Mathematics and Computer Science
2022; 7(6): 130-143
http://www.sciencepublishinggroup.com/j/mcs
doi: 10.11648/j.mcs.20220706.15
ISSN: 2575-6036 (Print); ISSN: 2575-6028 (Online)

Application of Dynamic Programming to Revenue Management: The Optimum Validity Model's Test (S)

Felix Obi Ohanuba[*], Everestus Okafor Ossai, Precious Ndidiamaka Ezra, Martin Nnaemeka Eze

Department of Statistics, University of Nigeria, Nsukka, Nigeria

Email address:

felix.ohanuba@unn.edu.ng (Felix Obi Ohanuba), everestus.ossai@unn.edu.ng (Everestus Okafor Ossai), precious.ezra@unn.edu.ng (Precious Ndidiamaka Ezra), nnaemeka.eze@unn.edu.ng (Martin Nnaemeka Eze)

[*]Corresponding author

To cite this article:

Felix Obi Ohanuba, Everestus Okafor Ossai, Precious Ndidiamaka Ezra, Martin Nnaemeka Eze. Application of Dynamic Programming to Revenue Management: The Optimum Validity Model's Test (S). *Mathematics and Computer Science*. Vol. 7, No. 6, 2022, pp. 130-143. doi: 10.11648/j.mcs.20220706.15

Received: December 4, 2022; **Accepted**: December 19, 2022; **Published**: December 29, 2022

Abstract: A suitable decision plan is followed by an effective financial management to achieve optimality while investing in a competing stock portfolio. This study altered a Dynamic Programming (DP) model of Bellman. The modified model was used to solve a business problem. The problems of choosing a stock portfolio for optimal return among investors in financial markets have resulted in a financial crisis. Most financial analysts provide investors with incorrect and unvalidated investment information. The consequences were minimal optimum, no return, and an investment problem. The goals are to ensure optimality in investor returns, validate the results using two validity tests, and select the test that best validated the model. The silhouette and Dunn tests were used to validate the outcome result. The results of using Silhouette reduced computational complexity and produced a more robust and validated return. The k-means clustering (an aspect of unsupervised machine learning) provided better statistical evaluation and information on the investment pattern. In comparison to previous work, the introduction of variables allowed for the best return at stage one. Finally, a validated investment report can help to avoid mistakes made by market analysts and investors when making investment decisions.

Keywords: Dynamic Programing, Stocks Portfolio, Reverse Algorithm, K-means Clustering, Dencision Making

1. Introduction

Revenue management in a financial stock portfolio has thrown the global economy for a loop. It has resulted in the collapse of many global stock markets as a result of a lack of information on where to invest and make the best return. The term "dynamic programming" is synonymous with optimization. The term can be used interchangeably with "planning" or "tabular method." As a result, the term "dynamic programming," abbreviated DP, derives from the term "mathematical programming," and it is defined as the mathematical optimization of variables in mathematics (or optimization of mathematical expressions) [1]. Dynamic programming, or DP, is the process of following an optimal decision rule when making a choice during a fund investment to achieve the best possible outcome, as the procedure entails.

According to a reviewed study in revenue management, there is a need to first state the choice-based network revenue management problem by formulating the underlying dynamic program and structuring a review based on its composition, components, or variables [2]. According to the literature, model modification is required to suit a specific (chosen) method. It is important to note that DP, like topological data analysis, can be combined with other methods on input datasets; the combined aspect of DP was used in decision-making, integrated Bellman's equation, and applied in estimating missing data [3]. The results of their study revealed that the DP had a significant advantage over the dataset's base model, and they recommended that future research be conducted to measure or test the significant differences in the parameters because there are closenesses or similarities in the state variables.

This study specifies the limit of the calculation and determines the number of times to run the model. The k-means clustering aspect of the evaluation test was discussed.

Code functions in the R programming language were used; the algorithm presented followed the step-by-step computation to minimize complexity. The topological data analysis helped to integrate the k-clustering algorithm, which explains what was done in the computation, and finally removed the outliers and displayed compactness in the resultant patterns.

The main goal of revenue management in passenger air transport was stated to be the duty of optimally selling a constant and perishable inventory in a specified time period; the company should have named or defined their products in relation to their services; the idea was to pilot the availability of the products (tickets) over the set time to produce the maximum profits (revenues) [4]. There are numerous revenue management systems, but their research concentrated on passenger air transportation, covering the state-of-the-art, key contributions, and challenges. According to their research, incremental sales in the airline industry would allow airlines to directly respond to their customers (including travel agents). Onboard shopping priority, group booking, home delivery, seat reservation, and other options are also suggested. They concluded that combining revenue management approaches with other sectors would create a favorable environment for future research.

Dynamic programming, or DP, was used in decision-making to design an effective way of distributing the limited fund (resources) among the group of projects to reduce the overall losses in project investment [5]. The DP has a lot more applications that are yet to be explored. Natural settings, it was stated, contribute to the importance of investment security with adequate networks, consistency of returns, management competency, a stable environment, and legal and regulatory controls [4].

The tabular method introduced in this study would produce a robust result and detailed procedures for solving the financial problem stage by stage while observing the recursive procedures; the network flow was also identified. According to Hamdy, DP can find the best solution to an n-variable problem by decomposing it into n stages of single variable sub-problems [6]. Hamdy shows in his work that the stage varies depending on the optimization problem. Dynamic programming does not provide computational details for optimizing each stage as the tabular method does. Markowitz in 1952 [7] developed a theory on portfolio selection and fund allocation strategy, particularly among various stocks, so that investors can get the most out of their portfolios; making the most profit is the primary goal of financial management, planning, and investment.

A study in dynamic programming used Bellman Richard's general DP model to solve a company's existing financial problems, but there was no flow route in their recursive stage, so the method was compounded and difficult to understand [8]. In contrast to the tabular method, where the paths of flow are distinct, the outcome benefit at each investment stage was unknown in Bellman's modeling procedure. Furthermore, their study did not show optimality at any stage, and detailed procedures of optimization of the dynamic programming problem are necessary for investors to have an idea of their profit or return at each stage of flow. This study incorporated both the optimality principle, which uses a recursive procedure to solve a company's financial problems and revealed the extent of validity of the result; the expected returns at each stage of implementation were shown so that the investor can see how their investment grows at each stage.

Another article presented the findings of a case study of the decision-making processes of investment managers in developing economies, and the article stated that natural settings contribute to the importance of investment security [9]. Furthermore, adequate networks, consistent returns, management competency, a stable environment, and legal and regulatory controls are key investment decision-making concepts. There was no validation or significant test to measure differences between the dependent variables. Modern Portfolio Theory (MPT) provides a solution for allocating funds rationally among various stocks. Markowitz's research identified the key to achieving the best returns with the least amount of risk in portfolios with varying asset risk. The MPT has been used to select a portfolio-asset with the lowest risk of investing; the theory is critical during the selection process; however, it does not provide the expected outcome from a specific stock. This selection assists investors in minimizing risk during portfolio selection or decision making; the study explained the procedures in MPT [10]. This study revealed MPT theories in assigning weight to the portfolio as a tool for investment decision making; covariances between assets with the lowest variance were also explained.

Haugh and Kogan [11] focused on American option pricing and portfolio optimization problems when the underlying state space is high-dimensional. Consult these works [12-20] for more information on dynamic programming and validity testing. This study clearly describes the procedure of Markowitz's Portfolio Theory in stock selection; apply a tabular method in solving the stock allocation portfolio problem in DP, and detect shape and existing cluster in the investment. The use of a tabular method in revenue management is uncommon, and evaluation of the results is rarely performed. This research topic was inspired by two factors. Investors have been misled when attempting to choose where to invest their resources among various stocks. When it comes to allocating resources for projects, government agencies struggle. The procedure of the method proposed in this study is simplified, concise, and follows the recursive processes of the principle of optimality. Bellman's dynamic programming model was modified to fit our approach.

