

뉴스 감성 분석을 반영한 포트폴리오 선정 모형의 투자 성과 분석: 한국 주식시장을 중심으로*

박주미¹ · 정민수¹ · 김홍선² · 김성문^{1†}

¹연세대학교 경영대학 경영학과, ²LG 에너지솔루션

Adjusted Portfolio Selection Model Reflecting News Sentiment Analysis: Focusing on South Korean Stock Market*

Joomi Park¹ · Minsu Chung¹ · Hongseon Kim² · Seongmoon Kim¹

¹School of Business, Yonsei University, ²LG Energy Solution

■ Abstract ■

This paper proposes an adjusted portfolio selection model: SA-M portfolio selection model. SA-M portfolio model is based on Markowitz's mean-variance portfolio selection model, but it adjusts the total proportion of capital invested in risky assets considering the result of news sentiment analysis. We applied KR-FinBERT to news headline data which were gathered from 8 different economic newspapers. The performance of this model was evaluated by using historical monthly stock return data of South Korean stock market. An investment was held from September 2012 to August 2023. As a result of the investment, SA-M portfolio selection model outperformed benchmark portfolios, suggesting the usefulness of sentimental analysis in terms of developing investment strategies.

Keywords : Sentiment Analysis, Mean-Variance Portfolio Selection Model, Investment Performance Analysis, News Headline, KOSPI, South Korean Stock Market

논문접수일 : 2024년 07월 25일 논문게재확정일 : 2024년 11월 30일

논문수정일 : 2024년 11월 27일

* 이 논문은 2024년도 연세대학교 연구비의 지원을 일부 받아 수행된 연구임(2024-22-0187). 또한, 연세대학교 경영연구소의 연구비 지원을 일부 받음.

† 교신저자, kims@yonsei.ac.kr

1. 서 론

보편적인 주식 투자자는 낮은 위험으로 높은 수익률을 달성할 수 있는 포트폴리오를 이상적으로 여길 것이다. 그러나 일반적으로 수익률과 변동성은 비례하기에 수익성과 안정성을 동시에 추구하기는 어렵다. 마코위츠(Harry M. Markowitz)[39]는 이러한 문제를 해결할 수 있는 포트폴리오 선정 모형(Portfolio Selection Model)을 고안하였다. 마코위츠 포트폴리오 선정 모형은 과거의 주가 데이터를 사용하여 입력값인 기대수익률과 분산을 계산한 후, 투자자의 최저 요구 기대수익률을 만족하는 수준에서 가장 분산이 작은 포트폴리오를 구성한다. 즉, 마코위츠 포트폴리오 선정 모형을 통해 수익성은 높이고 위험은 낮은 투자를 할 수 있다. 그러나 마코위츠 포트폴리오 선정 모형은 입력값에 따른 민감도가 높아, 입력값의 작은 변화에도 매우 다른 포트폴리오 투자 비중이 구해질 수 있다는 문제가 있다. Jobson et al.[30]는 과거 데이터를 기반으로 구한 예측값과 실제값 간의 차이로 인한 예측오차가 발생하는 문제를 제시하였다. Michuad[40]는 예측오차를 고려하였을 때 마코위츠 포트폴리오 선정 모형으로 구한 포트폴리오가 최적이라고 할 수 없음을 지적하였다.

입력값 추정의 정확도를 개선하고자 다양한 방법론이 고안되었다. 불확실성을 고려하여 베이지안식(Bayesian) 접근법[15, 31, 43, 46], 강건 최적화(Robust Optimization)[14, 19, 23, 25, 26] 등이 제안되었다. 또한 기대수익률을 고려하지 않는 최소 분산 포트폴리오(Minimum Variance Portfolio)가 제안되었다[29]. 반면 실증 실험을 통해 예측 오차의 내재에도 불구하고 마코위츠 포트폴리오 선정 모형이 유효함을 보여준 선행 연구들이 있다[2, 5, 6, 10, 32]. 국내의 주식시장에서의 실험 결과는 불완전 정보 하에서도 마코위츠 모형이 여전히 좋은 투자 성과를 낼 수 있음을 제시하였다.

이 외에도 단순한 형태의 마코위츠 포트폴리오 선정 모형으로는 현실의 주식시장을 충분히 반영하지 못한다는 문제가 있기에 제약조건을 추가하는 연구

가 진행되었다. Jagannathan et al.[29]는 비중 벡터를 제한하여 예측 정확도를 향상시킬 수 있음을 제시하였으며, DeMiguel et al.[20]는 포트폴리오의 놈(norm)을 제한하는 제약조건을 추가해 투자 성과를 향상시킬 수 있음을 보였다. 더 나아가 거시경제지표의 주가 예측력을 이용하여 직전 월의 거시경제지표가 당월의 주식수익률에 미치는 사전적 기대를 제약조건에 반영하는 포트폴리오 선정 모형이 연구되었다. 양지혜 외[8]는 건화물시황 운임지수(Baltic Dry Index)와 이자율을 각각 반영한 포트폴리오 선정모형을 설계하여 포트폴리오의 수익성과 안정성을 개선하였다. 이렇듯 주가 예측력을 지닌 추가 정보를 제약조건으로 반영하여 포트폴리오의 성과를 개선하는 방법이 연구되고 있다.

한편 텍스트의 감성을 분석하여 추가적인 정보를 얻는 방법이 주목을 받고 있다. 이는 자연어처리 기법(Natural Language Processing, NLP)의 일종인 감성 분석(Sentimental Analysis)을 통해 실현 가능하다. 감성 분석은 텍스트 데이터에서 분위기, 감정, 의견 등의 주관적인 정보를 추출하는 방법으로, 시장 분석, 리뷰 분석, 소셜 미디어 분석 등에 널리 사용된다. 더 나아가 뉴스에 감성 분석을 적용하여 주가의 변동을 예측하고자 하는 시도가 활발하게 이루어지고 있다. 뉴스 감성은 주가 선행성과 주가 후행성을 모두 갖는다. Mo et al.[41]은 뉴스 감성이 시장 수익률에 영향을 미치기까지 평균 5일 정도의 시간 지연을, 시장 수익률이 뉴스 감성에 영향을 미치기까지 평균 1일 정도의 시간 지연을 갖으며 결과적으로 뉴스 감성과 시장 수익률 간에 상호 피드백 메커니즘이 발생함을 입증하였다. 즉, 뉴스는 투자자의 의사결정에 영향을 미치며, 뉴스에 대한 투자자의 반응이 주가의 변동을 야기한다[24]. 더 나아가 Kalyani et al.[33], Li et al.[37], Nemes et al.[42]은 뉴스 감성 분석을 통해 주가 예측 성능을 개선할 수 있음을 제시하였다.

그러나 범용 사전으로는 금융 분야에서 용어가 사용되는 맥락을 파악하기 어려워 텍스트의 감성을 잘못 분류할 수 있다는 문제점이 있다[38]. 이러한 문

제는 2018년 구글이 개발한 BERT(Bidirectional Encoder Representations from Transformers) [21]를 활용하여 개선되었다. BERT는 범용 언어 모델로 개발되었으나, 특정 분야의 데이터를 추가적으로 사전 학습한다면 해당 분야에 적합하게 적용 가능하다. 이에 따라 기존 BERT에 금융 데이터 사전학습 및 미세 조정(fine-tuning)을 적용해 금융 분야에 특화된 자연어처리가 가능한 FinBERT가 개발되었다[12, 44].

국내에서도 한국어 텍스트에 대한 금융특화 자연어 처리를 할 수 있는 KR-FinBERT가 고안되었다[35]. KR-FinBERT를 통해 한국 주식시장에 감성 분석을 적용하는 연구가 진행되고 있다. 장주현 외[9]는 기사 헤드라인의 감성 분석을 통해 KOSPI 지수 예측 성과를 개선하였으며, 김대겸 외[1]는 뉴스 데이터에서 추출한 감성 표현을 어텐션(attention) 기반 모델에 추가하는 추가 예측 방법을 제안하였다. 김성진 외[3]는 증권사 리포트를 감성 분석한 후 종목별 점수를 추출하여 상위 종목을 매수하는 투자 전략을 수립하였다. 이러한 선행연구 결과는 KR-FinBERT가 한국 주식시장의 전망 예측 및 투자자의사결정에 사용될 수 있음을 보여준다.

