Received May 20, 2020, accepted June 5, 2020, date of publication June 12, 2020, date of current version June 25, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3001944*

# A Framework to Perform Asset Allocation Based on Partitional Clustering

**FLÁVIO GABRIEL DUARTE AND LEANDRO NUNES DE CASTRO**

Natural Computing and Machine Learning Laboratory (LCoN), Mackenzie Presbyterian University, São Paulo 01302-000, Brazil

Corresponding author: Leandro Nunes de Castro (lnunes@mackenzie.br)

This work was supported in part by CNPq, in part by Capes, in part by Fapesp under Grant Proc. n. 2016/08352-1, and in part by MackPesquisa.

**ABSTRACT** Over the past years, many approaches to perform asset allocation have been proposed in the literature. Most of them tackle this problem as an optimization task, where the goal is to maximize return, whilst minimizing the risk. However, such approaches require the inversion of a positive-definite covariance matrix, usually resulting in the concentration of allocation, instability and low performance. Some methods have been recently introduced to solve this problem by facing it as a clustering problem. This paper introduces a framework for asset allocation based on partitional clustering algorithms. The idea is to segment the assets into clusters of correlated assets, allocate resources for each cluster and then within each cluster. The framework allows the use of different partitional clustering algorithms, intragroup and intergroup allocation methods. Also, various assessment criteria are considered, and a specialized initialization method is proposed for the clustering algorithm. The framework is evaluated with the Brazilian Stock Exchange (B3) data from the period 12/2005 to 04/2020. Different initialization methods are used for the clustering algorithm together with two intergroup and two intragroup techniques, resulting in five experimental scenarios. The results are compared with the Ibovespa index, the mean-variance model of Markowitz, and the risk-parity model recently proposed by López de Prado.

**INDEX TERMS** Asset allocation, partitional clustering, framework, mean-variance model, risk-parity model.

## I. INTRODUCTION

The work of Markowitz [1] is precursor in introducing a mean variance model for portfolio selection, resulting in a model capable of increasing portfolio returns and reducing risks. Various techniques have been introduced to improve the portfolio allocation process [2], [3], and more recently, machine learning models have been used to realize asset allocation [4], [5].

Traditional methods of portfolio optimization typically assume that future returns and expected risks need to be incorporated into the optimization process. A wrong estimate of future return and risk leads to a non-optimal portfolio as a result of optimization [3].

The existence of more than one objective to be optimized simultaneously, as well as the a priori unfamiliarity of the problem decision surface, makes the application of AI techniques very convenient for this task. Among these, multi-objective evolutionary algorithms have received much attention from the community as efficient tools for portfolio optimization [6].

Optimization methods that depend on future return and risk values may compromise the optimization result if the estimates are wrong. The algebraic constraint where increasing the correlation matrix size implies increasing the conditioning number can cause numerical errors that make the matrix unstable [4].

As an alternative to optimization methods, portfolio allocation techniques known as *risk parity* have become popular, where allocations are made based on predicted risk without the need to incorporate expected future returns [7]. However, in these methods it is still necessary to invert the positive definite covariance matrix that can lead to numerical errors and instability [4].

One way of performing portfolio allocation using risk parity without the need to invert the convariance matrix is

The associate editor coordinating the review of this manuscript and approving it for publication was Nadeem Iqbal.

by applying a hierarchical clustering algorithm, as introduced in [4] and later extended by [5]. In the approach proposed here, we use a partitional clustering algorithm, known as $k$-medoids, for portfolio selection.

While a hierarchical clustering performs a hierarchical decomposition of the dataset, a partitional clustering segments the dataset into a number of clusters. The proposed method allows for flexibility in portfolio construction, as both intragroup and intergroup information allow the application of different allocation techniques.

Inspired by the idea of making portfolio allocations using clustering algorithms and considering different ways of structuring the portfolio allocation process, this paper proposes a framework of stock portfolio allocation using partitional clustering algorithms. The *framework* corresponds to a sequential process of applying different building blocks so that, at the end of the process, a portfolio allocation is obtained. In summary, the main contributions of the paper are:

- A framework for asset allocation based on partitional clustering;
- Performing asset allocation without explicitly solving an optimization problem;
- Possibility of using different partitional clustering algorithms within the framework;
- Possibility of using different methods to perform intra- and intergroup allocation;
- Extensive performance evaluation with different configurations and comparison with other state of the art approaches.

From the works of de Prado [4] and Raffinot [5], which apply clustering algorithms to perform portfolio allocation, it was proposed to use fuzzy partitional clustering algorithms for allocation in our research. The idea of using fuzzy clustering was to consider cluster membership as a measure of the allocation of each asset.

This paper is organized as follows. In Section III we review the two main concepts of the paper: the asset allocation problem, and the clustering task. In Section IV we introduce the proposed framework, detaching the allocation process flow and each of its main phases: asset selection, the calculation of the correlation and distance matrices, the clustering method, the intragroup and integroup allocation methods, and some performance evaluation metrics used. In Section II we present a brief survey of clustering methods used to solve the asset allocation problem, and Section V brings the performance evaluation of the framework and its comparison with the mean-variance model of Markowitz [1], the Ibovespa index, and the Hierachical Risk-Parity method of de Prado [4]. The paper is concluded in Section VI with some general comments on the results obtained and some future trends for research.

## II. RELATED WORKS

By observing the optimization proposed by Markowitz [1] and the boundary created by the method, it is natural to apply multiobjective algorithms for solving such tasks. In the search for contributions made to this area, it is possible to observe that optimization-based approaches are widely studied according to the review by Ponsich *et al.* [6].

As clustering can be considered an optimization problem [8], and portfolio selection is originally an optimization task, it can also be solved using clustering algorithms. Therefore, this section will focus only on those related works based on clustering solutions for the portfolio allocation problem.

de Prado [4] and Raffinot [5] used as cluster preprocessing the concepts proposed in Mantegna [9], who was the first to propose and apply time series clustering algorithms from the correlation matrix to solve the portfolio allocation problem. The proposed method proved to be robust to find which groups of actions are most exposed and affected by economic variables. Following the concept of hierarchical clustering, de Prado [4] proposed a hierarchical allocation method using the information contained in the covariance matrix and created more diverse portfolios than those created with traditional methods. The results obtained were more diversified and less volatile portfolios.

From the work of de Prado [4], Raffinot [5] tested the following hierarchical clustering methods: *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward's Method* and *Directed Bubble Hierarchical Tree*. The author concluded that portfolios created from hierarchical clustering are more diversified and have better risk-adjusted returns.

Marti *et al.* [10] reviewed financial market clustering by commenting on the work of Mantegna [9] and how the technique he defined became the basis for time series clustering. The review also comments on clustering applications for portfolio selection, trading, risk management and financial policy making.

Nanda *et al.* [11] applied $k$-means, self-organizing map and *fuzzy C-means* to stocks. The authors applied these algorithms not only to time series from the correlation matrix, as is the aim of this paper, but also to perform a fundamentalist and balance analysis of a company using the data obtained from prices. The results obtained showed that the $k$ -means algorithm generated more compact groupings and, when compared with the market index, it presented better results.

Tola *et al.* [12] have applied cluster analysis for portfolio optimization. Based on Random Matrix Theory concepts, they used the hierarchical Single Linkage and Average Linkage algorithms to perform clustering. They also contributed a technique to indicate the optimal size of a portfolio. The results obtained were commented by the authors on the question of filtering technique difference of each algorithm. The tests showed consistent results in the evaluation period compared to the test period.

Dose and Cincotti [13] applied hierarchical clustering algorithms to build portfolios that aim to reflect the same performance of a market index, where the clustering result is an optimized portfolio that can replicate the index result in question.

## III. ASSET ALLOCATION AND CLUSTERING

### A. ASSET ALLOCATION

Portfolio selection is an activity that has been widely researched, especially among investment professionals and academics [14]. It is a research theme that has intersection with the areas of Finance, Mathematics and Computing [6], [15].

The first work involving portfolio selection was published by Markowitz [1], who proposed a mathematical model for the selection of asset portfolios containing the idea of investment diversification. Through diversification it is possible to reduce risk without decreasing the expected return. Markowitz proposed that an investor should seek to maximize portfolio return while variance, which is considered risk, should be minimized [16]. The ideas proposed by Markowitz became what came to be called the *Modern Portfolio Theory*, and the Markowitz model is also known as the *mean variance model*.

Fabozzi *et al.* [17] comment that one of the main aspects of Markowitz's work is not to assess asset risk in isolation, but to assess the contribution this asset makes to portfolio risk.

Markowitz's mean-variance work assumes a set of $N$ assets, where $X_i$ is the percentage of asset $i$ in the portfolio. The mean variance model has two constraints:

$$X_i \geq 0, \tag{1}$$

$$\sum_{i=1}^{N} X_i = 1. \tag{2}$$

Equation 1 states that the portfolio only has assets with a non-negative percentage, and Equation 2 adds that the sum of all assets in the portfolio must be 1.

The portfolio *Expected Return* ($E$) is calculated from the sum of expected returns on assets weighted by the percentage of assets in the portfolio:

$$E = \sum_{i=1}^{N} X_i \mu_i, \tag{3}$$

where $X_i$ is the percentage of asset $i$ in the portfolio, $\mu_i$ is the return of asset $i$, and $E$ is the expected return on the portfolio.

The expected risk ($V$) of the portfolio is given by the variance of assets and their correlation:

$$V = \sum_{i=1}^{N} \sum_{j=1}^{N} X_i X_j \sigma_{ij}, \tag{4}$$

where $X_i$ and $X_j$ are the percentages of assets $i$ and $j$ in the portfolio, $\sigma_{ij}$ is the covariance between $X_i$ and $X_j$, and $V$ is the portfolio variance.

The goal of Markowitz' mean-variance model is to minimize Eq. 4 (**min** $V$), whilst maximizing Eq. 3 (**max** $E$), subject to the constraints presented in Eqs. 1 and 2. The main thesis defended by Markowitz is portfolio diversification and its benefits [1]. With the model proposed above, the investor has a set of combinations of $E$ and $V$, where each portfolio has different expected values, according to the percentage of each asset in each portfolio. Given a set of $(E, V)$ portfolio combinations, the investor should choose the portfolio that has the best ratio between $E$ and $V$, which is the largest $E$ to the lowest $V$ possible.

Fabozzi *et al.* [17] point out that the Modern Portfolio Theory is a normative theory, describing patterns and norms of behavior that investors should pursue when allocating portfolios. According to them, the *asset pricing theory* [18] formalizes the relationship between risk and return that should exist between assets. Also, they point out that the mean-variance model and the asset pricing model form a robust portfolio selection framework. The authors also point out that the mean-variance model is a model independent of any asset pricing theory.

de Prado [4] comments that Markowitz [1] understood that for various risk levels there are optimal portfolios in terms of risk-adjusted returns, so the concept of *efficient frontier*.

