

뉴스 기사의 감성을 사용한 기업 이익과 주식 수익률 예측*

최보미** · 야마다 아키히코*** · 양철원****

〈요 약〉

본 연구는 기업에 대한 뉴스 기사의 텍스트 정보가 향후 기업 이익과 수익률을 예측할 수 있는지 검증하였다. 이를 위해 코스피200 지수에 포함된 기업의 뉴스 기사 약 50만개를 수집한 후, KOSELF(Korean Sentiment Lexicon for Finance) 감성 사전을 사용하여 텍스트에 나온 긍정어와 부정어의 수를 측정하는 방법으로 뉴스 기사의 감성변수를 계산하였다. 이익공표일 이전의 뉴스 기사의 감성은 기업의 이익을 통계적으로 유의미하게 예측하였다. 특히 부정적 감성이 긍정적 감성보다 더 강한 예측력을 가지고 있었다. 이는 뉴스 기사에 기업의 미래 이익에 대한 의미있는 정보가 내포되어 있음을 의미한다. 더 나아가 뉴스 기사의 감성과 미래 주식 수익률과의 관계를 살펴보았다. 뉴스 기사의 감성이 이익공표일 이후의 시장 반응에 대해서는 예측력을 가지고 있지 않았지만, 다음 날의 추가 변화와는 유의미하게 연관되어 있었다. 지금까지의 결과들은 뉴스 기사의 텍스트가 기업의 미래 이익과 주식 수익률에 대한 유의미한 정보를 포함하고 있음을 보여준다. 또한, 뉴스 정보가 시장 참여자들에게 영향을 주고 있음을 시사한다.

주제어 : 뉴스 기사, 감성분석, 기업 이익, 수익률, 예측력

논문접수일 : 2024년 11월 13일 논문수정일 : 2024년 11월 28일 논문게재확정일 : 2024년 11월 29일

* 본 논문에 대해 유인한 조연을 해 주신 2024년 재무관리학회 정기학술세미나 참석자들에게 감사드립니다.
연구팀(비정형 데이터 기반 핀테크 융합연구팀)은 2024년도 대학혁신지원사업에 의한 단국대학교 연구중심학과 및 학제 간 융합연구팀 지원사업의 지원을 받아 본 연구를 수행하였습니다.

** 제1저자, 단국대학교 경영학부 석사과정, E-mail: 20infinite02@naver.com

*** 공동저자, 서울대학교 기초교육원 강의조교수, E-mail: ayamada5413@snu.ac.kr

**** 교신저자, 단국대학교 경영경제대학 경영학부 교수, E-mail: yang@dankook.ac.kr

I. 서론

금융시장에서 가장 중요한 관심은 정보가 가격에 반영되는 과정이다. 그렇다면, 다양한 참가자들에 의해 다양한 정보가 생성되어 유통되고 있는 금융시장에서 어떤 정보가 가격에 반영되는 의미있는 정보인가? 지금까지는 주로 정량적인 정보가 분석대상이었지만, 최근 빅데이터 기술의 발달로 인해 비정형 자료(unstructured data)에 대한 관심이 증대되었다. 그 중 대표적인 것이 인간의 언어로 표현된 텍스트(text)이다. 인간은 언어를 사용하여 자신의 생각을 정교하게 전달할 수 있는 존재이며, 이는 금융시장 참여자들도 동일하게 적용된다.

본 연구는 사람들이 가장 쉽게 접할 수 있는 기업에 대한 뉴스 기사를 분석함으로써 이 텍스트 안에 기업의 이익이나 수익률을 예측할 수 있는 정보가 내포되어 있는지 검증하고자 한다. 이를 위해 2023년 5월부터 2024년 5월까지 코스피200 지수에 포함된 기업의 뉴스 기사를 수집하였다. 네이버페이 증권이 뉴스공시를 통해 최종적으로 약 50만 개의 뉴스 기사 표본을 선정할 수 있었다.

수집된 뉴스 기사의 텍스트를 대상으로 긍정과 부정을 구분하는 감성 분석(sentiment analysis)을 적용하여 뉴스 기사를 해석하였다. 감성 분석은 감성 사전을 사용하여 텍스트에 나온 긍정어와 부정어의 수를 측정하는 방법을 사용하였다. 이 방법은 얼마나 정확한 감성 사전을 사용하느냐에 따라 성과가 좌우된다. 초기 연구는 하버드 대학에서 제공한 일반 감성 사전(General Inquirer)을 주로 사용하였지만, 이후 Loughran and McDonald(LM, 2011)는 일반 감성 사전이 재무분야에 적합하지 않음을 보여주며, 재무분야에 특화된 새로운 감성 사전(LM 사전)을 제시하였다. 본 연구에서는 LM의 논리를 한국어에 적용하여 개발된 KOSELF(Korean Sentiment Lexicon for Finance) 한국어 재무 감성 사전을 사용하여 뉴스 기사의 감성을 측정하였다(조수지 외 2인, 2021; Cho, Lee and Yang, 2024).

이후 뉴스를 통해 추출된 감성이 기업의 미래 이익이나 주식 수익률을 예측할 수 있는지 검증하였다. 기업의 미래 이익은 분기별로 발표되는 이익의 비기대치 변수인 SUE(Standardized Unexpected Earnings)를 사용하였다. 이는 Bernard and Thomas(1989)가 제안한 방법에 따라 계절성 랜덤 워크 모형을 사용하여 예측한 기대치를 발표 이익에서 제거하는 방법으로 산출되었다. 주식 수익률 변수는 이익발표일을 포함한 다음 날의 누적초과 수익률(CAR, cumulative abnormal return)과 뉴스 발표일 다음 날의 수익률 두 가지를 사용하였다. 최종적으로 기업 뉴스 기사의 감성과 기업 이익 및 수익률과의 관계를 회귀분석을 사용하여 검증하였다.

본 연구는 다음과 같은 점에서 기존 문헌에 기여할 수 있으리라 사료된다. 첫째, 코스피200 기업에 대한 약 50만 개의 한국어 뉴스 기사라는 대규모 표본을 사용하여 뉴스 기사 텍스트의 예측력을 검증하고 있다. 금융 분야에서 텍스트 정보가 기업의 미래 이익이나 주식 수익률에 대한 예측력을 가지고 있고, 또는 기업의 재무적 곤경이나 부도를 예측하는 데 도움을 줄 수 있다는 연구들이 다수 존재한다(Li, 2010; Kearney and Liu, 2014; Loughran and McDonald, 2016, 2020; Gentzkow, Kelly, and Taddy, 2019). 하지만, 대부분의 연구는 영어 텍스트를 기반으로 하는 미국 기업에 관한 연구들이며, 한국의 기업에 대한 연구는 부족할 실정이다. 한국 연구로는 이주화, 류두진(2021)의 연구가 있지만 기업 6개만을 선택적으로 분석하였다. 본 연구는 코스피200 기업이라는 더 폭넓은 기업들을 대상으로 광범위한 연구를 실행했다는 점에서 차별점을 지닌다. 최남현 외 2인(2023)는 337개의 기업의 뉴스 기사를 대상으로 하고 있지만, 신규 상장된 공모기업에 한정되어 있다. 본 연구는 이익공표일을 대상으로 하여 기업의 이익과 주식 수익률을 중심으로 분석하고 있다는 점에서 차이가 난다. 또한, 한국어는 언어 특성상 영어와 다른 언어이기도 하며, 한국 기업은 미국과 다른 시장환경 가운데 있다. 새로운 언어와 새로운 시장에서도 기존 연구의 결과를 확인할 수 있다면 텍스트 분석의 유용성에 대한 더 강한 확신을 가질 수 있을 것이다.

둘째, 본 연구를 통해 뉴스 정보의 역할에 대해 고찰할 수 있다. 재무 연구들은 뉴스가 아직 가격에 반영되지 않는 사적 정보를 가지고 있는 것으로 주장하기도 하며, 뉴스 발표가 정보비대칭을 완화하는데 도움을 주기 때문에 가격이나 투자자의 거래행태에 영향을 주는 것으로 보고하는 등 다양한 결과가 있다(Tetlock, Tsechansky, and Macskassy, 2008; Fang and Peress, 2009; Fang, Peress and Zheng, 2014; Calomiris and Mamaysky, 2019). 한국의 뉴스 텍스트를 분석함으로써 한국 주식시장에서 뉴스의 역할에 대한 더 발전된 답을 줄 수 있을 것이다.

본 연구는 다음과 같이 구성되어 있다. 제Ⅱ장에서는 선행연구 결과를 제시하며, 제Ⅲ장에서는 자료와 방법론을 설명한다. 제Ⅳ장에서는 실증분석 결과를 평가하고, 제Ⅴ장은 결론을 제시한다.

Ⅱ. 문헌연구

최근 빅데이터 기술의 활성화와 함께 재무분야에서의 정성적 정보 즉, 텍스트를 분석하는 연구들이 많이 진행되고 있다. 재무분야에서 텍스트 분석은 주로 텍스트의 긍정 또는 부정 감성을 추출하여 사용하는 감성 분석이 주를 이룬다. 초기 연구들은 주로 감성 사전(pre-

defined dictionary)을 활용하여 긍정 및 부정 단어 비율을 계산하는 방식으로 이루어졌으며, Harvard의 General Inquirer(GI)와 같은 범용 감성 사전 또는 금융 특화 감성 사전을 주로 사용하였다(Tetlock, 2007; Tetlock, Tsechansky, and Macskassy, 2008; Bollen, Mao, and Zeng, 2011; Loughran and McDonald, 2011; 양철원, 2021; 조수지 외 2인, 2021). 더 나아가서 사전을 사용한 방법 외에도 머신러닝 기법을 활용하여 감성을 분류하는 시도가 이루어지고 있다. 감성 분석 알고리즘을 통해 뉴스의 긍정과 부정의 감정을 학습시키고, 주가 변동 예측 모델을 구성하는 연구가 증가하고 있다. 이러한 접근은 나이브 베이즈(Naïve Bayes)나 서포트 벡터 머신(SVM), 딥러닝(Deep Learning)과 같은 기법을 사용하여 텍스트 데이터를 학습하고 분류하는 방법을 포함하며, 주가 예측에서 감성 분석의 효율성을 높이는 데 중점을 두고 있다(Li, 2010; Li, Lundholm and Minnis, 2013; Buehlmaier and Whited, 2018; 김용석, 조성욱, 2019; 서한백 외 3인, 2014).

