

# 감성 분석이 결합된 MVO 전략 : 심층 분석과 적용 가능성

감다살 팀

김민경

(경상대학 경제통상학부)

김강민

(경상대학 경제통상학부)

장혜영

(인문대학 고고인류학과)

최원준

(IT대학 전자공학부)

## < A B S T R A C T >~

금융 시장은 언제나 불확실성을 내포하고 있다. 오랜 기간 동안 투자자들은 불확실성 사이에서 최적의 포트폴리오 구성을 위해 다양한 전략을 개발해 왔다. 그중에서도 Markowitz의 평균-분산 최적화(MVO) 이론은 자산배분의 핵심적인 방법론으로 널리 사용되어 왔다. 그러나 MVO는 과거 수익률 데이터에 대한 의존성과 코너 솔루션 문제 등 몇 가지 중요한 한계점을 지닌다. 이러한 한계를 극복하기 위해 본 연구는 감성 분석(Sentiment Analysis)을 MVO에 결합한 새로운 자산배분 전략을 제안한다.

본 연구에서는 한국어 금융 도메인에 특화된 KR-FinBERT 모델을 활용하여 금융 뉴스로부터 시장 심리를 추출하였다. 산출된 감성지수가 단순 노이즈가 아닌 코스피 지수에 유의미한 영향을 미친다는 것을 확인하기 위해 상관관계 분석, 그레인저 인과관계 검정, 충격 반응 함수 등을 적용하였다. Savitzky-Golay 필터를 사용해 평활화된 감성지수는 육안으로 관찰하기에 코스피 지수를 유의미하게 선행하는 경향을 보였다. 또한, 그레인저 인과관계 검정에서는 감성지수가 코스피 지수를 유의미하게 예측함을 통계적으로 확인하였다. 이를 바탕으로 충격 반응 함수를 통해 감성지수의 충격이 코스피 지수에 미치는 효과를 분석하였고, 해당값으로 감성지수의 시차별 가중치를 설정하여 MVO 모델의 자산배분 비율에 반영하였다.

실험은 2016년 10월 부터의 데이터를 활용하여 2017년 10월부터 2024년 4월까지 기간으로 KOSPI200 각 섹터 ETF를 활용하여 백테스트를 진행하였다. 감성지수를 적용한 MVO 모델은 전통적인 MVO모델과 선행연구에서 제시되어진 감성지수를 활용한 블랙-리터만 모델보다 수익률, 샤프비율, MDD, 정보비율 등 여러 성과지표에서 우수한 결과를 보였다. 추가적으로, 해당 방법을 사용하여 코스피200 섹터 ETF를 추정하는데 가장 유의미한 데이터는 코스피 태그 데이터임을 확인하였다.

본 연구는 감성 분석을 통해 시장 심리의 동적 변화를 효과적으로 파악하고, 이를 자산배분 전략에 반영함으로써 전통적인 MVO 모델의 한계를 극복할 수 있음을 시사한다. 향후 연구에서는 다양한 뉴스 소스와 소셜 미디어 데이터를 활용하여 감성지수의 정확성을 높이고, 다른 시장이나 자산군에 대한 적용 가능성을 탐색함으로써 본 연구의 범용성을 검증할 수 있을 것이다.

## 1. Introduction

금융 시장은 다양한 요인에 의해 변동성이 커지며, 투자자들은 이러한 변동성 속에서 최적의 포트폴리오를 구성하기 위해 노력한다. Markowitz가 제안한 MVO(Mean-Variance Optimization)이론은 포트폴리오 관리에서 가장 중요한 자산배분 전략 중 하나이다 [1]. MVO는 자산배분 이론의 기본 개념 중 하나로, 기대수익률과 공분산을 고려하여 주어진 목표 수익률 하에서 위험을 최소화하는 포트폴리오 구성을 목표로 한다. 이론적으로 MVO는 완벽한 자산배분 전략 중 하나라고 볼 수 있다. 그러나 MVO는 여러 장점에도 불구하고 몇 가지 주요한 한계점을 가지고 있다. 특히, 과거의 수익률 데이터를 기반으로 한 MVO 모델은 미래의 시장 수익률 및 변동성을 정확히 반영하지 못할 수 있으며, 이는 최적화된 포트폴리오의 성능을 저해할 수 있다.

전통적인 MVO 모델은 자산 간의 상관관계를 충분히 반영하지 못하고, 특정 자산에 과도한 비중을 할당하는 '코너 솔루션' 문제를 발생시키기도 한다 [2]. 이는 분산 투자의 원칙에 반하며, 포트폴리오의 위험이 특정 자산에 집중되어 불균형이 발생할 수 있다. 또한, MVO는 시장의 동적 변화를 반영하지 못하는 정적 모델로, 실시간으로 변화하는 투자자 감정이나 시장 심리를 반영하는 데 한계가 있다.

이러한 한계를 극복하기 위해 본 연구는 감정 분석(Sentiment Analysis)을 MVO에 결합하는 방안을 제안한다. 감정 분석은 자연어 처리(NLP) 기법을 사용하여 금융 뉴스나 기타 텍스트 데이터에서 투자자 감정과 시장 심리를 추출하고 이를 정량화하는 방법이다. 본 연구에서는 한국어 금융 도메인에 특화된 LLM 모델인 KR-FinBERT [3]를 활용하여 감정 지수를 산출하고, 이를 MVO 모델에 반영하여 보다 동적이고 유연한 포트폴리오 최적화를 시도하였다.

감정 지수를 기반으로 MVO의 위험자산과 무위험 자산 사이의 자산배분 비율을 조정함으로써, 전통적인 MVO 모델의 한계인 코너 솔루션 문제를 완화하고, 시장의 변동성에 대응하는 포트폴리오를 구성하였다. 실험 결과, 감정 분석을 결합한 MVO 모델은 전통적인 MVO 모델에 비해 샤프 비율(Sharpe Ratio)과 최대 손실(MDD) 등 다양한 측면에서 더 우수한 성과를 보였다. 특히, 한국 경제 관련 뉴스 데이터에서 추출한 감정 지수를 활용한 모델이 가장 높은 성능을 나타내었으며, 이는 감정 분석이 포트폴리오 최적화에 효과적으로 활용될 수 있음을 시사한다.

## 2. Related Work

MVO가 가진 실무적인 한계를 극복하고자 다양한 연구들이 진행되어 왔다. 이중 가장 널리 알려진 연구는 Black, F과 Litterman의 블랙-리터만 모형이다 [4]. 블랙-리터만 모형은 투자자의 주관적 견해와 시장의 균형 상태를 결합하여 기대 수익률을 도출함으로써 MVO의 한계인 코너해 문제와 불안정한 자산배분 결과를 개선했다. 또한 Gerber Statistic(Gerber et al, 2022)에서는 MVO의 보완 방법으로 Gerber statistic을 제안하였다 [5]. Gerber Statistic은 두 자산의 수익률이 동시에 어떤 threshold를 기준으로 같은 방향, 다른 방향, 또는 그 사이에 있는지를 쌍으로 구분하여 자산 간의 동조성을 파악하는 통계량으로, 보다 강건한 공분산 행렬을 만들어 MVO를 보완하고자 했다. Robust Portfolio Optimization(Qiu et al, 2015)는 MVO의 한계인 기대수익률의 불확실성을 극복하기 위해 기대수익률을 특정 숫자로 정해놓은 것이 아닌 불확실성 집합(uncertainty set)으로 정의해 최악의 상황을 가정하고 최적 자산배분을 실시해 MVO를 보완하고자 하였다 [6]. 이렇듯 MVO의 한계를 극복하고자 현대까지 다양한 연구가 진행 중에 있다.