To evaluate portfolio performance, Sharpe [19] proposed using an indicator, which was called the *reward-to-variability ratio*, and became known as *Sharpe Ratio* (*SR*):

$$SR = \frac{\mu - r_f}{\sigma}, \tag{5}$$

where $\mu$ is the return of the series over the period, $r_f$ is the return of the risk-free rate and $\sigma$ is the standard deviation of the analyzed series.

### B. CLUSTERING AND THE K-MEDOIDS ALGORITHM

*Clustering* is a term used to designate numerical methods of multivariate data analysis for the purpose of discovering groups of data. It can be defined as the organization of a set of objects into groups based on their similarity [20]. Grouping is the process of partitioning a dataset into subsets so that the objects in each group share common characteristics, usually proximity, based on some measure of similarity or distance. A group can be defined according to the internal cohesion and/or external isolation of its objects [20].

The concept of *natural cluster*, introduced by Carmichael and Julius [21], which defines that natural groups are those that share continuous, relatively dense and object-populated regions of the space, and these regions are surrounded by relatively empty regions.

Most clustering algorithms focus on obtaining $k$ groups of $n$ similar objects according to some pre-established criteria. Finding an optimal clustering solution is considered an NP-hard problem, because the number of clusters in a dataset is not known a priori and not all datasets have natural cluster separations [22]. Furthermore, different initializations of the clustering algorithms may result in different performances.

Clustering algorithms can be roughly divided into: partitional, hierarchical, density-based, and graph-based [20]. The most well-known hierarchical algorithms are single-link and complete-link, and both can be implemented in various forms. DBSCAN is an implementation of density-based algorithms and the minimum spanning tree is an implementation used for graph-based clustering solutions [20].

The most commonly used partitioning algorithms are $k$-means, $k$-medoids and variations of both [20]. This work proposes the use of a $k$-medoids partitional algorithm, in its variation known as *Partitioning Around Medoids* (PAM) [23]. The choice of $k$-medoids is due to the fact that the distance matrix is obtained from the correlation between the time series of assets. Thus, the distance between the objects (assets) and a prototype that is not another data object does not exist and, therefore, the prototype must be a data object (medoid). Since the $k$-medoids algorithm operates by receiving an already calculated distance matrix as input, this algorithm is suitable to be applied to this task.

#### 1) THE *K*-MEDOIDS ALGORITHM

The $k$-medoids algorithm is a clustering method similar to $k$-means, but whose prototype is always a data object. While $k$-means is sensitive to noisy and discrepant data objects, because of the average that the algorithm uses to define and update groups, $k$-medoids is more robust to both problems, because instead of using the average of objects to represent the center of the group, $k$-medoids uses an object from the group itself to represent the center. The group center object in $k$-medoids is named *medoid* and represents the most central object in the group based on the minimum sum of distances for the other objects in the group [24].

A common implementation found in the literature for $k$-medoids is called *partitioning around medoids* (PAM), proposed by Kaufman and Rousseeuw [23]. PAM consists of selecting $k$ objects to be the initial medoids. Then, each object is associated with the group whose medoid is the closest. Repositioning of the medoids is performed by minimizing the objective function, which is the sum of the distances of objects from the group to the medoid. Following the repositioning of the medoids, the process of assigning the objects to the groups and recalculating the medoid placement is performed iteratively until the algorithm has converged and no further medoid and cluster updates happen [24]. The $k$-medoids algorithm pseudocode is presented in Algorithm 1. PAM operates by minimizing the intracluster distance [25] and most partitioning clustering algorithms aim to minimize the quadratic error between objects in a group and their prototype (intracluster distance).

**Algorithm 1** $k$-medoids PAM Algorithm Pseudocode

```
def kmedoids(k, data):
    medoids = select_medoids(k,data)
    medoids_copy = medoids
    while not medoids == medoids_copy:
        groups = calc_groups(data,
            medoids,k)
        medoids_copy = update_medoids(
            groups,medoids)
        medoids = medoids_copy
    return groups
```

Reynolds *et al.* [26] point out that in the $k$-medoids algorithm there is no need to recalculate the distance between objects and the prototype at each interaction, since distances are already calculated in the distance matrix, reducing the computational cost when compared to $k$-means.

#### 2) FUZZY CLUSTERING

Fuzzy clustering is a technique that can be applied to the $k$-means and $k$-medoids algorithms, which aims to assign a membership degree for each object in relation to each cluster. While crisp methods indicate whether or not an object belongs to a group, in fuzzy clustering every object belongs to all groups, but with different membership degrees [20].

A feature of fuzzy clustering is that for each object $i$ in the $j$-th group, there is a $\mu_{i,j}$ value that corresponds to the membership degree of the object to the group. The sum of the membership degrees of each object to all groups must be equal to 1, which is also the maximum membership value of an object to a group [20]:

$$\sum_{j=1}^{k} \mu_{i,j} = 1 \tag{6}$$

For the $k$-means algorithm, the centroid positioning is calculated considering the membership degree of the objects to the group. However, for the $k$-medoids algorithm, whose prototype is an object from the dataset, the membership degree does not interfere with the definition of the group center, since the membership degree is determined after the medoids are calculated. To calculate the membership degree of the objects in a cluster found by $k$-medoids, Equation 7 [25], [27]:

$$\mu_{ij} = \frac{(\frac{1}{r(x_i,v_j)})^{1/(m-1)}}{\sum_{k=1}^{c}(\frac{1}{r(x_i,v_k)})^{1/(m-1)}}, \tag{7}$$

where $x_i$ is the object whose membership is being calculated, $m > 1$ the *fuzzyfication* parameter (a coefficient that weights the membership degree), $r(.,.)$ a measure of the distance between assets, $v_j$ the group's medoid, and $v_k$ the other medoids.

### C. THE IBOVESPA INDEX

The Ibovespa index is the result of a theoretical portfolio prepared according to a methodology that aims to be an indicator of the average performance of the most representative and tradable assets of the Brazilian stock market [28]. The Ibovespa index is reevaluated and rebalanced every four months.

For an asset to be included in the Ibovespa index, it must, within the term of three previous portfolios and meet the following criteria:

- Be in the top 85% of assets in the tradability index.
- Have a presence of, at least, 95% in trading sessions.
- Have a participation of financial volume greater than 0.1% of the total stock market.
- Not traded for less than R$ 1.00 (one real) per share.

The weighting of an asset in the Ibovespa index is based on the market value of the shares available for trading by the companies, with a participation limit based on the liquidity of the asset.

## IV. A FRAMEWORK TO PERFORM ASSET ALLOCATION BASED ON PARTITIONAL CLUSTERING

de Prado [4] proposed using *machine learning* for portfolio allocation introducing a technique called *Hierarchical Portfolio Construction* (HPC), which uses a hierarchical clustering algorithm to perform allocation based on the clustering result.

Sharpe [18] noted that while conceptually portfolio optimization is simple, involving the creation of efficient boundaries, the solutions are complex. This problem is usually solved with parametric quadratic programming, which uses inverse arrays for finding the solution [18], [29].

de Prado [4] noted that in portfolio optimizers, as correlated assets are added to this matrix used for portfolio optimization, the matrix conditioning number increases, making the inverse calculation unstable. The proposed partitioning clustering portfolio allocation method does not need to calculate the inverse matrix, as the allocation process consists of the following steps:

1) Perform clustering from the correlation matrix between assets;
2) Perform intergroup allocation; and
3) Perform intragroup allocation.

### A. THE ALLOCATION PROCESS FLOW

The framework starts by calculating the correlation matrix between assets and then their distance matrix from the correlation. Then, a partitional clustering algorithm is applied, which divides the assets into different clusters according to their similarity or dissimilarity. Once segmented, it is necessary to perform the allocation within each group (intragroup) and, finally, the allocation among groups (intergroup), where the final portfolio allocation will be made taking into account the allocations of each group. After allocation, the performance of the generated portfolio is evaluated. This flow is summarized in Figure 1 and each of the building blocks will be described separately in the following sections.

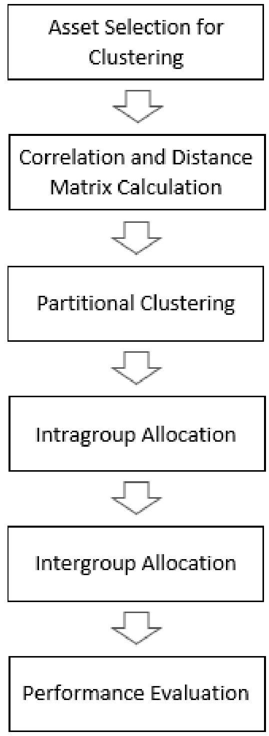


**FIGURE 1. Allocation process flow.**

### B. PREPROCESSING: CALCULATION OF CORRELATION AND DISTANCE MATRIX

To cluster financial time series, it is necessary to perform some transformations with the series. According to Marti *et al.* [10], a technique for time series clustering was defined by Mantegna [9] for hierarchical clustering and has been widely adopted ever since. The technique can be defined as follows:

- Let $N$ be the number of assets to be clustered.
- Let $P_i(t)$ be the price of asset $i$ at time $t$, where $1 \leq i \leq N$.
- Let $r_i(t)$ be the return in time $t$ of asset $i$, where:

$$r_i(t) = \log P_i(t) - \log P_i(t-1). \tag{8}$$

- For each pair of assets $(r, s)$, the correlation $\rho(r, s)$ is calculated:

$$\rho(r, s) = \frac{\sum (r - \overline{r})(s - \overline{s})}{\sqrt{\sum (r - \overline{r})}\sqrt{\sum (s - \overline{s})}} \tag{9}$$

- The correlation coefficients $\rho(r, s)$ are converted into distance $d(r, s)$:

$$d(r, s) = \sqrt{2(1 - \rho(r, s))} \tag{10}$$

### C. CLUSTERING

After the preprocessing steps that generate the asset correlations distance matrix, Mantegna [9] applies a *minimum spanning tree algorithm* to find hierarchical clusters. The same steps are used by de Prado [4] and will be employed in our framework with the application of a partitional clustering algorithm.

In the framework, instead of using the log return (Eq. 8), the simple net return is used, and the distance equation is the same as the one adopted by de Prado [4]:

$$d(r_i, s_i) = \sqrt{\frac{1}{2}(1 - \rho(r_i, s_i))} \tag{11}$$

After distance calculation, clustering is performed using a partitional clustering algorithm, such as $k$-medoids [20].