Two variables were introduced into the DP model. system control and state control variables. The first (state control variable) was used to check the boundary limit, and the second (system control variable) helped to produce an optimal return at stage one. There was no optimal value at stage one in the previous work by The choice of portfolio selection for optimal return among investors in the financial markets has resulted in a financial crisis. Financial analysts aggravate the problem when they feed investors wrong and unvalidated information on investments. The effect mostly led to minimal or no optimal return, and sometimes to investment bankruptcy. The study's aims are to ensure global optimality in investors' returns and validate the results of the outcome using two validity tests. This study compared and chose the test that best fit the modified model. The Silhouette results reduced computational complexity and produced a more robust and validated return. The k-means clustering (an aspect of unsupervised machine learning) provided better statistical evaluation and information on the investment pattern.

The tabular approach was not considered in the majority of the literature. The advantage of this approach is that the procedures follow a simple, descriptive format that can be easily understood when explained to investors. The validity test authenticated the results, which most analysts failed to implement into their model. The most important advantage of this study is that errors made by market analysts and investors in their investment choices can be avoided by checking the validity of their outcome or model.

2. Material and Methodology

The methodology used in this study is known as the tabular method. The method was applied to the solution of an investment problem in which a company or investor purchases four stocks. The problem depicts the relationship between each stock's return and investment. The investment is measured in units, and the total amount invested by the company is \$6000. Modern portfolio theory explained the fundamental principles of choosing among competiting stocks. In its computation, the solution follows the iterative and recursive procedures of the dynamic programming principle. In this study, the introduced variable(s) assisted in achieving a global optimal investment return; the Bellman's model was further modified to suit the technique and to allow an optimal return at stage one, as explained in the following model:

$$f_k(S_k) = \max\{g_k(U_k) + f_{k+1}(S_{k+1})\} + x_k + x_{k+1},$$
$$0 \le x_k, x_{k+1} \le 10,\ 0 \le U_k \le S_k,\ k = n, n\text{-}1, ?\ ,1,\ \text{ and } n \ge k.$$

where S is the total investment, n is the Item number of the portfolio, U_k is the Decision variable, investment assigned to item K), $g_k(U_k)$ is the Stage objective function, return of U_k, S_k is the State variables (i.e., investment of item k to item n , $S_{k+1} = S_k$ - U_k is the state transition equation, $f_{n+1}(S_{n+1})$ = Maximize return of S_k , X_{k+1} is the control variable introduced and which tentatively lie between 0 to 10. The range is arbitrarily chosen for this study, and the study is limited to stage four. If we let X_{k+1}= 0 and $f_{n+1}(S_{n+1}) = 0$, we can recover the original model. The development of this modified DPP model is presented as follows.

Given x_j to be a vector containing a variable to be selected at stage j, where j represents the various stages; n is the last stage. The general Dynamic Programming model can be stated as:

$$\left.\begin{aligned} &\max z = \sum_{j=1}^{n} f_j(x_j) \\ &\quad \sum_{j=1}^{n} a_j x_j \le b;\ a_j \ge 0\ \forall_j\ j = 1,2 \ldots, n \\ &\quad x_j \ge 0\ \ \forall_j\ j = 1,2 \ldots, n, \text{ all } x_j \text{ integers} \end{aligned}\right\} \quad (1)$$

The methodology for its computation that yields optimal solutions, and indeed all alternative optimal solutions, is unique. In the model, a_j, and b are assumed to be integers in (1). The only restriction limits the selection of physical dimensions used in measuring a_j and $b.$ The procedure could start from the first nth stage (number) to the last nth number and vice vasa in the maximization of z, provided that the method makes it possible to examine every set of feasible x_j while the recursive principle is observed. This study started from the backward process (i.e., from where $k = 4$ to $k = 1$ or see stages 1 and 4 in section 4.2.1).

Denote z* the absolute maximum of equation 1; this is defined in (2) as follows

$$z^* = \max_{x_1,\ldots,x_n} \sum_{j=1}^{n} f_j(x_j), \quad (2)$$

where the maximization carried out for non-negative integers, x_j must satisfy (3)

$$\sum_{j=1}^{n} a_j x_j \le b. \quad (3)$$

We proceed by selecting a value of x_j, holding x_n fixed, and maximizing z over the remaining variables, $x_1, \ldots, x_{n-1}$. The values of $x_1, \ldots, x_{n-1}$ which maximize Z under these conditions will depend on the value of x_j selected. If we do this for every allowable value of x_n, then $Z*$ will be the largest Z value obtained, and we thus find a set of x_j which maximize Z. Firstly, x_j is selected and the expression is computed as shown (4).

$$\max_{x_1,\ldots,x_{n-1}} \left\{ \sum_{j=1}^{n} f_j(x_j) \right\} + x_k + x_{k+1} = f_n(x_n) + \max_{x_1,\ldots,x_{n-1}} \sum_{j=1}^{n-1} f_j(x_j). \quad (4)$$

$f_n(x_n)$ is independent of $x_1, \ldots, x_{n-1}$, and the maximization is taken over non-negative integers $x_1, \ldots, x_{n-1}$ satisfying

$$\sum_{j=1}^{n-1} a_j x_j \le b - a_n x_n.$$

Now, $\displaystyle \max_{x_1,\ldots,x_{n-1}} \sum_{j=1}^{n-1} f_j(x_j) + x_k + x_{k+1}.$

for non-negative integers that satisfies (2.4) depends on x_n or more specifically on $b - a_n x_n$.

Therefore,

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

Computing $W_{n-1}(b-a_n x_n)$ for every allowable value of x_n, (5) transforms to (6) as

$$Z^*=\max_{x_n}\left[f_n(x_n)+W_{n-1}(b-a_n x_n)\right]+x_k+x_{k+1}. \quad (6)$$

The variable, x_n can take on the values $0,1,\ldots,[b/a_n]$, where $[b/a_n]$ is the largest integer. An investment, a_n is assigned to the item n, and it is to be chosen based on the minimum value of the correlation of the item n. $W_{n-1}(b-a_n x_n)$ is a function to be maximized over the variables $x_1,\ldots,x_{n-1}$. The development of this model where the system control variable was not introduced can also be found in our previous study [21]. Ideally, the process utilizes the recursive relationship and solve from $f_n(s_n)$ of the first stage to the final $f_1(s_1)$ which is the maximum return on the issue, while portfolio allocation scheme is also optimal as reviewed in Ohanuba, F. O et al., and Hadley, G. [21, 22].

2.1. Markowitz Theory

Prof. Harry Markowitz, an economist, developed the theory in the 1950s with the goal of assisting investors in calculating risk based on their desired return. According to Modern Portfolio Theory (MPT), stocks with high correlations should be avoided. However, between the 1930s and 1950s, some scientists (Allais, De Finetti, Hicks, Marschak, and others) discussed elements of portfolio problems, but Harry Markowitz was the first to propose the theory of portfolio selection for risk aversion: von Neumann-work Morgenstern's on expected utility results provided evidence of economic justification for the model [23-25].

In this study, MPT was first applied in choosing the suitable stocks (those with the lowest correlation values) to invest in. The model in portfolio selection can be stated using the works of these authors; Tobin, Sharpe, and Board [26-29] as.

2.2. Minimize $X'VX$

$$\text{Subject to } \mathrm{X'h}=\mathrm{h_p} \quad (7)$$

$$\mathrm{X'e}-1$$

where X'VX is portfolio variance; the variable, X is the column of investment proportions in the risky assets, V is the positive semi-definite variance-covariance matrix of asset returns, h_p investor's target of return, h is the expected column vector asset returns, and e is the unit column vector.

The constraint to model restrictions of short investment is

$$\mathrm{X}\geq 0 \quad (8)$$

where 0 = column vector. The qaudratic equation in nonlinear mathematical programming is related to (8). Similar idea of augumentation was adopted in case 2 of the tabular method's solution in this work when we assumed a control variable to be a particular value. Some study focused on the attributes in an item purchased (invested on) rather than the item itself according to Board, J. L et al. [29].

