Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 1334–1343

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Clustering algorithms for Risk-Adjusted Portfolio Construction

Diego León[1] *, Arbey Aragón[2], Javier Sandoval[1], Germán Hernández[2], Andrés Arévalo[2], and Jaime Niño[2]

[1] Universidad Externado de Colombia, Bogotá, Colombia
{diego.leon, javier.sandoval}@uexternado.edu.co
[2] Universidad Nacional de Colombia, Bogotá, Colombia
{aaragonb, gjhernandezp, ararevalom, jhninop,}@unal.edu.co

Abstract

This paper presents the performance of seven portfolios created using clustering analysis techniques to sort out assets into categories and then applying classical optimization inside every cluster to select best assets inside each asset category.

The proposed clustering algorithms are tested constructing portfolios and measuring their performances over a two month dataset of 1-minute asset returns from a sample of 175 assets of the Russell 1000® index. A three-week sliding window is used for model calibration, leaving an out of sample period of five weeks for testing. Model calibration is done weekly. Three different rebalancing periods are tested: every 1, 2 and 4 hours. The results show that all clustering algorithms produce more stable portfolios with similar volatility. In this sense, the portfolios volatilities generated by the clustering algorithms are smaller when compare to the portfolio obtained using classical Mean-Variance Optimization (MVO) over all the dataset. Hierarchical clustering algorithms achieve the best financial performance obtaining an adequate trade-off between accumulated financial returns and the risk-adjusted measure, Omega Ratio, during the out of sample testing period.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: Portfolio Construction, Return Correlation Matrix, Clustering Algorithms.

1 Introduction

In finance, portfolio construction or asset allocation is one of the most frequent problems practitioners solve every day. Portfolio Theory by Markovitz [10] introduced the problem faced by investors on a daily basis, in a framework called mean-variance as an optimization problem, specifically minimizing portfolio variance at a given level of expected or minimum required return. Markovitz summarized the solution space using the minimum variance frontier or more precisely, the positive slope section commonly known as the efficient frontier.

*Corresponding author

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.185

Markovitz's framework is presented from two perspectives. Given a portfolio's expected variance σ_p, find the maximum return μ_p; on the other hand, given a portfolio's expected return μ_p, find the minimum variance σ_p. The two approaches are consistent. The optimization problem is built assuming a portfolio consisting of n risky stocks, and positive definite VCV matrix σ, where VCV is a Variance-Covariance matrix. The objective is to find a weight vector w that minimizes portfolio's total variance. This approach is known as the mean-variance optimization (MVO).

Hence, Markovitz's portfolio optimization find a global minimum for the following objective function:

$$f(w) = V(w, \sigma) = w'\sigma w \tag{1}$$

subject to the general constraint, portfolio weights must sum up to one.

This model is commonly associated with the Capital Asset Pricing Model (CAPM) developed by William Sharpe [14]. This fact partially explained why Harry Markowitz shared the Nobel Prize in 1990 with Sharpe. However, the two models are used for different purposes by financial practitioners.

CAPM theory considers Markowitz's model from a microeconomics perspective to discover price formation of financial assets. In this model, the central concept is that market portfolio is uniquely defined. In Markowitz's model, portfolio optimization depends on expected or preferred returns and risks. Furthermore, the optimal portfolio is not unique and depends on investor's risk aversion. As a consequence, these two models could give two different approaches to the asset allocation problem. While the CAPM theory is the principal pillar for passive management, Markowitz's model is the central technique to start actively managing a portfolio if a practitioner believes that the information set is not unique or homogeneously spread across market participants.

Analytically, let us consider a universe of n risky assets. Let $\mu = \mathbb{E}[R]$ and $\sum = \mathbb{E}[(R-\mu)(R-\mu)^T]$ be the vector of expected returns and the asset's return covariance matrix. Classical MVO assumes that a portfolio with weights $w = (w_1, w_2, ..., w_n)^T$ of n risky assets is fully invested, meaning that $\sum_{i=1}^{n} w_i = 1$, hence, short sells are not allowed, i.e. $w > 0$. When we have n stock prices time series: S_{ij} at time j for security i, the log returns are $R_{ij} = \ln \frac{S_{ij}}{S_{ij-1}}$, then is possible to denote $R = (R_1, \cdots, R_n)$ the asset returns vector. Portfolio's return is equal to $R(w) = \sum_{i=1}^{n} w_i R_i$. In a matrix form, it is obtained $R(w) = w^T R$, portfolio's expected return is:

$$\mu(w) = \mathbb{E}[R(w)] = \mathbb{E}[w^T R] = w^T \mathbb{E}[R] = w^T \mu \tag{2}$$

