Malaysian Journal of Mathematical Sciences 20(3): 1215–1240 (2026)
https://doi.org/10.47836/mjms.20.3.16

Malaysian Journal of Mathematical Sciences

Journal homepage: https://mjms.upm.edu.my

On Optimizating Portfolio Returns via A Modified Dynamic Programming (DP) Model Based on Bellman's Equation: Identifying Finest Cluster Shape

Ohanuba, F. O.*[1], Ossai, E. O.[1], and Mba, E. M.[1]

[1]*Department of Statistics, University of Nigeria Nsukka, Nigeria*

E-mail: felix.ohanuba@unn.edu.ng
**Corresponding Author*

Received: 30 May 2025
Accepted: 15 December 2025

Abstract

Effective financial management leads to the development of suitable decision plans that aim for optimal results when investing in a competitive stock portfolio. This research adapted and utilized a dynamic programming (DP) model developed by Bellman to address the financial issue. There are challenges in selecting an investment that produces maximum returns; the problem has led to economic crisis among investors in financial markets. Many financial analysts offer investors accurate and unverified investment details, leading to suboptimal or no returns and various investment issues. The objectives of this study are to maximize investor returns and to confirm the findings by employing two statistical validity tests to determine the most suitable test for this study. Two tests (Silhouette and Dunn) have been utilized for result validation. By utilizing Silhouette, the computation became simpler and produced a more reliable and superior result. The method of k-means clustering demonstrated superior statistical evaluation, optimal fit, and improved investment patterns. Compared with earlier studies, the introduction of a system variable in this study yielded the highest return in the initial stage. In the end, the purpose of verified investment reports was achieved, minimize mistakes frequently made by financial managers and stack holders during their investment plans and in choosing portfolios.

Keywords: dynamic programming; portfolio; recursive method; decision making; statistical finance.

1 Introduction

The capital market was described as an institution that has the facility through which an economy could mobilize and allocate the resources, and take informed risk(s) with the aim of transforming savings into productive investment. Investment in stock offers a platform where demands in financial assets meet the supply in the resource-flow facilitation from saver with surplus liquidity to investor(s) in the productive section [15]. The research problem lies with the problem of devising a better way of solving a dynamic programming problem (DPP) involving investment in portfolio and other stocks in the stock market for optimality. Inadequate revenue management during portfolio investment (selection) has caused economic disruptions worldwide. This has led to negative impacts to the worldwide market due to inaccurate data, lack of essential information, and expertise on where and how to invest for optimal returns. Dynamic programming originated from mathematical programming similar to optimization and utilizes mathematical variables according to [8]. When choosing a fund investment, DP refers to the process or strategy of adhering to an optimal decision rule to attain the greatest potential result.

Firstly, it is necessary to define those underlying dynamics according to the composition, constituents, or variables to fashion out the constrains and choice of portfolio [43]. The review specified that the model used requires to be modified to align with the specified goals it was set to achieve. According to some research, DP can be combined with other models to generate better decision outcomes, particularly when combined with the Bellman equation [26]. According to the statement, the primary objective of revenue management in passenger air transport is to sell a consistent and perishable inventory as efficiently as possible within a given time frame. Since the goal was to verify the availability of the products over the designated time period, the company should have named or defined its items in connection with its services, the corporation should have identified its items according to its portfolio [22]. Their finding reveals a maximum sale in air transport. Their research indicates that airlines would be able to respond directly to their consumers, including travel brokers, if they saw an increase in sales. In addition, seating reservations, home delivery, group booking, shopping priority on board, and other choices are recommended. They came to the conclusion that future research might benefit from integrating revenue management techniques with those of other industries. To minimize the overall losses in projects, dynamic programming was employed in the decision-making process to create an efficient method of allocating the limited resources across the group of projects [47].

There are many more uses of the DP that have not yet been investigated. The shortcomings in financial management are caused by natural environments and additional limitations, such as insufficient management skills and an inadequate or unclear network [39]. The Colombian capital market recorded downfall in GDP performance 2010 and 2022, which has prompted it to suffer depth and liquidity lost during the period. The studies of Sankar in 2008 demonstrates how mathematical programming and game theory provide structured frameworks for improving decision-making by modeling optimization and strategic interactions [27]. GarcÃa and associate in 2017 [13] compare genetic algorithms and tabu search heuristics for index tracking optimization under cardinality constraints, showing how each approach performs in balancing portfolio accuracy and computational efficiency. The authors recently found that optimized portfolio in Colombia performed better than the common investment funds by generating better returns while possessing a balanced risk; they adopted a bi–objective mean-variance model in the analysis of 17 stocks between 2009 to 2024, and concluded that the model is flexible in the economic settings.

In contrast to the dynamic programming method and models in [45], the tabular method through staged technique gave details in the computation of the problem. To help investors get the most out of their chosen portfolios, [23] proposed a selection theorem. The main goal of financial management, planning, and investment is to maximize profit [23]. The study of [45] has not shown the point of optimum and explicit process in each stage of the DP network, but this study has filled the gap by displaying the optimum results in each n-stage. Besides, the degree of validity of the results was demonstrated by this study; the investor could observe the growth of their investment at each level by viewing the expected returns at each implemented stage to view their investment growth. For every investment, the resulting patterns were also generated. In their investment study, [3] stressed the value of having a strong network; however, their findings were not validated.

The technique for judiciously distributing money among several stock portfolios is offered by modern portfolio theory (MPT). Though, it does not give the anticipated results from a particular stock, the theory is crucial during the selection process, and the procedure helps investors minimize risk; the current study described the MPT processes [31]. Weight was apportioned to each portfolio to make decision, and the smallest variances between assets were also used. The study of [28] focused on the management aspect of risk and disaster; they discovered the effective hybrid tool to use and itemized the early warning rules to be implemented by the stakeholders. Grinald [17] used Markov dynamic programming to market value maximization, developing a structured approach that links portfolio decision-making with dynamic optimization techniques and offering insights into how investors can systematically manage risk and return over time.

Some authors that worked on dynamic programming with tests of validity. The first, Harry Markowitz in 1952 propounded Modern Portfolio Theory, revealing how investors can maximize returns for a given level of risk through diversification across assets, a principle that today underpins index funds, risk management strategies, and the way institutional and individual investors construct portfolios [23]. Markowitz in 1957 also pointed out how to comprehend individual securities, the uncertainty of returns, and the correlation among different security returns for portfolio investment creation [32]. Rousseeuw in 1987 used a graph to represent how satisfactory object lies within its assigned cluster with respect to other cluster of similar kinds [40]. The article by Brandt et al. [9] and Yao et al. [46] argue about digital assets and that BRICS equity markets are interconnected through both static and dynamic relationships, stressing how cryptocurrencies and equities influence each other over time.

Klein with coauthors provide an academic synthesis demonstrating how revenue management has advanced through theoretical generalizations and broadened industry applications, seeking to optimize pricing strategies and resource utilization [31]. Bellman and Kalaba in 1957 articulate how dynamic programming provides a rigorous methodological framework for optimizing decision making processes within statistical communication theory, thereby advancing the efficiency and structure of information transmission analysis [12]. Aharon, and coauthors investigate quantile frequency connectedness in emerging AI and IoT token markets, showing how these digital assets exhibit complex interdependencies across different time horizons and risk levels, thereby offering insights into market dynamics and systemic risk in the evolving fintech landscape [7]. Dunn establishes that well separated clusters form the basis for defining optimal fuzzy partitions, thereby advancing clustering methodology by providing a formal framework for partitioning data into overlapping groups with clear separation criteria [20]. The present study applied a tabular technique to solve dynamic programming problem and clearly demonstrated processes in modern portfolio theory. In this study, the DP model was expanded to include two variables. These are the variables for state and system control. The state control variable, which was the second variable, helped achieve the best outcome at stage one.

2 Related Works

In the stock investment business (market), lots of investment means and methods exist. The more common is for an investor to choose a single firm where fundamental analysis can be used to analyze firms [15]. The study is an extension of the stochastic mean–semi variance portfolio selection model, where liquidity is a major measure of portfolio performance in emerging market; the L–R fuzzy number balances return, risk, and liquidity effectively, hence underlining the main trade-offs, and found that in the Colombian market, a higher level of liquidity involves higher risk and lower profitability. Another active way of management is by adopting a strategic diversified portfolio of the firms [13]. Yu and associate in 2016 set up a dynamic programming model to guide environmental investment decision-making in coal mining, offering a structured approach to balance economic efficiency with ecological sustainability [39].

More recently, portfolio selection theory has according to Bueno et al. [15] incorporated approaches that transcend the traditional mean-variance model by Markowitz [23] in order to respond to the complexities of emerging markets. Though, their method has been applied to the Colombian stock market. Among the developments, fuzzy models have highlighted themselves as being particularly relevant in capturing of intrinsic uncertainty in financial markets with the help of credibility theory and use of fuzzy logic. Those techniques allow the representation of ambiguity and subjectivity in the estimation of important financial variables, and helping to make decisions in highly complex and uncertain environments: the method (a multi–objective credibilistic model) was applied in the South American market [15]. Zhang, Li, and Guo [34] in 2018 review portfolio selection within the Markowitz mean-variance framework, highlighting extensions that address uncertainty and practical decision-making challenges in modern finance. The study of Kumar et al. [44] in 2023 introduce a sustainable portfolio selection method using an intuitionistic fuzzy framework. Their approach balances multiple objectives–risk, return, and sustainability-while addressing uncertainty in investor preferences, offering a concise yet flexible model for building portfolios aligned with financial and environmental goals.

