1	0,0084	0,2556
	ARIMA(1,0,1)	ϕ_1	-0,9499	0,0000**
		θ_1	0,9674	0,0000**
PTBA	ARIMA(0,0,1)	θ_1	0,0732	0,0109**
	ARIMA(1,0,0)	ϕ_1	0,0719	0,0118**
	ARIMA(1,0,1)	ϕ_1	-0,1033	0,7898
		θ_1	0,1759	0,6462
		φ_1	0,0549	0,1111
		φ_2	0,1379	0,0055

Based on parameter significance tests and the smallest AIC values, the best ARIMA model for ITMG stock is the White Noise ARIMA model, ARIMA(1,0,1) for ADRO stock, and ARIMA(0,0,1) for PTBA stock. The ARIMA models for each stock satisfy the white noise assumption but do not meet the normality assumption. The presence of heteroskedasticity is tested using the Lagrange Multiplier Test. The test results show that all p-values at lags 3, 5, and 7 are less than 0.05, indicating that the residuals exhibit heteroskedasticity. GARCH modeling is therefore employed to address the variance heteroskedasticity present in the residuals. In identifying the GARCH model, the ACF and PACF plots of the squared residuals from the ARIMA models are used, as these plots provide insights into the variance patterns and help in determining appropriate GARCH model specifications.

Table 3. Parameter Estimation of ARIMA Modelling

Stock	Model	Parameter	Coef	P-Value
ITMG	ARMA *White Noise* GARCH(1,0)	ω	0,0000	0,0000
		φ_1	0,9511	0,0000
	ARMA(0,0) GARCH(1,1)	ω	0,0000	0,1113
		φ_1	0,0198	0,0000**
	ARMA(0,0) GARCH(0,1)	ω	0,0000	0,0000**
		β_1	0,9990	0,0000
	ARMA(0,0) GARCH(2,0)	ω	0,0004	0,0000
		φ_1	0,0549	0,1111
		φ_2	0,1379	0,0055

Stock	Model	Parameter	Coef	P-Value
ADRO	ARMA(1,1) GARCH(1,0)	θ_1	0,8089	0,0000
		ϕ_1	0,9476	0,0000
		ω	0,0000	0,0000
		φ_1	0,8905	0,0000
	ARMA(1,1) GARCH(1,1)	θ_1	0,3570	0,1940
		ϕ_1	-0,4369	0,0961
		ω	0,0001	0,0000**
		φ_1	0,2102	0,0000**
		β_1	0,6099	0,0000**
	ARMA(1,1) GARCH(0,1)	θ_1	-0,9525	0,0000
		ϕ_1	0,9674	0,0000
		ω	0,0000	0,0000
		β_1	0,9982	0,0000
	ARMA(1,1) GARCH(2,0)	θ_1	0,3896	0,0438
		ϕ_1	-0,4726	0,0092
		ω	0,0004	0,0000
		φ_1	0,2093	0,0000
		φ_2	0,1999	0,0000
PTBA	ARMA(0,1) GARCH(1,0)	θ_1	-0,8815	0,0000
		ω	0,0000	0,5950
		φ_1	0,9987	0,0000**
	ARMA(0,1) GARCH(1,1)	θ_1	0,0000	0,9991
		ω	0,0004	0,0000
		φ_1	0,1673	0,0002
		β_1	0,0003	0,9987
	ARMA(0,1) GARCH(0,1)	θ_1	0,7246	0,0116
		ω	0,0000	0,0000
		β_1	0,9989	0,0000**
	ARMA(0,1) GARCH(2,0)	θ_1	0,0000	0,9989
		ω	0,0004	0,0000
		φ_1	0,1678	0,0002
		φ_2	0,0000	0,9989

This study adopts the principle of parsimony by prioritizing relatively simple GARCH models as comparative candidates, including GARCH(1,0), GARCH(1,1), GARCH(0,1), and GARCH(2,0). These models are evaluated to identify the most appropriate specification that adequately captures volatility dynamics while maintaining model simplicity. Based on the initial identification of GARCH models for each stock and a comparison of their Akaike Information Criterion (AIC) values, the best GARCH model for ITMG stock is GARCH(1,1), for ADRO stock is GARCH(1,1), and for PTBA stock is GARCH(0,1).

