Iranian Journal of Numerical Analysis and Optimization
Vol. 16, No. 3, 2026, pp 852–875
https://doi.org/10.22067/ijnao.2026.95106.1716
https://ijnao.um.ac.ir/

Research Article

A probabilistic approach to portfolio selection with Risk-Aversion coefficient estimated via logistic regression

L. Chin*, A. Sukmana, E. Chendra and A.P. Wijaya

Abstract

This work presents a probabilistic risk-aversion approach to portfolio selection that combines a logistic regression-based risk assessment framework with a mixed-integer optimization algorithm with mean semi-absolute deviation as the risk criterion. It allows us to incorporate uncertainty in the portfolio selection analysis as well as performance analysis by Monte Carlo

*Corresponding author

Received 27 August 2025; revised 18 December 2025; accepted 18 April 2026

Liem Chin
Center for Mathematics and Society, Department of Mathematics, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia. e-mail: chin@unpar.ac.id

Agus Sukmana
Center for Mathematics and Society, Department of Mathematics, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia. e-mail: asukmana@unpar.ac.id

Erwinna Chendra
Center for Mathematics and Society, Department of Mathematics, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia. e-mail: erwinna@unpar.ac.id

Andreas Parama Wijaya
Center for Mathematics and Society, Department of Mathematics, Faculty of Science, Parahyangan Catholic University, Bandung, Indonesia. e-mail: a.p.wijaya@unpar.ac.id

How to cite this article
Chin, L., Sukmana, A., Chendra, E. and Wijaya, A.P., A probabilistic approach to portfolio selection with Risk-Aversion coefficient estimated via logistic regression. *Iran. J. Numer. Anal. Optim.*, 2026; 16(3): 852-875. https://doi.org/10.22067/ijnao.2026.95106.1716

simulation and compares the portfolios with different levels of risk over time instead of applying a deterministic system of selection based on historical data. The technical indicators of the LQ45 index predict the risk-aversion coefficient and can steer the model to the direction of the market. Under cardinality, quantity, and liquidity constraints, seven portfolios constructed from major Indonesian stock indices can be optimized with a Branch-and-Bound method that operates in GEKKO Python. Simulation results on 10,000 runs are taken to give the portfolios probabilistic rankings of return and risk. The probability values computed using the Monte Carlo method are plotted against the sample size to illustrate the convergence behavior of the algorithm. Eventually, to evaluate the robustness of the algorithm, variations in the probability values are examined by adding normally distributed noise into the model input parameters.

AMS subject classifications (2020): Primary 91G10; Secondary 91G60.

Keywords: Logistic regression; Mean semi-absolute deviation; Monte-Carlo simulation; Portfolio selection; Risk-aversion coefficients.

1 Introduction

After Markowitz introduced the mean-variance portfolio theory [13], many research topics have been based on it. For example, studies by Anagnostopoulos and Mamanis [1] and Cesarone, Scozzari, and Tardella [3] built upon the theory, while Simaan [17] compared it to other theories, such as mean absolute deviation (MAD). There is also research in which portfolio risk is measured using MAD, mean semi-absolute deviation (MSAD), or Gini's mean difference (GMD) instead of the variance used by Markowitz.

Konno and Yamazaki [11] and Konno and Wijayanayake [10] chose MAD as a measure of portfolio risk for two reasons: It is a more straightforward approach, and the Markowitz model, which uses a covariance matrix, is considered impractical due to computational difficulties and implementation complications. MAD accounts for all deviations of a portfolio's return rate from its expected value, both smaller and larger than the expected value. However, a rational investor might only consider deviations below the expected value as actual risk. In other words, variability in the return rate of a portfolio above the average should not be penalized, as investors are more concerned with poor performance (underperformance) than with excessive performance (overperformance). For this reason, MSAD is preferred over MAD, as seen in the work of Sortino and Meer [18]. Although they did not explicitly use the term MSAD, their article introduced the concept of downside deviation, which is the basis of MSAD. They argued that investors are more concerned with the risk of loss, so risk measurement should focus on deviations below the average. Other researchers who have used MSAD as a measure of portfolio risk include Fang, Lai, and Wang [4] and Yixuan and Zhongfeng [21].

Using GMD for risk measurement, as done by Ji, Lejeune, and Prasad [8] and Jiang, Ji, and Chang [9], is actually better than MSAD. This is because GMD is calculated by averaging all absolute differences between every pair of data points in a data set, whereas MSAD is calculated by averaging the absolute deviations of returns that are below the average return. However, this increases the computational complexity of GMD at $O(n^2)$, compared to MSAD, which is only $O(n)$. For this reason, we chose MSAD as our portfolio risk measure over GMD.

Apart from risk measurement, there are other things that need to be considered in forming a portfolio, that is, risk-aversion adjustment. This adjustment is exhibited by Jiang, Ji, and Chang [9] and Grant, Kwon, and Satchell [6]. The risk-aversion coefficient is needed to reflect the market trend [9]. Finally, there are several practical constraints that need to be considered in portfolio selection: Cardinality (the number of assets in a portfolio) [1, 3, 14, 20], quantity (the proportion of funds for each asset) [1, 3, 20], and transaction cost [4].

In general, research on portfolio optimization uses deterministic methods in formulating problems [22]; therefore, an alternative approach is needed to consider uncertainty and address the problem probabilistically, as proposed by Shadabfar and Cheng [15] and Xiaoxia [19]. The main objective of this research is to provide a consideration for investors by adopting Shadabfar-Cheng's algorithm. Our research introduces several key methodological enhancements and extensions compared to the work of Shadabfar and Cheng [15]. The main differences are as follows.

1. The portfolio risk measure is calculated using the MSAD, rather than the portfolio variance.
2. We incorporate a risk-aversion coefficient that reflects the market trends.
3. The model is extended by incorporating cardinality, quantity, transaction cost, and liquidity constraints.
4. We display the proportion of funds for each asset, which is essential for investors but is often not discussed in many researches.

We use Monte Carlo simulation to present a probabilistic solution for the individual return and risk of portfolios [16].

The remainder of this paper is organized as follows. Section 2 begins with a brief introduction to the fundamental methods. Section 3 then describes the portfolio optimization model we constructed and the probabilistic approach methodology employed in this research. Section 4 is dedicated to the discussion and analysis of the research results, including the convergence of Monte Carlo simulation and robustness of the algorithm to random perturbations. Finally, Section 5 provides the conclusion based on the obtained findings.

2 Overview

2.1 Logistic regression

Logistic regression is a type of statistical analysis used to model the relationship between one or more independent variables (predictors) and a single categorical or binary dependent variable. Unlike linear regression, which predicts a continuous value, logistic regression predicts the probability that an event occurs, with the output being a value between 0 and 1.

In the context of this study, logistic regression is used to predict the direction of a stock index movement and to determine the risk-aversion coefficient α, as demonstrated by Jiang, Ji, and Chang [9]. Although Jiang et al. actually used both logistic regression and XGBoost for this prediction, the accuracy of the two methods was not significantly different (XGBoost was only 1.01% more accurate than logistic regression). For this reason and because logistic regression is easier to implement and interpret, this study uses logistic regression to predict the direction of the movement of the stock index.

