

# BERT를 활용한 뉴스 기사 감성분석과 블랙-리터만 모형을 결합한 자산 배분 전략 제안\*

김동재\*\* · 석상익\*\*\* · 문형빈\*\*\*\*

## 〈요 약〉

본 연구에서는 투자자의 주관적인 판단을 최소화하기 위해 뉴스 기사에 내포된 감성정보를 딥러닝 모형을 활용하여 추출한 후, 이를 블랙-리터만 모형에 적용하는 객관적인 자산 배분 전략을 제시한다. 연구 결과, 뉴스 기사로부터 추출한 감성을 자산별 기대수익에 대한 전망에 활용하여 포트폴리오를 구성하는 경우 시장 포트폴리오, 동일비중 포트폴리오, 평균분산 포트폴리오보다 수익성이 개선된 것을 확인하였다. 특히 감성이 긍정도 부정도 아닌 중립인 기사를 제외할 경우 수익성이 더욱 개선된다는 점을 보였으며, 통합된 감성정보의 극성을 반영하여 포트폴리오를 구성할 경우 강도를 반영하는 경우보다 자산 배분의 수익성이 개선되는 것을 확인하였다. 본 연구는 자산 배분 전략을 구성할 때 기대수익에 대한 투자자의 주관적인 판단을 활용함에 따라 발생할 수 있는 한계를 극복하는 방안으로서 대중에게 공개된 뉴스 기사의 객관적인 정보를 적용하는 자산 배분 전략 구성 방법을 제안하고, 그 활용 가능성과 우수성을 확인하였다는 측면에서 의의를 가진다.

주제어 : 딥러닝, BERT, 뉴스 기사, 감성분석, 블랙-리터만 모형, 자산 배분 전략

논문접수일 : 2023년 08 월 16일 논문수정일 : 2023년 10월 05일 논문게재확정일 : 2023년 10월 15일

\* 본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(과제번호: 2022R1F1A1076189, RS-2023-00242528).

\*\* 제1저자, 부경대학교 경영학부 학사과정, E-mail: carrytheeast@gmail.com

\*\*\* 공동교신저자, 울산대학교 경영대학 경영학부 조교수, E-mail: siseok@ulsan.ac.kr

\*\*\*\* 교신저자, 부경대학교 데이터정보과학부 빅데이터융합전공 조교수, E-mail: hbmoon@pknu.ac.kr

## I. 서 론

1952년 Markowitz가 평균분산모형(Mean-Variance Model, 이하 MV 모형)을 제시한 이래로 이 모형은 반세기 이상 동안 최적 자산 배분의 이론적 기준으로 사용되어 왔다. 하지만 MV 모형은 수익률, 표준편차 등 투입 변수에 대한 추정오차에 매우 민감하게 반응하며 표본 외(out-of-sample) 성과가 낮은 등 실무적인 측면에서는 여러 가지 취약점이 지적되었고, 이를 개선하기 위해 다양한 연구가 진행되어 왔다(Jobson and Korkie, 1981; Michaud, 1989; Best and Grauer, 1991; Birtten-Jones, 1999; Drobetz, 2001; He and Litterman, 2002). 특히 1990년에 발표된 블랙-리터만 모형(Black-Litterman 모형, 이하 BL 모형)은 일부 종목의 비중 편중, 입력 변수에 대한 민감성, 추정오차 극대화 등의 문제를 극복하는 동시에 분산투자의 장점을 유지하는 정교한 자산 배분 전략으로 널리 인용되고 있다.

BL 모형에서는 베이지안 통계에 기반하여 시장으로부터의 정보와 투자자의 견해를 결합하여 새로운 수익률과 공분산 추정치를 만들어 내고 이를 다시 MV 모형의 투입 변수로 사용하므로, 더욱 견고하고 안정적인 포트폴리오가 산출된다. 즉, BL 모형을 통해 시장정보를 통해 추정된 균형 시장 포트폴리오(equilibrium market portfolio)에 투자자의 상대적 전망을 반영하여 투자 비중을 안정적으로 조정할 수 있다.

이때 투자자 전망은 투자자가 자산의 기대수익에 대해 갖는 기대나 의견을 말하며, 이를 추정하는 방법에 대해 다양한 연구가 진행되어 왔다(Beach and Orlov, 2007; Didenko and Demicheva, 2013; Kara et al., 2019; Sokolov et al., 2021). 하지만 자산 배분 전략을 수립하는 과정에서 사적 정보와 같은 투자자의 주관적인 견해가 과도하게 포함될 경우 투자 상황을 지나치게 낙관적이거나 비관적으로 판단하는 과신(overconfidence) 또는 귀인 편향(attribution bias)과 같은 문제가 발생할 수 있다(Daniel et al., 1998; Daniel et al., 2002). 따라서 투자자는 객관적인 정보를 활용하여 개별 자산에 대한 전망을 추정해야 하며 이를 위한 적절한 방법이 필요하다.

이에 따라 본 연구에서는 BL 모형에서 투자자의 주관적인 전망을 대체할 수 있는 객관적인 방안으로서 뉴스 기사에 포함된 정보에 대한 감성분석을 제안하고자 한다. 뉴스 기사는 시장에 공개되어 있으며 시의성 있는 객관적인 정보로 특히 주가에 직간접적인 영향을 미칠 수 있는 정보를 포함하고 있다는 특징이 있다(Mahajan et al., 2008; Li et al., 2014; Shah et al., 2019; Sousa et al., 2019; Sokolov et al., 2021; Sonkiya et al., 2021). 따라서 최적 자산 배분 전략을 도출함에 있어 뉴스 기사를 활용하여 개별 자산에 대한 신속하고 객관적인 전망을 추정할 수 있다. 구체적으로, 본 연구에서 제안하는 객관적인 전망에 기반한

최적 자산 배분 전략은 다음의 과정을 통해 도출된다. 먼저, 딥러닝 기반의 감성분석 모형을 이용하여 뉴스 기사에 포함된 감성정보를 긍정(positive), 부정(negative), 중립(neutral)으로 분류하고, 이를 취합하여 각 기업에 대한 감성점수를 계산한다. 그 후 개별 기업에 대한 감성점수를 기반으로 해당 기업에 대한 투자자 전망을 추정하고, 이를 BL 모형에 통합하여 최적 포트폴리오를 추출한다. 이때 감성점수 산출 시 중립 정보의 반영 여부, 감성점수 산출 방식(극성(polarity) 또는 강도(intensity))에 따른 4가지 모형을 구성하고, 각각의 모형을 활용한 최적 포트폴리오를 도출한다. 마지막으로 각 모형의 성능을 평가하고 수익성 관점에서 최적의 모형을 제안한다. 본 연구는 BL 모형의 투자자 전망을 뉴스 기사 감성분석에 기반한 객관적인 전망으로 대체하여 투자자의 주관에 최대한 배제함으로써 과신이나 귀인 편향과 같은 문제를 최소화할 수 있는 최적의 자산 배분 전략을 제안하였다는 점에서 의의가 있다.

본 연구는 다음과 같이 구성된다. 먼저 제Ⅱ장에서는 자산 배분 전략 및 감성분석에 대한 선행연구를 검토하고 본 연구의 차별성과 구체적인 의의를 도출한다. 다음으로 제Ⅲ장에서는 연구모형을 구체적으로 기술하고 분석에 사용된 데이터에 대한 정보를 제시하며, 제Ⅳ장에서는 연구모형에 대한 실증분석 결과를 제시한다. 끝으로 제Ⅴ장에서는 본 연구를 요약하고 시사점을 도출하며 본 연구의 한계점 및 추후 연구주제를 제안한다.

## Ⅱ. 이론적 배경 및 관련 연구

### 1. 전통적인 자산 배분 전략 모형

Markowitz(1952)는 투자에 대한 위험을 과거 가격의 변동성으로 정의하고 자산간 상관관계를 이용하여 위험을 분산시키는 분산투자를 증명하였다. 이를 바탕으로 기대수익률과 변동성을 고려한 MV 모형을 개발하였으며, 자산 배분의 중요성을 강조하였다. 그러나 현실에서는 MV 모형에 대한 가정 중 일부가 충족되지 않거나 실무적으로 적용하는 과정에서 몇 가지 한계점을 보였다. 일례로 MV 모형을 적용할 경우 과거 수익률을 통해 최적해를 산출하는 과정에서 기대수익률이 높은 자산이나 음의 상관계수를 가진 자산으로 포트폴리오의 비중이 편중되는 코너 해(corner solution)가 산출되는 경우가 빈번하였으며, 기대수익률이나 분산의 추정치에 대한 작은 차이에도 포트폴리오의 최적해가 민감하게 반응하는 문제점을 보였다(Idzorek, 2005; Walters, 2014).