3. Solution for Selecting Portfolio Model

In the absent of a short investment, the solution can be rewritten as

$$\mathrm{L}=\tfrac{1}{2}\mathrm{X'VX}-\lambda_1(\mathrm{X'h}-\mathrm{h_p})-\lambda_2(\mathrm{X'e}-1). \quad (9)$$

From (9), we define the first order conditions.

$$\mathrm{VX}=\lambda_1\mathrm{h}+\lambda_2\mathrm{e}.$$

This showed a straight-line relation between the target returns, h and their covariances, VX.

Make X the subject of the formular, to obtain

$$\mathrm{X}=\lambda_1\mathrm{V}^{-1}\mathrm{h}+\lambda_2\mathrm{V}^{-1}\mathrm{e}=\mathrm{V}^{-1}[\mathrm{he}]\mathrm{A}^{-1}, \quad (10)$$

where

$$A=V^{-1}\left[he\right]\left[\mathrm{h_p}\,1\right]'=V^{-1}\left[he\right]\left[\mathrm{h_p}\,1\right]'=\begin{bmatrix}\mathrm{h'V^{-1}}h\,\mathrm{h'V^{-1}}e\\ \mathrm{h'V^{-1}}e\;\mathrm{e'V^{-1}}e\end{bmatrix}.$$

Substite (10) into the portfolio variance (X′VX), gives

$$V_p=\left[h_p\,1\right]A^{-1}\left[\mathrm{h_p}\,1\right]' \quad (11)$$

Then, portfolio standard deviation, S_p (i.e., $S_p=\sqrt{V_p}$) is

$$S_p=\left(\left[h_p\,1\right]A^{-1}\left[\mathrm{h_p}\,1\right]'\right)^{1/2}=\begin{pmatrix}h_p^2/A & h_p/A\\ \\ h_p/A & 1/A\end{pmatrix}^{1/2}. \quad (12)$$

Next is the implementation of the tabular method, and finally the validity test of the analysis. The tabular procedure obeys the optimality rule while maintaining a recursive procedure [8].

Modern Portfolio Theory (MPT) emphasizes utilizing the best returns with the least amount of risk on portfolios with varying asset risks, but this is dependent on how one divides up his investment pile. The procedures for carrying out the MPT are as follows:

1) Data collection.
2) Creation of the Markowitz efficient frontier (i.e., $\sigma_p=\sqrt{\sigma_p^2}$).
3) Creation of market portfolio and market line.
4) Creation of the optimal portfolio; this is the plot that represents steps 1 to 3, that represents point of tangent between the Markowitw efficient frontier and the capital market [30]. The whole technique for this study integrated the validity analysis to ensure completeness

of the process and validity of the outcome. For more reading on portfolio selection of Markowitz, see Board, J. L et al.; Markowitz, H.; Zhang, Y et al.; Ismail, A. and Pham H.; Dai, Z. and Kang J. [29-33].

Analytically, if there are N number of securities; Let r_{it} be the anticipated return at time, t per dollar invested in security i. Let d_{it} be the rate at which the return on the i^{th} security is selected at time t be discounted back to he present; Let X_i be the relative amount invested in security, i. Then the discounted anticipated return of the portfolio is

$$R = \sum_{i=1}^{\infty}\sum_{i=1}^{N} d_{it} r_{it} X = \sum_{i=1}^{N} X_i (\sum_{i=1}^{\infty} d_{it} r_{it}) \tag{13}$$

$R_i = \sum_{i=1}^{N}\sum_{i=1}^{\infty} d_{it} r_{it}$, is then the discounted return asset of the i^{th} security.

$R = \sum X_i R_i$, where R_i is independent of X_i.

Since $X_i \geq 0$ for all i and $\sum X_i = 1$, R is a weighted average of R_i with the X_i as non-negative weights and to maximize R, we let $X_i = 1$ for i with maximum R_i. If several $R_{a_o}, a = 1,2,\ldots,k$ are maximum, then all allocations with $\sum_{a=1}^{k} a_0 = 1$ maximize R [7].

Theorem on Modern Portfolio of Return states that:

The portfolio of the return is the proportion of the combined weights of the aggregated assets' values.

$$R = \sum X_i R_i,\ E(R_p) = \sum_i w_i^2 E(R_i) = \text{the expected return,} \tag{14}$$

where R_p is the return obtained the portfolio, R_i is the return on asset i and, w_i is the weighting of the component asset "i" in the portfolio.

The Portfolio volatility is a function of the correlations, ρ_{ij} of the component assets, for asset pairs (i, j).

Portfolio return variance:

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j \rho_{ij},\ \rho_{ij} = 1 \text{ for } i = j \tag{15}$$

where ρ_{ij} is the correlation coefficient between the returns on assets i and j.

Taking the square root of (15), Portfolio return volatility (statistically called standard deviation) is obtained in (16)

$$\sigma_p = \sqrt{\sigma_p^2} \tag{16}$$

The diagram below depicts the Modern Portfolio Theory in action. Typically, the target return is plotted versus the portfolio standard deviation. When the curve aligns with the capital market line, as shown in Figure 1, the market indicates an open risk in which any investor can participate.

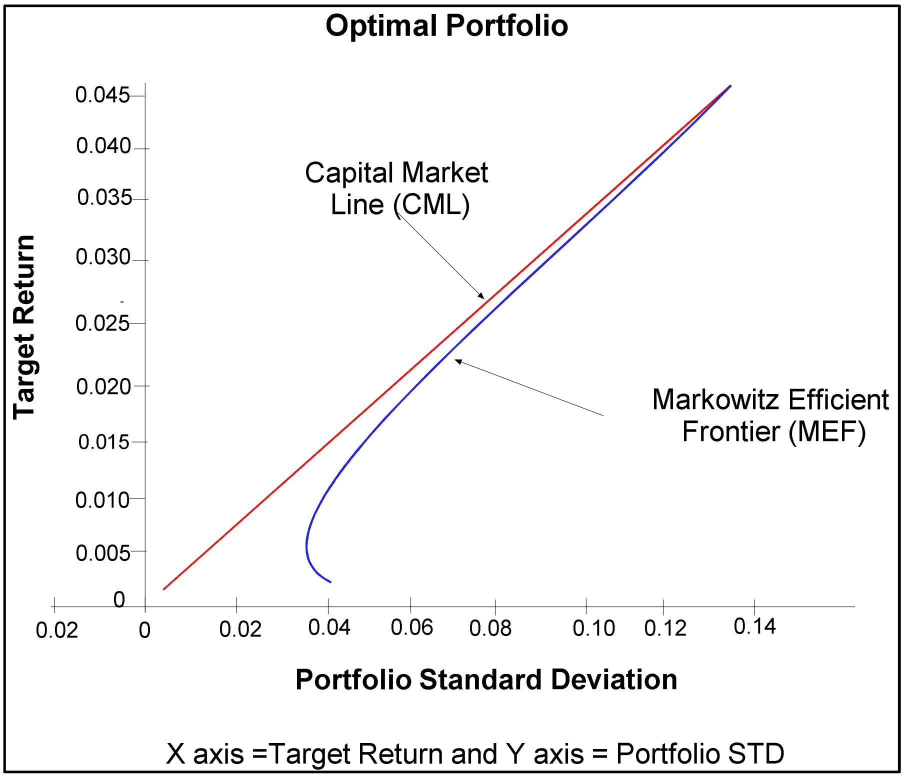

Figure 1. *Modern Portfolio Theory.*

If there is no risk-free asset, the efficient frontier of MPT is formed at the hyperbola; if there is a risk-free asset, the efficient frontier is formed at the straight line as revealed in Yao, D et al. [8], and shown in Figure 2.