분석 대상으로는 뉴스 기사와 같은 언론 기사의 텍스트를 통한 감성 분석으로 주가 수익률의 예측력을 검증하고자 하는 시도가 주목받고 있다. 대표적으로 Tetlock(2007)은 Wall Street Journal에 게시된 칼럼의 감성을 GI를 활용하여 추출하였으며, 부정적 어조가 익일 Dow Jones 지수의 낮은 수익률과 주식 거래량에 영향을 미친다고 결론내렸다. 또한, Antweiler and Frank(2004)는 온라인 금융 포럼의 텍스트 감성을 통해 시장 수익률을 예측했으며, Das and Chen(2007)은 뉴스의 긍정 및 부정적 감성이 향후 수익률의 방향성을 예측할 수 있음을 발견하였다. 이들은 온라인 텍스트가 투자자 심리에 영향을 미쳐 주가에 반영된다고 설명하며, 언론 기사와 소셜 미디어가 시장 예측에 있어서 중요한 정보 원천이 될 수 있음을 밝혔다. 이러한 연구는 뉴스와 같은 대중 미디어가 주가 예측력에 중요한 역할을 한다는 것을 실증적으로 뒷받침한다. Tetlock, Tsechansky, and Macskassy(2008)은 미국 개별 기업의 월스트리트 저널과 다우존스의 뉴스 기사를 활용하여 텍스트의 부정적 감성이 분기별 이익과 수익률에 대한 예측력을 가지고 있음을 보여주었다. Heston and Sinha (2017)는 90만 건 이상의 톰슨로이터 뉴스 기사를 분석하여, 일간 뉴스가 이후 1-2일 동안의 주식 수익률을, 주간 뉴스는 한 분기 동안의 주식 수익률을 예측할 수 있음을 확인하였다. 긍정적인 뉴스는 주식 수익률에 빠르게 반영되지만 부정적인 뉴스는 오랜 시간 지연된 반응을 보였다. Calomiris and Mamaysky(2019)는 51개 국가를 대상으로 뉴스 정보가 미래의 국가별 수익률과 변동성, 하락폭을 예측하며, 신흥 시장에 대한 뉴스 기사에는 더 많은 정보가 포함되어 있음을 발견하였다.

최근 한국에서는 SNS 댓글의 뉴스 감성 분석을 통해 시장 예측력을 검토한 연구 결과가 있다. 김명진 외 3인(2020)의 연구에서는 네이버 주식 토론방의 데이터를 활용하여 SNS상

투자자의 감성이 일정 수준의 주가 방향성을 예측할 수 있음을 확인하였다. 장은아 외 2인(2020)은 미국 다우존스지수 예측을 위해 뉴스 정보의 감성을 다양한 거시경제지표와 조합하여 가장 효과적인 지표를 제시하였다. 뉴스 정보의 감성 분석은 최신 자연어처리 기법인 BERT와 NLTK VADER를 사용하고, 예측모델은 주가예측모델로 적합하다고 알려진 딥러닝 예측모델 LSTM을 적용하였다. 이주화, 류두진(2021)은 한국 기업 6개를 대상으로 네이버 금융 종목토론실과 네이버 뉴스 기사를 감성 분석하여 기대지수를 도출하고, 주식 수익률에 대한 설명력을 분석하였다. 회귀분석 결과, 종목토론실 텍스트 정보가 네이버 뉴스 기사보다 높은 설명력과 유의성을 갖고 있음을 발견하였다. 최남현 외 2인(2023)는 신규 상장된 337개 공모기업을 대상으로 상장 전 시점 개인투자자의 관심을 반영하는 댓글과 해당 뉴스 기사를 활용하여 공모주 상장 직후 시초가 형성에 영향을 미치는지 분석하였다. 김동재 외 2인(2023)은 New York Times가 제공하는 뉴스 기사에 내포된 감성 정보를 딥러닝 모형을 활용하여 추출한 후, 이를 블랙-리터만 모형에 적용한 자산 배분 전략을 제시하였다.

이처럼 뉴스 텍스트의 정성적 정보가 재무 지표와 결합하면, 주가와 기업 수익 예측에 있어 유용함을 알 수 있다. 감성 분석이 제공하는 긍정 및 부정적 신호는 투자자의 심리와 시장 반응을 더욱 세밀하게 반영함으로써, 전통적인 정량적 지표와는 다른 관점에서 기업의 재무성과를 예측하는 데 기여할 수 있다.

Ⅲ. 자료 및 방법론

1. 자료

표본은 한국 주식시장 상장기업 중 2023-2024년 기준 코스피200 지수에 포함된 기업을 대상으로 하였다. 표본기간은 2023년 2분기부터 2024년 2분기까지 5분기의 기간으로 설정하였다. 뉴스 기사는 네이버페이 증권(<https://finance.naver.com/>)의 뉴스공시에서 수집하였고, 해당 기간 동안 표본기업을 대상으로 발행된 뉴스 기사를 파이썬을 사용한 크롤링 작업을 통해 추출하였다.

시장 자료는 선행연구를 바탕으로 대상기업의 시가총액(MV), 장부가 대비 시장가 비율(B/M, Book-to-Market ratio), 거래량(TV), 당기순이익, 애널리스트 예측 순이익, 종가 자료를 FnGuide에서 제공하는 DataGuide를 통해 수집하였다. 시장 초과수익률 산출을 위한 일별 시장 수익률(market return) 자료 또한 FnGuide를 통해 수집하였다.

<표 1>은 표본기업들의 특성변수에 대한 요약 통계량을 보여주고 있다. 이 통계량은 코스피200 소속 기업의 5개 분기의 관측치를 사용하였으며, 각 기업의 시계열 평균을 바탕으로 횡단면 분포를 계산한 결과이다. 기업 규모를 나타내는 시가총액(MV)(단위: 천만원)의 평균은 9천억 원이며, 표준편차는 2천 4백억 원이다. B/M의 평균은 1.38이며, 표준편차는 0.40이다. 거래량(TV)(단위: 주)의 평균은 약 90만 주며, 표준편차는 1백만 주이다.

<표 1> 표본기업의 요약통계량

본 표는 표본인 KOSPI200 기업의 특성변수에 대한 요약 통계량을 보여주고 있다. 표본기간은 2023년 2분기부터 2024년 2분기까지의 관측치이다. 하위표본별로 시가총액(MV), 장부가 대비 시장가 비율(B/M, Book-to-Market ratio), 거래량(TV)이며, FnGuide를 통해 수집하였다. 시가총액(MV)의 단위는 천만원이고, 거래량(TV)의 단위는 주이다.

	N	Mean	Median	Std.	1Q	3Q
MV	928	902,371	266,711	240,587	735,257	1,055,971
B/M	900	1.3785	1.0327	0.3958	1.0876	1.5936
TV	928	884,422	218,369	1,058,273	406,969	893,148

2. 뉴스 기사의 감성변수 구축

2023년 5월부터 2024년 5월까지의 KOSPI200 기업에 대한 뉴스 기사를 네이버페이 증권을 통해 수집하였다. 뉴스 기사는 다음과 같은 필터링 과정을 거쳤다. 첫째, 기사 내용의 첫 25단어 내에 기업명이 포함되어 있고, 둘째, 기사 전체에서 기업명이 최소 두 번 이상 언급되며, 셋째, 기사 길이가 최소 30단어 이상인 기사만을 표본으로 선정하였다. 총 502,737개의 수집된 기사 중 필터링을 거쳐 최종적으로 497,621개의 기사를 실증분석에 활용하였다.

다음으로 수집된 뉴스 기사의 감성을 계산하였다. 부정어 비율(NEG), 긍정어 비율(POS), 그리고 어조(TONE) 3개 변수는 각 뉴스 기사를 통해 기업의 전반적인 감성을 계산한 측정치이다. NEG와 POS는 각각 총 단어 수에 대한 부정 및 긍정적 단어의 비율을 나타내며, 이는 각각 식 (1)과 (2)에서 보여준다. 식 (3)의 TONE은 POS에서 NEG를 차감한 값으로 텍스트의 전반적인 감성을 평가하는 지표이다.

$$NEG = \frac{Negative\ Word\ Count}{Total\ Word\ Count} \tag{1}$$

$$POS = \frac{Positive\ Word\ Count}{Total\ Word\ Count} \tag{2}$$

$$TONE = POS - NEG \tag{3}$$

감성변수인 NEG와 POS에서 부정어와 긍정어를 분류하기 위해서 감성 사전이 필요하다. 본 연구에서는 Cho, Lee, and Yang(2024)이 한국어 금융 재무분야의 감성 분류를 위해 개발한 KOSELF(Korean Sentiment Lexicon for Finance) 감성 사전을 사용하였다. Cho, Lee, and Yang(2024)는 조수지 외 2인(2021)의 초기 KOSELF 버전을 확장하여 새로운 사전을 제안하였다. 초기 KOSELF는 단일(uni-gram)단어만을 포함하고 있으며, 단어 선택에 있어 연구자의 주관적인 판단이 개입된다는 한계를 지니고 있었다. 확장된 KOSELF는 두 단어 조합(bi-gram) 구문을 포함하여 “수익 증가/감소” 또는 “비용 증가/감소”와 같은 미세한 어조의 차이를 포착할 수 있도록 하였다. 또한, 나이트 베이즈 통계를 활용하여 자동화된 단어 선택 절차를 적용함으로써, 보고서가 축적됨에 따라 새로운 단어를 추가하여 KOSELF의 개선이 용이하게 이루어질 수 있도록 하였다. 초기 사전은 47개의 부정적인 단어와 48개의 긍정적인 단어로 구성된 단어 목록이지만, 확장된 KOSELF에는 두 단어 조합(bigram)을 포함하여 최종적으로 140개의 긍정적인 단어와 129개의 부정적인 단어로 구성되어 있다. 본 연구에서는 확장된 KOSELF를 사용하여 뉴스 기사의 감성을 측정하였다.

