

# 베이지안 변수선택 기법을 이용한 발틱건화물운임지수(BDI) 예측

한상우

전북대학교 무역학과 석사과정

김영민

전북대학교 무역학과 조교수

## Forecasting the Baltic Dry Index Using Bayesian Variable Selection

Xiang-Yu Han<sup>a</sup>, Young Min Kim<sup>b</sup>

<sup>a</sup>Department of International Trade, Jeonbuk National University, South Korea

<sup>b</sup>Department of International Trade, Jeonbuk National University, South Korea

Received 07 September 2022, Revised 27 October 2022, Accepted 30 October 2022

### Abstract

Baltic Dry Index (BDI) is difficult to forecast because of the high volatility and complexity. To improve the BDI forecasting ability, this study apply Bayesian variable selection method with a large number of predictors. Our estimation results based on the BDI and all predictors from January 2000 to September 2021 indicate that the out-of-sample prediction ability of the ADL model with the variable selection is superior to that of the AR model in terms of point and density forecasting. We also find that critical predictors for the BDI change over forecasts horizon. The lagged BDI are being selected as an key predictors at all forecasts horizon, but commodity price, the clarksea index, and interest rates have additional information to predict BDI at mid-term horizon. This implies that time variations of predictors should be considered to predict the BDI.

**Keywords:** Baltic Dry Index, Out-of-sample forecasting, Bayesian variable selection

**JEL Classifications:** C11, C53, F17, L91

<sup>a</sup> First Author, E-mail: [hansangwoo0306@gmail.com](mailto:hansangwoo0306@gmail.com)

<sup>b</sup> Corresponding Author, E-mail: [kimym.econ@gmail.com](mailto:kimym.econ@gmail.com)

© 2022 The Korea Trade Research Institute. All rights reserved.

## I. 서론

발틱건화물운임지수(Baltic Dry Index 이하 BDI)는 영국 런던에 기반을 둔 발틱해운거래소(Baltic Exchange)에서 건화물(Dry Cargo) 운임지수로 사용되어온 BFI 지수(Baltic Freight Index)를 대체하고 1985년 1월 4일 운임 수준을 기준(=1,000)으로 산정하여 주요 노선을 운항하는 벌크선(Bulk Carrier)들의 운임을 평균해 가중치를 적용한 세계 해운업의 경기상황을 나타내는 종합운임지수이다. BDI 지수는 포장 없이 벌크선으로 운송하는 석탄, 철광석, 건축자재, 곡물 등의 원자재에 대한 운임을 나타내는 지표로 원자재의 운송량이 증가하면 BDI 지수는 상승하고 운송량이 감소하면 BDI 지수는 하락한다.

항공 산업의 발전으로 항공 화물량이 증가한 건 사실이지만 석탄, 철광석, 원유, 곡물 등의 필수 원자재는 여전히 해운을 통해 이루어지고 있다. 벌크선이 운송하는 화물은 생산의 기반이 되는 원자재이기 때문에 벌크선이 활발하게 움직인다는 것은 원자재에 대한 수요가 증가하고 있다는 것을 나타낸다. 즉, 생산이 활발하게 진행되고 있음을 의미하고 벌크선의 움직임이 활발하지 않다면 반대로 해석할 수 있다. 그리고 원자재의 수요가 증가함에 따라 벌크선의 운임이 상승하여 BDI 지수가 상승하는데, 이는 원자재의 운송을 위한 벌크선의 수요가 증가한다는 것을 의미한다. 시장 상황에 따라 운임이 결정되는 BDI 지수는 원자재의 운송량이 증가하면 상승하여 해운업체의 이익 증가로 이어지고 수익이 늘어난 해운업체들은 선박 발주량을 증가시킨다. 이렇게 수주가 늘어나면 조선업의 경기 또한 좋아지게 될 뿐만 아니라 조선업체는 선박을 만들기 위한 철강 수요의 증가로 인해 철강업체의 경기 또한 좋아지게 된다. 게다가 BDI 지수는 경기 선행지표로서 경제활동을 나타내는 중요한 지표일 뿐만 아니라 미래의 경제성장이나 인플레이션을 예측하는 지표로도 사용되고 있다. 따라서 선박금융 업무를 수행하는 기관뿐만 아니라 실물경제의 현황과 동향을 파악해야 하는 금융 및 정부 기관들도 해운시장에 대한 분석 및 정확한 예측은 아주 중

요하다고 할 수 있다.

해운금융시장에서는 해상운임의 변동성을 헤지(Hedge)하기 위해 주요 항로별 운임에 대한 선도거래가 이루어지고 있다. 해상운임선도거래란 현재 시점에서 계약당사자 간 합의된 가격으로 미래 시점의 해상운임 또는 용선료를 매매하는 거래를 뜻한다. 이때 발틱해운거래소에서 발표한 항로별 운임지수를 사용한다. 하지만 적지 않은 나라의 국적선사와 화주는 해상운임선도거래를 거의 활용하고 있지 않다. 이는 해상운임의 특성상 높은 변동성으로 인해 위험관리 측면에서 거래의 대부분이 장외시장(Off Board Market)에서 화주가 가진 화물정보와 선사가 가진 선박정보의 교환을 통해 직접 거래를 하여 계약을 성사시키거나 중개기관을 통해 제한된 거래가 이루어지며 무엇보다 해상운임선도거래를 활용하기 위해 해운시장의 전망에 대한 자체적인 예측역량이 부족하기 때문이다. 따라서 BDI 지수의 정확한 예측을 통해 해운선사의 이해 관계자 및 화주는 해상운임 변동 위험에 대비하여 파생금융상품인 해상운임선도거래를 활용하여 계약을 성사시켜야 한다.

BDI 지수 및 기타 해상운임과 관련된 대표적인 연구는 다음과 같다. Cullinane, Mason and Cape (1999)은 ARIMA 모형을 통해 1993년 BFI 지수(Baltic Freight Index) 개정 전과 후의 자료를 사용하여 1993년의 개정이 BFI 지수의 움직임에 큰 영향을 미치지 않았음을 보여 주었다. Bildirici, Kay i kç i and Onat (2015)은 MS-VAR 모형을 활용하여 BDI 지수와 경제성장률 간의 관계를 분석한 결과, BDI 지수가 미국 GDP 성장률의 위기를 나타내는 지표로 사용할 수 있음을 확인하였다. Xu, Yip and Marlow (2011)은 AR-GARCH 모형을 이용하여 건화물 시장에서의 운임과 선박 공급량 사이의 관계를 분석하고 확인하였다. Lin, Chang and Hsiao (2019)은 VAR-BEKK-GARCH-X 모형을 사용하여 BDI 지수가 해운기업의 주가, 미국의 통화 및 원자재 가격에 영향을 미친다는 것을 검증하였다. Ruan, Wang, Lu and Qin (2016)에서는 교차 상관관계 분석을 통해 BDI 지수와 원유 가격(서부 텍사스유, 브렌트유) 간의 관계를 확인하였다.

