Expert Systems with Applications 38 (2011) 14952–14959

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Distributing weights under hierarchical clustering: A way in reducing performance breakdown

Jin Zhang [a,*], Dietmar Maringer [b]

[a] Center for Computational Finance and Economic Agents, School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom
[b] Faculty of Economics and Business Administration, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Asset allocation
Clustering technique
Sharpe ratio
Evolutionary approach
Heuristic optimization

ABSTRACT

This paper proposes a clustering asset allocation scheme which provides better risk-adjusted portfolio performance than those obtained from traditional asset allocation approaches such as the equal weight strategy and the Markowitz minimum variance allocation. The clustering criterion used, which involves maximization of the in-sample Sharpe ratio (SR), is different from traditional clustering criteria reported in the literature. Two evolutionary methods, namely Differential Evolution and Genetic Algorithm, are employed to search for such an optimal clustering structure given a cluster number. To explore the clustering impact on the SR, the in-sample and the out-of-sample SR distributions of the portfolios are studied using bootstrapped data as well as simulated paths from the single index market model. It was found that the SR distributions of the portfolios under the clustering asset allocation structure have higher mean values and skewness but approximately the same standard deviation and kurtosis than those in the non-clustered case. Genetic Algorithm is suggested as a more efficient approach than Differential Evolution for the purpose of solving the clustering problem.

1. Introduction

Generally speaking, traditional asset allocation strategies can be classified into two categories: parametric approaches (e.g. the Markowitz allocations); and parameter-free allocations (e.g. the equal weight strategy). According to the well-known Markowitz theory, rational investors should always prefer the 'efficient' portfolios which yield the highest return at any given risk level. However, most of the time, a precise estimation of asset return properties (such as expected return, variance and covariance) may be difficult to obtain. As empirically observed financial data is quite noisy, the estimates derived from such data may be unreliable. Moreover, the accuracy of estimates may rely on not only the number of assets but also the available number of observations, which is particularly important to the Markowitz allocations. If a portfolio contains hundreds of assets, the high dimensionality may hinder an accurate estimation of the dependency structure of assets (i.e. covariance under the Markowitz framework), which in the literature is usually referred to as the 'curse of dimensionality'. The above problems may result in suboptimal portfolios if investors still apply the traditional asset allocations to manage portfolios, especially large ones. For instance, the Sharpe ratio (SR) of portfolios, which is a risk-adjusted performance measure based on the first two moments of returns, may not be optimal.

Many researchers have suggested different approaches for improving the estimation of return moments and portfolio performance. For example, Harris and Yilmaz (2007) combined the return-based and the range-based measures of volatility to improve the estimate of the multivariate conditional variance–covariance matrix. On the other hand, one may simply adopt parameter-free allocations (e.g. the equal weight (EW) investment strategy of Windcliff & Boyle (2004)) which are independent of those return moment measures. In addition to methods from mathematics and finance, approaches from computer science have also been considered by researchers. For instance, Pattarin, Paterlini, and Minerva (2004) employed a clustering technique to analyze mutual fund investment styles; and Lisi and Corazza (2008) proposed an active fund management strategy which selected stocks after clustering equities. The clustering techniques considered in the above studies still comply with the traditional clustering criterion, i.e. minimizing the dissimilarity between the cluster members while maximizing the dissimilarity between clusters.

This study, however, proposes a different clustering criterion to the traditional one. The proposed clustering criterion segments assets by maximizing the in-sample SR of portfolios. Two main benefits are expected from using this clustering asset allocation scheme. First, the dimensionality in the asset allocation problem will decrease, as the cluster size can be controlled by using a cardinality constraint. Thus, when portfolio managers apply the

* Corresponding author. Tel.: +44 7722787008.
E-mail addresses: jzhangf@essex.ac.uk (J. Zhang), dietmar.maringer@unibas.ch (D. Maringer).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.05.052

Markowitz allocations to managing large portfolios, the 'curse of dimensionality' problem may be avoided. Secondly, the out-of-sample SR of a portfolio which is constructed under the clustering structure shall be better than the portfolio SR in the non-clustered case by using a same asset allocation.

Provided there is no structural break between the in-sample and the out-of-sample periods and the clustering structure is optimal, one should observe the same clustering impact on both the in-sample and out-of-sample SRs. The first four moments of the SR distribution are considered in order to study the clustering impact on the SR; and a rational investor should prefer the SR distributions with high mean, high skewness, low standard deviation and low kurtosis. To construct the SR distribution, simulated portfolio returns are first generated by using the portfolio weights based on simulated asset returns, and then the simulated SR values which are calculated from the portfolio returns constitute the SR distribution. In this study, the EW strategy and the Markowitz minimum variance portfolio (MVP) allocation are adopted to distribute asset weights; and two approaches are used to provide the simulated asset returns for both the in-sample and out-of-sample SR studies, i.e. the traditional bootstrap method and the single index market model.