The selection of the GARCH(1,1) model for ITMG and ADRO stocks is supported by the statistical significance of the GARCH parameters, as indicated by p-values below 0.05. This suggests that both the ARCH and GARCH effects play an important role in explaining the volatility structure of these stocks. In contrast, for PTBA stock, the GARCH(1,1) model initially considered as the primary candidate exhibits insignificant

parameter estimates. Consequently, a more detailed evaluation is conducted by comparing the AIC values across all alternative GARCH models that possess statistically significant parameters.

Based on this comprehensive comparison, the GARCH(0,1) model is identified as the most suitable model for PTBA stock, as it produces the smallest AIC value among the competing models. This result indicates that the volatility dynamics of PTBA stock are sufficiently captured by a simpler GARCH structure, emphasizing the importance of balancing model complexity and goodness of fit in volatility modeling.

3.3 Extreme Value Theory Modelling

The standardized squared residuals obtained from the ARIMA-GARCH modeling do not follow a normal distribution and exhibit extreme values. Therefore, Extreme Value Theory (EVT) with the Peak Over Threshold (POT) approach is employed to model the probability of future extreme events. Prior to determining the threshold value, the standardized squared residuals from the ARIMA-GARCH model are multiplied by −1. This step is necessary because the residuals are return-type data, where extreme values occur on the left tail of the distribution and represent losses. To identify the optimal threshold, the Mean Residual Life Plot (MRLP) method is used. In this plot, the y-axis represents the mean excess loss, which measures the average magnitude of excess losses once a threshold is exceeded, while the x-axis represents the threshold value, serving as a cutoff that separates normal observations from extreme observations. Based on the MRLP results, 61 extreme observations are identified for each stock, with threshold values of 1.620132 for ITMG, 1.699751 for ADRO, and 1.567710 for PTBA.

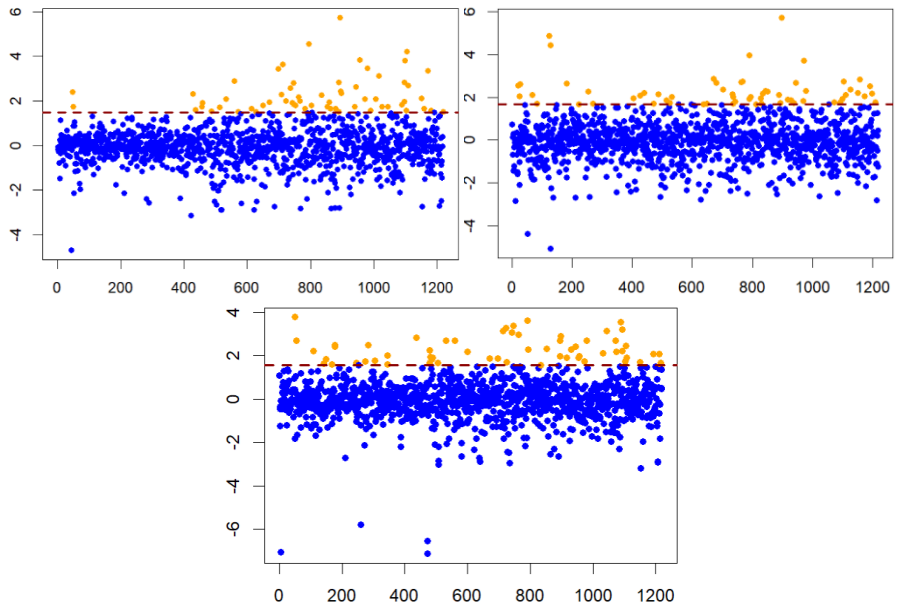


Figure 4. Extreme Data

The distribution of extreme data is represented by orange-colored points, while observations within the threshold are indicated by blue-colored points. The Generalized Pareto Distribution (GPD) is then used to estimate the extreme value parameters. The following presents the results of the parameter estimation for all stocks analyzed.

