뉴스 빅데이터를 이용한 해운뉴스지수 개발†

Developing the Maritime News Index using News Big Data Analysis

박성화* · 권장한** · 김태일***
Park, Sung Hwa · Kwon, Jan Han · Kim, Tae Il

목 차

Ⅰ. 서 론
Ⅱ. 선행연구 및 본 연구의 차별성
Ⅲ. 해운뉴스지수 작성 및 유용성 검정
Ⅳ. 결 론

〈Abstract〉

This paper uses news articles to measure credit risk in Korea's shipping industry. We develop the Maritime News Index(MNI) which reflects the frequency of news articles in 12 newspapers that contain the terms related to 6C's of credit. The dynamic of this index displays spikes prior to or during elevated risk periods. We find that the MNI is significantly correlated with the maritime business survey index (BSI). The results from the forecasting regressions indicate that the MNI is a strong predictor of the growth in freight transportation. Evidence suggests that the index may provide useful information for forecasting market situations.

Key Words: Maritime News Index, Big Data Analysis, Recession Sentiment Index, Maritime Business Survey Index, Forecasting Regression

† 본 연구는 한국해양수산개발원의 연구개발적립금사업인 「재무정보를 활용한 해운기업의 부실예측 연구」의 일부 내용을 발췌하여 재작성함. 해운뉴스지수키워드 감수에 도움을 주신 MBN 엄해림 기자에게 감사드린다.
* 제1저자, 한국해양수산개발원 전문연구원, shpark83@kmi.re.kr
** 제2저자, 한양대학교 대학원 경제금융학과 박사과정, plugkwon@hanayang.ac.kr
*** 교신저자, 한국해양수산개발원 연구위원, ktizorro@kmi.re.kr

Ⅰ. 서 론

뉴스 빅데이터 분석은 언론의 정보전달자 역할을 강조한다.[1]언론의 기능 중 여론을 공공의 관심사로 전환시켜 의사 결정에 영향을 미치는 여론 감독(public opinion supervision) 기능을 분석의 기본 전제로 한다.[2] 언론사 보도는 객관성과 공정성을 기본가치로 하는 보도 규범이 존재하기 때문에 유사한 사건이 발생하였을 때 기자들이 공통적으로 사용하는 어휘가 존재하고,[3] 뉴스기사는 시장성과, 유관기관 전문가의 의견, 정책 논의 등 다양한 정보원을 기초로 여론을 형성하므로[4] 뉴스 기사에서 특정 단어의 반복적 사용은 관심 사건에 대한 강도를 측정하는 수단이 될 수 있다.

해운산업의 리스크 측정에는 어려움이 존재한다. 개별기업의 리스크를 감시하고 평가하기 위해 대차대조표나 손익계산서 등의 재무지표를 사용하는 방법은 분석이 용이하고 결과가 명확하다는 장점이 있지만 과거 실적을 바탕으로 미래를 평가하는 것이기 때문에 급변하는 환경과 새로운 정보에 대한 대응에는 취약하다는 단점이 있다. 재무지표 분석방법의 단점을 보완하기 위해 시장정보를 사용하여 개별기업 리스크를 측정하고 기업의 현 상태뿐만 아니라 미래 리스크에 대한 전망을 반영한 위험예측에 대한 시도가 이루어지고 있다.[5] 시장정보는 재무지표에 비해 시계열이 연속적이고 빈도가 잦다는 장점이 있다.[6] 하지만 해운시장의 경우 전체 해운산업에서 상장회사의 수가 적고 규모의 불균형이 존재하기 때문에 시장정보를 활용하는 데에 한계가 존재한다. 또한 주요 시장성과에 대한 집계자료의 시계열이 짧아 계량분석에 용이하지 않은 경우가 대부분이다. 대표적인 해운시황 척도인 발틱 운임지수, 물동량과 같은 실물경제변수는 해운기업들의 영업활동에 관한 정보를 지니고 있지만, 해운기업의 신용위

1) Alexopoulos and Cohen (2009), p.4.
2) 손태규(2013), p.111.
3) 이완수·박재영(2014), p.10.
4) Bontempi, Golinelli, and Squadrani (2013), p.3.
5) 예금보험공사(2015), p.1.
6) Burton and Seale(2005), pp.93-94.

험은 반영하지 못하고 가격이 결정되는 시장의 대상이 세계 해운시장이기 때문에 우리나라 해운 산업의 고유 리스크에 대한 설명력은 낮을 것이라 판단할 수 있다. 또한 신용평가사나 금융기관의 신용등급자료 역시 접근에 제약이 있어 활용에 어려움이 따른다.[7)]

이와 같은 자료의 한계 외에도 해운산업 고유의 영업활동 역시 해운산업 신용위험 측정에 어려움을 주는 요인이다. 해운기업의 주요 영업활동은 선박 운용을 통한 해상운송이다. 해상운송은 세계무역운송의 약 90%를 차지할 정도로 세계운송에서 비중이 크고 세계 경기와 밀접한 관련이 있다. 실제로 2008년 글로벌 금융위기로 시작된 세계 경제의 침체로 인해 발틱운임지수는 약 93% 하락 하였고 운임의 급락은 해운기업의 주요 자산인 선박가치의 하락으로 이어졌다.[8)] 이러한 해운산업의 특성을 고려한다면 해운기업의 위험은 기업 고유의 리스크 외에도 세계시황, 환율, 유가, 이자율 등과 같은 시장여건에도 영향을 받는다고 할 수 있을 것 이다.[9)] 한편, 해운업은 기업 간에도 밀접한 관련이 있는데 주요 해운사들은 전략적 제휴를 통해 항로와 일정을 공유하여 운송비를 최소화하고 있고 국내 해운기업들은 경영 여건에 차이가 있음에도 불구하고 선도기업의 대응전략을 추종하는 경향이 있다.[10)]

이와 같은 해운산업 신용위험 측정의 어려움을 보완하기 위해서 본 연구는 뉴스 빅데이터 분석방법을 사용하여 우리나라 해운산업의 신용위험을 측정하고자 하였다. 뉴스 빅데이터 분석이 지닌 장점은 확장성과 경제성이다. 뉴스 빅데이터 분석을 통한 신용위험 측정은 재무지표, 시장자료와 같은 집계자료에 비해 지수 작성에 대한 경제적 비용이 낮고 키워드 변경을 통한 지수 개선이 다른 지표에 비해 용이하다.