The study by Wang et al. [41] introduced a new fuzzy portfolio selection model that combines three-way decision theory and cumulative prospect theory to better reflect the wealth of investors and improve decision-making under uncertainty. Three-way decisions (TWD) in the work of Wang et al. [41] is a decision-making framework that categorizes options into three regions, based on thresholds. The background of their study entails; traditional portfolio models, such as Markowitz mean-variance framework, which focus on maximizing returns and minimizing risk using probabilistic assumptions, real-world investment decisions that often involve uncertainty and imprecision, and cumulative prospect theory (CPT) [16]. The work by Wang et al. [41] also used a novel particle swarm optimization algorithm to solve the complex optimization problems, and the algorithm decomposed the entire swarm into several sub-swarms; the three-way boundary decision model reduces decision risk better than the traditional two-way decisions.

A three-stage methodology is proposed by Aharon et al. [2]: Stage I is ethical screening, Stage II is sustainability scoring and risk evaluation, Stage III is portfolio selection optimization; with the aim to maximize satisfaction (i.e., how well the portfolio does in all three dimensions). Aharon et al. [2] used the electronic constraint method to solve the multiobjective model, and compared the results of their method to the existing approaches to that there is a meaningful set of efficient portfolios among which different investor preferences can select. The works of Ali et al. [6] investigate how renewable energy tokens, a novel class of digital assets, interact with the stock markets of BRICS countries (Brazil, Russia, India, China, South Africa). Using advanced econometric models (TV–VAR), the authors analyzed return spill over, and volatility transmission between these tokens and traditional equity markets. The findings suggest that renewable energy tokens offer diversification benefits, especially during periods of market turbulence, making them attractive for sustainable portfolio construction. Also Ali fellow authors in 2024 further explore the static and dynamic connectedness between GCC equity markets and renewable energy cryptocurrencies, showing how FinTech-driven integration shapes cross-market linkages and offers new insights into diversification and systemic risk [5].

The study by Aharon et al. [2] states that the inclusion of both AI and IoT tokens in investment portfolios is producing increasing sophisticated systems. Given the rapidly growing interest in combining AI technologies with IoT in parallel with the block chain and cryptocurrency markets, it was suggested, that there is necessity to examine AI and IoT with respect to investment portfolios. Shoaib et al. [5] examine how renewable energy tokens (i.e digital or crypto–like assets related to renewable energy) interact with the equity markets of BRICS countries using Time Varying parameter Vector AutoRegressive models (TVP–VAR) to study both static and dynamic connectivity between assets. According to Aharon et al. [**?**], there is substantial connectivity between renewable energy of such tokens and BRICS stock markets; and there is diversification potential of such tokens in investment portfolios with the BRICS markets acting as the transmitters of return spillovers.

Ruzanna and Noriszura [1] analyzed the dependent structure existing between two return series of the Kuala Lumpur Composite Index (KLCI) and the capitalization weighted index (EMAS) for a period of eight years. The results show that the Student's t copula is suitable to represent the dependence structure of KLCI–EMAS pair. Their finding revealed that there is bit strong dependence at upper and lower tails of the two series (KLCI–EMAS), both experienced concurrent shocks. The study of Supandi et al. [36] used portfolio optimization to solve a sensitivity problem in a mean-variance model to its input (i.e., mean vector and covariance matrix of returns); the findings produced an outcome that was used to construct superior portfolios in comparison to the mean-variance portfolio and robust portfolio optimization.

The study by Ali et al. [5] use the QVAR model to examine the nexus between NFTs and equity markets, revealing significant connectedness at extreme quartiles. Their portfolio analysis highlights that including NFTs provides diversification advantages, with higher hedge ratios between tokens and stock markets but the implication of the study is that, past connectedness does not necessarily predict future behavior, especially in fast-evolving markets (crypto, tokens, etc). Ali, Al-Nassar, and Naveed in 2024 investigate static and dynamic relationships between digital assets and BRICS equity markets, revealing interconnected patterns that enhance understanding of cross-market linkages and systemic risk in global finance [4]. Aharon and associate examine quantile frequency connectedness in emerging AI and IoT token markets, highlighting complex interdependencies across risk levels and time horizons that deepen understanding of systemic risk and market dynamics in digital finance [2]. They examined both static (average) and dynamic (time-varying) connectedness, under extreme market conditions, and then assessed the implications for portfolio construction and hedging using DCC-GARCH models, the study discovered that the connectedness among NFTs and BRICS equities is more pronounced in extreme tails than at median or mean levels, meaning that under stress, the coupling is stronger.

The current study seeks to evaluate the conclusions using two validity tests and guarantees global optimality in investors' returns. In this investigation, the test that best fit the updated model was compared and selected. In addition to producing a more reliable and confirmed return, the Silhouette findings decreased processing complexity. The investment pattern was better, statistically evaluated by the k-means clustering, and the best patterns in the clusters were obtained.

3 Materials and Methods

The original Bellman model before adjustment is

$$f_k(S_k) = \max\{g_k\ (U_k) + f_{k+1}(S_{k+1})\}, \quad 0 \leq U_k \leq S_k.$$

It is the reverse dynamic programming equation. The parameters are:

i) S = Total Investment.

ii) n = Item number of the portfolio.

iii) U_k = investment assigned to item k (called decision variable).

iv) $g_k\ (U_k)$= State objective function (i.e., return of).

v) S_k = State variable (i.e. investment of item k to item n).

vi) $f_k(S_k)$ = Maximized return of S_k.

This study introduces the controlled variable, x_k, and the state-controlled variable x_{k+1} into the model increased the number of parameters to 8 instead of the initial number of 6, and the model transforms to (1). Equation (27) is the stage by stage application of the model. The process uses the principle of optimality, which is the specific mechanism by which it improved the investment patterns and minimized mistakes made by financial managers and stakeholders. The underlying principle enables us to derive function equations for sequential processes. From the principle, an optimal policy $(x_1, x_2, ..., x_n)$ has the property that whatever the initial state and decision are, the remaining decisions $(x_2, x_3, ..., x_n)$ must constitute an optimal policy for the n_1 stage process, starting with the stage U_1 which results from the first decision, x_1. Two conditions that be met in order to apply the principle of optimality: Separability of the objective function and state separability property. The Separability of the objective function implies that the objective function can be expressed as the sum of single variable functions (i.e.$f_1(x_1, x_2, ..., x_n) = f_1(x_1), f_2(x_2), ..., f_n(x_n)$). The separability property means, given the current state, an optimal policy for the remaining stages is independent of the policy adopted in the previous stages. The introduced variable was possible due to the optimality principle and was carried out to achieve the optimum results of the state variable, as shown in (1).

This study utilized the tabular method to address an investment issue involving a firm's acquisition of six stock portfolio. The issue illustrates relationships linking investment with its returns. The cooperation invested a total of $6000, and the investments are measured in units. The vital conditions of choosing were itemized in the modern portfolio theory of Markowitz. The computation of the solution adheres to [45] iterative and recursive dynamic programming principle. The following model explains how model of Bellman was adjusted to fit its methodology, and permit an operation of optimality in the first stage. In this study, the additional variables helped achieve a worldwide optimal investment return in the equation (1):

$$f_k(S_k) = max\ \{g_k(U_k) + f_{k+1}(S_{k+1})\} + x_k + x_{k+1} \tag{1}$$

where x_{k+1} and x_k lie between 0 and 10; U_k lies between 0 and S_k; $k = n, n-1, ..., 1$ and $n > k$ where S_k = state variables, U_k = decision variable, S_{k+1} =$S_k - U_k$ = state dynamic expression, n = number of portfolios, k = number of investments, $g_k(U_k)$ = dynamic expression for the decision variable, $f_k(S_k)$ = optimum results of state variable. When $S_{k+1} = S_k - U_k$ and $f_{k+1}(S_{k+1})$ are zero, then $f_{k+1}(S_{k+1}) = \max\{g_k\ (U_k)\} + x_k$ in our previous work is recovered. The evolution of the adjusted model of DPP model is shown below: Let x_j represent a vector, j = levels and n number of observations. One way to express the broad dynamic model is as show in the equation (2):

$$\max\ z = \sum_{j=1}^{n} f_j(x_j) \quad \text{subject to}$$

$$\left\{ \begin{array}{l} \sum_{j=1}^{n} a_j x_j \leq b, \\ x_j \geq 0 \quad \forall\ j = 1, \ldots, n\ \forall\ x_j \end{array} \right\} \tag{2}$$

In the computation, these variables, a_j and b in (2) are integers and the recursive process for solving problems began with nth stage (which started with k = 4 and stopped with k = 1) and could start from the first stage during the maximization of the variable, z from (1). The maximization is defined in the equation (3) as

$$z^* = \max_{\{x_1,\ldots,x_n\}} \sum_{j=1}^{n} f_j(x_j) \tag{3}$$

In this model, x_j represents the allocation for item j.

The maximum of (3) for the variable x_j must satisfy (4),

$$\sum_{j=1}^{n} a_j x_j \leq b. \tag{4}$$

Firstly, x_j is selected in the maximum return in the recursive process with computation of the equation as shown in the equation (5):

$$\max_{x_1,\ldots,x_{n-1}} \left(\sum_{j=1}^{n} f_j x_j \right) + x_k + x_{k+1} = f_n x_n + \max_{x_1,\ldots,x_{n-1}} \sum_{j=1}^{n-1} f_j(x_j) \tag{5}$$

The optimum result,$f_n(x_n)$ does not depend on $x_1, ..., x_{n-1}$ to be satisfied. Therefore,

$$(W_{n-1}\ (b - a_n x_n) = \max_{x_1,\ldots,x_{n-1}} \sum_{j=1}^{n-1} f_j(x_j) + x_k + x_{k+1}. \tag{6}$$

Computing $W_{n-1}\ (b - a_n)$ for each permissible variable, x_n, in (6) reduces to equation (7):

$$z^* = \max_{x_n}[f_n(x_n) + (W_{n-1}\ (b - a_n)] + x_k + x_{k+1}. \tag{7}$$

The variable, x_n takes large integer from the series, $0, 1, ..., \frac{b}{a_n}$ [29] gave details on the model evolution from the first stage to the highest stage. The study of [18] also described the model in its transition step without recursive process. The augmentation ideal of this study was shown when the value of the control variable was increased to 10 in case 2 of the tabular technique. [21] in their study focused on the attributes associated with the investment items.