Starting from the initial medoids, the $k$-medoids algorithm is run, verifying to which group each object belongs, being the group formed by the objects that present the shortest distance to the group's medoids. After the first cluster partitioning, all objects in each group are checked to find the one with the shortest distance to all objects in the group. This process allows the recalculation of the medoids positioning. As the algorithm is iterative, the group membership and recalculation of the medoids are performed until it has converged and there is no further update to be made.

In general, the $k$-medoids algorithm is sensitive to the initial medoids, i.e., different choices of the initial medoids result in different data partitions. Thus, the framework allows the use of different initialization methods of prototypes (medoids). Two ways to initialize the medoids will be considered: 1) a standard method of random initialization; and 2) a novel heuristic method that takes as initial medoids the assets that have the lowest volatility observed in the set. The heuristic initialization method takes as the initial medoids those assets with the lowest standard deviation of the past series of prices. This results in lower volatility assets composing the portfolio. According to Baker *et al.* [30], portfolios composed of low volatility stocks have higher returns and lower drawdowns.

### D. INTRAGROUP ALLOCATION

Once the groups and objects belonging to each of the clusters are identified, it is necessary to perform *intragroup allocation*, which aims to allocate the resources in each of the assets that make up the group.

In this work two intragroup allocation methods are proposed: a *fuzzy method*, which calculates the membership degree of each object to each group; and a *performance proportional method*, which allocates based on a performance indicator, giving greater weight to objects with higher indicators, and vice versa.

#### 1) FUZZY INTRAGROUP ALLOCATION (FIA)

In clustering tasks, a fuzzy process can be used to calculate the membership of each object to each group. This method is used by applying Equation 7 to find the membership degree of the $i$-th object to the $j$-th group. The membership degree indicates that the closer the object of the medoid, the greater its relevance to that group, while the farther from the medoid, the lower the relevance to the group. Since the medoid is the object that has the shortest average distance to all objects in the group, it is understood that it is the object that best represents the group and, therefore, its membership degree to the group is maximum. After applying the equation, we have the membership degree for all objects to all groups, where $\mu_{i,j}$ represents the membership information of the object $i$ to the group the object belongs to, represented by $j$.

#### 2) PERFORMANCE PROPORTIONAL INTRAGROUP ALLOCATION (PPAA)

The selection mechanism named as *Performance Proportional Intragroup Allocation* was inspired by the *roulette wheel* selection technique in Genetic Algorithms [31]. It is appropriate for defining the allocation to be made to each group in the proposed framework, because it considers a performance indicator of the (assets) objects and weighs the relevance among all objects in proportion to that indicator. In the case of intragroup allocation, it is considered the performance of the assets in the period analyzed, making the assets that performed better more relevant in the group, while those that performed worse had lower relevance in the group.

Equation 12 represents the operation of the PPAA method [32]:

$$\mu_{i,j} = \frac{f_{i,j}}{\sum_{c=1}^{n} f_{c,j}}, \tag{12}$$

where $\mu_{i,j}$ represents the importance of the $i$-th object to the $j$-th cluster, according to the $f$ metric used to calculate the membership degree of each object in the group. In this paper, the $f$ metric used was the past performance of the assets. In the sum, $n$ is the number of assets in the $j$-th group.

### E. INTERGROUP ALLOCATION

Once the groups, their objects and the intragroup allocation have been made, the intergroup allocation must be performed. The result of the intergroup allocation will be the final portfolio allocation. The intergroup allocation considers the information present in each object of each group, found from the use of the previously presented methods, which are:

- **fuzzy method**: Each object $i$ has a $\mu_{i,j}$, which represents the membership degree for the $j$-th group.
- **performance proportional method**: Each object $i$ has a $\mu_{i,j}$, which presents relevance to the $j$-th group found from a metric previously defined.

Initially two intergroup allocation methods will be proposed: *balanced capital allocation*; and *performance proportional intragroup allocation*.

#### 1) BALANCED CAPITAL ALLOCATION (BCA)

A balanced intergroup capital allocation is the one that divides the amount invested evenly across all groups, i.e., each group receives the same amount of capital. DeMiguel *et al.* [33] presented this strategy as a naïve diversification rule, but which tends to perform well with a good *sharpe ratio* and return compared to a traditional capital allocation optimization strategy:

$$C_j = \frac{C_t}{k}, \tag{13}$$

where $C_j$ is the capital for each group, $C_t$ is the total capital and $k$ the number of groups. The method proposed by DeMiguel *et al.* [33] is expressed by Equation 13, but the final portfolio allocation should total all available capital for investment, then normalize according to:

$$w_{i,j} = \frac{\mu_{i,j}}{k} \tag{14}$$

where $w_{i,j}$ is the percentage of the $i$-th asset in the final portfolio, $\mu_{i,j}$ is obtained from the intergroup allocation, and $k$ is the number of groups.

#### 2) PERFORMANCE PROPORTIONAL INTERGROUP ALLOCATION (PPEA)

The performance proportional intragroup allocation method, used in the context of intragroup allocation, will also be used in the context of intergroup allocation. While in the intragroup allocation this method was applied to give relevance to objects within groups, in the case of intergroup allocation this method is used to give relevance to groups. In the case of intergroup allocation, the performance of all assets in the group in the period analyzed is considered, so that the groups with the best performance have greater relevance in the final allocation, while those with the worst performance have lower relevance in the group.

The performance proportional allocation is made according to Equation 15 and in the same way as Balanced Capital Allocation, this method normalizes allocations between groups so that the final investment will total all available capital for investment.

$$w_{i,j} = \frac{\mu_{i,j}}{\sum_c^n \mu_{c,j}} * \frac{f_j}{\sum_g^m f_g}, \tag{15}$$

where $w_{i,j}$ is the percentage of the $i$-th asset in the final portfolio, $\mu_{i,j}$ is obtained from the intergroup allocation, $\mu_{c,j}$ is used to find the sum of the $j$-th group values, $f_j$ is the metric used to apply to the $j$-th group, and $f_g$ is used to find the sum of the metrics for all groups.

After using the BCA and PPEA methods, the final portfolio is composed and ready for investment.

### F. PERFORMANCE MEASURES

To assess the proposed strategies, the following performance measures will be used:

- **Annualized Return**: This is the rate of return scaled to one year.

$$R = \frac{p(t)}{p(i)}^{(1/n)} - 1 \tag{16}$$

where $p(t)$ is the accumulated value of the portfolio at the observed time, $p(i)$ is the initial value and $n$ is the number of years.

- **Annualized Volatility**: This is a statistical measure of the dispersion of returns for an asset.

$$\sigma_p = \sigma_{daily} * \sqrt{252} \tag{17}$$

where $\sigma_p$ is the volatility of the series in the period, $\sigma_{daily}$ is the standard deviation of the series of daily returns in the period and $\sqrt{252}$ is the annualization of the standard deviation. This measure informs the realized portfolio volatility.

- **Sharpe Ratio**: This is a measure that indicates the risk-adjusted return on investment [19]. The indicator is obtained according to Equation 5:

$$SR = \frac{\mu - r_f}{\sigma}, \tag{18}$$

- **Turnover Ratio**: This is the measure that indicates the portfolio change with each rebalancing.

$$TR = \sum_{i=1}^{N} |w_i(t) - w_i(t-1)| \tag{19}$$

where $w_i(t)$ is the weight of the asset in the current month and $w_i(t-1)$ is the weight of the asset in the previous month.

- **Maximum Drawdown**: This is the number that indicates in percent the greatest fall from peak to valley in the return series. The equation below, presented by Magdon-Ismail [34], is used to find the Maximum Drawdown:

$$MDD(t) = \sup_{t\in[0,T]} [\sup_{s\in[0,t]} X(s) - X(t)], \tag{20}$$

where $X(t)$ represents the portfolio value at time $t$ and $X(s)$ the largest portfolio value in the past at time $t$.

## V. PERFORMANCE ASSESSMENT

The proposed framework in this paper has six sequential phases, namely:

1) Asset Selection;
2) Calculate the correlation and distance matrix between assets;
3) Apply the partitional clustering algorithm;
4) Perform intragroup allocation;
5) Perform intergroup allocation;
6) Evaluate performance.

To evaluate the application of the proposed framework, in Phase (1) it will be used assets traded in the Brazilian Stock Exchange (B3) and which are part of the Ibovespa index in the respective allocation evaluation period. The tests will be performed considering the period from 12/2005 until 04/2020. In Phase (2), the correlation, according to Equation 9, and the distance will be calculated according to Equation 11. The partitional clustering algorithm (Phase (3)) selected was $k$-medoids with two distinct ways of initializing prototypes. Intragroup allocation (Phase (4)) used the FIA and PPAA methods, and intergroup allocation (Phase (5)) used the BCA and PPEA methods. The *framework* performance (Phase (6)) considered the measures presented in Section IV-F.

### A. EXPERIMENTAL METHODOLOGY

Since Phases (3), (4), and (5) of the framework allow for some variations, different combinations for them resulted in different evaluation settings, as follows:

- **Configuration 1**: Random selection of initial medoids, Fuzzy Intragroup Allocation (FIA), and intergroup Balanced Capital Allocation (BCA).
- **Configuration 2**: Use of the volatility heuristics for medoids initialization, Fuzzy Intragroup Allocation (FIA), and intergroup Balanced Capital Allocation (BCA).

