

Black-Litterman Portfolio with K-shape Clustering

Yeji Kim · Poongjin Cho[†]

School of Computing, Gachon University

K-shape 군집화 기반 블랙-리터만 포트폴리오 구성

김예지 · 조풍진[†]

가천대학교 AI소프트웨어학부

This study explores modern portfolio theory by integrating the Black-Litterman portfolio with time-series clustering, specifically emphasizing K-shape clustering methodology. K-shape clustering enables grouping time-series data effectively, enhancing the ability to plan and manage investments in stock markets when combined with the Black-Litterman portfolio. Based on the patterns of stock markets, the objective is to understand the relationship between past market data and planning future investment strategies through backtesting. Additionally, by examining diverse learning and investment periods, it is identified optimal strategies to boost portfolio returns while efficiently managing associated risks. For comparative analysis, traditional Markowitz portfolio is also assessed in conjunction with clustering techniques utilizing K-Means and K-Means with Dynamic Time Warping. It is suggested that the combination of K-shape and the Black-Litterman model significantly enhances portfolio optimization in the stock market, providing valuable insights for making stable portfolio investment decisions. The achieved sharpe ratio of 0.722 indicates a significantly higher performance when compared to other benchmarks, underlining the effectiveness of the K-shape and Black-Litterman integration in portfolio optimization.

Keywords : Portfolio Theory, Black-Litterman Portfolio, Time-series Clustering, K-shape Clustering, Investment Strategy

1. 서 론

포트폴리오 최적화 문제는 자산 간의 효율적인 조합을 통해 최소한의 위험으로 최대한의 수익을 추구한다. 블랙-리터만(Black-Litterman) 포트폴리오는 투자자의 주관적인 전망을 기반으로 한 포트폴리오 최적화 방법론이다[5]. 이 모델은 마코위츠(Markowitz)의 포트폴리오 이론[9]을 기반으로 하되, 투자자의 뷰와 시장의 불확실성 사이의 균형을 맞추는 것을 중점으로 한다. 블랙-리터만 포트폴리오 구성에서 가장 핵심적인 단계는 뷰를 효과적으로 반영하는 것이며, 머신러닝 기법을 적용한다면 포트폴리오의 수

익성을 높이는 뷰를 찾는 데 결정적인 역할을 할 것으로 기대한다.

군집화는 비슷한 특징을 가진 데이터들을 모아주는 머신러닝 기법으로서, 같은 군집에 속한 주식들은 같은 뷰를 가진다고 가정할 수 있다. 군집화 방법에 따라 다양한 군집을 구성할 수 있는데, K-Means 군집화 방법과 같은 일반적인 군집화 방법은 주식 가격 데이터와 같은 시계열 데이터에 적합하지 않다[11]. 시계열 데이터는 시간의 흐름에 따른 패턴을 포착하는 것이 중요하고, 이에 K-shape 군집화와 같은 시계열 데이터에 특화된 군집화 방법들이 개발되었다[13]. 따라서, 시계열 군집화에 기반하여 블랙-리터만 포트폴리오의 뷰를 설정한다면, 주식 시계열의 패턴을 더욱 잘 파악한 포트폴리오 전략을 얻을 수 있을 것이다. 시계열 군집화 기반 포트폴리오 구성의 자세한 과정은 다음과 같다.

먼저, 시계열 데이터의 군집화를 위해 K-shape 방법론을 적용하여 주식들의 패턴을 군집화한다. 각 군집은 유사한 주식 가격 움직임 패턴을 보이는 주식들로 구성되며, 군집별로 과거 moving window 기간 동안의 평균 수익률을 구하고 이를 블랙-리터만 뷰로 설정한다. 이에 따라, 같은 움직임 패턴을 보여 같은 군집에 속하게 된 종목은 동일한 뷰를 가지게 되고, 이러한 방식으로 투자자의 주관적인 뷰와 시장의 동향을 함께 반영하여 포트폴리오의 균형을 찾아간다.

본 연구의 결과는 투자자들에게 더욱 효율적이고 전략적인 포트폴리오 구성 방법을 제시할 것으로 기대한다. 시계열 데이터에 특화된 군집화와 포트폴리오 이론을 결합한 연구는 이전에 시도되지 않았으며, 이는 기존의 포트폴리오 최적화 이론을 확장하는 중요한 역할을 한다. 또한, 군집화 방법은 정답이 없다고 알려져 있어 모델의 튜닝이 어렵지만, 군집화 기반 투자전략의 수익성을 통해 모델의 튜닝이 가능하다. 이는 복잡한 시장 데이터 분석에 새로운 관점을 제시하고, 포트폴리오 관리 분야에 새로운 패러다임을 가져온다. 본 연구는 자산관리 및 펀드 투자 분야에 혁신적인 기여를 기대하며, 실제 투자 전략 수립에 직접 적용 가능한 새로운 통찰을 제공한다. 이어지는 제2장에서는 선행연구와 이론적 배경을 설명하고, 제3장에서는 시계열 군집화 기반 포트폴리오 구성 방법론을 소개한다. 제4장에서는 백테스팅을 통해 수익성을 살펴보고, 마지막으로 제5장에서는 결론을 유도하고 추후 연구 과제들을 알아본다.

2. 이론적 배경

2.1 선행연구

포트폴리오를 구성하는데 있어서 군집화를 활용한 많은 연구가 있었고, Cho et al.[6]는 군집화를 통해 비슷한 가격 움직임을 가지는 두 종목을 찾아 페어 트레이딩 포트폴리오를 구성하였다. 또한, Wu et al.[19]는 하방 리스크를 피하는 추세 특성들을 바탕으로 주식들을 군집화하여 마코위츠 포트폴리오를 구축하였고, Tayali[17]는 군집화를 바탕으로 카디널리티 및 박스 제약조건을 고려한 현실적인 포트폴리오를 설계하였다. 시계열 데이터에 특화된 군집화 방법을 활용한 포트폴리오 연구는 존재하지 않았지만, 관련하여 Moon and Kim[10]은 반응형 Dynamic Time Warping(DTW)을 통해 비슷한 패턴의 주식을 그룹화하여 포트폴리오를 최적화하였고, Putra et al.[14]는 B-Spline을 통해 고차원 시계열 데이터를 보간하고 군집화를 통한 포트폴리오를 구성함으로써 군집화의 수행 시간

을 줄여주었다. Aghabozorgi and Teh[1]는 개별 주식들의 공동 움직임을 평가하기 위해 모양 기반의 세 단계 군집화 방법을 제안하였다. 이 방법은 시계열 간의 정확한 거리 계산보다는 모양을 기반으로 군집을 형성하는데 주력했다. 이는 시계열 군집화의 기초가 되는 모델이라고 간주할 수 있고, K-shape 군집화의 Shape-Based Distance(SBD) 기반 접근은 시계열 데이터의 구조적 유사성을 더 정밀하게 포착할 수 있는 보완적인 기법으로 활용될 수 있다.