그러나, 인터넷이 널리 보급됨에 따라 투자자들은 인터넷 뉴스 등과 같은 다양한 매체를 통해 많은 양의 정보를 접하고 투자하기 시작했다. 이러한 이유로 뉴스 등의 다양한 매체에서부터 추출된 감성을 활용하여 투자기법을 개선하고자 하는 연구 또한 다양하게 존재했다. Hung, M.-C(2024)는 BERT 모델을 사용하여 뉴스 감성을 측정하고 GRU 모델을 사용하여 주가를 예측한 전략이 기타 자산배분 전략들보다 우수한 성능을 보임을 입증하였다 [7]. Colasanto (2022) 또한 금융신문에서 얻어낸 평균 감정 점수를 사용하여 주가

를 예측하고 동적으로 블랙-리터만 모델을 개선하고자 시도하였다 [8]. 이러한 노력은 국내에서도 지속적으로 관찰해 볼 수 있다. 김동재 (2023)의 경우 블랙 리터만 모형에서 투자자의 주관적인 전망을 뉴스 기사에 포함된 감성 점수로 대체하여 블랙리터만 모형을 개선하고자 시도하였다 [9]. 또한, 강두원 (2022)는 SNS 데이터 내 감성을 활용하여 소비자들의 기업 인식 데이터로 주가 예측 모델을 구성하였다 [10]. 그러나, 블랙리터만 모델을 사용하기 위해서는 개별 자산 혹은 자산군들에 해당하는 감성지수가 필요하다. 개별 자산에 해당하는 감성지수를 도출하기에는 데이터가 부족하거나, 특정 전문용어에 치우쳐져 감성지수를 정확하게 반영하지 못하게 되는 단점이 존재한다. 이러한 점을 개선하고자 본 논문에서는 코스피 지수 자체의 감성 점수를 활용해 무위험 자산과 위험자산 사이의 자산배분 비율을 조정하고자 한다.

### 3. Methodology

#### 3.1 MVO(평균-분산 최적화)

본 연구에서는 2011년 1월부터 2024년 4월까지의 KOSPI 200 건설, 철강, 에너지, 정보기술, 금융, 경기소비재, 헬스케어 섹터 ETF의 종가지수를 활용하였다. 해당 지수들은 시장환경에 따라 주가지수가 각각 다르게 변동하고, 코스피 지수와 유사하게 움직이기 때문에, MVO의 자산 배분에 적합하다고 판단하였다. 각 자산의 월별 수익률 데이터를 기반으로 Markowitz의 평균-분산 최적화 이론을 적용하여, 포트폴리오의 기대수익률을 최대화하고 리스크를 최소화하는 최적의 자산 배분을 도출하였다. MVO 모델은 다음의 수학적 최적화 문제를 풀어 자산 배분을 결정하였다.

$$\min_w \left( \frac{1}{2} w^\top \sum w - \lambda \mu^\top w \right)$$

[수식 1] MVO 모델

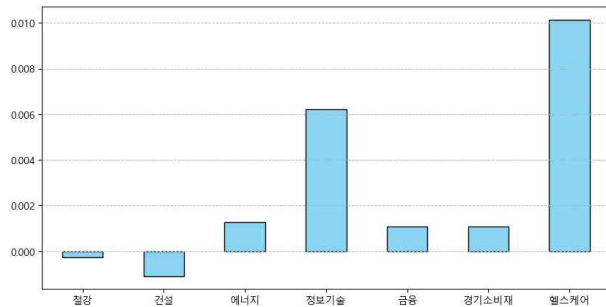
$w$ : 자산 배분 벡터

$\sum$ : 자산 간의 공분산 행렬

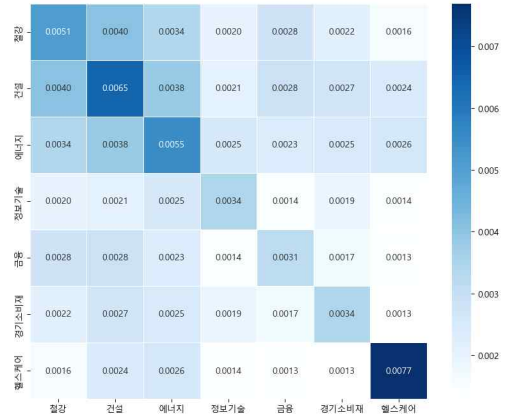
$\mu$ : 기대 수익률 벡터

$\lambda$ : 투자자의 리스크 허용도

해당 MVO 과정에서 기대수익률의 경우 데이터 수집 기간 동안의 평균 수익률을 사용하였다. 그 결과 [그림 1]처럼 헬스케어와 건설이 변동성이 크게 나타났으며, 정보기술과 헬스케어의 수익률이 가장 좋게 나타났다.



[그림 1] ETF별 기대 수익률



[그림 2] ETF별 공분산 행렬

위 데이터를 사용하여 무위험 자산 수익률을 3.5%로 잡고 앞서 설명한 공식대로 mvo weight를 계산한 결과 다음과 같은 weight가 추출되었다.

[2.82105022e-09, 7.35392569e-09, 5.44350608e-09, 7.86609304e-01, 9.20433771e-10, 1.07676183e-09, 2.13390673e-01, 5.64405253e-09]

무위험 자산에 자산배분이 극히 일부만 할당되는 코너해 문제가 발생한 것을 알 수 있다. 또한, 한달간의 리밸런싱 기간을 잡고 2018년부터 2024년까지 MVO 전략을 사용하였을 때 6년간 총 수익률은 약 -35%로 좋지 않게 나타난 것을 확인할 수 있다. 우리는 이를 개선하고자 감성 지수를 사용할 예정이다.

### 3.2 감성 데이터 수집 및 전처리

뉴스 데이터는 코스피에 관한 시장의 반응을 적나라하게 보여준다. 시장의 감성이 부정적일 경우 “추락하는 주식에 날개가 없다. 삼성전자, 연일 신저가 찍어” 등과 같은 뉴스가 주를 이루고 있으며, 시장의 감성이 긍정적일 경우 “기관, LG에너지솔루션 6일 연속 매수 행렬! AI도 상승 전망!” 등 과 같은 뉴스가 주를 이루는 모습을 보여준다. 이러한 점을 고려했을 때, 뉴스 데이터는 단순한 정보 제공을 넘어 시장의 전반적인 분위기와 감성을 파악하는 데 중요한 역할을 한다고 할 수 있다. 이에, 뉴스 데이터를 사용하여 시장의 감성을 파악하는 것이 적절하다고 판단하였다. 다만, 뉴스매체는 객관적인 사실만을 전달하고자 하지만, 뉴스 매체별로 시장의 반응을 다르게 해석하고 뉴스를 내는 경향이 있다. 따라서, 뉴스 매체를 세 가지로 나누어 데이터들을 수집하였다. 한국에서 코스피에 관한 광범위한 뉴스를 다루고 있는 사이트는 대표적으로 한국경제 뉴스와 매일 경제 뉴스 두 가지 종류가 있으며, 외국 뉴스를 다루고 있는 데이터는 대표적으로 블룸버그에서 제공하는 해외 매크로 뉴스 데이터가 존재한다. 따라서 세 종류의 뉴스매체를 대상으로 크롤링을 사용하여 데이터를 수집하였다.