**TABLE 1.** Comparison of strategies using a correlation window of 12 months for Configuration 1.

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 8.0% [6.4%, 10.2%] | 26.1% [25.9%, 26.3%] | 51.9% [50.7%, 53.3%] | -53.3% [-54.2%, -51.6%] | -0.1 [-0.2, -0.0] |
| 4 | 8.7% [7.7%, 10.2%] | 26.3% [26.1%, 26.5%] | 60.1% [55.9%, 62.5%] | -54.6% [-56.6%, -51.1%] | -0.1 [-0.1, -0.0] |
| 5 | 7.7% [5.9%, 10.8%] | 26.3% [25.9%, 26.6%] | 66.7% [64.6%, 68.6%] | -54.8% [-61.6%, -52.0%] | -0.1 [-0.2, 0.0] |
| 6 | 8.2% [7.1%, 9.3%] | 26.4% [26.1%, 26.8%] | 71.4% [67.9%, 73.0%] | -53.5% [-56.0%, -50.1%] | -0.1 [-0.1, -0.0] |
| 7 | 7.3% [4.2%, 9.4%] | 26.6% [26.0%, 27.2%] | 76.3% [73.7%, 78.4%] | -55.1% [-58.6%, -51.6%] | -0.1 [-0.2, -0.0] |
| 8 | 7.8% [4.7%, 9.4%] | 26.5% [26.2%, 26.9%] | 80.3% [78.7%, 81.8%] | -54.4% [-56.6%, -52.0%] | -0.1 [-0.2, -0.0] |
| 9 | 8.4% [6.0%, 10.0%] | 26.5% [26.2%, 26.8%] | 82.7% [80.7%, 85.1%] | -53.5% [-55.2%, -51.2%] | -0.1 [-0.2, -0.0] |
| 10 | 8.3% [6.6%, 10.8%] | 26.6% [26.0%, 26.9%] | 85.2% [84.4%, 86.6%] | -53.9% [-58.8%, -48.8%] | -0.1 [-0.1, 0.0] |
| 11 | 9.2% [7.2%, 12.1%] | 26.7% [26.0%, 27.1%] | 87.3% [85.6%, 89.1%] | -54.0% [-56.9%, -51.4%] | -0.0 [-0.1, 0.1] |
| 12 | 8.3% [6.3%, 10.0%] | 26.6% [26.2%, 27.0%] | 89.1% [88.4%, 90.7%] | -53.5% [-56.2%, -52.3%] | -0.1 [-0.2, -0.0] |
| 13 | 8.2% [6.1%, 10.8%] | 26.6% [26.2%, 26.9%] | 90.3% [88.7%, 92.3%] | -54.0% [-58.8%, -51.4%] | -0.1 [-0.2, 0.0] |
| 14 | 8.5% [6.2%, 10.3%] | 26.6% [26.1%, 26.9%] | 91.2% [90.0%, 92.2%] | -53.9% [-56.6%, -52.4%] | -0.1 [-0.2, -0.0] |
| 15 | 9.3% [7.0%, 11.6%] | 26.7% [26.1%, 27.1%] | 92.1% [90.8%, 93.4%] | -53.8% [-55.9%, -52.6%] | -0.0 [-0.1, 0.0] |
| 16 | 8.8% [7.8%, 10.5%] | 26.6% [26.1%, 26.9%] | 92.5% [91.4%, 93.5%] | -54.1% [-57.0%, -51.9%] | -0.1 [-0.1, 0.0] |
| 17 | 8.9% [7.5%, 10.3%] | 26.5% [26.3%, 27.1%] | 92.8% [91.9%, 95.3%] | -53.3% [-56.8%, -51.5%] | -0.1 [-0.1, -0.0] |
| 18 | 9.0% [7.2%, 10.9%] | 26.6% [25.9%, 26.9%] | 92.9% [91.1%, 94.3%] | -53.3% [-56.0%, -50.0%] | -0.1 [-0.1, 0.0] |
| 19 | 9.7% [7.7%, 11.9%] | 26.4% [26.1%, 26.8%] | 92.5% [91.4%, 93.8%] | -52.8% [-54.9%, -50.8%] | -0.0 [-0.1, 0.1] |
| 20 | 8.9% [8.0%, 10.4%] | 26.5% [26.0%, 26.8%] | 92.1% [90.0%, 94.4%] | -53.4% [-55.4%, -50.8%] | -0.1 [-0.1, 0.0] |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

- **Configuration 3**: Use of volatility heuristics for medoid initialization, Fuzzy Intragroup Allocation (FIA) and Performance Proportional Intergroup Allocation (PPEA).
- **Configuration 4**: Use of volatility heuristics for medoid initialization, Performance Proportional Intragroup Allocation (PPAA) and intergroup Balanced Capital Allocation (BCA).
- **Configuration 5**: Use of volatility heuristics for the medoid initialization, and Performance Proportional Intragroup and Intergroup Allocation.

The portfolio allocation process using partitional clustering is performed monthly, with rebalancing in the first business day of each month. For the execution of the clustering algorithm, at the time of rebalancing it will be considered the closing price information in a previous period of 12 or 24 months. From the information of these periods, the correlation is calculated. Since it is not possible to know a priori the optimal number of groups in partitional clustering, portfolio allocations will be performed using values of $k$ between 3 and 20. At each rebalancing, an asset set is required to calculate the correlation between the assets and then perform the partitional clustering. The assets that will be part of each rebalancing are those that are part of the theoretical Ibovespa index portfolio in the respective rebalancing assessment month.

For Configuration 1, the results present the mean, worst and best values using the following format: Average [Worst, Best], all taken over 10 runs. For all other configurations, the proposed volatility-based heuristics is used to initialize the medoids and, thus, the algorithm becomes deterministic. In such cases, the single value obtained for each configuration is presented. For comparison purposes, the results of the Ibovespa index for the period, the performance of a portfolio using the mean-variance (MVP) model proposed by Markowitz [1] and the Hierarchical Risk Parity (HRP) model proposed by de Prado [4] are also presented. The portfolios were built by analyzing 12 and 24 month correlation windows and monthly rebalancing. The tests were always performed using a 100% purchased portfolio and the the results were organized in tables considering each value of $k \in [3, 20]$.

### B. RESULTS AND DISCUSSION

#### 1) CONFIGURATION 1: RANDOM SELECTION OF INITIAL MEDOIDS, FUZZY INTRAGROUP ALLOCATION (FIA), AND INTEGROUP BALANCED CAPITAL ALLOCATION (BCA)

Table 1 presents the results for strategies that use 12 months of asset price information to calculate the correlation and Table 2 presents the results for the same configuration, but the strategies were calculated using 24 months of price information to calculate the correlation. Since the $k$ -medoids algorithm is stochastic and sensitive to the initial condition, 10 simulations were performed for each $k$ value with the initialization of the medoids.

From the analysis of the results of the strategies, it is possible to observe that the performance is greatly impacted by the initial medoids, because there is a big difference between the best and the worst result for the same value of $k$.

It is possible to observe in Tables 1 and 2 that the average test result is better than the Ibovespa index. Volatility remained close to that observed in the Ibovespa index and the *Turnover ratio* increases as the value of $k$ increases, but the average return is better in strategies with a higher $k$ value. The average drawdown is close to the Ibovespa index and, in some settings, this value is higher.

#### 2) CONFIGURATION 2: USE OF VOLATILITY HEURISTICS FOR MEDOIDS INITIALIZATION, FUZZY INTRAGROUP ALLOCATION (FIA), AND INTERGROUP BALANCED CAPITAL ALLOCATION (BCA)

Given the sensitivity of the $k$-medoids algorithm to the initial medoids, an initialization heuristic was proposed taking as

**TABLE 2. Comparison of strategies using a correlation window of 24 months for Configuration 1.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 9.2% [7.2%, 12.0%] | 26.4% [26.1%, 26.7%] | 50.8% [48.4%, 52.7%] | -53.2% [-56.9%, -51.2%] | -0.0 [-0.1, 0.1] |
| 4 | 8.9% [6.6%, 11.7%] | 26.4% [26.0%, 26.7%] | 60.4% [58.9%, 62.0%] | -53.3% [-56.5%, -51.3%] | -0.1 [-0.1, 0.1] |
| 5 | 9.2% [7.4%, 12.1%] | 26.7% [26.3%, 27.1%] | 66.5% [62.6%, 70.0%] | -53.2% [-56.6%, -49.7%] | -0.0 [-0.1, 0.1] |
| 6 | 8.9% [5.4%, 11.2%] | 26.5% [26.2%, 26.8%] | 71.2% [70.2%, 73.1%] | -53.7% [-58.7%, -51.0%] | -0.1 [-0.2, 0.0] |
| 7 | 8.6% [7.3%, 10.7%] | 26.7% [26.2%, 27.3%] | 77.4% [75.4%, 78.5%] | -54.8% [-60.2%, -50.9%] | -0.1 [-0.1, 0.0] |
| 8 | 9.2% [7.0%, 11.5%] | 26.7% [26.3%, 27.2%] | 80.0% [76.7%, 81.9%] | -54.6% [-58.7%, -51.1%] | -0.0 [-0.1, 0.0] |
| 9 | 10.2% [8.1%, 11.9%] | 26.8% [26.3%, 27.0%] | 82.3% [80.9%, 84.3%] | -54.3% [-57.9%, -52.2%] | -0.0 [-0.1, 0.1] |
| 10 | 8.6% [5.3%, 11.3%] | 26.8% [26.5%, 27.1%] | 85.3% [83.6%, 86.9%] | -55.0% [-59.1%, -50.6%] | -0.1 [-0.2, 0.0] |
| 11 | 9.9% [7.1%, 12.4%] | 26.6% [26.0%, 27.1%] | 87.3% [86.0%, 88.8%] | -53.7% [-57.3%, -50.1%] | -0.0 [-0.1, 0.1] |
| 12 | 8.7% [6.9%, 10.4%] | 26.6% [26.1%, 27.1%] | 87.5% [84.0%, 90.2%] | -53.9% [-56.4%, -52.3%] | -0.1 [-0.1, 0.0] |
| 13 | 9.3% [7.0%, 10.9%] | 26.8% [26.2%, 27.4%] | 89.3% [88.2%, 90.9%] | -53.9% [-57.3%, -50.6%] | -0.0 [-0.1, 0.0] |
| 14 | 9.8% [8.8%, 12.7%] | 26.7% [26.3%, 27.0%] | 90.7% [89.2%, 92.6%] | -53.8% [-55.1%, -52.1%] | -0.0 [-0.1, 0.1] |
| 15 | 9.1% [6.3%, 11.5%] | 26.7% [26.1%, 27.2%] | 90.9% [89.2%, 92.4%] | -54.2% [-58.5%, -51.3%] | -0.0 [-0.2, 0.0] |
| 16 | 9.2% [5.5%, 12.3%] | 26.7% [26.2%, 27.1%] | 91.5% [90.1%, 92.7%] | -53.7% [-58.2%, -52.0%] | -0.0 [-0.2, 0.1] |
| 17 | 9.0% [6.9%, 12.3%] | 26.5% [26.1%, 26.9%] | 91.7% [89.2%, 93.6%] | -53.9% [-55.9%, -51.0%] | -0.1 [-0.1, 0.1] |
| 18 | 8.9% [6.4%, 10.0%] | 26.5% [26.2%, 26.8%] | 90.9% [89.0%, 93.0%] | -54.2% [-55.3%, -51.9%] | -0.1 [-0.2, -0.0] |
| 19 | 9.5% [8.5%, 10.6%] | 26.7% [26.3%, 26.9%] | 91.2% [89.5%, 92.4%] | -53.9% [-56.4%, -52.1%] | -0.0 [-0.1, 0.0] |
| 20 | 9.4% [6.5%, 11.1%] | 26.4% [26.1%, 26.6%] | 90.7% [89.8%, 92.6%] | -52.6% [-55.2%, -50.1%] | -0.0 [-0.1, 0.0] |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