블랙-리터만 포트폴리오는 투자자의 전망을 반영할 수 있는 투자전략으로서, 효과적인 뷰를 제시하기 위한 여러 연구들이 등장하였다. Wang and Aste[18]는 포트폴리오의 내재된 비정상성을 해결하고자 Inverse Covariance Clustering (ICC) 방법을 통해 시장의 상황을 판별하고 이를 블랙-리터만 뷰에 반영하였고, Sahamkhadam et al.[16]는 꼬리 의존성을 반영한 Copula를 통해 투자자 전망을 평가하였다. 딥러닝 모델을 통해 투자자의 뷰를 예측한 연구도 존재하였는데, Barua and Sharma[2]는 CNN-BiLSTM 모델을 통해 주가가격을 예측하여 뷰를 설정하였고, Rezaei et al.[15]는 CEEMD-CNN-LSTM 모델로 수익을 예측하여 뷰를 반영하였다. Barua and Sharma[3]는 블랙-리터만 모델에 투자자의 정서를 반영하였다. 투자자의 두려움과 탐욕을 나타내는 기술적 지표들 통해 정서적 요인을 수치화하고, 이를 뷰 구축에 반영함으로써 포트폴리오 관리에 심리적 요소를 중요한 변수로 포함하였다. 뷰를 구축하는 데는 CEEMDAN-GRU 모델을 통해 심리 변수를 예측하였고, 투자자 정서를 수치화하는 기법은 뷰를 구성하는데 있어 추가적으로 고려할 수 있는 요소이다. Pang and Karan[12]는 Conditional Value at Risk(CVaR)을 통해 주식 수익률의 비정상성을 모델링함으로써 더욱 현실적인 투자 환경을 반영하였다. 이외에도 수익률 분포의 비정상성을 반영하기 위한 통계적인 모델들이 도입되었지만, 자산 간의 상호작용을 포착하는 데는 한계가 있었다. 본 연구에서는 군집별 평균 수익률을 토대로 투자자의 뷰를 설정하여 K-shape 군집화와 블랙-리터만 모델을 결합한 포트폴리오 최적화 방법론을 제안하였다. 시계열 군집화를 통한 포트폴리오 최적화는 자산을 유사한 수익률 패턴을 공유하는 그룹으로 세분화함으로써, 포트폴리오 내 자산들의 상호작용을 세밀하게 분석할 수 있다.

2.2 시계열 데이터 군집화

시계열 데이터는 시간에 따른 연속적인 값들의 집합으로, 순서와 시간 간의 패턴을 고려하지 않으면 시간 정보가 손실될 수 있다. 이는 시장의 변동성, 최근의 이벤트, 외부 정보와 같은 요인들이 시간의 경과와 함께 최근의 시장 동향에 크게 영향을 미치기 때문이다. 따라서, 이러한 시계열 데이터의 복잡한 특성과 변화하는 동향을 정확

하게 포착하고 반영하기 위해 이에 적합한 군집화 방법이 필요하다.

군집화는 데이터의 집합을 동질적인 그룹, 즉 군집으로 분할하는 비지도 학습 방법론을 의미한다. 각 군집은 내부의 데이터 포인트들이 서로 유사한 특징을 가지며, 다른 군집의 데이터 포인트와는 구별되는 특징을 지니게 된다. 군집화의 목적은 주어진 데이터를 유사한 특징을 가지는 객체끼리 군집을 구성함으로써, 데이터의 구조나 패턴을 찾아내는 것이다. 대표적인 군집화 알고리즘으로는 K-Means 군집화가 널리 알려져 있다.

K-Means 군집화는 주어진 데이터를 K 개의 군집으로 묶는 기법으로, 각 군집의 중심점을 기반으로 한 데이터 포인트들의 평균 거리를 측정한다[8]. 평균 거리를 기반으로 비슷한 특성을 지닌 데이터들을 K 개의 동일한 군집에 할당하고, 각 군집의 중심과 해당 군집에 속하는 데이터 포인트 간의 거리의 합을 최소화한다. 수식으로 표현하면, K-means 군집의 목적함수 J 는 다음과 같다.

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

여기서 μ_i 는 i 번째 군집의 중심점, C_i 는 i 번째 군집에 속하는 데이터 포인트의 집합, $\|x - \mu_i\|$ 는 데이터 포인트 x 와 그 데이터 포인트가 속한 군집의 중심점 μ_i 사이의 유클리드 거리를 나타낸다. K-Means 군집화 알고리즘은 먼저 각 데이터 포인트를 가장 가까운 중심점에 할당한 뒤, 할당된 데이터 포인트들의 평균을 기반으로 중심점을 다시 계산한다. 이 두 과정을 반복하여 군집화를 수행한다.

이처럼 K-Means는 간결하면서도 효과적인 군집화를 위한 주요 도구로 사용된다. 하지만 K-Means 군집화는 시계열 데이터의 복잡한 패턴과 시간에 따른 변화의 연속성을 완벽하게 잡아내지 못한다는 점에서 한계가 존재한다. K-Means 군집화는 각 데이터 포인트의 거리를 기반으로 군집화를 수행하는데, 이때 모든 데이터 포인트는 동등한 중요도를 가진다. 그러나 시계열 데이터에서는 최근의 데이터 포인트가 과거의 데이터보다 더 큰 영향을 미치는 경우가 많다. 또한, 시간의 흐름에 따른 데이터의 패턴을 찾아내지도 못한다. 따라서 K-Means와 같은 전통적인 군집화 방법을 사용하면 최근의 중요한 정보나 패턴을 제대로 반영하지 못할 수 있다. 이러한 단점을 극복하기 위해 시계열 데이터에 특화된 군집화 방법이 요구되고, 대표적으로 Dynamic Time Warping(DTW)와 K-shape 군집화 방법이 있다.

DTW는 두 시계열 데이터 간의 유사도를 측정할 때, 시간적인 변화에 따른 왜곡의 최소화 방법을 제공한다[4]. 예를 들어, 주식 A의 과거 6개월 가격 움직임이 주식 B의 현재 4개월 가격 움직임과 유사할 경우, 시간의 길이와 시

점이 달라도 DTW를 사용하여 유사도를 계산할 수 있다. 이러한 특성 때문에 시계열 데이터의 길이나 패턴이 약간 다르더라도 비슷한 군집을 구성할 수 있다. 두 시계열 $X = \{x_1, x_2, \dots, x_m\}$ 와 $Y = \{y_1, y_2, \dots, y_n\}$ 가 주어졌을 때, DTW 거리는 다음과 같이 계산된다.

$$DTW(i,j) = d(x_{i,y_j}) + \min\{DTW(i-1,j), DTW(i,j-1), DTW(i-1,j-1)\} \quad (2)$$

$d(x_i, y_j)$ 는 x_i 와 y_j 사이의 거리를 나타내며, 보통 유클리드 거리로 계산된다. 하지만 이 과정은 계산적으로 매우 복잡하며, 시계열 데이터의 길이가 길어질수록 상당한 시간이 소요된다는 단점이 존재한다. 이러한 DTW의 계산적 제약을 극복하기 위해, K-shape 군집화가 도입되었다. K-shape 군집화는 시계열 데이터에 특화된 군집화 방법으로, 시계열 데이터의 모양(shape)에 기반하여 군집화를 수행한다. 기본 원리는 시계열 데이터 간의 유사도를 최대화하는 동시에 각 군집 내의 시계열들이 패턴에 따라 군집화될 수 있도록 한다. 자세한 과정은 다음과 같다. 첫째, 초기 중심값으로 k 개의 무작위 시계열 데이터를 선택한다. 둘째, 각 시계열 데이터 x_i 에 대해, 가장 유사한 중심 s_j 를 찾아 그 군집에 할당한다. 이때 유사도는 SBD를 사용하는 데, SBD의 계산은 다음 단계들로 이루어진다.

먼저 두 시계열 데이터 간의 교차상관계수(cross-correlation)를 계산하여 두 데이터 간의 최적의 정렬을 찾는다. 교차상관계수는 두 시계열 데이터 간의 상관관계를 시간 지연을 고려하여 측정하는 방법으로, 두 시계열 사이의 시간적 관계나 패턴의 일치 정도를 알 수 있다. 수식은 다음과 같다.

$$CC_\tau(x,y) = \sum_{t=1}^T x(t) \cdot y(t+\tau) \quad (3)$$

여기서 τ 는 시간 차이(lag)를 나타내며, $x(t)$ 는 시계열 x 의 t 시점에서의 값이며, $y(t+\tau)$ 는 시계열 y 의 $t+\tau$ 시점에서의 값을 나타낸다. 최적의 정렬에서 얻어진 교차상관계수를 기반으로 두 시계열 데이터 간의 SBD를 다음과 같이 계산할 수 있다. 이 때, SBD 값이 낮을수록 두 시계열 데이터는 더 유사하다고 판단된다.

$$SBD(x,y) = 1 - \frac{\max_\tau CC_\tau(x,y)}{\sqrt{norm(x) \cdot norm(y)}} \quad (4)$$

셋째, 각 군집에 할당된 시계열 데이터들의 SBD를 기반으로 새로운 중심값을 계산한다. 이 과정에서도 역시 교차상관계수를 활용하여 군집 내의 최적의 정렬을 찾아 SBD

를 계산한다. 마지막으로, 중심값의 위치 변화가 없거나 미리 설정한 반복 횟수에 도달하면 알고리즘을 종료한다.