3.1 Implementation pseudo-code (compact)

To integrate clusters into Bellman DP as state label, estimate cluster transition matrix from rolling windows.

For each rolling window: compute features (mean-ret, vol, skew, mean-corr, liquidity, tail) standardize features pca = PCA (n-components -> keep 90%variance) X = pca.transform(features) for k in 2 ... 10: run K Means (nth clusters=k, init='k-means++', nth_init $= 30$) compute silhouette, CH, DB choose candidate k based on indices + economic tests assign cluster labels C-t to assets estimate $P(C_{t+1}|C_t)$ and cluster-specific return stats solve modified Bellman DP with state(W_t, C_t) simulate forward to next re-balancing point Rolling window: 120 trading days (quarterly) and test 60 & 252 as sensitivity. Re-balancing frequency: monthly (test weekly/quarterly sensitivity). K search: k=2 ...8 if N small (<30 assets), otherwise up to 10. n_init $= 30$, max_(iter) $= 300$, tol $=$ 1e $- 4$. PCA variance retained: 90% report exact number of components.

3.2 Methodological novelty:

1. Prior portfolio studies either: (i) use static clustering for stratification, (ii) use regime models (Markov switching) without explicit clustering, or (iii) apply DP but on smooth parametric state spaces.

2. Contribution of this study: explicit, data-driven clustering (K-means) to define discrete states, combined with a Bellman recursion conditional on cluster label and cluster transition probabilities. It also connects the hybrid ML + stochastic control novelty. In essence, DP breaks a complex optimization problem into smaller sub-problems, solves each recursively, maintaining the principles of optimality cautiously, and combines them to achieve a global optimum.

3.2.1 Markowitz's investment theorem

Harry Markowitz propounded an investment theory in the 1950s to assist economists and investors identify risk level during investment. Richard Bellman in 1950s developed Dynamic Programming (DP) as a mathematical optimization framework for solving multi-stage decision problems as documented in [23]. The principle underlying DP is the Principle of Optimality, which has been stated and explained in the section on methodology, an optimal policy has the property that, whatever the initial state and decision, and the remaining decisions must constitute an optimal policy with respect to the state resulting from the first decision. In dynamic market behavior, Bellman recursive structure allows re-optimization at each stage as new market information arrives. The study of [24] propose the theory of portfolio selection in a manner that averts risk. In this study, MPT proposes investing on varieties of assets with low correlation values. These listed studies were used in statement (articulation) of portfolio selection. Tobin expands the theory of portfolio selection by providing a structured framework that integrates risk-return trade–offs with investor preferences, offering a foundational basis for modern portfolio optimization under uncertainty [38]. Sharpe in 1970 examines portfolio theory by integrating it with capital market analysis, providing a systematic framework that explains how risk, return, and market equilibrium interact to guide optimal investment strategies [35]. Board, Sutcliffe, and Ziemba in 2009 revisit the Markowitz mean–variance model of portfolio selection, emphasizing its enduring role as a foundational framework for balancing risk and return in modern investment decision-making

[21]. The formula is in equation (8):

$$\begin{aligned} \text{Min} \quad & X'VX \\ \text{s.t.} \quad & X'h = h_p \\ & X'e = 1 \end{aligned} \tag{8}$$

Therefore, $X \geq 0, X'VX$ = variance, Min = minimize, s.t. = subject to, and X = a matrix whose dimension equals number of the proportions an unsure asset, K, V = dispersion matrix, h_p= target result, h = return of the model asset means and e = unit.

3.3 The solutions of selection of portfolio model

The solution for the selected model in the MPT can be recast as follows if there are large investments. This is shown in equation (9):

$$L = (0.5)X'VX - \lambda_1(X'h - h_p) - \lambda_2(X'e - 1) \tag{9}$$

The first order principle in equation (9): is defined as

$$VX = \lambda_1 h + \lambda_2 e$$

This indicates that its variance has a linear relationship the covariance. Put X at the center of the equation (9) to get (10):

$$X = \lambda_1 V^{-1}h + \lambda_2 V^{-1}e = V^{-1}[he]A^{-1} \tag{10}$$

Therefore,

$$A = \begin{bmatrix} \mathbf{h}^{\mathrm{I}}\mathbf{v}^{-1}\mathbf{h} & \mathbf{h}^{\mathrm{I}}\mathbf{v}^{-1}\mathbf{e} \\ \mathbf{h}^{\mathrm{I}}\mathbf{v}^{-1}\mathbf{e} & \mathbf{e}^{\mathrm{I}}\mathbf{v}^{-1}\mathbf{h} \end{bmatrix}$$

Substituting equation (10) in $X'VX$ gives

$$A = \mathbf{v}^{-1}\left[\mathbf{h}_p \quad 1\right] A^{-1} \left[\mathbf{h}_p \quad 1\right]^{\mathrm{I}} \tag{11}$$

The standard deviation of the portfolio is obtained from equation (11) to produce (12):

$$S_p = \left(\left[h_p\ 1\right] A^{-1} \left[h_p\ 1\right]^T\right)^{\frac{1}{2}} = \begin{bmatrix} \dfrac{h_p^2}{A} & \dfrac{h_p}{A} \\ \dfrac{h_p}{A} & \dfrac{1}{A} \end{bmatrix}^{\frac{1}{2}}$$

$$S_p = \left(h_p^2 \cdot \frac{1}{A} + 2 \cdot h_p \cdot \frac{1}{A} + \frac{1}{A}\right)^{\frac{1}{2}} \quad \text{i. without the matrix form} \tag{12}$$

The tabular method's implementation comes next, then the analysis's validity test, and lastly, the pattern recognition. Steps 1 through step 3 are represented by the four primary processes for doing the MPT: gathering the data set, generating the efficient frontier, creating market and portfolio lines, and creating the portfolio optimum. This symbolizes Markowitz point of tangency between the capital market and the efficient frontier in [23]. Validity analysis was incorporated into the entire methodology for this study to guarantee the process completeness and the validity of the results. Further detailed reading of portfolio selection of the markowitz procedure can be found [?] present a robust version of the Markowitz mean-variance portfolio model that addresses uncertainty in the covariance matrix, offering a more resilient approach to portfolio optimization. Zhang and his coresearchers in 2018 provide a comprehensive review of portfolio selection problems within Markowitz's mean-variance framework, synthesizing a wide range of studies to highlight its theoretical foundations, practical applications, and limitations. Their work underscores the framework's enduring relevance in modern investment analysis while also pointing to areas where extensions and refinements have been developed to address evolving financial complexities [48]. In the procedure, when we have securities, r_{it} as an expected result per dollar, d_{it} is the rate of a specific security, i^{th} from N invested securities, then the portfolio discounted and expected portfolio return, R of the investment is in the equation (13):

$$R = \sum_{i=1}^{\infty} \sum_{i=1}^{N} d_{it} r_{it}$$

$$X_i = \sum_{i=1}^{N} X_i \left(\sum_{i=1}^{\infty} d_{ir} r_{ir} \right), \quad X_i \geq 1 \quad \forall i \tag{13}$$

where,

$$R_i = \sum_{i=1}^{\infty} \sum_{i=1}^{N} d_{it} r_{it}$$

is the relative asset for the i^{th} security,

$$R = \sum X_i R_i,$$

where R_i & X_i are independent.

To maximize R, $\sum X_i$ is set to 1, then R reduces to an average weight R_i with X_i (where X_i is a non-negative weight of the i^{th} security). According to [23], several securities can be maximized by selecting the group and setting the sum of the selected group to 1. Two theorems of portfolio selection and volatility, according to [35], state that The joint weights of assets and the expected portfolio return are equal and give rise to (14):

$$R = \sum_i X_i R_i, \quad E(R_p) = \sum_i w_i^2 E(R_i) \tag{14}$$

where R_p = portfolio return from R_i, R_i is the return on i asset, w_i = weight of the portfolioś component asset. For a pair of assets, i and j, the volatility of portfolio is an expression of their correlations, and mathematically gives rise to equation (15):

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_{i\, j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}, \quad \forall\, i = j \tag{15}$$

Obtaining square root of (15), the equation reduces to the standard deviation of the process shown in an Equation (16):

$$\sigma_p = \sqrt{\sum_i w_i^2 \sigma_i^2 + \sum_{i\, j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij}}, \quad \forall\, i = j. \tag{16}$$

Diagram in modern portfolio shows the plot of the target return against standard deviation of the portfolio. Financial market presents a risky system once the curve coincides with the capital market line. i.e. when both lines form a tangent, and so investors can participate. The Figure 1 shows the plot of expected return against the standard deviation in the MPT.