단순히 주가를 예측하는 것에서 더 나아가 감성 분석의 결과를 기존 포트폴리오 선정 모형과 결합하고자 하는 연구가 진행되고 있다. Colasanto et al.[18]은 FinBERT를 통해 종목별 감성 점수를 추출한 후 이를 Black-Litterman model과 결합하여 포트폴리오의 성과를 향상시켰다. Hung et al.[28]은 Black-Litterman model에 GRU(gated recurrent unit) 등의 딥러닝 기술을 적용하였으며 이때 뉴스 감성 분석 결과를 활용하여 모델의 과적합(overfitting) 문제를 해소하였다.

그러나, 그 유효성에도 불구하고 한국어 금융 텍스트에 대한 감성 분석 결과를 전통적인 포트폴리오 선정 모형과 결합하여 한국 주식시장에 적용하는 연구는 아직 진행되지 않았다. 불완전 정보에도 불구하고 마코위츠 포트폴리오 선정 모형을 한국 주식시장에 적용해 우수한 투자 성과를 얻은 여러 실증 연구 결

과[2, 4, 6]를 고려할 때, 마코위츠 포트폴리오 선정 모형에 감성 분석 정보를 추가한다면 유의미한 투자 성과 개선이 있을 것으로 예상된다.

본 연구는 KOSPI 지수에 대한 감성 분석을 진행한 후 그 결과로 한국 주식시장의 전망을 판단하여 위험 자산 투자 비중을 조정하는 포트폴리오 선정 모형을 설계하고, 그 투자 성과를 살펴보고자 한다. 시장 전체의 상황을 나타내는 하나의 감성지수를 사용하여 전체 종목에 일괄적으로 적용하였기에 Black-Litterman model보다 비교적 형태가 단순한 마코위츠 포트폴리오 선정 모형을 사용하였다. 본 연구가 제시하고자 하는 포트폴리오 선정 모형은 기존 마코위츠 포트폴리오 선정 모형에 감성 분석의 결과를 반영하는 제약조건을 추가하여 장의 상황에 따른 능동적인 투자가 가능하도록 한다. 즉, 감성 분석의 결과 긍정적인 전망이 예측될 경우 위험자산에 공격적으로 투자하며, 부정적인 전망이 예측될 경우 위험자산 투자비중을 줄이고 무위험자산 투자비중을 높여 보수적으로 투자한다. 본 연구는 포트폴리오 선정 모형의 우수성을 입증하기 위해 실증실험을 진행하였다. 2012년 9월부터 2023년 8월까지의 11년을 투자 기간으로 설정하였으며, 투자 시작 시점의 시가총액을 기준으로 한국 주식시장의 30개 종목을 선정하여 그 월간 주식수익률 데이터를 가지고 실험을 진행해 투자 성과를 확인하였다.

본 연구는 다음과 같은 차별점을 갖는다. 한국 주식시장에 감성 분석을 적용하여 주가를 예측하는 연구가 진행되고 있지만, 감성 분석을 이용해 한국 주식 포트폴리오를 구성하고 그 성과를 평가하는 연구는 초기 단계로 김성진 외[3] 등 일부만 존재한다. 본 연구는 감성 분석의 결과로 위험자산 투자비중을 조정하는 포트폴리오를 제시함으로써, 한국 주식시장에서 포트폴리오를 선정할 때 감성 분석이 유효한 투자 전략인지를 확인하고자 한다. 또한 전통적인 포트폴리오 이론인 마코위츠 포트폴리오 선정 모형에 감성 분석 정보를 추가하여 투자 성과를 개선하는 융합 연구는 아직 연구되지 않았다. 본 연구는 감성 분석의 결과를 반영하는 제약조건을 추가하여 기존 마코위츠 포트폴리오 선정 모형의 성과를 유의미하게 개

선택할 수 있음을 제시하고자 한다. 본 연구를 통해 한국 주식시장에서 포트폴리오 모델과 감성 분석의 결합이 가지는 잠재력을 보이고자 한다. 더 나아가 추후 더 복잡한 포트폴리오 전략으로 확장될 가능성을 제시하며, 후속 연구를 위한 기초적인 토대를 마련하는 학문적 기여를 이루고자 한다.

본 논문의 구성은 다음과 같다. 제2장에서는 감성 분석 과정과 이를 반영한 포트폴리오 선정 모형을 제안한다. 이어 제3장에서는 한국 주식시장을 대상으로 하는 실증 실험의 설계를 설명한다. 제4장에서는 해당 포트폴리오의 투자 성과를 벤치마크와 비교해 제시한다. 마지막으로 제5장에서는 본 연구의 결론과 한계점을 정리하고 앞으로의 연구 방향에 대하여 논의한다.

2. 모형 설계

제2장은 본 연구에서 사용하는 포트폴리오의 구성을 제시한다. 2.1절은 본 연구의 바탕이 되는 마코위츠 포트폴리오 선정 모형을 소개한다. 2.2절은 감성 분석을 통해 감성지수를 도출하는 과정을 보여준다. 2.3절은 감성지수를 통해 마코위츠 투자비중을 보정한 SA-M 포트폴리오 선정 모형을 제안한다.

2.1 마코위츠 포트폴리오 이론

본 연구는 마코위츠(Markowitz)[39] 포트폴리오(Simple Markowitz Model, 이하 SMM 포트폴리오)를 기반으로 한다. SMM 포트폴리오의 구성은 다음과 같다.

- N = 포트폴리오의 투자대상 종목의 수
- $\omega_{i,t}$ = 주기 t 에 종목 i 에 투자하는 비율
($i=1, 2, \dots, N$), ($t=1, 2, \dots, T$)
- $\mu_{i,t}$ = 주기 t 에 종목 i 의 기대수익률
($i=1, 2, \dots, N$), ($t=1, 2, \dots, T$)
- $\sigma_{ij,t}$ = 주기 t 에 종목 i 와 j 의 수익률에 대한 공분산 ($i, j=1, 2, \dots, N$), ($t=1, 2, \dots, T$)

- K_t = 주기 t 에 포트폴리오에 설정한 최저요구수익률 ($t=1, 2, \dots, T$)
- V_t = 주기 t 에 포트폴리오의 수익률에 대한 분산(기대위험도) ($t=1, 2, \dots, T$)

$$\text{Minimize } V_t = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij,t} \omega_{i,t} \omega_{j,t} \quad (1)$$

$$\text{subject to } \sum_{i=1}^N \mu_{i,t} \omega_{i,t} \geq K_t \quad (2)$$

$$K_t = \frac{1}{N} \sum_{i=1}^N \mu_{i,t} \quad (3)$$

$$\sum_{i=1}^N \omega_{i,t} = 1 \quad (4)$$

$$\omega_{i,t} \geq 0 \quad \forall i, t \quad (5)$$

식 (1)은 포트폴리오의 수익에 대한 분산을 최소화한 최적의 투자비율을 구하고자 하는 SMM 포트폴리오의 목적함수를 나타낸다. 식 (2)는 포트폴리오가 최소 기대수익률을 만족해야 함을 나타낸다. 이때 필요한 주기 t 의 최소 기대수익률 K_t 은 Jung et al.[32]이 제시한 식 (3)을 통해 결정된다. K_t 는 기대수익률 $\mu_{i,t}$ 의 평균값으로 두어, 시장 전망에 따라 유동적인 최소 기대수익률을 가질 수 있다. 식 (4)는 모든 가용금액을 투자해야 함을, 식 (5)는 공매도를 허용하지 않음을 보여준다.

2.2 감성 분석

본 연구는 SMM 포트폴리오에 감성지수를 더하여 투자 효과를 개선한 새로운 포트폴리오를 제안하고자 한다. 감성 분석을 통한 감성지수의 도출 과정은 다음과 같다.

본 연구는 한국언론진흥재단의 뉴스통합 DB인 빅카인즈(BIGKinds)[11]를 활용하여 2010년 8월 1일부터 2023년 7월 31일까지의 뉴스 데이터를 수집하였다. 경제일간지(매일경제·머니투데이·서울경제·아시아경제·아주경제·파이낸셜뉴스·한국경제·헤럴드경제) 8종을 대상으로 “코스피”를 키워드로 넣고 검색하여 뉴스 기사를 선정하였다. 중복 기

사를 제외하였으며, 기호 및 특수문제 제거, 괄호 내의 부가적인 메타데이터 제거 과정을 거쳤다. 뉴스 헤드라인의 특성상 텍스트 데이터가 짧고 명사 중심으로 구성되어 형태소 분석이나 토큰화는 생략하였다. 총 298,914건의 기사가 최종적으로 감성 분석 대상이 되었다.