According to James et al. [7], logistic regression can be expressed as

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i X_i, \tag{1}$$

where $X = (X_1, X_2, \ldots, X_k)$ are the k predictors. Equation (1) can be rewritten as

$$p(X) = \frac{e^{\beta_0 + \sum_{i=1}^{k} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{k} \beta_i X_i}}.$$

The probability value $p(X)$ is then used as an input for the risk-aversion coefficient α in the portfolio optimization model [9].

2.2 Branch & Bound method

The Branch and Bound algorithm is a divide-and-conquer methodology for solving discrete optimization problems. We adopted the robust method from [12], which is often partially responsible for successfully solving complex problems.

Consider the integer linear program (ILP):

$$\begin{aligned} z_P := \max \quad & \sum_{j=1}^{n} c_j y_j \\ \text{subject to :} \quad & \\ & \sum_{j=1}^{n} a_{ij} y_j \leq b_i, \text{for } i = 1, 2, \ldots, m, \\ & y_j \in \mathbb{Z}, \text{for } j = 1, 2, \ldots, n. \end{aligned}$$

We assume that the feasible region of the linear programming (LP) relaxation of P is a bounded set. Consequently, the set of feasible solutions for a subproblem P' is a finite set.

1. **Upper Bounds:** Solve the linear programming relaxation of the subproblem P'. Let y^* be its optimal solution. The optimal objective function value of the LP relaxation of P' serves as an upper bound for $z_{P'}$.

2. **Branching Rule:** Select a variable y_k that is not an integer, that is, $y_k^* \notin \mathbb{Z}$. The branching is performed by creating two new subproblems: (a) P' with the additional constraint $y_k \leq \lfloor y_k^* \rfloor$, and (b) P' with the additional constraint $y_k \geq \lceil y_k^* \rceil$.

3. **Lower Bounds:** If the solution y^* of the LP relaxation of P' happens to be in S' (the set of feasible solutions to the integer linear program P'), then the objective value $z^* := c(y^*)$ constitutes a lower bound on $z_{P'}$. This bound can be further refined. Even if the optimal solution is not feasible for P', feasible solutions to P' can be visited during the process of solving its LP relaxation (e.g., using the primal simplex method). Each of these solutions also provides a lower bound on $z_{P'}$. Furthermore, it may be possible to perturb the solution of the LP relaxation to obtain a feasible solution to P'; for instance, if the coefficients a_{ij} are all nonnegative, then rounding down the components of y^* provides a feasible solution.

3 Problem formulation and methodology

3.1 Portfolio model

In this study, portfolio risk is measured using the MSAD, as proposed by Fang, Lai, and Wang [4].

Assume that a portfolio consists of n assets and that r_{ij} represents the return of the ith asset in the jth period, with $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. It is also assumed that $m > n$. Let y_i, for $i = 1, 2, \ldots, n$, denote the proportion of investment for the ith asset, with $\sum_{i=1}^{n} y_i = 1$. With these proportions, we assume that investors can buy assets in real numbers.

The semi-absolute deviation of the portfolio return with respect to the expected return in period $j = 1, \ldots, m$ is defined as follows [4]:

$$v_j(y) = \min\left(0, \sum_{i=1}^{n}(r_{ij} - \bar{r}_i)y_i\right) = \frac{1}{2}\left(\sum_{i=1}^{n}(r_{ij} - \bar{r}_i)y_i - \left|\sum_{i=1}^{n}(\bar{r}_i - r_{ij})y_i\right|\right)$$

with $\bar{r}_i$ being the average return of the ith asset.

Thus, the MSAD (mean semi-absolute deviation) of the portfolio return can be expressed as:

$$V(y) = \frac{1}{m}\sum_{j=1}^{m} v_j(y) = \frac{1}{2m}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}(r_{ij} - \bar{r}_i)y_i - \left|\sum_{i=1}^{n}(\bar{r}_i - r_{ij})y_i\right|\right),$$

where $y = (y_1, y_2, \ldots, y_n)^\top$ and $y_i \in [0, 1]$.

The portfolio return without transaction cost is given by $\sum_{i=1}^{n} \bar{r}_i y_i$. If we assume that the transaction cost for each asset is γ_i. Then the total transaction cost is $\sum_{i=1}^{n} \gamma_i y_i$, so we have to subtract the previous return by this cost to get the net of return portfolio, that is,

$$\sum_{i=1}^{n} \bar{r}_i y_i - \sum_{i=1}^{n} \gamma_i y_i.$$

The portfolio optimization model in which risk is measured using MSAD is given by

$$\max \quad z_1 = \sum_{i=1}^{n} \bar{r}_i y_i - \sum_{i=1}^{n} \gamma_i y_i, \tag{2a}$$

$$\min \quad z_2 = \frac{1}{2m}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}(r_{ij} - \bar{r}_i)y_i - \left|\sum_{i=1}^{n}(\bar{r}_i - r_{ij})y_i\right|\right), \tag{2b}$$

subject to

$$\sum_{i=1}^{n} s_i = K, \tag{2c}$$

$$0 \leq y_i \leq \delta_i s_i, \quad i = 1, \ldots, n, \tag{2d}$$

$$l_i \leq \delta_i \leq u_i, \quad i = 1, \ldots, n, \tag{2e}$$

$$\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \gamma_i y_i = 1, \tag{2f}$$

$$\sum_{i=1}^{n} \lambda_i y_i \geq L, \tag{2g}$$

$$s_i \in \{0, 1\}, \quad i = 1, \ldots, n, \tag{2h}$$

$$y_i \geq 0, \quad i = 1, \ldots, n. \tag{2i}$$

The portfolio model (2) can be explained as follows:

1. The functions z_1 (2a) and z_2 (2b) represent the portfolio return and risk, respectively. Since investors naturally seek to maximize returns and minimize risk, the problem is formulated as one of maximizing returns and minimizing risk.

2. The constraint $\sum_{i=1}^{n} s_i = K$ (2c) is a **cardinality constraint**, where K denotes the number of assets in the portfolio. The binary variable s_i (2h) indicates whether an asset i is included in the portfolio, thereby allowing investors to limit the number of assets.

3. The constraints $0 \leq y_i \leq \delta_i s_i$ (2d) and $l_i \leq \delta_i \leq u_i$ (2e), for $i = 1, \ldots, n$, are **quantity constraints**. They ensure that if an asset i is in the portfolio, its investment proportion is limited to a range between 0 and δ_i. The investor's profile determines the values of l_i and u_i.

4. The constraint $\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \gamma_i y_i = 1$ (2f) means that the entire investment fund is used for portfolio construction after deducting transaction fees.