본 연구에서는 K-shape 군집화의 하이퍼파라미터인 K 의 값에 따라 어떻게 군집화 결과가 변화하는지, 그리고 이 변화가 투자 전략의 수익률에 어떠한 영향을 미치는지를 살펴볼 것이다. 이를 위해 다양한 K 값에 대하여 실험하여 최적의 군집 수를 도출하려고 한다. 또한, 다른 군집화 방법론과의 비교를 통해 K-shape 군집화가 어떻게 주식 시장의 패턴을 더욱 정확하게 파악하고 이를 투자전략에 효과적으로 적용할 수 있는지에 대한 방안을 제시한다.

2.3 포트폴리오 이론

본 논문은 시계열 데이터 군집화를 기반으로 한 블랙-리터만 포트폴리오 투자전략을 제시한다. 자산을 어떻게 구성할 것인지, 어떤 자산에 어떤 비율로 투자할 것인지 결정하는 과정은 투자의 핵심 문제 중 하나이다. 이러한 문제를 해결하기 위해 여러 이론들이 제시되어 왔으며, 이 중에서도 포트폴리오 이론은 주요한 역할을 하였다. 본 연구의 목표는 이러한 포트폴리오 이론을 시계열 데이터 군집화와 연계하여 새로운 투자 전략을 탐색하는 것이다.

포트폴리오 이론의 중심에는 자산의 예상 수익률과 위험 사이의 균형을 찾는 것에 있다. 전통적으로, 마코위츠 포트폴리오 이론이 이 균형을 찾기 위한 주요 방법으로 널리 사용되었다. 이 이론은 최적의 자산 조합을 결정하기 위해 예상 수익률과 위험을 고려한다. 포트폴리오의 예상 수익률 μ 와 분산 σ_p^2 는 다음과 같이 계산된다.

$$\mu = \sum_{i=1}^n w_i r_i \quad (5)$$

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_i \sigma_j \rho_{ij} \quad (6)$$

이 때, r_i , w_i , w_j , σ_i , σ_j , ρ_{ij} 는 자산 i 의 예상 수익률, 자산 i, j 의 포트폴리오 가중치, 자산 i, j 의 표준편차, 자산 i, j 간의 상관계수를 의미한다.

마코위츠 모델은 투자자들이 다양한 자산 간의 상관관계에서 목표 수익률을 달성할 수 있는 최소 위험의 포트폴리오를 선택할 수 있도록 해주는 장점이 있다. 그러나 이 이론은 모든 투자자가 동일한 정보와 가정 하에 결정을 내리는 것을 전제로 한다는 한계를 지니고 있다. 실제 시장 환경에서는 정보의 비대칭성, 시장의 변동성, 그리고 개별 투자자의 주관적인 판단은 포트폴리오 선택에 큰 영향을 준다. 특히, 마코위츠 이론은 과거 데이터에 의존하며 미래의 변동성을 완벽하게 반영하지 못한다. 이에 동일 정보의 전제와 미래 예측의 한계를 적절히 반영하기 위한

새로운 접근법의 필요성이 대두되었고, 이에 블랙-리터만 포트폴리오 모델이 제안되었다.

블랙-리터만 모델은 투자자의 주관적인 의견이나 시장에 대한 견해인 뷰를 포트폴리오 최적화 과정에 포함시켜, 보다 현실적인 투자 환경에 적합한 투자 전략을 도출할 수 있도록 돕는다. 블랙-리터만 모델의 예상 수익률 r_{exp} 과 가중치 \hat{w} 는 다음과 같다.

$$r_{exp} = ((\tau\Sigma)^{-1} + P^T\Omega^{-1}P)^{-1}((\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}Q) \quad (7)$$

$$\hat{w} = \frac{\Sigma^{-1}r_{exp}}{1^T\Sigma^{-1}r_{exp}} \quad (8)$$

이 때, τ , Σ , P , Ω , π , Q 는 시장정보 대비 뷰의 가중치, 자산 간의 공분산 행렬, 뷰 행렬, 뷰의 불확실성 행렬, 기대 수익률, 뷰가 반영된 기대 수익률을 의미한다.

뷰는 투자자가 보유하고 있는 시장에 대한 주관적인 의견이나 예상을 나타낸다. 뷰는 대체로 특정 자산 또는 자산군에 대한 예상 수익률이나 성과에 대한 정보를 제공한다. 다음은 투자자의 주관적인 뷰를 구체적이고 효과적으로 반영하는 방법을 세 가지 제시한다.

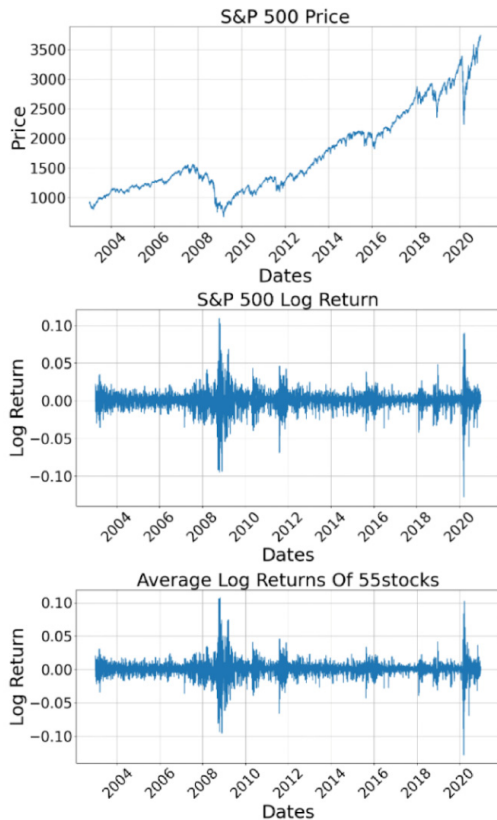
전망을 나타내는 방식에는 크게 절대적 전망(Absolute View)과 상대적 전망(Relative View)이 있다[7]. 절대적 전망은 특정 자산이나 자산군에 대한 직접적인 예상 수익률을 나타내는데, 예를 들면, 다가오는 12개월 동안 IT부문이 연간 10%의 수익률을 기록할 것이라는 형태로 제시되는 견해가 있다. 반면, 상대적 전망은 한 자산이나 자산군이 다른 특정 자산이나 자산군에 비해 얼마나 성과를 나타낼 것인지 상대적인 수익률에 대한 예상으로 표현한다. 금융 부문이 다음 해에 헬스케어 부문에 비해 더 높은 수익률을 기록할 것이라는 예상을 예로 들 수 있다. 이렇게 투자자가 시장에 대한 자신만의 예상이나 의견을 가지고 있을 때, 포트폴리오 최적화와 자산 배분 전략을 결정하는데 있어 핵심적인 역할을 한다. 또한, 포트폴리오 운용자는 자신의 전망에 대한 확신의 정도, 즉 신뢰도(confidence level)로도 표현할 수 있다. 이 신뢰도는 주관적 전망의 분산을 나타내는 대각행렬로 표현되며, 전망에 대한 확신의 정도와 역의 관계를 갖게 된다.