이후 KR-FinBERT[35]를 사용하여 감성지수를 산출하였다. KR-FinBERT는 한국어 NLP(자연어처리)가 사전 훈련된 KR-BERT-Medium에 경제 신문 등의 금융 텍스트를 학습시켜 한국 금융시장 관련 언어를 처리할 수 있도록 확장한 모듈이며, 텍스트를 입력받아 바로 결과를 출력할 수 있는 구조로 설계되어 있다. KR-FinBERT는 뉴스 기사 내용을 긍정(positive), 부정(negative), 중립(neutral)로 분류한다. 텍스트에 대해 각 클래스에 해당하는 확률값을 출력해, 가장 높은 값을 가진 클래스로 분류되며, 96.3%의 정확도를 가지고 있는 것으로 밝혀져 있다 [35].

2023년 6월 30일의 뉴스 기사 67건에 대해 직접 감성을 분류하여 KR-FinBERT의 분석 결과와 비교 해본 결과는 <표 1>과 같다. 긍정(positive) 클래스의 민감도(sensitivity)는 92.31%, 특이도(specificity)는 87.04%였다. 부정(negative) 클래스의 민감도는 85.71%, 특이도는 91.67%였다. 중립(neutral) 클래스의 민감도는 74.47%, 특이도는 90.0%였다. 민감도는 실제로 해당 값인 데이터(양성 데이터) 중 모델이 해당 값으로 예측한 비율, 특이도는 실제로 해당 값이 아닌 데이터(음성 데이터) 중 모델이 해당 값이 아니라고 예측한 비율을 뜻한다.

<표 1> 2023년 6월 30일의 뉴스 헤드라인 감성 분석 Confusion Matrix

실제값\예측값	예측값 긍정	예측값 중립	예측값 부정
실제값 긍정	12	1	0
실제값 중립	7	35	5
실제값 부정	0	1	6

이후 식 (6)과 같이 d 일의 m 번째 기사에 대해 감

성지수 $\gamma_{m,d}$ 를 부여했다. $\gamma_{m,d}$ 는 KR-FinBERT가 해당 기사를 positive로 분류한 점수 $c_{m,d,Pos}$ 에 1을 곱한 수와 negative로 분류한 점수 $c_{m,d,Neg}$ 에 -1을 곱한 수를 더하여 산출하였다. 총 298,914건의 기사에 대하여 $\gamma_{m,d}$ 의 평균값은 0.042584였으며, $\gamma_{m,d} > 0$ 인 기사가 153,251건, $\gamma_{m,d} < 0$ 인 기사가 145,662건, $\gamma_{m,d} = 0$ 인 기사가 1건으로 나타났다.

이어서 d 일의 기사 총 M 개에 대하여, $\gamma_{m,d}$ 의 평균으로 해당 일자의 감성지수인 Γ_d 를 구하였다. 이 과정은 식 (7)에 나타나 있다. 총 4,748일에 대하여 Γ_d 의 평균은 0.040389이었으며, $\Gamma_d > 0$ 인 일자가 3,018일, $\Gamma_d < 0$ 인 일자가 1,730일로 나타났다.

Γ_d 를 구한 방식과 유사하게, 식 (8)과 같이 t 번째 투자주기에 속하는 총 D 개의 일자 d 에 대하여, Γ_d 의 평균값으로 투자 주기 t 의 감성지수 S_t 를 도출했다. 투자 실험 기간과 최초 감성 참조 주기를 더한 총 156개 주기에 대해 그 감성지수를 구하였으며, S_t 의 평균은 0.040387이었다. $S_t > 0$ 인 주기가 123회, $S_t < 0$ 인 주기가 33회로 나타났다.

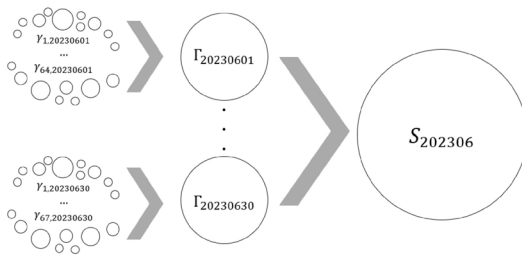
- $\gamma_{m,d}$ = 일자 d 에 기사 m 의 감성지수
($m = 1, 2, \dots, M$), ($d = 1, 2, \dots, D$)
- $c_{m,d,Pos}$ = 일자 d 에 기사 m 이 positive로 분류된 점수
($m = 1, 2, \dots, M$), ($d = 1, 2, \dots, D$)
- $c_{m,d,Neu}$ = 일자 d 에 기사 m 이 neutral로 분류된 점수
($m = 1, 2, \dots, M$), ($d = 1, 2, \dots, D$)
- $c_{m,d,Neg}$ = 일자 d 에 기사 m 이 negative로 분류된 점수
($m = 1, 2, \dots, M$), ($d = 1, 2, \dots, D$)
- Γ_d = 일자 d 의 감성지수 ($d = 1, 2, \dots, D$)
- S_t = 주기 t 의 감성지수 ($t = 1, 2, \dots, T$)

$$\gamma_{m,d} = 1 * c_{m,d,Pos} + 0 * c_{m,d,Neu} + (-1) * c_{m,d,Neg} \quad (6)$$

$$\Gamma_d = \frac{1}{M} \left(\sum_{m=0}^M \gamma_{m,d} \right) \quad (7)$$

$$S_t = \frac{1}{D} \left(\sum_{d=0}^D \Gamma_d \right) \quad (8)$$

<그림 1>은 2023년 6월의 감성지수 S_{202306} 을 구하는 과정을 도식화한 것이다. 2023년 6월 1일의 기사는 총 64개로, 각각의 기사에 대해 감성지수 $\gamma_{1,20230601}, \dots, \gamma_{64,20230601}$ 를 구하였다. 이후 기사 별 감성지수의 평균으로 2023년 6월 1일의 감성지수 $\Gamma_{20230601}$ 를 구하였다. 같은 방법으로 2023년 6월의 모든 일자에 대하여 총 30개의 일자 별 감성지수를 구하였다. 이렇게 구한 일자 별 감성지수 $\Gamma_{20230601}, \dots, \Gamma_{20230630}$ 의 평균값으로 2023년 6월의 감성지수 S_{202306} 를 구하였다.



<그림 1> 2023년 6월의 감성지수 도출과정

<표 2>는 t 주기의 KOSPI 지수 변화율과 감성지수 s_{t-1} 의 상관계수를 나타낸 것이다. KOSPI 지수 변화율이 양수인 투자기간 대비 음수인 투자기간에서 상관계수가 상대적으로 유의하게 나타났다. 이는 감성지수가 특히 하락장을 예측해내는데 유효함을 보여준다. 즉, 감성지수가 낮을 경우 장의 상황이 나쁘다고 예측하고 위험자산 투자 비중을 줄이는 것이 합리적인 의사결정임을 확인할 수 있다.

<표 2> t 주기의 KOSPI 지수 변화율과 감성지수 s_{t-1} 의 상관계수

	전체 투자 기간	KOSPI 변화율이 양수인 투자 기간	KOSPI 변화율이 음수인 투자 기간
상관계수	0.503675	0.052205	0.477508

<그림 2>는 전 주기 중 s_t 가 가장 높았던 202104 주기와 s_t 가 가장 낮았던 202206 주기의 뉴스 데이터로 제작한 워드 클라우드이다. ‘코스피,’ ‘코스닥,’

‘증시,’ ‘외국인’ 등 등장 빈도가 높으며 일반적으로 사용되는 중립 주식 용어는 불용어(stop word)로 분류하였다. 워드 클라우드를 통해 2021년 4월의 데이터는 ‘상승’의 단어 등장 비중이 ‘하락’에 비해 높으며, 2022년 6월의 데이터는 이와 반대로 ‘하락’의 등장 비중이 더 높음을 확인할 수 있다.



<그림 2> 감성지수 최고 주기(a)와 감성지수 최저 주기(b)의 워드 클라우드 비교

2.3 SA-M 포트폴리오 선정 모형

본 연구는 식 (1)~(5)를 통해 구한 SMM 포트폴리오 투자비중 $\omega_{i,t}$ 를 식 (6)~(8)에서 구한 감성지수 s_t 로 보정하는 감성보정-마코위츠 포트폴리오(이하 SA-M 포트폴리오) 선정 모형을 제안하고자 한다. $\omega_{i,t}$ 를 s_t 로 보정하는 방법은 다음과 같다.