Black and Litterman(1990)은 자산 배분 전략에서 분산투자의 장점을 유지하면서도 MV 모형의 여러 문제점을 해결할 수 있는 BL 모형을 발표하였다. CAPM의 균형 가정 및 역

최적화에 기반한 BL 모형은 기대수익률과 같은 추정된 입력 변수가 결과에 미치는 영향을 배제했다. 또한, 균형 가정의 시장 중립적 수익률을 통해 특정 종목군에 대한 코너 해 문제를 보완하였고 투자자 전망을 선형적으로 반영할 수 있는 자산 배분 모형을 제시하였다. 이후 구체적인 구현을 위하여 신뢰수준과 거래비용 등을 반영하여 BL 모형을 확장하는 연구가 이루어졌다. Idzorek(2005)은 투자자 전망에 대한 신뢰수준(confidence level)을 백분율(0%~100%)로 수치화하여 나타냈으며, 분산 및 불확실성을 추정하는 방식을 단순화하였다. Meucci(2010)은 기존 BL 모형의 접근법에 대한 증명과 함께 일반화에 대해 논의하였으며, 질적 정보를 정량적인 투자자 전망으로 통합하는 방법을 제시하였다. Walters(2014)는 선행연구를 기반으로 추정 방법과 매개변수에 대한 해석과 사례를 보여주었으며, Kolm et al.(2021)은 거래비용과 투자자 전망의 형성 구조와 같은 실질적인 구현을 고려하여 모형을 확장하였으며, 추가로 베이지안 추정에 대한 해석을 제시하였다.

한편으로 주가 예측을 통해 투자자 전망을 더욱 정확하고 편향되지 않게 추정하여 BL 모형과 결합하는 연구가 진행되었다. 투자자 전망 추정을 위해 주로 주가 데이터를 사용하여 시계열성을 학습하는 모형이 활용되었으며, 이 외에도 주가 데이터에서 파생된 기술적 지표 및 거시경제지표를 함께 이용하였다. Beach and Orlov(2007)은 1988년부터 2003년까지 20개국의 MSCI지수와 거시경제요인의 월별 데이터를 이용하여 시계열 기반의 회귀 모형인 GARCH 모형을 추정한 후 BL 모형과 결합하였다. Kara et al.(2019)은 이스탄불 증권거래소의 BIST-30 지수와 다우존스지수의 소수 기업을 대상으로 GARCH 모형과 기계학습 기반의 서포트벡터 회귀 모형을 결합하는 BL 모형을 제시하였다. 또한, Didenko and Demicheva(2013)은 기계학습 기반의 앙상블 모형인 랜덤포레스트를 사용한 BL 모형을 제시하였다.

그러나 주가 데이터를 기반으로 한 자산 배분 전략은 시장에 대한 제한적인 정보 사용으로 인한 한계점이 있다. 시장은 연속적으로 변화하며 돌발적인 이슈나 투자자 심리로 인해 변동이 발생한다. 따라서 단순히 과거 주가만 사용할 경우 시장 변동에 대한 즉각적인 대응에 어려움이 있으며, 시장 상황에 부합하지 않는 포트폴리오가 구성될 수 있다. 이와 같은 시계열 정보의 한계를 보완하기 위해 뉴스 기사나 소셜 미디어의 투자자 감성정보를 활용하는 연구가 진행되고 있다.

## 2. 금융 분야에서 감성분석의 활용

금융 분야에서 감성분석은 투자자들이 특정 주식이나 산업, 시장에 대한 긍정적 또는 부정적인 감성을 가졌는지를 파악하기 위해 사용된다. 예를 들어, 특정 자산에 대하여

긍정적인 감성이 많이 나타나는 뉴스 기사나 소셜 미디어 게시물이 증가한다면 투자자들이 해당 자산에 대해 긍정적인 태도를 가질 가능성이 커지므로, 가격 상승이 예상된다. 이를 실증적으로 보이기 위해 Mahajan et al.(2008)은 2005년 8월부터 2008년 7월까지의 데이터를 이용하여 봄베이 증권거래소의 SENSEX지수와 뉴스 기사 간의 관계를 분석하였다. 해당 연구는 latent dirichlet allocation(LDA)을 기반으로 시장에 미치는 주요 이벤트를 식별하고 상승과 하락에 대한 분류 예측 시스템으로 이용하였으며, 분석 결과 60% 내외의 예측 정확도를 보여주었다. 또한 Li et al.(2014)은 홍콩 증권거래소를 대상으로 2003년 12월부터 2008년 3월까지의 뉴스 데이터를 활용하여 주식가격과 뉴스 기사의 관계를 분석하였다. 해당 연구는 단어 빈도 기반의 Bag of words 방식을 사용하여 뉴스 기사의 감성을 분류하였다. 2007년부터 2008년 3월까지의 데이터를 활용하여 뉴스 기사와 주가의 관계를 평가하였다. 분석 결과 뉴스 기사를 이용한 감성분석이 주가 예측의 정확도를 향상하는 데에 도움을 준다고 제시하였다. 선행연구 결과는 뉴스 기사와 주가가 유의미한 관계가 있음을 보여주었다.

한편 자연어 처리 분야에서는 ELMO(Peters et al., 2017)와 GPT(Radford et al., 2018) 같은 사전학습 언어 모델과 Transformers(Vaswani et al., 2017)와 같은 딥러닝 아키텍처가 좋은 성능을 보이며 많이 활용되고 있다. 특히 인코더-디코더 기반 Transformers의 메인 아이디어를 채택하여 BERT(Devlin et al., 2018)와 같은 언어 모델이 개발되었으며, 이외에도 최근 개발된 고성능 모델들은 Transformers 아키텍처를 기반으로 하고 있다. FinBERT(Araci, 2019)는 재무/금융 분야에서 자연어처리를 활용하기 위하여 BERT를 기반으로 개발되었으며, 주요 작업(downstream task)으로 감성분석을 진행한다. FinBERT는 금융 도메인 TRC2-Financial 데이터셋을 사전학습한 후, Finance PhraseBank(Malo et al., 2014) 데이터셋과 FiQA task 1의 FiQA sentiment 데이터셋을 이용하여 검증되었다. 해당 모형은 검증 데이터셋 모두에서 state-of-the-art를 달성하였으며, 특히 분류 작업에서 15%의 정확도 향상을 보여주었다.

최근 이와 같은 언어 모델을 이용하여 뉴스 기사에 포함된 정보를 추출할 후 주가와 의 관계를 분석하는 연구가 활발히 진행되고 있다. Sokolov et al.(2021)은 BERT를 기반으로 1998년부터 2019년까지 S&P500 기업의 뉴스 기사를 분석하여 ESG 신호를 추출하고, 이를 BL 모형과 결합하였으며, 그 결과 벤치마크 포트폴리오 대비 성과가 개선된다는 것을 보였다. Sousa et al.(2019)는 사전 학습된 BERT를 기반으로 직접 구축한 데이터셋을 추가학습하여 뉴스 기사의 감성과 다우존스지수의 관계를 분석하였으며, 예측 결과 F-score에서 72.5%, 주가 분류 예측에서 69%의 정확도를 보여주었다. 한편 Sonkiya et al.(2021)은 Apple Inc을 대상으로 사전 학습된 BERT를 이용한 뉴스 기사의 감성분석 결과와 기술 지표 및 특정

국가의 주가 지수 등을 생성적 적대 신경망(GAN)과 결합하여 주가를 예측하였다. 분석 결과 시계열 기반의 모형인 LSTM, GRU, ARIMA와 비교하였을 때 GAN 모형이 가장 우수한 성과를 기록하였으며, 특히 기술적 지표 외에 투자자의 감성분석 결과가 시장 상황에 유의한 영향을 미친다고 제시하였다. 이와 같은 선행연구 결과는 뉴스 기사에 포함된 정보가 기업의 주가와 유의미한 관계가 있음을 보여주고 있으며, 특히 뉴스 기사의 감성을 이용할 경우 주가 예측의 정확도가 높아진다는 것을 시사한다. 따라서 본 연구에서는 뉴스 기사에 포함된 감성정보가 자산 배분 전략에 유용한 정보를 제공한다는 가정을 기반으로, 이를 BL 모형에서의 투자자 전망에 활용하는 방안을 제안하고자 한다.

### Ⅲ. 방법론

#### 1. 연구모형

본 연구에서는 자산 배분 전략과 투자자 전망을 통합할 수 있는 Black and Litterman (1990)의 BL 모형을 기반으로 자산 배분 모형을 구성한다. BL 모형은 최적 비중인 균형 포트폴리오(equilibrium portfolio)와 과거 주가 수익률의 공분산을 사용하여 사전분포를 추정하는 역 최적화(reverse optimization)과정을 거쳐 내재 균형 수익률(implied equilibrium return)을 계산한다. 그러나 실무적으로 모든 자산에 대한 최적 포트폴리오를 관찰하기는 불가능하므로 시가총액에 비례한 포트폴리오, 즉 시장 포트폴리오를 최적 포트폴리오로 구성한다. BL 모형의 내재 균형 수익률을 구하는 역 최적화 과정은 다음 식 (1)과 같다.

$$\Pi = \lambda \Sigma \omega_{mkt} \quad (1)$$

여기서  $\Pi$ 는 내재 균형 수익률을 나타내며,  $\lambda$ 는 시장의 위험회피계수(risk aversion)를 나타낸다. 본 연구에서는 Walters (2014)를 참고하여 시장 포트폴리오의 평균 수익률을 분산으로 나누어 시장의 위험회피계수로 정의하였다.  $\Sigma$ 는 과거 수익률의 공분산(covariance)을 나타내며,  $\omega_{mkt}$ 는 시가총액 가치에 비례한 시장 포트폴리오를 나타낸다. 본 연구에서는 매매일 기준 최근 5개년의 주가 데이터를 사용하여 과거 수익률의 공분산을 산출하였다.