Table 4. GPD Estimation Parameter

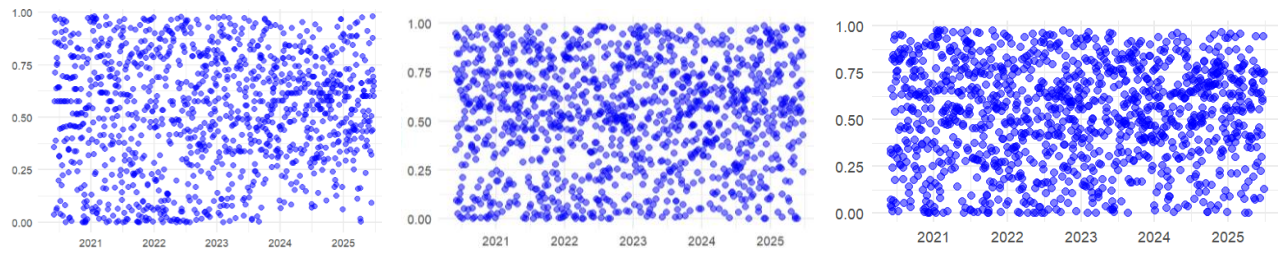
Stock	Scale	Shape
ITMG	0,65540698	0,05376682
ADRO	0,49780430	0,16618740
PTBA	1,03093400	-0,39668870

The EVT parameter estimation results indicate that PTBA stock has the largest scale parameter value, suggesting the highest level of extreme volatility, while ADRO exhibits the narrowest dispersion of extreme values, implying lower risk, and ITMG lies in a moderate position. A negative shape parameter indicates a bounded-tail distribution of the Weibull type, meaning that extreme risk is relatively controlled. In contrast, a positive shape parameter signifies a heavy-tailed Fréchet-type distribution, implying a higher probability and magnitude of extreme losses that are theoretically unbounded.

Thus, it can be concluded that the extreme risk of PTBA stock can be relatively well controlled, even though extreme losses tend to be large when events exceed the threshold value. These findings are also supported by the results of the goodness-of-fit tests, which confirm that the GPD adequately fits the extreme value data for all stocks. The estimated scale and shape parameters are subsequently used in the cumulative distribution function (CDF), which serves as the transformation method prior to further modeling of the data using copula techniques.

3.4 Copula Modelling

Copula modeling is employed to capture the dependence structure among stocks by combining their marginal distributions into a joint distribution according to Sklar's Theorem. Since copulas operate on the Uniform $[0,1]$ scale, the residuals are transformed using the Probability Integral Transformation (PIT). A semi-parametric marginal distribution is adopted, in which residuals below the selected threshold are modeled using the Normal cumulative distribution function (CDF) to represent the central part of the distribution, while residuals exceeding the threshold are modeled using the Generalized Pareto Distribution GPD CDF to capture the tail behavior. The Normal CDF and GPD CDF are joined at the selected threshold to form a continuous marginal distribution before applying the PIT. This approach is justified because the GPD is theoretically valid only for exceedances over a sufficiently high threshold, whereas the Normal distribution provides an adequate approximation for the bulk of the residuals. Consequently, the resulting marginal distribution preserves both the regular behavior of the data and the extreme tail characteristics. Finally, the transformed observations are mapped onto the Uniform $[0,1]$ interval, providing a common probabilistic scale for copula estimation.



(a) (b) (c)

Figure 5. Scatterplot After Transformation

This process also results in the formation of three portfolios: Portfolio A, which consists of ITMG and ADRO stocks; Portfolio B, which consists of ITMG and PTBA stocks; and Portfolio C, which consists of ADRO and PTBA stocks. The portfolio construction was limited to pairwise combinations because this study employed bivariate Archimedean copulas, which model the dependence structure between two random variables. Consequently, the analysis was conducted separately for each stock pair. Compared with multivariate copulas, bivariate copulas require fewer parameters, provide more stable and computationally

efficient estimation, and offer easier interpretation while effectively capturing nonlinear and tail dependence between asset pairs.