해운업의 세계경제 및 기업 간 연관성을 고려하고 해운산업 고유의 영업활동을 반영하기 위해 해운뉴스지수 작성에 사용된 키워드는 Grammenos(2010)가

7) 박성화·김태일·권장한(2017), p.829.
8) Albertijn, Bessler and Drobetz (2011), p.70.
9) 박성화·김태일·권장한(2017), p.824, Marsy, Olugbode and Pointon (2010), p.454.
10) 나정호·김광원·방희석 (2015), p.187.

제안한 신용평가의 6C(6C's of credit)에서 이론적 근거를 두었다. Grammenos (2010)는 해운기업의 위험평가 시 고려해야 할 신용평가의 6C(6C's of credit)를 제안하였다. 신용평가의 6C는 해운기업의 신용평가에서 고려해야할 재무적 요인인 자본구조(Capital) 외에도 비재무적 요인인 경영능력(Character and Capacity)과 기업구조(Company), 세계 시장과의 밀접한 연관성을 나타내는 시장조건(Conditions), 해운기업의 주요담보인 선박에 관한 담보능력(Collateral)을 고려하여 신용평가를 해야 한다고 하였다. 해운뉴스지수에 사용된 키워드는 신용평가 6C의 5개 영역에 부합하는 26개의 키워드를 포함하는 신문기사로 구성되었다.

해운뉴스지수에 사용된 언론기사는 한국언론진흥재단의 빅카인즈(Big Kinds)의 자료를 사용하였다. 빅카인즈는 개방형 뉴스빅데이터 분석시스템으로 종합일간지 및 경제지, 지역일간지, 방송사 총 42개의 언론사의 기사를 제공한다.[11)]

빅카인즈에서 제공하는 언론사 중 종합일간지 및 경제지 12개사를 중심으로 1991년 1월부터 2018년 3월까지 신문기사 18,909건을 사용하였다. 지수 작성에 사용된 신문기사는 신용평가의 6C의 영역별 위험이 증가하였을 때 사용빈도가 높을 것이라 예상되는 특정 단어가 포함된 신문기사로 구성하였다.

본 연구의 목적은 우리나라 해운산업 전반의 신용위험을 측정하는 해운뉴스지수를 개발·소개하고, 작성된 지수를 바탕으로 해운시장 전반의 위험을 관리·감독하는 데 있다. 그동안 반복되어 왔던 해운업의 위기에도 불구하고 정부는 사후적인 정책으로 대응해왔다. 해운산업이 우리나라 국가경제에 기여하는 부분을 고려한다면 해운산업 전반의 위험을 사전에 감지하고 관리하여 위기를 미연에 방지하는 선제적인 정책 대응이 요구된다 할 수 있다. 본 연구에서 고안된 해운뉴스지수는 해운산업의 신용위험을 나타내는 지표로 해운시장의 신용위험을 감지하여 정책당국은 물론 기업의 효율적 의사결정에 유용한 정보를 제공할 것으로 기대된다.

11) 빅카인즈에 대한 자세한 내용은 최윤희(2016)를 참고하기 바란다.

본 연구의 구성은 다음과 같다. 제2장에서 해운기업의 위험과 뉴스를 이용하여 경제데이터를 분석한 선행연구를 소개하고 본 연구의 차별성을 제시한다. 제3장에서는 해운뉴스지수의 작성방법에 대해 논의하고 예측회귀분석을 통해 해운뉴스지수의 유용성을 점검한다. 마지막으로 제4장에서는 결론과 시사점을 제시한다.

Ⅱ. 선행연구 및 본 연구의 차별성

1. 선행연구

해운산업의 위험에 관한 연구는 위기예측모형을 활용하여 해운기업의 위험에 영향을 주는 변수를 탐색하고 해운산업 위기를 분석한 연구[12]와 위험관리 측면에서 해운기업의 위험관리 요인을 분석하고 효율적인 위험관리 방안을 모색하는 연구[13]로 분류할 수 있다.

위기예측모형의 대표적인 방법인 신호접근법은 특정 위기에 대해 설명력이 높은 변수를 선행지표로 설정하여 선행지표가 임계치(threshold)를 넘어설 경우 위기신호로 판별하는 비모수적 방법이다. 이재규·김일수(2017)는 신호접근법을 활용하여 우리나라 해운시장의 위기를 예측하고 해운시장 위기에 유의미한 설명력을을 지닌 변수를 찾고자 하였다. 해운시장 위기를 측정하기 위한 변수로 자국통화의 절하률과 선박연료유 가격증가율의 가중치로 표시된 해운시장 압력지수를 산정하여 해운시장 압력지수가 일정 임계치를 초과할 때 해운시장 위기가 발생한 것으로 간주하였다. 연구결과, 컨테이너처리실적, 부산연료유 가격변동이 해운시장 위기를 예측하는 선행지표임을 확인할 수 있었다.