5. The constraint on equation (2g) is called a **liquidity constraint**. The liquidity of the ith stock, denoted as λ_i and its calculated as the ratio of the average number of shares traded to the number of listed shares. Investors set the number L for this liquidity constraint

6. The constraint $y_i \geq 0$, for $i = 1, \ldots, n$ (2i), prohibits short-selling in portfolio construction.

The model (2) is a multi-objective optimization problem. To simplify it, the model is transformed into a single-objective optimization problem, as follows:

$$\min \Bigg[-\alpha \left(\sum_{i=1}^{n} \bar{r}_i y_i - \sum_{i=1}^{n} \gamma_i y_i \right) + (1-\alpha) \left(\frac{1}{2m} \sum_{j=1}^{m} \left(\sum_{i=1}^{n} (r_{ij} - \bar{r}_i) y_i - \left| \sum_{i=1}^{n} (\bar{r}_i - r_{ij}) y_i \right| \right) \right) \Bigg].$$

The coefficient $\alpha \in [0,1]$ is referred to as the risk-aversion coefficient, and the investor typically determines its value. However, in this study, α is derived from the predicted value $p(x)$ of a logistic regression model.

If the value of α is greater than or equal to 0.5, the market is predicted to be in a bullish trend. In contrast, if it is less than 0.5, a bearish trend is expected. For $\alpha = 0$, the probability of a market downturn is 1. Consequently, the model is equivalent to a risk-minimization strategy, indicating that the investor is highly risk-averse during a downward market trend. Conversely, when $\alpha = 1$, the probability of a market upturn is 1. The model, therefore, becomes a return-maximization strategy, which means the investor becomes more return-oriented during an upward trend.

By eliminating the absolute value function in the risk term, we can transform the initial optimization model into

$$\min \quad -\alpha R + (1-\alpha) \frac{1}{m} \sum_{j=1}^{m} u_j$$

subject to

$$\begin{aligned}
R = r^\top y - \Gamma^\top y = \sum_{i=1}^{n} \bar{r}_i y_i - \sum_{i=1}^{n} \gamma_i y_i, \\
u_j + \sum_{i=1}^{n} (\bar{r}_i - r_{ij}) y_i \geq 0, \quad j = 1, 2, \ldots, m, \\
u_j \geq 0, \quad j = 1, 2, \ldots, m, \\
\sum_{i=1}^{n} s_i = K, \\
0 \leq y_i \leq \delta_i s_i, \quad i = 1, \ldots, n, \\
l_i \leq \delta_i \leq u_i, \quad i = 1, \ldots, n, \\
\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \gamma_i y_i = 1, \\
\sum_{i=1}^{n} \lambda_i y_i \geq L,
\end{aligned}$$

$$s_i \in \{0,1\}, \quad i = 1, \ldots, n,$$
$$y_i \geq 0, \quad i = 1, \ldots, n.$$

where $r = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_n)^\top$ represents the vector of average returns for assets over a certain period, $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)^\top$ represents the vector of transaction cost for assets, R represents the net portfolio return, and $u_j = \sum_{i=1}^{n}(r_{ij} - \bar{r}_i)y_i$ is an auxiliary non-negative variable.

3.2 Study design

The assets used to analyze the portfolio model in this study are stocks, and all adjusted closed stock and LQ45 index price data were obtained from Yahoo Finance using Python software.

In general, there are two stages to solve this portfolio optimization problem:

1. **Deterministic portfolio formation.** This stage begins with selecting a target portfolio, predicting the risk-aversion coefficient, and constructing an optimum portfolio based on historical data. The final output of this stage is the proportion of funds for each stock, as well as the portfolio's return and risk.

2. **Probabilistic modeling of portfolio optimization.** This stage is conducted to provide investors with an alternative consideration in selecting their portfolio by looking at the best possible portfolios from various existing target portfolio selections.

The discussion continues with a detailed explanation of each stage and is summarized in Figure 1.

We will analyze seven portfolios in this study, which consist of stocks included in seven different indices in Indonesia: IDX30, IDXQ30 (IDX Quality 30), IDXV30 (IDX Value 30), IDXG30 (IDX Growth 30), ECONOMIC30 (IDX Cyclical Economic 30), JII (Jakarta Islamic Index), and IDXSHAGROW (IDX Sharia Growth). Each of these indices is composed of 30 stocks, with the list accessible on the Bursa Efek Indonesia website [23].

To form these seven portfolios, the risk-aversion coefficient first needs to be predicted using a logistic regression model. According to Jiang, Ji, and Chang [9], there are five critical technical indicators used as predictors to forecast the risk-aversion coefficient. However, for simplification, we only used four out of these five indicators: EMV(10) (Ease of Movement), MACD(2,3,3) (Moving Average Convergence/Divergence), RSI(10) (Relative Strength Index), and ADX(3) (Average Directional Movement Index). We can find all the formulas for these four indicators in [9].