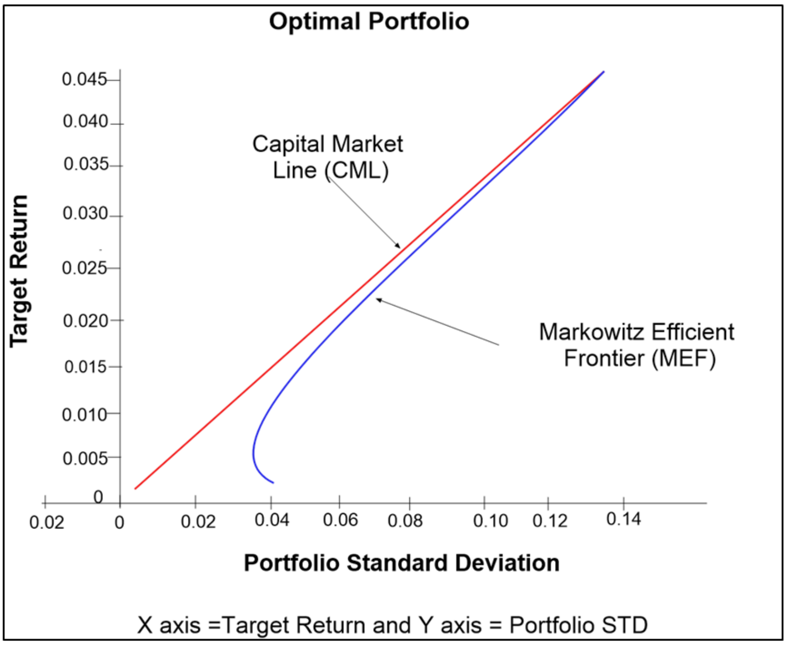

Figure 1: Modern Portfolio Theory Graph

If the risk is not safe for investors to participate, the hyperbola will form the MPTś efficient frontier. But when the assets are free of risk, the straight line in Figure 2 will form the efficient frontier.

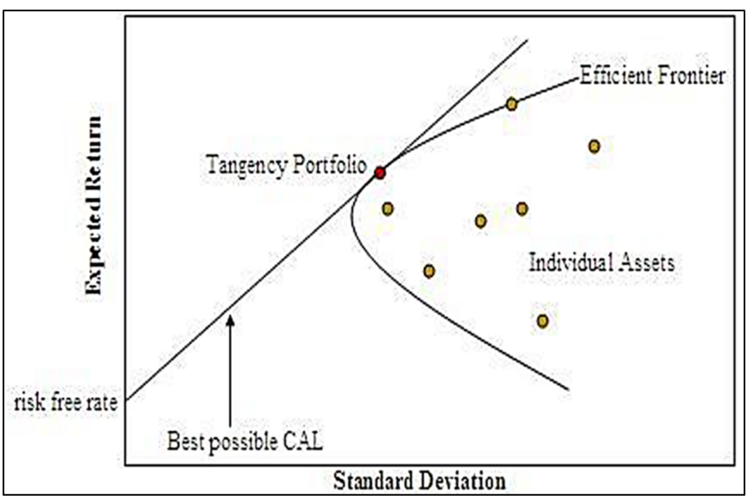

Figure 2: Modern Portfolio Theory Graph

Portfolio of two assets is possible and the result of $E(R_p)$ is in (17):

$$\begin{aligned} E(R_p) &= W_a E(R_a) + W_b E(R_b) \\ &= W_a E(R_a) + (1 - W_a) E(R_b) \end{aligned} \tag{17}$$

and the Portfolio variance is in (18):

$$\sigma_p^2 = W_a^2 \sigma_a^2 + W_b^2 \sigma_b^2 + 2 W_a W_b \sigma_a \sigma_b \rho_{ab} \tag{18}$$

where a = variable from the first portfolio, b = variable from the second portfolio, ab is a combined variable from the correlation. From (18), three assets can be obtained as shown in (19) and (20).

$$E(R_m) = W_a E(R_a) + W_b E(R_b) + W_c E(R_c) \tag{19}$$

Its variance can be deduced from (14), to give (20).

$$\sigma_p^2 = W_a^2 \sigma_a^2 + W_b^2 \sigma_b^2 + W_c^2 \sigma_c^2 + 2 W_a W_b \sigma_a \sigma_b \rho_{ab} + 2 W_a W_c \sigma_a \sigma_c \rho_{ac} + 2 W_b W_c \sigma_b \sigma_c \rho_{bc} \tag{20}$$

where c = variable from the third portfolio. A Similar derivation procedure is applied in obtaining more asset returns.

3.3.1 Interpolating of missed values

In this study, formulas in Excel were used to resolve the missing values prior to the analyses. The Excel formula provided in (21) was used to obtain the interpolation expression, f_x , for this investigation.

$$f_x = B2 + \frac{\$E2 - \$B2}{n+1} \tag{21}$$

where $B2$ = value of the cell before the missing value, $E2$ = value of the cell after the missing value, n = number of all the cells of missing value, $n + 1$= distance separating $B2$ and $E2$ = Excel command to fix empty columns at $n \geq 2$.

3.3.2 Classical metric distance and cluster validity index calculation

Dunn [11] stated that Euclidean distance is an effective distance metric for determining the Silhouette and Dunn indices, respectively, and that each distance reduces to a metric if the listed three conditions are satisfied. The three conditions are (i) condition of symmetry, (ii) condition of triangle inequality, and (iii) condition of non-negative values. If i, j are points with variables, x_i, then $d(x_i, x_j)$ = Euclidean matrix of the points. The distance of the points is obtained from the expression in (22).

$$d(x_i, x_j) = [(x_{i2} - x_{j2}) + \cdots + (x_{iq} - x_{jq})]$$

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{q} (x_{ik} - x_{jk})^2} \tag{22}$$

The distance separating i and j are obtained in (23).

$$d(x_i, x_j) = \sum_{k=1}^{6} |x_{ik} - y_{ik}| \tag{23}$$

The variables, $k = 6$ (this represents the number of portfolios at point x_i and y_i respectively. The important factors that influence how a process is validated, particularly, when complex datasets are involved and can be found in the these three studies; [33], & and [19] cited in the literature. If x_i and y_j are points on c_i and c_j, then the nth variables for Dunn matrix that separates the points, i and $j$are obtained below in (24).

$$D_I = \frac{\min_{1 \leq i < j \leq n} d(c_i, c_j)}{\max_{1 \leq i \leq n} d(c_i)} \tag{24}$$

Therefore, the numerator represents the inter group distances that exist in the clustered points;the denominator is the maximum average that separates them. If N=set of n objects,d = dissimilarity. This study deals with clusters associated with partitions.After reprocessing the datasets, the test's methodology will enable the data to be systematically categorized into homogeneous elements in various groups; this is not done by hand or at random [42]. After pre-processing the datasets, the test's methodology will enable the data to be systematically categorized into correlative elements in various groups; this is not done by hand, and according to [42] freeze-thaw cycles influence fracture mode classification in marble under multi-level cyclic loads, providing insights into material degradation mechanisms and structural reliability in cold environments, and [33]. For a variable where $x_i \in X$, the silhouette index (width) is shown in (25):

$$S_I(C, d) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}. \tag{25}$$

From(25), a(i) is the sum of matrix distance of mean points, and b(i) is the sum of minimum distance of mean points. The mean values between, i and j with additional data in (25) and the closest cluster are denoted by $a(i)$and $b(i)$, respectively. Silhouette Score (average): We used primary guide. Interpretation: greater than 0.5 is good; between 0.25 to 0.5 reasonable; less than 0.25 is weak. The score we obtained using is show in Table 7. Silhouette Test was used to measure or evaluate cluster compactness and separation (i.e., how distinct and cohesive the obtained clusters are), and the procedure in choosing number of clusters k:

1. Perform clustering (e.g., K means) on standardized data for a chosen number of clusters k.
2. For each data point i, Compute $a(i)$ = the average distance between i and all other points in the same cluster (measures cohesion).
3. Compute $b(i)$ = the lowest average distance between i and all points in other clusters (measures separation).
4. Compute the silhouette index for point i, for the formula, see (25). The $s(i)$ ranges from -1 to 1. Near 1 is well clustered (distinct and cohesive). Near 0 is on the boundary between clusters. Near -1 is likely misclassified.
5. Compute the overall silhouette score as the average of all $s(i)$ values.
6. Interpretation: Compare average silhouette scores for different k values. The optimal k is usually where the silhouette score is maximized (indicating best separation and compactness).
7. Visualization: Plot silhouette values for each cluster to see if clusters are uniformly strong or if one cluster is weak (low or negative values).

3.3.3 Cluster validity measures

The act of breaking down a collection of items into smaller groups is known as cluster analysis. Every set is a cluster, with classes in certain clusters corresponding to each other and classes inside others batches that differ. Cluster analysis is an unsupervised learning technique aimed to identify the best way to classify set of data. It also showed that two algorithms might be joined using a combined dissimilarity measure using k-prototype algorithms [14] . The process of clustering involves grouping of data objects into several groups or clusters, with the goal of creating homogeneity of variance. Numerous academic fields, including medicine, biology, mathematics, physics, engineering, data mining, and more, use clustering as a data mining tool.Some clustering techniques developed in [37] include partitioning, density based, hierarchical,and k-means. It is important to note that validity measurements of any system are usually associated with a cluster's inner stable component [10]. Dunn Test. The Dunn Test (also called Dunn post-hoc test) is a non-parametric multiple comparison procedure used after a Kruskal Wallis test indicates significant differences among several groups (clusters). It tells you which specific pairs of clusters differ significantly in their median values.In the context of your portfolio optimization study, it helps assess whether clusters of assets differ significantly in returns, volatility, or other performance measures, thereby validating the economic meaningfulness of the clusters. Step by step procedure to construct Dunn Test:

1. Confirm global significance (i.e., Check if there exist a statistically significant difference among groups).
2. Rank all observations; Combine all data points from all clusters into one dataset; assign ranks from 1 to N(smallest to largest values) and compute the average rank.
3. Compute Dunn's test statistic for each cluster pair, and for each pair of clusters i and j in (26):

$$Z_{ij} = \frac{\left|\overline{R}_i - \overline{R}_j\right|}{\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \tag{26}$$

Where: N = total number of observations, $n_i, n_j =$ number of observations in clusters I and j. $\overline{R}_i,\ \overline{R}_j =$ mean ranks of the two clusters. 4. Compute p-values; compare each Z_{ij} to the standard normal distribution (two-tailed). 5. Adjust for multiple comparisons; we use Bonferroni where we multiply each p-value by 2 (i.e., number of comparisons). 6. Interpret results; If adjusted $p_{ij} < 0.05$: clusters $i$and j differ significantly. The results are summarized in a pairwise significance matrix (Dunn coefficient matrix), where each cell shows the adjusted p-value or significance symbol (e.g., *, **). Significance interpretation: *p < 0.05 (significant)* **p < 0.01 (highly significant)** We obtained 4 highly significant values as shown in the Table 1.