국내 문헌으로는 Lee Sung-Yhun and Ahn Ki-Myung (2018)은 VECM 모형으로 추정된 실증분석결과에 따르면 리보금리, 환율, 벌크선 발주 선복량, 그리고 중국경제성장률이 BDI 지수에 영향을 미치는 것으로 나타났다. Kim Bu-Kwon, Kim Dong-Yoon and Choi Ki-Hong (2019)은 Copula 모형을 사용하여 선형별 건화물 운임지수와 서부 텍사스유 가격 변화율 사이의 의존구조를 분석하였다. Kim Chang-Beom (2011)에서는 엔/달러 환율과 다우존스지수가 건화물 운임지수에 영향을 미친다는 것을 검증하였다. 마지막으로 Bae Sung-Hoon, Ha Young-Mok and Park Keun-Sik (2018)은 중국 대두 수입량, 철광석 가격, 석탄 가격, 석탄 물동량, 중국 PMI 지수가 BDI 지수에 영향을 미치는 것을 확인하였다.

이러한 BDI 지수와 관련된 연구의 중요성에도 불구하고 1) 높은 변동성과 2) 지수가 결정되는 과정의 복잡성으로 인해 정확한 예측이 매우 어렵다고 알려져 있다. 이를 극복하기 위해서 기존 문헌에서는 BDI 지수의 예측 정확도를 높이기 위해 다양한 모형을 고려해왔다. Zeng, Qu, Ng and Zhao (2016)은 경험적모드 분해법(Empirical Mode Decomposition, EMD)과 인공신경망(Artificial Neural Networks, ANN)을 결합한 EMD-ANN 모형을 사용하여 ANN 및 VAR 모형에 비해 BDI 지수에 대한 예측력이 더 우수함을 보여주었다. Tsioumas, Papadimitriou, Smirlis and Zahran (2017)은 중국 철강 생산량, 신조선 발주량과 DBECI 지수를 추가적인 예측변수로 사용하여 VARX 모형이 ARIMA 모형과 비교했을 때 BDI 지수에 대한 예측력이 더 우수한 것을 확인하였다. Bae Sung-Hoon, Lee Gun-Woo and Park Keun-Sik (2021)은 ANN, RNN, LSTM과 같은 딥러닝 모델을 통해 BDI 지수를 예측 및 비교한 결과, LSTM 모형의 예측력이 가장 뛰어난 것으로 나타났다.

BDI 지수 예측과 관련된 국내 문헌은 다음과 같다. Mo Soo-Won(2010)은 ARIMA 모형과 HP Filter를 사용하는 경우 임의보행 모형보다 예측력이 더 우수한 것을 확인하였다. Kim Hyeong-Jun, Ryu Doo-Jin and Cho-Hoon

(2019)은 같은 방법을 활용하여 BDI 지수의 다양한 예측시계(1, 2, 3개월 후)를 설정하여 분석한 결과, 임의보행 모형보다 상대적으로 예측력이 더 우수하다는 것을 확인하였다. Kim Hyun-Sok and Chang Myung-Hee (2014)는 VAR 모형에 환율과 산업생산지수를 BDI 지수를 예측하는 데 포함하여 일반적인 VAR 모형과 비교했을 때 베이지안 VAR 모형의 예측력이 더 높은 것으로 나타났다.

일반적으로 BDI 지수 예측 연구에서 국제유가, 환율, GDP, 산업생산량과 같은 금융변수를 예측변수로 사용해왔기 때문에 BDI 지수는 다양한 거시 및 금융변수와 밀접한 관계를 갖는다고 볼 수 있다. 하지만 BDI 지수가 거시 및 금융변수들과 밀접한 관계를 갖는다고 해서 이론에 기반하여 예측변수에 모두 포함하면 불필요한 변수들로 인해 발생하는 과도적합(Overfitting) 문제를 초래한다. 또한 연구자가 기존 연구를 통해 예측변수를 선형적으로 선택한다면 중요한 변수가 누락될 가능성이 존재한다. 결국 BDI 지수 예측 연구는 전형적인 예측변수 불확실성 문제에 직면하고 있으며 불확실성 문제를 고려하지 않는다면 정확한 예측을 기대하기 어렵다. 따라서 과도적합 및 변수누락 문제를 동시에 해결하면서 예측력을 극대화하기 위해 최적 예측변수들을 찾아내는 변수선택 기법이 필요하다. 기존 연구와 달리 본 연구에서는 BDI 지수의 예측력을 향상시키기 위해 베이지안 변수선택 기법(Bayesian Variable Selection)을 적용한 자기회귀시차분포(Autoregressive Distributed Lag 이하 ADL) 모형과 자기회귀(Autoregressive 이하 AR) 모형을 이용하여 BDI 지수의 분포예측을 시행하였다. BDI 지수 외에 총 16개의 예측변수를 사용하였고 모든 자료는 월별 자료이며 자료의 기간은 2000년 1월부터 2021년 9월까지이다. 추정결과, 2016년 10월부터 2021년 9월까지 최근 5년의 자료를 대상으로 모든 예측시계를 종합적으로 고려하였을 때, 본 연구에서 제시한 베이지안 변수선택 기법을 적용한 ADL 모형의 예측력이 점예측과 분포예측 모두에서 AR 모형에 비해 우수하였다. ADL 모형의 실제 예측결과, BDI 지수는 현재 시점에서 향후 1년 동안

8% 내외의 수준을 유지할 것으로 예측된다. 특히, 2022년 1월과 5월은 7%로 하락할 것으로 예측되었다. 또한 ADL 모형의 추정결과로부터 각 예측시계에 따라 선택된 예측변수들을 살펴보면 단기 예측에서는 전기의 BDI 지수가 중요한 예측변수로 선택되었고 중기 예측에서는 BDI 지수, 철광석 가격, 석탄 가격, 밀 가격, 국제 유가, 리보금리, 클락슨 지수가 중요한 예측변수로 선택되었다. 그리고 장기 예측에서는 BDI 지수, 클락슨 지수, 철광석 가격이 BDI 지수 예측에 중요한 예측변수로 작용하였다.

추가적으로 베이지안 변수선택 기법을 사용한 대표적인 선행연구로 George and McCulloch (1997)의 연구에서는 베이지안 변수선택 기법의 사전분포의 가정과 사후분포의 계산방법에 대해 설명하였다. Kim Young-Min and Lee Seo-Jin (2020)는 베이지안 변수선택 기법을 사용하여 1년 미만의 단기 예측에서 캐나다 달러와 영국 파운드에 대해 베이지안 변수선택 기법이 약 60%의 정확도로 환율 변화율을 예측하였고 환율에 대해 중요한 예측 변수는 시간에 따라 달라지며 국가마다 다르다는 것을 확인하였다. Lee Chang-Hoon, Kang Kyu-Ho and Ann Ji-Hee (2020)는 정확한 서울 아파트 매매 가격지수 예측분포를 도출하기 위해 베이지안 변수선택 기법을 사용하여 표본 외 예측을 통해 벤치마크모형들과 예측력을 비교 및 평가하였다. Koo Byung-Soo (2020)는 모형의 불확실성을 고려하기 위해 동태적 Nelson-Siegel 모형에 베이지안 변수선택 기법을 적용하여 수익률 곡선을 추정하고 미래 수익률에 대한 시장기대를 추정하였다.