<표 2>는 최종 표본인 497,621개 뉴스 기사의 감성변수에 대한 요약 통계량을 보여주고 있다. NEG의 평균은 5.50%이며, POS의 평균은 6.51%이다. 뉴스 기사에서 긍정적인 단어의 비율이 부정적인 단어보다 살짝 많음을 보여준다. 따라서, 둘을 차감한 TONE 변수의 평균은 1.01%로 긍정적인 감성으로 나타난다. 표준편차에서는 NEG(4.30%)가 POS(3.96%)보다 크다. 긍정적인 감성은 뉴스 기사에 고르게 분포하는 것에 비해, 부정적인 감성은 뉴스마다 편차가 더 심함을 의미한다.

<표 2> 뉴스 기사 감성변수의 요약통계량

본 표는 전체 뉴스 기사의 감성변수에 대한 요약 통계량을 보여주고 있다. NEG는 전체 뉴스 기사에서 부정적인 단어가 차지하는 비율을 의미하며, POS는 긍정적인 단어가 차지하는 비율을 의미한다. TONE은 긍정적인 단어의 비율에서 부정적인 단어의 비율을 차감하여 전반적인 어조를 측정한 값이다.

	N	Mean	Median	Std.	1Q	3Q
NEG	497,621	0.0550	0.0462	0.0430	0.0240	0.0764
POS	497,621	0.0651	0.0600	0.0396	0.0370	0.0875
TONE	497,621	0.0101	0.0134	0.0617	-0.0238	0.0483

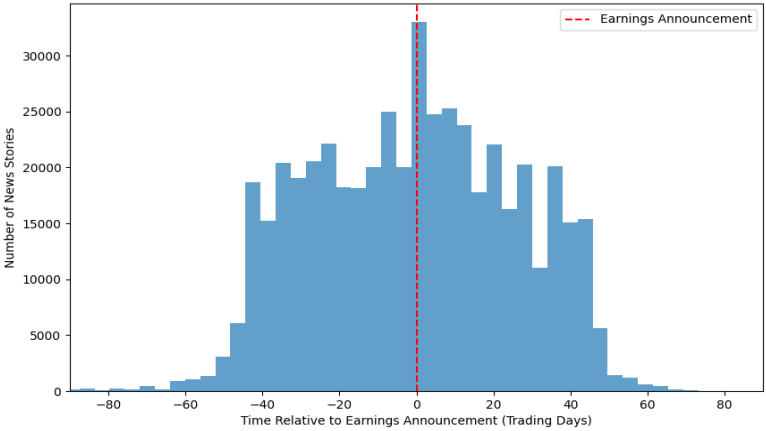
3. 뉴스 기사와 이익공표일

[그림 1]은 이익공표일 전후 90 거래일 동안의 뉴스 분포의 히스토그램을 보여주고 있다. 각 기업의 이익공표일은 KRX가 제공하는 KIND(한국투자공시시스템)와 금융감독원이

운영하는 DART(전자공시시스템)를 통해 수집하였다. 가로축은 기업의 이익공표일 전후 기간을 나타내며, 세로축은 뉴스 기사의 개수를 표시한다. 그림에서 뉴스 기사의 개수는 뉴스 발표 날짜와 가장 가까운 이익공표일을 매칭하여 계산하였다. 이러한 접근 방식은 이익공표일을 중심으로 미디어 보도가 집중되는 현상을 시각화하는 데 유리하며, 주요 사건 전후에 뉴스 기사가 급증하는 패턴을 보다 명확하게 보여준다.

[그림 1] 이익공표일 주변의 뉴스 기사수의 분포

본 그림은 이익공표 전후 90 거래일 동안의 뉴스 기사 분포를 나타낸 히스토그램이다. 가로축은 기업의 이익공표 전후 기간이고, 세로축은 뉴스 기사의 개수이다. 빨간색 점선은 이익공표일을 의미하며, 각 뉴스 기사는 해당 날짜와 가장 가까운 이익공표일과 매칭하여 계산하였다. 이를 통해 이익공표일을 중심으로 뉴스 보도가 집중되는 현상을 시각화하였다.



그림을 보면, 이익공표일을 중심으로 뉴스 기사의 빈도가 급격히 증가하는 것을 확인할 수 있다. 특히 이익공표일에 뉴스 기사가 가장 많은데, 총 497,621개의 기사 중 63,484개 (12.76%)가 당일에 집중되어 있다. 또한, 이익공표 전후 20일 구간에는 전체의 47.46%인 236,272개의 기사가, 전후 40일 구간에는 약 85.14%에 해당하는 423,859개의 기사가 분포되어 있다. 이익공표 당일에 뉴스 보도가 가장 집중되며, 대부분의 뉴스가 이익공표일로부터 40일 이내에 발표되고 있다. 이는 이익공표에 대한 시장의 기대가 뉴스 기사의 증가로 나타난 결과라 할 수 있다. 또한, 기업의 실적 공표가 투자자들의 중요한 의사결정 요소로 작용하고 있음을 추측할 수 있다.

4. 가설 및 분석방법론

본 연구에서는 뉴스 텍스트의 정보력과 관련하여 두 개의 가설을 설정하였다. 첫 번째

가설은 뉴스 텍스트의 정보가 기업의 미래 이익을 예측할 수 있을 것이라는 가설이며, 두 번째는 미래 주식 수익률을 예측할 수 있다는 가설이다.

첫 번째 가설은 미국의 Tetlock, Tsechansky, and Macskassy(2008)의 연구에 의해서 지지되고 있다. 이들은 미국 월스트리트 저널과 다우존스의 뉴스 기사에서 추출한 기업에 대한 부정적 감성이 분기별 이익에 대한 예측력을 가지고 있음을 보여주었다. 본 연구에서는 부정적 감성(NEG)뿐만 아니라, 긍정적 감성(POS) 및 이를 조합한 어조(TONE) 변수까지도 검증하였다. 뉴스의 감성이 기업의 미래 이익을 예측하는지 확인하기 위해 다음과 같은 회귀분석을 실행하였다.

$$SUE_{i,t} = \alpha + \beta SENT_{i,t-1} + \gamma Control_{i,t-1} + \epsilon_{i,t-1} \quad (4)$$

여기서, SUE는 표준화된 비기대 이익을 나타내며, SENT는 NEG, POS, TONE의 감성변수를 의미한다. Control은 통제변수를 나타낸다. 통제변수로는 전 분기의 SUE, 기업 규모(MV), 장부가 대비 시장가 비율(B/M), 거래량(TV) 변수들을 사용하였다.

만약 TONE이나 POS의 회귀계수가 양(+)의 유의한 값을 갖는다면, 기업의 미래 이익에 대한 예측력이 있다고 결론 내릴 수 있을 것이다. NEG의 경우에는 미래 이익과 반대로 움직여 음(-)의 유의한 회귀계수가 나올 것으로 예상된다.

식 (4)의 회귀분석에서 사용되는 종속 변수 SUE (Standardized Unexpected Earnings)는 다음과 같이 계산한다. SUE는 기업의 분기별 표준화된 비기대이익으로, Bernard와 Thomas(1989)가 제안한 방법에 따라 계절성 랜덤워크 모형(random walk model)을 기반으로 산출되었다. 구체적으로 다음과 같이 정의된다:

$$SUE_t = \frac{UE_t - \mu_{UE}}{\sigma_{UE}} \quad (5)$$

$$UE_t = E_t - E_{t-4} \quad (6)$$

여기서, UE_t (Unexpected Earnings)는 해당 분기의 비기대이익이며, E_t 는 분기별 실제 수익을 나타낸다. UE_t 는 계절적 변동을 제거하고 동일한 기간의 성과를 비교하기 위해 전년 동기 대비 E_{t-4} 를 차감하여 계산하였다. 식에서 μ_{UE} 와 σ_{UE} 는 이전 12분기 동안의 비기대이익의 평균과 표준편차이다. 즉, SUE는 특정 분기의 실제 기업이익이 시장 기대를 얼마나 벗어났는지를 측정하는 지표이다.

두 번째 가설은 뉴스 텍스트의 정보가 기업의 주식 수익률을 예측할 수 있을 것이라는 가설이다. 이는 두 가지 가설로 다시 나누어진다. 첫째, 이익공표 전 뉴스의 감성 지수가 이익공표일 이후의 주식 수익률을 예측한다는 가설이다. 이를 위해 뉴스의 감성이 이익공표 이후의 시장의 반응을 예측하는지, 즉 누적초과수익률과 관계있는지를 확인하기 위해 회귀분석을 실행하였다. 구체적인 회귀식은 다음과 같다.

$$CAR_{i,t} = \alpha + \beta SENT_{i,t-1} + \gamma Control_{i,t-1} + \epsilon_{i,t-1} \quad (7)$$

여기서, CAR는 누적초과수익률(CAR, Cumulative Abnormal Returns)을 나타내며, SENT는 NEG, POS, TONE의 감성변수를 의미한다. Control은 통제변수를 나타낸다. 통제변수로는 이익공표일 이전의 CAR와 기업 규모(MV), 장부가 대비 시장가 비율(B/M), 거래량(TV) 변수들을 사용하였다. 누적초과수익률(CAR)은 특정 사건 전후의 비정상 수익률(AR, Abnormal Returns)을 누적한 값으로, 주가 변동성을 측정한다. 구체적으로 CAR는 다음과 같은 방식으로 계산된다:

$$CAR_{t_1,t_2} = \sum_{t=t_1}^{t=t_2} (R_t - E(R_t)) \quad (8)$$

여기서, R_t 는 거래일 t일 주식의 실제 수익률이고, $E(R_t)$ 은 주식의 기대수익률이다. 이 둘을 뺀 값이 비정상 수익률(AR)이다. 본 연구에서 주요 변수로 사용된 $CAR_{0,1}$ 는 이익공표일부터 1일 후까지의 누적초과수익률을 나타낸다. 이를 통해 기업의 주가가 이익공표를 통해 시장 대비 얼마나 더 상승 또는 하락하였는지 측정할 수 있다.