Table 1: Table 1: values of adjusted p-values.

Clusters	**C1**	**C2**	**C3**	**C4**
C1	-	0.021 *	0.342	0.005 **
C2	0.021	-	0.048 *	0.267
C3	0.342	0.048 *	-	0.009 **
C4	0.005 **	0.267	0.009 **	-

* adjusted p-values

3.4 Return and investment problem

In an attempt to determine the ideal portfolio by logical money allocation to optimize investment return, a corporation invested \$60,000 in four stocks. The relationship between each stock's returns measured in unit (i.e., \$10,000 per unit) and investment (\$10,000 per unit) following market research and professional forecasting is displayed in Table 1. The financial issue of a corporation above is described in the study by [3]. For this study,one unit of the investment is valued at \$10,000.

3.4.1 Solutions and result(s)

The recursive solution was obtained using the preceding values, and the maximum number of rows in the range of 0 to 6 determines the optimal unit. At a maximum investment value of 60, it can reach a maximum of 140. It will reach a new maximum value of 150 units, or \$1,500,000 in dollars, when the value of the system control variable, which is 10, is added. Every step is carried out in the same way. This study's methodology involved integration of the two variables (system variable and control variable) into the basic model in (26). The increment is made so that the recursive process limit is not exceeded. This allows for the potential implementation of optimality from stage one and at step one [29]. The introduction of the variables has also improved the outcomes of the investment, as shown in Table 1-5. The Silhouette and Dunn tests were introduced to show how valid this study is and to investigate which of the two tests will work best in the investment problem that relates to the one in this study. Resolving the Corporation's Issue with the Modified Model The process started with a recursive iteration of a backward procedure using the system control and state control variables in the dynamic programming problem by [7]. Four stages completed the whole procedure. The first stage is where the item number of the portfolios, n, is 4, which is the maximum limit; this implies that the limit is exceeded at k = 5. For example, k+1 = 5 is outside the first stage and, statistically, lying within the rejection zone of the process, while n = k = 4 is the maximum in the first stage. To recover the original DPP model, we observe the following: Suppose

$$f_k(S_k) = \max\left\{g_k(U_k) + f_{k+1}(f_{k+1})\right\} + x_k + x_{k+1}, \quad 0 \le U_k \le S_k,$$
$$k = n, n-1, \ldots, 1, \qquad 0 \le x_k, x_{k+1} \le 10.$$

And given that x_{k+1}, x_k are 0, this function, $f_{n+1}(S_{k+1}) + x_k + x_{k+1} = 0$, then, the previous expression, i.e. $f_k(S_k) = \max g_k(U_k)$ is recovered. For this study, the highest level or stage for the solution of the whole process is made up of $n = k = 4$, and is in the first stage. The implication is to identify irrelevancy in the process, and so any solution that fall outside the region, i.e. $x_k \le 10$ is discarded.

Using the table method to solve the modified dynamic programming model

The study made the following assumptions at each stage: $n = 4, x_k = 10, \&\ x_{k+1} = 5$. The reason for this assumption is to specify the boundary at each stage and to obtain the maximum value at each level. The procedure in all the stages follows the principle of optimality which uses recursive/backward process where the next stage comes from the immediate past optimal return/result from the past stage.

The First-Stage

Taking into account $n = 4$, which entails buying the 4th item obtained in (27).

$$f_4(S_4) = \max\{g_4(u_4) + f_5(s_5)\} + x_4 + x_5 \quad 0 \leq u_4 \leq S_4, \tag{27}$$

Equation 27 shows the stage-by-stage application of the model. The function, $f_5(s_5)$ this time fall beyond limit border, so $S_5 = 0$. The condition is applied in the introduction of variable at $k \leq 3$ at $x_4 = 10$.

Therefore,

$$\begin{aligned}
&\text{If} \quad S_4 = 0, \quad f_4(0) = 0 + 0 = 0, \quad g_4(u_4) = 0.\\
&\text{If} \quad S_4 = 1, \quad f_4(1) = 60 + 0 = 60, \quad g_4(u_4) = 60.\\
&\text{If} \quad S_4 = 2, \quad f_4(2) = 80 + 0 = 80, \quad g_4(u_4) = 80.\\
&\text{If} \quad S_4 = 3, \quad f_4(3) = 100 + 0 = 100, \quad g_4(u_4) = 100.\\
&\text{If} \quad S_4 = 4, \quad f_4(4) = 120 + 0 = 120, \quad g_4(u_4) = 120.\\
&\text{If} \quad S_4 = 5, \quad f_4(5) = 130 + 0 = 130, \quad g_4(u_4) = 130.\\
&\text{If} \quad S_4 = 6, \quad f_4(6) = 140 + 0 = 140, \quad g_4(u_4) = 140.
\end{aligned}$$

Table 2: Table 2: Optimum Result and Optimum item in First Stage

$S_4 \backslash U_4$	0	1	2	3	4	5	6	$\max\{g_4(u_4)\}$	$f_4(S_4)$	$U_4^*(U_4, U_5)$
0	0							0	10	$(0,0)$
1	0	60						60	70	$(1,0)$
2	0	60	80					80	90	$(2,0)$
3	0	60	80	100				100	110	$(3,0)$
4	0	60	80	100	120			120	130	$(4,0)$
5	0	60	80	100	120	130		130	140	$(5,0)$
6	0	60	80	100	120	130	140	140	150	$(6,0)$

Due to the introduction of the system variable, the optimal return at this level is \$1,500,000 in the last column and row of the Table 2. Also, its unit was changed.

The Second Stage

Buying $S_3(S_3) = 0, 1, 2, 3, 4, 5, 6$ from 2 portfolios (including the third with forth) that yields the highest returns on the investment divided between the two stocks, assuming $k = 3$. Here in (28), we have

$$f_3(S_3) = \max\{g_3(U_3) + f_4(S_4)\} + x_4, \quad 0 \leq U_3 \leq S_3, \quad x_4 = 10. \tag{28}$$

The outcome is obtained in the Table 3, following a similar working procedure.

Table 3: Table 3: Optimum Result and Optimum Item in the Second Stage

$S_3 \backslash U_3$	0	1	2	3	4	5	6	$\max\{g_3(U_3)+f_4(S_4)\}$	$f_3(S_3)$	$U_3^*(U_3,U_4)$
0	10							10	20	$(0,0)$
1	70	50						70	80	$(0,1)$
2	90	120	120					120	130	$(2,0)$
3	110	160	190	170				190	200	$(2,1)$
4	130	180	210	240	200			240	250	$(3,1)$
5	140	190	230	260	270	210		270	280	$(4,1)$
6	150	200	250	280	290	280	230	290	300	$(4,2)$

The best option in this situation is $(U_3, U_4) = (4, 2)$, which is an investment of 60,000 divided between two stocks, including $20,000$ (on the 4th) and $40,000$ (on the 3rd) stocks, respectively. This level produces an optimum of $\$3,000,000$.

Third Stage

Given $k = 2$, the three stocks (which entail the second, the third, and the fourth) should be invested in $S_2(S_2) = 0,1,2,3,4,5,6$ for realization of optimal result. The case gives the (29).

$$f_2(S_2) = \max\{g_2(U_2) + f_3(S_3)\} + x_3, \quad 0 \le U_2 \le S_2, \quad x_3 = 10 \tag{29}$$

The outcome is obtained in Table 4, following a similar working procedure.

Table 4: Optimum Result and Optimum Items in Third Stage

$S_2 \backslash U_2$	0	1	2	3	4	5	6	$\max\{g_2(U_2)+f_3(S_3)\}$	$f_2(S_2)$	$U^*(U_2,U_3,U_4)$
0	20							20	30	$(0,0,0)$
1	80	40						80	90	$(0,0,1)$
2	130	120	80					130	140	$(0,2,0)$
3	200	170	160	100				200	210	$(0,2,1)$
4	250	240	210	180	110			250	260	$(0,3,1)$
5	280	290	280	230	190	120		280	290	$(1,3,1)$
6	300	320	330	300	240	200	130	330	340	$(2,3,1)$

In this instance, the best product, $(U_2, U_3, U_4) = (2, 3, 1)$, generated by investing 60,000 on the 1st one, then 20,000 on the 2nd one, and $30,000$ on the 3rd stock. So, optimum in this stage gives $\$3,400,000$.