[표 1] 매일경제 사이트에서 제공하는 코스피 검색 데이터 [11]

예시 문장
<p>[외국인, 실적부진 대형주로 몰린다...왜?]</p> <p>최근 외국계 자금이 국내 증시에 든든한 버팀목으로 작용하고 있다. 코스피 지수가 2000선 돌파를 시도하고 있는 가운데 외국인은 지난 10일부터 11거래일 연속 순매수를 보이면서 지수를 끌어올리고 있기 때문이다. 다만 매수세는 SK하이닉스, POSCO 등 1분기 실적 전망치가 하향 조정되고 있는 대형 종목에 집중적으로 쏠리고 있어 외국인의 이같은 행보에 관심이 모이고 있다. ...</p>

[표 2] 한국경제 사이트에서 제공하는 코스피 태그 데이터 [12]

예시 문장
<p>[코스피 상장사 현금배당 10년 만에 15조→41조]</p> <p>유가증권시장 상장사의 현금배당 규모가 최근 10년간 세 배 가까이 늘어난 것으로 나타났다. 순이익 대비 배당금 비중을 나타내는 배당성향도 가파르게 상승해 미국과 일본 등을 앞섰다. 20일 한국상장회사협의회에 따르면 유가증권시장 상장사의 지난해 합산 현금배당액은 41조2000억원을 기록했다. 현금배당은 2014년만 해도 15조5000억원에 그쳤다. 하지만 이후 꾸준히 증가해 2020년 40조원대에 올라섰고 최근 2~3년간 41조원을 넘나들고 있다. ...</p>

[표 3] 블룸버그 사이트에서 제공하는 뉴스 데이터 [13]

예시 문장
<p>[오늘의 5가지 이슈: 美고용둔화, 핼코 채권 낙관]</p> <p>미국 고용시장이 식고 있다는 추가적 신호에 미국채 금리는 하락세를 이어가 10년물 금리가 장중 한때 8bp 가까이 후퇴해 5월 중순래 최저치인 4.31%까지 밀렸다. 4거래일 동안 약 30bp 하락한 셈이다. 스왑시장은 확실한 연준의 첫 25bp 인하 예상 시기를 12월에서 11월로 앞당겼고, 연내 인하 기대치를 44bp로 높였다.BMO Capital Markets의 Ian Lyngen는 이번 지표 약세가 9월 인하 근거를 도울 수 있다고 진단했다. 국제유가 하락 역시 디스인플레이션 진전을 뒷받침하는 분위기다. ...</p>

[표 1]은 한국 경제 사이트에서 제공하는 코스피 태그 데이터이다. 주로 코스피와 직접적인 관련이 있는 데이터 ('코스피, 외국인 매도에 약세', '코스피 2800 아래서 출발' '자본시장법 개정안 발의' 등)를 다루고 있다. 기사는 사이트에서 제공하고 있는 2016.07.08. ~ 2024.06.31. 까지 총 7,583개의 기사를 기반으로 추출하였다.

[표 2]는 매일경제 사이트에서 제공하는 코스피 검색 데이터이다. 이는 본문이나 기사 제목에 코스피가 들어가는 모든 종류의 데이터('우크라 살상무기 지원 검토', '밸류업 업은 중간배당', '원금 보장 ELB' 등)를 다루고 있으며, 한국 경제에서 제공하는 코스피 태그 데이터 보다 많은 데이터를 다루고 있다. 그러나, 불필요한 기사도 포함되어 있다는 단점이 있다. 기사는 2016.03.22. ~ 2024.06.05. 까지 총 25,577개의 기사를 기반으로 추출하였다.

[표 3]은 블룸버그 서비스에서 제공하는 한국어 뉴스 사이트에서 받아온 기사 데이터이다. 주로 해외 매크로 지표(미국 PCE, FOMC 금리결정 등)뉴스를 다루고 있으며, 하루에 총 5 개의 주요 뉴스가 생성된다. 기사는 사이트에서 제공하는 2020.11.17. ~ 2024.06.05 일간의 데이터를 수집하였으며, 총 4,045(809\*5)개의 기사를 기반으로 추출하였다.

수집된 뉴스 데이터는 감성 분석 모델에 적합하도록 전처리 과정을 거쳤다. 이 과정은 텍스트 내 불필요한 기호 및 HTML 태그 제거, 띄어쓰기 및 특수문자 수정 등을 포함하였다. 전처리된 데이터는 감성 점수를 부여하기 위한 입력으로 사용되었다. 전처리된 데이터는 KR-finBERT 모델을 통해 감성 점수를 부여받았다. 각 뉴스 기사에 대해 긍정(1), 중립(0), 부정(-1)의 세 가지 범주로 분류되었으며, 이는 이후 감성 분석에서 사용될 감성지수를 계산하는 데 활용되었다.

### 3.3 KR-finBERT 모델

KR-FinBERT는 BERT(Bidirectional Encoder Representations from Transformers) 모델을 한국어 금융 도메인에 맞춰 미세 조정한 모델이다 [3]. BERT는 문맥을 양방향으로 학습하는 트랜스포머 기반 모델로, 자연어 처리 분야에서 강력한 성능을 발휘한다. 본 연구에서는 이 모델을 사용하여 각 뉴스 기사의 감성 상태를 분석하고 점수를 부여하였다. 감성 분석은 각 뉴스 문장을 입력으로 받아 긍정(1), 중립(0), 부정(-1)의 세가지 상태로 분류하는 작업이다. 긍정으로 판단된 문장, 중립으로 판단된 문장, 부정으로 판단된 문장은 [표 4]와 같다.

[표 4] KR-Finbert로 판별한 긍정, 부정, 중립 문장 예시

긍정 문장 예시
인공지능(AI) 혁명의 핵심에 있는 반도체업체 엔비디아의 실적 발표를 하루 앞두고 뉴욕증시에서 S&P 500 지수가 또다시 사상최고치를 다시 썼다 올해 들어 벌써 24번째 신기록 경신이다 {'label': 'positive', 'score': 0.9994427561759949}
부정 문장 예시
또한 모든 중국산 제품에 최대 60%의 관세와 전 세계에서 미국으로 들어오는 모든 수입품에 10% 관세를 부과하겠다는 공약을 내놓았다 {'label': 'negative', 'score': 0.6739509701728821}
중립 문장 예시
스왑 트레이더들은 연초만해도 올해 연준이 기준금리를 25bp씩 6번 넘게 인하할 것으로 기대했으나, 이제는 총 33bp로 베팅을 낮춰 1차례 인하에 그칠 것으로 내다보고 있다 {'label': 'neutral', 'score': 0.9984210729598999}

부여된 감성 점수를 바탕으로 각 뉴스의 감성지수를 계산하였다. 긍정, 중립, 부정의 총합을 바탕으로 감성지수는 다음과 같은 공식으로 산출되었다.

$$Sentiment\ score = \frac{\sum Positive - \sum Negative}{\sum Positive + \sum Negative}$$

[수식 2] 감성지수

이 공식은 뉴스별 집계된 긍정과 부정 문장의 비율을 반영하며, 코스피 지수와의 상관관계를 분석하기 위한 중요한 지표로 사용되었다.