**TABLE 3. Comparison of strategies using a correlation window of 12 months for Configuration 2.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 8.4% | 25.6% | 41.4% | -53.7% | -0.1 |
| 4 | 9.9% | 25.4% | 46.9% | -53.0% | -0.0 |
| 5 | 11.3% | 24.3% | 49.9% | -50.4% | 0.0 |
| 6 | 11.5% | 24.0% | 55.2% | -49.6% | 0.0 |
| 7 | 10.8% | 23.7% | 58.0% | -47.9% | 0.0 |
| 8 | 10.3% | 23.8% | 59.4% | -48.8% | -0.0 |
| 9 | 10.2% | 24.1% | 62.6% | -50.3% | -0.0 |
| 10 | 9.1% | 23.9% | 62.2% | -49.0% | -0.1 |
| 11 | 10.6% | 23.9% | 62.1% | -50.6% | 0.0 |
| 12 | 10.7% | 23.7% | 63.1% | -50.0% | 0.0 |
| 13 | 10.6% | 23.5% | 64.3% | -49.2% | 0.0 |
| 14 | 10.5% | 23.4% | 63.9% | -47.7% | 0.0 |
| 15 | 10.9% | 23.3% | 61.6% | -48.1% | 0.0 |
| 16 | 10.5% | 23.7% | 61.3% | -50.2% | 0.0 |
| 17 | 10.8% | 23.6% | 61.1% | -50.7% | 0.0 |
| 18 | 10.9% | 23.7% | 60.0% | -50.4% | 0.0 |
| 19 | 11.1% | 23.7% | 59.0% | -50.7% | 0.0 |
| 20 | 10.8% | 23.6% | 58.2% | -49.7% | 0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

**TABLE 4. Comparison of strategies using a correlation window of 24 months for Configuration 2.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 9.4% | 26.1% | 32.9% | -52.5% | -0.0 |
| 4 | 9.2% | 25.2% | 42.6% | -51.9% | -0.0 |
| 5 | 9.0% | 24.7% | 44.7% | -52.1% | -0.1 |
| 6 | 10.0% | 24.8% | 47.8% | -51.8% | -0.0 |
| 7 | 9.9% | 24.7% | 51.2% | -52.3% | -0.0 |
| 8 | 9.7% | 24.2% | 52.3% | -52.7% | -0.0 |
| 9 | 9.2% | 24.1% | 53.7% | -50.2% | -0.0 |
| 10 | 10.1% | 24.1% | 54.8% | -50.1% | -0.0 |
| 11 | 10.6% | 24.1% | 53.1% | -51.1% | 0.0 |
| 12 | 10.2% | 24.2% | 51.5% | -53.4% | -0.0 |
| 13 | 11.1% | 24.1% | 51.6% | -50.9% | 0.0 |
| 14 | 11.7% | 23.9% | 51.0% | -49.4% | 0.1 |
| 15 | 10.5% | 23.7% | 49.7% | -52.1% | 0.0 |
| 16 | 11.7% | 23.9% | 48.9% | -50.8% | 0.1 |
| 17 | 11.5% | 23.9% | 47.8% | -52.9% | 0.0 |
| 18 | 10.8% | 23.9% | 46.9% | -52.3% | 0.0 |
| 19 | 10.3% | 24.0% | 47.2% | -52.2% | -0.0 |
| 20 | 10.6% | 23.9% | 45.8% | -50.5% | 0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

initial medoids the assets that presented the lowest volatility in the analyzed sample for the correlation calculation period.

Table 3 shows the test results that were performed using 12 months of price information to calculate the correlation. For intragroup allocation, the balanced allocation of capital between groups was used and for the intergroup allocation the fuzzy process was used. Strategies with values of $k$ between 3 and 20 were tested. Table 4 uses the same settings with 24 month correlation analysis.

It is possible to observe from Table 4 that as the value of $k$ increases, the return increases, and the *drawdown* decreases. It is also possible to observe that portfolio volatility is lower when the $k$ value is higher and the *turnover ratio* of the tested portfolios decreases as the $k$ value increases. When compared to Configuration 1 tests, this value is smaller.

When comparing the test results with the Ibovespa index, the returns are higher, the volatility is lower and the *drawdown* is also lower, being possible to observe the phenomenon cited by [30] and the allocation performing the expected behavior, according to the adopted heuristic.

It is understood that the higher return and lower drawdown portfolios for the highest values of $k$ are the result of lower maximum allocation in a single asset. The maximum allocation a stock can have in a portfolio, in cases where the group consists of only a single stock, is $1/k$, where $k$ is the number of groups. Thus, the higher the value of $k$, the lower the maximum allocation.

**TABLE 5.** Comparison of strategies using a correlation window of 12 months for Configuration 3.

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 8.8% | 24.9% | 40.9% | -54.1% | -0.1 |
| 4 | 10.1% | 24.6% | 46.2% | -53.0% | -0.0 |
| 5 | 11.6% | 23.7% | 49.4% | -49.9% | 0.1 |
| 6 | 11.9% | 23.5% | 55.1% | -48.5% | 0.1 |
| 7 | 11.1% | 23.2% | 58.0% | -47.0% | 0.0 |
| 8 | 10.3% | 23.4% | 59.0% | -47.6% | -0.0 |
| 9 | 10.1% | 23.6% | 62.1% | -48.9% | -0.0 |
| 10 | 9.7% | 23.5% | 61.7% | -48.8% | -0.0 |
| 11 | 11.3% | 23.4% | 62.1% | -50.2% | 0.0 |
| 12 | 11.5% | 23.2% | 62.9% | -49.7% | 0.0 |
| 13 | 11.2% | 23.1% | 64.1% | -48.7% | 0.0 |
| 14 | 11.2% | 23.0% | 63.8% | -48.2% | 0.0 |
| 15 | 11.5% | 22.9% | 61.6% | -48.6% | 0.0 |
| 16 | 10.9% | 23.3% | 61.6% | -49.9% | 0.0 |
| 17 | 11.1% | 23.1% | 61.3% | -49.8% | 0.0 |
| 18 | 11.2% | 23.2% | 60.1% | -49.8% | 0.0 |
| 19 | 11.4% | 23.2% | 59.3% | -49.9% | 0.0 |
| 20 | 11.0% | 23.1% | 58.5% | -49.7% | 0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

The results observed in the 24-month window tests, in Table 4, also show portfolios with smaller *turnover*, higher return and lower *drawdown* as the value of $k$ increases.

### 3) CONFIGURATION 3: USE OF VOLATILITY HEURISTICS FOR MEDOIDS INITIALIZATION, FUZZY INTRAGROUP ALLOCATION (FIA), AND PERFORMANCE PROPORTIONAL INTERGROUP ALLOCATION (PPEA)

Test Configuration 3 is very similar to Configuration 2, with the difference in the intergroup allocation. While in Configuration 2 the allocation was balanced against the number of groups, in Configuration 3 the allocation is inspired by genetic algorithms and the allocation ratio in each group follows the logic of the genetic operator known as roulette selection [31]. In this case, the group's performance is calculated considering that capital was allocated in a balanced manner across all assets of that group in the prior period used for the cluster analysis. After calculating group performance, capital is allocated to each group, where the best performing group receives the most allocation, while the worst performing group receives the least.

Table 5 presents results using 12 months of information and price for correlation calculation, while Table 6 presents results using 24 months of correlation.

When looking at the results from the Configuration 3, portfolios using the 12-month correlation window compared to the Configuration 2, it can be seen that the change in intergroup allocation improved the portfolio results. The produced portfolios have higher return and lower volatility, turnover and drawdown. Compared to the Ibovespa index, the results of the portfolios generated are significantly better from the point of view of return and risk.

**TABLE 6.** Comparison of strategies using a correlation window of 24 months for Configuration 3.

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 9.7% | 25.4% | 32.3% | -53.5% | -0.0 |
| 4 | 9.0% | 24.4% | 41.2% | -52.7% | -0.1 |
| 5 | 8.9% | 24.1% | 44.0% | -53.1% | -0.1 |
| 6 | 9.6% | 24.1% | 47.4% | -52.6% | -0.0 |
| 7 | 9.1% | 24.0% | 51.2% | -53.5% | -0.1 |
| 8 | 9.1% | 23.6% | 52.6% | -52.8% | -0.1 |
| 9 | 9.1% | 23.5% | 54.4% | -50.1% | -0.1 |
| 10 | 9.7% | 23.6% | 55.3% | -49.9% | -0.0 |
| 11 | 10.8% | 23.5% | 54.0% | -50.8% | 0.0 |
| 12 | 9.9% | 23.8% | 52.6% | -52.9% | -0.0 |
| 13 | 10.9% | 23.7% | 53.0% | -51.0% | 0.0 |
| 14 | 11.3% | 23.5% | 52.6% | -49.9% | 0.0 |
| 15 | 10.4% | 23.2% | 51.5% | -51.8% | 0.0 |
| 16 | 11.7% | 23.3% | 50.6% | -50.0% | 0.1 |
| 17 | 11.1% | 23.4% | 49.1% | -52.9% | 0.0 |
| 18 | 10.8% | 23.4% | 48.4% | -52.3% | 0.0 |
| 19 | 10.1% | 23.6% | 48.7% | -52.7% | -0.0 |
| 20 | 10.6% | 23.4% | 47.5% | -50.5% | 0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

**TABLE 7.** Comparison of strategies using a correlation window of 12 months for Configuration 4.

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 8.5% | 25.1% | 39.0% | -53.0% | -0.1 |
| 4 | 9.9% | 24.9% | 42.5% | -52.4% | -0.0 |
| 5 | 10.5% | 24.2% | 43.7% | -49.9% | 0.0 |
| 6 | 10.6% | 23.7% | 47.4% | -48.9% | 0.0 |
| 7 | 10.4% | 23.7% | 48.5% | -49.4% | 0.0 |
| 8 | 9.8% | 23.7% | 49.5% | -49.5% | -0.0 |
| 9 | 10.0% | 23.9% | 50.9% | -50.5% | -0.0 |
| 10 | 9.1% | 23.8% | 50.2% | -50.0% | -0.1 |
| 11 | 10.1% | 23.7% | 49.7% | -49.9% | -0.0 |
| 12 | 9.8% | 23.5% | 50.2% | -49.4% | -0.0 |
| 13 | 9.9% | 23.5% | 49.3% | -49.4% | -0.0 |
| 14 | 10.2% | 23.5% | 49.9% | -49.1% | -0.0 |
| 15 | 10.9% | 23.4% | 48.7% | -49.3% | 0.0 |
| 16 | 10.0% | 23.7% | 48.5% | -50.7% | -0.0 |
| 17 | 10.4% | 23.7% | 47.5% | -51.1% | -0.0 |
| 18 | 10.3% | 23.7% | 46.9% | -50.7% | -0.0 |
| 19 | 10.5% | 23.7% | 46.2% | -50.2% | 0.0 |
| 20 | 10.2% | 23.7% | 45.3% | -49.9% | -0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

### 4) CONFIGURATION 4: USE OF VOLATILITY HEURISTICS FOR MEDOIDS INITIALIZATION, PERFORMANCE PROPORTIONAL INTRAGROUP ALLOCATION (PPIA) AND INTERGROUP BALANCED CAPITAL ALLOCATION (BCA)

Configuration 4 of the tests is similar to Configuration 2, both using the initial medoid selection heuristic from lower volatility stocks and the intragroup balanced allocation method. In this configuration, portfolios were created using the intergroup performance proportional method and capital is allocated to each group using Equation 13.