Stage Four

When $k = 1$, investment in the six items, $(S)_1(S_1 = 0,1,2,3,4,5,6)$ yields highest return on investment at this level involving the 1st, 2nd, 3rd, and 4th. The case gives in the (30)

$$f_1(S_1) = \max\{g_1(U_1) + f_2(S_2)\} + x_2, \quad 0 \le U_1 \le S_1, \quad x_2 = 10. \tag{30}$$

Table 5: Table 5: Optimum Result and Optimum Items in Fourth Stage

$S_2 \backslash U_2$	0	1	2	3	4	5	6	$\max\{g_1(U_1)+f_2(S_2)\}$	$f_1(S_1)$	$U_1^*(U_1,U_2,U_3,U_4)$
0	30							30	40	$(0,0,0,0)$
1	90	40						90	100	$(0,0,0,1)$
2	140	130	100					140	150	$(0,0,2,0)$
3	210	180	240	130				240	250	$(0,0,2,1)$
4	260	250	310	210	160			260	270	$(0,0,3,1)$
5	290	300	360	270	250	170		360	370	$(2,1,1,1)$
6	340	330	390	340	300	260	170	390	400	$(2,0,3,1)$

Here, U_1, U_2, U_3, U_4 is equal to $(2, 0, 3, 1)$, and represent the optimum of the sum of investment for the four stocks: $60,000$, which includes 20,000 in the first stock, 0 in the second, $30,000$ (for the 3rd stock), and $10,000$ (for the 4th stock). lastly, we compared to the optimal return of $\$3,400,000$ from the previous analysis, the optimum this time is $\$4,000,000$.

4 The Outcome of The Validity Tests

Libraries needed for the calculation were imported in the first stage, and the second step was finished by importing the resulting data into the Python software. The cluster is recalculated till the optimal results are produced in groups. Step 3 involved computing the data points and evaluating the algorithm to make sure the process was sustained to prevent error. The greatest outcomes among the clusters, $k = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10)$ were likewise shown by the optimum values, and the ideal value was attained at k = 2. Table 5 displays the Silhouette index, which was plotted in Figure 4 and ranged from 0.7 to 0.9. The best group (i.e., 70% to 94%) had the best performance at $k = 2$. The validity test measures perfect values that fall within the accepted region using a special benchmark. The Silhouette results from this investigation showed that the problem analysis was done correctly. With all of its values greater than 0.5, the group $k = 2$. has performed well when compared to the benchmark.

4.1 The statistical comparison of models validity

In comparing the K-means + DP approach to prior literature, we structure the comparison across three axes: A. Methodological novelty Prior portfolio studies: (i) used static clustering for stratification, (ii) used regime models (Markov switching) without explicit clustering, or (iii) applied DP but on smooth parametric state spaces. Your contribution: explicit, data-driven clustering (K-means) to define discrete states, combined with a Bellman recursion conditional on cluster label and cluster transition probabilities. Emphasize the hybrid ML + stochastic control novelty. B. Empirical performance Report where you outperform: higher out-of-sample Sharpe, lower draw down, better downside protection or improved tail metrics. Use the same transaction-cost assumptions as comparable papers (state them). Also show where you do not outperform (e.g., in low volatility, stable markets), honest comparisons strengthen credibility. C. Interpretability and implementation Many advanced models (deep RL, GMM with latent states) have opaque state definitions. K-means gives practically interpretable clusters (list top assets in each cluster, cluster feature profiles); useful for portfolio managers. It is possible to compare the statistics that serve as the foundation for validity models of a metric form , as shown in (23), as shown here. The study in [30] have revealed that the silhouette index (S_I) and dun index D_i are the two main statistics for measuring the validity of a test. From the diagram, the ideal number of groups were compared, and the best one was selected. In order to examine the suitability of two models, (S_I) and D_i, the likelihood ratio test, L_r can used, which is defined in the equation (31):

$$L_r = -2\log\left(\frac{L_{SI}}{L_{DI}}\right) \sim \chi^2_\alpha(w), \tag{31}$$

where w is the number of likelihood parameters. The likelihood, L_{SI} , and L_{D_I} has been represented as the likelihood of the models, S_I and D_I respectively in (27). The models have been defined in equations (17) and (18). This expression, $\chi_(\pm)^2(w)$ indicates probability proportion, L_r tends into chi-square function having a significant level given as $\pm$, and the same distribution has some weight parameter, w, possibly related to the sample size. A decision not to accept during the analysis is taken if $L_r < \chi^2_\pm(w)$, which shows rejection of a claim. The implication of turning down a claim indicates that the model, S_I is a better fit compared to D_I. The parameter, w, is then obtained. The best cluster is found at $k = 2$, where the highest values were observed, as Table 6 illustrates. The steps outlined in subsection 3.1 and 3.2, were followed in the generation of the cluster results in this investigation. for Silhouette and Dunn tests, respectively.

Table 6: Silhouette Output Score and their Validity

State	**2**	**3**	**4**	**5**	**6**	**Validity Score**
Gombe	0.759	0.602	0.548	0.352	0.401	0.6541
Ondo	0.624	0.600	0.679	0.349	0.341	0.4123
Abia	0.736	0.608	0.494	0.533	0.327	0.4465
Enugu	0.755	0.556	0.416	0.330	0.375	0.6395
Adamawa	0.810	0.710	0.379	0.319	0.435	0.8186
Akwa-Ibom	0.923	0.790	0.710	0.779	0.692	0.9597

Table 7: Table 7: Dunn table on the produced clusters for 6 investments

Stage (No Inv.)	$K = 1$	$K = 2$	$K = 3$	$K = 4$	Ave. Validity test for D_i
1	0.511	0.798	0.780	0.799	0.722
2	0.499	0.910	0.840	0.811	0.765
3	0.523	0.710	0.697	0.598	0.632
4	0.439	0.888	0.712	0.711	0.688
5	0.490	0.811	0.721	0.692	0.678
6	0.521	0.893	0.819	0.689	0.731
Range	$(0.0 - 0.5)$	$(0.7 - 0.9)$	$(0.6 - 0.8)$	$(0.5 - 0.8)$	$(0.6 - 0.8)$

The results in Tables 6 and 7, makes it challenging to choose the best test because their outcomes fell within close limits of 0 and 1.It indicates that similarities exist between their experiments' structural patterns.Using Python computer scripts, the researchers computed the plots of the two tests and acquired the plots of the six investments, along with their clusters. After that, the best was chosen using the outcome criteria. The investment plots for DI and SI are displayed in Figures 3 and 4, respectively. In Figure 3, the Silhouette (SI) diagram attained optimality precisely in cluster k = 2 for the 6 items, contrary to the Dunn (DI) diagram in Figure 4. The Silhouette was selected as the best fit for the model by the researchers using the criterion in this section because of the homogeneity across the optimality class.

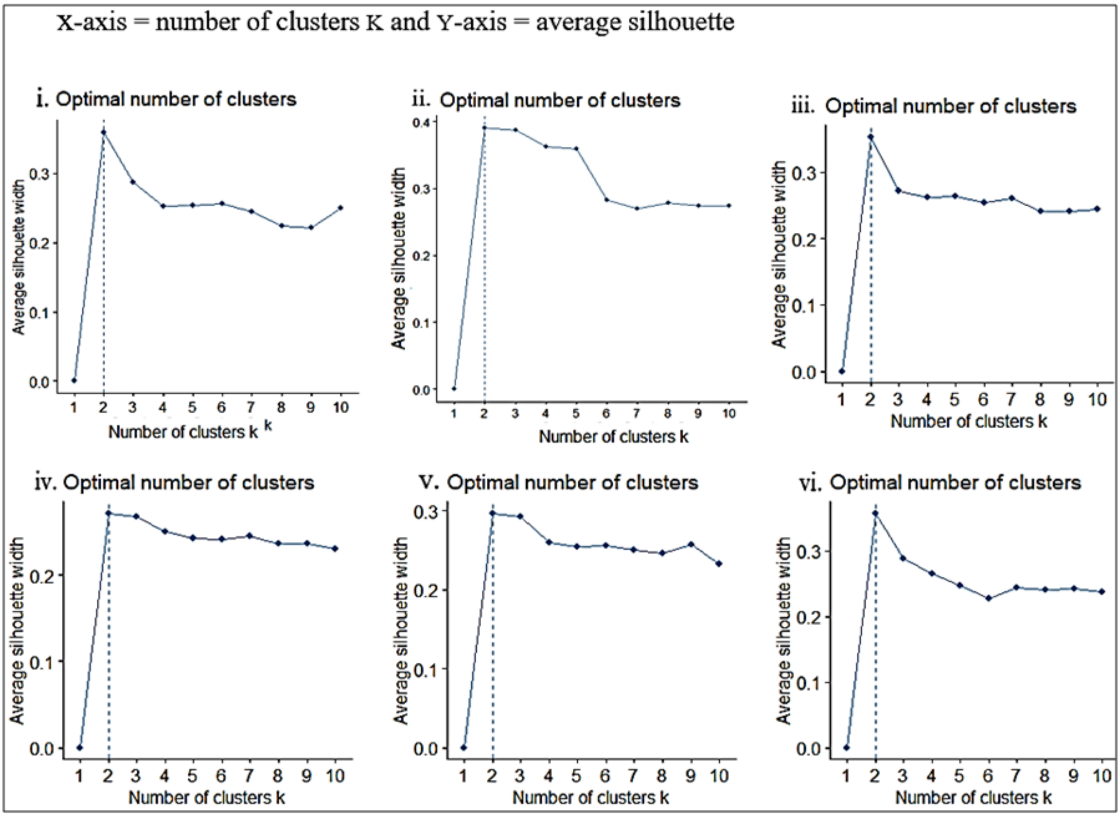

Figure 3: Silhouette plot on the six items

The valid clustering coefficients at k = 2 were computed and SVM were employed to automate, and categorize features into the two optimal classes. According to [25], by locating the crucial information, known as support vectors, the SVM algorithm, which is embedded in machine learning, creates a mathematical representation of the intended categorization. When compared to conventional clustering procedures, the test of validity made the datasets more connected. Additionally, the pattern made by A(SI) in figure 5, which is located on the left-hand side, clearly shows two cluster groups in contrast to the pattern made by Dunn index B(DI). The six investments are used to highlight the close pattern (shape) of both tests as shown in Figure 5.