BL 모형에 사용되는 투자자 전망은 자산의 기대수익률에 대해 투자자가 갖는 기대나 의견을 말한다. 투자자 전망의 수준은 여러 자산에 대해 상대적으로 조절하는 상대적 전망과 단일 자산에 대한 절대적 전망으로 나눌 수 있으며, 이러한 투자자 전망에 대한 식은 다음과 같다.

$$P\mu = Q + \epsilon, \quad \epsilon \sim N(0, \Omega) \quad (2)$$

여기서  $P$ 는 전체  $N$ 개 자산 중 전망이 존재하는  $K$ 개 자산에 대한 전망을 나타내는 전망행렬( $K \times N$ )이다.  $\mu$ 는 각 자산에 대한 기대수익률( $N \times 1$ )을 나타낸다. 투자자가 설정한  $Q$ 는 투자자 전망에 대한 기대수익률( $K \times 1$ )을 나타내며,  $\epsilon$ 는 투자자 전망의 오차항을 나타낸다. 투자자 전망 오차항의 분산인  $\Omega$ 는 투자자 전망에 대한 불확실성으로 알려져 있으며, 각 전망에 대해 독립적이고 대각 외 요소가 0인 대각행렬( $K \times K$ )이다.

본 연구에서는 뉴스 기사에 포함되어 있는 감성정보를 활용하여 BL 모형의 투자자 전망을 추정하고, 이를 자산 배분에 적용하는 모형을 제안한다. 먼저, 뉴스 기사에 포함된 감성정보를 특정 기업의 투자자 전망에 활용하기 위하여, Meucci(2010)의 방법론을 따라 투자자 전망에 대한 기대수익률( $Q$ )을 식 (3)과 같이 정의하였다.

$$Q = \Pi + \beta\sigma \quad (3)$$

여기서  $\Pi$ 는 특정 자산의 내재 균형 수익률을 나타내며,  $\beta$ 와  $\sigma$ 는 각각 해당 자산에 대한 투자자 전망 계수와 과거 수익률의 표준편차를 나타낸다. 즉, 투자자가 각 자산의 수익률에 대해 내재 균형 수익률( $\Pi$ ) 대비 표준편차( $\sigma$ )의 몇 배만큼 변할 것으로 전망( $Q$ )하는지를 투자자 전망 계수( $\beta$ )를 통해 결정한다. 본 연구에서는 뉴스 기사의 감성정보를 활용하여 투자자 전망 계수를 도출한다.

투자자 전망 계수를 추정하기 위해 우선 뉴스 기사에 포함된 감성정보를 Araci(2019)의 FinBERT를 사용하여 추출하였다. 해당 모델은 입력데이터가 긍정( $s_{positive}$ ), 중립( $s_{neutral}$ ) 혹은 부정( $s_{negative}$ )으로 분류될 확률을 출력하는데, 이를 이용하여 각 기사의 감성점수를 도출한다. 이 후 개별 기사의 감성점수를 기업 단위로 통합하기 위하여, 각 감성점수의 극성에 대하여 1(긍정), 0(중립), -1(부정)의 가중치를 부여하고 식(4)과 같이 가중 합한 결과를 특정 기업에 대한 뉴스 정보로 활용한다.

$$\gamma = [\sum_1^h (s_{positive} \times (1) + s_{neutral} \times (0) + s_{negative} \times (-1))] / h \quad (4)$$

여기서  $\gamma$ 는 특정 기업별 뉴스 기사의 감성점수가 통합된 정보를 나타내며  $h$ 는 해당 기업에 대한 뉴스 기사의 개수를 나타낸다. 한편 모든 뉴스 기사가 주가와 연관되어 있는 내용을 포함한다고 보기는 어렵다. 본 연구에서 사용한 언어 모델인 FinBERT는 투자자 관점에서

금융 및 경제 기사에 대한 의견을 분류한 Finance PhraseBank 데이터셋을 추가로 학습한 모델이다. 해당 데이터셋은 전체 4845개의 문장으로 구성되어 있으며, 이중 59.4%가 중립으로 분류된 데이터다. 여기서 중립으로 분류된 데이터는 주가에 영향을 미치지 않는 정보를 의미한다(Malo et al.,2014). <표 1>은 Finance PhraseBank 데이터셋에서 중립으로 분류된 데이터의 예시를 나타내고 있다.

<표 1> Finance PhraseBank 중립데이터 예시

예시 문장
According to Gran, the company has no plans to move all production to Russia, although that is where the company is growing.
Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications, the statement said.
In Sweden, Gallerix accumulated SEK denominated sales were down 1 % and EUR denominated sales were up 11 %.

본 연구에서는 중립으로 분류된 기사의 사용 유무가 분석 결과에 미치는 영향을 고려하기 위해, 중립으로 분류된 기사를 제외하고 산출된 결과를 추가로 분석한다. 즉, 식(5)을 사용하여 중립적 기사가 배제된 점수를 계산한다.

$$\gamma' = \sum_1^{h'} (s_{positive} \times (1) + s_{negative} \times (-1)) / h' \tag{5}$$

여기서  $h'$ 는 중립적 기사를 제외하고 긍정 혹은 부정으로 분류된 뉴스 기사의 개수를 나타내며,  $\gamma'$ 와 이를 활용하여 계산된 통합 감성정보를 나타낸다.

다음으로, Idzorek(2005)와 Walters(2014)의 방식에 따라 투자자 전망의 신뢰수준을 모형에 반영하였다. 구체적으로, Walters(2014)의 불확실성 계수(coefficient of uncertainty)  $\alpha$ 를 도입하고, 투자자 전망 오차항의 분산을 다음 식 (6)과 같이 정의하였다.

$$\Omega = \alpha P \Sigma P^T, \alpha = (1 - conf) / conf \tag{6}$$

여기서  $conf$ 는 각 투자자 전망의 신뢰수준을 나타내며 백분율로 표현된다. 본 연구에서 투자자 전망에 대한 신뢰수준  $conf$ 을 구하는 방법은 다음 식 (7)과 같다.



$$conf = \begin{cases} Freq_{positive}/h, & \gamma > 0 \\ Freq_{neutral}/h, & \gamma = 0 \\ Freq_{negative}/h, & \gamma < 0 \end{cases} \quad (7)$$

여기서  $Freq_{positive}$ ,  $Freq_{negative}$ ,  $Freq_{neutral}$ 는 특정 기업에 대한 감성정보의 극성별 기사의 개수를 나타내며,  $h$ 는 해당 기업에 대한 전체 뉴스 기사의 개수를 나타낸다. 단, 중립적 기사를 제외하고 분석한 모형에서는 긍정 혹은 부정으로 분류된 뉴스 기사의 개수를 나타낸다. 식 (7)에 따르면, 신뢰수준  $conf$ 에 대한 계산은 통합된 감성정보( $\gamma$ )의 극성에 따라 달라진다. 예를 들어, 양의  $\gamma$ 값을 가진 기업이라면 투자자 전망에 대한 신뢰수준은 해당 기업에 대한 전체 기사 중 긍정적 기사의 비율로 표현된다. 따라서, 식 (7)로 산출되는 신뢰수준은 뉴스 기사 개수에 직접적으로 영향을 받는다. 예를 들어, 단일 뉴스 기사의 감성정보가 반영된 투자자 전망의 신뢰수준은 1로 나타나게 된다. 이는 투자 비중이 적은 수의 뉴스 기사에 의해 민감하게 바뀔 가능성이 있다는 것을 의미하며, 뉴스 기사 수가 충분한 기업에 대해서 위와 같은 연구모형을 적용하는 것이 적절하다는 것을 시사한다. 따라서, 본 연구에서는 뉴스 기사 수가 일정량 이상으로 충분한 기업을 분석대상으로 하였다. 한편, 중립적 기사를 제외하고 분석한 모형의 결과로  $\gamma'$ 이 0이 되는 경우,  $Freq_{neutral}$ 가 정의되지 않기 때문에 이 경우에 신뢰수준은 0으로 설정한다.