While its variance is equal to:

$$\begin{aligned}
\sigma^2(w) =& \mathbb{E}[(R(w) - \mu(w))(R(w) - \mu(w))^T] \\
=& \mathbb{E}[w^T R - w^T \mu)(w^T R - w^T \mu)^T] \\
=& \mathbb{E}[w^T (R-\mu)(R-\mu)^T w] \\
=& w^T \mathbb{E}[(R-\mu)(R-\mu)^T] w \\
=& w^T \sum w
\end{aligned} \tag{3}$$

On Figure 1, we have simulated 500 portfolios, with n=4 securities, and 1,000 observations for each asset. It is possible to see the minimum variance frontier, shown by the dotted green line, the positively sloped segment contains all the optimum portfolios for a given level of desired risk.

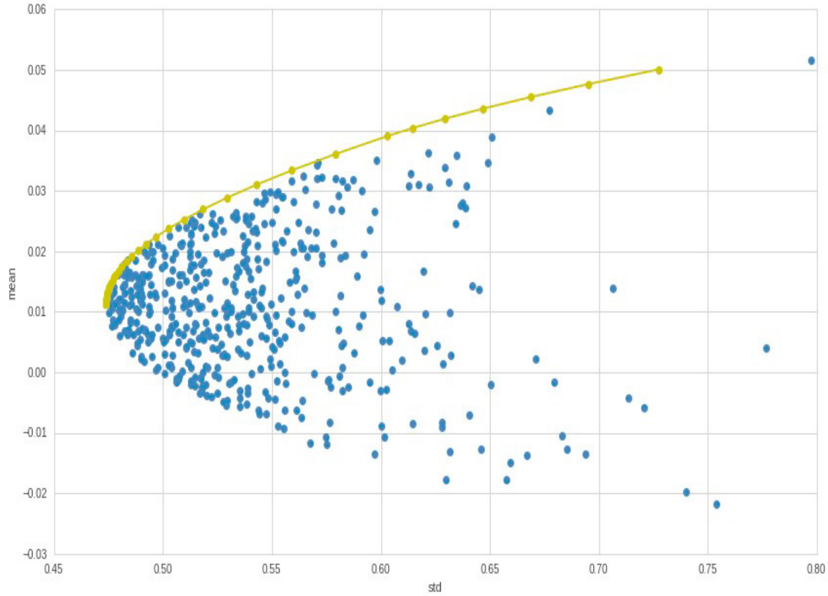

Figure 1: Optimized Markovitz portfolios.

Clustering analysis is a successful tool to partition large data sets into groups based on its features. This technique is commonly found in pattern discovery, classification and anomaly detection. In portfolio optimization, clustering analysis could also be a successful technique helping to select better portfolios based on unsupervised learning to preprocess asset selection. In recent years, clustering analysis has arisen as one of the most innovative methods in portfolio selection, as a multivariate statistical analysis method that reorganizes objects in a data set into homogeneous groups [1]. Clustering analysis is frequently used for grouping data objects in similar groups according to their similarities.

In this study, we tested several algorithms, which can be disaggregated into two main groups: partitional and hierarchical clustering [15]. Dataset object division into non-overlapping clusters is considered partitional clustering, whereas the presence of nested clusters organized as subclusters in a tree map is called a hierarchical clustering. The later is commonly displayed graphically using a diagram known as a dendrogram [15].

This work applies a set of correlation-based clustering algorithms, assuming that correlation coefficient between two time series is a measure of their similarity. Correlation-based clustering has been recently used to infer hierarchical structures among different asset classes from its correlation coefficient matrix [9] [4] [3]. It is important to mention that in high volatility periods the correlation as a measure of similarity does not hold up.