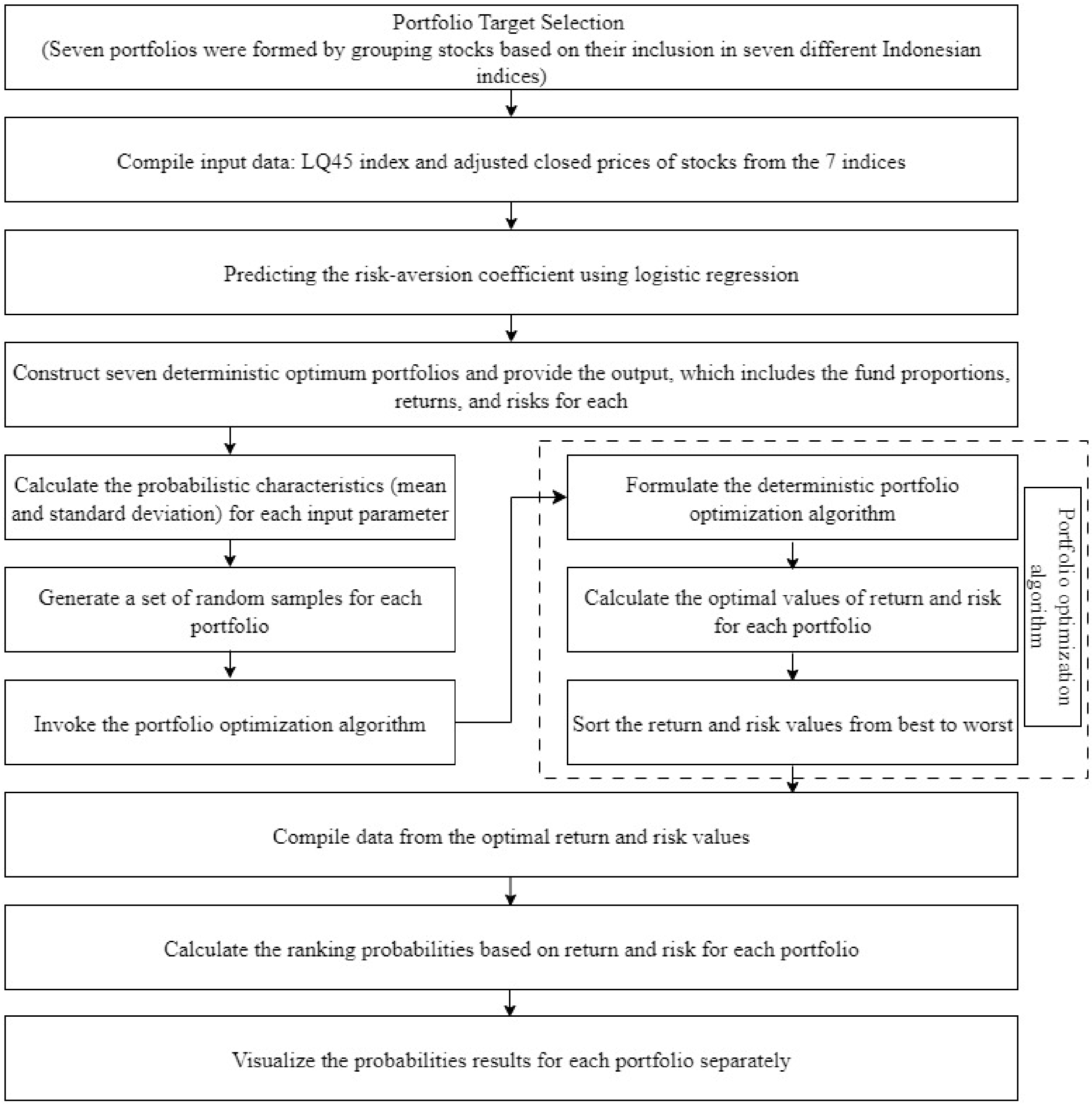

Figure 1: Scheme of the proposed algorithm.

The procedure for calculating the risk-aversion coefficient α is as follows:

1. Collect monthly data for the LQ45 index and use τ periods to build a logistic regression model to determine the risk-aversion coefficient, and a portfolio can finally be formed. In this study, we chose $\tau = 300$ months (25 years). This data is divided into training and testing data with a ratio of 4:1.

2. Based on the obtained LQ45 index values, four technical indicators are calculated and used as predictors to forecast the probability of the LQ45 index appreciating. In addition to the adjusted-closing price, we also need monthly data on the highest and lowest prices of the LQ45 index to calculate EMV(10) and ADX(3), as well as additional trading volume data to calculate EMV(10). The prediction result is then used as the risk-aversion coefficient α in the portfolio model.

The next step is to prepare the daily stock price and trading volume data from the seven indices targeted for portfolio selection. A total of 1,201 price data points are taken to obtain 1,200 daily return data points from these stocks. Meanwhile, we also need 1,200 trading volume and the number of listed stocks data for each stock to calculate their liquidity. Data on the number of listed stocks can be obtained from website IDX [23]. Then, the following steps are performed:

1. Calculate the daily return for each stock.

2. Assuming there are 20 trading days in a month then divide the 1,200 daily return data obtained in the previous step into 60 groups, with each group representing one month. Do this for each stock. Calculate the monthly return r_{ij} for every stock by summing the daily returns in each month. Therefore, each stock has 60 monthly return values.

3. Calculate the expected monthly return of the ith stock (i.e., $\bar{r}_i$) by averaging the monthly returns obtained in the previous step. Do this for all stocks.

4. With the same assumptions in the return calculation (there are 20 trading days in 1 month), we divide the 1,200 trading volume data into 60 groups with each group representing one month. Calculate the average monthly trading volume for every stock so that we can obtain the liquidity of each stock (λ_i) by calculating the ratio of average monthly trading volume to the volume of listed stock.

5. Solve the optimization model to form the seven portfolios discussed, using the GEKKO package in Python. This package already includes the Branch & Bound subroutine to solve the Mixed-Integer Linear Programming (MILP) problem, which is our portfolio optimization model. Given that seven indices are being discussed, the constructed portfolios are named A-G, where Portfolio A consists of stocks from the IDX30 index, Portfolio B consists of stocks from IDXQ30 index, and so on.

The output of this process is the proportion of funds, return, and risk for all portfolios. This output is used as a reference for investors to choose a portfolio that fits their profile. To achieve this goal, investors need to see the possibilities of the best portfolios from the seven existing portfolios. For this purpose, a probabilistic approach is taken using Monte Carlo simulation, and it is assumed that the daily return of each stock is a normally distributed random variable, as stated by Shadabfar and Cheng [15]. The stages are explained as follows:

1. From the historical data, calculate the mean (α_{ij}^k) and standard deviation (β_{ij}^k) of the daily return for every 20 days for each stock. Here, i denotes the ith stock index ($i \in \{A, B, C, D, E, F, G\}$ since seven indices are used in this study), j denotes the jth stock within an index ($j \in \{1, 2, \ldots, 30\}$ since each index consists of 30 stocks), and k is the kth month ($k \in \{1, 2, \ldots, 60\}$ since the data was taken over 60 months or 5 years). For example: α_{Aj}^1 means the average daily return in the 1st month (remember that each month is assumed to have 20 days) for stock j in stock index A.

2. Perform the following steps to form seven portfolios:

   (a) For $k \in \{1, 2, \ldots, 60\}$, generate 20 random numbers each (these random variables represent the daily return of a stock) from a normal distribution $N(\alpha_{ij}^k, \beta_{ij}^k)$. Thus, each stock has 1200 random numbers representing daily returns over five years.

   (b) Calculate the monthly return for each stock by summing the daily returns per month (each month has 20 days), so for each stock, there are 60 monthly return values.

   (c) Calculate the expected monthly return for each stock by averaging the monthly returns from step b).

   (d) Solve the portfolio optimization model to obtain the return and risk simultaneously. (This is the same as in the deterministic portfolio formation, meaning the same risk-aversion coefficient is used.) The output required at this stage is the return and risk of the seven portfolios formed.

3. Rank the seven portfolios based on the return obtained in end of step 2, so that we have a rank of 1 to 7 for the existing portfolio based on the return.

4. Rank the seven portfolios based on the risk obtained in end of step 2, so that we have a rank of 1 to 7 for the existing portfolio based on the risk.

5. Repeat steps 2 through 4 a total of N times (in this study, we choose $N = 10,000$). As a result, for each portfolio, the number of rankings from 1 to 7 is obtained, for both return and risk. Let these be denoted as n_{il} and m_{il}, where $l = 1, 2, \ldots, 7$. For instance, n_{F1} refers to the number of rank 1 outcomes for the return of Portfolio F over N simulation runs.

6. Calculate the respective probabilities of the rankings obtained in step 5, for both return and risk, using the formula

$$p_{il} = \frac{n_{il}}{N} \quad \text{and} \quad q_{il} = \frac{m_{il}}{N}.$$

We have visualize these results in the form of bar and pie charts. After that, the analysis is continued to see the convergence of the Monte Carlo simulation by creating a graph of the probability of return and risk for a specific portfolio and a specific ranking against the number of simulations N. In this study, Portfolio C for ranking five was chosen, and we ran N from 1 to 10,000.