Table 7 shows the results using 12 months of price information for correlation calculation while Table 8 presents results using 24 months of correlation.

**TABLE 8. Comparison of strategies using a correlation window of 24 months for Configuration 4.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 10.0% | 25.5% | 33.2% | -51.4% | -0.0 |
| 4 | 9.9% | 24.8% | 41.3% | -50.5% | -0.0 |
| 5 | 10.5% | 24.5% | 41.5% | -50.3% | 0.0 |
| 6 | 10.3% | 24.5% | 43.6% | -50.7% | -0.0 |
| 7 | 10.4% | 24.4% | 44.8% | -50.0% | 0.0 |
| 8 | 10.4% | 24.1% | 44.9% | -50.1% | 0.0 |
| 9 | 9.8% | 23.9% | 45.7% | -50.3% | -0.0 |
| 10 | 10.5% | 23.9% | 45.7% | -49.6% | 0.0 |
| 11 | 11.1% | 23.8% | 45.1% | -50.3% | 0.0 |
| 12 | 10.6% | 23.7% | 43.1% | -50.5% | 0.0 |
| 13 | 11.3% | 23.8% | 43.2% | -49.2% | 0.0 |
| 14 | 12.1% | 23.6% | 42.2% | -48.3% | 0.1 |
| 15 | 11.3% | 23.7% | 40.6% | -50.5% | 0.0 |
| 16 | 11.9% | 23.8% | 39.7% | -50.2% | 0.1 |
| 17 | 11.5% | 23.9% | 39.3% | -51.1% | 0.0 |
| 18 | 10.8% | 23.8% | 38.7% | -51.1% | 0.0 |
| 19 | 10.2% | 23.9% | 38.5% | -51.0% | -0.0 |
| 20 | 10.3% | 23.8% | 36.7% | -50.2% | -0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

When looking at the results from the Configuration 4, portfolios using the 12-month correlation window compared to Configuration 2, where both use intergroup balanced allocation, it can be seen that the results from Configuration 4 are only slightly better.

The portfolios produced in Configuration 4 have lower drawdown compared to the Ibovespa index, the results of the generated portfolios are significantly better from the point of view of return and risk.

### 5) CONFIGURATION 5: USE OF VOLATILITY HEURISTICS FOR MEDOID INITIALIZATION, PERFORMANCE PROPORTIONAL INTRAGROUP ALLOCATION (PPAA) AND PERFORMANCE PROPORTIONAL INTERGROUP ALLOCATION (PPEA)

Configuration 5 of the tests is similar to Configuration 3, both use the initial medoid selection heuristic from lower volatility stocks, and also use the performance proportional method to allocate intragroup and intergroup capital. In this case, the shares of each group that had the highest return in the analyzed window, will have greater allocation in the generated portfolio.

Table 9 presents the results of portfolios calculated from 12 months of correlation, while Table 10 presents the results for portfolios calculated from 24 months of correlation.

When compared with the Ibovespa index, return is higher, volatility and drawdown are lower for higher values of $k$, showing better risk-return ratio, which can be observed with larger values of the *sharpe ratio*.

### 6) THE FRAMEWORK PERFORMANCE IN ATYPICAL SCENARIOS

In the tested period, the stock market went through two major global crises. The Subprime crisis in 2008 and the crisis due to COVID-19 in 2020.

In the period 2008 and 2009, considering the recovery, the portfolios of Configurations 2 to 5, for most of the $k$ values, performed better than the Ibovespa index, showing a higher return, lower volatility and drawdown.

For the year 2020, the portfolios of the respective configurations showed a return close to the Ibovespa index and a higher drawdown. The poor performance in these periods, close to the market, is due to the fact that the framework is a systematic stock portfolio allocation tool, that is, the portfolio generated by the framework will always be allocated in stocks, being subject to wide variations, both positive and negative.

**TABLE 9. Comparison of strategies using a correlation window of 12 months for Configuration 5.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 8.9% | 24.7% | 36.9% | -53.2% | -0.1 |
| 4 | 10.3% | 24.4% | 40.9% | -52.3% | -0.0 |
| 5 | 10.8% | 23.8% | 42.2% | -49.4% | 0.0 |
| 6 | 10.8% | 23.4% | 46.1% | -48.4% | 0.0 |
| 7 | 10.6% | 23.4% | 47.2% | -48.9% | 0.0 |
| 8 | 9.8% | 23.4% | 48.4% | -48.9% | -0.0 |
| 9 | 10.0% | 23.6% | 49.8% | -49.8% | -0.0 |
| 10 | 9.5% | 23.5% | 49.4% | -49.3% | -0.0 |
| 11 | 10.6% | 23.4% | 49.2% | -49.2% | 0.0 |
| 12 | 10.3% | 23.3% | 49.8% | -49.0% | -0.0 |
| 13 | 10.3% | 23.2% | 48.8% | -48.9% | -0.0 |
| 14 | 10.5% | 23.3% | 49.3% | -48.7% | 0.0 |
| 15 | 11.2% | 23.2% | 48.0% | -48.8% | 0.0 |
| 16 | 10.3% | 23.4% | 48.2% | -50.3% | -0.0 |
| 17 | 10.5% | 23.4% | 47.3% | -50.6% | 0.0 |
| 18 | 10.6% | 23.4% | 46.5% | -50.2% | 0.0 |
| 19 | 10.7% | 23.4% | 46.1% | -50.0% | 0.0 |
| 20 | 10.4% | 23.4% | 45.2% | -49.6% | -0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

**TABLE 10. Comparison of strategies using a correlation window of 24 months for Configuration 5.**

| K | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|
| 3 | 10.3% | 25.1% | 29.6% | -52.2% | -0.0 |
| 4 | 9.8% | 24.4% | 37.8% | -51.2% | -0.0 |
| 5 | 10.2% | 24.2% | 38.5% | -51.3% | -0.0 |
| 6 | 10.3% | 24.1% | 40.7% | -50.8% | -0.0 |
| 7 | 10.5% | 24.0% | 42.9% | -50.8% | 0.0 |
| 8 | 10.7% | 23.7% | 43.5% | -49.8% | 0.0 |
| 9 | 10.2% | 23.6% | 44.8% | -49.5% | -0.0 |
| 10 | 10.6% | 23.6% | 45.1% | -49.3% | 0.0 |
| 11 | 11.6% | 23.6% | 44.6% | -50.2% | 0.1 |
| 12 | 10.9% | 23.6% | 42.8% | -50.2% | 0.0 |
| 13 | 11.3% | 23.6% | 43.2% | -49.3% | 0.0 |
| 14 | 11.9% | 23.5% | 42.0% | -48.9% | 0.1 |
| 15 | 11.3% | 23.5% | 40.7% | -50.7% | 0.0 |
| 16 | 11.8% | 23.5% | 39.6% | -50.0% | 0.1 |
| 17 | 11.2% | 23.5% | 39.0% | -51.3% | 0.0 |
| 18 | 10.8% | 23.5% | 38.6% | -51.4% | 0.0 |
| 19 | 10.2% | 23.6% | 38.2% | -51.5% | -0.0 |
| 20 | 10.4% | 23.5% | 36.6% | -50.4% | -0.0 |
| Ibovespa | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

### 7) PORTFOLIO COMPARISON

Tables 11 and 12 give the best results found for configurations 2 through 5, which use the same medoid initialization method. Table 11 presents the best results that used 12 months

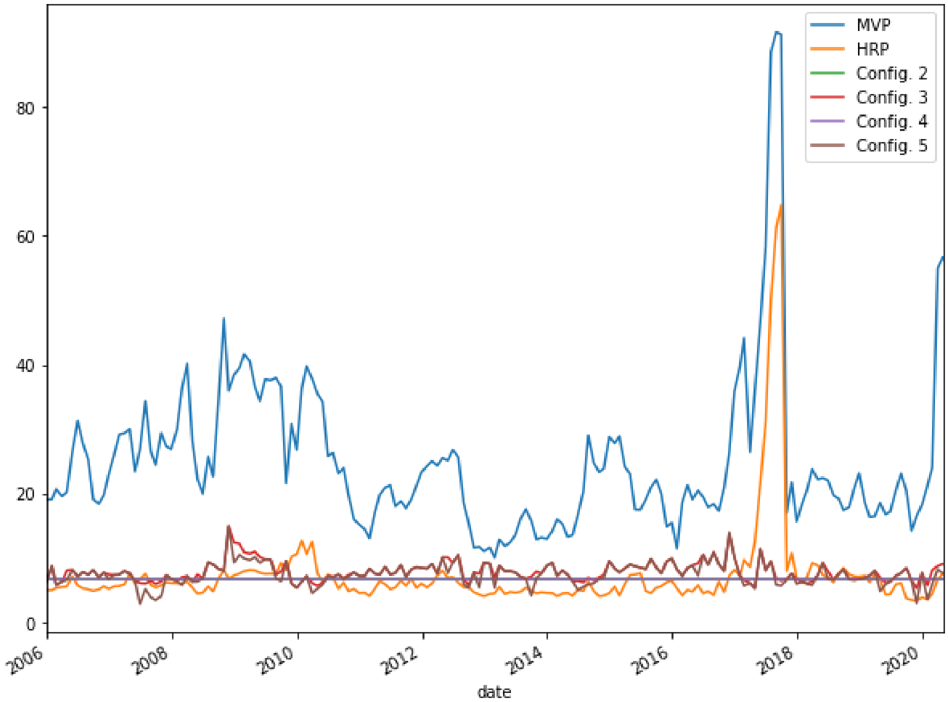


**FIGURE 2.** Monthly evolution of the maximal percentage of each portfolio.

**TABLE 11.** Best results for each configuration using a correlation window of 12 months.

| Config | k | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|---|
| Config. 2 | 19 | 11.1% | 23.7% | 59.0% | -50.7% | 0.0 |
| Config. 3 | 12 | 11.5% | 23.2% | 62.9% | -49.7% | 0.0 |
| Config. 4 | 15 | 10.9% | 23.4% | 48.7% | -49.3% | 0.0 |
| Config. 5 | 15 | 11.2% | 23.2% | 48.0% | -48.8% | 0.0 |
| Ibovespa | | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | | 9.1% | 22.3% | 22.5% | -47.7% | -0.1 |
| MVP | | 8.0% | 17.4% | 32.3% | -44.9 % | -0.1 |