---

12) 박성화·김태일·권장한(2017), 이재규·김일수(2016), Kuvussanos and Tsouknidis, (2016), Lozinskaia, Merikas, Merika, and Penikas (2017)

13) 나정호·김광원·방희석 (2015), 김재봉(2000), Albertijn, Bessler and Drobetz (2011), El-Masry, Olugbode and Pointon (2010)

신호접근법이 변수간 상관관계를 고려하지 않는 비모수적 방법인데 반해 이원선택형 모형(bianry choice models)은 변수간의 상관관계를 고려한 모수적 방법의 위기예측모형이다. 이원선택형 모형은 종속변수가 위기일 경우 1, 건전일 경우 0의 값을 지니는 이항 변수의 형태를 지닌다. 연구에 따라 위기판별의 기준이 상이한데, 박성화 외(2017)는 한국선주협회의 탈퇴여부와 법정관리 및 부도 여부를 위기로 판별하여 우리나라 해운기업의 부실에 관한 연구를 하였고 금융비용관련 재무지표가 해운기업의 부실에 유의한 설명력이 있음을 밝혔다. Kuvussanos and Tsouknidis(2016)는 해운기업 부실채권 결정요인에 관한 연구를 수행하였다. 해운기업의 채무불이행에 관한 설명변수로 Grammenos(2010)가 제시한 신용평가의 6C에 의거하여 변수를 설정하였고 해운기업의 채무불이행은 해운시황과 선박가치에 의해 결정된다고 주장하였다. Lozinskaia et al., (2017)은 해운기업의 부실예측에 관한 연구에서 해운기업의 부실에 관한 연구는 재무적 요인과 비재무적 요인이 모두 고려되어야 한다고 주장하였다.

해운기업의 위험관리에 관한 연구로는 나정호·김광원·방희석(2015)이 있다. 나정호 외(2015)는 해운기업의 위험은 대표적으로 시장 환경의 예측하지 못한 변화로 인한 시장위험, 기업의 운영과정에서 발생할 수 있는 운영위험, 현금흐름과 관련된 재무위험으로 분류할 수 있다고 하였다. 이러한 위험의 분류를 바탕으로 해운산업 종사자들을 대상으로 한 계층분석적 의사결정과정(analytic hierarchy process, AHP)을 수행하였다. 연구결과, 운임, 선가, 유가위험을 포함하는 시장위험이 다른 요소보다 중요하게 관리되어야 하는 것으로 나타났다. 효과적인 위험관리를 위해선 정부가 중심이 된 맞춤형 시황정보 제공이 필요하다고 하였다.

2. 본 연구의 차별성

해운산업 위험에 관한 연구는 해운기업의 위험이 재무적 요인뿐만 아니라 비재무적 요인에 큰 영향을 받고 거시경제여건과 같은 시장위험에 더 취약할

수 있음을 보여 준다. 이러한 시장위험은 개별기업의 관리 범위를 넘어서기 때문에 보다 효과적인 위험관리를 위해선 정부가 중심이 된 맞춤형 시황정보 제공의 필요성을 나정호 외(2015)는 주장하였다. 본 연구에서 고안된 해운뉴스지수는 기존 연구에서 해운기업 위험의 주요 요인으로 지목된 항목들을 고려한 키워드를 중심으로 작성되었다. 따라서 본 지수는 우리나라 해운산업 전반의 신용위험을 측정하여 정보를 제공하므로 시장 참여자들의 효율적인 의사결정에 도움을 줄 것 이다.

우리나라 해운산업 시황을 판단하고 전망하기 위해선 해운산업 경기에 대한 대표 변수가 필요하다고 할 수 있다. 이러한 필요성에 의거하여 한국해양수산개발원은 매월 해운업 경기실사지수를 발표하고 있다. 해운업 경기실사지수는 해운업 참여자들에게 현재 경기수준과 향후 전망 등을 설문조사하여 전반적인 해운경기 동향과 전망을 파악하는 경기 설문지표(business survey index)이다.[14] 해운업 경기실사지수는 실제 해운업 종사자들의 전망과 체감 경기를 판단하는 데 용이하지만 비교적 최근부터 공표되고 있어 시계열 분석에 용이하지 못하다는 한계가 있다.[15]

뉴스 빅데이터 분석을 활용한 지수 작성의 장점은 확장성과 경제성이다. 해운업 경기실사지수는 설문에 대한 시장 참여자의 응답으로 작성된다. 따라서 설문지 질문의 구성이 변경되지 않는 한 설문 외의 내용으로 지수를 작성하는 것은 현실적으로 불가능하다할 수 있다. 반면에 뉴스 빅데이터 분석방법은 키워드 구성을 변경함으로써 지수의 설명력을 개선하고 필요한 경우 키워드를 새로 구성하여 새로운 시황 지수를 개발하는 것에 대한 경제적 비용이 비교적 적은 장점이 있다.

뉴스 빅데이터 분석방법은 시공간에 구애받지 않고 장기 시계열 자료를 구축할 수 있다. 해운업 경기실사지수와 같은 시장 참여자의 설문자료는 설문과 집계가 이루어지지 않은 시점의 자료는 생성이 불가능하지만 뉴스 빅데이터 분석를 통해 작성된 해운뉴스지수는 뉴스 아카이브가 존재하는 시점부터 데이터

14) 한국해양수산개발원 2016년 11월 30일 보도자료.
15) 해운업 경기실사지수는 2013년 8월부터 발표되고 있다.

를 생성할 수 있어 장기 시계열 구축이 가능하다는 장점이 있다. 이러한 뉴스 빅데이터 분석의 장점을 바탕으로 본 연구에서 고안된 해운뉴스지수는 현재 우리나라 해운산업의 신용위험의 증감이 어떤 요인에 의한 것인지 분석할 수 있게 해주고, 더 나아가 발생 가능한 위험을 감지하고 선제적인 대응을 하는 데에도 유용한 정보를 제공할 수 있을 것이다.

뉴스를 이용하여 경제현상을 분석한 연구는 뉴스 기사를 활용하여 주식시장, 외환시장, 경기변동 등 다양한 시장을 분석하였다.[16] 해운뉴스지수는 뉴스 빅데이터 분석을 활용하여 해운산업의 위험을 감지하고 우리나라 해운시장에 유용한 정보를 제공하여 기업의 의사결정과 정부의 정책결정에 도움을 주고자 작성되었다.