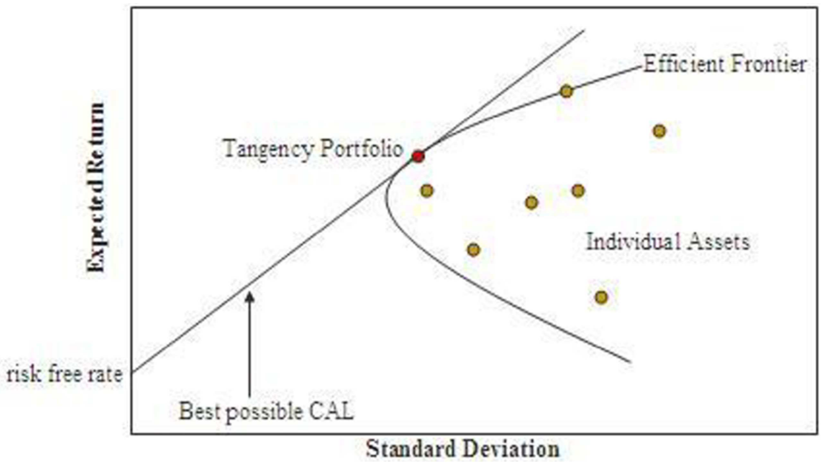

Figure 2. *Efficient Frontier for risk free assets.*

For a two-asset, Portfolio return is obtained from (14) as

$$E(R_p) = w_a E(R_a) + w_b E(R_b) = w_a E(R_a) + (1 - w_a) E(R_b), \tag{17}$$

and the Portfolio variance is

$$\sigma_p^2 = w_a^2 \sigma_a^2 + w_b^2 \sigma_b^2 + 2 w_a w_b \sigma_a \sigma_b \rho_{ab}$$

From (17) a three-asset Portfolio return is derivered as given in (18)

$$E(R_m) = w_a E(R_a) + w_b E(R_b) + w_c E(R_c), \tag{18}$$

and the Portfolio variance is derived from (14) as

$$\sigma_p^2 = w_a^2 \sigma_a^2 + w_b^2 \sigma_b^2 + w_c^2 \sigma_c^2 + 2 w_a w_b \sigma_a \sigma_b \rho_{ab} + 2 w_a w_c \sigma_a \sigma_c \rho_{ac} + 2 w_b w_c \sigma_b \sigma_c \rho_{bc} \tag{19}$$

For assets greater than three, the same derivation pattern is applied.

3.1. Metric Distance for Validity Tests

The statistical method of validating the research of this kind starts with data cleaning for missing values. Interpolation was used for data cleaning before proceeding to other analysis. The interpolation function, f_x in this study was obtained from Excel function given as

$$f_x = B2 + \frac{\$E2 - \$B2}{n+1} \tag{20}$$

where $B2$ = the value of the cell before the empty cell(s), depending on the position of the cell in excel spreadsheet, $E2$ = the value of the cell after the empty cell(s) (i.e., cells of the missing value),

n = the number of the empty cell(s) between ($B2$ and $E2$), $(n+1)$ = the gap between ($B2$ and $E2$), and \$ is used for fixing those empty columns once (where $n \geq 2$) by clicking and dragging from the edge of the first cell (of the missing value) where the function was first applied.

3.2. The Classical Metric Distance for Calculating Cluster Validity Index

The Euclidean distance or the Manhattan distance is an efficient distance measure used in calculating Silhouette index and Dunn index. The distance (space) calculation is referred to as a metric if it satisfies three properties: (i) symmetry, (ii) triangle inequality, (iii) positive value.

Let i and j be two points consisting of x'_i and x'_j p-variables, then, the Euclidean distance between i and j points ($d(x_i, x_j)$) can be calculated as

Euclidean: $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ and $y_j = (x_{j1}, x_{j2}, \ldots, x_{jp})$

$$d(x_i, x_j) = \left[(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{ip} - x_{jp})^2 \right]^{\frac{1}{2}}$$

$$d(x_i, y_j) = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |x_{i1} - x_{j1}|^2 \right)^{\frac{1}{2}} \; \forall \, x_i, x_j \in S$$

$$= \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \tag{21}$$

Let x_i and y_j be two points consisting of x'_i and y'_j n-variables, $|(x_i - y_j)|$ is the absolute value of distance between x_i and y_i. Then, the Manhattan distance between i and j points ($d(x_i, y_j)$) can be calculated as

$$d(x_i, y_j) = (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots + (x_{i6} - x_{j6})^2$$

$$= \sum_{k=1}^{6} |(x_{ik} - y_{ik})| \quad (22)$$

The variables $p = 6$ (i.e., the total number of portfolios for the investment, x_i and y_i are the values of the i^{th} variable, at points x and y, respectively.

Some of the important factors affecting specific validation approaches are based on two dimensions of datasets [34, 35].

Let x_i and y_j be two points consisting of c'_i and c'_j *n*-variables, then, the Dunn distance between *i* and *j* points ($d(c_i, c_j)$) can be calculated as

$$D_I = \frac{\min_{1 \le i < j \le n} d(c_i, c_j)}{\max_{1 \le i \le n} d(c_l)} \quad (23)$$

where $d(c_i, c_j)$ = inter-cluster distance metric between c_i and c_j clusters,

$d(c_l)$ = mean distance between all the pairs. The illustration for the cluster is shown in Figure 3. The figure represents three different cluster groups in a dataset. Intra-cluster distance metric is obtained within each group, but the mean distances are obtained between all the pairs.

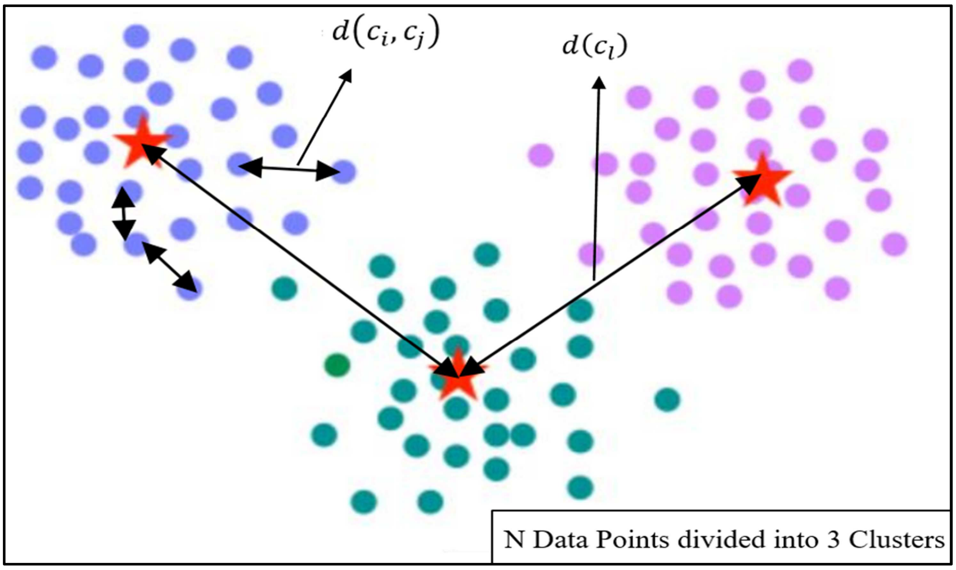

Figure 3. *N Data Points divided into 3 Clusters.*

Let X={$x_1, \ldots, x_n$} be a set of *n* object from space χ, *d* be a dis-similarity or distance over χ. This study deal with clustering that are partitions; it can equivalently be expressed by labels $l(1), \ldots, l(n) \in \mathbb{N}_k = 1, \ldots, k$ where $l(i) = r \Leftrightarrow x_i \in C_r$, $i \in \mathbb{N}_n$, and cluster sizes are denoted by $n_r = \sum_{i=1}^{n} 1(l(i) = r)$, $r \in \mathbb{N}_k$. The approach for the test will allows the data to be systematically classify homogenious elements into different groups after preprocessing the datasets; it is not manually or randomly [36]. The Silhouette index (width) for an observation $x_i \in X$ is

$$(C, d) = \frac{b(i) - a(i)}{max\ a(i), b(i)}, \quad (24)$$

where

$$a(i) = \frac{1}{n_{l(i)} - 1} \sum_{\substack{l(i)=l(j) \\ i \neq j}} d(x_i, x_j) \text{ and } b(i) = \min_{r \neq l(i)} \frac{1}{n_r} \sum_{l(j)=r} d(x_i, x_j)$$

$a(i)$ = average distance between *i* & *j* all other observations in the same cluster *C*.