Finally, the discussion is closed by evaluating the robustness of the proposed algorithm. Robustness evaluation is essential in Monte Carlo simulation because the accuracy of the results critically depends on the assumed probabilistic characteristics of the input parameters, which are often estimated with uncertainty. By evaluating the stability of the simulation outcomes under small perturbations in the input parameters, robustness evaluation ensures that the conclusions drawn from Monte Carlo simulations are not overly sensitive to estimation errors and remain reliable for decision-making. For this purpose, a normal distribution with an absolute coefficient of variation (ACV) is employed. The ACV is adopted because the mean returns of assets may take negative values. The ACV is defined as follows [2]:

$$ACV = \frac{\sigma}{|\mu|},$$

where, μ and σ are mean and standard deviation of the data, respectively. The mean of the normal distribution is assumed to be a scaled version of the mean of the initial random variable, that is

$$\mu_{f_k} = \phi\mu_k,$$

where μ_k denotes the mean of the initial random variable; ϕ represents the noise contribution factor; and μ_{f_k} is the mean of the normal distribution associated with the kth perturbation ($k = 1, 2, \ldots, N_{\text{pert}}$) and N_{pert} is the number of simulations used for robustness evaluation. The proposed algorithm is then applied to compute the optimal probabilistic solutions for the seven portfolios under noise inputs. Specifically, the probability ranking procedure for the seven portfolios is rerun using perturbed inputs, where each random number originally generated from the normal distribution $N(\alpha_{ij}^k, \beta_{ij}^k)$ is augmented by an additive noise term drawn from the normal distribution $N(\mu_{f_k}, \sigma_{f_k})$ with $\sigma_{f_k} = ACV \times |\mu_{f_k}|$.

4 Results and discussion

4.1 Portfolio analysis

To predict the direction of market movement, we used monthly price data for the LQ45 index from July 31, 2000, to July 31, 2025, covering 300 months. The logistic regression model predicted a probability of 0.9219 with an accuracy of 82.7%. Subsequently, this predicted probability was used as the risk-aversion coefficient α. Since the value is closer to 1, the market is expected to be bullish, leading investors to prioritize returns over risk in portfolio construction.

Seven portfolios were constructed using stocks from various indexes: IDX30, IDXQ30, IDXV30, IDXG30, ECONOMIC30, JII, and IDXSHAGROW. Each index consists of 30 stocks. The data consisted of adjusted closing prices collected over five years from July 29, 2020, to July 31, 2025, totalling 1201 data points, assuming 20 trading days per month. Stocks of companies that had their Initial Public Offering (IPO) after July 29, 2020, were excluded to ensure a complete dataset of 1201 data points. The specific stocks excluded were GOTO, MBMA, ADMR, AVIA, CMRY, TPAG, AMMN, and PGEO.

Five key parameters were established for portfolio formation and simulation:

1. The cardinality constraint was set at $K = 8$,
2. The transaction cost of every stock is the same, that is, 0.0015,
3. The number L of liquidity constraint is 0.05,
4. The lower and upper bounds for the quantity constraint for each stock i were $l_i = 0.025$ and $u_i = 0.5$, respectively, and
5. The number of simulations was 10,000.

The time consumption to run this simulation for seven portfolios is around 4.7 hours. The convergence rate of the standard Monte Carlo estimator was evaluated by analyzing the decay of the estimation error with respect to the sample size. This error decreases proportionally to number of simulations N, that is, $O(N^{-1/2})$ [5].

Table 1 compares the returns and risks of the equal-weighted $\frac{1}{n}$ portfolios (which are constructed by allocating the same weight to all stocks that fall within the index) with an alternative solution modeled in this paper. The findings suggest that the equal-weighted $\frac{1}{n}$ portfolios display strikingly low risk. This conclusion correlates to a great diversification, since the exposure of each asset is fixed and so the risk is minimized. But the significantly low risk of the equal-weighted

$\frac{1}{n}$ approach comes with small returns. This kind of return profile may not appeal to investors looking for higher returns, particularly in cases where moderate risk tolerance is considered appropriate. In contrast, the portfolios constructed using the proposed model demonstrate a significant improvement in returns across all portfolio configurations. Although this improvement in return is accompanied by an increase in risk, the trade-off remains reasonable and consistent with portfolio theory, which posits that higher expected returns generally require higher risk exposure. Furthermore, the proposed portfolios also have one additional benefit: There are just eight stocks in each portfolio.

The smaller number of assets implies lower implementation complexity and transaction costs than the equal-weighted $\frac{1}{n}$ strategy that requires investors to allocate capital over the index of stocks they own.

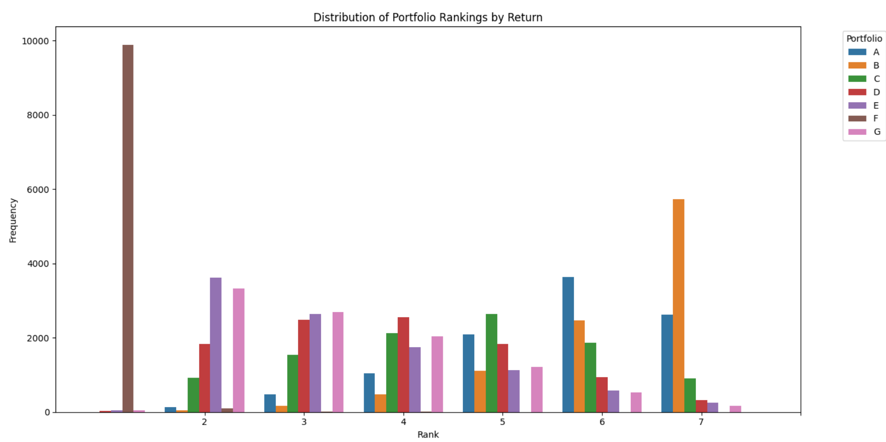

Figure 2: Distribution of return rankings (1-7) for all portfolios. Rank 1 indicates the highest return, while Rank 7 indicates the lowest return.

The overall results indicate that the proposed model strikes a better risk, return, practical feasibility alignment, thus an attractive substitute for those investors seeking improved performance without an unnecessarily difficult diversification process. From this table we can also infer that Portfolio F offers the highest return at 10.67% compared to other portfolios and the highest risk of 7.00%. However, such an investor might have additional considerations to make investment decisions following the approach which was discussed in Subsection 3.2.

Figures 2 and 3 illustrate the distribution of return and risk ranks (1 to 7) for all portfolios. From these two figures, we can see the frequency of a ranking of a portfolio for 10,000 simulations. For example, Portfolio F is the best scenario for returns and the worst for risk because the frequency of rank 1 for returns of this portfolio is the highest compared to other portfolios. Similarly, the frequency of rank 7 (which means the highest risk) is also the highest. Tables 2