본 연구에서는 수익률 시계열 데이터를 기반으로 블랙-리터만 포트폴리오를 구성하였다. 주어진 자산들을 군집화 기법들을 사용하여 15개의 군집으로 분류하고, 각 군집에 대한 뷰 Q 는 해당 군집 내 자산들의 최근 60영업일의 평균 수익률을 사용하였다. 뷰의 확실성을 나타내는 행렬 Ω 은 0.01의 값으로 초기화된 대각 행렬을 설정하였는데, 이는 투자자의 뷰에 대한 불확실성이 상대적으로 낮다는 가정을 반영하였다.

3. 시계열 군집화 기반 포트폴리오 구성

3.1 데이터 전처리

본 연구에서는 2003년 1월부터 2020년 12월까지의 S&P500 지수에 속한 주식 중 대표적으로 평가되는 55개 회사의 주식 데이터를 선택하여 포트폴리오를 구성한다. 미국 주식 시장은 Global Industry Classification Standard (GICS) 분류에 의해 11개의 섹터로 나뉘어지고, 각 섹터에서 시가 총액 상위 5개씩을 선택한다면 S&P500 지수를 대표할 수 있는 55개 종목을 구성할 수 있다. 실제로, 해당 55개 종목의 대표성을 확인하기 위해, 55개 종목의 평균 로그 수익률과 S&P500 지수의 로그 수익률을 확인할 수 있고, 그 결과는 <Figure 1>과 같다.



<Figure 1> Price and Log Returns of S&P500, and the Average Log Returns of 55 Stocks

이 때, 2003년 1월부터 2020년 12월까지 55개 주식에 대한 수정주가 데이터를 사용하였고, 로그 수익률은 다음의 수식을 통해 계산하였다.

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (9)$$

이 때, r_t 시점 t 에서의 로그 수익률, P_t 는 시점 t 에서의 주식 가격, P_{t-1} 는 시점 $t-1$ 에서의 가격을 나타낸다. <Figure 1>에 의하면, S&P500 지수의 로그 수익률과 섹터별로 뽑힌 55개 종목의 평균 로그 수익률은 시간에 따라 거의 같은 모습을 보이고 있다. 따라서, 해당 55개 종목은 S&P500 지수를 대표할 수 있고, 이 종목들을 통해 포트폴리오를 구성하고자 한다. <Table 1>은 이 55개 종목의 통계치를 나타낸다. Jarque-Bera 테스트와 augmented Dickey-Fuller 테스트 결과, 55개 종목 모두 수정주가 데이터가 비정상성이 나타났다. 따라서, 정상성을 띄도록 로그 수익률로 변환하여 사용할 수 있고, 군집화 시에는 0부터 1까지 값을 가지도록 최소-최대 정규화 과정을 추가적으로 거쳐 사용하였다.

3.2 시계열 군집화 기반 포트폴리오 백테스팅

포트폴리오의 구성은 3가지 단계로 이루어진다. 첫 번째는 유사도를 포착하는 단계로써, 현재를 t 로 기준 삼는다면, $t-60$ 에서 $t-1$ 까지의 시계열이 해당된다. 55개 종목에 대해 유사도 포착 단계 동안 60개 표준화 로그 수익률을 통해 군집화를 수행한다. 이 때, 시계열 데이터에 특화된 군집화 방법을 사용하기 위해 K-shape 군집화 방법을 사용하고, 적합한 군집의 수를 찾는 과정이 필요하다. 또한, 종목 간의 유사도를 포착하는데 있어 60개가 아닌 다른 개수에 대한 시계열을 통해서 민감도 분석을 수행할 수 있다.

두 번째는 같은 군집에 속한 종목들에 대해서 블랙-리터만 포트폴리오의 뷰를 같도록 설정한다. 블랙-리터만 포트폴리오의 투자자의 개인적인 예상치인 뷰를 설정할 수 있고, 이 연구에서는 각 군집의 평균 수익률을 뷰로 설정하였다. 같은 군집에 속한 종목들은 앞으로도 같은 수익률이 예상된다는 가정을 의미하는 뷰이고, 지난 60일 간의 평균 수익률이 앞으로도 유지될 것이라는 가정도 포함한다. 따라서, 각 군집별 평균 수익률을 바탕으로 Q 행렬을 구성하였고, 각 군집에 포함된 주식들에 대하여 동일한 가중치를 부여하였다. 즉, P 행렬은 군집 내 주식들이 동등한 가중치를 가지도록 설계하였다. 이 때, Ω 행렬은 뷰의 신뢰성을 나타내며, 본 연구는 모든 군집에 대해 동일한 값인 0.01로 설정하였다.

세 번째는 뷰를 반영한 블랙-리터만 포트폴리오 비중을 바탕으로 각 종목에 투자하는 백테스팅을 수행한다. 테스트 기간은 t 에서 $t+20$ 까지의 시계열에 해당하고, 테스트 기간 또한 민감도 분석을 수행할 수 있다. 20일 동안의 백테스팅이 끝나면 첫 번째 포트폴리오가 종료되고, 다시 두 번째 포트폴리오를 준비한다. 이를 리밸런싱이라고 하고, 20일을 주기로 리밸런싱을 수행한다. 영업일을 기준으로 데이터를 처리하였기 때문에 20영업일은 약 한 달을 의미하고, 2003년 1월부터 2020년 12월까지 총 216번의 리밸런싱을 수행하였다.

<Table 1> Price and Log Returns of S&P500, and the Average Log Returns of 55 Stocks

Sector	Stock	Mean	Std	Skewness	Kurtosis	J-B	ADF
Telecommunication Services (TELE)	VZ	0.0000	0.0157	0.14	6.67	8863***	-13.9***
	DIS	0.0003	0.0189	-0.10	9.82	19200***	-30.5***
	T	-0.0001	0.0164	0.02	7.21	10365***	-29.1***
	CMCSA	0.0002	0.0209	0.00	8.37	13938***	-14.8***
	EA	0.0002	0.0268	0.13	6.08	7376***	-11.8***
Consumer Discretionary (COND)	AMZN	0.0006	0.0331	0.45	12.15	29443***	-11.3***
	HD	0.0002	0.0196	-0.98	23.38	109676***	-17.3***
	MCD	0.0003	0.0145	-0.14	6.69	8931***	-51.8***
	NKE	0.0005	0.0189	-0.33	12.03	28931***	-18.8***
	SBUX	0.0006	0.0213	0.16	7.34	10746***	-12.2***
Energy (ENRG)	XOM	0.0001	0.0152	0.01	10.37	21433***	-15.8***
	CVX	0.0002	0.0160	0.05	10.55	22167***	-17.5***
	COP	0.0003	0.0189	-0.31	5.43	5955***	-16.8***
	EOG	0.0007	0.0239	-0.04	4.72	4447***	-26.6***
	SLB	0.0001	0.0223	-0.34	5.98	7224***	-29***
Health Care (HLCR)	JNJ	0.0002	0.0121	-0.61	16.01	51366***	-15.5***
	PFE	0.0001	0.0157	-0.25	5.50	6073***	-13***
	UNH	0.0007	0.0197	0.26	20.56	84249***	-16.4***
	MRK	0.0000	0.0173	-1.41	29.08	169989***	-15.1***
	ABT	0.0003	0.0150	-0.49	8.98	16240***	-28.1***
Industrials (INDS)	BA	0.0004	0.0189	-0.28	5.80	6757***	-32.3***
	MMM	0.0003	0.0145	-0.03	5.39	5781***	-15***
	UNP	0.0005	0.0177	-0.25	4.00	3230***	-18***
	HON	0.0002	0.0196	-0.23	14.02	39175***	-12.7***
	RTX	0.0003	0.0168	-1.22	26.65	142691***	24.7***
Materials (MTRL)	PPG	0.0002	0.0174	-0.05	5.54	6125***	-16.5***
	ECL	0.0004	0.0148	-0.10	6.83	9292***	-16.2***
	APD	0.0003	0.0173	-0.16	5.72	6531***	-19.5***
	SHW	0.0006	0.0178	-0.53	14.86	44207***	-12.8***
	NEM	0.0001	0.0255	0.19	4.69	4406***	-52.7***
Financials (FINC)	JPM	0.0001	0.0246	0.26	14.01	39170***	-11.6***
	BRK-B	0.0004	0.0140	0.73	12.40	31052***	-17.7***
	BAC	0.0000	0.0292	-0.34	27.24	147913***	-10.6***
	WFC	0.0002	0.0239	0.87	27.94	156133***	-10.6***
	C	-0.0004	0.0309	-0.54	42.03	352042***	-10.6***
Consumer Staples (CONS)	PG	0.0001	0.0136	-4.07	109.01	2379971***	-15.6***
	KO	0.0001	0.0130	0.03	9.11	16518***	-19.1***
	PEP	0.0002	0.0123	0.04	11.57	26671***	-18.2***
	WMT	0.0001	0.0151	0.10	7.05	9904***	-14.2***
	MO	0.0002	0.0235	-28.17	1428.99	407335173***	-70.1***
Information Technology (INFT)	AAPL	0.0007	0.0269	-4.36	117.99	2787784***	-15.1***
	MSFT	0.0001	0.0193	-0.13	9.64	18514***	-29.1***
	INTC	0.0000	0.0235	-0.47	9.03	16409***	-16.6***
	CSCO	-0.0001	0.0245	0.15	9.13	16607***	-12.3***
	ORCL	0.0001	0.0248	-0.01	8.52	14458***	-16.5***
Utilities (UTIL)	NEE	0.0004	0.0140	0.17	9.64	18539***	-20.4***
	DUK	0.0001	0.0154	-0.27	11.89	28196***	-13.3***
	D	0.0003	0.0133	-0.62	9.93	19952***	-28***
	SO	0.0002	0.0120	0.16	6.35	8053***	-52.6***
	EXC	0.0002	0.0162	-0.05	8.01	12784***	-12.6***
Real Estate (REES)	AMT	0.0004	0.0302	-0.51	24.77	122448***	-13.8***
	SPG	0.0004	0.0213	0.22	19.40	75033***	-12.3***
	CCI	0.0003	0.0299	-0.30	25.11	125671***	-13***
	PLD	0.0002	0.0250	-0.99	32.23	207696***	-11.7***
	PSA	0.0005	0.0187	0.11	16.09	51541***	-11.8***