앞서 구한 sentimental data는 [수식 3]처럼 월별 감성지수로 변환 한 후 Min- Max 정규

화 과정을 12개월 별로 이동시켜 적용 후 사용되었다.

$$monthly\ sentiment\ score = \frac{\sum \left( \frac{\sum (daily\ senti\ score)}{vm\ daily\ news} \right)}{vm\ monthly\ news}$$

$$\min\_senti\_score = \min(monthly\ sentimental\ score12\ monthly\ sentimental\ score1)$$

$$\max\_senti\_score = \max(monthly\ sentimental\ score12\ monthly\ sentimental\ score1)$$

$$SentiIndex = 0 : if (senti\ score < \min\_senti\_score)$$

$$SentiIndex = 1 : if (senti\ score > \max\_senti\_score)$$

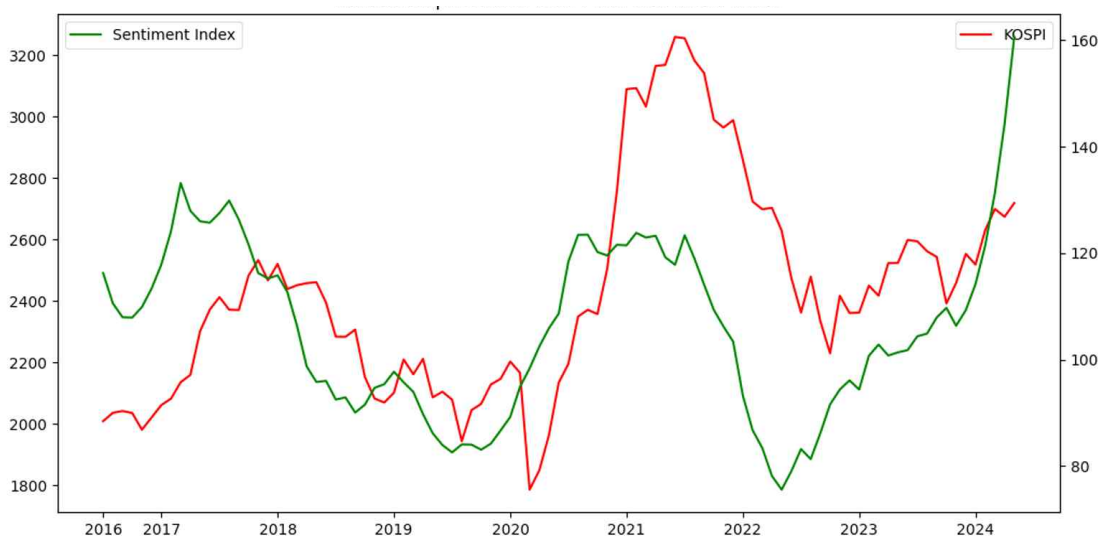
$$SentiIndex = (senti\ score - \min\ senti\ score) / (\max\ senti\ score - \min\ senti\ score)$$

[수식 3] 감성지수 정규화 과정

감성지수는 감성이 극도로 긍정적이었던 시기와 부정적이었던 시기 두 시기를 명확하게 구분 해야 할 필요가 있다. 위 과정을 통해 이상치의 변동성에 대한 영향을 최대한 살리면서, 다른 지표와의 비교 및 MVO 모델에서의 적용이 용이하게 만들었다.

### 3.4 Weight 산출

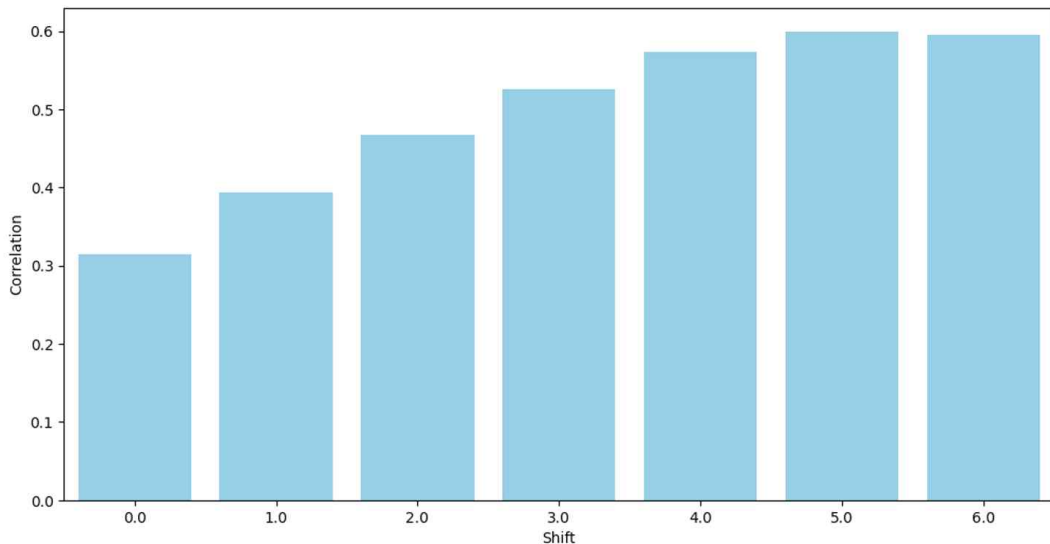
본 연구에서는 산출된 감성지수가 단순 노이즈가 아니며, 코스피 지수에 유의미한 영향을 끼친다는 것을 증명하기 위해 상관관계 분석과 인과관계 분석을 실시 하였다. 먼저, 감성지수 그 자체는 노이즈가 많아 육안으로 판별하기 쉽지 않다. 따라서 Savitzky-Golay 필터 [14]를 사용하여 감성지수를 평탄화하였다. Savitzky-Golay 필터는 이동 윈도우 내에서 저차 다항식을 데이터에 적합시켜 중앙값을 추정함으로써 작동한다. 그 결과, [그림 3]에서 관찰할 수 있듯이 감성지수가 코스피 종가를 선행하는 듯한 모습을 보여주었다.



[그림 3] 코스피 종가와 감성지수의 시계열 데이터



통계적으로 확인하기 위해 두 데이터의 상관관계 분석을 진행하였다. 상관계수를 산출한 결과 코스피 종가와 감성지수 Index의 상관계수는 0.31로 나타났으며, 감성지수를 0~6개월 이동하면서 상관계수를 산출한 결과 [그림 4]에서 관찰할 수 있듯이 감성지수를 5개월 이동하였을 때 상관계수가 0.6으로 가장 높은 결과를 보여주었고 그 이후 천천히 감소하는 모습을 보였다.



[그림 4] 코스피 종가와 감성지수 사이의 상관관계 분석

해당 결론을 통해 감성지수가 코스피 지수를 어느 정도 선행한다고 볼 수 있었다.