Exploring a new approach for asset allocation [8], López de Prado introduced a portfolio diversification technique called “Hierarchical Risk Parity” (HRP), which uses graph theory and hierarchical clustering. One of the main advantages of HRP is the possibility to compute a portfolio on an ill-degenerated or even a singular covariance matrix, with a lower computational cost. He adverted that the covariance matrix lacks the notion of hierarchy, which allows weights to vary freely in random ways. Furthermore, he provided a good example that illustrates the hierarchy concept for the asset allocation task: “...stocks could be grouped in terms of liquidity, size, industry and location, where stocks within a given group compete for allocations. In deciding the allocation to a large publicly-traded U.S. financial stock like J.P. Morgan, we will consider adding or reducing the allocation to another large publicly-traded U.S. bank like Goldman Sachs, rather than a small community bank in Switzerland, or a real estate holding in the Caribbean". This fact means that a correlation matrix makes no difference between the hierarchy of assets classes.

This paper aims at placing stocks into groups, or clusters, suggested by the raw data, not defined a priori, such that stocks in a given cluster tend to be similar to each other and stocks in different clusters tend to be dissimilar. In terms of financial theory, this purpose allows building highly diversified portfolios. In the following stage, this work proposes a risk-adjusted optimization using Sharpe Ratio for optimal asset allocation into each cluster.

The next section will present a short review on selected clustering algorithms used to build the risk diversified portfolio. Third section will explain the selected dataset and main topics of the computational experiments. The fourth explain with detail the risk-adjusted portfolio performance measure [6] [7]. Fifth section will show results focusing on the performance measures of portfolio selection based on each clustering algorithm using Markovitz optimized portfolios as benchmarks. Finally, conclusions and further research opportunities are presented.

2 Clustering Algorithms

We tested the potential of clustering in portfolio selection using different clustering algorithms. First Algorithm used is K-Means. This method [2], initializes cluster centers by selecting points within the dataset that are further away from each other in a probabilistic manner. This algorithm widely used in several applications because it is by far the one with the highest simplicity improving the speed of convergence. Mini Batch K-Means [13] is a K-Means variant, which uses mini-batches or subsets of input data to reduce computation time, while still attempting to optimize the same objective function, thereby converging faster than K-Means. However, the quality of the results is reduced. The Spectral clustering algorithm is especially efficient if the affinity matrix is sparse and it is used as a similarity metric for distance in a graph i.e. nearest-neighbor graph [11].

The Birch algorithm uses the Euclidean distance between points as a similarity metric, for a given data it builds a tree known as a Characteristic Feature Tree (CFT). It has two parameters, branching factor and threshold. Specifically, the branching factor limits the number of subclusters in a node and the threshold limits the distances between points. The approach of this algorithm is related to data reduction methods, given that it reduces the input data to a set of subclusters. One of its principal disadvantages is the scalability to high dimensional data [18]. This drawback could be a problem with huge financial datasets.

If hierarchical clustering is going to be applied to portfolio optimization, it is required a convenient distance metric such[9]:

$$D_{i,j} = \sqrt{2(1 - \rho_{i,j})} \tag{4}$$

where $D_{i,j}$ is the distance metric based on the Pearson correlation coefficient, $\rho_{i,j}$, between the i^{th} and the j^{th} asset.

This work tests also three categories of hierarchical clustering, specifically the three principal agglomerative clustering algorithms Average Linkage (AL), Complete Linkage (CL) and Ward's Method (WM). Each one of them is based on a linkage criteria to determine the metric

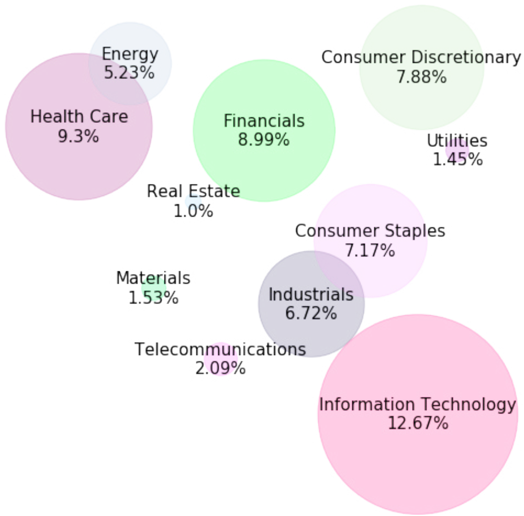

Figure 2: Asset Class by sector.