본 논문의 구성은 다음과 같다. 제II장에서는 BDI 지수를 예측하는데 사용되는 예측모형에 대해 설명한다. 제III장에서는 자료와 모형을 추정하고 예측하는 방법을 간략하게 요약하고 제IV장에서는 표본 외 예측결과와 본 연구에서 제시한 모형으로부터 생성된 BDI 지수의 실제 예측결과 및 예측시계에 따른 주요변수가 선택된 결과를 제시한다. 제V장에서는 결론으로 마무리를 짓는다.

## II. 모형설정 및 추정

### 1. 예측모형

본 연구에서는 BDI 지수의 표본 외 예측력을 향상시키기 위해 베이지안 변수선택을 적용한 ADL 모형을 고려하였다. ADL모형은 종속변수의 자기회귀 항과 강외생적인(Strictly Exogenous) 예측변수  $x_t$ 를 추가하여 추정하고 예측하는 모형이다.  $y_t$ 는  $t$ 시점의 BDI 지수이다. 구체적으로 BDI 지수는 아래와 같이  $h$ -기와 그 이전의 자기시차 또는 외생변수에 의해 주어진다.

$$y_t = \beta_0 + \sum_{j=1}^J \rho_j y_{t-j-h+1} + \sum_{k=1}^K \sum_{j=1}^J \phi_{k,j} z_{k,t-j-h+1} + e_t, \quad e_t \sim N(0, \sigma^2)$$

$k$ 는 예측변수의 수,  $x_{k,t}$ 는  $t$ 시점의  $k$ 번째 예측변수를 나타내고  $J$ 는 고려되는 최대 시차이다. 위 식을 단순화하여 다음과 같은 행렬식으로 표현할 수 있다.

$$Y = X\beta + e$$

여기서  $Y$ 는 BDI 지수의 시계열벡터,  $X$ 는 종속변수의 자기회귀항과 예측변수의 시차로 이루어진 행렬이며  $\beta = (\beta_0, \beta_{1,1}, \beta_{1,2}, \dots, \beta_{K+1,J-1}, \beta_{K+1,J})'$ 는 각 예측변수들의 계수벡터이다. 예측을 목적으로 하는 연구에서 불필요한 변수를 예측변수로 사용하여 예측하는 경우 정확도를 떨어뜨릴 수 있다. 베이지안 변수 선택기법은 중요한 예측변수들의 추정계수 값이 0이 되게 함으로써 종속변수에 영향을 미치지 않도록 하여 중요한 예측변수들만 활용하여 예측하는 방법론이다.

예측력 비교를 위해 본 연구에서는 가장 대표적인 시계열 예측모형인 AR 모형을 기준 모형(Benchmark Model)으로 사용한다. AR 예측모형은 아래와 같이 주어진다.

$$y_t = \beta_0 + \sum_{j=1}^J \rho_j y_{t-j-h+1}$$

AR 모형의 경우 BDI 지수 자료만 사용하여 예측을 시행하고 ADL 모형과의 예측력 비교를 통해 본 연구에서 고려하고 있는 예측변수가 BDI 지수 예측에 추가적인 정보를 포함하고 있는지 여부를 검증할 수 있다. AR 모형 역시 앞서 설명한 ADL 모형과 동일한 방식으로 예측 시계별로 최적 시차를 선택하기 위해 베이지안 변수선택 방법을 활용한다.

## 2. 사전분포

베이지안 방법론에서는 사전분포를 모형으로 취급하기 때문에 변수선택 방법론을 적용하기 위한 사전분포에 대해 설명하도록 하겠다. 먼저 특정 예측변수  $X$ 가 종속변수  $Y$ 를 예측하는데 중요한 변수라면 1, 중요하지 않다면 0인 값을 갖고 베르누이 분포를 따르는  $\gamma_{k,j}$ 를 도입한다. 즉,  $\gamma_{k,j}$ 는  $X$ 의 중요성 여부를 나타내는 지시함수(Indicator Function)이다. 따라서  $\gamma_{k,j}$ 의 사전분포는 아래와 같이 설정된다.

$$\Pr(\gamma_{k,i} = 1) = p, \Pr(\gamma_{k,i} = 0) = 1 - p$$

여기서  $p$ 는 하이퍼 파라미터(Hyper Parameter)로 해당 설명변수가 중요할 확률이고 반대로  $(1-p)$ 는 중요하지 않을 확률이다. 따라서  $\beta_{k,j}$ 는 다음과 같은 사전분포를 따른다.

$$\beta_{k,i} | \gamma_{k,i} \sim \gamma_{k,i} N(0, b_1) + (1 - \gamma_{k,i}) N(0, b_0)$$

이는  $\gamma_{k,j} = 1$ 이라면  $\beta_{k,j}$ 의 사전분포는 다음과 같고,

$$\beta_{k,i} | \gamma_{k,i} \sim N(0, b_1)$$

$\gamma_{k,j} = 0$ 일 때,  $\beta_{k,j}$ 의 사전분포는 아래와 같다는 것을 의미한다.

$$\beta_{k,i} | \gamma_{k,i} \sim N(0, b_0)$$

일반적으로  $b_0$ 와  $b_1$ 은 하이퍼 파라미터(Hyper Parameter)이고  $b_0, b_1 > 0$ ,  $b_1 > b_0$ 이며  $b_0$ 은 0에 가까운 값으로 설정하고 반대로  $b_1$ 은  $b_0$

에 비해 충분히 큰 값을 갖도록 설정한다.