기대수익률 $E(R_t)$ 은 자본자산가격결정모형(CAPM, Capital Asset Pricing Model)을 기반으로 다음과 같이 산출되었다.

$$E(R_t) = R_f + \beta_t (R_m - R_f) \quad (9)$$

여기서, R_f 는 무위험 이자율이며, 대응치로 국고채 1년물의 수익률을 사용했다. R_m 은 같은 기간의 시장 수익률이며, 코스피 수익률을 사용하였다. 베타(β_t)는 개별주식의 시장 민감도를 나타내며, 이일공표일 기준 (-230,-31) 거래일을 사용하여 개별주식의 수익률에 대해 시장 수익률을 회귀분석하여 추정하였다.

둘째, 뉴스의 감성 지수가 다음 날의 일별 수익률을 예측한다는 가설이다. 이는 미국의 Tetlock, Tsechansky, and Macskassy(2008)의 연구에 의해서 지지되고 있다. 이들은 미국 월스트리트 저널과 다우존스의 뉴스 기사가 발표된 다음 날의 수익률이 뉴스 텍스트의 감성과 유의미한 관계가 있음을 보여주고 있다. 구체적인 회귀식은 다음과 같다.

$$RET_{i,t} = \alpha + \beta SENT_{i,t-1} + \gamma Control_{i,t-1} + \epsilon_{i,t-1} \quad (10)$$

여기서, RET는 일별수익률을 나타내며, SENT는 NEG, POS, TONE의 감성변수를 의미한다. Control은 통제변수를 나타낸다. 통제변수로는 전 날의 수익률과 기업 규모(MV), 장부가 대비 시장가 비율(B/M), 거래량(TV) 변수들을 사용하였다.

위의 두 개의 가설에서 설명변수로 사용되는 뉴스 기사의 감성변수를 구축할 때, 예측력을 검증하기 위한 구축 기간에 대한 설정이 필요하다. 이를 위해 이익공표 전의 어느 기간에 유의한 정보가 생성되는지에 대한 고찰이 필요하다. [그림 1]를 보면, 이익공표일 전 40여 일 전부터 상당히 많은 뉴스의 증가가 발견되어 지속되는 것을 볼 수 있다. 이에 근거하여 본 연구에서는 장기간 뉴스와 단기간 뉴스로 나누어서 뉴스 기사의 정보력을 검증하였다. 첫째, 장기간은 기업의 이익공표 전 -30일부터 -3일까지 약 한 달 전에 해당하는 기간으로 설정하였다. 둘째, 단기간은 기업의 이익공표 전 -7일부터 -1일까지 약 일주일 전에 해당하는 뉴스 기사 감성을 사용하여 독립 변수를 구축하였다. 예를 들어, 특정 주식의 특정 분기의 NEG(-7,-1)은 이익공표일 기준 (-7,-1)에 해당 하는 기간에 발표된 모든 뉴스 기사 텍스트를 통합하여 부정어 수를 구하고 전체 단어 수를 나누어서 계산하였다. 본 연구의 표본이 KOSPI200 기업의 5분기이므로 이론적으로 1,000개의 관측치가 만들어져야 하지만, 뉴스가 관측이 안되는 경우를 가정하면 실제로는 800개 정도의 관측치를 가지게 된다. POS나 TONE에 대해서도 동일하게 계산하였다. 이에 대한 자세한 분포는 <표 3>에 제시하였다.

<표 3>은 회귀분석에 사용되는 변수들의 요약 통계량을 보여준다. 우선 이익공표일을 관측치로 한 회귀분석식 (4)와 (7)에 사용될 변수들은 다음과 같다. SUE의 평균은 -0.0864로 음(-)의 값을 보이며, 이는 표본기간 동안 전반적으로 이익충격(earning shock)이 많았음을 나타낸다. CAR(0,1)는 이익공표 당일과 그다음 날의 초과 수익률을 누적하여 분기별로 측정한 값이다. CAR(0,1)의 평균은 0.2634%로 양(+)의 값을 보이지만, 중앙값은 -0.0836%로 나타나 평균과 차이가 크다. 표준편차는 4.9971%로 변동성이 큰 것으로 보인다. 이는 이익공표 직후 투자자들의 반응이 기업마다 상이하게 나타나면서 누적초과수익률이 넓게 분포하며, 일부 기업은 큰 폭의 수익률 변화를 경험함을 의미한다. 반면, CAR(0,2)는 이익공표 당일과

그로부터 이틀 후까지의 누적초과수익률을 측정한 값이다. CAR(0,2) 역시 평균이 0.3208%로 양(+)의 값을 보이고, 중앙값은 -0.2640%로 음(-)의 값을 보인다. 표준편차는 7.0497%로 CAR(0,1)에 비해 높다. 이를 통해 이익공표 후의 주가 반응이 시간이 지남에 따라 더 큰 변동성을 나타낼 수 있음을 알 수 있다. 이는 이익공표 이후 추가적인 정보가 시장에 반영되면서 기업별로 주가 반응이 더욱 다양해지는 경향을 시사한다.

설명변수인 감성변수는 단기와 장기로 나뉜다. 단기 감성변수 NEG(-7,-1)의 평균 0.0585보다 POS(-7,-1)의 평균 0.0634가 더 크다. 따라서, 둘의 차이인 TONE(-7,-1)의 평균이 0.0049로 양(+)의 값을 갖는다. 장기 감성변수도 POS(-30,-3)의 평균 0.0650이 NEG(-30,-3)의 평균 0.0587보다 크며, TONE(-30,-3)의 평균이 0.0063으로 양(+)의 값을 갖는다. 장기 TONE 변수가 단기 TONE 변수보다 더 큰 특성을 보인다.

다음은 뉴스 발표 일자를 기준으로 한 회귀분석식 (10)에 사용될 변수들이다. RET는 뉴스 발표 다음 날의 원 수익률을 나타내며, 중앙값은 0에 가깝고 평균은 0.0636%로 나타났다. 이는 뉴스에 대한 시장 반응이 크지 않거나 특정 방향으로 편향되지 않음을 보여준다. RET_EX은 뉴스 발표 다음 날의 초과수익률이며, 원 수익률에서 시장 수익률을 차감하였다. RET_EX도 평균이 0.0180%로 양(+)의 값을 보이지만, 중앙값이 -0.1138%로 음(-)의 값을 가진다.

<표 3> 회귀분석 변수의 요약통계량

본 표는 회귀분석에 사용된 주요 변수들의 요약 통계량을 제시하고 있다. SUE는 비기대이익을 나타내며, CAR(0,1)은 이익공표 당일과 그 다음 날의 누적초과수익률, CAR(0,2)는 이익공표 당일과 그로부터 이틀 후까지의 누적초과수익률을 측정한 값이다. 감성변수는 단기와 장기로 구분되어 있다. 단기 감성변수인 NEG(-7,-1)와 POS(-7,-1)는 이익공표일 전 7일에서 1일 전까지의 부정 및 긍정 감성을 각각 나타내고, TONE(-7,-1)은 둘의 차이로 계산한 글 전체의 단기적인 어조를 나타낸다. 장기 감성변수인 NEG(-30,-3)와 POS(-30,-3)는 이익공표일 전 30일에서 3일 전까지의 부정 및 긍정 감성을 나타내며, TONE(-30,-3)은 장기적인 어조를 측정한 값이다. RET는 뉴스 발표 다음 날의 원 수익률을 나타내며, RET_EX는 뉴스 발표 다음 날의 초과수익률을 의미한다.

	N	Mean	Median	Std.	1Q	3Q
SUE	970	-0.0864	-0.0843	0.9119	-0.5022	0.3700
CAR(0,1)(%)	828	0.2634	-0.0836	4.9971	-2.2066	2.2267
CAR(0,2)(%)	828	0.3208	-0.2640	7.0497	-2.8618	2.8463
NEG(-7,-1)	762	0.0585	0.0579	0.0173	0.0501	0.0663
POS(-7,-1)	762	0.0634	0.0631	0.0141	0.0565	0.0698
TONE(-7,-1)	762	0.0049	0.0052	0.0239	-0.0065	0.0165
NEG(-30,-3)	805	0.0587	0.0574	0.0122	0.0520	0.0643
POS(-30,-3)	805	0.0650	0.0648	0.0104	0.0599	0.0700
TONE(-30,-3)	805	0.0063	0.0076	0.0172	-0.0018	0.0156
RET(%)	73,135	0.0636	-0.0060	2.5884	-1.2000	1.0800
RET_EX(%)	73,135	0.0180	-0.1138	2.3815	-1.1217	0.9405

IV. 실증분석 결과

1. 기업의 이익 예측

이 부분에서는 뉴스의 감성이 기업의 이익을 예측할 수 있는지 검증한다. 회귀분석의 종속 변수로 SUE를 사용한다. SUE가 양(+)의 값으로 클수록 Earning Surprise로, 음(-)의 값으로 클수록 Earning Shock으로 해석할 수 있다. 본 회귀분석에서는 장기 및 단기 두 가지 기간에 따른 감성 지수를 독립 변수로 설정하여 분석을 진행하였다.

<표 4>는 장기간의 감성 지수를 사용한 SUE의 회귀분석 결과를 보여준다. NEG(-30,-3)는 SUE에 대해 -4.4820의 음(-)의 회귀계수를 가지며, t-값이 -2.46로 통계적으로 유의하다. 이는 장기간의 부정적인 뉴스가 기업의 예상치 못한 이익 감소를 예측하는 정보를 내포하고 있음을 시사한다. 반면, POS(-30,-3)는 3.0394의 양(+)의 회귀계수를 보였으나, t-값이

<표 4> 표준화된 비기대이익(SUE)에 대한 장기간 뉴스 기사 감성의 회귀분석

본 표는 표준화된 비기대이익(SUE)을 종속 변수로 하여 장기 뉴스 감성 지수가 미치는 영향을 회귀분석한 결과를 보여준다. 독립 변수로는 부정 비율(NEG -30,-3), 긍정 비율(POS -30,-3), 전반적 어조(TONE -30,-3)가 사용되었으며, 통제 변수로 기업 규모(MV), 시장 가치 대비 장부 가치 비율(B/M), 거래량(TV), 이전 분기의 SUE가 사용되었다. 회귀분석의 정확성을 높이기 위해 MV, B/M, TV는 로그(log)로 전환하였고, 이전 분기의 비기대이익이 현재 분기에 미치는 영향을 반영하기 위해 Lag(SUE)를 사용하였다. R²는 모형의 설명력을 나타내며, 표본 수(Obs.)가 함께 제시된다. 표에서 *, **, ***은 각각 10%, 5%, 1% 수준에서의 통계적으로 유의함을 나타낸다.