Ⅲ. 해운뉴스지수 작성 및 유용성 검정

1. 해운뉴스지수 작성 개요 및 자료설명

본 장에서는 해운뉴스지수에 사용된 키워드와 작성방법, 작성된 해운뉴스지수를 다룬다. 뉴스빅데이터 관련 변수 작성의 어려움은 기사에 담긴 정보가 호재인지 악재인지 판단하는 것이다.[17] 특히 한국어의 경우 문장 구조상 역접의 관계가 많기 때문에 단순 키워드 검색만으로 호재 및 악재의 판단이 어렵다. 이러한 문제점을 인식하여 해운뉴스지수 작성된 사용된 키워드는 전문 언론인의 감수를 받아 문장에서 역접의 구조를 지니기 어려운 단어로 구성하였고 필요한 경우 직접 기사를 읽어 해운산업 위험의 증가에 대한 내용인지를 판단하였다.

해운산업 위험관련 뉴스는 한국언론진흥재단에서 제공하는 뉴스아카이브인

16) 이긍희 (2017), 이완수·박양수(2016), Alexopoulos and Cohen (2015), Baker, Bloom, and Davis (2016).
17) 강선아·김유신·최상현(2015), p.1169.

빅카인즈에서 추출하였다. 빅카인즈에서 제공하는 42개 언론사 중 종합일간지 및 경제지에 해당하는 12개 언론사의 1991년 1월부터 2018년 3월까지의 신문기사 수를 사용하여 지수를 작성하였다. 해운뉴스지수에 사용된 언론사와 수록기간은 <표 1>과 같다.[18)]

▌표-1. 지수작성에 사용된 언론사 ▐

신문사	수록기간
경향신문	1991년 1월 - 2018년 3월
국민일보	1991년 1월 - 2018년 3월
매일경제	1995년 1월 - 2018년 3월
문화일보	1996년 12월 - 2018년 3월
서울경제	1996년 10월 - 2018년 3월
서울신문	1991년 1월 - 2018년 3월
세계일보	1991년 1월 - 2018년 3월
파이낸셜뉴스	2001년 11월 - 2018년 3월
한겨레	1991년 1월 - 2018년 3월
한국경제	1995년 1월 - 2018년 3월
한국일보	1991년 1월 - 2018년 3월
헤럴드경제	2000년 1월 - 2018년 3월

2. 신용평가의 6C (6C's of Credit)[19)]

뉴스빅데이터를 활용하여 해운산업의 위험을 측정하기 위해 사용된 키워드는 Grammenos(2010)가 제안한 신용평가의 6C(6C's of Credit)를 이론적 근거로 설정했다. 해운사의 영업활동은 크게 두 가지로 분류할 수 있는데 하나는 보유한 선박을 운영하여 수익을 내는 영업활동이고 다른 하나는 선박의 구입과 매각을 통한 수익이다. 따라서 해운기업의 신용위험은 해운기업의 주요 담보인 선박가치의 변동이 주요인이 될 것이다. Grammenos (2010)는 전통적인 위험평

18) 빅카인즈에서 제공하는 종합일간지 및 경제지 중 내일신문의 경우 자료누락기간이 존재하여 분석에서 제외하였다.

19) 본 절은 Grammenos(2010)를 참고하여 작성하였다.

가방법 외에도 이러한 해운산업의 특성에 기인하여 해운기업의 신용평가에 고려해야 할 5개 영역의 6C로 경영능력(Character/Capacity), 자본구조(Capital), 기업구조(Company), 세계시황(Conditions), 담보능력(Collateral)을 제시하였다.

경영능력은 해당기업의 장기적인 계획과 전략을 평가한다. 경영능력 평가 시 고려되어야 할 부분으로 투자, 금융, 용선, 위험관리, 인적자원을 강조한다. 다른 경쟁기업에 비해 저비용 고수익의 경영효율성을 달성 여부를 중요시한다. 자본구조는 해운산업의 자본집약적 특성을 반영한다. 선주와 경영자 지분의 적정성, 재무구조, 재무제표 분석을 통해 기업의 신용위험을 평가한다. 기업구조는 투입과 산출에 의거하여 신용평가를 한다. 투입은 고용형태, 용선선주와의 계약, 용선선주 선택 등을 다루고 산출은 보험, 관리와 같은 운영비용을 평가한다. 시장점유율 역시 주요 고려대상이 된다. 시장조건은 종합적인 세계경제, 해운시장, 금융 및 정치 상황을 다룬다. 금융시장에서의 이자율과 무역에 대한 정부의 정책방향 역시 주요 고려대상이 된다. 세계경제는 세계 산업생산, 해상물동량 등을 평가하고 해운시장 평가항목으로는 수급분석 및 전망, 규제 등이 있다. 마지막으로 담보능력은 해운사의 주요자산인 선대의 구성 및 상태와 선박의 수명으로 구성되며 선박 외의 유가증권 역시 포함된다.

▌표-2. 신용평가의 6C에 의거한 해운뉴스지수 작성에 사용된 키워드 ▐

영역	주요 내용	키워드
경영능력 (Character/Capacity)	기업의 목표 및 장기전략	리스크, 둔화, 마이너스, 악화, 실패, 역대최저, 불경기, 적자, 부진, 손실
자본구조 (Capital)	재무구조	채권단, 법정관리, 부도, 부실, 매각, 불확실성
기업구조 (Comapany)	투입구조 및 산출구조	해고, 감축, 구조조정
시장조건 (Conditions)	금융시장, 세계경제, 해운시장	먹구름, 암울, 하락, 악재, 공급과잉
담보능력 (Collateral)	선대 및 유가증권	폐선, 가뭄

출처: Grammenos(2010), p.781, 김태일(2014), p.135.

이러한 신용평가의 6C의 5개 영역에 근거하여 영역별 위험이 증가했을 시에 사용될 수 있는 26개의 키워드를 선별하였다. 신용평가의 6C의 주요내용과 해운뉴스지수에 사용된 키워드는 <표 2>에 수록하였다.