Table 1: Summary of Stock Portfolios Performance and Composition

Portfolio A		Portfolio B		Portfolio C		Portfolio D		Portfolio E		Portfolio F		Portfolio G	
Stock Code	Prop.	Stock Code	Prop.	Stock Code	Prop.	Stock Code	Prop.	Stock Code	Prop.	Stock Code	Prop.	Stock Code	Prop.
ADRO	46.9%	ADRO	50%	ADRO	2.5%	BRIS	2.5%	ANTM	2.5%	ADRO	2.5%	ADRO	2.5%
AKRA	2.5%	AKRA	2.5%	AUTO	2.5%	ENRG	50%	BRIS	2.5%	AKRA	2.5%	BRIS	2.5%
AMRT	2.5%	AMRT	2.5%	ELSA	2.5%	ESSA	2.5%	BRMS	34.9%	ANTM	22.2%	DSNG	2.5%
ANTM	18.6%	BFIN	2.5%	ENRG	50%	ISAT	2.5%	BRPT	2.5%	BRMS	2.5%	ISAT	2.5%
BRPT	2.5%	BMRI	4.7%	HRUM	2.5%	ITMG	2.5%	ESSA	2.5%	MEDC	2.5%	ITMG	2.5%
ISAT	21.8%	BNGA	2.5%	ITMG	34.9%	MAPA	2.5%	FILM	50%	PANI	50%	MEDC	2.5%
MEDC	2.5%	BRIS	32.6%	MEDC	2.5%	MIDI	2.5%	MAPA	2.5%	PGAS	2.5%	RAJA	50%
PTBA	2.5%	PTBA	2.5%	SRTG	2.5%	SSIA	34.9%	PTBA	2.5%	PTRO	15.2%	SSIA	34.9%
Return: 3.42%		Return: 3.40%		Return: 4.56%		Return: 4.86%		Return: 5.55%		Return: 10.67%		Return: 5.60%	
Risk: 3.60%		Risk: 3.71%		Risk: 4.83%		Risk: 4.80%		Risk: 6.13%		Risk: 7.00%		Risk: 5.08%	
Equal-weighted $\frac{1}{n}$													
Portfolio A		**Portfolio B**		**Portfolio C**		**Portfolio D**		**Portfolio E**		**Portfolio F**		**Portfolio G**	
Return: 1.27%		Return: 1.01%		Return: 1.35%		Return: 1.88%		Return: 1.60%		Return: 2.37%		Return: 2.06%	
Risk: 0.41%		Risk: 0.37%		Risk: 0.46%		Risk: 0.38%		Risk: 0.45%		Risk: 0.44%		Risk: 0.44%	

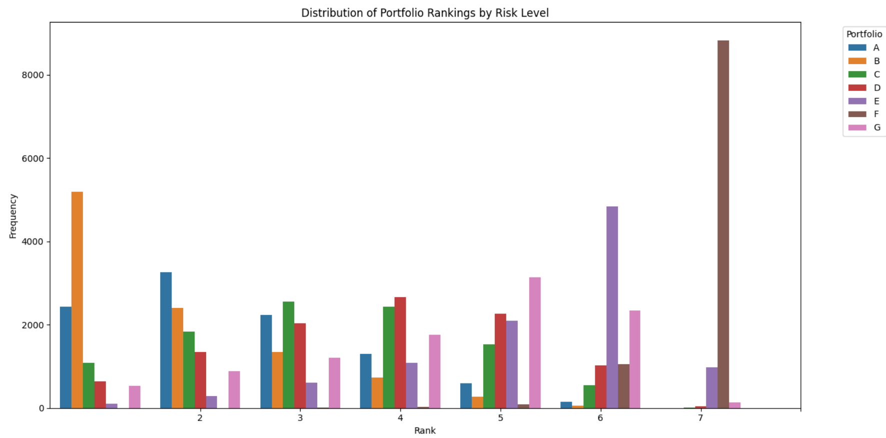

Figure 3: Distribution of risk rankings (1-7) for all portfolios. Rank 1 indicates the lowest risk, while Rank 7 indicates the highest risk.

and 3 show the estimated probabilities of return and risk, respectively. The results indicate that Portfolio F offers the highest probability (0.9884) of the best return but also presents the worst risk to investors with probability 0.8822. Figure 4 is a pie chart that visualizes the estimated probability of return and risk for each portfolio. From the two tables and the figure, we have more information about the seven portfolios. For instance, although the probability of achieving relatively low risk for Portfolio B is the highest (51.98%), but the probability of achieving the return is the lowest (0). Such information is also provide for all other portfolio ranks.

Table 2: Estimated probabilities of return for all portfolios, ranked from 1 (highest return) to 7 (lowest return).

Portfolio	**Return Rank**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
A	0	0.0131	0.0474	0.1051	0.2094	0.3634	0.2616
B	0	0.0052	0.0173	0.0472	0.1104	0.247	0.5729
C	0	0.0919	0.1537	0.2129	0.2639	0.1863	0.0913
D	0.0022	0.1839	0.2489	0.2562	0.1827	0.0934	0.0327
E	0.0051	0.3622	0.2633	0.1739	0.1126	0.058	0.0249
F	0.9884	0.0103	0.0009	0.0004	0	0	0
G	0.0043	0.3334	0.2685	0.2043	0.121	0.0519	0.0166

Table 3: Estimated probabilities of risk for all portfolios, ranked from 1 (lowest risk) to 7 (highest risk).

Portfolio	**Risk Rank**						
	1	**2**	**3**	**4**	**5**	**6**	**7**
A	0.2439	0.3264	0.2241	0.1305	0.06	0.0148	0.0003
B	0.5198	0.2398	0.1345	0.0732	0.0277	0.005	0
C	0.1089	0.1828	0.2558	0.2431	0.153	0.0549	0.0015
D	0.0635	0.1343	0.2032	0.2665	0.2257	0.1026	0.0042
E	0.0105	0.0285	0.0612	0.1085	0.2102	0.4834	0.0977
F	0.0001	0.0002	0.0004	0.0021	0.0093	0.1057	0.8822
G	0.0533	0.088	0.1208	0.1761	0.3141	0.2336	0.0141