- $\overline{s_{t-1}} = t-1$ 주기로부터 n 개월 이전 동안의 감성지수 $s_{t-2}, s_{t-3}, \dots, s_{t-1-n}$ 중 최대값 ($t=1, 2, \dots, T$)
- $\underline{s_{t-1}} = t-1$ 주기로부터 n 개월 이전 동안의 감성지수 $s_{t-2}, s_{t-3}, \dots, s_{t-1-n}$ 중 최소값 ($t=1, 2, \dots, T$)
- A_t = 주기 t 에 위험자산에 적용하는 감성 보정 계수(adjusting factor) ($t=1, 2, \dots, T$)
- $\omega_{i,t}$ = 감성 보정 후 최종적으로 주기 t 에 종목 i 에 투자하는 비율 ($i=1, 2, \dots, N$), ($t=1, 2, \dots, T$)
- $\Omega_{f,t}$ = 주기 t 에 무위험자산에 투자하는 비율 ($t=1, 2, \dots, T$)

$$A_t = \begin{cases} 1 & \text{if } S_{t-1} \geq \overline{S_{t-1}} \\ \frac{S_{t-1} - \underline{S_{t-1}}}{\overline{S_{t-1}} - \underline{S_{t-1}}} & \text{if } \underline{S_{t-1}} \leq S_{t-1} < \overline{S_{t-1}} \\ 0 & \text{if } S_{t-1} < \underline{S_{t-1}} \end{cases} \quad (9)$$

$$\overline{S_{t-1}} = \max\{S_{t-1-k} \mid k=1, 2, \dots, n\} \quad (10)$$

$$\underline{S_{t-1}} = \min\{S_{t-1-k} \mid k=1, 2, \dots, n\} \quad (11)$$

$$\Omega_{i,t} = \omega_{i,t} \cdot A_t \quad (12)$$

$$\Omega_{f,t} = 1 - A_t \quad (13)$$

식 (9)~(13)은 양지혜 외[8]를 참고하였다. 식 (9)는 주기 t 에 위험자산에 적용하는 감성 보정계수 A_t 의 도출식이다. 리밸런싱 시점에 이미 습득할 수 있는 정보만을 활용할 수 있기에, 당월의 투자 계획을 세우는 데에 있어 직전 월의 감성지수를 사용하였다. 직전 월의 감성지수가 높았을 경우 시장 상황이 좋을 뜻해 당월의 주식수익률이 높을 가능성이 있다. 이 경우, 위험자산에 투자하는 비중을 높인다. 반면 직전 월의 감성지수가 낮았을 경우, 시장 상황이 좋지 않다고 볼 수 있어 당월의 주식수익률 또한 낮을 수 있다. 이 경우, 위험 자산에 투자하는 비중을 낮춰 무위험자산에 투자한다. 감성 지수의 높고 낮음을 판단하는 데 있어 식 (10)~(11)을 사용하였다. 식 (10)

은 $S_{t-2}, S_{t-3}, \dots, S_{t-1-n}$ 중 최대값을 $\overline{S_{t-1}}$ 로, 식 (11)은 $S_{t-2}, S_{t-3}, \dots, S_{t-1-n}$ 중 최소값을 $\underline{S_{t-1}}$ 로 정의한다. 이렇게 구한 최대값과 최소값을 가지고 식 (9)에서 min-max 정규화를 진행한다. $0 \leq A_t \leq 1$ 이 성립하도록 $S_{t-1} \geq \overline{S_{t-1}}$ 일 경우 A_t 를 1로, $S_{t-1} < \underline{S_{t-1}}$ 일 경우 A_t 를 0으로 제한한다.

식 (12)는 앞서 구한 SMM 포트폴리오 투자비중 $\omega_{i,t}$ 를 감성 보정계수 A_t 로 곱하여 주기 t 에 종목 i 에 대하여 감성 보정된 최종 투자비중 $\Omega_{i,t}$ 를 구할 수 있음을 보여준다. 이 때, $0 \leq A_t \leq 1$ 에 따라 $0 \leq \Omega_{i,t} \leq 1$ 이며, $\Omega_{i,t}$ 는 항상 $\omega_{i,t}$ 와 같거나 작다. 식 (13)은 전체 자산에서 위험자산에 투자하지 않은 비중은 무위험자산에 투자함을 나타낸다.

<그림 3>은 감성지수 비교 참조 기간 $n=12$ 일 때 2012년 8월의 감성 보정계수 A_{201208} 및 2012년 9월의 감성 보정계수 A_{201209} 의 도출 과정을 나타낸 예시이다. A_{201208} 를 구할 때는 2011년 8월부터 2012년 7월까지의 감성지수를 활용하였으며, 그 다음 주기가 되어 A_{201209} 를 도출할 때는 비교 범위에서 2011년 8월의 감성지수가 빠지고 2012년 8월의 감성지수가 추가된 것을 확인할 수 있다.

Date	Score						
2011-08	-0.04574		S_{201208}	$\overline{S_{201208}}$	$\underline{S_{201208}}$	$\frac{S_{201208} - \underline{S_{201208}}}{\overline{S_{201208}} - \underline{S_{201208}}}$	A_{201208}
2011-09	0.020024						
2011-10	0.080768						
2011-11	0.00298						
2011-12	0.040507						
2012-01	0.076742						
2012-02	0.060538						
2012-03	0.099626						
2012-04	0.071193						
2012-05	0.026231						
2012-06	0.047764						
2012-07	0.035979						
2012-08	0.020116		S_{201209}	$\overline{S_{201209}}$	$\underline{S_{201209}}$	$\frac{S_{201209} - \underline{S_{201209}}}{\overline{S_{201209}} - \underline{S_{201209}}}$	A_{201209}
2012-09	0.118681						

<그림 3> 감성 보정계수 A_t 도출과정 예시

3. 실험 설계

3장은 앞서 제시한 SA-M 포트폴리오 선정 모형을 적용할 투자 실험 환경을 제시한다. 3.1절은 사용한 데이터와 세부적인 실험 설정을 소개하고 투자 알고리즘을 제시한다. 3.2절은 투자 성과 비교를 위해 사용될 벤치마크 포트폴리오를 설명한다.

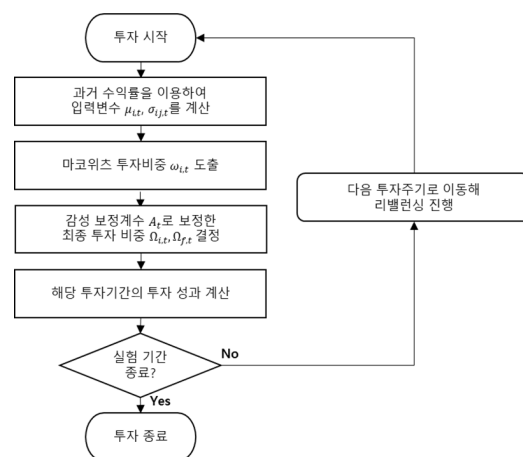
3.1 사용 데이터 및 투자 알고리즘

본 연구는 한국 주식시장을 실험 대상으로 두고 2012년 8월 말일을 기준으로 KOSPI 시가총액 상위 30개 주식 종목을 선정했다. 삼성물산, 우리금융, SK는 투자 기간 내의 합병 및 재편으로 인한 종목 변동이 있어 투자 대상에서 제외하였다. 이후 전체 투자 기간 종료 시까지 종목을 유지해 실험을 진행하였다. 실험 대상으로 선택한 종목은 <표 3>과 같다. <표 3>의 종목명은 2012년을 기준으로 작성하였다.

실험 기간은 2012년 9월 1일부터 2023년 8월 31일로, 총 11년 0개월이다. 매월 1일부터 말일까지의 1개월이 한 투자 주기를 이루며, 총 132개의 투자 주기를 가진다. 1개월마다 리밸런싱하여 투자 비중을 변경해 주었으며, 이에 맞춰 월간 주식수익률 데이터를 사용하였다. 1개월의 리밸런싱 기간은 양지혜 외[8] 등 선행연구 결과를 따라 결정하였다. 무위험 수익률의 경우 2012년부터 2023년까지의 한국 단기이자율 데이터를 사용하였다. 매년 6월 말일에 변경된 이자율 데이터가 공시되기 때문에, 당년 7월부터 후년 6월까지 1년 단위로 이자율이 적용되도록 설계하였다. 거래수수료의 경우 실제 한국 주식시장을 반영해 매수 시 거래금액의 0.015%, 매도시 거래금액의 0.215%로 적

용하였다. 주가, 단기이자율, KOSPI 지수 등 데이터는 공신력있는 금융 DB인 DATASTREAM에서 수집하였다.