used for the merge strategy. Agglomerative clustering uses a bottom-up approach to perform hierarchical clustering when each observation starts in its own cluster, and clusters are successively merged together. The AL algorithm minimizes the average of the distances between all observations of pairs of clusters and has been related to a version of spanning tree called Average Linkage Minimum Tree [16]. The Other algorithms covered, CL, minimizes the maximum distance between observations of pairs of clusters and finally, WM minimizes the sum of squared differences within all clusters. Although the CL and WM algorithms are not associated with spanning tree graphs, they could deliver remarkable results.

3 Data description and Experiment

All stock information used in this paper come from the Rusell 1,000® Index constituents, figure 2 shows the exposition by industrial sector of the stocks selected. The Rusell 1,000 stocks represent over 90% of the total market capitalization of all U.S. listed stocks, and it is considered a bellwether index for large-cap investing. This work uses closing prices for the 175 assets with the highest participation in the index from 12/09/2016 to 9/11/2016 using a 1-minute resolution. These 175 assets represent the 64.16% of the Rusell 1,000® Index, thereby, representing 71% of the total US equity market capitalization. The first graph in figure 3 shows the return correlation matrix before and after clustering for the 175 assets selected, clustering algorithms are executed using the data for first three weeks in a sliding window using a 1-minute return correlation matrix as an in-sample data. Afterward, we distribute assets inside each cluster relying on a Sharpe Ratio optimization [14]. During the next week, we keep the allocation fixed and then estimate the in-cluster distribution every one, two and four hours. This means that the clustering algorithms are executed each week with the data comprised of

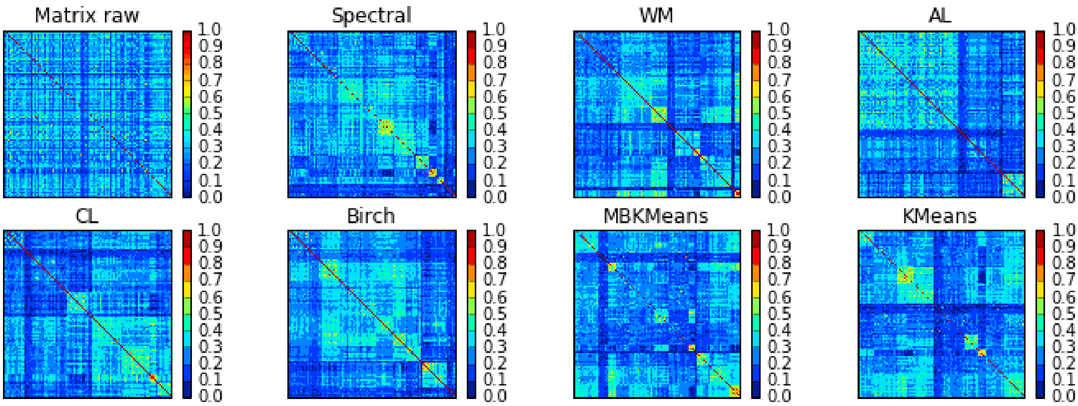

Figure 3: Example of the return correlation matrix before clustering and after running the seven clustering algorithms tested in this study.

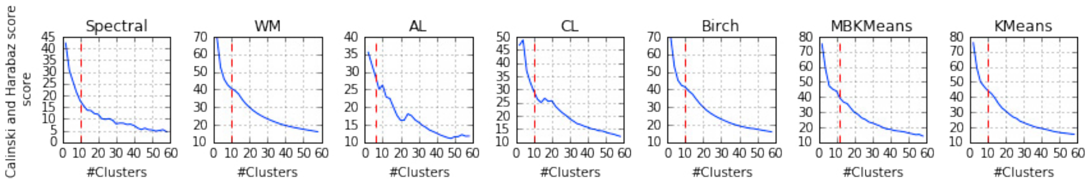

Figure 4: Calinski-Harabaz Score.

the past three weeks in a sliding window and the asset allocation is reconfigured with mentioned periodicities into each cluster by Sharpe ratio optimization.