베이지안 변수선택 방법론은 하이퍼 파라미터를 어떻게 설정하는지에 따라서 연구결과가 민감하게 변하는 경우가 종종 있다. 본 연구에서는 이러한 문제를 완화하기 위해 변수선택과정에서 데이터의 정보를 최대한으로 반영할 수 있도록 하이퍼 파라미터인  $p$ ,  $b_0$ ,  $b_1$ 에 대하여 계층적(Hierarchical) 사전분포를 사용한다. 이는 연구자의 주관에 따라 하이퍼 파라미터를 결정하는 것이 아니라 자료의 정보를 통해  $b_0$ 와  $b_1$ , 그리고  $p$ 를 업데이트하는 방법이다. 우선  $p$ 는 0과 1사이의 연속적인 확률을 나타내는 값이므로 다음과 같은 베타분포를 따른다.

$$p \sim \text{Beta}(c_0, d_0)$$

이 연구에서는  $c_0 = 2$ ,  $d_0 = 2$ 로 설정한다. 또한  $b_0$ 와  $b_1$ 에 대해 다음과 같은 역감마분포를 따른다고 가정한다.

$$b_0 \sim \text{IG}(\nu_{00}/2, \delta_{00}/2), b_1 \sim \text{IG}(\nu_{01}/2, \delta_{01}/2)$$

여기서  $b_0$ 의 사전평균은  $E(b_0) = \delta_{00}/(\nu_{00} - 2)$ 로 계산되는데 그 값을 0에 가깝게 하고  $b_1$ 의 사전평균은  $E(b_1) = \delta_{01}/(\nu_{01} - 2)$ 의 값보다 훨씬 작게 설정한다.

마지막으로 오차항  $e_i | \sigma^2 \sim N(0, \sigma^2)$ 은 베이지안 추정방법상 편의를 위해 조건부정규분포를 따르며  $\sigma^2$ 에 대한 사전분포는 일반적으로 다음과 같은 역감마분포를 따른다고 가정한다.

$$\sigma^2 \sim \text{IG}(\nu_0/2, \delta_0/2)$$

여기서 오차항의 분산에 대한 사전분포의 하이퍼-파라미터는 자료의 정보를 충분히 반영할 수 있도록 설정하였다. 즉  $\nu_0 = 5$ ,  $\delta_0 = 0.5$ 로 설정하였다.

## 3. 추정방법

본 연구에서는 결합사후분포(Joint Posterior Distribution)를 도출하기 위해 깃스샘플링

(Gibbs-Sampling) 알고리즘을 활용한다. 이는 특정 파라미터를 제외한 모든 파라미터와 데이터가 주어져 있을 때, 완전조건부분포(Full Conditional Distribution)로부터 샘플링하여 사후분포를 도출하는 방법이다.  $\theta$ 는 파라미터  $\beta, \sigma^2, b_0, b_1, p$ 의 집합이고  $k$ 번째 예측변수의 중요도에 해당하는  $\gamma_k$ 의 행렬은  $\Gamma = (\gamma_0, \gamma_{1,1}, \gamma_{1,2}, \dots, \gamma_{K+1, J-1}, \gamma_{K+1, J})$ 이다. 따라서  $\theta$ 와  $\Gamma$ 의 결합사후분포는 다음과 같이 도출된다.

$$\pi(\theta, \Gamma | Y) \propto f(Y | \beta, \sigma^2) \pi(\beta | \Gamma, b_0, b_1) \pi(\Gamma | p) \pi(p) \pi(\sigma^2) \pi(b_0) \pi(b_1)$$

모형 파라미터들의 사후분포 샘플링에 대한 단계를 요약하면 다음과 같다.

- 사후분포 샘플링 알고리즘
- 0 단계: 초기값과 샘플링 횟수 설정
  - 1 단계:  $\beta$ 를  $\beta | Y, \theta_{-\beta}, \Gamma$ 에서 샘플링
  - 2 단계:  $\sigma^2$ 을  $\sigma^2 | Y, \beta$ 에서 샘플링
  - 3 단계:  $b_0, b_1$ 을  $b_0, b_1 | \beta, \Gamma$ 에서 샘플링
  - 4 단계:  $p$ 를  $p | \Gamma$ 에서 샘플링
  - 5 단계:  $\Gamma$ 를  $\Gamma | \beta, b_0, b_1, p$ 에서 샘플링
  - 6 단계: 1-5단계를 반복

#### 4. 예측력 평가 방법

본 연구에서는 모형의 예측력을 평가하기 위해 일반적으로 예측력을 평가하는데 가장 널리 사용되는 평균 제곱근 오차(Root Mean Square Error, 이하 RMSE)와 베이지안 방법론에서 표준적으로 사용되는 로그 사후예측우도(Posterior Predictive Likelihood, 이하 PPL) 두 가지 방법을 사용한다.

먼저 RMSE는 점예측의 정확도를 측정하는 방법인데 수학적 정의는 다음과 같다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{t+1} - \hat{Y}_{t+1})^2}$$

여기서  $n$ 은 표본 외 예측 기간,  $Y_{t+1}$ 은 실제 값, 그리고  $\hat{Y}_{t+1}$ 은 예측치를 의미한다. RMSE의 값이 작다는 것은 실제값과 예측치의 차이

즉 예측 오차가 작다는 것을 의미하고 반대로 실제값과 예측치의 차이가 크다면 RMSE의 값은 커진다. 따라서 RMSE의 값이 작은 모형일수록 더 높은 예측력을 갖는다고 할 수 있다.

다음으로 PPL은 분포예측의 정확도를 측정하는 방법이며 사후예측밀도(Posterior Predictive Density, 이하 PPD)의 합으로 계산된다. 예를 들어  $h=1$ 인 경우 로그 PPL의 값은  $(Y_{t+1}, Y_{t+2}, \dots, Y_T)$ 의 조건부 결합예측밀도(Joint Predictive Density)를 의미한다. 먼저 PPD는 다음과 같이 정의한다.

$$f(Y_{t-1+h} | F_{t-1}, M) = \int f(Y_{t-1+h}, \theta | F_{t-1}, M) d\theta = \int f(Y_{t-1+h} | F_{t-1}, \theta, M) \pi(\theta | F_{t-1}, M) d\theta$$

여기서  $M$ 은 모형을 나타내고,  $F_{t-1}$ 은  $t-1$ 시점까지의 모든 정보를 의미한다. PPD의 계산 과정에서 적분을 계산하는 것은 매우 어렵고 대부분은 계산되지 않는다. 따라서 실질적으로 적분을 하지 않고 다음과 같이 수치적으로 근사하는 방법을 사용한다.

$$f(Y_{t-1+h} | F_{t-1}, M) = \frac{1}{n_1} \sum_{j=1}^{n_1} f(Y_{t-1+h} | F_{t-1}, \theta^{(j)}, M)$$

여기서  $n_1$ 은 반복시행 이후 burn-in을 제외한 실제 샘플링의 횟수이고  $(j)$ 는  $j$ 번째 반복시행에서 추출된 값을 나타내며  $\theta^{(j)}$ 는 사후분포로부터 추출된 값이다. 그러면 로그 PPL은 다음과 같이 계산된다.

$$\log PPL(M, h=1) = \log f(Y_{t+1}, Y_{t+2}, \dots, Y_T | F_T, M) = \sum_{t+1}^T \log f(Y_t | F_{t-1}, M)$$

$\log f(Y_t | F_{t-1}, M)$ 는  $t$ 시점의 PPD의 값이고 특정 시점의 표본 외 예측분포에 그 시점의 실제 값을 대입하여 계산된다. 그리고 실제값이 예측분포에서 높은 빈도로 나타나는 값이라면 PPD의 값은 더 커질 것이다. 따라서 PPL의 값이 큰 모형일수록 높은 예측력을 갖는다는 것을 의미한다.