$b(i)$ = average distance between *i* & *j* all other observations in the nearest. Each validity test has its benchmark for measurement. The Silhouette's range is between (−1 to +1), the Dunn index, lies between (0 to ∞). The index of Dunn should be maximized (large), which implies that low values are not considered.

3.2.1. Cluster Validity Indices

Cluster analysis is the process of dividing or grouping a set of observations (objects or datasets) into subsets. Each set represents a cluster, with groups within one cluster similar to one another and groups within other clusters distinct. According to Garcia-Dias et al. in 2020 [37], clustering analysis is a type of unsupervised learning that seeks to determine the most natural way of categorizing a dataset. Garcia-Dias et alwork.'s in 2020 also demonstrated that the k-prototype algorithms could join two algorithms using a combined dissimilarity measure (k-means and k-modes). Clustering is the process of categorizing a set of data objects into multiple groups or clusters so that objects within a cluster are highly similar but very dissimilar to objects in other clusters. Clustering is a data mining tool that is used in a variety of fields such as biology, engineering, mathematics, medicine, data mining, and so on. Density-based, partitioning, hierarchical, and k-means (a coefficient-based technique) are some clustering methods [38]. The validity measures are generally internal, stable, and biological, according to Brock et al. in 2008 [39].

3.2.2. Steps in Computating Validity Indices

The metric distance is then computed. The following are the five steps involved in the computation: 1. Importation of

Libraries 2. Data importation 3. Cluster centroids computation 4. Algorithm validation 5. Recalculation. Figure 4 illustrates the five steps. The relevant libraries for the computation were imported in the first step, which involved the use of code functions. The result of our dataset is then imported into Python programming language software; the second step is completed. In step 3, the data points were computed, and the algorithm was tested to ensure that the process was followed correctly. Finally, the cluster is recomputed until the optimum values in clusters are obtained. At k = 2, the optimum values also indicated the maximum values among the clusters (k = 1, 2, 3,4,5,6,7,8,9,10).

Algorithm psuedocode in k-means and silhouette analysis

Start

1 **Libraries' Importation:** Importation of pandas as PD, np, plt, K-means from sklearn cluster etc. // these are encoded in the computation spreadsheet.

2 **Importation of data:** read.csv(file.choose()) // import all csv files in folder one by one

3 **Computation of the cluster centroids:** Kmean.cluster_centers //compute the centroid of data points

4 **Testing of the algorithm:** // k-means and Silhouette analysis were used to test it

5 **Re-computation:** Re-compute the cluster until the best cluster value is achieved

Else

Repeat *4* and *5*

Figure 4. *The 5-Step procedure in the Validity Test computation.*

Intra-cluster validity test was adopted to evaluate the efficiency of our analysis. Silhouette widths were considered. The Euclidean distance is an efficient distance measure used in calculating Silhouette index in intra-cluster validity test.

4. Application to Financial Problem and Results

4.1. Return and Investment Problem

A company invested $60,000 in four stocks in the hopes of confirming the optimal portfolio through rational fund allocation to maximize investment return. The relationship between the return (unit: $10,000 per unit) and the investment (unit: $10,000 per unit) of each stock is shown in Table 1 after market investigation and expert forecast; the above is a company's financial problem.

Table 1. *Table of Return and Investment Problem.*

Item Return Investment	U_1	U_2	U_3	U_4
0	0	0	0	0
1	40	40	50	60
2	100	80	120	80
3	130	100	170	100
4	160	110	200	120
5	170	120	210	130
6	170	130	230	140

4.2. The Solution and Results

The previous values are used to derive the recursive solution. The best unit is determined by the highest values in the rows ranging from 0 to 6. Row 0 has no value, so 0 is used; row 1 has a maximum value of 60. The highest value is 280 units, which is $2,800,000, out of the maximum values (0, 60, 120, 180, 230, 260, and 280). (see Table 3). The same procedure is followed at each stage. In this study, the solution entails adding a system variable and a control variable to the model. The addition of system variables was intended to allow for optimal return at stage one when compared to previous work by Ohanuba, F. O et al. [21].

4.2.1. Solution and Results of the Financial Problem in the Modified Model

In this case, the system control variable, x_k and state control variable, x_{k+1} are used to solve the DP problem of Bellman's model.

1) S – total investment;
2) n – item number of portfolios;
3) U_k – decision variable, investment assigned to item *k;*
4) $g_k(U_k)$ – stage objective function, the return of U_k;
5) S_k – state variables, investment of item k to item n;
6) $S_{k+1} = S_k - U_k$ – state transition equation;
7) $f_k(S_k)$ – the maximum return of S_k;
8) x_{k+1} – state control variables;
9) X_k – system control variables.

In all cases, k is a fixed value (arbitrary value), and n varies at a different stage. Exampe, at the first stage, $n = 4$ is maximum and $k = 5$, is an outlier at stage one.

Therefore, the reverse dynamic programming equation can be recovered as follow:

$$f_k(S_k) = \max\{g_k(U_k) + f_{k+1}(S_{k+1})\} + x_k + x_{k+1}, \quad 0 \le x_k, x_{k+1} \le 10, \tag{25}$$

$0 \le U_k \le S_k$, $k = n, n\text{-}1, ?\ ,1$, and $n \ge k$.

If x_{k+1}= 0, x_k= 0 then, $f_{n+1}(S_{n+1}) + x_k + x_{k+1} = 0$, the original equation would be regained. The variable k = stages; the maximum stage in this problem is n = k = 4, which implies that every calculation outside this range is irrelevant ($k \leq n$). The value of k is determined by the value of n.

Solving the modified dynamic programming model in table method

This study assumed, $x_k = 10, x_{k+1}$ = 5, and n = 4 at every state of all the stages. Recall that n determines the value of k at every stage.

4.2.2. Results of the Recursive Procedure from k = 4 to k = 1

The First Stage

Given fthat $k = 4$, namely investing S_4 (S_4 = 0, 1, 2, 3, 4, 5, 6) in one stock (the fourth stock), in this case,

$$f_4(S_4) = \max\{g_4(u_4) + f_5(s_5)\} + x_4 + x_5,\ 0 \leq u_4 \leq s_4 \quad (26)$$

At this stage, $f_5(s_5)$ does not exist since, S_5 is outside the maximum stage (limit boundry). Hence $S_5 = 0$. Also, at $k = 4, x_{k+1} = x_5 = 0$ because it is out the limit bound not withstanding the initial assumption, x_{k+1} = 5. The assumption applies in the introduced variable at $k \leq 3$ that is at $x_4 = 10$.

Obviously, if $S_4 = 0$, $f_4(0) = 0 + 0 = 0, g_4(U_4) = 0$

if $S_4 = 1$, $f_4(1) = 60 + 0 = 60, g_4(U_4) = 60$

if $S_4 = 2$, $f_4(2) = 80 + 0 = 80, g_4(U_4) = 80$

if $S_4 = 3$, $f_4(3) = 100 + 0 = 100, g_4(U_4) = 100$

if $S_4 = 4$, $f_4(4) = 120 + 0 = 120, g_4(U_4) = 120$

if $S_4 = 5$, $f_4(5) = 130 + 0 = 130, g_4(U_4) = 130$

if $S_4 = 6$, $f_4(6) = 140 + 0 = 140, g_4(U_4) = 140$

Table 2. *Calculation of Optimal Return and Optimal item.*

S_4	$\mathrm{Max}\{g_4(u_4)\}$	$f_4(S_4)$	U_4
0	0	0	0
1	60	60	1
2	80	80	2
3	100	100	3
4	120	120	4
5	130	130	5
6	140	140	6

Optimal return at this stage is \$1,500,000 and the optimal item here is 6.