본 연구는 투자시점 t 에서 최근 1년($t-12$ 기부터 $t-1$ 기까지 총 12개월)의 월간 주식수익률을 가지고 종목 i, j 의 기대수익률 $\mu_{i,t}$ 및 공분산 $\sigma_{ij,t}$ 를 구해 SMM 포트폴리오 투자비중 $\omega_{i,t}$ 를 구하였다. 이후 min-max 정규화한 감성 보정계수 A_t 로 $\omega_{i,t}$ 를 보정하여 최종 투자비중 $\Omega_{i,t}$ 를 도출하였다. <그림 4>는 본 연구에 사용된 투자 알고리즘을 순서도로 정리한 것이다.



<그림 4> 투자 알고리즘 순서도

3.2 벤치마크 포트폴리오

본 연구에서 제안한 SA-M 포트폴리오의 투자성과를 비교할 벤치마크 포트폴리오로 SMM 포트폴리오, 최소 분산 포트폴리오(Mean-variance portfolio, 이하 MVP 포트폴리오), 매수 후 보유 포트폴

<표 3> 실험대상으로 선택한 투자대상 종목

삼성전자	현대차	POSCO	기아차	현대모비스	LG화학
삼성생명	현대중공업	신한지주	한국전력	SK이노베이션	SK하이닉스
KB금융	NHN	KT&G	SK텔레콤	S-Oil	LG전자
삼성화재	LG	LG생활건강	LG디스플레이	롯데쇼핑	KT
삼성중공업	현대글로비스	하나금융지주	고려아연	삼성엔지니어링	호남석유

리오 (Buy and hold portfolio, 이하 B&H 포트폴리오), 동일 투자 비중 포트폴리오(이하 1/N 포트폴리오)를 선정하였다.

MVP 포트폴리오는 포트폴리오의 분산을 최소화하는 목적 함수를 가지며, SMM 포트폴리오와는 달리 투자자의 최저요구 기대수익률을 반영하지 않는다. 변동성을 낮추는 것에 집중하는 만큼 위험 회피성이 뛰어나 많은 선행 연구에서 우수한 성과를 보여 주었다[13, 17, 27]. MVP 포트폴리오의 구성은 식 (14)~(16)과 같다.

$$\text{Minimize } V_t = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij,t} \omega_{i,t} \omega_{j,t} \quad (14)$$

$$\text{subject to } \sum_{i=1}^N \omega_{i,t} = 1 \quad (15)$$

$$\omega_{i,t} \geq 0 \quad \forall i, t \quad (16)$$

1/N 포트폴리오는 모든 종목에 동일한 비중을 투자해 분산 투자 효과를 극대화하는 모형으로, 그 성과가 우수함이 많은 실증 연구에서 제시되었다[22, 36]. SMM 포트폴리오의 경우 앞서 2장에서 수식 (1)~(5)로 제시한 모형을 그대로 사용하였다. B&H 포트폴리오의 경우 첫 투자 주기에서 SMM 포트폴리오의 최적 마코위츠 투자비중을 구해 투자한 후, 전체 투자기간 동안 매도 없이 동일한 투자 비중을 유지하였다. 벤치마크 포트폴리오는 모두 SA-M 포트폴리오와 동일한 리밸런싱 주기(1개월) 및 거래비용을 적용하여 객관적인 투자 성과 비교가 가능하도록 구성하였다. 벤치마크 포트폴리오 외에 KOSPI 지수를 벤치마크 지수로 선정하여, 실제 한국 주식시장에서 SA-M 포트폴리오가 유효성을 가짐을 볼 수 있도록 하였다.

4. 실험결과

4장은 벤치마크 대비 SA-M 포트폴리오의 투자 성과를 다양한 지표를 통해 제시한다. 누적수익률, 샤프 비율, 소르티노 비율, 칼마 비율, GPR 차원에서

SA-M 포트폴리오와 벤치마크의 성과를 비교한다.

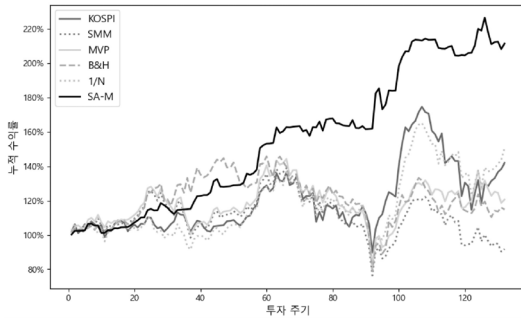
누적 수익률은 포트폴리오가 전체 투자 기간 동안 실현한 수익률로, 포트폴리오의 성과를 평가하는데 사용되는 대표적인 성과지표이다. SA-M 포트폴리오의 감성지수 비교 참조기간 $n=3, 6, 12, 24$ 로 설정하고 실험하여 얻은 누적수익률은 다음의 <표 4>와 같다.

<표 4> 감성지수 비교 참조기간 n 에 따른 누적수익률 비교

포트폴리오	누적수익률
SA-M ($n=3$)	211.44%
SA-M ($n=6$)	195.47%
SA-M ($n=12$)	188.96%
SA-M ($n=24$)	180.44%
1/N	150.60%
KOSPI	141.87%
MVP	120.63%
B&H	114.81%
SMM	91.74%

SA-M 포트폴리오는 다양한 감성분석 비교 참조기간 n 에 대하여 일관적으로 벤치마크 대비 우수한 누적수익률을 기록하였으며, 이를 통해 모형의 일반화 가능성을 확인할 수 있다. 가장 우수한 성과를 기록한 것은 최대값, 최소값 비교가 가능한 최소 기간인 3개월이었다. 즉, 감성지수를 포트폴리오 이론에 결합할 시, 최근의 변화에 가장 민감하게 반응할 수 있는 방향으로 반영했을 때 가장 좋은 성과를 낸다고 말할 수 있다. 이후의 성과지표에 대해서 SA-M 포트폴리오는 $n=3$ 으로 두고 벤치마크와 비교하였다.

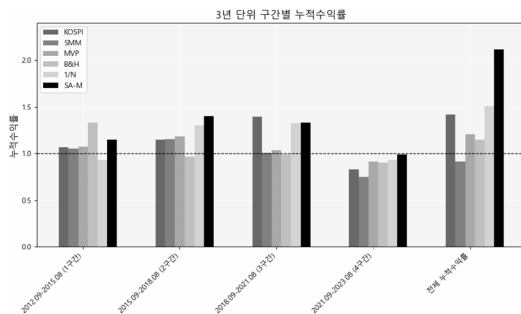
SA-M 포트폴리오와 벤치마크의 누적수익률 변화는 <그림 5>와 같다. 실험기간인 11년 동안 SA-M의 누적수익률은 211.44%로, 벤치마크 포트폴리오 대비 (SMM 91.74%, MVP 120.63%, B&H 114.81%, 1/N 150.60%) 매우 우수한 누적수익률을 기록했다. 또한 KOSPI 지수와 비교하여도(141.87%) 높은 수익률을 기록하였다.



〈그림 5〉 SA-M과 벤치마크 포트폴리오의 누적수익률 변화

〈그림 5〉에 제시된 누적수익률 변화 그래프를 보면, SA-M는 하락장에서 우수한 성과를 보였다. 특히 COVID-19 전후로 시장이 두 차례의 폭락을 겪으며 큰 변동성을 보일 때 SA-M은 감성 분석을 통해 불확실성을 예측하고 위험자산 투자 비중을 조정하여 효과적으로 대응했음을 확인할 수 있다.

〈그림 6〉은 전체 투자기간을 3년 단위로 나누어 총 4개 구간으로 구분하였을 때 각 구간 별 SA-M 포트폴리오 및 벤치마크 포트폴리오의 누적수익률 성과 비교를 나타낸 그래프이다. 전체 투자 기간이 11년이기에 마지막 구간은 3년이 아닌 2년으로 구성되었다. 1구간은 비교적 변동이 적은 보합장으로, 2구간은 상승장을 중심으로, 3구간은 보합 후 급상승장으로, 4구간은 하락장을 중심으로 이루어져 구간 별로 서로 다른 장의 특성을 반영하고 있다. 4개 구간에 대하여 SA-M 포트폴리오는 전체 6개의 포



〈그림 6〉 3년 단위 구간 별 SA-M 포트폴리오 및 벤치마크의 누적수익률 비교

트폴리오 중 항상 2위 이내의 성적을 기록하였다. 다양한 구간에 적용해보아도 꾸준히 좋은 성과를 얻을 수 있었으며, 이를 통해 SA-M 포트폴리오의 강건성을 확인할 수 있었다.