	(1)	(2)	(3)
Intercept	1.0239 (2.68)***	0.8278 (2.12)**	0.8771 (2.30)**
NEG(-30,-3)	-4.4820 (-2.46)**		
POS(-30,-3)		3.0394 (1.33)	
TONE(-30,-3)			3.5242 (2.61)***
log(MV)	-0.0377 (-1.33)	-0.0527 (-1.88)*	-0.0446 (-1.60)
log(B/M)	-0.1380 (-4.54)***	-0.1365 (-4.44)***	-0.1322 (-4.32)***
log(TV)	-0.0321 (-1.45)	-0.0349 (-1.57)	-0.0348 (-1.57)
Lag(SUE)	-0.3934 (-14.65)***	-0.3888 (-14.49)***	-0.3935 (-14.66)***
R ²	0.20	0.20	0.20
Obs.	970	970	970

1.33으로 나타나 긍정적 뉴스의 정보력이 통계적으로 유의하지 않음을 확인할 수 있었다. 한편, TONE(-30,-3)의 회귀계수는 3.5242이며, t-값이 2.61으로 유의한 결과를 나타낸다. 이는 장기간의 전반적인 뉴스 어조가 기업의 이익예측에 중요한 역할을 할 수 있음을 다시한 번 확인해 준다.

<표 5>는 단기간의 감성 지수를 사용한 SUE의 회귀분석 결과를 보여준다. NEG(-7,-1)는 SUE에 대해 -4.2755의 음(-)의 회귀계수를 가지며, t-값이 -2.97로 유의한 부정적 정보력을 지니고 있다. 이는 단기적으로도 부정적인 뉴스가 기업이익 감소를 예측할 수 있음을 의미한다. POS(-7,-1)는 3.5562의 양(+)의 회귀계수를 가지며 t-값이 1.93로 유의한 양(+)의 회귀계수를 가진다. 이는 단기간의 긍정적 뉴스는 기업의 이익 증가에 대한 유의한 정보가 포함되어 있음을 시사한다. 또한, TONE(-7,-1)은 3.6403의 회귀계수와 3.37의 t-값을 가지며, 유의한 양(+)의 관계를 나타낸다. 이로써 단기 뉴스 감성이 기업이익 예측에 유의한 역할을 할 수 있음을 보여준다.

<표 5> 표준화된 비기대이익(SUE)에 대한 단기간 뉴스 기사 감성의 회귀분석

본 표는 표준화된 비기대이익(SUE)을 종속 변수로 하여 단기 뉴스 감성 지수가 미치는 영향을 회귀분석한 결과를 보여준다. 독립 변수로는 부정 비율(NEG -7,-1), 긍정 비율(POS -7,-1), 전반적 어조(TONE -7,-1)가 사용되었으며, 통제 변수로 기업 규모(MV), 시장 가치 대비 장부 가치 비율(B/M), 거래량(TV), 이전 분기의 SUE가 사용되었다. 회귀분석의 정확성을 높이기 위해 MV, B/M, TV는 로그(log)로 전환하였고, 이전 분기의 비기대이익이 현재 분기에 미치는 영향을 반영하기 위해 Lag(SUE)를 사용하였다. R²는 모형의 설명력을 나타내며, 표본 수(Obs.)가 함께 제시된다. 표에서 *, **, ***은 각각 10%, 5%, 1% 수준에서의 통계적으로 유의함을 나타낸다.

	(1)	(2)	(3)
Intercept	1.0059 (2.64)***	0.8364 (2.17)**	0.8912 (2.34)**
NEG(-7,-1)	-4.2755 (-2.97)***		
POS(-7,-1)		3.5562 (1.93)*	
TONE(-7,-1)			3.6403 (3.37)***
log(MV)	-0.0360 (-1.28)	-0.0556 (-1.98)**	-0.0449 (-1.61)
log(B/M)	-0.1421 (-4.69)***	-0.1381 (-4.53)***	-0.1379 (-4.55)***
log(TV)	-0.0335 (-1.52)	-0.0345 (-1.56)	-0.0353 (-1.60)
Lag(SUE)	-0.3933 (-14.68)***	-0.3876 (-14.47)***	-0.3925 (-14.69)***
R2	0.20	0.19	0.20
Obs.	970	970	970

<표 4>와 <표 5>를 비교해 보면, NEG는 장기와 단기 모두에서 SUE에 일관되게 유의한 음(-)의 관계를 가진다. 이는 부정적 뉴스 감성이 기업의 미래 성과를 예측하는 중요한 정보 원천이 될 수 있음을 시사한다. 반면, POS는 단기간에서만 유의한 양(+)의 영향을 보이며, 장기에서는 통계적 유의성이 발견되지 않는다. 긍정적 뉴스의 경우, 정보력이 단기간에서 더 유의하게 작용하는 것으로 해석할 수 있다. TONE 변수 또한 장기와 단기 모두에서 유의한 양의 영향을 미쳤으나, 단기간에서의 영향력이 더 크다. 따라서 단기적인 뉴스가 기업이익 예측에 있어 더 유의미한 정보를 가지고 있다고 볼 수 있다. 본 연구의 결과는 미국에서 뉴스 기사의 감성이 분기별 이익을 예측할 수 있다는 Tetlock, Tsechansky, and Macskassy(2008)의 연구결과와도 일치한다. 이들은 부정적 감성변수만을 사용하였지만, 본 연구는 부정적 감성변수뿐 아니라 긍정적 감성변수를 사용하여서도 예측력을 확인하였다.

2. 주식수익률 예측

1) 이익발표 이후 주식수익률

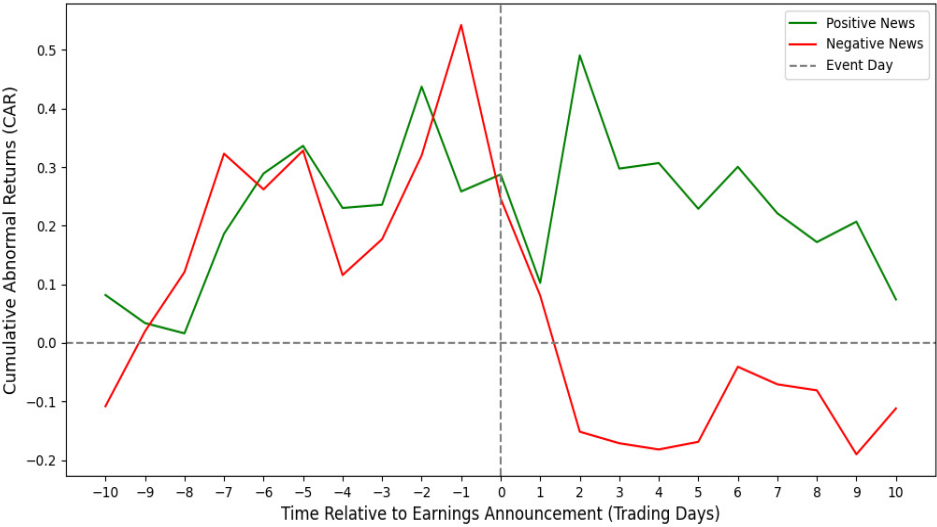
본 장은 뉴스 기사의 감성이 이익공표일의 수익률을 예측할 수 있는지 검증한다. 먼저 [그림 2]는 이익 공표 전후 10 거래일 동안 긍정 및 부정 감성변수에 따른 누적초과수익률(CAR)의 시계열 추이를 보여준다. 감성 분석을 통해 뉴스 기사의 감성을 계산하여 식 (3)에서 제시한 어조(TONE) 변수를 구축하였다. 전체 기업-분기 조합 관측치들의 TONE(-30-3) 변수를 정렬하여 상위 25%(긍정)와 하위 25%(부정)에 해당하는 뉴스를 대상으로 CAR를 계산했다. 긍정적 뉴스에 해당하는 기업-분기 표본의 CAR 평균값은 초록색 선으로, 부정적 뉴스에 해당하는 표본에 대해서는 빨간색 선을 사용하여 구분하였다.

[그림 2]를 보면, 이익공표 전 긍정적 뉴스를 가지고 있는 기업은 이익공표 후 CAR가 양(+)의 값을 가진다. 하지만, 이익공표 전 부정적 뉴스를 가진 기업은 공표 후 CAR가 큰 폭으로 감소하며 음(-)의 CAR를 보인다. 특히, 부정적인 감성의 뉴스의 영향력이 긍정적인 경우보다 더 명확함을 확인할 수 있다. 이익 공표 전 뉴스의 감성이 어느 정도 이익공표 후 수익률을 예측하는 것으로 보이지만, 통계적 유의성에 대한 구체적 검증은 회귀분석을 통해 확인할 필요가 있다.

회귀분석을 위한 종속변수로는 이익공표 후 수익률인 CAR(0,1)와 CAR(0,2)를 사용하였다. 이익공표일을 포함한 다음 거래일과 2일차 거래일까지의 수익률을 검증하였다. [그림 2]를 보면, 2일차까지의 수익률이 뉴스의 감성을 가장 잘 대변하는 것으로 보이기 때문이다. 회귀분석은 뉴스의 감성변수를 독립 변수로 설정하여 진행하였다.

[그림 2] 긍정과 부정 뉴스 기사로 구분한 이익공표일 주변의 누적초과수익률(CAR)

본 그림은 이익공표 전후 10 거래일 동안 긍정 및 부정적 감성에 따른 누적초과수익률(CAR)의 시계열 변화를 나타낸다. 가로축은 기업의 이익 공표 전후 10 거래일의 기간이고, 세로축은 누적초과수익률(CAR)의 평균값이다. 점선으로 표시된 부분은 이익공표일(Event Day)을 의미한다. 감성 분석을 통해 각 기업-분기 조합에 대한 어조(TONE)를 계산하고, 이를 기반으로 상위 25%에 해당하는 긍정적 뉴스와 하위 25%에 해당하는 부정적 뉴스에 대해 CAR를 계산하였다. 초록색 선은 긍정적 뉴스에 해당하는 CAR 평균값을 나타내며, 빨간색 선은 부정적 뉴스에 해당하는 CAR 평균값을 나타낸다.