Table 5. Kendall's Tau Test

Portfolio	τ	Test Statistic	P-Value
A	0,3496271	18,284	$< 2{,}2 \times 10^{-16}$
B	0,3716653	19,405	$< 2{,}2 \times 10^{-16}$
C	0,3708083	19,360	$< 2{,}2 \times 10^{-16}$

All portfolios exhibit test statistics greater than $Z_\alpha$and p-values less than 0.05; therefore, the null hypothesis $H_0$is rejected, leading to the conclusion that the two assets in each portfolio are not independent. Based on **Table 3.**, the Kendall's Tau (τ) coefficients for all portfolios are positive, indicating that the assets within each portfolio have a positive dependence structure. This implies that when one stock increases, the other stock in the portfolio tends to increase as well. Among the portfolios, Portfolio B has the largest Kendall's Tau coefficient, indicating the strongest dependence between its constituent stocks compared to the other portfolios.

This study employs the Archimedean copula family, including the Clayton, Frank, and Gumbel copulas, to model the dependence structure between assets. The optimal copula model is selected based on the minimum Akaike Information Criterion (AIC) value, which indicates the best balance between model fit and complexity. The following presents the parameter estimates and corresponding AIC values for each selected copula model.

Table 6. Copula Parameter Estimation

Portfolio	Best Copula	Parameter	AIC
A	Gumbel	1,555712	-375,7532
B	Gumbel	1,728990	-489,3500
C	Gumbel	1,651336	-436,7015

This indicates that Portfolio A, Portfolio B, and Portfolio C exhibit more dominant upper-tail dependence, making the Gumbel copula model more appropriate. Subsequently, data are generated based on the estimated copula parameters and then transformed back into return form for portfolio weighting and risk estimation.

3.5 Mean Variance Efficient Portfolio

The next stage involves applying the Mean Variance Efficient Portfolio (MVEP) method to determine the optimal portfolio weights that minimize risk while maximizing expected returns. The portfolio optimization is performed using simulated return data generated through Monte Carlo simulation based on the best-fitting copula model. Specifically, random samples are generated from the selected copula to preserve the dependence structure among the asset returns, and these simulated observations are subsequently transformed back to the original return scale using the estimated marginal distributions. By utilizing the simulated data, the optimization accounts for both the marginal behavior and the dependence structure of the asset returns under the fitted hybrid ARIMA-GARCH-EVT-Copula framework. This process produces statistically efficient portfolios that better represent the joint behavior of the assets under various market conditions, thereby providing a more robust investment strategy for pension fund risk management. The following presents the portfolio weights for all portfolios.

Although the Mean Variance Efficient Portfolio (MVEP) framework is traditionally developed under a mean–variance setting, it remains applicable in this study because the optimization is performed using Monte Carlo simulated returns generated from the fitted ARIMA-GARCH-EVT-Copula model rather than the original return series. The simulated data preserve the heavy-tailed marginal behavior through EVT and the dependence structure through the selected copula model. Consequently, the estimated mean vector and covariance matrix used in the MVEP optimization reflect these distributional characteristics, allowing the

optimization to account indirectly for extreme events while maintaining the computational simplicity and interpretability of the mean–variance framework. Nevertheless, because MVEP relies solely on the first two moments of the simulated returns, it may not fully capture tail risk in the optimization process. Therefore, alternative portfolio optimization methods based on coherent risk measures, such as Conditional Value-at-Risk (CVaR) optimization or Expected Shortfall optimization, could be explored in future studies.

Table 7. Weight Portfolio

Portfolio	Stocks	Weight
A	ITMG	62,11456%
	ADRO	37,88544%
B	ITMG	19,93126%
	PTBA	80,06874%
C	ADRO	8,668356%
	PTBA	91,33164%

This result suggests that the MVEP optimization favors PTBA because it provides a more favorable risk–return trade-off than the other stocks. Specifically, PTBA exhibits the lowest return volatility among the selected mining stocks while maintaining relatively strong financial fundamentals, allowing the optimization process to reduce overall portfolio variance without substantially sacrificing expected returns. In contrast, ADRO receives the smallest allocation due to its relatively high volatility, whereas ITMG is assigned a moderate weight depending on its contribution to portfolio diversification. Consequently, PTBA becomes the dominant component in the optimal portfolios as it offers the greatest contribution to risk reduction while maintaining portfolio efficiency under the simulated dependence structure generated by the selected Gumbel copula model.