수익률만으로 포트폴리오의 성과를 판단하기에는 부족하며, 포트폴리오의 변동성 등 다른 지표가 복합적으로 고려되어야 한다. 샤프 비율(Sharpe ratio)은 포트폴리오의 변동성과 수익률을 함께 고려하는 대표적인 성과지표이다. 샤프 비율은 수익률의 변동성 대비 수익률을 나타내며, 포트폴리오의 초과수익률을 표준편차로 나뉘 구할 수 있다. 샤프 비율을 구하는 방법은 식 (17)~(20)과 같다.

- r_t = t 번째 투자기간 동안의 포트폴리오 수익률 ($t = 1, 2, \dots, T$)
- r_t^f = t 번째 투자기간 동안의 무위험 수익률 ($t = 1, 2, \dots, T$)
- M_t = t 번째 투자기간 동안의 포트폴리오의 무위험수익률 대비 초과수익률 ($t = 1, 2, \dots, T$)
- σ_p = 포트폴리오의 수익률에 대한 표준편차

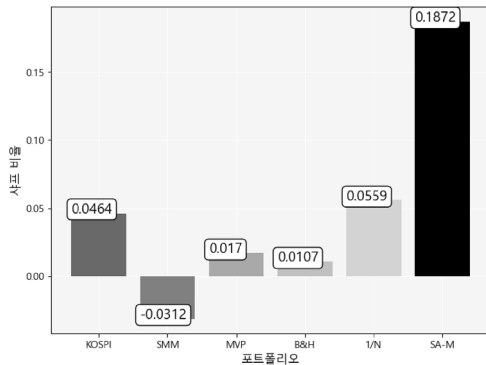
$$M_t = r_t - r_t^f \quad (17)$$

$$\tilde{M} = \frac{1}{T} \sum_{t=1}^T M_t \quad (18)$$

$$\sigma_p = \sqrt{\frac{\sum_{t=1}^T (M_t - \tilde{M})^2}{T}} \quad (19)$$

$$\text{Sharpe ratio} = \frac{\tilde{M}}{\sigma_p} \quad (20)$$

〈그림 7〉을 통해 SA-M 포트폴리오와 벤치마크의 샤프 비율을 확인할 수 있다. SA-M 포트폴리오의 샤프 비율은 0.1872로, 벤치마크(KOSPI 0.0464, SMM -0.0312, MVP 0.0170, B&H 0.0107, 1/N 0.0559) 대비 매우 우수하다. 이는 SA-M 포트폴리오를 통해 상대적으로 낮은 위험으로 높은 수익률을 얻을 수 있음을 의미한다.



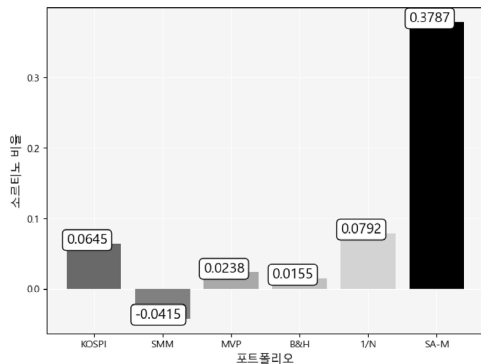
〈그림 7〉 SA-M 포트폴리오와 벤치마크의 샤프 비율

샤프 비율은 포트폴리오 수익률의 표준편차를 분모로 사용하기 때문에 포트폴리오의 수익률 하락뿐만 아니라 상승까지도 위험으로 집계되는 문제점을 가진다. 소르티노 비율(Sortino ratio)은 이러한 문제를 개선한 지표로, 포트폴리오의 초과수익률을 하방 표준편차로 나누어 수익률 하락의 경우만 위험으로 집계한다. 소르티노 비율을 구하는 방법은 식 (21)~(22)와 같다.

$$\sigma_{\rho}^d = \sqrt{\frac{\sum_{t=1}^T [\min\{M_t, 0\}]^2}{T}} \quad (21)$$

$$\text{Sortino ratio} = \frac{\tilde{M}}{\sigma_{\rho}^d} \quad (22)$$

- σ_{ρ}^d = 포트폴리오의 수익률에 대한 하방 표준편차



〈그림 8〉 SA-M 포트폴리오와 벤치마크의 소르티노 비율

〈그림 8〉을 통해 SA-M 포트폴리오와 벤치마크의 소르티노 비율을 확인할 수 있다. SA-M 포트폴리오의 소르티노 비율은 0.3787로, 벤치마크(KOSPI 0.0645, SMM -0.0415, MVP 0.0238, B&H 0.0155, 1/N 0.0792) 대비 뛰어난 성과를 보여준다. 특히 샤프 비율에서 보다 소르티노 비율에서 SA-M과 벤치마크 포트폴리오 간의 성과 격차가 늘어났다. 샤프 비율에서 집계된 SA-M 포트폴리오의 위험은 수익률 상승이 위험으로 과대 집계된 것이며, 하방 표준편차를 통해 구한 실제 위험은 더 낮음을 확인할 수 있다.

표준편차, 하방 표준편차 대신 최대 낙폭을 사용하여 최대 낙폭 대비 수익률을 볼 수도 있다. 칼마 비율(Calmar ratio)은 초과수익률을 포트폴리오의 최대 낙폭으로 나눈 것이다. 칼마 비율을 구하는 방법은 식 (23)~(25)와 같다.

- r_t^c = t 번째 투자기간까지의 누적수익률

$$(t = 1, 2, \dots, T)$$

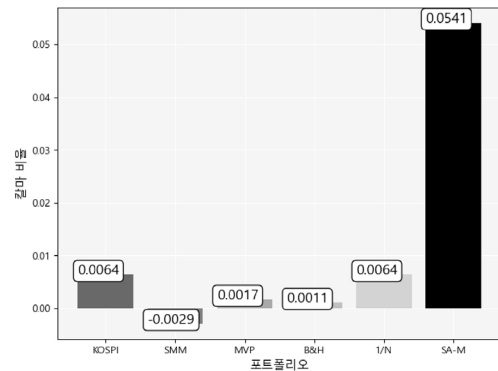
- D_t = t 번째 투자기간까지의 최대 낙폭

$$(t = 1, 2, \dots, T)$$

$$D_t = \frac{\max_{0 \leq v \leq t} (1 + r_v^c) - (1 + r_t^c)}{\max_{0 \leq v \leq t} (1 + r_v^c)} \quad (23)$$

$$\overline{D}_v = \max_{0 \leq v \leq t} D_v \quad (24)$$

$$\text{Calmar ratio} = \frac{\tilde{M}}{\overline{D}_v} \quad (25)$$



〈그림 9〉 SA-M 포트폴리오와 벤치마크의 칼마 비율

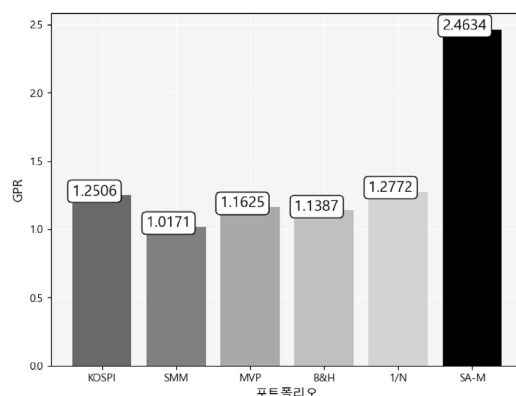
<그림 9>에서 SA-M 포트폴리오와 벤치마크의 칼마 비율을 확인할 수 있다. SA-M 포트폴리오의 칼마 비율은 0.0541로, 벤치마크(KOSPI 0.0064, SMM -0.0029, MVP 0.0017, B&H 0.0011, 1/N 0.0064) 대비 그 수치가 매우 우수하다. 이는 SA-M 포트폴리오가 상대적으로 수익률이 높을 뿐 아니라 최대 낙폭도 낮음을 보여주며, SA-M의 위험 자산 투자 비중 조정이 효과적임을 나타낸다. 벤치마크 포트폴리오들이 크게 급락을 겪을 때, SA-M은 감성 지수를 기반으로 위험자산 투자 비중을 조정하여 수익률 하락을 방어해 최대 낙폭이 작게 나타났다고 볼 수 있다.