This paper selected the optimal number of clusters when needed, using the Calinski-Harabaz(CH) score [5]. All non-hierarchical clustering technique needed an apriori definition of the number of clusters used. The Calinski-Harabaz score is the ratio of means between cluster dispersion and the dispersion inside every cluster. A higher value of this index shows dense, well-separated clusters. The performance of each clustering algorithm is influenced by the number of clusters formed. We found that the best selection based on the CH score was ten clusters for all algorithms as is shown in figure 4. In addition, this work also presents the Silhouette coefficient(SC) [12]. The SC score is bounded between -1 for incorrect clustering and +1 for highly dense clustering. Scores around zero indicate overlapping clusters, the score is higher when clusters are dense and well separated. Figure 5 shows that hierarchical clustering algorithms tend to have the best cluster configuration based on the SC. It is important to highlight that these algorithms generate autonomously the optimal number of clusters for grouping the data.

Fifth section will show trading performances of the seven selected portfolio strategies using clustering techniques and Markovitz optimization inside clusters.

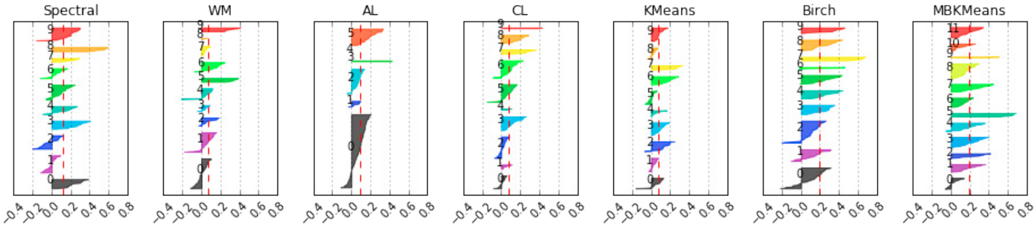

Figure 5: Silhouette Coefficient.

4 Portfolio Performance Measure

One of the most important measures to evaluate the performance of a financial asset is Omega ratio [6] [7], a relative measure of the probability of reaching a certain return, such as a minimum target. A high Omega value indicates a high probability that a minimum target return will be met or exceeded. The value of the Omega function at r, a threshold level, is the ratio of positive returns relative to r weighted by their probability and negative returns relative to r also weighted by their probability. If $F(x)$ is the cumulative density function of returns then:

$$\Omega(r) = \frac{\displaystyle\int_r^{\infty} (1 - F(x))dx}{\displaystyle\int_{-\infty}^{r} F(x)dx}. \tag{5}$$

The omega concept divides the expected return into two parts, gains and losses, or returns above the threshold level and those below it. Therefore, in simple terms, consider omega as the ratio of positive returns relative to negative returns. The Omega Ratio is a measure of performance that doesn't assume a normal distribution of returns and contains much more information about the return distribution, including the mean, variance, skew and kurtosis and is especially valuable for non-normal investment returns. It actually uses the real return distribution rather than a theoretical normal distribution.

The Omega ratio also considers all information available from return's history of a given asset, it can be used to rank potential investments addressing the investor's threshold level. However, the Omega based decisions are not static for at least two reasons: As return information is updated, the probability distribution will change and omega must be updated. As the asset manager's threshold level changes, the rankings among comparative investments may change. Therefore, the omega ratio allows asset managers to visualize the trade-off between risk and return at different threshold levels for various asset choices.

5 Results

Figure 6 shows that the best clustering technique is the Ward Method (WM), a subclass of hierarchical clustering techniques when portfolio performance is measured using the Omega Ratio. This model reached a positive return during the out-sample dataset when clusters were rebalanced every hour during the 160 transactions or portfolio rebalances. Figure 6 also shows

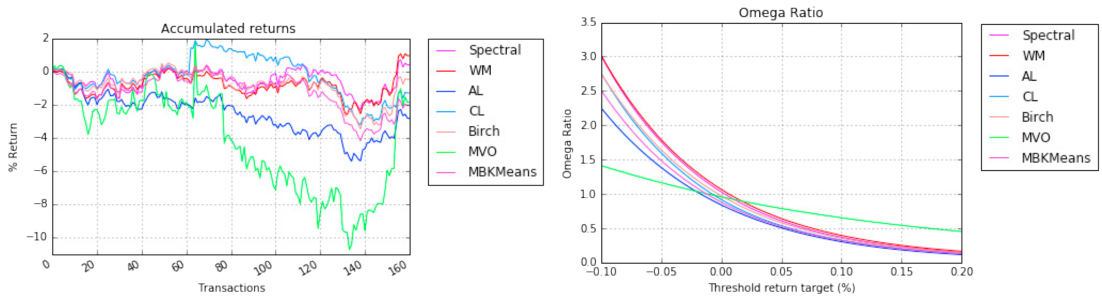

Figure 6: Accumulated returns by transaction and Omega ratio performance in one hour holding period.