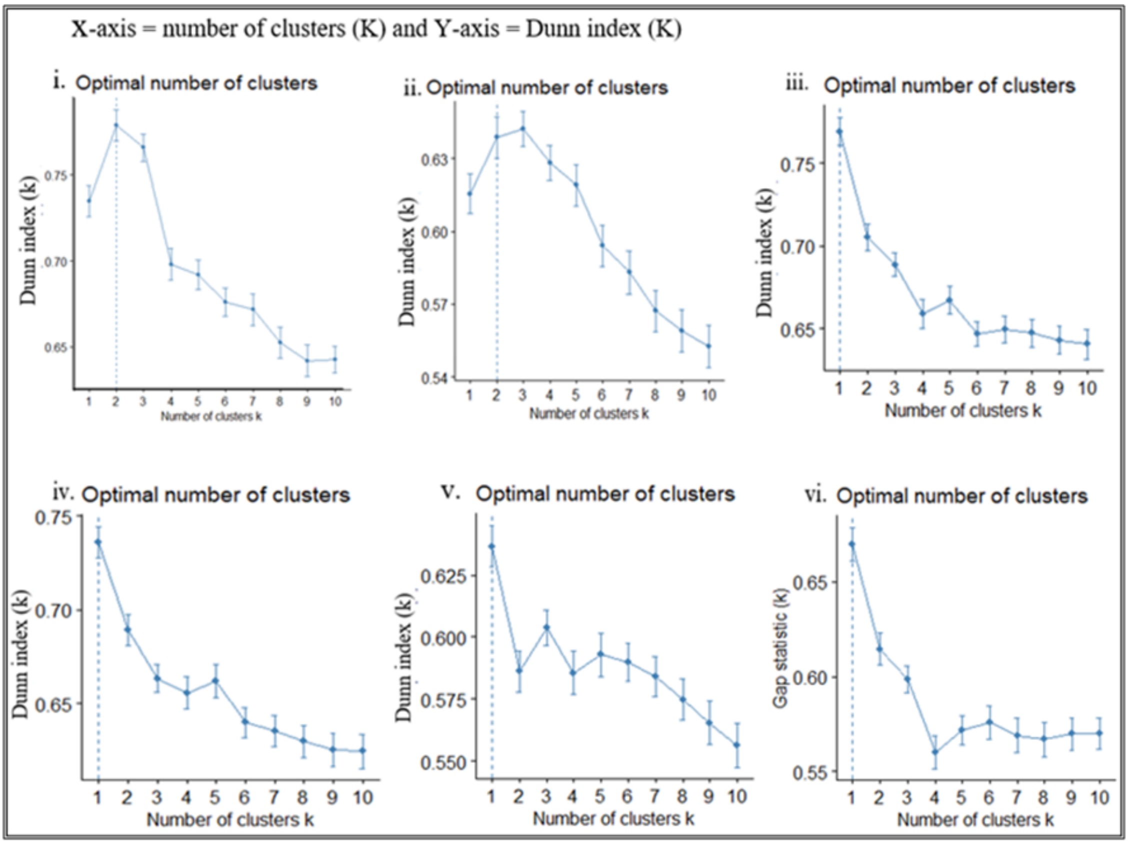

Figure 4: Diagram of Dunn plot for the items

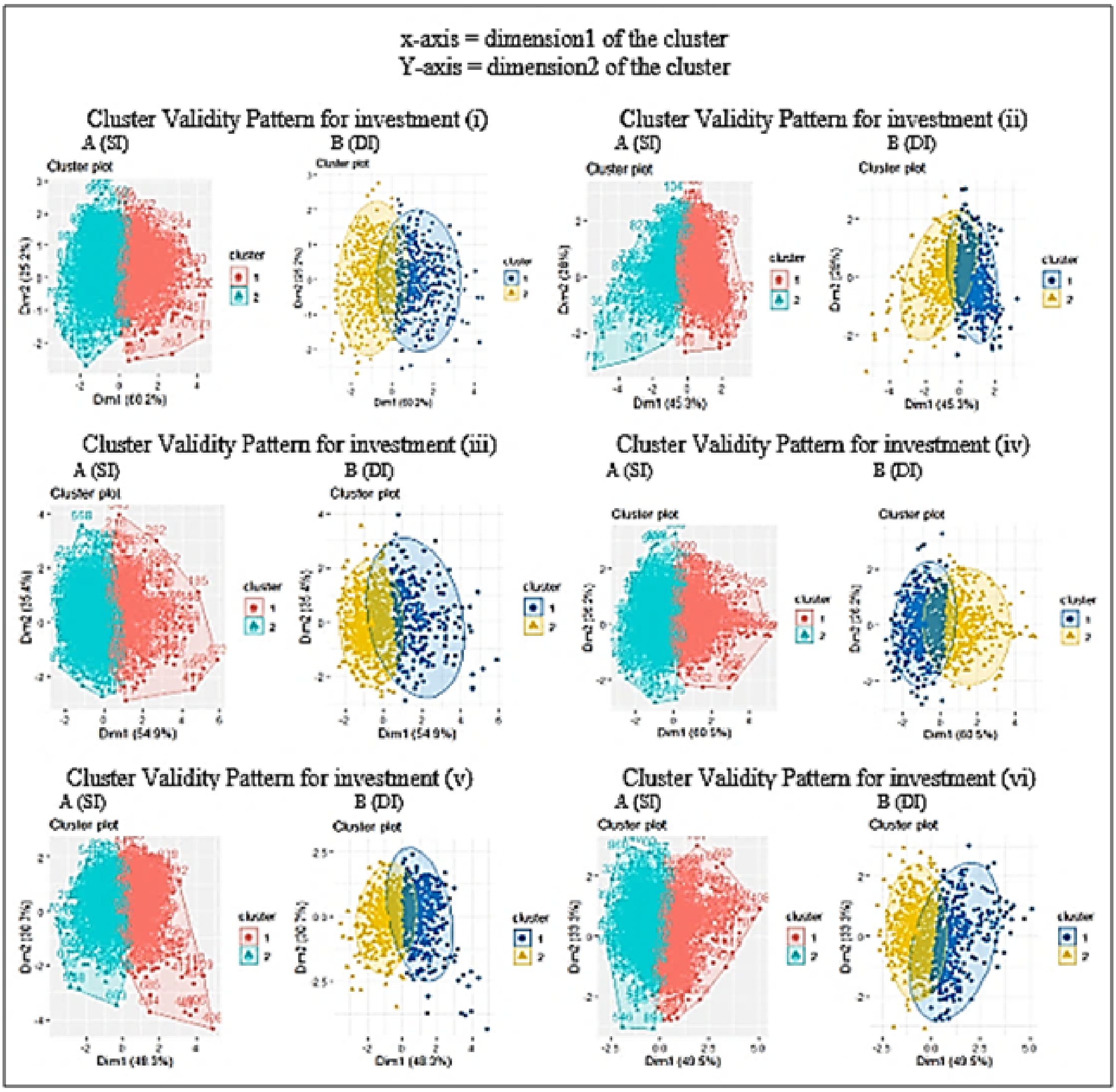

Figure 5: Comparison of patterns in cluster validity tests on the 6 investments

5 Conclusions

The adjusted in this study was found to produce the best return at stage 1. The model that best fit the investment returns was chosen after comparing between 2 tests of validity. Detailed examples were given on how the Modern Portfolio Theorem was developed and used when choosing which competitive stocks to invest in. This study described the methodology used to derive each stock's variances and correlations; selection is based on the appropriateness and values of the market. The model's equation and the process for choosing a portfolio successfully were supplied by this study. Investors need to carry out further optimization to get the expected return, especially where the market is free (See Figure 1). Among the study's innovations are a straightforward comparison that show the performance of the two tests in Figures $5-6$, and an assessment of the analysis to ensure the outcome analysis's accuracy. In group k = 2, the best probability value is 0.94, as indicated by the validity test and their results, which recorded 94% (high performance).Refer to Table 5. As a result, it is recommended that investors continue their investment strategy after the selection stage. They are encouraged to get an estimate of their expected returns and verify it before choosing to invest in the selected portfolio or assets. The practice will drastically reduce errors in financial management and save cost. Future studies could look into how to compare two or more models using the likelihood ratio test procedure. Future investigations into the application of validity testing to other financial models are made possible by this study.

Acknowledgement The author(s) like to acknowledge those that have contributed in one way or the other to this work. I acknowledge the Universiity of Nigeria, Nsukka for the basic training in Operations Research at the Universiti Sains Malaysia for research training I received for a period of three years. I appreciate the reviewers who scutinized this work to make it better and publishable. I thank Associate Prof. W. I. E. Chukwu and Prof. Mohd Tahir who always encouraged me and provided mentorship to me. Finally, I thank God Almighty for life and good health.

Conflicts of Interest The authors declare that no conflicts of interest to reveal.

References

[1] R. Ab Razak & N. Ismail (2014). Dependence measures in malaysian stock market. *Malaysian Journal of Mathematical Sciences*, *8*, 109–118.

[2] D. Y. Aharon, S. Ali & M. Naveed (2025). Who a (m) i? exploring quantile frequency connectedness in emerging ai and iot token markets. *The North American Journal of Economics and Finance*, *80*, 102497. https://doi.org/10.1016/j.najef.2024.102497.

[3] S. Akinwale & R. Abiola (2007). A conceptual model of asset portfolio decision making: a case study in a developing economy. *African Journal of Business Management*, *1*(7), 192–200. http://doi.org/10.5897/AJBM.

[4] S. Ali, N. S. Al-Nassar & M. Naveed (2024). Bridging the gap: uncovering static and dynamic relationships between digital assets and brics equity markets. *Global Finance Journal*, *60*, 100955. https://doi.org/10.1016/j.gfj.2024.100955.