## 5. 표본 외 예측

본 연구에서는 추정에 사용하는 시계열 자료를 통해 2016년 10월부터 2021년 9월까지 최근 5년의 자료를 대상으로 표본 외 자료를 예측하고 예측력을 비교하며 예측시계는 1개월부터 1년까지 고려한다. 또한 예측시계에 따라 단기·중기·장기 3가지로 구분되는데 단기는 1개월에서 4개월, 중기는 5개월에서 8개월, 장기는 9개월에서 12개월로 정의한다. 예측은 직접예측(Direct Forecasting)에 기반하는데 이는  $t$  시점까지의 정보로  $h$ 기 이후의 BDI 지수를 직접예측하는 방법이다. 직접예측은 잠재적으로 모형설정의 불확실성이 존재하고 설명변수의 확률과정이 불명확할 때 유용하다.

베이지안 변수선택 방법론은 중요하지 않은 변수의 사전분산  $b_0$ 의 사전평균을 어떻게 설정하는지에 따라 선택되는 변수가 변하는 경우가 존재한다. 만약  $b_0$ 의 사전평균값을 너무 작게 설정한다면 변수선택 과정에서 중요하지 않은 변수를 모형에서 제외시키지 못할 수 있으며, 반대로  $b_0$ 의 사전평균값을 크게 설정한다면 중요한 변수가 중요하지 않은 변수로 선택될 수 있기 때문에 변수누락 문제를 초래할 수 있다. 따라서 선택된 변수가 너무 많거나 적으면 모형의 예측력을 감소시키기 때문에 예측력을 극대화하는 최적의 하이퍼 파라미터를 찾기 위해  $\nu_{00}$ 를 1000으로 설정하는 대신  $b_0$ 의 사전평균은  $10^{-4}, 10^{-6}, 10^{-8}, 10^{-10}$ 으로 설정한다. 이를 통해 예측결과를 비교하여 최적의 사전평균을 찾고 보여준다.

최적의 예측을 위해 예측변수들의 단위조정에는 Z-score 방법과 미네소타 사전분포(Minnesota Prior) 두 방법을 적용하였다. Z-score 방법은 변수의 표본평균을 0, 표준편차는 1이 되도록 표준화하는 방법이며 미네소타 사전분포 방법은 사전분포가 지나치게 강하거나 약해지는 현상을 방지하기 위해 각 변수에  $(w_y/w_k)$ 를 곱하여 상대적으로 잡음이 많이 존재하는 변수에 작은 가중치를 부여하는 방법이다. 또한 예측변수에 내재된 BDI 지수에 대한 정보는 현재 시점의 예측변수가 아닌 예측변수의 시차값을 더 많이 포함할 가능성이 존재하

기 때문에 이러한 문제를 완화하기 위해 본 연구에서는 예측변수의 시차를 1개월부터 최대 2개월까지 고려한다. 마지막으로 윈도우 크기(Window Size)는 59개월, 71개월, 83개월, 95개월로 설정하여 각 예측시계별로 표본 외 예측에 적용하고 비교한다.

변수선택 기법을 적용하여 변수선택 과정에서 BDI 지수 예측에 중요한 예측변수가 너무 많이 선택된다면 중요한 예측변수의 예측력이 비효율적으로 추정되어 오히려 예측력을 감소시킬 수 있다. 이러한 문제를 완화하기 위해 본 연구에서는 먼저 예측변수들을 대상으로 변수선택 알고리즘을 적용한 후, 선택된 변수집합에 대해 필터링을 진행하는 방법을 활용하였다. 구체적인 추정과정 알고리즘은 Kang Kyu-Ho, Kim Jung-Sung and Shin Se-Rim (2021)을 참고하길 바란다.

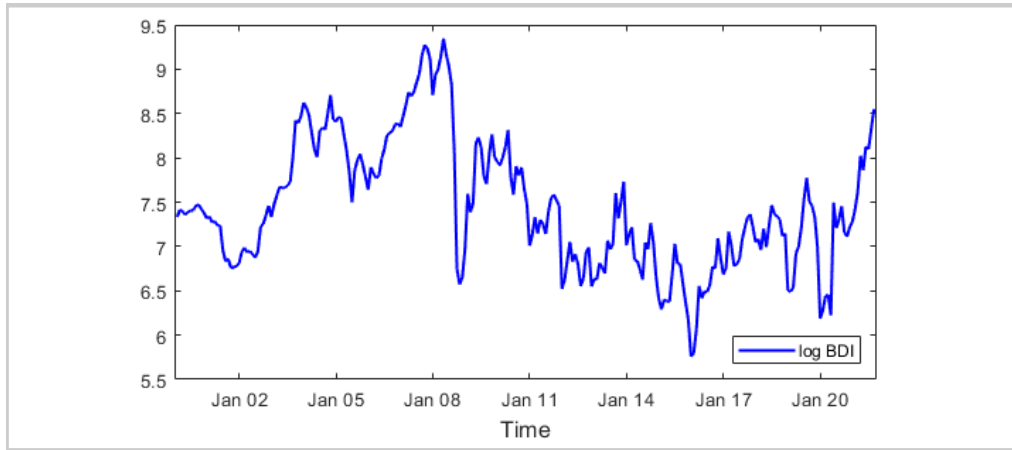
## III. 실증분석

### 1. 자료

본 연구에서 사용되는 관심변수인 BDI 지수는 월별 자료를 사용하며 해당 자료는 발틱해운거래소에서 수집하였다. <Table 1>에서는 BDI 지수의 예측을 위해 사용되는 예측변수들에 대한 출처와 구성에 대하여 요약하였다. BDI 지수 외에 총 16개의 예측변수를 사용하였고 연구에 사용되는 모든 자료는 월별 자료이며 자료의 기간은 2000년 1월부터 2021년 9월까지이다. 그리고 각각의 예측변수들은 자료의 특성을 기준으로 9개 그룹(해운시장, 외환시장, 금융시장, 금리, 원자재 가격, 생산량, 경제지표, 불확실성, 기후)으로 나누어진다.

<Fig. 1>은 종속변수인 BDI 지수의 시계열 자료를 그림으로 나타낸 것이다. 그림에서 나타나듯이 BDI 지수는 2000년 이후 중국의 급격한 경제성장과 더불어 원자재의 수요로 인해 급격하게 상승하여 2008년 5월 역대 최고치를 기록하였지만 당해 글로벌 금융위기로 인해 큰 폭의 하락을 경험하였다. 그리고 2009년 이후 BDI 지수의 상승세는 세계경제 회복기대의 증

**Fig. 1. Baltic Dry Index**



Notes: This figure shows the monthly log Baltic Dry Index from January 2000 to September 2021.