3.6 Value at Risk and Expected Shortfall

The application of the MVEP method is able to provide optimal stock combinations for each Gumbel copula–based portfolio, reflecting a balance between return and risk. However, prior to estimation, a Monte Carlo simulation of the generated data is conducted with 10,000 iterations in order to represent various possible future portfolio return scenarios that follow the underlying dependence structure. Subsequently, the expected portfolio return is calculated to reflect the combined performance of all stocks according to their investment weights. The following presents the Value at Risk and Expected Shortfall results for all portfolios.

Table 8. Value at Risk and Expected Shortfall Estimation

Portfolio	Confidence interval	Value at Risk	Expected Shortfall
A	90%	0,0276	0,0369
	95%	0,0338	0,0431
	99%	0,0482	0,0562
B	90%	0,0244	0,0337
	95%	0,0314	0,0397
	99%	0,0435	0,0512
C	90%	0,0262	0,0353
	95%	0,0328	0,0411
	99%	0,0466	0,0505

The VaR and ES values reported in **Table 8.** are expressed in log-return units, as the analysis was performed using daily log returns. Since the estimated values are relatively small, they can be approximately

interpreted as percentage losses. At the 90%, 95%, and 99% confidence levels, Portfolio A exhibits the highest VaR values (0.0276, 0.0338, and 0.0482, respectively), corresponding to estimated maximum daily losses of approximately 2.76%, 3.38%, and 4.82%. Similarly, Portfolio A records the highest ES values, indicating that the expected average loss beyond the VaR threshold is approximately 3.69%, 4.31%, and 5.62%, respectively. In contrast, Portfolio B consistently reports the lowest VaR and ES values, suggesting lower downside risk across all confidence levels. Moreover, Portfolio A also records the highest Expected Shortfall (ES) compared to the other two portfolios, implying that it suffers the largest average losses once the VaR threshold is exceeded.It can therefore be concluded that Portfolio B and Portfolio C represent the most stable options for investors with a moderate risk profile, whereas Portfolio A is more suitable for aggressive investors who are willing to bear higher risk in pursuit of greater returns.

3.7 Backtesting

Backtesting is used to evaluate the accuracy of VaR estimates by examining whether the observed violation rate is consistent with, or not significantly different from, its theoretical proportion.

Table 9. Backtesting Test

Portfolio	Confidence interval	Exceptions	LR	P-Value
A	90%	121	0,00739	0,93145
	95%	61	0,00004	0,99475
	99%	13	0,05321	0,81756
B	90%	120	0,03305	0,85572
	95%	59	0,06634	0,79673
	99%	13	0,05321	0,81756
C	90%	122	0,00009	0,99238
	95%	62	0,01893	0,89054
	99%	12	0,00300	0,95627

Based on the test results, all portfolios at the 90%, 95%, and 99% confidence levels show p-values greater than 0.05 and LR values smaller than the chi-square critical values, leading to the decision to fail to reject the null hypothesis H_0. This indicates that the estimated VaR values are valid at these confidence levels. These findings suggest that actual losses exceeding the VaR thresholds occur infrequently in the model and that the model is able to accurately reflect the level of risk in line with actual market dynamics. Overall, this demonstrates that the ARIMA-GARCH-EVT-Copula modeling approach is effective in capturing market volatility dynamics, heavy-tailed distribution characteristics, and extreme dependence among stocks. The model also successfully reflects market conditions that are sensitive to commodity price fluctuations and global economic uncertainty.