본 연구에서는 변동성 반영 상수( $c$ )를 사용하여 통합된 감성정보( $\gamma$ )와 결합한 뒤, 투자자 전망계수를 도출한다. 변동성 반영 상수는 식 (3)에서 투자자 전망에 대한 기대수익률에 과거 수익률의 표준편차( $\sigma$ )가 반영되는 정도를 결정한다. 예컨대 변동성 반영 상수가 1이라면, 투자자 전망이 긍정(부정)인 경우 기대수익률이 한 단위 표준편차만큼 상승(하락)할 것으로 전망한다는 것을 뜻한다. 본 연구에서는 현실적인 변동 정도를 반영하여 변동성 반영 상수로 0.1, 0.2, 0.3, 0.4 및 0.5를 사용하여 분석을 진행하였다. 투자자 전망계수는 다음 두 가지 방법으로 계산한다. 첫 번째 방법으로 통합된 감성정보( $\gamma$ )를 식 (8)을 이용하여 극성( $\rho$ )으로 변환한 후, 변동성 반영 상수  $c$ 를 곱하여 투자자 전망 계수  $\beta^1$ 로 정의한다( $\beta^1 = c\rho$ ). 따라서 첫 번째 방법은 통합된 정보가 가지는 극성을 이용하여 투자자 전망 수익률을 과거 주가 수익률의 변동성에 비례하여 조절한다.

$$\rho = \begin{cases} -1, & \gamma < 0 \\ 0, & \gamma = 0 \\ 1, & \gamma > 0 \end{cases} \quad (8)$$

통합된 감성정보를 극성으로 반영하는 방법은 사전에 정해진 과거 변동성만큼 투자자

전망 수익률을 조절하여 직관적으로 사용할 수 있다는 장점이 있으나, 정보의 방향성만 이용하기 때문에, 약한 긍정 또는 약한 부정과 같은 정보를 충분히 반영하지 못하는 한계가 있다. 따라서 두 번째 방법으로 통합된 감성정보( $\gamma$ )와 변동성 반영 상수( $c$ )를 곱하여 투자자 전망 계수  $\beta^2$ 로 정의하여 통합된 정보의 강도를 이용한다( $\beta^2=c\gamma$ ). 이를 통해 뉴스 기사에 반영된 모든 정보를 사용하여 정보의 방향성과 동시에 변동량을 반영하여 투자자 전망 수익률을 계산한다.

뉴스 기사의 감성점수를 이용한 모형으로 도출된  $P, Q, C, \Omega$ 와 내재 균형 수익률을 결합하여 최종적으로 BL 모형의 기대수익률과 투자 비중은 각각 식 (9), 식 (10)을 통하여 산출된다(Walters, 2014).

$$\Pi_{BL} = \Pi + \Sigma P^T [(P^T \Sigma P) + \Omega]^{-1} [Q - P \Pi] \tag{9}$$

$$\omega_{BL} = \Pi_{BL} (\lambda \Sigma)^{-1} \tag{10}$$

여기서  $\Pi_{BL}$ 은 모형을 통해 계산되는 블랙-리터만 모형의 사후 기대수익률을 의미하고,  $\omega_{BL}$ 은 모형을 통해 계산되는 최종 비중을 나타낸다. 최적화 과정에서는 제약조건으로 비중의 총합은 1로, 자산별 비중 범위는 0과 1 사이로 설정함으로써 공매도를 제외하였다.

본 연구에서는 감성점수가 통합된 정보의 반영 방법 간의 차이와 중립으로 분류된 기사의 사용 유무간의 차이를 확인하기 위하여 4가지의 연구모형을 분석하였다. <표 2>는 연구모형에 대한 전반적인 구성을 나타내고 있다.

<표 2> 모형 구성

모형	구성
Model 1	중립으로 분류된 기사가 포함된 데이터셋을 기반으로 통합된 감성정보의 극성을 반영하여 자산 배분한 모형
Model 2	중립으로 분류된 기사가 포함된 데이터셋을 기반으로 통합된 감성정보의 강도를 반영하여 자산 배분한 모형
Model 3	중립으로 분류된 기사가 제외된 데이터셋을 기반으로 통합된 감성정보의 극성을 반영하여 자산 배분한 모형
Model 4	중립으로 분류된 기사가 제외된 데이터셋을 기반으로 통합된 감성정보의 강도를 반영하여 자산 배분한 모형

2. 데이터

본 연구에서는 다우존스지수에 포함된 기업에 대한 2013년부터 2022년까지의 과거 주가

데이터와 2018년부터 2022년까지의 뉴스 기사를 수집하여 모형의 성능을 평가하였다. 뉴스 기사 빈도 수에 따른 편향을 통제하기 위해 다우존스지수에 포함된 기업 중 분석 기간 동안 연평균 기사 수가 20건 이상인 기업만 표본에 포함하였으며, 이를 통해 최종적으로 11개의 기업을 선정하였다.<sup>1)</sup> 분석대상 기업의 종목코드와 회사명, 시가총액은 <표 3>과 같다. 또한, 포트폴리오 성과평가 기간은 2018년 1월 2일부터 2022년 12월 30일까지 총 260주로 설정하였으며, 포트폴리오 구성(공분산, 기대수익률, 위험회피계수 등 도출)에는 매주 거래일 기준으로 직전 5년간의 주가 데이터를 사용하였다. 뉴스 기사의 신속성과 파급력을 고려하여 본 연구에서는 자산 배분 기간을 일주일로 설정하고, 매주 리밸런싱을 진행하여 총 260회의 자산 배분 과정을 거친다. 이때 직전 일주일 동안 게재된 해당 기업과 관련 뉴스 기사를 활용하여 포트폴리오를 구성한다.

<표 3> 분석대상 기업

종목코드와 회사명의 경우 CNBC에 제시된 정보(<https://www.cnbc.com/dow-30/>)를 활용하였으며 시가총액은 2023년 6월 30일 기준이다.

종목 코드	회사명	시가총액(단위: \$1B)
AAPL	Apple Inc	2977.91
AXP	American Express Co	127.15
BA	Boeing Co	127.74
DIS	Walt Disney Co	162.67
GS	Goldman Sachs Group Inc	108.03
JNJ	Johnson & Johnson	426.19
JPM	JPMorgan Chase & Co	418.41
MSFT	Microsoft Corp	2484.35
NKE	Nike Inc	173.82
TRV	Travelers Companies Inc	39.52
WMT	Walmart Inc	413.67

기업별 주식관련 데이터는 Yahoo Finance가 제공하는 재무관련 정보 서비스(yfinance API)를 통하여 수집하였으며, 전체 수집 기간 2013년부터 2022년까지 총 10년이다. 본 연구에서 사용한 과거 주가 수익률 데이터의 기업별 통계량은 <표 4>와 같다.

뉴스 기사를 수집하기 위해 New York Times가 제공하는 기사 관련 정보 서비스(NYT API)를 이용하였으며, 비즈니스 섹션에서 <표 3>의 회사명으로 검색된 기사의 요약(abstract)을 수집하였다. 수집 기간은 2018년부터 2022년까지 5년으로 수집된 기사는

1) 본 연구에서는 뉴스 기사 수에 따른 편향을 통제하기 위해 11개의 기업만 선정하였으나, 표본 선택이 결과에 영향을 미치지 않는다는 것을 보이기 위해 [부록]에 전체 다우 존스 기업에 대해 분석한 결과를 첨부하였다.

총 4,254건이며 기업별 최대 931건, 최소 135건을 수집하였다. 수집된 기업별 뉴스 기사 건수는 <표 5>와 같으며, 산업과 기업의 특성에 따라 수집된 뉴스 기사의 양이 차이를 보였다. 리밸런싱 시 특정 기업에 대한 뉴스 기사가 없는 경우에는 해당 기업에 대한 전망이 없는 것으로 반영하였다.

<표 4> 주가 수익률 통계량

연평균은 연율화된 일평균수익률을 의미하고 변동성은 연율화된 표준편차를 의미한다.

회사명	연평균	최대값	최소값	표준편차	개수
AAPL	25.835	-12.865	11.981	28.961	2,517
AXP	14.492	-14.819	21.882	29.741	2,517
BA	17.950	-23.848	24.319	38.767	2,517
DIS	9.139	-13.163	14.412	26.389	2,517
GS	14.340	-12.705	17.580	27.866	2,517
JNJ	11.225	-10.038	7.998	17.666	2,517
JPM	15.588	-14.965	18.012	26.743	2,517
MSFT	28.524	-14.739	14.217	26.944	2,517
NKE	20.775	-12.808	15.531	28.060	2,517
TRV	12.744	-20.800	13.290	22.633	2,517
WMT	9.655	-11.376	11.709	20.493	2,517

<표 5> 기업별 뉴스 기사 건수

종목코드	AAPL	AXP	BA	DIS	GS	JNJ	JPM	MSFT	NKE	TRV	WMT	합계
건	931	368	182	175	484	886	370	135	141	135	447	4254

## IV. 연구 결과

본 연구에서는 수집된 뉴스 기사 정보를 기반으로 일주일 단위로 자산 배분 전략을 수립하며, 매주 첫 거래일의 시가에 매매한다. 구체적으로, 거래일 전 주의 첫 거래일 오전 9시 이후부터 거래일 당일 오전 9시 전까지의 뉴스 기사를 기반으로 자산 배분을 진행한다. 이때 포트폴리오의 거래 수수료가 1,000달러 미만으로 추정되어 투자금액을 고려하였을 때 성과 비교에 큰 영향을 미치지 않아 계산에서 제외하였다.