**Table 1. Predictor List**

Category	Data	Abbreviation	Source
Shipping Market	Baltic Dry Index	BDI	Baltic Exchange
	Clarksea Index	CI	Clarkson
	Bulkcarrier Fleet Development	BFD	Clarkson
Production	China Steel Production	CSP	Clarkson
Financial Market	Dow Jones Industrial Average	DJI	Yahoo Finance
	Yuan/Dollar Exchange Rate	YUD	Fred
FX Market	Yen/Dollar Exchange Rate	YED	Fred
	Euro/Dollar Exchange Rate	DE	Fred
Interest rates	London Interbank Offered Rate	LIB	Fred
	Global Price of West Texas Intermediate	WTI	Fred
Commodity Price	Global Price of Coal	CO	Fred
	Global price of Iron Ore	IO	Fred
	Global Price of Wheat	WH	Fred
Weather	Oceanic Nino Index	ONI	NOAA
Economic Indicator	Global Economic Conditions Indicator	GECI	Baumeister et al. (2020)
	University of Michigan Index of Consumer Expectations	CE	University of Michigan
Uncertainty	Economic Policy Uncertainty Index	EPU	Economic Policy Uncertainty

Notes: This table summarizes the sources and composition of the predictors.

가로 판단된다. 또한 코로나 팬데믹이 본격적으로 시작된 2020년부터 BDI 지수는 하락하였

지만 불과 1년 만에 급격하게 상승하며 높은 변동성을 보여주고 있다.



서론에서 언급했듯이 BDI 지수는 다양한 거시변수 및 금융변수의 영향을 받을 수 있다. 기존 연구에서 BDI 지수에 대해 다양한 분석을 진행하였으며 그 결과를 통해 해운시장과 관련된 변수들은 BDI 지수와 밀접한 관련이 있다는 것을 확인할 수 있다. 따라서 본 연구에서는 클락슨 지수(Clarksea Index), 신조선 발주량(Bulk Carrier Fleet Development), 그리고 중국 철강 생산량 자료를 예측변수로 포함하였다. 클락슨 지수는 선박 유형에 대한 해운사 수익성 지표이고, 신조선 발주량은 새롭게 발주되는 선박의 재화중량톤수(Dead Weight Tonnage)로 해운 시장 예측에 자주 사용되고 있다.

벌크선의 운임에 영향을 미치는 주요요인은 원자재와 곡물의 화물량이다. 즉, 원자재나 곡물의 수요가 증가한다는 것은 BDI 지수가 상승한다는 것을 의미한다. 따라서 원자재와 곡물의 가격에 따라 수요 및 공급이 결정되기 때문에 대표적인 원자재와 곡물 시장 관련 변수로 국제 유가(서부 텍사스유), 철광석 가격, 석탄 가격, 밀 가격을 예측변수로 선정하였다.

위에서 설명한 대부분 예측변수는 기존의 선행연구에서 BDI 지수와 관련이 있다는 분석결과나 예측력 향상에 도움이 된다는 근거를 토대로 선택하였다. 사실 BDI 지수는 거시경제변수, 외환 및 금융변수 등 다양한 변수의 영향을 받는다. 따라서 기존연구와 달리 본 연구에서는 추가로 경기와 관련된 변수로 글로벌 경제지표(Global Economic Conditions Indicator), 경기를 선행하는 지표인 소비자 기대지수(Index of Consumer Expectations), 미국 경제정책 불확실성 지수(Economic Policy Uncertainty Index)를 포함하였다. 경제정책 불확실성 지수는 텍스트 마이닝 방식으로 불확실성(Uncertain)을 측정한 자료이다. 국제 무역에서 국가 간의 경제협력이 긴밀해짐에 따라 한 국가의 불확실성이 다른 국가의 경제에 미치는 파급효과가 매우 큰 것으로 알려져 있다. 미국 경제정책 불확실성은 실물경제에 영향을 미치기 때문에 BDI 지수 예측에 적합한 예측변수가 될 수 있으므로 BDI 지수의 예측변수로 선정하였다. 금융 관련 변수로는 유로/달러 환율, 엔/달러 환율, 위안/달러 환율, 리보금리, 다우존

스지수를 BDI 지수에 대한 예측변수로 고려하였다. 마지막으로 ONI 지수의 상승은 남반구에 집중 호우와 홍수를 일으켜 곡물의 작황에 악영향을 줄 수 있으며 이는 곡물의 가격에 영향을 줄 수 있다. 따라서 곡물의 해상물동량 변동은 BDI 지수에 영향을 미칠 수 있으므로 ONI 지수를 예측변수로 포함하였다.

## 2. 표본 외 예측결과

〈Table 2〉은 모형별 표본 외 예측결과를 나타낸다. 앞서 설명한 바와 같이 RMSE 값이 작으면 작을수록 해당 모형이 예측에 도움이 된다는 것을 나타내고 PPL 값은 크면 클수록 높은 예측력을 갖는다는 것으로 해석할 수 있다. 패널(a)는 ADL 모형 중 예측력이 가장 우수한 모형의 예측시계별 RMSE와 PPL 및 최적 하이퍼 파라미터를 나타낸다. 패널(b)는 ADL 모형 중 예측력이 두 번째로 우수한 모형과 AR 모형의 RMSE와 PPL 및 가중치를 나타내고 있다. 우선 패널(a)의 결과에 따르면 단기 예측력이 중기 및 장기 예측력에 비해 월등히 우수하다는 것을 확인할 수 있다. 또한 예측시계가 단기에서 중기로 갈수록 예측력이 저하되는 반면, 중기에서 장기로 갈수록 예측력이 점차 개선되는 것을 확인할 수 있다. 한편 패널(b)의 결과에서 ADL 모형 중 예측력이 두 번째로 우수한 모형과 AR 모형의 표본 외 예측결과를 비교했을 때, 단기와 장기 예측에서는 ADL 모형의 예측력이 AR 모형에 비해 우수하지만 중기에서는 AR 모형의 예측력이 더 우수한 것으로 나타났다. 하지만 예측력 차이가 크게 나타나지 않아 AR 모형의 예측결과도 결코 나쁘지 않은 것을 확인할 수 있다. 또한 두 모형 역시 단기 예측력이 장기 예측력에 비해 월등히 우수하며 예측시계가 단기에서 중기로 갈수록 예측력이 저하되는 반면, 중기에서 장기로 갈수록 예측력이 점차 향상되는 현상이 나타났다. 따라서 모든 예측시계에 대해 종합적으로 고려하였을 때, 본 연구에서 제시한 베이지안 변수선택 기법을 적용한 ADL(1st) 모형의 예측력이 AR 모형에 비해 우수한 것으로 보아 BDI 지수에 영향을 미치는 정보변수들이 존재하며 예측시계별로 BDI 지