The Second Stage

Given $k = 3$, namely investing S_3 (S_3 = 0, 1, 2, 3, 4, 5, 6) among the two stocks (third and fourth stocks), which makes maximum returns on the investment allocated to the two stocks. In this case,

$$f_3(S_3) = \max\{g_3(U_3) + f_4(S_4)\} + x_4,\ 0 \leq U_3 \leq S_3,\ x_4 = 10 \quad (27)$$

Table 3. *Calculation of Optimal Return and Item.*

S_3 \ U_3	0	1	2	3	4	5	6	$\mathrm{Max}\{g_3(U_3) + f_4(S_4)\}$	$f_3(S_3)$	$U_3^*(U_3, U_4)$
0	0							0	10	(0,0)
1	60	50						60	70	(0,1)
2	80	110	120					120	130	(2,0)
3	100	130	180	170				180	190	(2,1)
4	120	150	200	230	200			230	240	(3,1)
5	130	170	220	250	260	210		260	270	(4,1)
6	140	180	240	270	280	270	230	280	290	(4,2)

The optimal item in this case is $(U_3 U_4) = (4,2)$, namely investment allocated to two stocks, 60,000: including investment in the fourth stocks, 20,000, and in the third stock, 40,000. The optimal return at this stage is \$2,900,000.

Third Stage

Given namely $k = 2$, investing $S_2 (S_2 = 0,1,2,3,4,5,6)$ among the three stocks (second, third and fourth stocks), which makes the maximum return on the investment allocated to the three stocks. Here,

$$f_2(S_2) = \max\{g_2(U_2) + f_3(S_3)\} + x_3,\ 0 \leq U_2 \leq S_2,\ x_3 = 1 \quad (28)$$

The same calculation method was applied, and the results obtained in table as follows.

Table 4. *Calculation of Optimal Return and Item.*

S_2 \ U_2	0	1	2	3	4	5	6	$\mathrm{Max}\{g_2(U_2) + f_3(S_3)\}$	$f_2(S_2)$	$U_2^*(U_2, U_3, U_4)$
0	0							0	10	(0,0,0)
1	60	40						60	70	(0,0,1)
2	120	100	80					120	130	(0,2,0)
3	180	160	140	100				180	190	(0,2,1)
4	230	220	200	160	110			230	240	(0,3,1)
5	260	270	260	220	170	120		270	280	(1,3,1)
6	280	300	310	280	130	230	180	310	320	(2,3,1)

Optimal item in this case is $(U_2, U_3, U_4) = (2,3,1)$, namely investment allocated to the three stocks, 60,000: including investment in the second stock, 20,000, in the third stock, 30,000, and in the fourth stock, 10,000. Hence, optimal return is $3,200,000.

Fourth Stage

Given $k = 1$, namely investing $S_1(S_1 = S = 0,1,2,3,4,5,6)$ among the four stocks (first, second, third and fourth stocks), which makes the maximum return on the investment allocated to the four stocks. Here,

$$f_1(S_1) = \max\{g_1(U_1) + f_2(S_2)\} + x_2,\ \ 0 \le U_1 \le S_1,\ \ x_2 = 10. \tag{29}$$

Table 5. *Calculation of Optimal Return and Item.*

S_2 \ U_2	0	1	2	3	4	5	6	Max$\{g_1(U_1)+f_2(S_2)\}$	$f_1(S_1)$	$U_1^*(U_1U_2, U_3, U_4)$
0	0							0	10	(0,0,0,0)
1	60	40						60	70	(0,0,0,1)
2	120	100	100					120	130	(0,0,2,0)
3	180	160	160	130				180	190	(0,0,2,1)
4	230	220	220	190	160			230	240	(0,0,3,1)
5	270	270	280	250	120	170		280	290	(2,1,1,1)
6	310	310	330	280	230	230	170	330	340	(2,0,3,1)

The optimal item in this case is $U_1U_2U_3U_4 = (2,0,3,1)$, namely investment allocated to the four stocks, 60,000: including investment in the first stock, 20,000, in the second stock, 0, in the third stock, 30,000, and in the fourth stock, 10,000. Finally, the optimal return this time is $3,400,000 compared to $3,300,000, which is the optimal return obtained when the variables was not introduced into the model. The value, $3,400,000 is the global maximum result in this study since it give the highest return among all the returns.

5. Result of the Validity Test (s)

This study followed the five-step procedure explained in the algorithm psuedocode to complete the validation test of the tabular method (See Figure 4). In the first step, libraries useful for the computation were imported. After, the resultant data were imported into the Python programming language software to complete the second step. The computation of the data points was done in step 3, and the testing of the algorithm was run to ensure the process was maintained to avoid error. Finally, the cluster is recomputed until the opimum values were obtained in clusters. The optimum values also indicated the maximum values among the clusters k = (1,2,3,4,5,6,7,8,9,10). The optimum value is obtained at k = 2. The Silhouette coefficient (SI) results obtained in this study fell within the range (0.7 and 0.9) at the best group shown in Table 6 and illustrated in Figure 6. At k = 2, the results fell between 70% – 94% efficiency. The validity test(s) has its unique benchmark for measuring perfect values. The Silhouette results were obtained in this study revealed that there is no error in the analysis of the problem and the best clusters were obtained in group k = 2 (See Table 6). The group K=2 has all its value > 0.5, indicating good result considering the benchmark. The average validity values in the groups also fall within values approximately ≥ 0.5, implying robust results on the average Silhouette.

5.1. Statistical Comparison of the Validity Models

The statistics that form the bases which validity models of the type $d(x_i, y_j) = \sum_{k=1}^{6} |(x_{ik} - y_{ik})|$ can be compared as presented here, especially when data are obtained as point clouds. Two such statistics that have been used for validity and evaluation are Silhouette index (S_I) and Dunn index (D_i) [40]. The evaluation is compared, and the best is chosen based on the plot in the computation that accurately indicate the optimal number of clusters. The likelihood ratio test can compare the adequacy or best fit of two models S_i and D_i based on L_r defined as

$$L_r = -2log\left(\frac{L_{S_I}}{L_{D_I}}\right) \sim \chi_\alpha^2\,(w) \tag{30}$$

In (30), L_{S_I} and L_{D_I} represent the likelihood of S_I and S_I, respectively. The models (S_I and D_I) have been defined in (17) and (18). The degree of freedom in the chi square distribution is represented by w. The hypothesis test, L_r is used to investigate that the two models fit the data equally; reject this if $L_r < -\chi_\alpha^2\,(w)$. The rejection of the hypothesis of equality of good fit reveals that S_I is better fit than D_I. Hence, w is formulated as $w =$ number of parameters in S_I – number of parameters in D_I.

In this study, the results of the clusters were obtained after the five steps explained in section 3.2.1 for Silhouette and Dunn tests and are shown in Tables 6 and 7, respectively. The tables show that the best cluster occurs at the group k = 2, where highest values were recorded. It is difficult to choose the best test based on observation of the outcomes in the two tables, since the values of the two tests fall within the same range (0–1). This shows that the structural patterns are similar for the the two tests. The researchers now obtain the plots of the six investments and use the outcome criterium in the plots to choose the best.