3. 해운뉴스지수 산출방법

Baker, Bloom and Davis(BBD, 2016)는 거시경제 전반의 정책불확실성을 측정하기 위해 Economic Policy Uncertainty (EPU) 인덱스를 고안하였다. EPU 인덱스는 정책불확실성을 나타내는 뉴스기사의 빈도를 기반으로 작성되는데 해운뉴스지수 역시 BBD (2016)의 방법에 의거하여 작성되었다. 해운뉴스지수의 작성방법은 다음과 같다:

첫째, 2절에서 소개된 신용평가의 6C에 근거한 키워드를 중심으로 신문사별 월별 경제 기사 수를 수집한다. 이렇게 수집된 신용위험을 반영하는 i신문사의 t시점 기사 수를 $X_{i,t}$라 하면 $X_{i,t}$를 t시점의 i신문사 전체 경제관련 기사 수 $Total_{i,t}$로 나누어 비율화 된 신문기사 수의 시계열 자료 $Y_{i,t}$를 생성한다.[20]

$$Y_{i,t} = \frac{X_{i,t}}{Total_{i,t}} \tag{1}$$

둘째, 전체 표본기간을 기준으로 $Y_{i,t}$의 신문사별 분산 σ_i^2을 구한 다음, 식 (1)에서 생성된 $Y_{i,t}$를 표준편차로 나누어주는 표준화과정을 거친다. 이 과정을 거쳐 표준화된 시계열 자료 $Z_{i,t}$를 얻게 된다.

20) 수집된 기사가 중복되는 경우, 중복된 기사를 삭제하여 하나의 자료만 지수 작성에 이용하였다. 기사 수는 신문사의 규모에 따라 편차가 존재하므로 전체 경제 기사수로 나누어 단위를 통일하고자 하였다.

$$Z_{i,t} = \frac{Y_{i,t}}{\sigma_i} \tag{2}$$

셋째, 표준화된 $Z_{i,t}$의 월간 신문사 전체 평균 μ_t를 구한다.

$$\mu_t = \frac{1}{n}\sum_{i=1}^{n} Z_{i,t} \tag{3}$$

넷째, μ_t의 전체 표본기간 평균 M을 구한 후, 전체 평균을 100으로 변환하여 해운뉴스지수 MNI_t를 작성한다.

$$MNI_t = \mu_t \times \frac{100}{M} \tag{4}$$

식 (4)를 통해 작성된 해운뉴스지수는 평균이 100인 시계열 자료가 된다. 따라서, 100을 기준으로 해운뉴스지수가 100보다 높은 경우 시장의 신용위험이 증가된 상황으로, 100보다 낮은 경우 시장의 신용위험이 안정적인 것으로 해석할 수 있을 것이다.

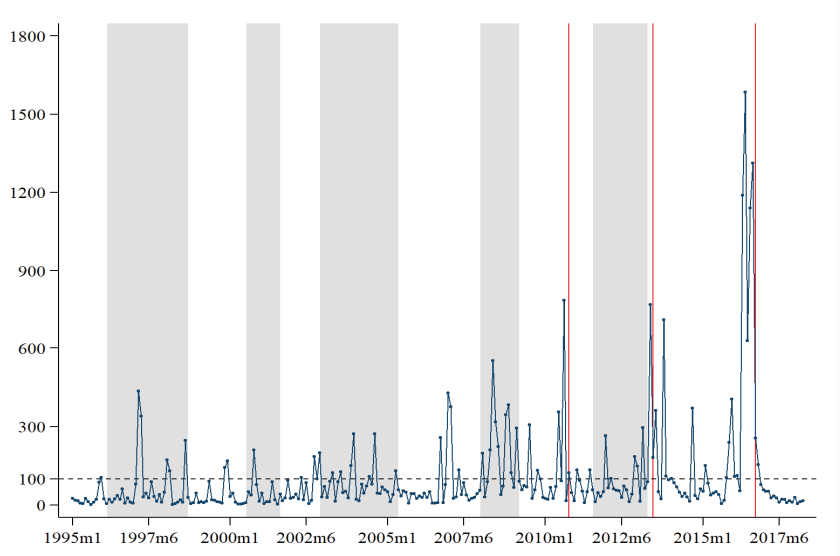
3. 해운뉴스지수와 해운업 경기실사지수 간의 교차 상관관계검정

<그림 1>은 1995년 1월부터 2018년 3월까지의 해운뉴스지수를 나타낸다. 해운뉴스지수는 1991년 1월부터 2018년 3월까지 우리나라 종합일간지 및 경제지 12개 언론사의 경제기사 중 해운기업의 위험을 담고 있는 18,909개의 신문기사를 활용하여 작성되었다. <그림 1>에서 검은 실선은 본 연구에서 고안된 해운 뉴스지수이며 회색 음영부분은 통계청의 경기순환 국면 중 수축기를 의미한다. 붉은 선은 우리나라 주요 해운사의 법정관리 신청시점이다.[21]

21) 대한해운 법정관리(2010년 10월), 팬오션 법정관리(2013년 6월), 한진해운 법정관리(2016년 9월) .

해운뉴스지수는 전반적으로 100 미만의 안정적인 모습을 보이지만 경기 수축기간에 상승하는 패턴을 보여 해운뉴스지수가 경기역행적(countercyclical)인 성격을 지님을 알 수 있다. 해운뉴스지수는 경기 수축기간 외에도 우리나라 해운업을 대표하는 해운사들의 법정관리 신청기간 이전에 큰 폭으로 상승하는 모습을 보이는 것을 확인할 수 있다. 해운산업 위기의 원인이 거시경제적인 요인이었던 1998년 외환위기와 2018년 글로벌 금융위기에는 해운뉴스지수가 위기에 후행하는 패턴을 보이지만 주요 해운사 법정관리 시점에는 해운뉴스지수가 법정관리에 선행하는 패턴을 보여 본 연구에서 고안된 해운뉴스지수가 우리나라 해운산업의 위험을 진단하고 예측하는 데 유용한 정보를 지니고 있음을 확인할 수 있다.