두 데이터의 정확한 인과관계를 검정해 보기 위해 그레인저 인과검정 방식을 사용하여 검증하였다 [15]. 그랜저 인과관계 검정은 한 시계열이 다른 시계열의 미래 값을 예측하는 데 유의미한 정보를 제공하는지를 판단하는 통계적 방법이다. 그레인저 인과관계 검정을 수행하기 전에, 시계열 데이터의 정상성(stationarity)을 확인하는 것이 필수적이다. 따라서 먼저, ADF 검정 [16]을 통해 두 시계열 데이터의 정상성을 검증해 본 결과 코스피 종가의 경우 테스트 통계량이  $-3.107$ 로 95% 임계값인  $-2.890$ 보다 작았으며, p-value 값이  $0.026$ 으로 5% 유의수준에서 정상 시계열임을 확인하였다. 감성지수의 경우에서도 마찬가지로 테스트 통계량이  $-3.531$ 로 95% 임계값  $-2.897$ 보다 작았으며, p-value  $0.007215$ 로 두 시계열 모두 정상성을 가짐을 확인하였다.

벤치마크로는 일반적으로 주가를 선행한다고 알려져있는 경기선행지수, M2, BTC 등의 데이터를 사용하였다. 검정 결과, [표 5]에서 확인할 수 있듯이 감성지수가 기타 벤치마크에 비해 더욱 유의미하게 코스피 지수를 선행함을 알 수 있었다.

[표 5] 그레인저 인과검정 결과

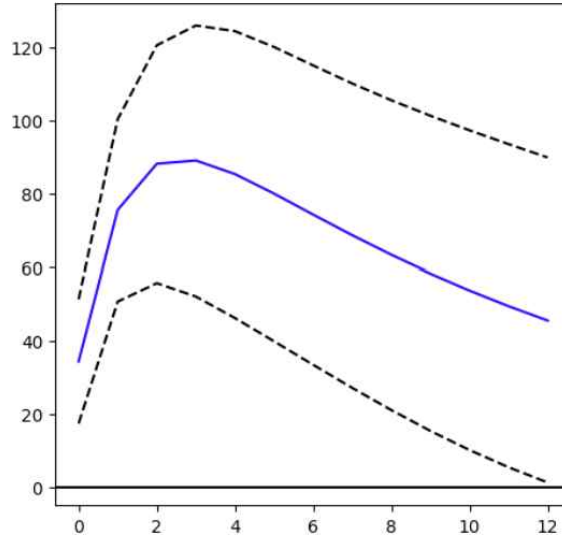
시차	경기선행지수	M2	BTC	sent-index	코스피->sent
1	0.5699	0.7070	0.6506	0*	0.8534
2	0.0383*	0.8049	0.3470	0*	0.3190
3	0.0678	0.4673	0.5562	0*	0.3600
4	0.2748	0.1211	0.5762	0.0001*	0.1832
5	0.3913	0.0204*	0.4863	0.0001*	0.0781
6	0.0609	0.0423*	0.5569	0*	0.0214*

이러한 결과는 감성지수가 코스피 지수의 선행 지표로서 유의미한 역할을 할 수 있음을 시사한다. 따라서 감성지수를 활용한 자산배분 전략이 포트폴리오의 성과를 향상시키는데 기여할 수 있음을 통계적으로 뒷받침한다.

인과관계분석 결과를 토대로 시장의 감성지수가 코스피 지수에 미치는 충격반응함수를 구현하였다 [17]. 충격반응함수는 감성지수에 한 단위의 충격이 발생하였을 때 코스피 지수가 시간에 따라 어떻게 반응하는지 보여준다. VAR 실험 결과 [표 5]에서 확인 가능하듯이 베이즈 정보기준(BIC)과 아카이케 정보기준(AIC) 모두 시차 1이 충격반응함수를 설명하는데 가장 적합한 것으로 나타났다. 따라서, 시차1을 기준으로 충격반응 함수를 살펴 본 결과 [그림 3]에서 보여지듯이 시차에 따른 코스피 지수의 반응은 34.39, 75.51, 88.09, 88.98, 85.27, 79.93, 74.21로 시차 3에서 정점을 찍고 그 이후 서서히 감소하는 모습을 나타내었다. 이는 시차 3에서 충격이 최대에 달한 후 감성지수가 코스피에 미치는 영향이 점차 감쇠함을 확인할 수 있다.

[표 6] 각 시차별 충격반응함수 결과

시차	AIC	BIC	FPE	HQIC
0	17.73	17.79	5.010e+07	17.75
1	14.76*	14.93*	2.568e+06*	14.83*
2	14.80	15.08	2.682e+06	14.92
3	14.78	15.08	2.637e+06	14.94
4	14.83	15.33	2.756e+06	15.03
5	14.83	15.45	2.755e+06	15.08
6	14.81	15.54	2.713e+06	15.10



[그림 5] 충격반응함수 결과

따라서, 우리는 해당 충격량을 기준으로 각 기간별 감성지수가 시차에 따라 자산배분에 미치는 영향을 Weight로 고려하였다. 감성지수는 사후적으로 산출될 수 밖에 없으므로 기간 0의 충격은 제외하였다. 최대 충격량을 가지는 시차 3 까지 해당하는 각 충격량은 [75.51, 88.09, 88.98]로 도출되었고 이를 통해 감성지수가 시차 3 까지 주가에 미치는 영향이 유사하다는 가정하에 [0.3, 0.35, 0.35]로 Weight를 설정하였다.

## 4. Experiments

### 4.1 실험 방식

해당 가중치를 활용하여 무위험 자산에 대한 투자 비중을 조정하였다. 이전 1개월, 2개월, 3개월 전의 감성지수에 각각 30%, 35%, 35%의 가중치를 부여하여 종합 감성지수를 계산하였다. 이 종합 감성지수를 기반으로, 무위험 자산에 배분할 비중을 다음과 같이 결정하였다.

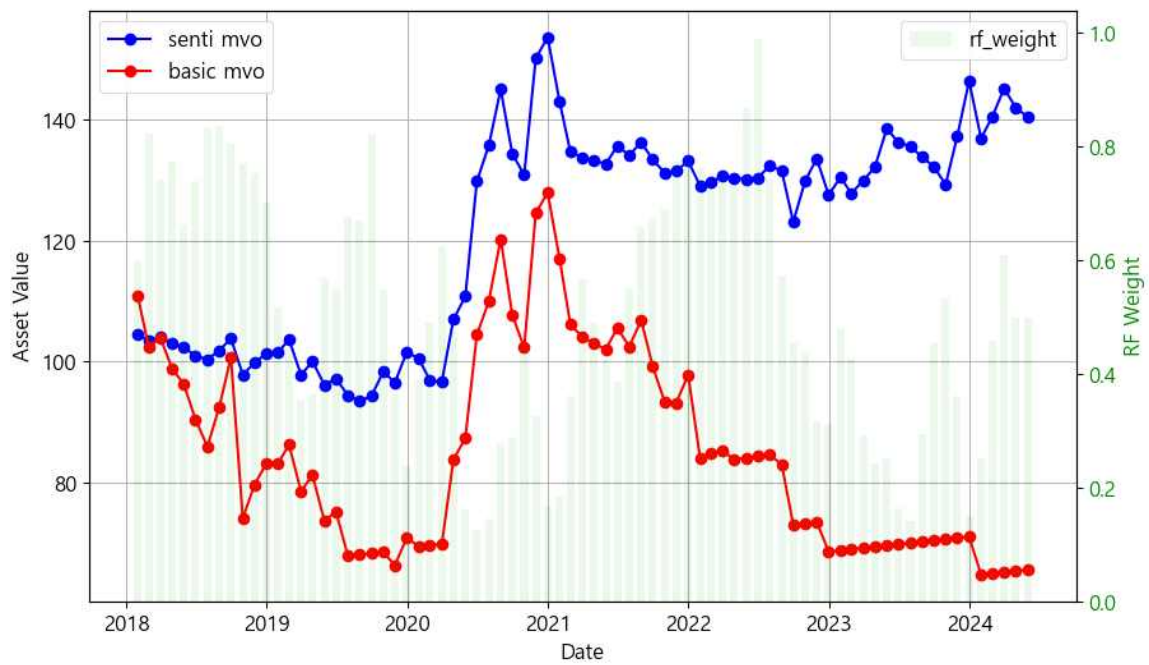
$$Risk-free Asset Weight = 1 - (Senti_{t-1} \times 0.3 + Senti_{t-2} \times 0.35 + Senti_{t-3} \times 0.35)$$