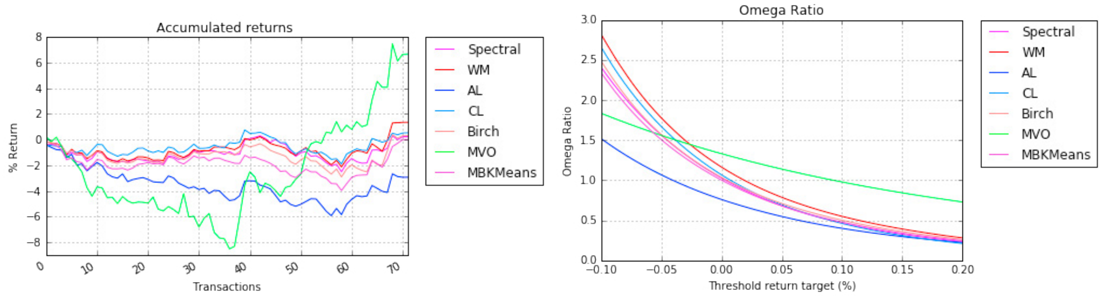

Figure 7: Accumulated returns by transaction and Omega ratio performance in two hours holding period.

that portfolios created using all the clustering techniques have better Omega ratios than the MVO portfolio. At threshold level 0%, WM shows the best Omega ratio of all models implying the best balance between return and risk. Rebalancing the portfolio every two hours, the MVO portfolio obtained the best result in terms of accumulated returns. The Omega ratio is higher for the MVO algorithm using a rebalancing period of two hours at a threshold level of 0%. However, figure 7 shows the instability in the portfolio returns generated using the MVO algorithm. The volatility of the MVO portfolio returns reaches the highest spread between minimum and maximum cumulative returns during the testing period. In contrast, clustering algorithms have a remarkable stability where WM and all hierarchical clustering algorithms show the lowest return volatility.

Figure 8 shows that the CL algorithm obtains the best accumulated returns in a rebalancing period of four hours. Similarly, it presents the best Omega ratio at a threshold level of 0%. This fact confirms the good behavior of hierarchical clustering algorithms in the portfolio selection task. The results indicate the benefits of using clustering algorithms in asset allocation. This results are in agreement with [8] and [17]. Following previous results, the clearest advantage of using hierarchical clustering algorithms to solve the asset allocation problem is to achieve a more stable diversified portfolio that shows a lower risk measured in terms of return volatility. Figure 9 shows that the benchmark algorithm for portfolio construction, MVO, has the worst volatility by far disregarding the rebalancing timing.

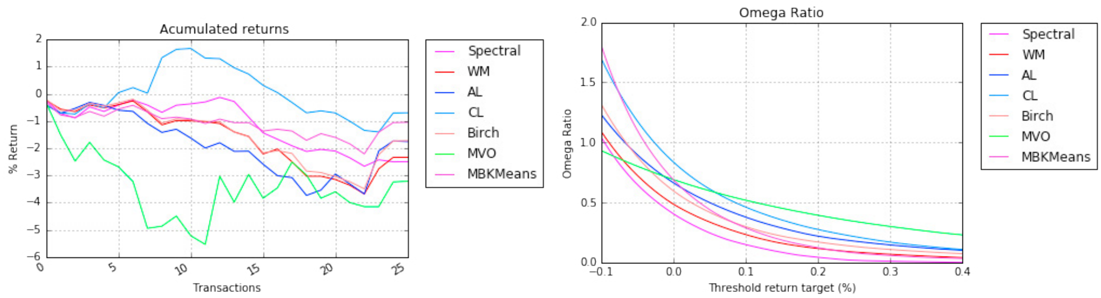

Figure 8: Accumulated returns by transaction and Omega ratio performance in four hours holding period.