**Table 2.** Out-of-sample Forecasting Performance Comparison

## (a) The Best ADL Model

1st ADL							
H	RMSE	InPPL	$J$	$b_0$	Size	Norm	Weight
1	0.272	-8.610	1	$10^{-8}$	83	Mine	0.690
2	0.369	-25.780	2	$10^{-10}$	71	Z-score	0.607
3	0.433	-35.849	1	$10^{-8}$	83	Z-score	0.863
4	0.498	-43.786	2	$10^{-8}$	83	Mine	0.908
5	0.580	-53.243	1	$10^{-8}$	83	Mine	0.386
6	0.572	-52.263	2	$10^{-8}$	95	Z-score	0.801
7	0.588	-53.877	2	$10^{-10}$	59	Mine	0.387
8	0.563	-51.421	2	$10^{-8}$	71	Z-score	0.639
9	0.563	-51.236	1	$10^{-8}$	83	Z-score	0.387
10	0.537	-48.746	2	$10^{-10}$	59	Mine	0.416
11	0.505	-46.531	1	$10^{-10}$	71	Mine	0.803
12	0.489	-46.301	1	$10^{-10}$	83	Mine	0.759

(b) 2<sup>nd</sup> Best ADL and AR Model

2nd ADL				AR		
H	RMSE	InPPL	Weight	RMSE	InPPL	Weight
1	0.280	-9.939	0.183	0.280	-10.300	0.127
2	0.376	-26.900	0.198	0.374	-26.915	0.195
3	0.456	-38.555	0.058	0.452	-38.238	0.079
4	0.521	-46.642	0.052	0.522	-46.928	0.039
5	0.570	-53.695	0.246	0.576	-53.288	0.369
6	0.582	-53.995	0.142	0.595	-54.903	0.057
7	0.591	-54.382	0.234	0.588	-53.899	0.379
8	0.583	-53.683	0.067	0.572	-52.198	0.294
9	0.571	-52.408	0.120	0.558	-50.992	0.493
10	0.538	-49.025	0.314	0.536	-49.177	0.270
11	0.523	-48.519	0.110	0.525	-48.753	0.087
12	0.502	-47.618	0.203	0.520	-49.311	0.037

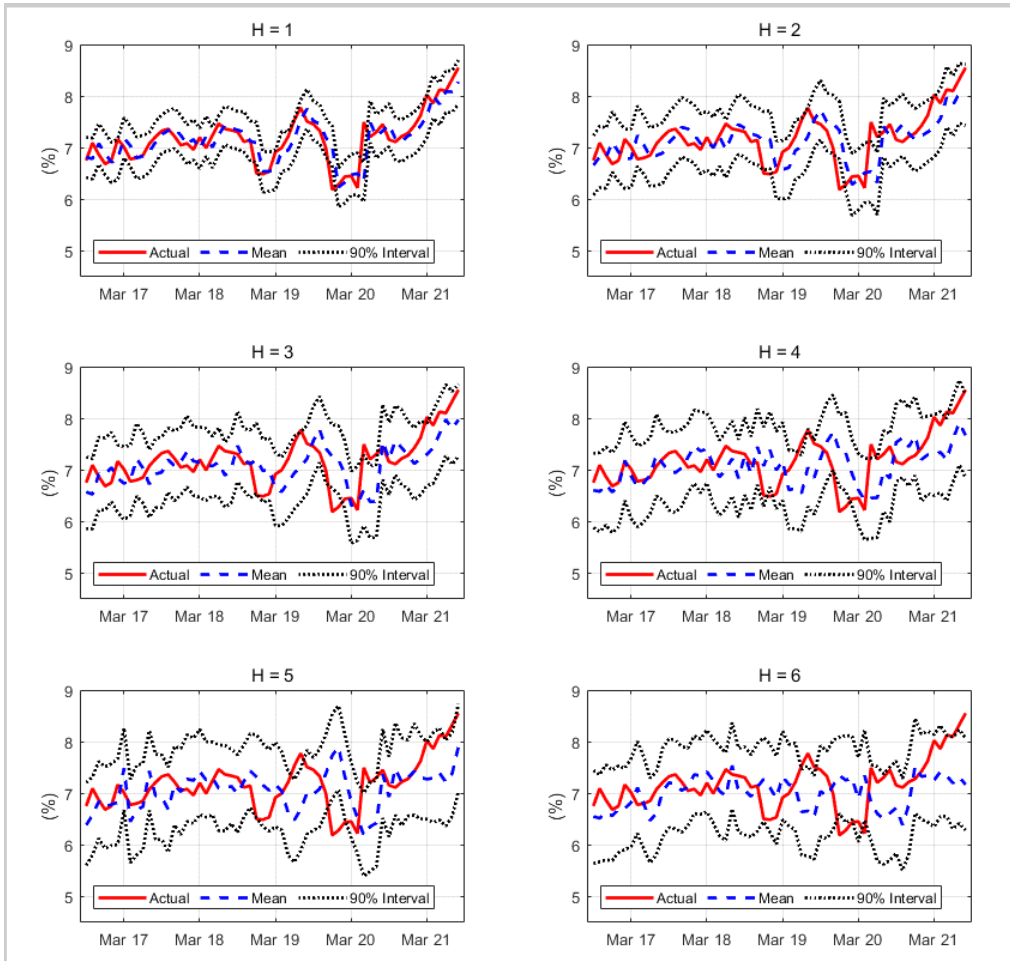
Notes: This table reports the RMSE and PPL of the model. Panel (a) reports the RMSE and PPL of the best ADL model along with the best hyper-parameters. Panel (b) reports the RMSE and PPL of the 2nd best ADL and AR models. InPPL is log PPL,  $J$  is the lag,  $b_0$  is the exponent of the prior mean of  $b_0$ , and Size is the window size. Norm is the normalization method. Weight is the model weight based on the PPL.

수 예측에 중요한 정보변수를 선택하는 것이 예측력 향상에 매우 중요하다는 것을 알 수 있다.

예측시계별 표본 외 예측결과를 RMSE와 PPL을 기준으로 비교해보면, 단기 예측에서 단

일변수 시계열 예측 모형인 AR 모형은 우수한 예측력을 보여주었지만 ADL(1st) 모형의 예측력이 더 우수한 것으로 나타났다. 예측시계가  $h=1$ 인 경우, AR 모형의 예측력은 ADL 모형에

Fig. 2. Out-of-sample Prediction for H=1~6



Notes: This figure plots the H-period-ahead out-of-sample forecasts generated from the best ADL model. The red solid line is the realized BDI, the blue dashed line is the posterior predictive mean, and dotted lines are 90% credibility intervals.