[수식 4] 무위험 자산 배분 비중

포트폴리오는 매월 감성지수에 따라 리밸런싱을 수행하였다. 감성지수가 높을 때는 위험 자산에 대한 비중을 높이고, 감성지수가 낮을 때는 무위험 자산에 대한 비중을 높이는 방식으로 포트폴리오 배분을 조정하였다. 해당 방식은 감성지수가 급격히 부정적일 때, 무위험 자산에 투자하는 비중이 높아지고, 감성지수가 긍정적일 때, 위험 자산에 투자하는 비중을 높이기 위해 고안되었다. 이는 자산배분 방식을 활용해 감성지수의 잦은 리밸런싱이라는 단점을 극복하였으며, 동시에 높은 주가 예측력이라는 장점을 그대로 가져올 수 있는 방식이다.

실험 결과 [그림 6]처럼 시장의 감성을 반영한 MVO가 기본 MVO 전략대비 수익률 측면에서 6년간 51% 더 우수한 성능을 보여주었다. 또한 포트폴리오의 성능을

평가하기 위해 샤프 비율(Sharpe ratio), 최대 낙폭(MDD, Maximum Drawdown), 소르티노 비율(Sortino ratio), 정보 비율(Information ratio)등의 지표를 사용하여 비교하였다 [18] [19] [20]. 이러한 지표들은 포트폴리오의 성과를 평가하기 위해 널리 사용되는 지표들이며, 투자 성과를 다양한 측면에서 비교가능하게 도와준다. 그 결과, [표 6]에서 관찰할 수 있듯이 샤프비율과 MDD 그리고 소르티노 비율에서도 각각 0.4343, -0.1985, 0.6752로 기존 MVO의 -0.1440, -0.4932, -0.192 대비 큰 개선점을 보였다. 3년간 자세한 감성지수와 수익률, 그리고 무위험 자산 비율은 Appendix A에서 확인 가능하다. 또한 각 섹터별 수익률과 비교해 보았을 때 또한 [그림 7]에서 나타난 것처럼 수익률의 경우 건설을 제외한 모든 섹터에서 6년간의 수익률이 높았다. [표 6]에서 관찰할 수 있듯이 샤프비율과 MDD 그리고 소르티노 비율 측면에서는 감성지수를 활용한 MVO가 가장 높은 수치를 기록하였다.



[그림 6] 감성지수를 활용한 MVO 전략과 기본 MVO 전략



[그림 7] KOSPI 200 ETF 지수별 수익률과 감성지수를 활용한 MVO 전략

[표 6] KOSPI 200 ETF와 MVO 전략의 성과 분석

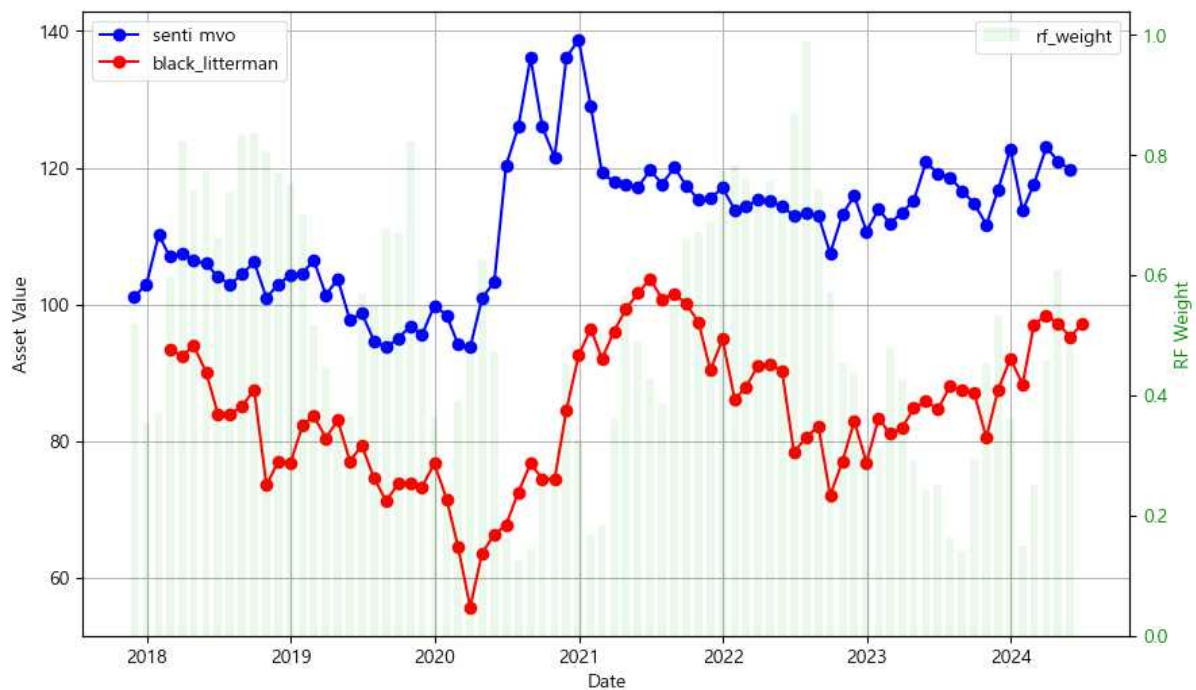
	샤프비율	MDD	소르티노비율	정보비율	수익률(6년)
MVO	-0.1440	-0.4932	-0.1932	nan	-35%
감성분석 MVO	0.4343	-0.1985	0.8635	0.6752	40%
철강	-0.0925	-0.4996	-0.1233	-0.0925	-27%
건설	0.2600	-0.4728	0.3398	0.2600	41%
에너지	0.0914	-0.3341	0.1306	0.0914	4.3%
정보비율	0.1069	-0.3016	0.1274	0.1069	1.8%
금융	0.0832	-0.4279	0.1019	0.0832	5.6%
경기소비재	-0.0951	-0.3618	-0.1077	-0.0951	-20%
헬스케어	0.0640	-0.3997	0.0860	0.0640	1.17%

이러한 결과를 통해 감성지수를 활용한 MVO가 전통적인 MVO방식 보다 뛰어난 성능을 보임을 확인하였다. 또한, 특정 기간 동안 MVO가 주식에 투자하지 않고, 무위험 자산에 모든 자산을 투자하는 코너 솔루션 문제가 발생한 것을 확인할 수 있는데, 이와 같은 문제점이 감성지수를 활용한 mvo에서는 나타나지 않아 코너 솔루션 문제를 일부 완화한 모습 또한 관찰할 수 있다.