4. CONCLUSIONS

This study estimated the extreme risk of LQ45 mining sector stocks using a hybrid ARIMA-GARCH-EVT-Copula approach. The sample consisted of three mining companies (ITMG, PTBA, and ADRO), selected based on their inclusion in the LQ45 index, the availability of historical data from 1 June 2020 to 30 June 2025, and strong financial fundamentals, as indicated by average Return on Equity (ROE) values above 12% and Debt-to-Equity Ratio (DER) values below 1. The results demonstrate that the ARIMA-GARCH models effectively capture autocorrelation and volatility clustering, while the EVT model accurately estimates tail risk. The dependence analysis indicates that the Gumbel copula provides the best representation of the dependence structure among the selected mining stocks. Furthermore, the Mean Variance Efficient Portfolio (MVEP)

assigns the largest weight to PTBA, resulting in a portfolio with lower extreme risk and supporting its suitability as a long-term investment alternative for pension funds.

From a practical perspective, the proposed hybrid framework provides pension fund managers and institutional investors with a systematic approach to evaluating mining stocks by simultaneously considering volatility, extreme downside risk, and dependence among assets. The findings suggest that PTBA and ITMG can be considered relatively more defensive investment alternatives, while the optimized portfolio offers a quantitative basis for portfolio allocation aimed at balancing return and risk under extreme market conditions. Consequently, the proposed methodology can support more informed long term investment decision making and portfolio risk management for pension funds.

5. ACKNOWLEDGMENTS

This section include gratitude to other contribution that support or involved in the research

6. REFERENCES

[1] I. M. Adnyana, "Manajemen Investasi Dan Portfolio," Lembaga Penerbitan Universitas Nasional (LPU-UNAS), Jl. Sawo Manila, No. 61. Pejaten, Pasar Minggu, Jakarta Selatan 12520, 2020. Accessed: Jul. 09, 2025. [Online]. Available: http://repository.unas.ac.id/3770/1/BUKU%20IINVESTASI%20REVISI%209%20OK.20.pdf

[2] Hartati and I. Saluza, "Aplikasi GARCH dalam Mengatasi Volatilitas Pada Data Keuangan," *Jurnal Matematika*, vol. 7, no. 2, p. 107, Dec. 2017, doi: https://doi.org/10.24843/jmat.2017.v07.i02.p87.

[3] A. M. Hakam and A. K. Jaya, "Estimating Conditional Value at Risk in Non-Cyclical Sector Companies Using the Extreme Value Theory Approach," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 159–175, Sep. 2024, doi: https://doi.org/10.20956/j.v21i1.35849.

[4] N. P. Iriani, M. S. Akbar, and Haryono, "Estimasi Value at Risk (VaR) Pada Portfolio Saham Dengan Copula," vol. 2, no. 2, p. 16079, Sep. 2013, doi: https://doi.org/10.12962/j23373520.v2i2.4653.

[5] F. Jin, J. Li, and G. Li, "Modeling the Linkages between Bitcoin, Gold, Dollar, Crude Oil, and Stock Markets: A GARCH-EVT-Copula Approach," *Discrete Dynamics in Nature and Society*, vol. 2022, no. 1, Jan. 2022, doi: https://doi.org/10.1155/2022/8901180.

[6] T. Eduardus, *Analisis Investasi dan Manajemen Portfolio*. Jogyakarta: BPFE, 2010.

[7] R. S. Hudson and A. Gregoriou, "Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns," *International Review of Financial Analysis*, vol. 38, pp. 151–162, Mar. 2015, doi: https://doi.org/10.1016/j.irfa.2014.10.008.

[8] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Cham Springer International Publishing, 2016.

[9] Mills, T. C, *Applied time series analysis : a practical guide to modeling and forecasting*. Academic Press, 2019.

[10] W. W.-S. Wei, *Time series analysis : univariate and multivariate methods*. Boston: Pearson Addison Wesley, Cop, 2006.

[11] S. Kotz and S. Nadarajah, *Extreme Value Distributions*. World Scientific, 2000.

[12] R. B. Nelsen, *An introduction to copulas*. New York: Springer, 2006.

[13] P. Trivedi and D. Zimmer, "the essence of knowledge Foundations and Trends ® in Econometrics Copula Modeling: An Introduction for Practitioners," 2005.

[14] N. A. Purnamasari, "Backtesting Untuk Value at Risk Pada Data Return Saham Bank Syariah Menggunakan Quantile Regression," 2017.