3.3 포트폴리오 수익성 지표

투자 전략의 성과 평가를 위한 다양한 지표들이 존재하고, 본 연구에서는 다음의 다섯 가지 지표를 사용한다. 샤프 비율(Sharpe Ratio, ShR)은 초과수익률에 대한 위험을 측정하는 데 널리 사용되는 지표이다. 포트폴리오의 수익률에서 무위험 수익률을 뺀 값, 즉 초과수익률을 포트폴리오의 표준편차로 나누어 계산한다. 이 지표는 위험 대비 수익률을 평가하는데 효과적이며, 샤프 비율이 높을수록 해당 포트폴리오가 위험 단위당 높은 수익을 제공하는 것을 의미한다. 소르티노 비율(Sortino Ratio, SoR)은 하방 리스크에 초점을 맞춘 성과지표로, 하락에 대한 리스크 대비 초과 수익률을 뜻한다. 샤프 비율과는 달리 소르티노 비율은 수익률의 하방 표준편차만을 고려하여 계산한다. 이익 승수(Profit Factor, PF)는 포트폴리오의 총 수익에 대한 총 손실의 비율을 나타낸다. 이 지표는 투자자에게 해당 포트폴리오가 얼마나 효과적으로 수익을 창출하는지에 대한 정보를 제공한다. 최대 손실 낙폭(Maximum Drawdown, MDD)은 주어진 기간 동안 포트폴리오 가치의 최대 하락률을 나타내는 지표로, 이 지표의 값이 크면 클수록 포트폴리오의 위험이 높아진다고 볼 수 있으며 이는 특정 시기에 투자자가 겪을 수 있는 최대 손실을 나타낸다. 칼마 비율(Calmar Ratio, CR)은 위험 대비 수익성을 측정하는 데 사용되는 지표로 연평균 수익률을 최대 손실 낙폭으로 나눈 값으로 계산된다. 높은 칼마 비율은 높은 수익을 가져오는 동시에 낮은 최대 손실 낙폭을 가진 포트폴리오를 의미한다. r_{rf} 는 무위험수익률이라고 하고, r_i , σ_i , σ_i^- , p_i , l_i , pv_i , lp_i 는 포트폴리오 i 의 수익률, 표준편차, 하방 표준편차, 총 수익, 총 손실, 피크, 피크 후 가장 낮은 수익률이라고 할 때, 위 지표들에 대한 수식은 다음과 같다.

$$ShR_i = \frac{r_i - r_{rf}}{\sigma_i} \quad (10)$$

$$SoR_i = \frac{r_i - r_{rf}}{\sigma_i^-} \quad (11)$$

$$PF_i = \frac{p_i}{l_i} \quad (12)$$

$$MDD_i = \frac{lp_i - pv_i}{pv_i} \quad (13)$$

$$CR_i = \frac{r_i}{MDD_i} \quad (14)$$

또한, 본 연구에서는 위의 주요 지표들 외에도 평균 수익률, 표준편차, 하방 표준편차, 총 수익, 총 손실, 수익 기간, 손실 기간 등의 추가 지표를 활용하여 투자 전략의 수익성을 평가하였다. 수익성 지표 계산 시, 벤치마크와의

쉬운 비교를 위해서 연율화된 수익률을 바탕으로 계산하였다.

해당 시계열 데이터 군집화 기반 포트폴리오의 수익성에 대한 상대적인 비교를 위해 4가지 벤치마크를 구성하였다. 첫 번째는 S&P500 지수에 대하여 매수 후 보유 방식의 벤치마크를 구성한다. 두 번째는 마코위츠 포트폴리오 이론을 바탕으로 투자한다. 세 번째는 K-Means 군집화 방법을 통해 각 주식을 군집화하고 그 군집을 바탕으로 블랙-리터만 포트폴리오를 구성한다. 네 번째는 DTW 거리를 바탕으로 한 K-Means 군집화 방법을 통해 각 주식을 군집화하고 블랙-리터만 포트폴리오와 비교를 수행한다.

또한, 해당 포트폴리오의 수익성이 꾸준히 유지되는지 확인하기 위하여 기간별로 수익성을 살펴보고, 머신러닝의 하이퍼파라미터에 강건한지, 학습 기간에 따라 강건한지 확인하기 위하여 민감도 분석을 수행한다. K-shape 군집화 방법은 군집의 수에 따라 다른 결과를 나타내고, 학습 기간에 따라서도 다른 결과를 나타낼 수 있다. 뿐만 아니라, 투자 기간에 따라서도 수익성이 달라질 수 있기 때문에 이에 따른 민감도 분석이 필요하다.

4. 백테스팅 결과 및 분석

4.1 수익성 평가

해당 포트폴리오의 수익성을 평가하고자 2003년 1월부터 2020년 12월까지 백테스팅을 진행하였다. 20영업일을 주기로 리밸런싱하여 18년간 매년 12번의 리밸런싱이 일어난다. 12개의 포트폴리오 수익률을 연율화하였고, 연간 수익률에 대한 주요 수익성 지표는 <Table 2>와 같다. 벤치마크로서 S&P500 지수의 매수 후 보유 전략, 마코위츠 포트폴리오, K-Means 군집화 기반 블랙-리터만 포트폴리오, DTW 거리를 통한 K-Means 군집화 기반 블랙-리터만 포트폴리오를 사용하여 수익성을 비교하였고, 각 지표마다 최고 값을 진하게 표시하였다.

벤치마크로 사용된 S&P500 지수는 연간 평균 수익률이 8.9%로 나타났고, 마코위츠 포트폴리오는 약간 낮은 8.4%로 S&P500에 근접한 수치를 보였다. 이에 비해, K-Means, DTW 기반 K-Means, 그리고 K-shape 군집화 포트폴리오는 모두 13% 이상의 연간 평균 수익률을 보였다.