#### 4.1.1 자산배분 알고리즘과 비교

감성지수를 활용한 전략으로 최근 대표되는 연구에서는 Black-Litterman 이론을 사용했다. 따라서 본 연구에서도 같은 데이터를 사용하여 감성지수를 활용한 블랙-리터만 전략과 수익률, 샤프비율, MDD, 소르티노비율, 정보비율면에서 비교를 수행하였다. 블랙리터만 전략은 각 자산별 투자자의 심리를 반영할 수 있는 전략이다. 따라서, 한국경제에서 제공하는 각 ETF 별 태그 데이터를 수집하여 이를 투자자 심리로 활용하였다. 또한, 최대한 유사한 환경으로 실험 하기위해 모델에 사용되는 파라미터 값은 널리 사용되는 값 혹은 MVO전략에서 사용한 값과 동일하게 실험하였다. 무위험 자산 수익률의 경우 3.5%로 가정하였으며 위험 회피계수는 2.5로 설정하였다. 시장 포트폴리오 가중치는 ETF를 사용한 관계로 모든 자산에 대해 동일 가중치를 사용하였으며  $\Omega$ 의 값은 0.05로,  $\tau$ 의 값은 0.02로 설정하였다.

위 방식을 사용해 백테스트를 진행한 결과 [표 7]에서 관찰할 수 있듯이 샤프비율은 0.13, MDD는 -0.41, 정보비율은 0.20, 소르티노 비율은 0.19로 전통적인 MVO방식 보다 높은 성능을 보였다. 그러나, 감성지수를 활용한 MVO 전략보다는 낮은 성능을 보였다. 이러한 결과는 각 자산군별 뉴스데이터가 금융도메인 뿐만 아니라 각 자산군에 특화된 용어가 많아 감성지수를 도출해 내는데 어려움을 겪었기 때문이라고 판단된다.



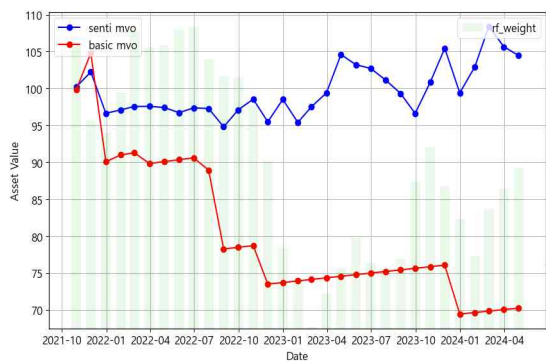
[그림 8] 감성지수를 활용한 블랙리터만 모델 전략

[표 7] KOSPI 200 ETF와 MVO 전략의 성과 분석

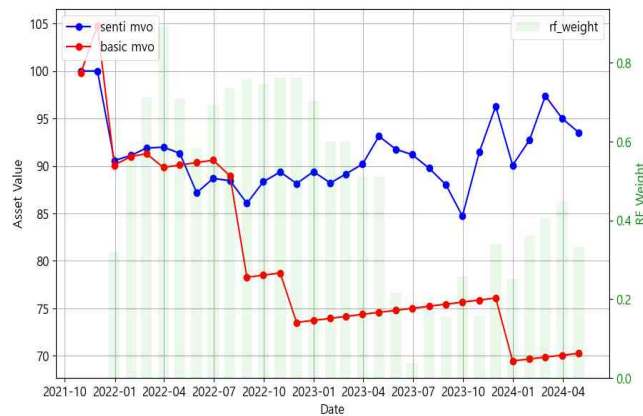
	샤프비율	MDD	소르티노비율	정보비율	수익률(6년)
MVO	-0.1440	-0.4932	-0.1932	nan	-35%
감성분석 MVO	0.4343	-0.1985	0.8635	0.6752	40%
블랙-리터만	0.1315	-0.4153	0.1922	0.2013	-4.73%

#### 4.2 감성 데이터 판별

앞서 데이터 설명에서 데이터의 종류를 3가지로 나누었다. 뉴스매체별로 데이터의 특성이 다른 만큼 어떤 종류의 데이터가 국내 자산 배분에 가장 적합한지 확인해볼 필요가 있다. [그림 8][그림 9][그림 10]에 각기 다른 데이터를 사용해 뽑아낸 감성지수를 사용한 MVO 전략과 전통적 MVO 전략과의 3년치 백테스트 결과를 확인 할 수 있다. 또한 실험 결과 [표 7]에서 나타나듯이, 한국경제 내 코스피 태그 데이터가 샤프 비율, 정보 비율, 소르티노 비율 등 다양한 측면에서 다른 뉴스 데이터 대비 가장 성능이 좋았음을 관찰할 수 있다. 이는 감성지수를 활용할 때, 노이즈가 섞이지 않은 코스피와 직접적인 영향이 있는 데이터 내부의 감성이 가장 중요하다는 것을 확인할 수 있다.



[그림 8] 매일경제 데이터를 활용한 MVO 전략 [그림 9] 한국경제 데이터를 활용한 MVO 전략



[그림 10] 블룸버그 데이터를 활용한 MVO 전략

[표 8] 실험 방식별 샤프비율, MDD, 소르티노비율, 정보비율 결과

	Senti(매일경제)	Senti(한국경제)	Senti(블룸버그)	감성분석 MVO	MVO
Sharpe Ratio	0.2216	0.3122	-0.1647	-0.2956	-0.9468
MDD	-0.0763	-0.0709	-0.1529	-0.0755	-0.3372
Sortino Ratio	0.3778	0.5241	-0.2289	0.4707	-0.7231
Information Ratio	1.3514	1.4403	0.9266	1.4586	nan

## 5. Conclusion

본 연구는 전통적인 평균-분산 최적화(MVO) 모델의 한계를 극복하기 위해 감성 분석을 결합한 새로운 포트폴리오 자산배분 전략을 제안하였다. 기존 MVO 모델은 과거 수익률 데이터에만 의존하여 미래의 시장 변동성을 정확히 반영하지 못하고, 코너 솔루션 문제로 인해 특정 자산에 과도한 비중을 할당하는 한계가 있었다. 이러한 문제를 해결하고자, 본 연구에서는 KR-FinBERT 모델을 활용하여 뉴스 데이터를 기반으로 감성



지수를 산출하였다. 구해진 감성점수를 통해 시장의 감성이 주가에 선행적인 영향을 미친다는 것을 통계적으로 검증하였다. 그레인저 인과검정과 충격반응함수를 통해 감성 지수가 코스피 지수에 유의미한 선행 지표임을 확인하였으며, 이를 자산배분 비율에 반영함으로써 포트폴리오의 성과를 향상 시키고자 하였다.