**TABLE 12.** Best results for each configuration using a correlation window of 24 months.

| Config | k | Return | Volatility | Turnover | Drawdown | Sharpe |
|---|---|---|---|---|---|---|
| Config. 2 | 16 | 11.7% | 23.9% | 48.9% | -50.8% | 0.1 |
| Config. 3 | 16 | 11.7% | 23.3% | 50.6% | -50.0% | 0.1 |
| Config. 4 | 14 | 12.1% | 23.6% | 42.2% | -48.3% | 0.1 |
| Config. 5 | 14 | 11.9% | 23.5% | 42.0% | -48.9% | 0.1 |
| Ibovespa | | 6.3% | 28.3% | - | -60.0% | -0.2 |
| HRP | | 9.1% | 22.7% | 19.5% | -48.7 % | -0.1 |
| MVP | | 8.0% | 17.4% | 30.9% | -42.3 % | -0.1 |

of information to calculate the correlation and Table 12 presents the best results that used 24 months of information to calculate the correlation.

In Configurations 2, 3, 4 and 5 you can see that as the value of $k$ increases, volatility decreases significantly, making these portfolios have a lower *drawdown*. Portfolios tested with lower $k$ values tend to be more concentrated, less diversified and exposed to fluctuations of few assets.

Starting from Configuration 2, the heuristic for the selection of initial medoids was always the same, using stocks with lower historical volatility to start the group search. Despite the changes in intragroup and intergroup allocation methods that result in portfolios with different results in Configurations 2, 3, 4, and 5, it can be concluded that the initial medoid selection heuristic had a significant impact on portfolio performance. When comparing the results of Configuration 1 portfolios with the Configuration 2 portfolios, it is possible to note the change in the result and the behavior of the portfolios. Thus, we can conclude that the medoids chosen to initialize the search for the algorithm are very important for the task execution, influencing the final result of the grouping and portfolio formation.

When comparing the portfolio result created using the Markowitz [1] model with the portfolio result in each configuration, it is possible to observe for MVP that the volatility and drawdown are lower. However, by looking at the capital distribution in the assets of the MVP portfolio compared to the HRP method and the test setup methods, it can be seen that the MVP portfolio concentrates most of the allocation on a few stocks, leaving the portfolio exposed to fluctuations of some stocks.

In Figure 2 you can see the monthly evolution of the maximum percentages in each portfolio. The Configurations 2, 3, 4 and 5 portfolios used for this analysis were created considering 12 month correlation information between assets and 15 groups as $k$ value in the $k$-medoids grouping. For portfolios created in Configurations 2 and 4, the intergroup balanced allocation method was used. Thus, the maximum weight an asset can have with this allocation is $1/k$, where grouping with $k$-medoids finds groups with only one asset. Then you can see in the graph a straight line for these two

**TABLE 13.** Portfolio.

| Ticker | MVP | HRP | Config. 2 | Config. 3 | Config. 4 | Config. 5 |
|---|---|---|---|---|---|---|
| ABEV3 | 17.9% | 4.2% | 6.1% | 5.3% | 3.2% | 2.7% |
| B3SA3 | 0.0% | 1.1% | 0.2% | 0.2% | 0.5% | 0.5% |
| BBAS3 | 0.0% | 0.4% | 0.3% | 0.3% | 0.3% | 0.3% |
| BBDC3 | 0.0% | 0.8% | 0.4% | 0.4% | 0.3% | 0.3% |
| BBDC4 | 0.0% | 0.6% | 0.5% | 0.5% | 0.3% | 0.3% |
| BBSE3 | 0.0% | 0.6% | 0.2% | 0.2% | 0.2% | 0.3% |
| BRAP4 | 0.0% | 0.6% | 1.1% | 1.1% | 0.9% | 0.9% |
| BRFS3 | 5.3% | 3.8% | 6.2% | 3.9% | 3.4% | 2.1% |
| BRKM5 | 3.6% | 3.2% | 6.7% | 8.6% | 6.7% | 8.6% |
| BRML3 | 0.0% | 1.6% | 0.2% | 0.2% | 0.4% | 0.4% |
| CCRO3 | 0.0% | 1.1% | 0.2% | 0.2% | 0.3% | 0.3% |
| CESP6 | 0.1% | 1.4% | 0.4% | 0.4% | 0.7% | 0.7% |
| CIEL3 | 0.0% | 1.5% | 6.0% | 5.7% | 3.0% | 2.8% |
| CMIG4 | 0.0% | 0.5% | 0.2% | 0.2% | 0.3% | 0.3% |
| CPFE3 | 0.0% | 1.2% | 0.5% | 0.5% | 1.2% | 1.1% |
| CPLE6 | 0.0% | 0.8% | 0.2% | 0.2% | 0.2% | 0.2% |
| CSAN3 | 0.1% | 1.7% | 6.7% | 9.7% | 6.7% | 9.7% |
| CSNA3 | 0.0% | 0.2% | 0.3% | 0.3% | 1.9% | 2.0% |
| CTIP3 | 13.4% | 4.7% | 5.8% | 6.1% | 2.2% | 2.3% |
| CYRE3 | 0.2% | 1.3% | 0.2% | 0.2% | 0.3% | 0.3% |
| ECOR3 | 0.0% | 0.7% | 0.2% | 0.2% | 0.3% | 0.3% |
| EGIE3 | 11.3% | 3.3% | 4.6% | 4.5% | 1.1% | 1.1% |
| EMBR3 | 2.0% | 1.7% | 0.4% | 0.2% | 1.7% | 0.9% |
| ENBR3 | 0.0% | 1.6% | 6.7% | 6.5% | 6.7% | 6.5% |
| EQTL3 | 14.4% | 4.9% | 6.7% | 9.0% | 6.7% | 9.0% |
| ESTC3 | 0.2% | 1.0% | 0.3% | 0.3% | 1.3% | 1.2% |
| FIBR3 | 10.3% | 3.1% | 1.2% | 0.6% | 1.2% | 0.6% |
| GGBR4 | 0.0% | 0.4% | 0.4% | 0.4% | 1.0% | 1.1% |
| GOAU4 | 0.0% | 0.3% | 0.3% | 0.3% | 0.6% | 0.6% |
| HYPE3 | 5.0% | 4.5% | 0.5% | 0.5% | 2.4% | 2.5% |
| ITSA4 | 0.0% | 1.1% | 0.7% | 0.7% | 0.3% | 0.3% |
| ITUB4 | 0.0% | 0.9% | 1.9% | 2.0% | 0.3% | 0.3% |
| JBSS3 | 0.4% | 1.4% | 0.5% | 0.3% | 3.3% | 2.0% |
| KLBN11 | 0.0% | 2.3% | 0.4% | 0.2% | 2.2% | 1.1% |
| KROT3 | 0.0% | 1.1% | 0.2% | 0.2% | 0.4% | 0.4% |
| LAME4 | 0.0% | 1.1% | 0.2% | 0.2% | 0.3% | 0.4% |
| LREN3 | 0.0% | 1.2% | 0.6% | 0.6% | 3.7% | 3.5% |
| MRFG3 | 4.9% | 3.3% | 0.4% | 0.3% | 1.0% | 1.0% |
| MRVE3 | 0.0% | 2.4% | 0.2% | 0.2% | 0.5% | 0.5% |
| MULT3 | 1.8% | 1.6% | 0.2% | 0.2% | 0.3% | 0.4% |
| NATU3 | 0.0% | 1.5% | 0.3% | 0.3% | 1.3% | 1.2% |
| PCAR4 | 0.0% | 1.6% | 0.3% | 0.3% | 0.7% | 0.6% |
| PETR3 | 0.0% | 0.3% | 1.7% | 1.6% | 1.2% | 1.1% |
| PETR4 | 0.0% | 0.3% | 3.6% | 3.3% | 1.1% | 1.0% |
| QUAL3 | 1.7% | 2.6% | 0.4% | 0.5% | 2.1% | 2.2% |
| RADL3 | 3.4% | 2.9% | 6.7% | 8.5% | 6.7% | 8.5% |
| RENT3 | 0.0% | 1.7% | 0.2% | 0.2% | 0.4% | 0.4% |
| RUMO3 | 0.0% | 0.7% | 0.1% | 0.1% | 0.2% | 0.2% |
| SANB11 | 0.0% | 1.0% | 0.3% | 0.3% | 0.4% | 0.4% |
| SBSP3 | 0.0% | 1.5% | 0.4% | 0.4% | 1.7% | 1.6% |
| SMLS3 | 0.0% | 1.2% | 0.4% | 0.4% | 1.0% | 0.9% |
| SUZB5 | 3.3% | 3.6% | 4.6% | 2.4% | 1.6% | 0.8% |
| TIMP3 | 0.0% | 1.6% | 0.2% | 0.2% | 0.7% | 0.7% |
| UGPA3 | 0.0% | 3.3% | 0.5% | 0.4% | 3.5% | 3.0% |
| USIM5 | 0.0% | 0.2% | 0.4% | 0.3% | 1.0% | 0.9% |
| VALE3 | 0.0% | 0.3% | 1.6% | 1.6% | 0.8% | 0.8% |
| VALE5 | 0.0% | 0.4% | 2.8% | 2.9% | 0.8% | 0.8% |
| VIVT4 | 0.2% | 3.3% | 0.2% | 0.2% | 0.3% | 0.3% |
| WEGE3 | 0.1% | 2.9% | 6.7% | 4.6% | 6.7% | 4.6% |

configurations, because in every month that grouping was performed, groups with a single asset were found.

In Configurations 3 and 5 it is possible to observe that the maximum percentages change with each execution of the grouping. This is due to the performance proportional intergroup allocation, where the best performing groups observed in the analyzed period have greater relevance in the final allocation. Although in intragroup allocation the two configurations use different metrics, the maximum weight presents small differences between the two configurations because the metric used for intergroup allocation was the same.