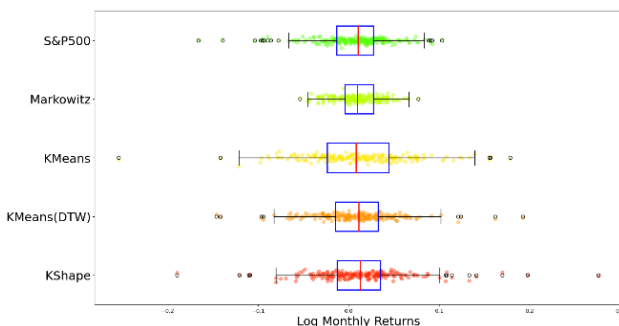
특히, K-shape 군집화 포트폴리오가 전체적으로 두드러진 성과를 보였다. 수익성 지표로 가장 많이 사용되는 샤프 비율 측면에서 0.722로 다른 포트폴리오와 비교하여 가장 높은 값을 보였으며, 이익 승수에서도 1.991로 가장 높은 성과를 나타냈다. 마코위츠 포트폴리오는 샤프 비율 0.607, 소르티노 비율 0.677, 그리고 MDD -0.51로, 다른

군집화 기반 포트폴리오 전략들에 비해 상대적으로 낮은 성과를 기록하였다. K-means 군집화 포트폴리오는 샤프 비율 0.69, 이익 승수 1.845, 그리고 MDD -0.432로 마코위츠에 비해 샤프 비율과 MDD에서 조금 더 높은 결과를 보였다. 그리고 DTW 기반 K-means 포트폴리오는 샤프 비율에서 0.717, 이익 승수에서 1.976로 높은 값을 보이며, K-shape 포트폴리오 다음으로 높은 성과를 나타냈다.

<Table 2> Annualized Profitability Metrics

Metrics	S&P500	Markowitz	K-means	K-means (DTW)	K-shape
Mean return	0.089	0.084	0.155	0.138	0.142
Standard deviation	0.105	0.195	0.196	0.165	0.170
Sharpe ratio	0.489	0.607	0.690	0.717	0.722
Sortino ratio	0.548	0.677	1.099	1.119	1.085
Downside deviation	0.122	0.095	0.123	0.107	0.115
Profit factor	1.651	1.872	1.845	1.976	1.991
Gross profit	3.977	3.245	6.072	5.031	5.129
Gross loss	2.409	1.733	3.293	2.555	2.577
MDD	-0.708	-0.51	-0.432	-0.372	-0.495
Calmar ratio	0.123	0.165	0.366	0.403	0.297
Profitable years	16	14	14	14	15
Unprofitable years	2	4	4	4	3

결론적으로, 시계열 군집화 기반 블랙-리터만 포트폴리오가 일반적인 K-Means 군집화 기반 블랙-리터만 포트폴리오보다 전반적으로 우수한 수익성을 보여주었으며, 가장 우수한 성과를 보인 것으로 평가된다. 이로부터 군집화 기반의 뷰 설정이 블랙-리터만 포트폴리오의 전반적인 수익성 향상에 기여했음을 확인할 수 있다. <Figure 2>는 월별 수익률에 대한 박스플롯이다.



<Figure 2> Boxplot of Monthly Returns

K-shape 군집화 방법은 가장 높은 중앙값을 보였고, 이

는 평균적인 월별 수익률이 높다는 것을 의미한다. 그 다음으로는 DTW 기반 K-Means 군집화가 높은 성과를 보였다. S&P500의 경우, 중앙값이 비교적 높게 나타났으나 이상치가 왼쪽으로 치우쳐져 있었고, 마코위츠 포트폴리오는 중앙값이 상대적으로 낮았지만, 이상치가 적고 하방 표준편차가 작아 안정적인 성과를 보였다. 마지막으로 K-means 군집화 방법의 중앙값은 가장 낮았으며, 분포의 범위도 가장 넓게 나타났다. 따라서, K-shape 군집화 방법이 평균적으로 높은 수익률을 보였고, K-means 군집화 방법은 가장 낮은 중앙값과 넓은 분포로 인해 상대적으로 불안정한 성과를 보이는 것으로 판단된다.

4.2 기간별 수익성 평가

시계열 군집화 기반 포트폴리오 전략의 샤프 비율이 벤치마크보다 높은 숫자를 기록하였지만, 기간에 따라 유효성에 차이가 있을 수 있다. 따라서, 해당 전략의 수익성이 꾸준할지 평가하기 위해 2003년부터 2020년까지 투자 기간을 4개 기간으로 나누어 수익성 지표를 계산하였다. 주식 데이터의 학습기간은 60일로 설정하여 이 기간 동안의 데이터를 바탕으로 군집화를 수행하였고, 이후 투자기간은 20일로 설정하여 이 기간 동안의 포트폴리오 수익률 및 위험성을 분석하였다. 군집의 개수는 총 15개로 설정하였다. <Table 3>은 기간별로 측정된 주요 수익성 지표를 나타낸다.

<Table 3> Sub-period Performance of Hierarchical Clustering

Period	Sharpe ratio	Sortino ratio	Profit factor	MDD	Calmar ratio
2003-2007	1.899	3.101	4.837	-0.117	2.262
2008-2012	0.021	0.027	1.081	-0.801	0.031
2013-2016	0.615	0.987	1.721	-0.111	0.898
2017-2020	0.833	1.331	1.981	-0.145	0.987

2003~2007년 기간 동안 포트폴리오는 전체적으로 높은 수익률을 보여주었다. 특히, 샤프 비율이 1.899, 소르티노 비율 3.101 및 이익 승수 4.837으로 모두 높은 값을 기록하였고, 최대 손실 낙폭은 -0.117로, 상대적으로 낮은 값을 보여주고 있다. 해당 기간 동안 고수익과 낮은 변동성을 함께 보였음을 확인할 수 있다. 칼마 비율 또한 2.262로 높은 값을 보여주어, 수익 대비 리스크가 적절하게 관리되었으며, 이 기간 동안 포트폴리오의 비교적 안정적인 성과를 나타낸다.

반면, 2008~2012년 기간의 지표는 글로벌 금융 위기의 영향을 반영하고 있는 것으로 보인다. 포트폴리오의 수익성 지표가 전체적으로 상당히 저하되었는데, 샤프 비율과 소르

티노 비율의 급격한 하락과 함께 최대 손실 낙폭의 값이 -0.801로 크게 증가하였다. 이는 해당 기간 동안의 큰 손실을 시사하며 포트폴리오 위험이 상대적으로 높았음을 나타낸다.

2013~2016년 기간은 이전의 급격한 하락에 비해 상대적으로 회복을 보여주었다. 성과지표들은 2008~2012년에 비해 상승하였으며, 샤프비율 0.615, 소르티노 비율 0.987까지 도달하였고, 최대 손실 낙폭은 -0.111로 측정되었다. 특히 칼마 비율 0.898으로, 포트폴리오의 안정성이 다시 회복되었음을 확인할 수 있다.

2017~2020년에는 지속적인 성장과 안정적인 성과를 보여주고 있다. 샤프 비율이 0.833으로, 고수익과 낮은 변동성의 조화를 나타내며, 소르티노 비율이 1.331로, 양호한 위험 대비 수익성을 보였다. 이 기간의 이익승수는 1.981, 최대 손실 낙폭은 -0.145로, S&P500 지수의 최대 손실 낙폭 값인 -0.708보다 훨씬 높은 수치를 보여준다. 또한, 다른 지표들도 일정한 수준을 유지하며 상대적으로 안정된 포트폴리오 성과를 보여준다. 따라서, 글로벌 금융 위기 이후 첫 구간만큼의 높은 성과를 보여주지는 않지만, 투자 기회가 여전히 남아있음을 확인할 수 있다.