Table 6. *Silhouette index of data points in generated clusters for 6 investments.*

Stage	**Number of Cluster (K)**				**Ave. Validity test for S_I**
No Inv.	**1**	**2**	**3**	**4**	
1	0.148	0.752	0.680	0.599	0.545
2	0.204	0.940	0.740	0.630	0.629
3	0.201	0.697	0.602	0.548	0.512
4	0.189	0.708	0.634	0.630	0.540
5	0.248	0.700	0.620	0.604	0.543
6	0.219	0.780	0.608	0.544	0.538
Range =	(0.0–0.2)	(0.7–0.9)	(0.6–0.7)	(0.5–0.6)	(0.5-0.6)

Table 7. *Dunn index of data points in generated clusters for 6 investments.*

Stage	**Number of Cluster (K)**				**Ave. Validity test for D_i**
No Inv.	**1**	**2**	**3**	**4**	
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)

5.2. Selection of the Best Validity Model Based on the Performance of the Plot

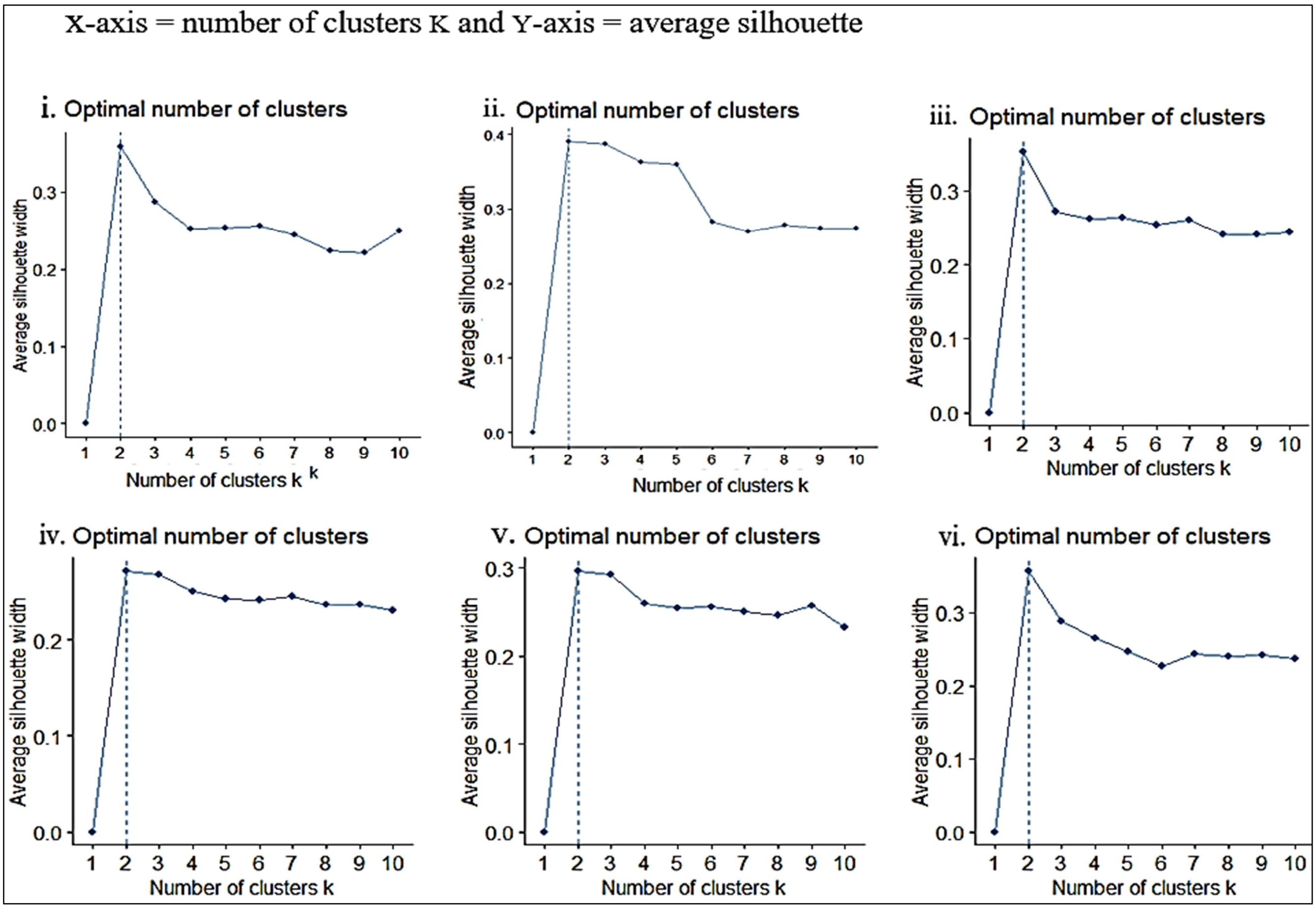

Figure 5. *Silhouette plot on the six investment items.*

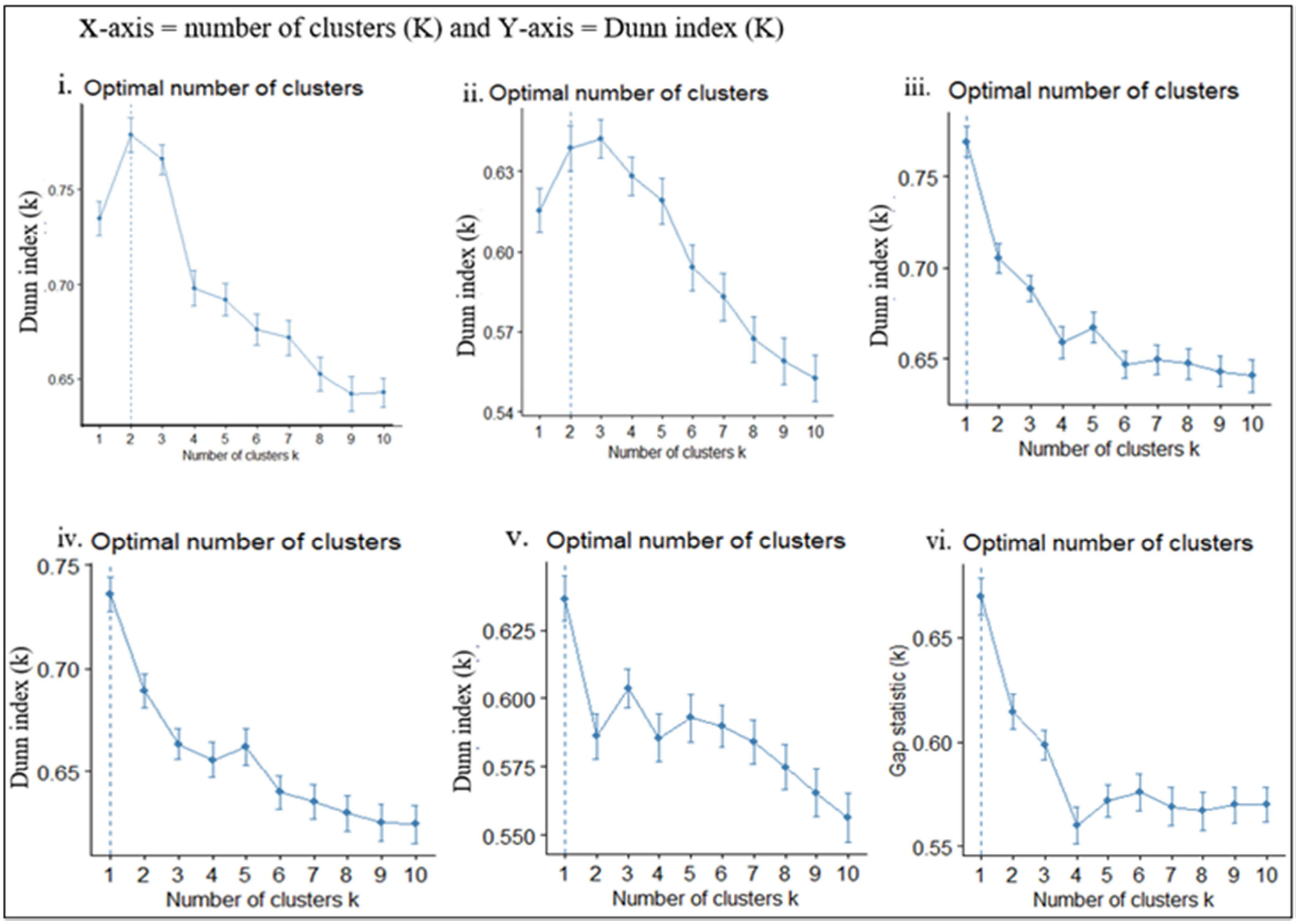

Figure 6. *Dunn index plot on the six investment items.*

The plots of the two tests were computed using Python programming codes. Figures 5 and 6 show the investment plots for S I and D I, respectively. In comparison to the Dunn (D I) plots in Figure 6, the Silhouette (S I) plots in Figure 5 were discovered to have attained the optimum value at exactly cluster k = 2 in all 6 items. In Dunn plots, the first two investments reached optimality at cluster k = 2, while the third through fifth investments did not but recorded optimality at cluster k = 1. As a result of the uniformity across the optimality class, the authors adopted the criterion in this section and chose the Silhouette as the best fit for the model. Dunn plot results revealed inconsistencies and errors in the validity measurement on the plots for those four investments.