그림-1. 해운뉴스지수



주: 분석기간 1991년 1월부터 2018년 3월까지 우리나라 종합 일간지 및 경제지 12개 언론사의 경제기사 중 해운기업 신용위험을 담고 있는 기사 수로 작성된 해운뉴스지수. 회색 음영은 통계청의 경기순환국면 중 수축기를 나타냄. 붉은 수직선은 우리나라 주요 해운사의 법정관리 신청 시점을 의미함.

해운뉴스지수가 우리나라 해운업 체감 경기 및 전망과 높은 상관성을 지닐 것으로 판단되기 때문에 해운경기를 대표하는 해운업 경기실사지수와 교차 상관관계(cross-correlation) 검정을 하였다. 해운뉴스지수는 해운산업 전반의 위험을 측정하는 지수이므로 종합 해운업 경기실사지수[22]와 교차 상관관계 검정을 진행하였다.

<표 3>은 해운뉴스지수와 시차별 해운업 경기실사지수 간 교차 상관관계 검정을 나타낸다. 해운뉴스지수와 해운업 경기실사지수는 시차와 무관하게 모두 음(-)의 상관관계를 보이는 것을 확인할 수 있다. 해운뉴스지수는 모든 해운업 경기실사지수의 3개월 이전과 통계적으로 유의한 상관관계를 지니며 상관계수 역시 높은 것으로 나타났다. 특히 동향과 전망 모든 항목에서 업황, 매출과 높은 상관관계를 보여 해운뉴스지수가 해운산업 전반의 시장상황과 기업의 매출부분과 높은 상관관계가 있음을 확인할 수 있었다.

▌표-3. 해운뉴스지수와 해운업 경기사실지수 간의 교차 상관관계 검정 ▐

시차 (h)	동향				전망			
	업황	채산성	자금사정	매출	업황	채산성	자금사정	매출
-3	-0.33**	-0.37***	-0.35***	-0.38***	-0.38**	-0.32**	-0.33**	-0.40***
-2	-0.33**	-0.24*	-0.17	-0.45***	-0.37***	-0.27*	-0.26*	-0.47***
-1	-0.22	-0.03	-0.03	-0.43***	-0.28**	-0.23*	-0.18	-0.40***
0	-0.23*	-0.05	-0.01	-0.41***	-0.32**	-0.19	-0.06	-0.43***
+1	-0.25*	-0.13	-0.04	-0.43***	-0.36***	-0.25*	-0.18	-0.45***
+2	-0.19	-0.05	-0.04	-0.40***	-0.29**	-0.17	-0.13	-0.43***
+3	-0.23*	-0.10	-0.13	-0.37***	-0.30**	-0.08	-0.05	-0.39***

주: 숫자는 t시점의 해운뉴스지수와 $t+h$시점의 해운업 경기실사지수 간의 교차 상관계수를 의미함. 추정기간은 2013:M8부터 2018:M2임. * p 〈 .10 ** p 〈 .05 *** p 〈 .01

22) 해운업 경기실사지수는 해운기업 업종을 나타내는 컨테이너, 건화물, 유조선 부문으로 구분되어 있으며 업황과 채산성, 자금사정, 매출에 관한 동향과 전망에 대한 질문으로 구성되어 있다.

4. 예측 회귀분석

해운뉴스지수의 예측력 검정을 위해 해상운송화물수입의 예측변수로써 해운뉴스지수와 발틱운임지수, 원달러 환율을 사용한 회귀분석을 실시하였다. 예측력 검정 모형은 식 (5)와 같다.

$$\triangle_h Y_{t+h} = \alpha + \beta_1 MNI_t + \beta_2 BDI_t + \beta_3 FX_t + \sum_{i=1}^{h} \gamma_i \triangle Y_{t-i} + \epsilon_{t+h}, \quad (5)$$

식(5)에서 Y_t는 해상운송화물수입을 의미한다. $\triangle_h Y_{t+h}$는 전망시계(forecast horizon) h 후 Y_{t+h}와 한 달 이전 Y_{t-1}의 연간증감률을 나타낸다.[23] $\triangle Y_{t-i}$는 연간증감률로 변환된 해상운송화물수입의 증감률의 i 시차값(lagged value)을 의미한다. MNI_t는 t 시점의 해운뉴스지수이며 BDI_t는 발틱운임지수를, FX_t는 원달러환율의 월평균값을 의미한다.[24] ϵ_{t+h}는 예측오차이다. 모형에서 사용된 해상운송화물수입과 원달러환율의 출처는 한국은행의 경제통계시스템이고 발틱운임지수는 Clarkson의 자료를 활용하였다. 분석 대상은 월간 자료이며 분석기간은 2006:M1부터 2018:M2이다. 월간 자료임을 고려하여 전망시계는 각각 3개월과 12개월로 설정하여 분석하였다.

<표 4>는 해운뉴스지수의 예측 회귀분석 결과를 나타낸다. 해운뉴스지수는 모든 예측시계에서 통계적으로 유의한 예측력을 지니는 것을 확인할 수 있다.

---

23) $\triangle_h Y_{t+h} = \frac{1200}{h+1} ln(\frac{Y_{t+h}}{Y_{t-1}})$

24) 자료의 안정성(stationary)을 확인하기 위해 Dickey-Fuller 검정을 수행하였다. 분석결과 해운뉴스지수는 안정적인 시계열로 나타나 원자료를 사용하였으며 발틱운임지수와 환율은 불안정(non-stationary) 시계열로 확인되어 전월대비 증감률로 변환하여 분석하였다.