4.3 민감도 분석

K-Means 군집과 K-shape 군집의 경우 하이퍼파라미터인 군집의 수에 따라 군집화의 결과가 달라진다. 군집의 수는 포트폴리오 구성 및 최적화 과정에 중요한 파라미터로 세분화된 주식 시장의 특징을 반영한다. 일반적으로는 실루엣 계수나 엘보우 기법 등을 통해 하이퍼파라미터 최적화를 수행하지만, 하이퍼파라미터 값에 따라 수익성 지표가 강건하지 않다면 투자전략의 위험성이 존재할 수 있다. 따라서, 군집의 수에 따라 민감도 분석을 수행하였고, 군집의 수를 5, 10, 15, 20으로 설정하여 포트폴리오 성과 지표의 변화를 분석하였다. 이때, 학습기간은 60일, 투자기간은 20일로 설정하였고, 결과는 <Table 4>와 같다.

<Table 4> Sensitivity Analysis by the Number of Clusters

K	Sharpe ratio	Sortino ratio	Profit factor	MDD	Calmar ratio
5	0.457	0.667	1.58	-0.62	0.18
10	0.534	0.744	1.723	-0.62	0.214
15	0.722	1.085	1.991	-0.495	0.279
20	0.644	0.893	1.87	-0.513	0.266

전체적으로 군집의 개수가 증가함에 따라 샤프 비율과 소르티노 비율은 초기에 증가하는 경향을 보였다. 특히 군집의 수가 5에서 15로 증가할 때, 샤프 비율이 0.457에서 0.722, 소르티노 비율이 0.667에서 1.085로 급격히 상승하는

것을 확인할 수 있다. 그러나 군집의 수가 20일 때는 샤프 비율이 0.644로 약간 감소하였는데, 이는 군집이 너무 세분화되어 포트폴리오의 위험을 증가시킬 수 있는 것으로 보인다. 최대 손실 낙폭과 칼마 비율 역시 군집의 수가 15일 때 가장 좋은 성과가 나타났다. 따라서, 군집의 수가 15일 때 포트폴리오의 안정도가 가장 잘 조절됨을 알 수 있다.

군집화의 학습기간과 투자기간에 따라서도 포트폴리오의 성능에 차이가 발생할 수 있고, 이에 따른 민감도를 평가하기 위해 다양한 학습 기간(20일, 40일, 60일) 및 투자 기간(20일, 40일, 60일) 조합에 대하여 실험을 진행하였다. 이 때, 군집의 개수는 15로 설정하여 포트폴리오 최적화를 수행하였다.

학습 기간은 군집화의 학습 및 블랙-리터만 모델의 뷰 설정에 사용되는 데이터의 범위를 의미한다. 이 기간에 수집된 데이터는 시장의 동향을 충분히 파악하여 이를 잘 반영하는 예상 수익률 및 리스크를 포착한다. 투자 기간은 포트폴리오 전략의 적용기간을 나타낸다. 이 기간은 단기 투자와 장기투자의 적합성에 따라 다를 수 있으며, 학습 데이터를 기반으로 해당 투자기간 동안 최적의 포트폴리오 구성을 테스트하였고, 결과는 <Table 5>와 같다.

<Table 5> Sensitivity Analysis by the Moving Windows

Moving window	Sharpe ratio	Sortino ratio	Profit factor	MDD	Calmar ratio
(20,20)	0.276	0.384	1.376	-0.863	0.078
(20,40)	0.388	0.334	1.811	-0.521	0.179
(20,60)	0.518	0.465	2.418	-0.391	0.308
(40,20)	0.427	0.503	1.567	-0.569	0.18
(40,40)	0.401	0.357	1.805	-0.392	0.231
(40,60)	0.309	0.211	1.752	-0.678	0.119
(60,20)	0.722	1.085	1.991	-0.495	0.279
(60,40)	0.682	0.618	2.377	-0.517	0.257
(60,60)	0.553	0.377	2.412	-0.601	0.218

본 연구에서는 민감도 분석을 통해 다양한 학습기간 및 투자기간 조합을 조사하였다. 제시된 표에서 볼 수 있듯이, 여러 조합 중 학습 기간 60일, 투자 기간 20일 조합이 다른 모든 조합에 비해 가장 우수한 성능을 보였다. 이 조합의 샤프비율은 0.722로, 모든 조합 중 가장 높은 값을 기록하였다. 또한, 소르티노 비율 역시 1.085로 최고 수치를 보였으며, 이익 승수 값은 1.991로 안정적인 수익률을 나타내었다. 학습 기간이 60일인 경우, 샤프비율은 모든 투자 기간과의 조합에서 전반적으로 높은 값을 나타냈는데, 20일 0.722, 40일 0.682, 60일 0.553을 기록하였다.

반면, 학습 기간이 20일인 조합은 전체적으로 성능이 다소 낮았는데, 특히 학습 기간 20일, 투자 기간 20일 조합은

샤프 비율이 0.276, 소르티노 비율 0.384로 가장 낮은 결과를 기록했다. 이 조합은 학습 기간과 투자 기간이 동일하게 20일로 짧아, 시장의 전반적인 트렌드나 패턴을 충분히 학습하지 못하고, 빠른 투자 기간 동안의 빠른 변동을 적절히 대응하지 못했을 것으로 판단된다. 나머지 기간 역시 40일과 60일 각각 샤프 비율이 0.388, 0.518로 상대적으로 낮은 수치를 보였다.

K-shape 군집화와 학습기간 및 투자기간 튜닝을 적용한 결과, 주식 시장의 복잡한 패턴과 변동성을 효과적으로 포착하고 분석할 수 있었으며, 이로 인해 달성된 Sharpe Ratio는 0.722로, 다른 기법들에 비해 상당히 높은 성과를 보여주었다. 특히, 학습기간과 투자기간의 조합이 포트폴리오 성과에 큰 영향을 미친다는 점을 확인할 수 있었는데, 이는 학습 기간과 투자 기간이 달라짐에 따라 주식 시장의 변동성과 시장 패턴의 반영도가 다르다는 것을 의미한다. 분석 결과를 종합하면, 충분한 학습 기간과 적절한 투자 기간을 설정함으로써, 시장의 변동성에 유연하게 대응하면서도 안정적인 수익률을 도출할 수 있었다.

5. 결론

본 연구는 주식 시장의 패턴을 깊이 분석하는 과정에서 두 주요 방법론을 결합하였다. 첫째, K-shape 군집화를 통해 주식 시장의 시계열 데이터를 효과적으로 그룹화하였다. 이를 통해 주식 시장의 여러 동적 패턴과 그 패턴들의 특성을 정밀하게 파악할 수 있었다. 둘째, 블랙-리터만 포트폴리오 최적화 방법을 활용하여 군집화된 패턴에 근거한 투자 방향성을 설정하였다. 이렇게 구성된 포트폴리오는 과거 시장 데이터를 기반으로 미래의 시장 동향을 예측하고, 적합한 투자 방향성을 도출하였다. 또한, 백테스팅을 수행하여 통해 해당 방법론의 효과와 유효성을 검증하였다.

K-shape 군집화는 시계열 데이터의 복잡성 속에서도 주식 시장의 동적 패턴을 명확히 분류하고 해석할 수 있게 하였다. 이러한 군집화의 결과는 블랙-리터만 포트폴리오 구성에 결정적인 역할을 하였으며, 이 과정에서 포트폴리오의 탄력성을 향상시켜 다양한 시장 조건에 유연하게 대응할 수 있는 구조를 마련하였다. 특히, 기간별 수익성 평가, 민감도 분석 등을 통해 포트폴리오의 수익률과 위험을 균형있게 조절하는 전략을 탐색하였다. 이는 주식 시장의 포트폴리오 최적화에 어떠한 영향을 미치는지에 대한 중요한 통찰을 제공하며, 이러한 접근법이 포트폴리오 관리 전략의 새로운 방향을 제시함을 시사한다.