<표 6>은 장기간의 감성 지수를 사용하여 이익공표를 기점으로 측정한 CAR(0,1)와 CAR(0,2)의 회귀분석 결과를 보여준다. CAR(0,1)는 이익공표 당일과 다음날의 누적초과수익률을 나타내며, 기업의 단기적인 주가 반응을 포착하는 데 중점을 둔다. 반면, CAR(0,2)은 이익공표 당일과 그로부터 이틀 후까지의 누적초과수익률을 측정하여, 기업의 단기적인 주가 반응뿐만 아니라 추가적인 정보가 주가에 반영되는 과정을 파악하는 데 유용하다. 이 두 변수를 모두 분석함으로써 이익공표 전후의 단기적 주가 반응을 다각도로 파악할 수 있다.

CAR(0,1)의 회귀분석 결과에서, 이익공표일 전 장기간의 부정적인 뉴스 감성을 나타내는 NEG(-30,-3)는 -1.8332의 음(-)의 회귀계수를 가지지만, t-값이 -0.21로 유의미하지 않다. POS(-30,-3)는 -11.6216의 음(-)의 회귀계수를 가질뿐더러 통계적으로 유의하지 않다. 이에 따라서 뉴스 기사의 전반적인 어조를 나타내는 TONE(-30,-3)도 유의하지 않은 음(-)의 회귀계수를 가진다. CAR(0,2)의 회귀분석 결과 역시 세 가지 감성변수 모두에서 통계적 유의성이 발견되지 않는다.

<표 6> 이익공표일 누적초과수익률(CAR)에 대한 장기간 뉴스 기사 감성의 회귀분석

본 표는 이익공표일 전 장기 감성 지수가 기업의 이익공표일 누적초과수익률에 미치는 영향을 회귀 분석한 결과를 나타낸다. 종속 변수는 이익공표 당일과 그 다음 날의 누적초과수익률을 나타내는 CAR(0,1)와 CAR(0,2)을 사용하였다. 독립 변수로는 장기간의 부정 비율(NEG -30,-3), 긍정 비율(POS -30,-3), 뉴스의 전반적 어조(TONE -30,-3)를 사용하였다. 통제 변수로는 기업 규모(log(MV)), 시장 가치 대비 장부 가치 비율(log(B/M)), 거래량(log(TV)), 비기대 이익(SUE) 및 이전 주가의 누적초과수익률 변수 AR(-1), AR(-2), CAR(-30,-3)를 사용하였다. MV, B/M, TV는 변수의 정규화를 위해 로그(log)로 변환하였으며, 현재 수익률에 대한 지연 효과를 반영하기 위해 이전 주가 변화 변수와 SUE를 통제변수로 사용하였다. R²는 모형의 설명력을 나타내며, 표본 수(Obs.)가 함께 제시된다. 표에서 *, **, ***는 각각 10%, 5%, 1% 수준에서의 통계적으로 유의함을 나타낸다.

	CAR (0,1)			CAR (0,2)		
	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	0.9467 (0.40)	1.3076 (0.54)	0.9504 (0.40)	1.1247 (1.08)	1.8297 (1.24)	1.4140 (1.15)
NEG(-30,-3)	-1.8332 (-0.21)			-7.4340 (-0.53)		
POS(-30,-3)		-11.6216 (-0.94)			16.9671 (0.85)	
TONE(-30,-3)			-2.3865 (-0.35)			9.5550 (0.88)
AR(-1)	0.0817 (1.09)	0.0801 (1.07)	0.0813 (1.09)	0.0766 (0.76)	0.0761 (0.75)	0.0753 (0.74)
AR(-2)	0.0848 (1.17)	0.0836 (1.16)	0.0833 (1.15)	0.0441 (4.34)***	0.4378 (4.30)***	0.4389 (4.31)***
CAR(-30,-3)	0.0091 (0.93)	0.0092 (0.94)	0.0092 (0.94)	-0.0217 (-1.62)	-0.0215 (-1.61)	-0.0213 (-1.59)
SUE	0.2518 (1.58)	0.2576 (1.62)	0.2557 (1.61)	0.1598 (0.63)	0.1609 (0.63)	0.1645 (0.65)
log(MV)	-0.0905 (-0.52)	-0.0760 (-0.44)	-0.0957 (-0.55)	-0.3215 (-1.15)	-0.2777 (-0.99)	-0.3106 (-1.12)
log(B/M)	-0.1382 (-0.78)	-0.1596 (-0.89)	-0.1455 (-0.82)	-0.3630 (-1.28)	-0.3864 (-1.36)	-0.3808 (-1.34)
log(TV)	0.0620 (0.46)	0.0684 (0.50)	0.0609 (0.45)	0.1159 (0.53)	0.1308 (0.60)	0.1199 (0.55)
R ²	0.01	0.01	0.01	0.03	0.03	0.03
Obs.	828	828	828	828	828	828

<표 7>은 단기간의 감성 지수를 사용하여 이익공표를 기점으로 측정한 CAR(0,1)와 CAR(0,2)의 회귀분석 결과를 보여준다. CAR(0,1)의 회귀분석 결과에서, NEG(-7,-1)는 -7.6514의 음(-)의 회귀계수를 가지지만, t-값이 -1.12로 통계적으로 유의하지 않다. POS(-7,-1)는 -5.4093의 음(-)의 회귀계수를 보일 뿐아니라 유의미하지 않다. TONE(-7,-1) 또한 2.2928의 통계적으로 유의하지 않다. 따라서, 단기간 감성변수는 이익공표일을 포함한 그 다음 날의 누적초과수익률(CAR)과 유의미한 관계를 가지지 않는 것으로 보인다.

CAR(0,2)의 회귀분석 결과에서도 통계적으로 유의미한 감성변수는 없다.

<표 7> 이익공표일 누적초과수익률(CAR)에 대한 단기간 뉴스 기사 감성의 회귀분석

본 표는 이익공표일 전 단기 감성 지수가 기업의 이익공표일 누적초과수익률에 미치는 영향을 회귀 분석한 결과를 나타낸다. 종속 변수는 이익공표 당일과 그 다음 날의 누적초과수익률을 나타내는 CAR(0,1)와 CAR(0,2)을 사용하였다. 독립 변수로는 단기간의 부정 비율(NEG -7,-1), 긍정 비율(POS -7,-1), 뉴스의 전반적 어조(TONE -7,-1)를 사용하였다. 통제 변수로는 기업 규모(log(MV)), 시장 가치 대비 장부 가치 비율(log(B/M)), 거래량(log(TV)), 비기대 이익(SUE) 및 이전 주가의 누적초과수익률 변수 AR(-1), AR(-2), CAR(-30,-3)를 사용하였다. MV, B/M, TV는 변수의 정규화를 위해 로그(log)로 변환하였으며, 현재 수익률에 대한 지연 효과를 반영하기 위해 이전 주가 변화 변수와 SUE를 통제변수로 사용하였다. R²는 모형의 설명력을 나타내며, 표본 수(Obs.)가 함께 제시된다. 표에서 *, **, ***은 각각 10%, 5%, 1% 수준에서의 통계적으로 유의함을 나타낸다.

	CAR (0,1)			CAR (0,2)		
	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	1.0099 (0.42)	0.9991 (0.42)	0.9057 (0.38)	1.3706 (1.14)	1.2157 (1.10)	1.2284 (1.11)
NEG(-7,-1)	-7.6514 (-1.12)			-10.7062 (-0.98)		
POS(-7,-1)		-5.4093 (-0.63)			2.4565 (0.18)	
TONE(-7,-1)			2.2928 (0.47)			6.5821 (0.82)
AR(-1)	0.0812 (1.09)	0.0791 (1.05)	0.0826 (1.10)	0.0780 (0.77)	0.0788 (0.78)	0.0802 (0.79)
AR(-2)	0.0833 (1.15)	0.0853 (1.18)	0.0835 (1.16)	0.4382 (4.31)***	0.4419 (4.34)***	0.4399 (4.32)***
CAR(-30,-3)	0.0089 (0.91)	0.0091 (0.94)	0.0090 (0.93)	-0.0225 (-1.69)*	-0.0220 (-1.65)*	-0.0222 (-1.66)*
SUE	0.2484 (1.56)	0.2608 (1.64*)	0.2484 (1.56)	0.1489 (0.58)	0.1514 (0.59)	0.1417 (0.56)
log(MV)	-0.0674 (-0.38)	-0.0816 (-0.47)	-0.0918 (-0.53)	-0.2664 (-0.95)	-0.3113 (-1.11)	-0.2976 (-1.07)
log(B/M)	-0.1335 (-0.75)	-0.1421 (-0.80)	-0.1367 (-0.77)	-0.3493 (-1.24)	-0.3564 (-1.26)	-0.3494 (-1.24)
log(TV)	0.0569 (0.42)	0.0659 (0.49)	0.0575 (0.42)	0.1157 (0.53)	0.1187 (0.55)	0.1116 (0.51)
R ²	0.01	0.01	0.01	0.03	0.03	0.03
Obs.	828	828	828	828	828	828

SUE와 CAR에 대한 회귀분석 결과는 큰 차이를 보이는데, 이는 두 변수의 특성 차이에 의해 기인하는 것이라 사료된다. SUE는 기업이익에 대한 변수로, 뉴스의 긍정적 및 부정적 감성이 모두 실적에 대한 유의한 예측력을 가지고 있다. 기업 이익은 갑작스럽게 변동되기 보다는 기업의 꾸준한 영업 활동에 의해서 결정이 되는 특성이 있다. 또한, 기자들은 취재를

통해서 기업의 상황을 어느 정도 파악할 수 있다. 이런 이유 때문에 뉴스 기사가 향후 발표될 이익에 대한 정보를 비교적 정확하게 내포할 수 있다. 따라서, 뉴스의 감성지수(NEG, POS, TONE)가 기업 실적 변수인 SUE와 유의미한 관계를 가지는 것이 이상한 일은 아니다.