### 1. 감성분석 결과

언어 모델의 감성분석 모형은 각각 긍정, 부정, 중립이 포함된 극성일 확률을 반환하며,

분석기간 동안 분류된 뉴스 기사의 감성분석 결과는 <표 6>와 같다. 해당 표는 분석기간 동안 수집된 종목별 뉴스 기사 개수와 감성분석 결과인 극성별 개수 및 비율을 나타내고 있다.

<표 6> 뉴스 기사 감성분석 결과

괄호 안은 각 감성이 차지하는 비중을 나타낸다.

종목코드	부정	중립	긍정	합계
AAPL	372(40.0%)	395(42.4%)	164(17.6%)	931(100%)
AXP	154(41.8%)	169(45.9%)	45(12.2%)	368(100%)
BA	84(46.2%)	85(46.7%)	13(7.1%)	182(100%)
DIS	65(37.1%)	91(52.0%)	19(10.9%)	175(100%)
GS	194(40.1%)	226(46.7%)	64(13.2%)	484(100%)
JNJ	356(40.2%)	414(46.7%)	116(13.1%)	886(100%)
JPM	148(40.0%)	178(48.1%)	44(11.9%)	370(100%)
MSFT	42(31.1%)	68(50.4%)	25(18.5%)	135(100%)
NKE	74(52.5%)	57(40.4%)	10(7.1%)	141(100%)
TRV	61(45.2%)	56(41.5%)	18(13.3%)	135(100%)
WMT	188(42.1%)	181(40.5%)	78(17.4%)	447(100%)
합계	1738(40.9%)	1920(45.1%)	596(14.0%)	4254(100.0%)

전체 기사 중 중립으로 분류된 기사는 약 45.1%를 차지하고 있으며, 모든 기업에서 중립으로 분류된 기사가 차지하는 비율은 40% 이상임을 알 수 있다. 이는 중립으로 분류된 기사가 연구모형에 미치는 영향을 고려해야 함을 시사한다.

2. 연구모형별 결과

앞서 <표 2>에서 제시한 바와 같이 Model 1과 Model 2는 중립으로 분류된 기사가 포함된 데이터셋을 활용하여 투자자 전망을 형성하는 모형이다. 단, Model 1은 통합된 감성정보의 극성을 반영하며, Model 2는 통합된 감성정보의 강도를 반영한다. 또한, Model 3과 Model 4는 각각 Model 1과 Model 2의 통합된 감성정보를 반영하는 방법은 동일하나, 중립으로 분류된 기사가 제외된 데이터셋을 활용하여 투자자 전망을 형성한다. <표 7>은 2019년 10월 28일을 기준으로 총 39건의 기사를 이용하여 형성한 연구모형별 블랙-리터만 모형의 사후추정치와 투자 비중을 보여주고 있다.

<표 7>을 통해 연구모형의 구성에 따라 형성하는 포트폴리오의 차이를 발견할 수 있다. 특히 종목코드 AAPL, TRV, WMT 등과 같이 통합된 감성정보( $\gamma$ )가 아주 작은 경우에, 기대수익률과 최종 비중에서 Model 1과 Model 2 간의 차이가 확연하게 나타난다는 것을

&lt;표 7&gt; 포트폴리오 최적화 예시

$W_{mkt}$ 은 시가총액기준 비중을 의미하며,  $\Pi$ 는 시장에서 관측되는 자산비중( $W_{mkt}$ )을 기반으로 계산된 내재수익률을 의미한다.  $\gamma$ 와  $conf$ 는 각각 통합된 감성정보와 신뢰도를 나타내며,  $\gamma'$ 와  $conf'$ 는 중립으로 분류된 뉴스 기사가 제외된 통합된 감성정보와 신뢰도를 나타낸다.  $\Pi_{BL}$ 은 모형을 통해 계산되는 사후 기대수익률을,  $\omega_{BL}$ 은 모형을 통해 계산되는 최종 비중을 나타낸다. 해당 결과에 설정된 변동성 반영 상수는 0.1이다.

종목코드	$W_{mkt}$ (%)	$\Pi$	$\gamma$	$conf$	$\gamma'$	$conf'$	Model 1		Model 2		Model 3		Model 4	
							$\Pi_{BL}$	$\omega_{BL}$ (%)	$\Pi_{BL}$	$\omega_{BL}$ (%)	$\Pi_{BL}$	$\omega_{BL}$ (%)	$\Pi_{BL}$	$\omega_{BL}$ (%)
AAPL	25.44	0.240	-0.092	0.500	-0.092	0.500	0.229	19.93	0.237	24.66	0.227	23.13	0.237	25.61
AXP	2.31	0.137	no view	no view	no view	no view	0.132	0.66	0.131	1.67	0.127	2.15	0.130	2.21
BA	5.36	0.176	-0.347	0.625	-0.525	0.800	0.162	0.53	0.170	4.07	0.156	1.04	0.165	3.28
DIS	6.28	0.129	0.757	1.000	0.757	1.000	0.149	15.12	0.144	15.50	0.149	20.56	0.144	16.56
GS	1.87	0.184	-0.814	1.000	-0.814	1.000	0.160	0.00	0.165	0.00	0.160	0.00	0.165	0.00
JNJ	8.75	0.102	-0.368	0.538	-0.618	0.875	0.097	1.72	0.099	6.89	0.088	1.03	0.093	3.93
JPM	9.67	0.169	-0.622	0.857	-0.886	1.000	0.152	0.00	0.157	3.07	0.148	0.89	0.150	0.79
MSFT	27.45	0.234	no view	no view	no view	no view	0.233	24.81	0.232	26.67	0.227	27.97	0.231	27.59
NKE	3.67	0.150	0.201	1.000	0.757	1.000	0.174	10.00	0.155	5.23	0.174	13.88	0.168	10.77
TRV	0.79	0.107	0.082	1.000	no view	no view	0.124	13.33	0.108	4.16	0.100	0.70	0.101	0.73
WMT	8.41	0.102	0.025	1.000	no view	no view	0.122	13.89	0.102	8.07	0.100	8.66	0.102	8.52

확인할 수 있다. 또한, 종목코드 TRV, WMT 등에서 중립으로 분류된 기사의 사용 유무에 따라 형성된 기대수익률과 최종 비중의 차이를 확인할 수 있다. 한편, 분석 과정에서 종목코드 AAPL의 Model 4 결과에서 투자자 전망으로 설정된  $\gamma'$ 이 음(-)의 값임에도 투자 비중이 증가하는 현상이 관찰되었다. 이는 투자 비중이 사후 기대수익률( $\Pi_{BL}$ )의 상대적인 크기에 의해 결정되었기 때문인 것으로 해석된다. 즉, AAPL 관련 뉴스 기사 감성분석 결과로  $\gamma'$ 가 음수가 되었으나 이를 반영한 사후 기대수익률의 크기가 다른 종목에 비해 상대적으로 커 투자 비중이 증가하게 된 것이다.

### 3. 평가지표 및 벤치마크 포트폴리오

본 연구에서는 위험지표로 연평균 표준편차인 변동성과 최대낙폭지수(Max Drawdown; 이하 MDD)를, 위험 조정 수익률로 샤프비율(Sharpe Ratio; 이하 Sh.R), 소르티노비율(Sortino Ratio; 이하 So.R)을 사용하였으며, 수익성 지표로 총 수익률과 연평균 수익률을 사용하였다. 이때 계산의 편의를 위해 무위험이자율은 0으로 가정한다. 위험 조정 수익률 중 샤프비율은 투자 성과 비교에 가장 많이 사용되는 지표이며, 포트폴리오 수익률과 변동성을 반영한 성과지표이며, 포트폴리오의 연평균수익률을 연간 변동성으로 나누어 계산한다. 소르티노비율은 샤프비율과 마찬가지로 수익률과 변동성을 반영한 성과지표지만 상승과 하락에 대한 전반적인 변동성을 고려하는 샤프비율과 달리 소르티노비율은 하락에 대한 변동성(하락위험)만 고려하는 차이점이 존재한다. 따라서 위험에 대해 더 보수적으로 평가할 수 있는 장점이 있으며 변동성이 높은 포트폴리오를 평가할 때 샤프비율보다 유리하다. 구체적으로 소르티노비율은 포트폴리오의 연평균수익률을 연간 하락 변동성으로 나누어 계산한다. 최대낙폭지수는 투자기간 동안 발생한 자산 가격의 최대 하락폭을 측정한 비율이며, 투자자가 직면할 수 있는 잠재적 손실을 평가할 수 있다. 구체적으로 최대낙폭지수는  $\frac{T-P}{P}$ 으로 계산되며,  $T$ 는 기간 동안 최고점을 나타내며,  $P$ 는 최고점 이후 기간의 최저점을 나타낸다.