▌표-4. 해운뉴스지수 예측 회귀분석 결과▐

구분	3개월 예측시계			12개월 예측시계		
	(1)	(2)	(3)	(4)	(5)	(6)
해운뉴스지수	-0.041*	-0.044**	-0.044**	-0.029**	-0.030**	-0.029**
	(-1.766)	(-2.228)	(-2.324)	(-2.079)	(-2.247)	(-2.262)
발틱운임지수		0.061***	0.053***		0.020***	0.017**
		(4.112)	(3.754)		(2.759)	(2.554)
환율			-0.336***			-0.139*
			(-2.585)			(-1.861)
상수	5.397	6.149	6.428	-0.847	-0.320	-0.229
	(0.810)	(0.975)	(1.074)	(-0.152)	(-0.060)	(-0.044)
관측치	139	139	139	121	121	121
R^2	0.088	0.210	0.252	0.241	0.286	0.310

주: 해운뉴스지수의 예측 회귀분석(forecasting regression) 결과. 숫자는 계수를, 괄호안은 z값을 나타냄. (1)열부터 (3)열의 종속변수는 예측시계 3개월의 해상운송화물수입의 연간증감률이고 (4)열부터 (6)열은 예측시계 12개월의 해상운송화물수입의 연간증감률임. 해상운송화물수입 변화율의 시차값 계수는 생략함. 표준오차는 오차항의 자기상관을 고려한 Newey and West (1987)의 표준오차를 Newey and West (1994)의 the automatic lag selection method를 통하여 계산함. 분석기간은 2006:M1부터 2018:M12임. * p 〈 .10 ** p 〈 .05 *** p 〈 .01

3개월 예측시계의 경우, 해운뉴스지수만 포함된 예측모형인 (1)열에서 해운뉴스지수의 계수의 예측력이 발틱운임지수와 환율을 통제한 (2), (3)열에 비해 계수의 통계적 유의성 및 R^2가 낮은 것으로 나타나는 데, 이는 단기에서 해운뉴스지수의 예측력이 환율과 발틱운임지수를 통제한 모형에 비해 낮음을 의미한다. 이와 같은 결과는 환율과 발틱운임지수가 우리나라 해운산업이 아닌 거시경제 여건에 의해서 결정되고 이러한 거시경제 여건의 변동이 해운업의 위험요인으로 인식되어 여론을 통해 공론화되는 데까지는 시차가 존재하기 때문으로 해석 할 수 있다. 한편, 환율과 발틱운임지수를 통제한 모형에서 해운뉴스지수의 예측력이 통계적으로 유의한 것으로 나타나는 것은 해운뉴스지수가 환율과 발틱운임지수가 설명하지 못하는 부분에 관한 예측력을 지니는 것으로 해석할 수 있다.

12개월 예측모형의 경우, 해운뉴스지수는 발틱운임지수와 환율 통제 여부와 무관하게 통계적으로 유의한 예측력을 지닌 것을 확인할 수 있으며 환율의 경우, 12개월 예측 모형에서 예측력의 통계적 유의성이 낮은 것으로 나타났다.

Ⅳ. 결 론

본 연구는 뉴스 빅데이터를 활용하여 우리나라 해운업의 신용 위험을 측정하는 해운뉴스지수를 개발·소개하였다. 반복되는 해운산업의 위기에도 불구하고 해운산업의 위험을 감지하는 대표 변수가 부존하다는 점에 착안하여 우리나라 해운산업의 위험을 감지하고 미연에 방지하는 데에 유용한 정보를 지닌 지수를 고안하였다.

해운뉴스지수는 뉴스 빅데이터분석을 사용하여 우리나라 해운산업의 신용위험을 나타내는 지수이다. 지수 작성을 위해 Grammenos(2010)의 신용평가의 6C에 의거한 5개부문의 26개 키워드가 포함된 언론기사의 빈도를 사용하였다. 뉴스 자료는 한국언론진흥재단에서 제공하는 빅카인즈의 12개 종합일간지 및 경제지의 1991년 1월부터 2018년 3월까지 18,909건의 신문기사를 활용하여 지수를 작성하였다.

해운뉴스지수는 우리나라 경기 수축기뿐만 아니라 주요 해운기업 위기에 크게 상승하는 경기 역행적(countercyclical) 패턴을 보이는 것을 확인할 수 있었다. 주요 해운경기지수인 해운업 경기실사지수와의 교차 상관관계 검정에서 해운뉴스지수는 해운업 경기실사지수와 유의한 음(-)의 상관관계를 지니는 것으로 나타났으며 특히 업황 및 매출 지수와 통계적으로 유의한 상관관계를 보였다. 해상운송화물수입과의 예측 회귀분석을 통해 해운뉴스지수가 해상운송화물수입에 통계적으로 유의한 예측력을 지니는 것으로 나타나 해운뉴스지수가 해운시장의 위험을 감지하고 예측하는데 도움이 되는 정보를 지니고 있음을 확인할 수 있었다.

해운산업의 위험을 측정하기 위해 새로운 접근 방법인 뉴스 빅데이터 분석을 사용했다는 점이 본 연구가 지닌 의의라 할 수 있을 것이다. 해운뉴스지수는 우리나라 해운시황에 대한 우려와 발생 가능한 위험을 조기에 식별하여 시장에 대한 정부의 효과적인 관리·감독은 물론, 해운기업의 선박매매 및 항로개척과 같은 기업의 효율적 의사결정에 도움을 줄 것으로 기대된다.

해운뉴스지수가 해운 산업 위험을 보다 효과적으로 감지하기 위해선 본 연구에서 제시된 키워드 구성뿐만 아니라 해운산업 위험을 포괄적으로 내포하는 키워드를 발굴하고 지수 작성 방법 역시 개선되어야 할 것이다. 이를 위한 정책당국과 해운산업 참여자 및 학계의 논의가 요구된다.