본 연구에서는 시가총액 기준 상위 55개의 대형주를 선별하여 분석을 진행하였다. 이 대형주들은 높은 시가총액으로 인해 큰 규모의 투자 변동에도 가격 안정성이 유지될

가능성이 높으며, 이는 시뮬레이션 결과가 실제 시장 상황과 크게 벗어나지 않음을 의미한다. 따라서, 대형 연기금과 같은 대형 기관 투자자들도 사용할 수 있는 실용적인 전략이다. 대형 연기금의 경우 장기적인 자본 성장과 리스크 관리를 추구해야 하는데, 해당 전략을 통해 자산 배분을 효과적으로 적용할 수 있다. 대형주들은 그 안정성에 비례하여 상대적으로 높은 수익률을 기대하기 어렵다는 단점을 지니고 있지만, 리스크 관리와 자본 보존에 중점을 둔 장기 투자에는 적합한 투자전략으로 기대한다. 실제 금융 기관에서는 백테스팅을 통해 이 전략의 유효성을 입증할 수 있으며, 다양한 시장 조건과 종목 풀을 통해 포트폴리오 성과를 실증적으로 검증할 수 있다.

본 연구의 방법론은 포트폴리오 최적화를 넘어 금융 산업의 리스크 관리, 자산 관리, 신용 평가 등 다양한 분야에 적용할 수 있다. K-shape 군집화와 블랙-리터만 모델의 결합은 복잡한 금융 시계열 데이터를 이해하고, 전략적인 자산 배분 결정을 최적화하는 데 있어 높은 적응성을 제공한다. 특히, 고변동성 시장 조건에서 리스크를 세밀하게 분석하고 관리하는 데 있어 군집의 특성을 사용할 수 있다. 자산 관리 부문에서는 개인의 투자 행태 및 정서, 경제 사이클에 따른 성과를 분석해 포트폴리오를 구성할 수 있다 [3]. 신용 평가 분야에서는 K-shape 군집화를 활용해 채무자의 신용 위험을 보다 정밀하게 평가하고, 시장 변동성과 대출 상환 능력 사이의 상호작용을 고려함으로써 대출 조건의 설정과 신용 리스크 관리를 개선할 수 있다.

향후 연구에서는 기관 투자자의 시장 전망을 더욱 세밀하게 반영하는 것이 필요하다. 이는 기관 투자자의 특정 시장에 대한 기대 수익률, 장기 및 단기 시장 예측, 그리고 각 자산에 대한 뷰의 강도와 방향성 등을 분석함으로써, 포트폴리오 최적화 모델을 실제 시장 상황에 더욱 밀착시킬 수 있게 한다. 또한, 운용기금의 다양한 특성을 반영하여 블랙-리터만 모형의 파라미터를 조정하는 것 역시 중요한 과제로 남아 있다.

더 나아가, 다양한 시계열 군집화 방법론과 포트폴리오 최적화 모델을 결합하여 비교 연구를 수행하는 것도 필요하다. 특히, 데이터의 특성과 성격에 따라 어떤 방법론이 더 효과적인지를 판단하는 기준을 세워, 연구의 방향성을 더욱 강화할 수 있을 것이다. 본 연구에서 제시한 방법론의 활용성을 더욱 확장하고 깊이 있게 연구함으로써, 포트폴리오 최적화의 실질적 효과를 극대화하는 방법에 대한 통찰을 얻을 수 있을 것이다.

Acknowledgement

This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korea gov-

ernment (MSIT) (No. 2021R1F1A1062917).

This work was supported by the Gachon University research fund of 2023.(GCU-202300670001)

References

- [1] Aghabozorgi, S. and The, Y. W., Stock Market Co-movement Assessment Using a Three-phase Clustering Method, *Expert Systems with Applications*, 2014, Vol. 41, No. 4, pp. 1301-1304.
- [2] Barua, R. and Sharma, A. K., Dynamic Black Litterman Portfolios with Views Derived Via CNN-bilstm Predictions, *Finance Research Letters*, 2022, Vol. 49, pp. 103111.
- [3] Barua, R. and Sharma, A. K., Using Fear, Greed and Machine Learning for Optimizing Global Portfolios: A Black-Litterman Approach, *Finance Research Letters*, 2023, Vol. 58, p. 104515.
- [4] Berndt, D. J. and Clifford, J., Using Dynamic Time Warping to Find Patterns in Time Series, in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 1994, ACM, pp. 359-370.
- [5] Black, F. and Litterman, R., Global Portfolio Optimization, *Financial Analysis Journal*, 1992, Vol. 48, No. 5, pp. 28-43.
- [6] Cho, P., Lee, M., and Song, J. W., Clustering-driven Pair Trading Portfolio Investment in Korean Stock Market, *Journal of Korean Society of Industrial and Systems Engineering*, 2022, Vol. 45, pp. 123-130.
- [7] Choi, Y. M., Liu, W., and Jhang, H., A Study on Strategic Asset Allocation using the Black-Litterman Model of Global Investors, *The Korean Journal of Financial Management*, 2021, Vol. 38, No. 4, pp. 293-314.
- [8] Hartigan, J. A. and Wong, M. A., Algorithm as 136: A k-means Clustering Algorithm, *Journal of the Royal Statistical Society. Series C*, 1979, Vol. 28, No. 1, pp. 100-108.
- [9] Markowitz, H., Portfolio Selection, *The Journal of Finance*, 1952, Vol. 7, No. 1, pp.77-91.
- [10] Moon, K. S. and Kim, H., Efficient Asset Allocation Based on Prediction with Adaptive Data Selection, *Economic Computation & Economic Cybernetics Studies & Research*, 2023, Vol. 57, No. 1, pp. 57-72.
- [11] Niennattrakul, V., Ratanamahatana, C. A., On Clustering Multimedia Time Series Data Using k-means and Dynamic Time Warping, In *2007 International Conference on Multimedia and Ubiquitous Engineering*, 2007, IEEE, pp. 733-738.
- [12] Pang, T. and Karan, C., A Closed-form Solution of the Black-Litterman Model with Conditional Value at Risk, *Operations Research Letters*, 2018, Vol. 46, No. 1, pp. 103-108.
- [13] Paparrizos, J. and Gravano, L., k-shape: Efficient and Accurate Clustering of Time Series, In *Proceedings of 2015 ACM SIGMOD International Conference on Management of Data*, 2015, ACM, pp. 1855-1870.
- [14] Putra, Y. E., Saepudin, D., and Aditsania, A., Portfolio Selection of Kompas100 Stocks Index Using b-spline Based Clustering, *Procedia Computer Science*, 2021, Vol. 179, pp. 375-382.
- [15] Rezaei, H., Faaljou, H., and Mansourfar, G., Intelligent Asset Allocation Using Predictions of Deep Frequency Decomposition, *Expert Systems with Applications*, 2021, Vol. 186, p. 115715.
- [16] Sahamkhadam, M., Stephan, A., and Östermark, R., Copula-based Black-litterman Portfolio Optimization, *European Journal of Operational Research*, 2022, Vol. 297, No. 3, pp. 1055-1070.
- [17] Tayali, S. T., A Novel Backtesting Methodology for Clustering in Mean-variance Portfolio Optimization, *Knowledge-Based Systems*, 2020, Vol. 209, p. 106454.
- [18] Wang, Y. and Aste, T., Dynamic Portfolio Optimization with Inverse Covariance Clustering, *Expert Systems with Applications*, 2023, Vol. 213, pp. 118739.
- [19] Wu, D., Wang, X., and Wu, S., Construction of Stock Portfolios Based on k-means Clustering of Continuous Trend Features, *Knowledge-Based Systems*, 2022, Vol. 252, pp. 109358.

ORCID

Yeji Kim | <https://orcid.org/0009-0008-2357-3887>

Poongjin Cho | <https://orcid.org/0000-0002-1844-0472>