실험 결과, 감성 지수를 활용한 MVO 모델은 전통적인 MVO 모델과 감성지수를 활용한 블랙리터만 모델에 비해 수익률, 샤프 비율, 최대 손실(MDD) 등 다양한 성과 지표에서 우수한 결과를 보였다. 특히, 코스피 지수와 직접적으로 관련된 한국경제의 뉴스 데이터를 활용한 감성 지수가 가장 우수한 성과를 보여주었다.

본 연구의 의의는 감성 분석을 활용하여 전통적인 포트폴리오 최적화 모델의 한계를 보완하고, 시장의 감성 변화를 효과적으로 반영한 자산배분 전략을 제시하였다는 데 있다. 감성 지수를 통해 투자자들의 심리와 시장 분위기를 정량화하고 이를 투자 결정에 반영하였을 뿐만 아니라 선행연구인 블랙-리터만 모델과의 비교 실험을 통해 제안된 모델의 우수성을 입증하였다.

## 6. References

1. Markowitz, H. (1952). Portfolio selection. *\*The Journal of Finance\**, 7(1), 77-91.
2. Walters, J. (2014). The Black-Litterman Model in Detail. [blacklitterman.org](http://blacklitterman.org); Boston University, Metropolitan College, Department of Computer Science. (Original work published 2009, revised 2014).
3. Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063. <https://doi.org/10.48550/arXiv.1908.10063>
4. Black, F., & Litterman, R. (1991). Asset allocation combining investor views with market equilibrium. *\*The Journal of Fixed Income\**, 1(2), 7-18.
5. Gerber, S., Markowitz, H., Ernst, P., Miao, Y., & Sargen, P. (2021). The Gerber Statistic: A Robust Co-Movement Measure for Portfolio Optimization. SSRN. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3880054](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3880054)
6. Qiu, H., Han, F., Liu, H., & Caffo, B. (2015). Robust portfolio optimization. *Advances in Neural Information Processing Systems*, 28.
7. Hung, M.-C., Hsia, P.-H., Kuang, X.-J., & Lin, S.-K. (2024). Intelligent portfolio construction via news sentiment analysis. *International Review of Economics & Finance*, 89(Part A), 605-617.
8. Colasanto, F., Grilli, L., Santoro, D., & Villani, G. (2022). BERT's sentiment sc

ore for portfolio optimization: A fine-tuned view in Black and Litterman model. *Neural Computing and Applications*, 34, 17507-17521.

9. Kim, D., Seok, S., & Moon, H. (2023). Asset allocation strategy based on news article sentiment analysis using BERT and Black-Litterman model. *The Korean Journal of Financial Management*, 40(5), 155-180.

10. Kang, D.-W., Yoo, S.-Y., Lee, H.-Y., & Jeong, O.-R. (2022). A study on deep learning-based stock price prediction using news sentiment analysis. *Journal of the Korea Society of Computer and Information*, 27(8), 31-39.

11. 매일경제. (n.d.). 코스피 검색 결과. <https://www.mk.co.kr/search?word=코스피>

12. 한국경제. (n.d.). 코스피 태그 결과. <https://www.hankyung.com/tag/코스피>

13. 블룸버그뉴스. (n.d.). 한국어 뉴스 <https://www.bloomberg.co.kr/blog/category/news/>

14. Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *\*Analytical Chemistry\**, 36(8), 1627-1639. <https://doi.org/10.1021/ac60214a047>

15. Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *\*Journal of Economic Dynamics and Control\**, 2, 329-352. [https://doi.org/10.1016/0165-1889\(80\)90069-X](https://doi.org/10.1016/0165-1889(80)90069-X)

16. Fuller, W. A. (1976). *\*Introduction to statistical time series\**. John Wiley & Sons.

17. Sims, C. A. (1980). Macroeconomics and reality. *\*Econometrica\**, 48(1), 1-48.

18. Sharpe, W. F. (1966). Mutual fund performance. *\*Journal of Business\**, 39(S 1), 119-138. <https://doi.org/10.1086/294846>

19. Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *\*Journal of Investing\**, 3(3), 59-64. <https://doi.org/10.3905/joi.3.3.59>

20. Grinold, R. C., & Kahn, R. N. (1999). *\*Active portfolio management: A quanti*

tative approach for producing superior returns and controlling risk\* (2nd ed.).  
McGraw Hill.

## 7. Appendix A.

date	revenue	rf_weight	sent_score	norm_senti
2018-01-01 00:00:00	4.561412193	0.597252566	85.14150943	0
2018-02-01 00:00:00	-1.099521003	0.823969144	113.7583893	0.405836901
2018-03-01 00:00:00	0.5796021	0.741804561	105.0672182	0.282581048
2018-04-01 00:00:00	-0.880358794	0.77318277	107.6419214	0.319094797
2018-05-01 00:00:00	-0.721954525	0.663325279	96.52096343	0.161380359
2018-06-01 00:00:00	-1.341270513	0.740999347	72.72727273	0
2018-07-01 00:00:00	-0.593612178	0.831833695	102.1108179	0.354328166
2018-08-01 00:00:00	1.460568868	0.837218424	92.43421053	0.237640594
2018-09-01 00:00:00	1.957533402	0.804692964	78.51239669	0.069761234
2018-10-01 00:00:00	-5.809825349	0.771882564	103.0162413	0.461731102
2018-11-01 00:00:00	2.052414571	0.75389003	90.4040404	0.373931624
2018-12-01 00:00:00	1.540405257	0.701798195	102.6143791	0.63222725
2019-01-01 00:00:00	0.192062965	0.517849871	100	0.664684014
2019-02-01 00:00:00	2.108845658	0.44843919	99.1416309	0.64376406
2019-03-01 00:00:00	-5.663870211	0.35295184	93.33333333	0.590183816
2019-04-01 00:00:00	2.251124828	0.364988029	71.04477612	0
2019-05-01 00:00:00	-3.85771816	0.568118243	97.07207207	0.814078923
2019-06-01 00:00:00	1.104328026	0.549211987	75.17241379	0.129103801
2019-07-01 00:00:00	-2.878925798	0.676341237	68.27094474	0
2019-08-01 00:00:00	-1.005453423	0.669886047	83.60655738	0.441372334
2019-09-01 00:00:00	0.873489312	0.82240197	102.7667984	0.992820815
2019-10-01 00:00:00	4.314635699	0.547673438	83.66701791	0.443112441
2019-11-01 00:00:00	-1.876076035	0.365098665	98.00569801	0.861980502
2019-12-01 00:00:00	5.267361946	0.23882921	85.64356436	0.503614718
2020-01-01 00:00:00	-1.054378213	0.392133054	71.89054726	0.104928626
2020-02-01 00:00:00	-3.592274379	0.490563085	86.91588785	0.540498093
2020-03-01 00:00:00	-0.273668148	0.624860402	114.7613762	1
2020-04-01 00:00:00	10.66461451	0.474100649	121.9978746	1
2020-05-01 00:00:00	3.562311873	0.160825668	99.47643979	0.580816643
2020-06-01 00:00:00	17.1800473	0.125755007	140.4682274	1
2020-07-01 00:00:00	4.581600743	0.146714175	109.2592593	0.567726555
2020-08-01 00:00:00	6.845046003	0.276396208	109.5074456	0.548529759
2020-09-01 00:00:00	-7.342750914	0.286736778	107.5125209	0.519439759
2020-10-01 00:00:00	-2.629745224	0.453478363	141.3349663	1