GPR(Gain to Pain ratio)은 양의 수익률 총합을 음의 수익률 총합의 절대값으로 나누어 구할 수 있다. GPR을 통해 전체 투자기간에서 음의 수익률 대비 양의 수익률의 크기를 확인할 수 있다. 식 (26)을 통해 GPR을 구할 수 있다.

$$GPR = \frac{\sum_{t=1}^T \min\{0, r_t\}}{\sum_{t=1}^T \max\{0, -r_t\}} \quad (26)$$

<그림 10>은 SA-M 포트폴리오와 벤치마크의

GPR을 나타낸다. SA-M 포트폴리오의 GPR은 2.4634로, 벤치마크(KOSPI 1.2506, SMM 1.0171, MVP 1.1625, B&H 1.1387, 1/N 1.2772) 대비 성과가 매우 뛰어나다. GPR을 통해 SA-M의 양의 수익률 총합이 음의 수익률 총합의 절댓값 대비 두 배 이상임을 확인할 수 있다. 즉, SA-M을 통해 투자자가 얻는 이익이 손실보다 큼을 알 수 있다.



<그림 10> SA-M 포트폴리오와 벤치마크의 GPR

<표 5>는 SA-M 포트폴리오와 벤치마크의 주요 성과지표 순위를 나타낸 것이다. SA-M 포트폴리오

<표 5> SA-M과 벤치마크 포트폴리오의 주요 성과지표 순위

	KOSPI	SMM	MVP	B&H	1/N	SA-M
누적 수익률	141.87% (3)	91.74% (6)	120.63% (4)	114.81% (5)	150.60% (2)	211.44% (1)
표준편차	0.0472 (5)	0.0426 (3)	0.0419 (2)	0.0454 (4)	0.0477 (6)	0.0233 (1)
하방 표준편차	0.0339 (6)	0.0320 (4)	0.0300 (2)	0.0314 (3)	0.0336 (5)	0.0115 (1)
최대 낙폭	0.3438 (2)	0.4508 (5)	0.4298 (4)	0.4591 (6)	0.4182 (3)	0.0807 (1)
샤프 비율	0.0464 (3)	-0.0312 (6)	0.0170 (4)	0.0107 (5)	0.0559 (2)	0.1872 (1)
소르티노 비율	0.0645 (3)	-0.0415 (6)	0.0238 (4)	0.0155 (5)	0.0792 (2)	0.3787 (1)
칼마 비율	0.0064 (3)	-0.0029 (6)	0.0017 (4)	0.0011 (5)	0.0064 (2)	0.0541 (1)
GPR	1.2506 (3)	1.0171 (6)	1.1625 (4)	1.1387 (5)	1.2772 (2)	2.4634 (1)

는 누적수익률, 표준편차, 하방 표준편차, 최대 낙폭, 샤프 비율, 소르티노 비율, 칼마 비율, GPR까지의 모든 성과지표에서 1위를 기록하였다. 이를 통해 SA-M 포트폴리오가 수익성, 안정성 등 다방면에서 매우 뛰어나며, 감성 분석의 결과를 포트폴리오 구성에 활용하는 방법이 실효성 있는 투자 전략임을 확인할 수 있다.

5. 결 론

본 연구는 마코위츠 포트폴리오 선정 모형을 통해 구한 투자비중을 뉴스 헤드라인 감성 분석을 통해 산출한 감성지수로 보정해주는 SA-M 포트폴리오 선정 모형을 제안하였다. 본 연구는 2012년 9월부터 2023년 8월까지 총 11년의 투자기간에 대해 투자 실험을 진행하여 SA-M 포트폴리오 선정 모형의 유효성을 확인하였다. 시가총액을 기준으로 한국 주식시장에서 30개 종목을 선정하여 월간 주식수익률을 가지고 실험을 진행하였으며, 그 결과 SA-M 포트폴리오의 성과가 벤치마크 대비 우수함을 확인할 수 있었다. SA-M 포트폴리오의 누적수익률은 211.44%로, 벤치마크 대비(KOSPI 141.87%, SMM 91.74%, MVP 120.63%, B&H 114.81%, 1/N 150.60%) 뛰어난 성과를 보여주었다. 누적 수익률뿐만 아니라 샤프 비율, 소르티노 비율, 칼마비율, GPR 모두에서 벤치마크 대비 월등한 수치를 기록하여 수익률이 높을 뿐만 아니라 변동성이 낮고 안정적인 투자 전략임을 확인할 수 있었다. SMM, MVP 등 벤치마크로 선정된 기존 포트폴리오 선정 모형은 시장이 비정상적인 상황에서 빠르게 변화하는 시장을 반영하지 못하는 한계를 가지고 있다. 일례로 COVID-19 당시 시장이 급변했음에도 불구하고 계속 과거의 주가 데이터에 의지하여 최적화를 하다 보니 수익률 하락 방위에 취약했으며, 이는 낮은 투자성과로 이어졌다. 이와 달리, 본 연구가 제안한 SA-M 모형은 뉴스 헤드라인 감성 분석 결과를 통해 급격한 하락장이 발생했음을 감지하고 무위험자산 투자비중을 늘리고 위험자산 투자비중을 줄여 수익률을 성공적으로 방어할 수 있었다.

감성지수를 활용하여 위험자산 투자비중을 결정하는 SA-M이 효과적인 투자 전략임을 실증적으로 입증하였다는 점이 본 연구의 의의이다. 특히 벤치마크 포트폴리오로 SMM 포트폴리오, MVP 포트폴리오, 1/N 포트폴리오 등 많은 선행 연구에서 성과가 우수했던 포트폴리오를 선정했던 것을 고려할 때, SA-M 포트폴리오의 벤치마크 대비 성과가 더욱 유의미하다고 할 수 있다. 본 연구는 기존의 마코위츠 포트폴리오 선정 모형과 자연어처리 기법인 감성 분석을 결합하였다는 점에서 차별점을 갖는다. 해외 주식시장을 대상으로 전통적인 포트폴리오 모델에 감성 분석 결과를 적용하는 연구가 점차 진행되고 있으나[18, 28], 아직 한국 주식시장에서는 감성 분석을 통한 주가 예측 연구[1, 9] 정도가 주로 진행되고 있다. SA-M 포트폴리오는 감성 분석을 통해 도출한 감성지수가 한국 주식시장 상황을 예측하는 지표로 활용될 수 있으며, 마코위츠 포트폴리오의 수익성 및 안정성을 유의미하게 개선할 수 있음을 시사한다. 본 연구는 기본 마코위츠 모형에 뉴스 기사로부터 추출한 단일 시장 감성 점수를 결합하는 방법만으로도 투자 성과를 크게 개선할 수 있음을 보였다. 직관적이고 객관적인 모델만으로도 충분한 성과를 내며, 한국 주식시장에서 포트폴리오 모델과 감성분석의 결합이 가진 잠재력을 확인할 수 있었다. 이는 더 나아가 좀 더 복잡한 포트폴리오 전략으로도 확장될 가능성을 가지고 있기에, 본 연구는 후속 연구를 위한 기초적인 토대를 마련했다는 점에서 학문적 기여를 지닌다.

연구의 한계점은 다음과 같다. 본 연구는 코스피에 대한 뉴스 헤드라인의 감성 분석만 진행하여 시장을 포괄적으로 보여주는 단일 감성지수를 도출하였다. 그러나 코스피의 등락이나 시장 상황에 반응하는 민감도는 산업 또는 종목 특성에 따라 차이가 있을 수 있다. 따라서 산업별, 또는 종목별 감성 분석을 통해 복수의 감성지수를 가지고 투자비중을 보정한다면 유의미한 투자 성과 개선이 이루어질 것으로 기대한다. 추후 복수의 감성지수를 가지고 종목의 절대적 전망 및 종목 간 상대적 전망을 반영한 전망행렬을 구성하여 Black-Litterman model에 적용하고자 한다. 또한 본

연구는 뉴스 헤드라인만을 감성 분석의 대상으로 두었다는 점에서 분명한 한계를 가진다. 뉴스 본문이나 댓글 내용 등을 포함한다면 더 풍부한 정보 가치를 갖는 감성지수를 도출할 수 있을 것이다. 개선된 감성지수를 Black-Litterman 등 더욱 복잡한 형태의 포트폴리오에 결합하여 한국 주식시장을 대상으로 실험해보는 연구를 향후 이어갈 예정이다.