6. Conclusion

In this study, the Bellman model of the DPP was successfully modified to yield the optimal return at stage 1. After comparing the outcome plots of the two validity tests, the study selected the model that provided the best fit for the investment returns. Illustrations of the development and application of the Modern Portfolio Theorem in the selection of investments for competing stocks. This paper explained how the variances and correlations of each stock are derived; selection is contingent on the market's values and suitability; and the Modern Portfolio Theory (MPT) provided a selection strategy for investors and financial managers. This study provided the equation for the model and the procedure for the successful selection of a robust portfolio. Even when the market is unrestricted and the curve is aligned with the capital, as depicted in Figure 1, investors must perform additional optimization to realize the anticipated return. The research offered Lastly, the evaluation of the analysis which need not be negleted in order to ensure the accuracy of the outcome analysis; these are the novelties of this study. The validity tests and their results revealed a high level of effectiveness of 94%; the best probability value in group k = 2 is 0.94. (See Table 6). Therefore, investors are advised not to stop their investment planning at the selection phase. Before deciding to invest in the chosen portfolio or assets, they are advised to obtain an estimate of their anticipated returns and validate it. Errors or mistakes in financial management could be drastically reduced as a result of this study. Future research may investigate the statistical test procedure of the likelihood ratio for comparing two or more models. This study paves the way for future research on the application of validity tests to additional financial management models.

Conflict of Interest

The author(s) declared that there is no contradiction of interest.

References

[1] Bellman, R., Dynamic programming princeton university press princeton. New Jersey Google Scholar, 1957.

[2] Wu, D., R. H. Kwon, and G. Costa, A constrained cluster-based approach for tracking the S&P 500 index. International Journal of Production Economics, 2017. 193: p. 222-243.

[3] Nelwamondo, F. V., D. Golding, and T. Marwala, A dynamic programming approach to missing data estimation using neural networks. Information Sciences, 2013. 237: p. 49-58.

[4] Klein, R., et al., A review of revenue management: recent generalizations and advances in industry applications. European Journal of Operational Research, 2020. 284 (2): p. 397-412.

[5] Yu, S. and S. Gao, A dynamic programming model for environmental investment decision-making in coal mining. Applied Energy, 2016. 166: p. 273-281.

[6] Hamdy, A. T., Operations research an introduction. 7th ed. 2002, New Jersey: Prentice Hall.

[7] Markowitz, H., Portfolio selection. The journal of finance, 1952. 7 (1): p. 77-91.

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

[9] Akinwale, S. and R. Abiola, A conceptual model of asset portfolio decision making: a case study in a developing economy. African Journal of Business Management, 2007. 1 (7).

[10] Omisore, I., M. Yusuf, and N. Christopher, The modern portfolio theory as an investment decision tool. Journal of Accounting and Taxation, 2011. 4 (2): p. 19-28.

[11] Haugh, M. B. and L. Kogan, Duality theory and approximate dynamic programming for pricing American options and portfolio optimization. Handbooks in operations research and management science, 2007. 15: p. 925-948.

[12] Bellman, R. E. and S. E. Dreyfus, Applied dynamic programming. Vol. 2050. 2015: Princeton University Press.

[13] Polson, N. G. and M. Sorensen, A simulation-based approach to stochastic dynamic programming. Applied Stochastic Models in Business and Industry, 2011. 27 (2): p. 151-163.

[14] Fang, J., et al., Sourcing strategies in supply risk management: An approximate dynamic programming approach. Computers & Operations Research, 2013. 40 (5): p. 1371-1382.

[15] Richard, C., Market Value Maximization and Markov Dynamic Programming.. Journal of Management Sciences, 1983. Vol. 29: p. 15-32.

[16] Brandt, W., Goyal A., Santa-Clara, P. and Stroud, R, A Simulation Approach to Dynamic Portfolio Choice with an Application to Learning About Return Predictability.. Society of finance Study., 2003. Vol 18: pp. 835-838.

[17] Vulcano, G., G. Van Ryzin, and C. Maglaras, Optimal dynamic auctions for revenue management. Management Science, 2002. 48 (11): p. 1388-1407.

[18] Yao, D. D., S. Zhang, and X. Y. Zhou, Tracking a financial benchmark using a few assets. Operations Research, 2006. 54 (2): p. 232-246.

[19] Yusuf, I. and N. Christopher, I,(2012), The modern portfolio theory as an investment decision tool. Journal of Accounting and Taxation, 2012. 4 (2): p. 19-28.

[20] Sankar, K., Mathematical Programming and Game Theory for Decision Making,. 2008. Vol. 1: pp. 2-13.

[21] Ohanuba, F. O, P. Ezra, and N. Eze, On the Financial Optimization in Dynamic Programming via Tabular Method. American Journal of Operational Research, 2020. 10 (2): p. 30-38.

[22] Hadley, G., Nonlinear and Dynamic Programming. 1st edition ed, ed. F. edition. 1964, New York city. USA: Addison-Wesley publishing Inc.

[23] Sharpe, W. F., A simplified model for portfolio analysis. Management science, 1963. 9 (2): p. 277-293.

[24] Sharpe, W. F., Mutual fund performance. The Journal of business, 1966. 39 (1): p. 119-138.

[25] Markowitz, H. M., Portfolio Selection: E cient Diversification of Investments. Cowles Foundation Monograph, 1959. 16.

[26] Tobin, J., Liquidity preference as behavior towards risk. The review of economic studies, 1958. 25 (2): p. 65-86.

[27] Tobin, J., The theory of portfolio selection. The theory of interest rates, 1965. 364: p. 364.

[28] Sharpe, W. F., Portfolio theory and capital markets. 1970: McGraw-Hill College.

[29] Board, J. L., C. Sutcliffe, and W. T. Ziemba, Portfolio selection: Markowitz mean-variance model. 2009.

[30] Markowitz, H., Portfolio Selection in The Journal of Finance Vol. 7. 1952.

[31] Zhang, Y., X. Li, and S. Guo, Portfolio selection problems with Markowitz's mean–variance framework: a review of literature. Fuzzy Optimization and Decision Making, 2018. 17 (2): p. 125-158.

[32] Ismail, A. and H. Pham, Robust Markowitz mean-variance portfolio selection under ambiguous covariance matrix. Mathematical Finance, 2019. 29 (1): p. 174-207.

[33] Dai, Z. and J. Kang, Some new efficient mean–variance portfolio selection models. International Journal of Finance & Economics, 2021.

[34] Rousseeuw, P. J., Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 1987. 20: p. 53-65.

[35] Handl, J., J. Knowles, and D. B. Kell, Computational cluster validation in post-genomic data analysis. Bioinformatics, 2005. 21 (15): p. 3201-3212.

[36] Wang, Y., et al., Investigation on the effect of freeze-thaw on fracture mode classification in marble subjected to multi-level cyclic loads. Theoretical and Applied Fracture Mechanics, 2021. 111: p. 102847.

[37] Garcia-Dias, R., et al., Clustering analysis, in machine learning. 2020, Elsevier. p. 227-247.

[38] Tan, P.-N., M. Steinbach, and V. Kumar, Data mining cluster analysis: basic concepts and algorithms. Introduction to data mining, 2013. 487: p. 533.

[39] Brock, G., et al., clValid, an R package for cluster validation. Journal of Statistical Software (Brock et al., March 2008), 2011.

[40] Ohanuba, F. O., M. T. Ismail, and M. K. M. Ali, Topological Data Analysis via Unsupervised Machine Learning for Recognizing Atmospheric River Patterns on Flood Detection. Scientific African, 2021: p. e00968.