포트폴리오의 성과를 비교하기 위한 벤치마크 포트폴리오는 시장 포트폴리오(market capitalization weighted portfolio; 이하 MKT)와 동일비중 포트폴리오(equally weighted portfolio; 이하 EW) 그리고 MV 모형을 이용한 평균분산 포트폴리오를 사용하였다. 시장 포트폴리오는 자산의 시가총액이 차지하는 비율과 비례하여 자산의 비중을 설정하는 전략으로, BL 모형에서 시장 포트폴리오는 투자자 전망이 없는 경우 형성되는 균형 포트폴리오를 의미한다. 동일비중 포트폴리오는 모든 자산을 동일한 비중으로 설정하는

포트폴리오로 최소한의 정보를 이용하여 형성되는 기본적인 전략이며, 복잡한 다른 자산 배분 전략보다 더 나은 성과를 보인다(Xing et al., 2018). 각 벤치마크 포트폴리오는 연구모형과 동일한 방식으로 매주 첫 주식 거래일의 시가에 자산 배분을 실시하였다.<sup>2)</sup>

#### 4. 포트폴리오 성과

본 연구의 목적은 뉴스 기사를 이용한 자산 배분 전략 구성 방법을 제안하고 활용 가능성을 보이는 것이다. 따라서 변동성 반영 상수( $c$ )를 임의로 설정하지 않고, 0.1에서부터 0.5까지 0.1씩 증가시키며 분석을 진행하였다. <표 8>은 벤치마크 포트폴리오와 연구모형을 이용한 포트폴리오의 성과를 나타내고 있다.

<표 8>에 따르면 벤치마크 포트폴리오 중에서 동일비중 포트폴리오는 변동성이 가장 작았으나, 전체적으로 저조한 성과를 기록하였다. 평균분산 포트폴리오는 높은 변동성과 함께, 수익성 지표와 소르티노비율에서 우수한 성과를 보이며 수익성이 뛰어난 포트폴리오를 형성하였다. 시장포트폴리오는 최대낙폭지수와 샤프비율에서 우수한 성과를 보였다. 시장 포트폴리오의 높은 샤프비율과 평균분산 포트폴리오보다 낮은 소르티노비율은 투자자가 고려하는 위험 요소에 따라 투자 성과를 판단하는 결과가 변동될 수 있음을 보여주고 있다.

본 연구에서는 수익성 지표로 총 수익률과 연평균 수익률을 사용하였다. 수익성 지표는 중립으로 분류된 기사를 포함한 모형에 비해 중립으로 분류된 기사를 제외한 Model 3과 Model 4가 뛰어난 성과를 보였다. 그리고 통합된 정보의 극성을 반영한 Model 1과 Model 3이 통합된 정보의 강도를 반영한 모형에 비해 뛰어난 수익성을 기록했다. 특히,  $c$ 가 증가함에 따라 수익성 지표가 증가하는 경향을 보였으며, 따라서 변동성 반영 상수인  $c$ 가 0.5인 Model 3이 가장 높은 수익성 성과를 기록하였다. 해당 모형의 총 수익률은 평균분산 포트폴리오에 비해 약 105%p 높은 314.494%로 나타났으며, 연평균수익률은 평균분산 포트폴리오의 연평균수익률에 비해 약 6.4%p 높은 36.808%로 나타났다.

또한, 통합된 정보를 반영하는 방법에 따라 변동성의 차이를 보였으며, 통합된 감성정보의 강도를 반영한 Model 2와 Model 4가 극성을 반영한 모형에 비해 비교적 안정적인 포트폴리오를 형성하였다. 변동성은  $c$ 가 증가함에 따라 증가하는 경향을 보였으며, 통합된

2) 언급한 벤치마크 포트폴리오 외에도 텍스트 정보를 활용한 다른 선행 연구 포트폴리오도 비교해야 하지만 직접적인 비교의 어려움과 구현의 한계 등으로 인해 벤치마크에서 제외하였다. 예컨대 Sokolov et al.(2021)의 경우 텍스트 정보와 BL 모형을 결합하였으나, 텍스트 정보는 ESG 신호를 추출하기 위해 사용되었다. 본 연구에서는 보편적인 전망을 추정하기 위해 텍스트 정보를 활용하였으므로 Sokolov et al.(2021)의 결과와 직접 비교하기에는 한계가 있다. 또한 구체적인 ESG 분류기에 대한 제한되어 있어 모형을 재현하기에 어려움이 있다.



정보를 반영하는 방법에 따른 차이는  $c$ 가 증가함에 따라 증가하는 경향을 보였다. 연구모형 중에서 변동성이 가장 낮은 모형은  $c$ 가 가장 낮은 0.1인 수준에서 발견되었으며,  $c$ 가 0.1인 Model 2는 동일비중 포트폴리오에 비해 약 1.56만큼 높은 것으로 나타났다. 반면에, 최대낙폭지수가 가장 낮은 모형은  $c$ 가 가장 높은 0.5인 수준에서 발견되었으며,  $c$ 가 0.5인 Model 2는 시장 포트폴리오에 비해 약 1.7만큼 낮은 것으로 나타났다. 최대낙폭지수는  $c$ 가 증가함에 따라 Model 1을 제외한 모형에서 감소하는 경향을 보였다. 가장 높은 위험성 지표인 기록한 연구모형은  $c$ 가 0.5인 수준에서 발견되었으며, Model 1의 최대낙폭지수와

<표 8> 모형별 포트폴리오 성과 비교

MDD는 최대낙폭지수를 의미하고, Sh.R은 샤프비율을, SOR은 소르티노비율을 의미한다. 연평균은 연율화된 일평균수익률을 의미하고 변동성은 연율화된 표준편차를 의미한다.  $c$ 는 변동성 반영 상수를 의미하며, 각 연구모형에서  $c$ 를 0.1부터 0.5까지 변화시키면서 모형의 성능을 평가하였다.

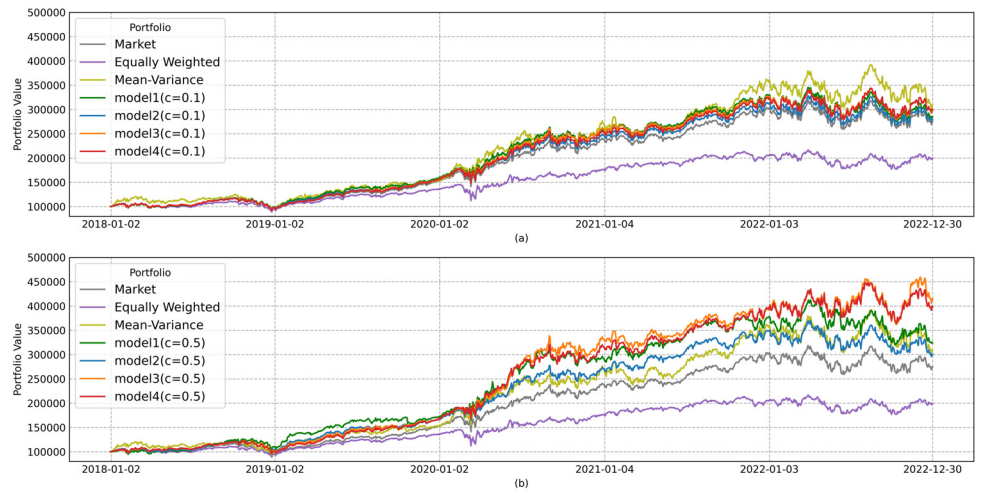
모형	총 수익률	연평균 수익률	변동성	MDD	Sh.R	So.R
Market	174.614	25.465	22.965	<b>21.693</b>	<b>1.109</b>	1.486
Equally Weighted	99.577	17.403	21.587	22.974	0.806	1.056
Mean-Variance	<b>209.441</b>	<b>30.431</b>	<b>28.855</b>	25.810	1.055	<b>1.538</b>
c=0.1						
Model 1	184.549	26.467	23.349	21.595	1.134	1.513
Model 2	179.360	25.959	<b>23.185</b>	21.325	1.120	1.489
Model 3	198.488	27.745	23.588	21.969	1.176	1.586
Model 4	199.689	27.777	23.353	22.131	1.189	1.599
c=0.2						
Model 1	196.926	27.623	23.626	21.349	1.169	1.561
Model 2	186.098	26.593	23.305	21.432	1.141	1.516
Model 3	222.545	29.881	24.095	21.834	1.240	1.690
Model 4	220.913	29.617	23.653	22.495	1.252	1.683
c=0.3						
Model 1	212.395	28.992	23.878	21.618	1.214	1.621
Model 2	194.499	27.352	23.401	20.860	1.169	1.555
Model 3	258.262	32.720	24.423	21.707	1.340	1.846
Model 4	243.278	31.421	23.840	22.405	1.318	1.781
c=0.4						
Model 1	216.175	29.370	24.100	22.016	1.219	1.621
Model 2	198.824	27.757	23.519	20.018	1.180	1.567
Model 3	291.429	35.165	24.700	21.478	1.424	1.976
Model 4	269.940	33.447	24.026	22.126	1.392	1.886
c=0.5						
Model 1	223.598	30.055	24.373	22.738	1.233	1.634
Model 2	199.771	27.863	23.600	<b>19.920</b>	1.181	1.566
Model 3	<b>314.494</b>	<b>36.808</b>	24.992	21.031	<b>1.473</b>	<b>2.059</b>
Model 4	298.017	35.456	24.196	21.780	1.465	1.981

Model 3의 변동성은 각각 22.738, 24.992를 기록하였으나, 이는 평균분산 포트폴리오의 성과에 비해 낮은 수준으로, 비교적 안정적인 포트폴리오를 형성하였다.