반면, CAR는 이익공표일 후의 초과수익률을 나타내는데, 주가는 기업 실적 외에도 금리, 경제 지표, 투자자 심리 등 다양한 외부 요인의 영향을 받는다. 이러한 특성으로 인해 뉴스 텍스트에 내포된 감성이 이익공표 후의 주가 움직임인 CAR를 정확히 예측한다는 것은 쉬운 일이 아니다. 이런 이유 때문에 뉴스 감성과 CAR 사이에는 유의미한 관계가 발견되지 않았을 것이라 판단된다.

2) 뉴스 발표 후의 주식수익률

이 부분의 <표 8>은 일별 감성 지수와 뉴스가 공표된 다음 날의 주식 수익률을 주요 변수로 사용한 회귀분석 결과를 제시한다. 이는 이전의 <표 6>과 <표 7>에서 이익공표일을 기준으로 감성 지수와 누적 초과 수익률(CAR) 간의 관계를 분석한 것과는 구분된다. <표 6>과 <표 7>은 기업의 이익발표일 전에 발행된 뉴스가 실적 공표 후 주가에 미치는 영향을 중심으로 회귀분석을 수행하였다. 반면 <표 8>과 <표 9>는 실적 발표와 무관하게 전체 표본기간의 일별 뉴스를 활용하여 뉴스 감성이 다음 날의 수익률에 미치는 영향을 분석하였다. 실적 발표와 같은 특정한 사건 기간이 아니더라도 일상적인 뉴스의 긍정적 또는 부정적 감성이 향후 주식 수익률과 연관되는지 검증하였다.

<표 8>은 뉴스의 감성변수가 다음 날의 주가 수익률(RET)에 미치는 영향을 분석한 결과를 제시한다. 종속 변수는 뉴스 기사가 공표된 다음 날의 수익률이다. NEG는 -1.3815의 음(-)의 회귀계수를 가지며, t-값이 -14.71로 나타나 통계적으로 유의한 음(-)의 관계를 나타낸다. POS도 1.1729의 양(+)의 회귀계수 (t-값=11.49)로 통계적으로 유의미하다. 긍정적 감성이 다음 날 주가에 대한 예측력이 있음을 시사한다. TONE 역시 1.1485의 양(+)의 회귀계수로 통계적으로 유의미하다.

다음 날의 초과 수익률(RET_EX)을 종속변수로도 분석을 시행하였다. NEG는 -1.2388의 음(-)의 회귀계수를 가지며, t-값은 -14.16으로 유의한 음(-)의 관계를 나타낸다. POS는 0.9717의 양(+)의 회귀계수를 보이며, t-값은 10.22로 나타나 유의한 양(+)의 관계를 가진다. TONE 또한 0.9971의 양(+)이 회귀계수로 통계적으로 유의미하다.

이러한 유의미한 관계는 두 가지로 해석할 수 있다. 첫째, 뉴스 텍스트에 향후 주가를 예측하는 정보가 포함되어 있다는 예측력의 관점으로 이해할 수 있다. 둘째는 뉴스가 다음 날 주가 수익률에 미친다는 영향력으로 해석할 수 있다. 뉴스의 정보에 의해 투자자들이

반응한다면 이들 거래에 의해 향후 주가가 움직일 수 있다.

<표 8> 익일 수익률에 대한 뉴스 기사 감성의 회귀분석

본 표는 일별 뉴스 감성 지수가 주가 수익률에 미치는 영향을 회귀 분석한 결과를 제시한다. 종속 변수로는 뉴스 기사가 공표된 다음 날의 수익률((RET)(%)과 이에서 시장수익률을 차감한 초과수익률(RET_EX)(%)을 사용하였다. 독립 변수로는 부정 비율(NEG), 긍정 비율(POS), 전반적 어조(TONE)가 사용되었다. 또한, 통제변수로 사용된 AR(0), AR(-1), AR(-2)은 이전 날짜들의 초과수익률을 나타내며, CAR(-30,-3)은 이익공표일 전후 30일부터 3일까지의 누적초과수익률이다. 또한, 기업 규모(log(MV)), 시장 가치 대비 장부 가치 비율(log(B/M)), 거래량(log(TV))이 사용되었으며, 이 변수들은 회귀분석의 정확성을 높이기 위해 로그(log)로 전환되었다. R²는 모형의 설명력을 나타내며, 표본 수(Obs.)가 함께 제시된다. 표에서 *, **, ***은 각각 10%, 5%, 1% 수준에서 유의함을 나타낸다.

	RET			RET_EX		
	(1)	(2)	(3)	(1)	(2)	(3)
Intercept	0.2709 (4.75)***	0.1973 (3.45)***	0.2148 (3.76)***	0.0194 (0.37)	-0.0417 (-0.80)	-0.0290 (-0.56)
NEG	-1.3815 (-14.71)***			-1.2388 (-14.16)***		
POS		1.1729 (11.49)***			0.9717 (10.22)***	
TONE			1.1485 (17.58)***			0.9971 (16.39)***
AR(0)	-0.0073 (-5.12)***	-0.0075 (-5.22)***	-0.0074 (-5.21)***	-0.0075 (-5.68)***	-0.0077 (-5.79)***	-0.0076 (-5.78)***
AR(-1)	-0.0203 (-11.86)***	-0.0206 (-12.02)***	-0.0202 (-11.80)***	-0.0217 (-13.68)***	-0.0219 (-13.82)***	-0.0216 (-13.62)***
AR(-2)	-0.0003 (-0.14)	-0.0010 (-0.48)	-0.0006 (-0.29)	-0.0081 (-4.36)***	-0.0086 (-4.68)***	-0.0083 (-4.51)***
CAR(-30,-3)	0.0002 (0.85)	0.0002 (0.80)	0.0002 (0.69)	0.0007 (2.84)***	0.0007 (2.78)***	0.0006 (2.70)***
SUE	-0.0022 (-0.57)	-0.0020 (-0.51)	-0.0024 (-0.63)	0.0116 (3.20)***	0.0118 (3.26)***	0.0114 (3.14)***
log(MV)	-0.0155 (-3.98)***	-0.0202 (-5.18)***	-0.0177 (-4.53)***	-0.0107 (-2.93)***	-0.0147 (-4.05)***	-0.0126 (-3.46)***
log(B/M)	0.0273 (5.87)***	0.0308 (6.63)***	0.0301 (6.47)***	0.0263 (6.07)***	0.0293 (6.76)***	0.0287 (6.64)***
log(TV)	0.0042 (1.29)	0.0035 (1.09)	0.0041 (1.29)	0.0147 (4.92)***	0.0141 (4.72)***	0.0147 (4.91)***
R2	0.01	0.01	0.01	0.01	0.01	0.01
Obs.	496,912	496,912	496,912	496,912	496,912	496,912

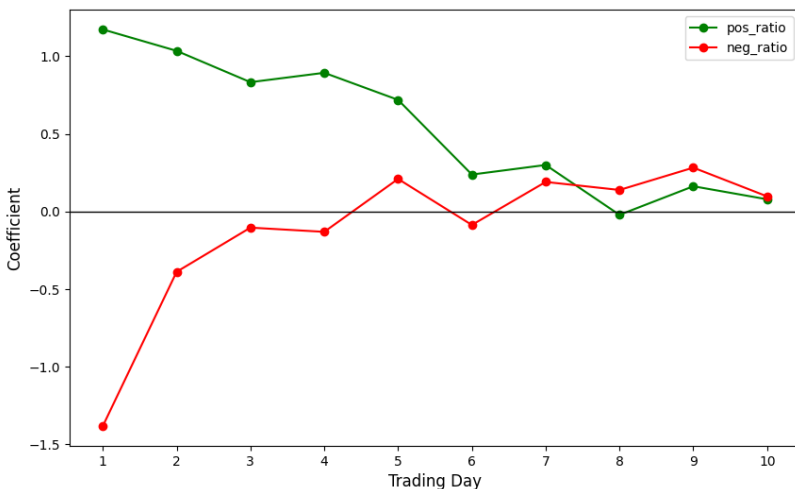
본 연구의 결과는 뉴스 기사의 감성이 향후 주식 수익률과 연관되어 있다는 Tetlock, Tsechansky, and Macskassy(2008)과 Heston and Sinha(2017), Calomiris and Mamaysky (2019)의 결과와도 일치한다. 특히 Heston and Sinha(2017) 결과와 같이, 부정적 뉴스는

다음 날 주식 수익률에 음(-)의 관계를, 긍정적 뉴스는 양(+)의 관계를 가짐을 보여준다. 또한, 부정적 뉴스의 회귀계수가 긍정적 뉴스에 비해 큰 결과를 보이는데, 이는 부정적 뉴스가 주식 수익률에 미치는 영향이 긍정적 뉴스보다 더 강함을 시사한다. 이는 투자자들이 부정적인 정보에 더 민감하게 반응하는 특성을 반영한 결과로 해석할 수 있다.

마지막으로 이익공표일 이후의 CAR를 분석한 <표 6>, <표 7>과 익일 일별 수익률을 사용한 <표 8>의 결과를 비교해보면, <표 8>과 같이 매일의 뉴스에 대해 시장이 즉각적으로 반응하기 때문에, 이익공표일이라는 특정기간에만 유의한 반응을 보일 이유가 없고, 따라서 <표 6>, <표 7>의 유의하지 않은 결과가 나온 것으로 해석할 수 있다. 이는 뉴스를 통해 이익공표일 이전에 이익관련 정보가 유출됨으로써 주가에 이미 반영되기 때문일 수도 있다. 어쨌든 뉴스 기사가 시장 참여자들에게 중요한 정보로 작용할 수 있음을 알 수 있다. 위의 결과들은 일별 감성 지수에 따라 시장이 즉각적으로 반응하고, 그로 인해 주가가 단기적으로 변동될 수 있음을 보여준다. 이익공표일과 같은 특정 사건일이 아니더라도 일별 정보 흐름이 주가 변화에 상당한 영향을 미칠 수 있음을 시사한다.