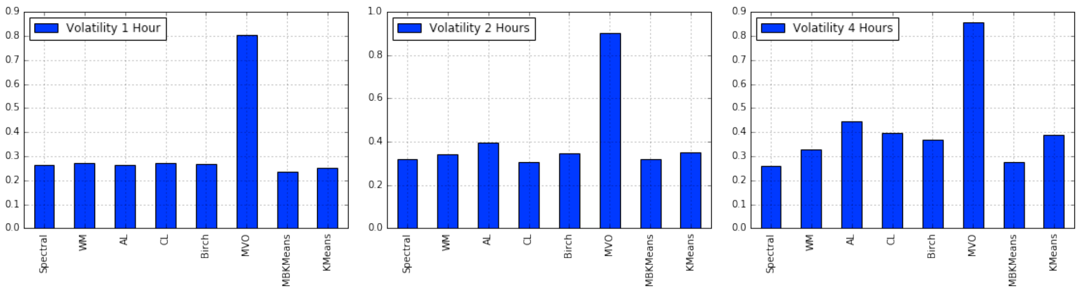

Figure 9: Volatility by holding period.

6 Concluding remarks and further research

Machine learning methods have recently been used in finance because they have a special advantage when learning from data and adapting to financial market changes. Clustering algorithms as a machine learning method allow practitioners to identify and incorporate fast changes inherent to the dynamics of financial markets.

The MVO algorithm treats all variables as interrelated, assuming a complete cluster. In other words, traditional asset allocation does not recognize the complexity immerse in the data. This work presents some novel, robust and flexible methods with visual interpretations to construct risk-adjusted portfolios. One of the main purposes of the study was testing the adequate timing for rebalancing a portfolio, the findings suggest that a small holding period, around one hour, shows the best financial performance. As a consequence, an adequate timing for portfolio rebalancing is important to incorporate all changes and new information presented in the market data. Clustering methods showed a better trade-off between return and risk than MVO algorithm. The empirical results indicate that hierarchical algorithms have a better performance when building diversified portfolios measured by the Omega ratio. One of the most important results is the stable behavior of clustering-based portfolios addressing a special issue in financial markets, the volatility.

The complexity of the decisions in portfolio selection leads us to propose a further research with larger data sets, optimizing within clusters by the Omega ratio instead of the Sharpe Ratio, exploring different distance measures such the Chebychev, Manhattan or Minkowski distances. We could also incorporate investor's preferences using fuzzy decision theory included in Fuzzy Clustering applied to portfolio selection.

References

[1] Mark S Aldenderfer and Roger K Blashfield. Cluster analysis. sage university paper series on quantitative applications in the social sciences 07-044. 1984.

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[3] Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, and Rosario N Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.

[4] Giovanni Bonanno, Fabrizio Lillo, and Rosario N Mantegna. High-frequency cross-correlation in a set of stocks. 2001.

[5] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[6] Con Keating and William F Shadwick. An introduction to omega. Technical report, The Financial Development Centre London, Working Paper, 2002.

[7] Con Keating and William F Shadwick. A universal performance measure. *Journal of performance measurement*, 6(3):59–84, 2002.

[8] Marcos López de Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, 2016.

[9] Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197, 1999.

[10] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[11] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. In *NIPS*, volume 14, pages 849–856, 2001.

[12] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[13] David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pages 1177–1178. ACM, 2010.

[14] William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

[15] Pang-Ning Tan et al. *Introduction to data mining*. Pearson Education India, 2006.

[16] Michele Tumminello, Claudia Coronnello, Fabrizio Lillo, Salvatore Micciche, and Rosario N Mantegna. Spanning trees and bootstrap reliability estimation in correlation-based networks. *International Journal of Bifurcation and Chaos*, 17(07):2319–2329, 2007.

[17] Hannah Cheng Juan Zhan, William Rea, and Alethea Rea. An application of correlation clustering to portfolio diversification. *arXiv preprint arXiv:1511.07945*, 2015.

[18] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: A new data clustering algorithm and its applications. *Data Mining and Knowledge Discovery*, 1(2):141–182, 1997.