위험 조정 수익률이 가장 높았던  $c$ 가 0.5인 Model 3은 시장 포트폴리오의 샤프비율을 0.364만큼 상회하는 1.473로 나타났으며, 소르티노비율은 평균분산 포트폴리오의 소르티노비율을 0.521만큼 상회하는 2.059로 나타났다. 대체적으로 높은 수익성 지표를 보인 Model 3이 높은 위험조정수익률을 보이며, 이와 반대로 낮은 수익성 지표를 보인 Model 2가 낮은 위험조정수익률을 보였다.

[그림 1]은 벤치마크 포트폴리오와 함께 변동성 반영 상수가 0.1인 모형과 0.5인 모형을 이용한 포트폴리오의 누적된 자산의 변화를 각각 보여주고 있다(컬러 그림은 <http://kfma.org/38>에서 볼 수 있음).

[그림 1] 포트폴리오별 누적 자산 금액( $c=0.1$  vs.  $c=0.5$ )



[그림 1]에 따르면 변동성 반영 상수가 0.1보다 0.5인 경우에 변동폭이 증가하는 현상이 발견되었으며, 이를 통해 변동성 반영 상수가 모형에 미치는 영향을 확인할 수 있다. 동일한 변동성 반영 상수를 사용하는 모형들 간에도 차이가 관찰되었으며, 변동성 반영 상수가 0.5일 경우 모형들 사이의 성과 차이가 더욱 크게 관찰되었다. 구체적으로, 중립으로 분류된 기사의 사용 유무에 따른 모형 간 차이는 특히 2021년 하반기부터 관찰되었으며, 중립으로 분류된 기사를 제외한 모형은 증가하는 반면에 중립으로 분류된 기사를 포함한 모형은 감소하는 경향이 관찰되었다. 그리고 통합된 정보를 반영하는 방법에 따른 모형 간에 변동

폭의 차이도 관찰되었으며, 통합된 정보의 강도를 반영한 모형의 변동 폭이 상대적으로 좁게 나타났다. 이러한 경향은 특히 2020년 하반기부터 2022년 하반기까지와 2022년 사사분기 등에서 관찰되었다.

## V. 결 론

본 연구에서는 뉴스 기사에 내포된 감성정보를 딥러닝 모형을 활용하여 추출한 후 이를 블랙-리터만 모형과 통합하는 방법을 통해 과신이나 귀인 편향과 같은 투자자의 주관적인 판단에서 발생할 수 있는 문제점을 극복하는 객관적인 자산 배분 전략을 제시하고 있다. 또한 수집된 뉴스 기사의 감성정보를 기반으로 블랙-리터만 모형의 투자자 전망을 설정하는 과정에서 통합된 정보를 반영하는 방법과 긍정도 부정도 아닌 중립으로 분류된 기사의 사용 유무에 따라 차이를 보이는 연구모형을 구성하였으며, 변동성 반영 상수를 조절하여 비교하고 평가하였다.

실험 결과, 연구모형은 벤치마크 포트폴리오에 비해 우수한 결과를 보였다. 특히 변동성 반영 상수가 0.5인 Model 3은 뛰어난 수익성을 보이며 수익성 지표와 위험 조정 수익률에서 벤치마크 포트폴리오의 성과를 모두 상회하는 결과를 보였다. 이러한 결과는 뉴스 기사의 감성정보를 활용한 자산 배분 전략이 수익성과 같은 투자 성과를 개선할 가능성을 보여준다. 특히, 중립으로 분류된 기사를 제외한 데이터셋을 기반으로 형성된 포트폴리오가 중립으로 분류된 기사를 포함한 데이터셋을 기반으로 형성된 포트폴리오에 비해 뛰어난 수익성을 기록하였으며, 동시에 중립으로 분류된 기사를 제외한 모형은 변동성 반영 상수와 관계없이 위험 조정 수익률이 모든 벤치마크 포트폴리오의 성과를 뛰어넘는 결과를 보였다. 이러한 결과는 투자 상황에서 영향력이 있는 정보를 선별해서 사용해야 하는 것을 시사하며, 불필요한 정보를 제거하면 투자 성과가 개선될 수 있는 가능성을 확인하였다. 또한, 뉴스 기사의 통합된 감성정보의 극성을 반영한 모형을 이용한 포트폴리오가 통합된 감성정보의 강도를 반영한 모형을 이용한 포트폴리오에 비해 높은 수익률과 높은 변동성을 기록하였다. 이는 명확한 방향성이 있는 정보의 강도를 반영하면서 발생한 오차에서 비롯된 것으로 판단된다. 마지막으로, 변동성 반영 상수가 증가함에 따라 모든 연구모형에서 수익성이 개선될 가능성을 확인하였으며, 중립으로 분류된 기사를 제외하거나 감성정보를 강도로 반영한 모형에서 최대낙폭지수가 감소할 가능성을 확인하였다. 이는 감성정보를 최대한 반영할 수록 투자 성과가 개선될 수 있는 가능성을 나타내며, 이러한 시도는 뉴스 정보를 효과적으로 활용하여 우수한 포트폴리오를 구성할 수 있는 접근을 보여주었다.

본 연구에서는 주관적인 판단으로 발생하는 문제점을 극복하기 위하여 뉴스 기사 기반의 자산 배분 전략을 제시하였다. 뉴스 기사의 감성정보를 자산 배분 전략의 정보 수단으로 활용함으로써 포트폴리오의 성과향상 가능성을 보여주었으며 특히 수익성의 향상을 관찰할 수 있었다. 이는 뉴스 기사 기반의 자산 배분 전략이 시장참여자에게 더 나은 투자 의사 결정을 제공할 수 있음을 보여주었으며, 동시에 과거 가격 데이터만을 사용하는 전통적인 자산 배분 전략의 시장 환경에 대응하기 어려운 한계점을 극복하였다. 또한, 중립으로 분류된 기사의 사용 유무와 통합된 정보의 반영 방법과 같은 뉴스 데이터의 활용 가능성을 확장시키는 기초적인 지식을 제공하고 향후 뉴스 데이터를 활용한 투자 연구에 기여함에 의의가 있다.

본 연구의 한계와 추후 연구 주제는 다음과 같다. 첫째, 분석대상의 범위가 다우존스인덱스에 포함된 기업으로 구성되어 있어 상대적으로 한정된 시장과 기업에 대해 분석 결과가 제공되었다. 따라서 일반화를 위해서 다양한 시장과 기업을 대상으로 한 연구가 추가로 필요하다. 특히 시가총액 등 기업의 규모에 따라 본 연구모형의 성능을 확인함으로써 뉴스 기사 감성분석 기반의 자산 배분 전략 방법론의 적용 가능 환경을 면밀히 확인할 수 있을 것이다. 둘째, 뉴스 기사에 기반한 방법론의 구조로 인해 언론이 활성화되지 않은 시장이나 정보가 부족한 기업의 경우, 뉴스 기사 외의 다른 매체를 고려할 수 있다. 또한 과거 가격 및 재무 데이터 기반의 자산 가치평가모델과 결합하여 활용된다면 보다 정확하고 신뢰성 있는 자산 배분 전략을 개발할 수 있다고 판단된다.

## 참 고 문 헌

- Alexander, D. and D. Svetlana, "Application of Ensemble Learning for Views Generation in Meucci Portfolio Optimization Framework," *Review of Business and Economics Studies*, 1, (2013), 100-110.
- Araci, D., "Finbert: Financial Sentiment Analysis with pre-trained Language Models," arXiv preprint arXiv:1908.10063, (2019).
- Baker, H. K. and V. Ricciardi, "How Biases Affect Investor Behaviour," *European Financial Review*, (2014), 7-10.
- Beach, S. L. and A. G. Orlov, "An Application of the Black-Litterman Model with EGARCH-M-derived Views for International Portfolio Management," *Financial Markets and Portfolio Management*, 21, (2007), 147-166.
- Best, M. J. and R. R. Grauer, "On the Sensitivity of Mean-variance-efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results," *Review of Financial Studies*, 4(2), (1991), 315-342.
- Black, F. and R. Litterman, "Asset Allocation: Combining Investor Views with Market Equilibrium," *Goldman Sachs Fixed Income Research*, 115(1), (1990), 7-18.
- Britten-Jones, M., "The Sampling Error in Estimates of Mean-variance Efficient Portfolio Weights," *Journal of Finance*, 54(2), (1999), 655-671.
- Daniel, K., D. Hirshleifer, and A. Subrahmanyam, "Investor Psychology and Security Market Under- and Overreactions," *Journal of Finance*, 53(6), (1998), 1839-1885.
- Daniel, K., D. Hirshleifer, and S. H. Teoh, "Investor Psychology in Capital Markets: Evidence and Policy Implications," *Journal of Monetary Economics*, 49(1), (2002), 139-209.
- Devlin, J., M. W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv preprint arXiv:1810.04805, 2018.
- Drobtz, W., "How to Avoid the Pitfalls in Portfolio Optimization? Putting the Black-Litterman Approach at Work," *Financial Markets and Portfolio Management*, 15(1), (2001), 59.
- He, G. and R. Litterman, "The Intuition Behind Black-Litterman Model Portfolios," Available at SSRN 334304, 2002.