The proposed clustering asset allocation scheme is introduced starting with the following technical terms. Assets within a cluster are called 'cluster members', the portfolio which is constructed by using the members in the same cluster is referred to as a 'cluster portfolio', and these cluster portfolios are combined to form a 'terminal portfolio'. In other words, the proposed approach first of all segments assets into a series of disjoint clusters according to the clustering structure. Next, a set of cluster portfolios is constructed by using an asset allocation on the basis of the cluster members in different clusters. Finally, the terminal portfolio is constructed by adopting the same asset allocation based on those cluster portfolios. Two population-based evolutionary methods (Differential Evolution and Genetic Algorithm) are used to tackle the clustering problem with the SR maximization design. Fig. 1 briefly describes this clustering asset allocation procedure in a case of three clusters with eleven assets.

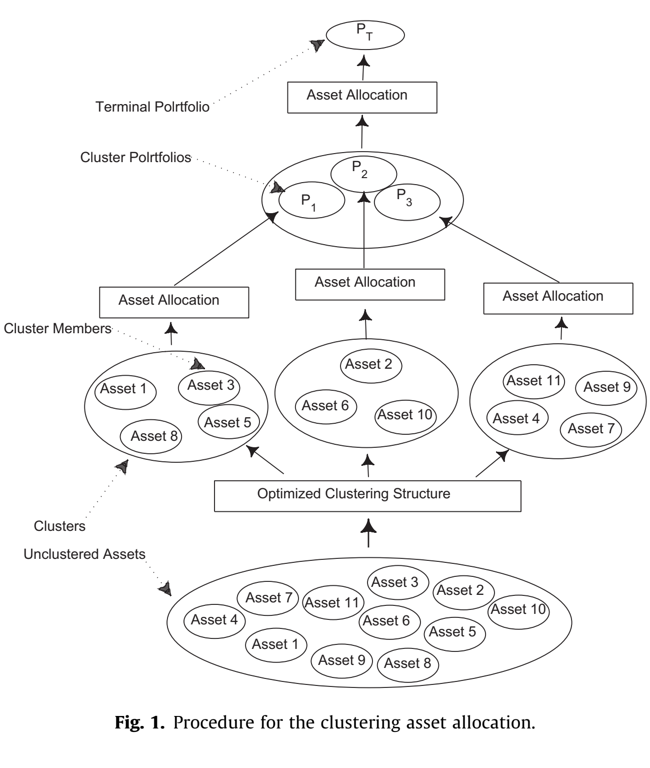

Fig. 1. Procedure for the clustering asset allocation.

The paper is organized as follows. Sections 2 and 3 introduce the clustering optimization problem and the two asset allocation approaches. Section 4 describes the two evolutionary algorithms for tackling the clustering problem. The experimental results are presented and discussed in Section 5. Section 6 concludes the study.

2. The optimization problem

Suppose there are N stocks considered for the asset allocation problem. The optimization problem is to identify a clustering structure $\mathcal{C}$ (i.e. a union of subsets $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_G$), so that the portfolio SR based on such a cluster structure is maximized given a cluster number G. In this study the cluster number G is manually assigned, and G is an integer number within a range $1 \leqslant G \leqslant N$ as empty clusters are not considered. When G is equal to the number of either *1* or N, there is no clustering effect thus the clustering asset allocation problem becomes an ordinary asset allocation problem. The optimization objective of the clustering problem can be described as

$$\max_{\mathcal{C}} SR = \frac{r_P - r_f}{\sigma_P}, \tag{1}$$

where $\mathcal{C}$ denotes the optimal clustering structure, r_P is the average daily return of the portfolio, σ_P is the standard deviation of the portfolio return over the evaluation period, and r_f refers to as the risk-free return. As with traditional clustering problems, the union of segmented assets $\mathcal{U}$ represents the collection of assets, and there is no intersection between two different clusters. Let $\mathcal{C}_g$ denote the gth cluster of assets, then the above constraints can be written as

$$\bigcup_{g=1}^{G} \mathcal{C}_g = \mathcal{U}, \tag{2}$$

$$\mathcal{C}_g \cap \mathcal{C}_h = \emptyset, \quad g \neq h. \tag{3}$$

Let $\widetilde{N}^{\min}$ and $\widetilde{N}^{\max}$ denote the minimum and maximum asset numbers allowed in a cluster respectively, then the following cardinality constraints are employed to limit the dimensionality of clusters:

$$\widetilde{N}^{\min} \leqslant \sum_{j=1}^{N} I_{j\in\mathcal{C}_g} \leqslant \widetilde{N}^{\max} \quad 1 \leqslant g \leqslant G, \tag{4}$$

$$\text{where} \quad I_{j\in\mathcal{C}_g} = \begin{cases} 1 & \text{if } j \in \mathcal{C}_g, \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

$$\text{with} \quad \begin{cases} \widetilde{N}^{\min} = \lceil \frac{N}{2G} \rceil, \\ \widetilde{N}^{\max} = \lceil \frac{3N}{2G} \rceil. \end{cases} \tag{6}$$

Eq. (5) corresponds to an indicator function showing whether asset j belongs to cluster g. The above optimization problem is hard to solve by using traditional optimization methods. Brucker (1978) pointed out that the clustering problem turns out to be non-deterministic polynomial-time hard (NP-hard) when the cluster number G becomes higher.

3. Asset allocation methods

3.1. Weight constraints

As with traditional asset allocation problems, the budget constraint must be met while using the proposed clustering asset allocation. The sum of cluster member weights in a cluster should be equal to 1, and likewise the sum of cluster portfolio weights. As