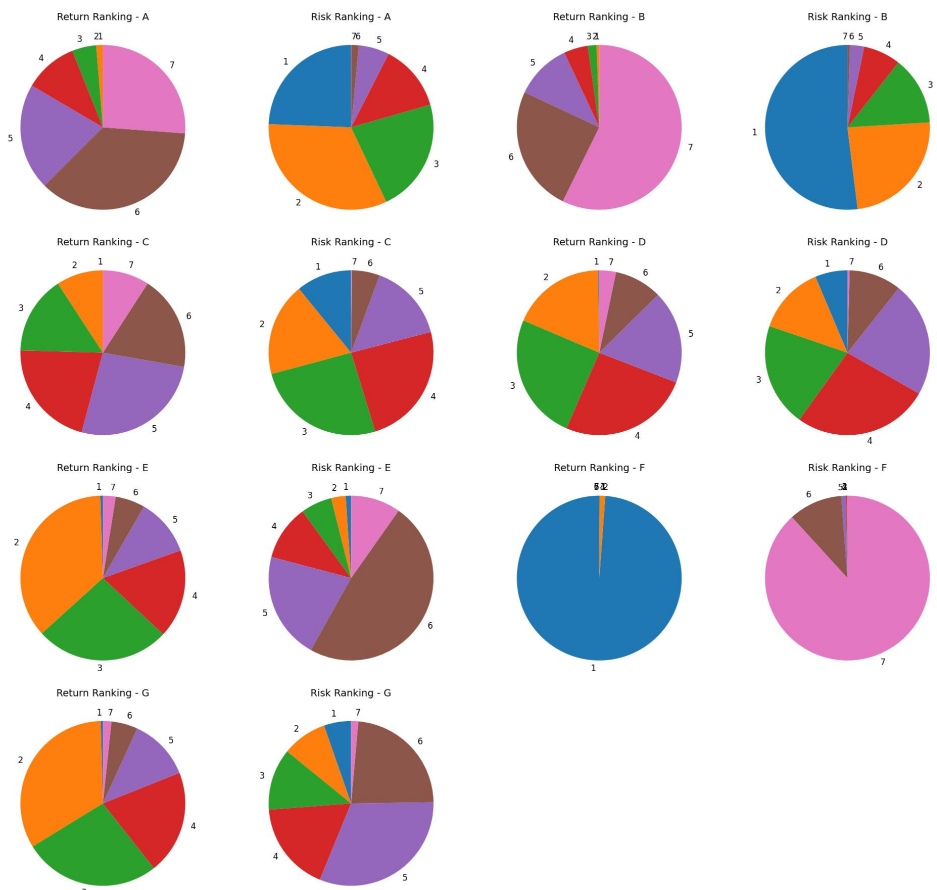

Figure 4: Estimated probability of return and risk for each portfolio, illustrating the distribution of return and risk rankings from 1 to 7.

4.2 Convergence and robustness evaluation

To examine the convergence of the Monte Carlo method in this portfolio optimization problem's probabilistic approach, Figures 5 and 6 show the graphs of return and risk probability values, respectively, against the number of random samples (simulations) for Portfolio C with a rank of 5. The blue curve represents the probability values for Portfolio C's return and risk for 1 to 10,000 simulations, while the dashed red line indicates the probability values for 10,000 simulations. At small random sample sizes, there are significant fluctuations in probability values. However, as the random sample size increases, these fluctuations decrease. Finally, after using approximately 6,000 random samples, the probabilities converge, and the fluctuations become increasingly small and can be ignored. Based on this result, all 10,000 simulations in the previous experiment are considered sufficient.

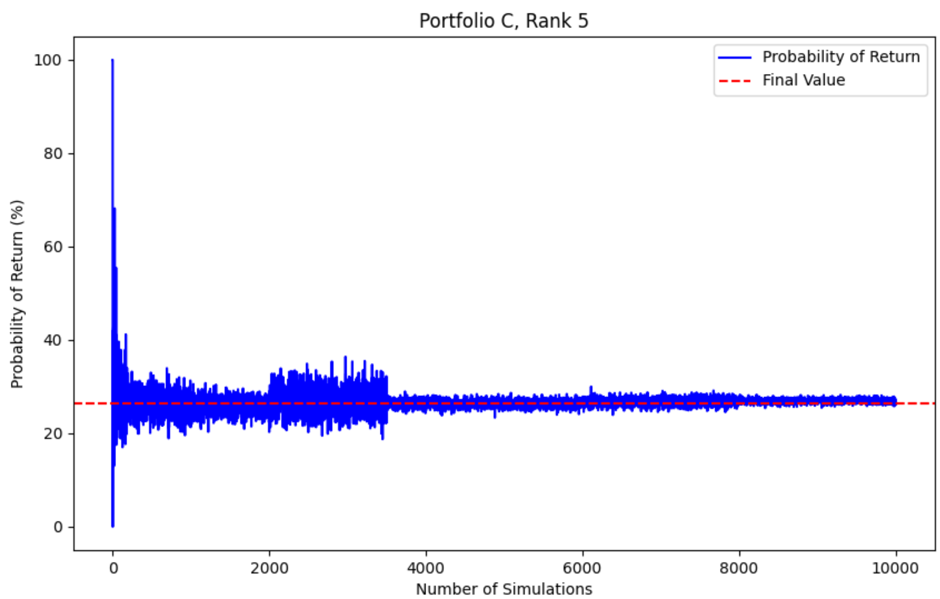

Figure 5: Convergence of the estimated probability of return against the number of simulations for portfolio C (Rank 5).

Besides convergence, the next important thing needed in Monte Carlo simulation is robustness evaluation. This evaluation is essential in Monte Carlo simulation because the accuracy of the results critically depends on the assumed probabilistic characteristics of the input parameters, which are often estimated with uncertainty. By evaluating the stability of the simulation outcomes under small perturbations in the input parameters, robustness analysis ensures that

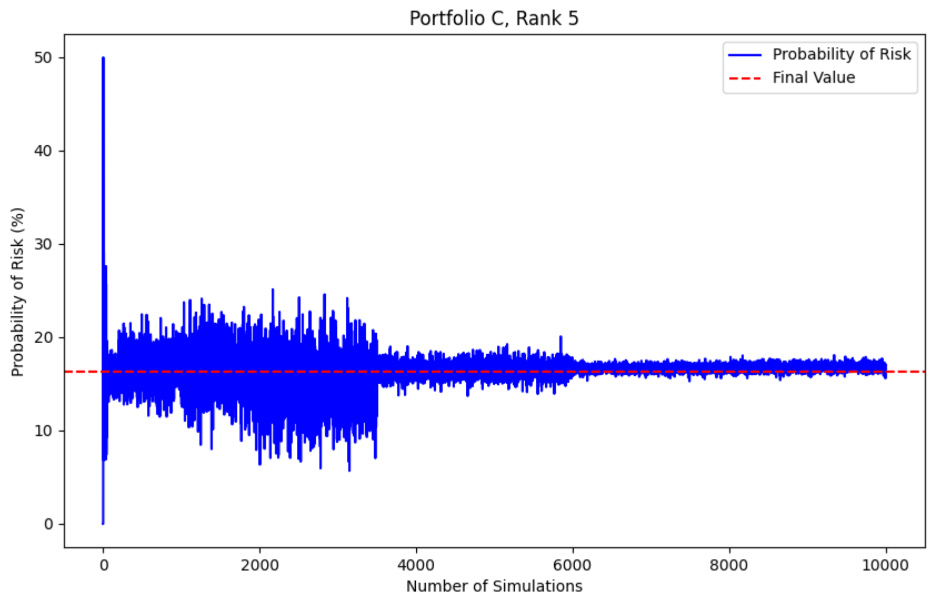

Figure 6: Convergence of the estimated probability of return against the number of simulations for portfolio C (Rank 5).

the conclusions drawn from Monte Carlo simulations are reliable for decision-making. For robustness evaluation, we use $N_{\text{pert}} = 100$. Figure 7 illustrates the box-plots of the return and risk probabilities for Portfolio C with rank 5, respectively. From Figure 7(a), the median value remains stable at approximately 0.261. The interquartile range is relatively narrow, indicating that more than 50% of the simulated probabilities are concentrated within a small interval. This limited dispersion suggests that the estimated return probability exhibits low sensitivity to input perturbations. The presence of only two outliers further indicates that extreme return realizations are rare and do not materially influence the overall distribution. Similarly in Figure 7(b), the median is approximately 0.158, and the interquartile range remains small compared to the overall range of the data, implying that the variability of risk probability is well controlled under input uncertainty. Although several upper outliers are observed, their occurrence is infrequent and does not reflect systematic instability in the risk estimation. Thus, our proposed algorithm is quite robust in estimating the probability of return and risk.