- Idzorek, T., "A Step-by-step Guide to the Black-Litterman Model: Incorporating User-specified Confidence Levels," *Forecasting Expected Returns in The Financial Markets*, Academic Press, (2005), 17-38.
- Jobson, J. D. and R. M. Korkie, "Putting Markowitz Theory to Work," *Journal of Portfolio Management*, 7(4), (1981), 70-74.
- Kara, M., A. Ulucan, and K. B. Atici, "A Hybrid Approach for Generating Investor Views in Black - Litterman Model," *Expert Systems with Applications*, 128, (2019), 256-270.
- Kolm, P. N., G. Ritter, and J. Simonian, "Black-Litterman and Beyond: The Bayesian Paradigm in Investment Management," *Journal of Portfolio Management*, 47(5), (2021), 91-113.
- Li, X., H. Xie, L. Chen, J. Wang, and X. Deng, "News Impact on Stock Price Return Via Sentiment Analysis," *Knowledge-Based Systems*, 69, (2014), 14-23.
- Mahajan, A., L. Dey, and S. M. Haque, "Mining Financial News for Major Events and Their Impacts on the Market," *IEEE*, 1, (2008).
- Malo, P., A. Sinha, P. Korhonen, J. Wallenius, and P. Takala, "Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts," *Journal of the Association for Information Science and Technology*, 65(4), (2014), 782-796.
- Markowitz, H., "Portfolio Selection," *Journal of Finance*, 7(1), (1952), 77-91.
- Meucci, A., "The Black-litterman Approach: Original Model and Extensions," *Encyclopedia of Quantitative Finance*, (2010).
- Michaud, R. O., "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?," *Financial Analysts Journal*, 45(1), (1989), 31-42.
- Peters, M. E., W. Ammar, C. Bhagavatula, and R. Power, "Semi-supervised Sequence Tagging with Bidirectional Language Models," arXiv preprint arXiv:1705.00108, 2017.
- Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-training," 2018.
- Shah, D., H. Isah, and F. Zulkermine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," *International Journal of Financial Studies*, 7(2), (2019), 26.
- Sokolov, A., K. Caverly, J. Mostovoy, T. Fahoum, and L. Seco, "Weak Supervision and Black-litterman for Automated Esg Portfolio Construction," *Journal of Financial*

*Data Science*, 3(3), (2021), 129-138.

Sonkiya, P., V. Bajpai, and A. Bansal, "Stock Price Prediction using BERT and GAN," arXiv preprint arXiv:2107.09055, 2021.

Sousa, M. G., K. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for Stock Market Sentiment Analysis," *IEEE*, (2019).

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems*, 30, 2017.

Walters, J., "The Black-Litterman Model in Detail," Available at SSRN 1314585, 2014.

Xing, F. Z., E. Cambria, and R. E. Welsch, "Intelligent Asset Allocation Via Market Sentiment Views," *IEEE*, 13(4), (2018), 25-34.

<부 록>

<부록 표 1> 다우존스 인덱스 포함 기업을 대상한 모형 성능평가 결과

Model 1~4의 경우, c=0.5를 적용한 결과임.

모형	총수익률	연평균수익률	변동성	MDD	Sh.R	So.R
Market	112.864	18.581	20.242	18.213	0.918	1.208
Equally Weighted	89.928	16.502	22.575	24.681	0.731	0.961
Mean-Variance	133.980	22.961	27.612	24.493	0.832	1.211
Model 1	90.207	16.838	23.731	27.694	0.710	0.950
Model 2	105.287	18.142	21.940	23.214	0.827	1.121
Model 3	151.187	23.346	23.263	24.096	1.004	1.381
Model 4	161.954	24.106	22.292	20.807	1.081	1.470

<부록 표 2> 기간별(2020년 이전 vs. 2020년 이후) 모형 성능 비교

Model 1~4의 경우, c=0.5를 적용한 결과임.

2020년 이전

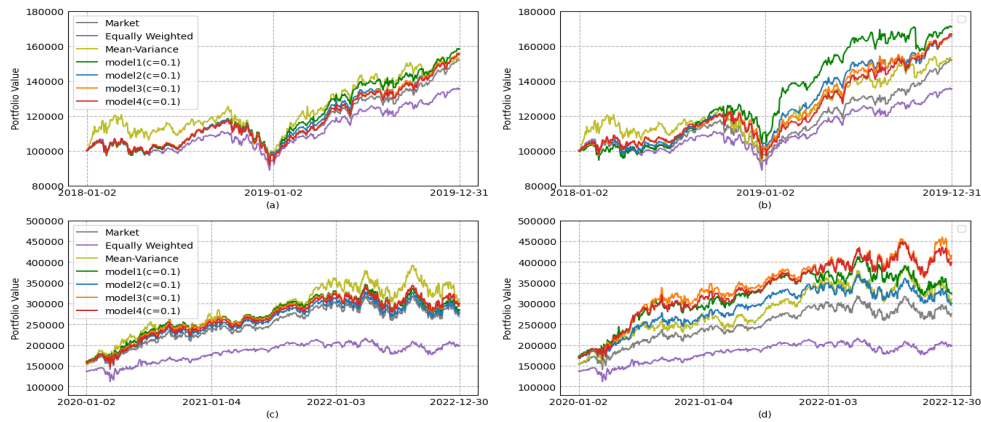
모형	총수익률	연평균수익률	변동성	MDD	Sh.R	So.R
Market	52.035	24.845	16.554	21.693	1.501	1.900
Equally Weighted	35.418	17.582	15.019	20.055	1.171	1.471
Mean-Variance	52.135	26.500	23.068	25.810	1.149	1.549
Model 1	71.109	33.123	19.572	17.663	1.692	2.099
Model 2	65.809	30.555	17.509	19.355	1.745	2.282
Model 3	66.398	30.994	18.395	21.031	1.685	2.224
Model 4	66.961	31.141	18.086	21.780	1.722	2.212

2020년 이후

모형	총수익률	연평균수익률	변동성	MDD	Sh.R	So.R
Market	77.756	25.247	26.394	18.679	0.957	1.342
Equally Weighted	45.449	16.795	25.035	22.974	0.671	0.912
Mean-Variance	99.605	32.330	32.152	23.351	1.006	1.543
Model 1	86.811	27.573	27.123	22.738	1.017	1.409
Model 2	77.863	25.455	26.913	19.920	0.946	1.292
Model 3	144.455	40.000	28.561	20.636	1.401	2.045
Model 4	134.067	37.621	27.531	18.657	1.367	1.925



[부록 그림 1] 기간별(2020년 이전 vs. 2020년 이후) 모형 성능 비교



# Asset Allocation Strategy based on News Article Sentiment Analysis using BERT and Black-Litterman Model\*

Dongjae Kim\*\* · Sang-ik Seok\*\*\* · HyungBin Moon\*\*\*\*

〈Abstract〉

This study introduces an asset allocation strategy that utilize sentiment analysis from news articles through a deep learning model. The derived sentiments are then integrated into the Black-Litterman model, offering a systematic approach to mitigate the inherent subjectivity in investor decisions. Empirical findings from this study reveal the superiority of this proposed approach compared with conventional benchmark portfolios such as market, equal-weighted, and mean-variance portfolios. In particular, the exclusion of articles generating neutral sentiment (neither positive nor negative) further enhances the profitability of the portfolio. In addition, the study shows that constructing portfolios based on the polarity of sentiment, rather than considering its positive or negative intensity, improves profitability. The significance of this study lies in its introduction of a novel framework for constructing asset allocation strategies. By utilizing objective information from publicly available news articles, it effectively circumvents the limitations tied to subjective investor judgment in predicting expected returns. The demonstrated feasibility and superiority of this data-driven approach in asset allocation strategies underscore its potential to revolutionize current practices.

Keywords : Deep Learning, BERT, News Article, Sentiment Analysis, Black-Litterman Model, Asset Allocation Strategy

\* This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Project No.: 2022R1F1A1076189, RS-2023-00242528).  
\*\* First author, Division of Business Administration, Pukyong National University, Undergraduates, E-mail: carrytheeast@gmail.com  
\*\*\* Co-Corresponding author, College of Business Administration, School of Management, University of Ulsan, Assistant Professor, E-mail: siseok@ulsan.ac.kr  
\*\*\*\* Corresponding author, Division of Data and Information Sciences, Major of Big Data Convergence, Pukyong National University, Assistant Professor, E-mail: hbmoon@pknu.ac.kr