투고일 2018. 04. 27
1차 심사일 2018. 06. 21
게재확정일 2018. 06. 28

■■ 참고문헌

1. 강선아, 김유신, 최상현(2015),「텍스트마이닝을 이용한 사회 이슈 찬반 분류에 관한 연구」,『한국데이터정보과학회지』, 제26권 제5호, 한국데이터정보과학회, pp.1167-1173.
2. 김재봉(2000),「해운산업의 운임리스크 관리에 관한 연구」,『리스크 管理硏究』, 제11권, 제1호, 한국리스크관리학회, pp.157-187.
3. 김재봉(2007),「해운산업의 전사적 리스크관리에 관한 연구」,『국제해양문제연구』, 제19권, 한국해양대학교 국제해양문제연구소, pp.171-188.
4. 김태일(2014),「해운기업 부도예측을 위한 통태적 모형 실증연구」, 한국외국어대학교 박사학위논문, pp.132-135.
5. 나정호, 김광원, 방희석(2015),「논문 : 해운기업의 위험관리 전략에 관한 연구」,『해운물류연구』, 제85권, 한국해운물류학회, pp.185-207.
6. 박대민(2016),「장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성」,『韓國 言論學報』, 제60권 제5호, 한국언론학회, pp.353-407.
7. 박성화, 김태일, 권장한(2017),「한국 해운기업의 부실예측 연구」,『해운물류연구』, 제97권, 한국해운물류학회, pp.823-852.
8. 손태규(2013),「언론의 기능과 역할에 대한 사법부의 인식과 판단」,『美國憲法硏究』, 제24권 제3호, 미국헌법학회, pp.99-141.
9. 안희준, 전승표, 최종범(2010),「남북관계 관련 뉴스가 주식시장에 미치는 영향」,『韓國經濟의 分析』, 제16권 제2호, 한국금융연구원, pp.199-231.
10. 예금보험공사, “FDIC의 시장정보 활용실태 및 시사점”, 조사분석자료, 2005. 11.29.
11. 이긍희(2017),「뉴스 빅데이터를 이용한 경기 판단: 빅카인즈 뉴스 경기지수의 개발」,『통계연구』, 제22권 제2호, 통계청, pp.67-94.
12. 이완수, 박양수(2016),「경제 정보에 대한 비대칭적 반응」,『韓國 言論學報』, 제60

권 제1호, 한국언론학회, pp.165-201.

13. 이완수, 박재영(2013), 「방송뉴스의 언어와 표현」, 『방송과 커뮤니케이션』, 제14권 제1호, 문화방송, pp.5-46.
14. 이재규, 김일수(2016), 「신호접근법에 의한 한국의 해운시장 위기 지표 분석 」, 『貿易 硏究』, 제12권 제5호, 한국무역연구원, pp.485-496.
15. 이형석, 김기석(2006), 「DEA 모형을 이용한 우리나라 해운업체의 정태적·동태적 효율성 분석」, 『大韓經營學會誌』, 제19권 제4호, 대한경영학회, pp.1197-1217.
16. 장영태(1994), 「몬테 카를로 시뮬레이션을 이용한 해운프로젝트의 투자위험측정연구」, 『Ocean and Polar Research』, 제16권 제2호, 한국해양연구원), pp.157-173.
17. 최윤희(2016), 「[특별기획 Ⅱ: 데이터저널리즘과 DB 큐레이션] 진화하는 빅카인즈, 데이터저널리즘의 플랫폼으로 기여」, 『조사연구』, 제28권, 한국조사기자협회, pp.89-97.
18. 한국해양수산개발원, “2018년 3월 해운업 경기실사지수 (해운업 BSI) 조사결과”, 2018.04.04.
19. Albertijn, S., Bessler, W. and Drobetz, W.(2011), “Financing Shipping Companies and Shipping Operations: A Risk-Management Perspective”, *Journal of Applied Corporate Finance*, (Wiley Online Library) Vol.23(4), pp.70-82.
20. Alexopoulos, M. and Cohen, J.(2015), “The power of print: Uncertainty shocks, markets, and the economy”, *International Review of Economics and Finance*, Vol.40, pp.8-28.
21. Alexopoulos, M. and Cohen, J.(2009), “Uncertain Times, Uncertain Measures”, *University of Toronto Working Paper*
22. Baker, S.R., Bloom, N. and Davis S., J.(2016), “Measuring Economic Policy Uncertainty”, *The Quarterly Journal of Economics*, Vol.131(4), pp.1593-1636.
23. Bontempi M.E., Golinelli, R. and Squadrani, M.(2016), “A New Index of Uncertainty Based on Internet Searches : A Friend or Foe of Other Indicators?”, Working Paper
24. El-MasryAhmed, A., Olugbode, M. and Pointon, J.(2010), “The exposure of

shipping firms' stock returns to financial risks and oil prices: A global perspective", *Maritime Policy and Management*, Vol.37(5), pp.453-473.

25. Fuentes-albero, C., Gilchrist, S. and Zakrajsek, E.(2016), "The Macroeconomic Impact of Financial and Uncertainty Shocks", *European Economic Review*, Vol.88, pp.185-207
26. Grammenos C.(2010), "Revisiting credit risk, analysis and policy in bank shipping finance", *The Handbook of Maritime Economics and Busines*, (London), LLP Publications, p.781.
27. Lozinskaia, A., Merikas, A., Merika, A., and Penikas, H.(2017), "Determinants of the probability of default: the case of the internationally listed shipping corporations", *Maritime Policy and Management*, Vol.44(7), pp.837-858.
28. Menachof, D., A. and Dicer, G.N.(2001), "Risk Management Methods for the Liner Shipping Industry", *Maritime Political Management*, Vol.28(2), pp.141-155.
29. Nam, H.J. and An, Y.H.(2017) "Default Risk and Firm Value of Shipping & Logistics Firms in Korea", *Asian Journal of Shipping and Logistics*, Vol.33(2), pp.61-65.
30. Newey, W.K. and West, K.D.(1987). "Hypothesis Testing with Efficient Method of Moments Estimation", *International Economic Review*, pp.777-787.
31. Newey, W.K. and West, K.D.(1994). "Automatic Lag Selection in Covariance Matrix Estimation", *The Review of Economic Studies*, Vol.61(4), pp.631-653.
32. Oltedal, H., Wadsworthz, E.(2010), "Risk perception in the Norwegian shipping industry and identification of influencing factors", *Maritime Policy and Management* Vol.37(6), pp.601-623.
33. Kmenta, J.(1986), *Elements of Econometrics*, (N.Y.), Macmillan Publishing Company.
34. Stopford M.(2013), *Maritime economics*. Routledge.
35. www.bigkinds.or.kr. (한국언론진흥재단, 빅카인즈, 2018년 4월 13일 검색).