[5] S. Ali, M. Naveed, M. Youssef & I. Yousaf (2024). Fintech-powered integration: Navigating the static and dynamic connectedness between GCC equity markets and renewable en-

ergy cryptocurrencies. *Resources Policy*, *89*, 104591. https://doi.org/10.1016/j.resourpol.2024.104591.

[6] S. Ali, M. Umar, M. Naveed & S. Shan (2024). Assessing the impact of renewable energy tokens on BRICS stock markets: A new diversification approach. *Energy Economics*, *134*, 107523. https://doi.org/10.1016/j.eneco.2024.107523.

[7] R. Bellman & R. Kalaba (2019). Dynamic programming and statistical communication theory. *Journal of the National Academy of Sciences*, *43*(8), 749–751. https://doi.org/10.1073/pnas.43.8.749.

[8] R. E. Bellman & S. E. Dreyfus (2015). *Applied dynamic programming*. Princeton university press, New York.

[9] M. W. Brandt, A. Goyal, P. Santa-Clara & J. R. Stroud (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies*, *18*(3), 831–873. https://doi.org/10.1093/rfs/hhi019.

[10] G. Brock, V. Pihur, S. Datta & S. Datta (2008). clvalid: An R package for cluster validation. *Journal of Statistical Software*, *25*(4), 1–22. https://doi.org/10.18637/jss.v025.i04.

[11] J. C. Dunn (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, *4*(1), 95–104. https://doi.org/10.1080/01969727408546059.

[12] J. Fang, L. Zhao, J. C. Franso & T. Van Woensel (2013). Sourcing strategies in supply risk management: An approximate dynamic programming approach. *Computers & Operations Research*, *40*(5), 1371–1382. https://doi.org/10.1016/j.cor.2012.08.014.

[13] F. García, F. Guijarro & J. Oliver (2018). Index tracking optimization with cardinality constraint: a performance comparison of genetic algorithms and tabu search heuristics. *Neural Computing and Applications*, *30*(8), 2625–2641. https://doi.org/10.1007/s00521-017-2816-9.

[14] R. Garcia-Dias, S. Vieira, W. H. L. Pinaya & A. Mechelli (2020). Clustering analysis. In *machine learning*, pp. 227–247. Elsevier,. https://doi.org/10.1016/B978-0-12-815739-8.00013-2.

[15] J. Gonzalez-Bueno, F. Garcia, J. L. Garcés-Bautista, J. Oliver & R. Tamošiūnienė (2024). A multi-objective credibilistic approach to portfolio optimization: balancing return, risk, and liquidity in emerging and developed markets. *Intelektinė ekonomika*, *18*(2), 284–307.

[16] J. González-Bueno, R. Tamošiūnienė, C. G. Morales & G. Rueda-Barrios (2025). Applying the mean-variance framework: portfolio optimization and comparative performance analysis in the emerging Colombian capital market. *Business, Management and Economics Engineering*, *23*(1), 164–188. https://doi.org/10.3846/bmee.2025.18864.

[17] R. C. Grinold (1983). Market value maximization and markov dynamic programming. *Management Science*, *29*(5), 583–594. https://doi.org/10.1287/mnsc.29.5.583.

[18] G. Hadley (1964). *Nonlinear and Dynamic Programming*. Addison-Wesley Publishing Inc., New York, USA 1st edition.

[19] J. Handl, J. Knowles & D. B. Kell (2005). Computational cluster validation in post-genomic data analysis. *Bioinformatics*, *21*(15), 3201–3212. https://doi.org/10.1093/bioinformatics/bti517.

[20] M. B. Haugh & L. Kogan (2007). Duality theory and approximate dynamic programming for pricing american options and portfolio optimization. *Handbooks in operations research and management science*, *15*, 925–948. https://doi.org/1016/S0927-0507(06)15015-9.

[21] C. S. J. L. Board & W. T. Ziemba (2009). Portfolio selection: Markowitz Mean-Variance Model. In S. Board & Ziemba (Eds.), *Encyclopedia of Optimization*, pp. 2990–2996. Springer, New York, NY.

[22] R. Klein, S. Koch, C. Steinhardt & A. K. Strauss (2020). A review of revenue management: Recent generalizations and advances in industry applications. *European journal of operational research*, *284*(2), 397–412. https://doi.org/10.1016/j.ejor.2019.06.006.

[23] H. Markowitz (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

[24] H. M. Markowitz (1959). *Portfolio Selection: Efficient Diversification of Investments* volume 16 of *Cowles Foundation Monograph*. Yale University Press, New Haven.

[25] T. A. Moughal (2013). Hyperspectral image classification using support vector machine. In *Journal of Physics: Conference Series*, volume 439 pp. 012042. IOP Publishing,. https://doi.org/10.1088/1742-6596/439/1/012042.

[26] F. V. Nelwamondo, D. Golding & T. Marwala (2013). A dynamic programming approach to missing data estimation using neural networks. *Information Sciences*, *237*, 49–58. https://doi.org/10.1016/j.ins.2009.10.008.

[27] S. Neogy, R. B. Bapat, A. Das & T. Parthasarathy (2008). *Mathematical programming and game theory for decision making* volume 1. World Scientific, Singapore.

[28] F. O. Ohanuba, M. T. Ismail & M. K. M. Ali (2023). Application of topological data analysis to flood disaster management in nigeria. *Environmental Engineering Research*, *28*(5). https://doi.org/10.4491/eer.2022.411.

[29] F. Ohanuba, P. Ezra & N. Eze (2020). On the financial optimization in dynamic programming via tabular method. *American Journal of Operational Research*, *10*(2), 30–38. https://doi.org/10.5923/j.ajor.20201002.02.

[30] F. Ohanuba, M. T. Ismail & M. M. Ali (2021). Topological data analysis via unsupervised machine learning for recognizing atmospheric river patterns on flood detection. *Scientific African*, *13*, e00968. https://doi.org/10.1016/j.sciaf.2021.e00968.

[31] I. Omisore, M. Yusuf & N. Christopher (2012). The modern portfolio theory as an investment decision tool. *Journal of Accounting and Taxation*, *4*(2), 19–28. https://doi.org/10.5897/JAT11.036.

[32] N. G. Polson & M. Sorensen (2011). A simulation-based approach to stochastic dynamic programming. *Applied stochastic models in business and industry*, *27*(2), 151–163. https://doi.org/10.1016/0377-0427(87)90125-7.

[33] P. J. Rousseeuw (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

[34] A. B. Ruiz, R. Saborido, J. D. Bermudez, M. Luque & E. Vercher (2020). Preference-based evolutionary multi-objective optimization for portfolio selection: A new credibilistic model under investor preferences. *Journal of Global Optimization*, *76*(2), 295–315. https://doi.org/10.1007/s10898-019-00782-1.

[35] W. F. Sharpe (2000). Portfolio theory and capital markets.

[36] E. D. Supandi, D. Rosadi & A. Abdurakhman (2017). Improved robust portfolio optimization. *Malaysian Journal of Mathematical Sciences*, *11*(2), 239–260.

[37] P.-N. Tan, M. Steinbach & V. Kumar (2013). Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*, *487*(533), 58.

[38] J. Tobin et al. (1965). The theory of portfolio selection. *The theory of interest rates*, *364*, 364.

[39] D. Tokic. The passive investment bubble 2020. https://doi.org/10.1002/jcaf.22443.

[40] G. Vulcano, G. Van Ryzin & C. Maglaras (2002). Optimal dynamic auctions for revenue management. *Management Science*, *48*(11), 1388–1407. https://doi.org/10.1287/mnsc.48.11.1388.269.

[41] X. Wang, B. Wang, T. Li, H. Li & J. Watada (2023). Multi-criteria fuzzy portfolio selection based on three-way decisions and cumulative prospect theory. *Applied Soft Computing*, *134*, 110033. https://doi.org/10.1016/j.asoc.2023.110033.

[42] Y. Wang, B. Zhang, S. Gao & C. Li (2021). Investigation on the effect of freeze-thaw on fracture mode classification in marble subjected to multi-level cyclic loads. *Theoretical and Applied Fracture Mechanics*, *111*, 102847. https://doi.org/10.1016/j.tafmec.2020.102847.

[43] D. Wu, R. H. Kwon & G. Costa (2017). A constrained cluster-based approach for tracking the s&p 500 index. *International Journal of Production Economics*, *193*, 222–243. https://doi.org/10.1016/j.ijpe.2017.07.018.

[44] S. Yadav, A. Kumar, M. K. Mehlawat, P. Gupta & V. Charles (2023). A multi-objective sustainable financial portfolio selection approach under an intuitionistic fuzzy framework. *Information Sciences*, *646*, 119379. https://doi.org/10.1016/j.ins.2023.119379.

[45] F. Yan & F. Bai (2009). Application of dynamic programming model in stock portfolio-under the background of the subprime mortgage crisis. *International Journal of Business and management*, *9*(3), 178–182.

[46] D. D. Yao, S. Zhang & X. Y. Zhou (2006). Tracking a financial benchmark using a few assets. *Operations Research*, *54*(2), 232–246. https://doi.org/10.1287/opre.1050.0260.

[47] S. Yu, S. Gao et al. (2016). A dynamic programming model for environmental investment decision-making in coal mining. *Applied energy*, *166*, 273–281. https://doi.org/10.1016/j.apenergy.2015.09.099.

[48] Y. Zhang, X. Li & S. Guo (2018). Portfolio selection problems with markowitz's mean–variance framework: a review of literature. *Fuzzy Optimization and Decision Making*, *17*(2), 125–158. https://doi.org/10.1007/s10700-017-9266-z.