By observing the maximum percentage of portfolios obtained from the optimization proposed by Markowitz (MVP), it can be noted that the portfolios obtained concentrate a very high percentage on a single asset. The average over the maximum percentage months for MVP is 22%. The high concentration in 2017 for the MVP portfolio is explained by the natural characteristic of the model, as optimization considers risk and return. As historical volatility was used in this portfolio, the optimization model concentrated more allocation on an asset that had very low volatility due to a delisting dynamics, that is, exit from the stock market. HRP allocation, in turn, considers the hierarchical structure and volatility between nodes. The same company because of its volatility had greater relevance in HRP allocation.

Table 13 shows the allocation of a specific month for the purpose of comparing the methods. The portfolios for Configurations 2, 3, 4 and 5 were generated using 12 months of correlation and 15 groups, as shown in Figure 2. The portfolios date is August 2016. In the MVP portfolio, the sum of the percentage of the 5 largest positions is 67.3%, while in the HRP it is 22%; in Settings 2, 3, 4 and 5 these values are 33.3%, 42.2%, 33.3% and 42.3%, respectively.

It is possible to observe from Table 13, that in the MVP portfolio many assets are not allocated, that is, they have 0.0% of weight in the final allocation of the portfolio. When compared to the other portfolios, this portfolio is less diversified. Although at the end of the test period analyzed the return on the MVP portfolio is satisfactory, the concentration on a few stocks, due to optimization, can lead the portfolio to have large losses if the most relevant stocks have sharp losses.

## VI. CONCLUSIONS AND FUTURE RESEARCH

In this paper we introduced a framework to perform asset allocation based on partitional clustering. The framework is composed of six main steps: i) asset selection; ii) correlation and distance matrix calculation; iii) clustering; iv) intragroup allocation; v) integroup allocation; and vi) assessment.

This approach is different from the traditional ones, which are usually based on maximizing the return whilst minimizing the risk. One problem with the traditional optimization-based approaches is that they require calculating the inverse of the covariance matrix, a process that is prone to errors.

In the framework proposed, different partitional clustering algorithms can be used, together with various intragroup and intergroup allocation methods. These make the framework flexible in terms of the techniques used in each of its main steps.

In the experiments performed here, we chose the $k$-medoids partitional clustering algorithm with different values of $k$, and introduced a new medoid initialization technique that significantly improved the algorithm's performance. Furthermore, we tested a fuzzy and a performance proportional intragroup allocation method, and a balanced capital or performance proportional intergroup allocation method.

The framework was tested using the Brazilian B3 data from 2005 to 2020 and its results compared with those of the classical Markowitz model (MVP), the Ibovespa Index, and the Hierarchical Risk Parity (HRP) model. For a correlation window of 12 months the framework was capable of obtaining the best average performance in terms of return, but with slightly higher volatility, turnover, drawdown and sharpe ratio when compared with the other methods.

In terms of future research, it is possible to run further experiments with other partitional clustering algorithms and test with data from other stock markets.

## REFERENCES

[1] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] F. Black and R. Litterman, "Global portfolio optimization," *Financial Analysts J.*, vol. 48, no. 5, pp. 28–43, 1992.
[3] R. O. Michaud, "The markowitz optimization enigma: Is 'optimized' optimal?" *Financial Analysts J.*, vol. 45, no. 1, pp. 31–42, 1989.
[4] M. L. de Prado, "Building diversified portfolios that outperform out of sample," *J. Portfolio Manage.*, vol. 42, no. 4, pp. 59–69, Summer 2016, doi: 10.3905/jpm.2016.42.4.059.
[5] T. Raffinot, "Hierarchical clustering-based asset allocation," *J. Portfolio Manage. Multi-Asset*, vol. 44, no. 2, pp. 89–99, 2018, doi: 10.3905/jpm.2018.44.2.089.
[6] A. Ponsich, A. L. Jaimes, and C. A. C. Coello, "A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications," *IEEE Trans. Evol. Comput.*, vol. 17, no. 3, pp. 321–344, Jun. 2013.
[7] R. Clarke, H. De Silva, and S. Thorley, "Risk parity, maximum diversification, and minimum variance: An analytic perspective," *J. Portfolio Manage.*, vol. 39, no. 3, pp. 39–53, 2013.
[8] M. Grötschel and Y. Wakabayashi, "A cutting plane algorithm for a clustering problem," *Math. Program.*, vol. 45, nos. 1–3, pp. 59–96, Aug. 1989.
[9] R. N. Mantegna, "Hierarchical structure in financial markets," *Eur. Phys. J. B*, vol. 11, no. 1, pp. 193–197, Sep. 1999.
[10] G. Marti, F. Nielsen, M. Bińkowski, and P. Donnat, "A review of two decades of correlations, hierarchies, networks and clustering in financial markets," 2017, *arXiv:1703.00485*. [Online]. Available: http://arxiv.org/abs/1703.00485
[11] S. R. Nanda, B. Mahanty, and M. K. Tiwari, "Clustering indian stock market data for portfolio management," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8793–8798, Dec. 2010.
[12] V. Tola, F. Lillo, M. Gallegati, and R. N. Mantegna, "Cluster analysis for portfolio optimization," *J. Econ. Dyn. Control*, vol. 32, no. 1, pp. 235–258, Jan. 2008.
[13] C. Dose and S. Cincotti, "Clustering of financial time series with application to index and enhanced index tracking portfolio," *Phys. A, Stat. Mech. Appl.*, vol. 355, no. 1, pp. 145–151, Sep. 2005.
[14] P. N. Kolm, R. Tütüncü, and F. J. Fabozzi, "60 years of portfolio optimization: Practical challenges and current trends," *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 356–371, Apr. 2014.
[15] R. Mansini, W. Ogryczak, and M. G. Speranza, "Twenty years of linear programming based portfolio optimization," *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 518–535, Apr. 2014.
[16] M. Rubinstein, "Markowitz's 'portfolio selection': A fifty-year retrospective," *J. Finance*, vol. 57, no. 3, pp. 1041–1045, 2002.
[17] F. J. Fabozzi, F. Gupta, and H. M. Markowitz, "The legacy of modern portfolio theory," *J. Investing*, vol. 11, no. 3, pp. 7–22, Aug. 2002.
[18] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, Sep. 1964.
[19] W. F. Sharpe, "Mutual fund performance," *J. Bus.*, vol. 39, no. 1, pp. 119–138, 1966.
[20] L. N. de Castro and D. G. Ferrari, *Introdução à Mineração de Dados: Conceitos básicos, Algoritmos e Aplicações*. São Paulo, Brazil: Saraiva, 2016.
[21] J. W. Carmichael and R. S. Julius, "Finding natural clusters," *Systematic Biol.*, vol. 17, no. 2, pp. 144–150, Jun. 1968.
[22] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, "The planar k-means problem is np-hard," *Theor. Comput. Sci.*, vol. 442, pp. 13–21, Jul. 2012.
[23] L. Kaufman and P. Rousseeuw, *Clustering by Means Medoids*. Amsterdam, The Netherlands: North Holland, 1987.
[24] X. Jin and J. Han, *K-Medoids Clustering*. Boston, MA, USA: Springer, 2010, pp. 564–565.
[25] R. Krishnapuram, A. Joshi, and L. Yi, "A fuzzy relative of the k-medoids algorithm with application to Web document and snippet clustering," in *Proc. FUZZ IEEE Int. Fuzzy Systems. Conf.*, vol. 3, Aug. 1999, pp. 1281–1286.
[26] A. P. Reynolds, G. Richards, and V. J. Rayward-Smith, "The application of k-medoids and PAM to the clustering of rules," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Exeter, U.K.: Springer, 2004, pp. 173–178.
[27] J. C. Bezdek, "Objective function clustering," in *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer, 1981, pp. 43–93.
[28] BM&F Bovespa. (2015). *Metodologia do Indice Ibovespa—B3*. [Online]. Available: http://www.b3.com.br/pt_br/
[29] H. Markowitz, "The optimization of a quadratic function subject to linear constraints," *Nav. Res. Logistics Quart.*, vol. 3, nos. 1–2, pp. 111–133, Mar. 1956.
[30] M. Baker, B. Bradley, and J. Wurgler, "Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly," *Financial Analysts J.*, vol. 67, no. 1, pp. 40–54, Jan. 2011.
[31] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning," *Addion Wesley*, vol. 1989, no. 102, p. 36, 1989.
[32] A. Lipowski and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 6, pp. 2193–2196, Mar. 2012.
[33] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?" *Rev. Financial Stud.*, vol. 22, no. 5, pp. 1915–1953, 2007.
[34] M. Magdon-Ismail, A. F. Atiya, A. Pratap, and Y. S. Abu-Mostafa, "On the maximum drawdown of a Brownian motion," *J. Appl. Probab.*, vol. 41, no. 1, pp. 147–161, Mar. 2004.

**FLÁVIO GABRIEL DUARTE** received the M.Sc. degree in electrical engineering and computing from the Graduate Program in Electrical Engineering and Computing, Mackenzie Presbyterian University, and the degree in information system from the Centro Universitário de Brusque—UNIFEBE. He works as an Quant Analyst at Pandhora Investimentos and previously worked as an Investment Analyst at Ágora and Rio Bravo Investimentos. He has experience in computer science, focusing on information systems, systems analysis and development, web development, and web scraping.

**LEANDRO NUNES DE CASTRO** received the B.Sc. degree in electrical engineering from the Federal University of Goias, the M.Sc. degree in electrical engineering from Unicamp, the Ph.D. degree in computer engineering from Unicamp, and the M.B.A. degree in strategic business management from the Catholic University of Santos. He was a Research Associate with the University of Kent, Canterbury, from 2011 to 2002, a Research Fellow at Unicamp, from 2002 to 2006, a Visiting Lecturer with University Technology Malaysia, in September 2005, and a Visiting Professor with the University of Salamanca, in 2014. He is currently a Professor with the Graduate Program in Electrical Engineering and Computing, Mackenzie Presbyterian University, where he founded and leads the Natural Computing and Machine Learning Laboratory (LCoN). He is acknowledged by the Brazilian Research Council as a Leading Researcher in computer science and was recognized, in 2011, as one of the most cited authors in computer science from the country. He has eight books published, from among four were authored or coauthored and the others organized, and has published over 220 scientific articles. His main research interests include natural computing and data mining, emphasizing artificial immune systems, neural networks, swarm intelligence, evolutionary algorithms, and real-world applications. He was the Founder Chief Editor of the *International Journal of Natural Computing Research* and a member of the committee of several conferences.