[그림 3] 뉴스발표 후 수익률에 대한 뉴스 기사 감성의 회귀계수의 추이

본 그림은 뉴스발표 후 10 거래일의 주식 수익률에 대한 긍정 및 부정적 감성의 회귀계수를 보여준다. 가로축은 뉴스발표 후 거래일을 표시하며, 세로축은 회귀계수이다. 초록색 선(pos_ratio)은 긍정적 감성의 회귀계수를, 빨간색 선(neg_ratio)은 부정적 감성에 대한 회귀계수를 나타낸다.



[그림 3]은 뉴스발표 후 10 거래일의 주식 수익률에 대한 긍정 및 부정적 감성의 회귀계수를 보여준다. <표 8>의 회귀분석에서 발견되는 감성의 유의성이 얼마나 지속되는지 확인하기

위해 뉴스발표 후 1거래일부터 10거래일까지의 주식수익률을 종속변수로 사용한 회귀분석을 각각 실행하였다. <표 8>에서 확인할 수 있듯이, 다음 날의 주식수익률에 대한 회귀계수의 규모는 부정적 감성(NEG)이 긍정적 감성(POS)보다 더 크다. NEG의 회귀계수는 2거래일 후의 수익률까지만 의미가 있고, 3거래일 후부터는 거의 0에 가까게 된다. 반면, POS의 회귀계수는 5거래일 수익률까지는 의미있게 지속되다가 이후부터는 0에 접근한다. 부정적인 뉴스는 강하고 빠르게 수익률에 반영되는 반면 긍정적인 뉴스는 약하지만 더 느리게 반영되는 특성이 있는 것으로 보인다.

V. 결 론

본 연구는 한국어 뉴스 기사의 감성이 기업의 비기대이익(SUE)과 이익공표일의 누적초과수익률(CAR)을 예측하는 도구로 활용될 수 있는지를 검증하였다. 분석 결과, 부정적 및 긍정적 감성은 이익발표일의 SUE에 대해 일관된 예측력을 보였다. 이는 뉴스가 기업의 기본적 재무 상태와 관련된 중요한 정보를 전달하며, 시장의 기대와 투자자의 행동에 영향을 미친다는 가설과 일치한다. 하지만, CAR에 대해서는 유의미한 결과를 발견하지 못하였다. 이는 주가가 기업이익에 비해 거시경제 상황, 시장 심리 등 다양한 요인의 영향을 받으며 더 예측하기 어렵기 때문이라 사료된다. 특히, 부정적 뉴스는 긍정적 뉴스에 비해 더 명확하고 강한 예측력을 보여주었다. 한편, 뉴스 기사 감성은 발표 다음 날의 주식 수익률과는 유의미한 관계를 가지고 있었다. 이는 뉴스가 주식 수익률에 대한 정보를 가지고 있을 뿐 아니라, 투자자들의 거래행태에도 영향을 주어 주가에도 영향을 미칠 수 있음을 시사한다.

본 연구는 여러 한계점도 존재한다. 첫째, 표본 기간이 짧다는 점이 있다. 네이버 증권사에서 제공하는 뉴스 기사가 최대 1년 치로 제한되어 있어 데이터 수집에 한계가 있다. 향후 더 표본을 늘려서 장기간 표본을 사용하여 뉴스의 정보력을 분석할 필요가 있다. 또한, 다른 정보매체와의 상호관계로 연구 범위를 넓힐 수도 있다. 기업의 실적발표에 대해서는 애널리스트들도 많은 관심을 가지며 매 분기 보고서를 발간하고 있다. 이들 간의 정보력의 우위나 상호관계를 연구한다면 흥미로울 것이다.

참 고 문 헌

- 김동재, 석상익, 문형빈, “BERT를 활용한 뉴스 기사 감성분석과 블랙-리터만 모형을 결합한 자산 배분 전략 제안”, 재무관리연구, 제40권 제5호, 2023, 155-180.
- 김명진, 류지혜, 차동호, 심민규, “SNS 감성 분석을 이용한 주가 방향성 예측”, 한국전자거래학회지, 제25권 제4호, 2020, 61-75.
- 김용석, 조성욱, “한국어 텍스트 분석과 적용: 머신러닝을 통한 증권발행신고서의 비정형화된 텍스트 분석: 머신러닝을 통한 증권발행신고서의 비정형화된 텍스트 분석”, 한국증권학회지, 제48권 제2호, 2019, 215-35.
- 서한백, 김은총, 최상문, 정재만, “BERT 어조에 기반한 애널리스트 보고서 텍스트 정보의 유용성”, 재무관리연구, 제41권 제5호, 2024, 47-84.
- 양철원, “애널리스트 보고서 제목의 정보력 검증: 텍스트 어조를 중심으로”, 재무관리연구, 제38권 제3호, 2021, 1-38.
- 이주화, 류두진, “뉴스/소셜 미디어 텍스트와 투자자 기대”, 경영학연구, 제50권 제2호, 2021, 533-555.
- 장은아, 최희련, 이홍철, “BERT를 활용한 뉴스 감성분석과 거시경제지표 조합을 이용한 주가지수 예측”, 한국컴퓨터정보학회논문지, 제25권 제5호, 2020, 47-56.
- 조수지, 김홍규, 양철원, “기업 재무분석을 위한 한국어 감성사전 구축”, 한국증권학회지, 제50권 제2호, 2021, 135-170.
- 최남현, 김강산, 장희수, “뉴스와 댓글 텍스트를 활용한 공모주 상장일 시초가 형성 예측 및 영향성 분석: XGBoost 기반”, 재무관리연구, 제40권 제1호, 2023, 81-104.
- Antweiler, W. and M. Z. Frank, “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards,” *Journal of Finance*, 59(3), (2004), 1259-1294.
- Bernard, V. L. and J. K. Thomas, “Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?,” *Journal of Accounting Research*, 27, (1989), 1-36.
- Bollen, J., H. Mao and X. Zeng, “Twitter Mood Predicts the Stock Market,” *Journal of Computational Science*, 2(1), (2011), 1-8.
- Buehlmaier, M. M. and T. M. Whited, “Are Financial Constraints Priced? Evidence from Textual Analysis,” *The Review of Financial Studies*, 31(7), (2018), 2693-2728.
- Calomiris, C. W. and H. Mamaysky, “How news and its context drive risk and returns around the world,” *Journal of Financial Economics*, 133(2), (2019), 299-336.

- Cho, S. J., K. K. Lee, and C. W. Yang, "Does the textual tone of analyst reports have valuable information? Korean evidence," *Asia-Pacific Journal of Financial Studies*, 53(3), (2024), 349-389.
- Das S. and M. Chen, "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, 53(9), (2007), 1375-1388.
- Fang, L. H., J. Peress, and L. Zheng, "Does media coverage of stocks affect mutual funds' trading and performance?," *Review of Financial Studies*, 27(12), (2014), 3441-3466.
- Fang, L. and J. Peress, "Media coverage and the cross-section of stock returns," *Journal of Finance*, 64(5), (2009), 2023-2052.
- Gentzkow, M., B. Kelly, and M. Taddy, "Text as data," *Journal of Economic Literature*, 57(3), (2019), 535-574.
- Heston, S. L. and N. R. Sinha, "News vs. sentiment: Predicting stock returns from news stories," *Financial Analysts Journal*, 73(3), (2017), 67-83.
- Kearney, C. and S. Liu, "Textual sentiment in finance: A survey of methods and models," *International Review of Financial Analysis*, 33, (2014), 171-185.
- Li, F., "The Information Content of Forward-Looking Statements in Corporate Filings-A Naïve Bayesian Machine Learning Approach," *Journal of Accounting Research*, 48(5), (2010), 1049-1102.
- Li, F., R. Lundholm, and M. Minnis, "A Measure of Competition Based on 10-K Filings," *Journal of Accounting Research*, 51(2), (2013), 399-436.
- Li, F., "Textual analysis of corporate disclosures: A survey of the literature," *Journal of Accounting Literature*, 29(1), (2010), 143-165.
- Loughran, T. and B. McDonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *Journal of Finance*, 66(1), (2011), 35-65.
- Loughran, T. and B. McDonald, "Textual analysis in accounting and finance: A survey," *Journal of Accounting Research*, 54(4), (2016), 1187-1230.
- Tetlock, P. C., "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *Journal of Finance*, 62(3), (2007), 1139-1168.
- Tetlock, P. C., M. S. Tsechansky, and S. Macskassy, "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *Journal of Finance*, 63(3), (2008), 1437-1467.

THE KOREAN JOURNAL OF FINANCIAL MANAGEMENT
Volume 41, Number 6, December 2024

Forecasting Corporate Earnings and Stock Returns using Sentiment in News Articles*

Bomi Choi** · Akihiko Yamada*** · Cheol-Won Yang****

〈Abstract〉

We investigate whether the text of news articles about companies can predict future corporate earnings and stock returns. We collect news articles of companies included in the KOSPI 200 index and calculate the sentiment variable of news articles by measuring the number of positive and negative words in the text using a sentiment dictionary of KOSELF(Korean Sentiment Lexicon for Finance). The sentiment of news articles prior to the earnings announcement date is statistically significant in predicting corporate earnings, with negative sentiment having stronger predictability than positive sentiment. This suggests that news articles contain meaningful information about a firm's future earnings. We further examine the relationship between news sentiment and future stock returns. News sentiment does not have predictive power for the market's reaction to earnings announcements, but it is significantly associated with the next day's stock returns. These results show that the text of news articles contains meaningful information about a firm's future earnings and stock returns, suggesting that news information influences market participants.

Keywords : News Articles, Sentiment Analysis, Earnings, Stock Return, Predictability

* We express our gratitude to the participants in annual meetings of the Korean Financial Management Association for providing valuable feedback on this paper. Our research team was supported through the Research-Focused Department Promotion & Interdisciplinary Convergence Research Project as a part of the Support Program for University Development for Dankook University in 2024.

** First Author, Graduate Student, School of Business Administration, Dankook University.
E-mail: 20infinite02@naver.com

*** Co-Author, Assistant Professor. Faculty of Liberal Education, Seoul National University.
E-mail: ayamada5413@snu.ac.kr

**** Corresponding Author, Professor of Finance, School of Business Administration, Dankook University,
E-mail: yang@dankook.ac.kr