참 고 문 헌

- [1] 김대겸, 윤민혁, 조영진, 최용훈(2023), “뉴스의 감성 분석을 사용한 주가 예측 방법론,” 한국통신학회논문지, 48(6), 748-750.
- [2] 김성문, 김홍선(2009), “한국 주식시장에서 비선형계획법을 이용한 마코위츠의 포트폴리오 선정 모형의 투자 성과에 관한 연구,” 경영과학, 26(2), 19-35.
- [3] 김성진, 박준표, 김선미, 김세영, 김예진, 김물결, 성창환, 홍주환, 이용재(2023), “에널리스트 리포트의 NLP 분석을 통한 주식 가격 움직임 예측 및 자산 배분 전략 개발,” 대한산업공학회/한국경영과학회 2023년 춘계 공동학술대회, 제주, 4340-4349.
- [4] 김홍선, 정종빈, 김성문(2013), “한국 주식시장에서 마코위츠 포트폴리오 선정 모형의 입력 변수의 정확도에 따른 투자 성과 연구,” 한국경영과학회지, 38(4), 35-52.
- [5] 박경찬, 김홍선, 김성문(2016), “비중 상한 제약조건에 따른 포트폴리오 성과에 대한 투자 비중 분석,” 경영과학, 33(4), 51-64.
- [6] 박경찬, 정종빈, 김성문(2013), “지수가중이동평균법과 결합된 마코위츠 포트폴리오 선정 모형 기반 투자 프레임워크 개발: 글로벌 금융위기상황 하 한국 주식시장을 중심으로,” 한국경영과학회지, 38(2), 75-93.
- [7] 양지혜, 김홍선, 이순봉, 김성문(2021), “4분기 GDP 효과를 반영한 포트폴리오 선정 모형의 투자 성과 분석,” 한국경영과학회지, 46(2), 23-43.
- [8] 양지혜, 이순봉, 김성문(2021), “Baltic Dry Index와 Interest Rates를 반영한 포트폴리오 선정 모형의 설계와 투자성과 분석,” 경영과학, 38(4), 81-105.
- [9] 장주현, 김재윤(2023), “KR-FinBERT 뉴스 감성분석을 활용한 KOSPI 주가지수 예측,” 한국통신학회 동계종합학술발표회, 평창, 1142-1143.
- [10] 최재호, 정종빈, 김성문(2013), “마코위츠 포트폴리오 선정 모형을 기반으로 한 투자 알고리즘개발 및 성과평가: 미국 및 홍콩 주식시장을 중심으로,” 경영과학, 30(1), 73-89.
- [11] 한국언론진흥재단, 빅카인즈. (www.bigkinds.or.kr)
- [12] Araci, D.(2019), “FinBERT: Financial sentiment analysis with pre-trained language models”, arXiv preprint arXiv:1908.10063.
- [13] Behr, P., Guettler, A., and Miebs, F.(2013), “On portfolio optimization: Imposing the right constraints,” Journal of Banking & Finance, 37(4), 1232-1242.
- [14] Ben-Tal, A., Margalit, T., and Nemirovski, A.(2000), “Robust modeling of multi-stage portfolio problems,” High Performance Optimization(Frenkm, H., Roos, K., Terlaky, T. and Zhang, S.), Springer, New York, U.S., 303-328.
- [15] Black, F. and Litterman, R.(1992), “Global portfolio optimization,” Financial Analysts Journal, 48(5), 28-43.
- [16] Chopra, V.K. and Ziemba, W.T.(1993), “The effect of errors in means, variances, and covariances on optimal portfolio choice,” Journal of Portfolio Management, 19(2), 6-11.
- [17] Clarke, R., De Silva, H., and Thorley, S.(2006), “Minimum-variance portfolios in the US equity market,” Journal of portfolio management, 33(1), 10-24.
- [18] Colasanto, F., Grilli, L., Santoro, D., and

- Villani, G.(2022), "BERT's sentiment score for portfolio optimization: a fine-tuned view in Black and Litterman model," *Neural Computing and Applications*, 34(20), 17507-17521.
- [19] DeMiguel, V. and Nogales, F.J.(2009), "Portfolio selection with robust estimation," *Operations Research*, 57(3), 560-577.
- [20] DeMiguel, V., Garlappi, L., Nogales, F.J., and Uppal, R.(2009), "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management Science*, 55(5), 798-812.
- [21] Devlin, J., Chang, M.W., Lee, K., and Toutanova, K.(2018), "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*.
- [22] Duchin, R., and Levy, H.(2009), "Markowitz versus the Talmudic portfolio diversification strategies," *The Journal of Portfolio Management*, 35(2), 71-74.
- [23] Fabozzi, F.J., Huang, D., and Zhou, G.(2010), "Robust portfolios: Contributions from operations research and finance," *Annals of Operations Research*, 176(1), 191-220.
- [24] Fung, G.P.C., Yu, J.X., and Lam, W.(2003), "Stock prediction: Integrating text mining approach using real-time news," 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings., Hong Kong, 395-402.
- [25] Ghaoui, L.E., Oks, M., and Oustry, F.(2003), "Worst-case value-at-risk and robust portfolio optimization: A conic programming approach," *Operations research*, 51(4), 543-556.
- [26] Goldfarb, D. and Iyengar, G.(2003), "Robust portfolio selection problems," *Mathematics of operations research*, 28(1), 1-38.
- [27] Haugen, R.A. and Baker, N.L.(1991), "The efficient market inefficiency of capitalization-weighted stock portfolios," *The Journal of portfolio management*, 17(3), 35-40.
- [28] Hung, M., Hsia, P., Kuang, X., and Lin, S.(2024), "Intelligent portfolio construction via news sentiment analysis", *International Review of Economics & Finance*, 89(A), 605-617.
- [29] Jagannathan, R. and Ma, T.(2003), "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *The Journal of Finance*, 58(4), 1651-1683.
- [30] Jobson, J.D. and Korkie, B.(1980), "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Association*, 75(371), 544-554.
- [31] Jorion, P.(1985), "International portfolio diversification with estimation risk," *Journal of Business*, 259-278.
- [32] Jung, J. and Kim, S.(2015), "An adaptively managed dynamic portfolio selection model using a time-varying investment target according to the market forecast," *Journal of the Operational Research Society*, 66(7), 1115-1131.
- [33] Kalyani, J., Bharathi, P., and Jyothi, P.(2016), "Stock trend prediction using news sentiment analysis," *arXiv preprint arXiv:1607.01958*.
- [34] Kaya, M.Y., and Karsligil, M.E.(2010), "Stock price prediction using financial news articles," 2010 2nd IEEE International Conference on Information and Financial Engineering, 2010. Proceedings., Chongqing, 478-482.
- [35] Kim, E. and Shin, H.(2022), "KR-FinBert: Fine-tuning KR-FinBert for Sentiment Analysis", GitHub repository. (<https://huggingface.co/snunlp/KR-FinBert-SC>).
- [36] Levy, H., and Duchin, R.(2010), "Markowitz's

- mean-variance rule and the Talmudic diversification recommendation,” In Handbook of portfolio construction(Guerard, J.B.), Springer, Boston, U.S., 97-123.
- [37] Li, X., Xie, H., Chen, L., Wang, J., and Deng, X.(2014), “News impact on stock price return via sentiment analysis,” Knowledge-Based Systems, 69, 14-23.
- [38] Loughran, T., and McDonald, B.(2011), “When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” Journal of Finance, 66(1), 35-65.
- [39] Markowitz, H.(1952), “Portfolio Selection,” Journal of Finance, 7(1), 77-91.
- [40] Michaud, R.O.(1989), “The Markowitz Optimization Enigma : Is ‘Optimized’ Optimal?”, Financial Analysts Journal, 45(1), 31-42.
- [41] Mo, S.Y.K., Liu, A., and Yang, S.Y.(2016), “News sentiment to market impact and its feedback effect,” Environmental Systems and Decisions, 36(2), 158-166.
- [42] Nemes, L., and Kiss, A.(2021), “Prediction of stock values changes using sentiment analysis of stock news headlines,” Journal of Information and Telecommunication, 5(3), 375-394.
- [43] Pastor, L., and Stambaugh, R.F.(2000), “Comparing asset pricing models: An investment perspective,” Journal of Financial Economics, 56(3), 335-381.
- [44] Yang, Y., Uy, M.C.S., and Huang, A.(2020), “FinBERT: A pretrained language model for financial communications”, arXiv preprint arXiv:2006.08097.
- [45] Yuanyuan, Z., Xiang, L., and Sini. G.(2018), “Portfolio selection problems with Markowitz’s mean-variance framework: a review of literature,” Fuzzy Optimization and Decision Making, 17(2), 125-158.
- [46] Zellner, A. and Chetty, V.K.(1965), “Prediction and decision problems in regression models from the Bayesian point of view,” Journal of the American Statistical Association, 60(310), 608-616.