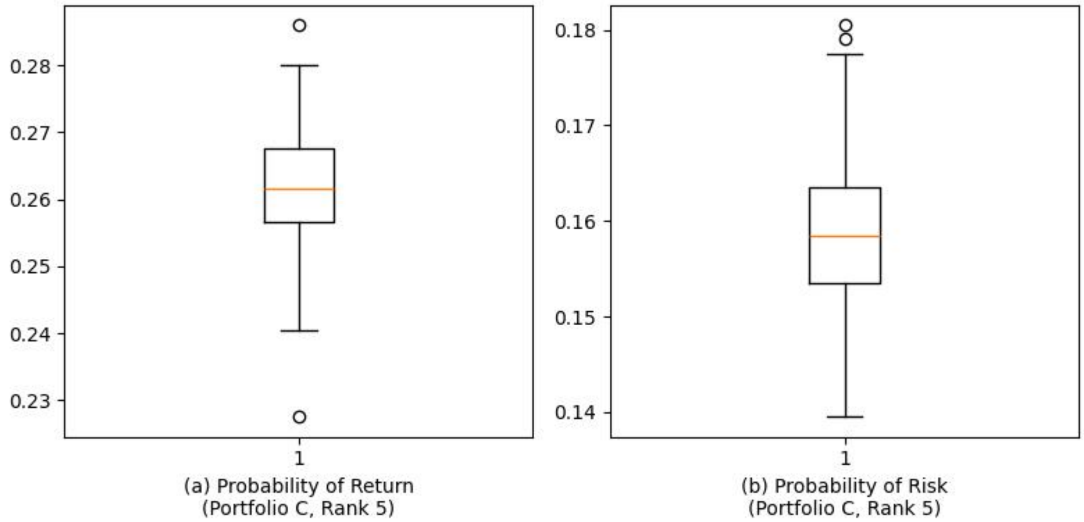

Figure 7: Box-plots of the simulated probabilities for (a) return and (b) risk of Portfolio C with rank 5 under input perturbations.

5 Conclusions

The algorithm proposed in this paper provides an alternative insight for investors with a probabilistic approach to selecting portfolios, including constraints on the number of assets, the quantity and liquidity of each asset, and also transaction cost within the portfolio. Based on five years of data collected between July 29, 2020, and July 31, 2025, with a total of 1,201 data points, investors can not only estimate their expected returns and risks, but also view the probability of an optimal portfolio that is appropriate by considering market trend conditions through the risk-aversion coefficient. A Monte Carlo simulation was used 10,000 times to determine these probabilities, as the convergence graph showed that the simulation had already converged after 6,000 iterations. Eventually, the absence of prominent distributional shifts and the small number of outliers confirm that the proposed portfolio model exhibits satisfactory robustness against small perturbations in the input parameters.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the ”Lembaga Penelitian dan Pengabdian kepada Masyarakat” (LPPM), Parahyangan Catholic University, for this study (grant number III/LPPM/2025-02/74-P). Authors are also grateful to their anonymous referees and editor for their constructive comments.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Anagnostopoulos, K.P. and Mamanis, G. *The mean-variance cardinality constrained portfolio optimization problem: An experimental evaluation of five multiobjective evolutionary algorithms*, Expert Syst. Appl. 38 (2011), 14208–14217.

[2] Bindu, K.H, Mukkamala, R., Dey, N. and Rao, C. R. *Coefficient of variation and machine learning applications*, CRC Press, 2019.

[3] Cesarone, F., Scozzari, A. and Tardella, F. *A new method for mean-variance portfolio optimization with cardinality constraints*, Ann. Oper. Res. 205 (2013), 213–234.

[4] Fang, Y., Lai, K. K. and Wang, S.Y. *Portfolio rebalancing model with transaction costs based on fuzzy decision theory*, Eur. J. Oper. Res. 175(2) (2006), 879–893. https://doi.org/10.1016/j.ejor.2005.05.020

[5] Glasserman, P. *Monte Carlo methods in Financial Engineering*, Springer. (2003), 1–3.

[6] Grant, A., Kwon, O. K. and Satchell, S. *Properties of risk aversion estimated from portfolio weights*, J. Asset Manag. 25 (2024), 427–444.

[7] James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. *An introduction to statistical learning with applications in Python*, Springer. (2023), 138–144.

[8] Ji, R., Lejeune, M.A. and Prasad, S.Y. *Properties, formulations, and algorithms for portfolio optimization using Mean-Gini criteria*, Ann. Oper. Res. 248 (2017), 305–343.

[9] Jiang, Z., Ji, R. and Chang, K. C. *A machine learning integrated portfolio rebalance framework with risk-aversion adjustment*, J. Risk Financial Manag. 13(7) (2020), 155.

[10] Konno, H. and Wijayanayake, A. *Mean-absolute deviation portfolio optimization model under transaction costs*, J. Oper. Res. Soc. Jpn. 42(4) (1999), 422–435.

[11] Konno, H. and Yamazaki, H. *Mean-absolute deviation portfolio optimization model and its applications to Tokyo Stock Market*, Manag. Sci. 37(5) (1991), 519–531.

[12] Lee, J. *A first course in combinatorial optimization*, Cambridge University Press: New York, United States of America. (2004), 177–180.

[13] Markowitz, H. *Portfolio selection*, J. Finance. 7(1) (1952), 77-91.

[14] Mousavi, A. and Michailidis, G. *Cardinality constrained mean-variance portfolios: a penalty decomposition algorithm*, Comput. Optim. Appl. 90 (2025), 631–648.

[15] Shadabfar, M. and Cheng, L. *Probabilistic approach for optimal portfolio selection using a hybrid Monte Carlo simulation and Markowitz model*, Alex. Eng. J. 59 (2020), 3381–3393.

[16] Shadabfar, M. and Wang, Y. *Approximation of the Monte Carlo sampling method for reliability analysis of structures*, Math. Probl. Eng. 2016(1) (2016), 5726565.

[17] Simaan, Y. *Estimation risk in portfolio selection: The mean variance model versus the mean absolute deviation model*, Manag. Sci. 43(10) (1997), 1437–1446.

[18] Sortino, F. A. and Meer, R.v.d. *Downside risk: Capturing what's at stake in investment situations*, J. Portf. Manag. 17(4) (1991), 27–31.

[19] Xiaoxia, H. *Portfolio analysis: from probabilistic to credibilistic and uncertain approaches*, Springer. (2010), 11–60.

[20] Yaman, I. and Dalkılıç, T. E. *A hybrid approach to cardinality constraint portfolio selection problem based on nonlinear neural network and genetic algorithm*, Expert Syst. Appl. 169 (2021), 114517, 294–305.

[21] Yixuan, L. and Zhongfeng, Q. *Mean semi-absolute deviation for uncertain portfolio optimization problem*, J. Uncertain Syst. 6(4) (2021), 299–307.

[22] Yufei, S., Grace, A., Lay, T. K., Guanglu, Z. *Portfolio optimization using a new probabilistic risk measure*, J. Ind. Manag. Optim. 11(4) (2015), 1275–1283.

[23] https://www.idx.co.id/id/data-pasar/data-saham/indeks-saham, last accessed July 24, 2025.