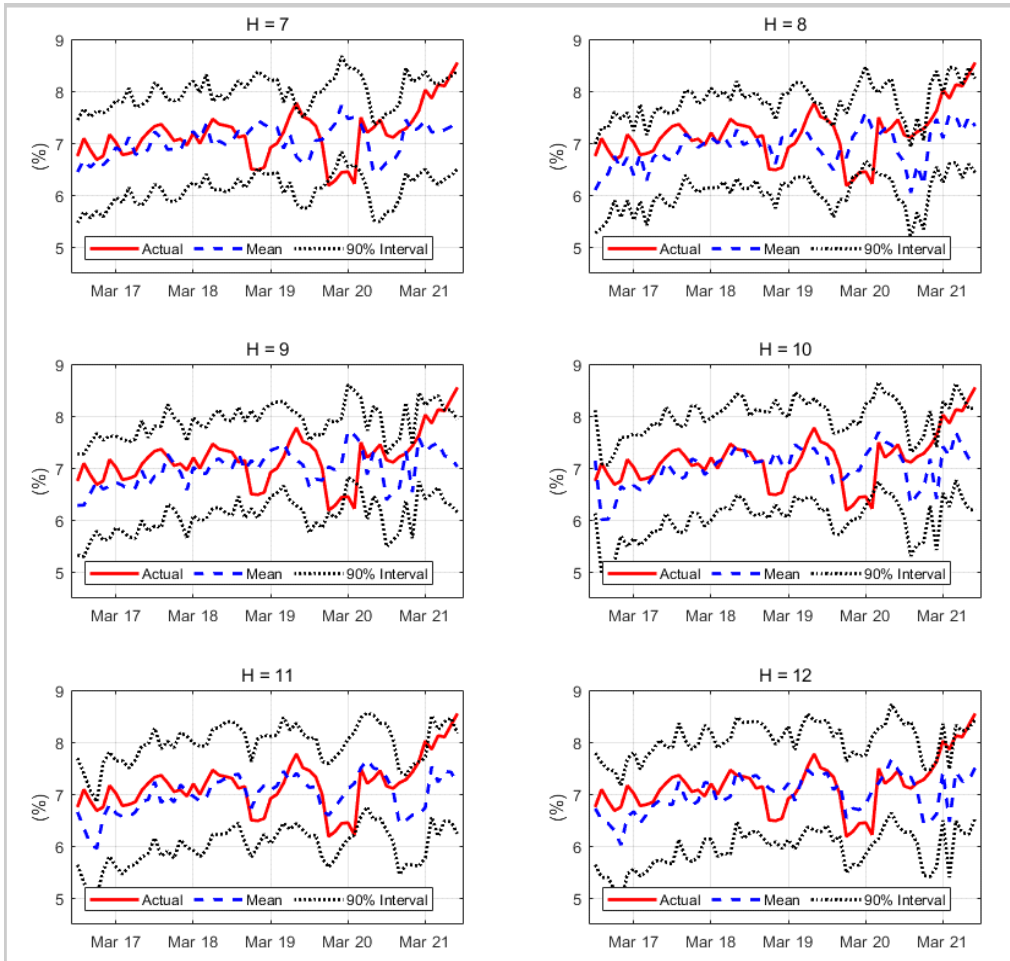
비해 RMSE 기준 2.9%로 감소한 것으로 나타났지만 PPL 은 19.6% 감소한 것으로 나타났다. 또한 ADL(2nd) 모형보다 PPL 기준 3.3% 감소한 것으로 나타났다.

본 연구의 목적상 분포예측이 정확해야하며 점예측과 달리 분포예측은 변동 가능폭을 예측할 수 있고 위험관리 측면에서 점예측에 비해 상대적으로 유용하기 때문에 PPL을 사용하여 예측력을 평가하는 것이 바람직하다.

중기 예측의 경우, 두 ADL 모형의 예측력이

모두 AR 모형에 비해 우수한 것을 <Table 2>을 통해 확인할 수 있다. 예측시계가  $h=6$ 인 경우, AR 모형의 예측력은 ADL(1st) 모형에 비해 RMSE 기준 4% 감소한 것으로 나타났지만 PPL 은 5% 감소한 것으로 나타났다. 하지만 <Fig. 2>와 <Fig. 3>에 나타나듯 ADL 모형의 예측력은 중기로 갈수록 급격하게 악화되는데 중기 ( $h=6$ ) 예측과 단기( $h=1$ ) 예측결과를 비교해보면 예측력이 많이 차이나는 것을 확인할 수 있다. 따라서 중기 예측에서는 본 연구에서 제시

Fig. 3. Out-of-sample Prediction for H=7~12



Notes: This figure plots the H-period-ahead out-of-sample forecasts generated from the best ADL model. The red solid line is the realized BDI, the blue dashed line is the posterior predictive mean, and dotted lines are 90% credibility intervals.

한 모형에 추가적인 예측변수를 고려하여 예측력을 향상시켜야 하는 것으로 판단된다.

장기 예측에서도 마찬가지로 두 ADL 모형의 예측력이 AR 모형에 비해 우수한 것으로 나타났다. 예측시계가  $h=12$ 인 경우, AR 모형의 예측력은 ADL(1st) 모형에 비해 RMSE 기준 6.3%로 감소한 것으로 나타났지만 PPL은 6.5% 감소한 것으로 나타났다. 하지만  $h=9$ ,  $h=10$ 인 경우, 분포예측에서 ADL(1st) 모형의 예측력이 더 우수하지만 점예측에서는 AR 모형의 예측력

이 더 우수한 것으로 나타났다. 즉, RMSE가 작다고 해서 PPL이 높은 것은 아니며 RMSE는 PPL에 비해 꼬리 예측정확도를 과소평가하는 성질이 있기 때문이다. 거듭 강조컨대 연구 목적상 분포예측이 정확해야하며 위험관리 측면에서 예측시계가  $h=9$ ,  $h=10$ 인 경우 역시 ADL 모형이 AR 모형보다 낫다고 판단된다. 따라서 모든 예측시계를 종합적으로 고려하였을 때 단일변수 시계열 모형인 AR 모형에 비해 본 연구에서 제시한 모형의 예측력이 더 우수한

**Table 3.** Selected Variables: ADL models

H	ADL(1st)	ADL(2nd)
1	const, BDI(-1),	const, BDI(-1)
2	const, BDI(-1),	const, BDI(-1)
3	const, BDI(-1), GECI(-1), ONI(-1), BFD(-1)	const, BDI(-1), CI(-1)
4	const, BDI(-1), DE(-1), DE(-2), YED(-1), YED(-2), LIB(-1), CE(-1), WTI(-1), WTI(-2), WH(-1), CSP(-1), CI(-2)	const, BDI(-1)
5	const, BDI(-1), EPU(-1), IO(-1)	const, BDI(-1), WTI(-1), CI(-1)
6	const, BDI(-1), BDI(-2), CO(-1), WH(-1), LIB(-2), WTI(-2), IO(-2), CI(-2)	const, BDI(-1), YUD(-1), CECI(-1), EPU(-1)
7	const, BDI(-1)	const, BDI(-1)
8	const, BDI(-1), BDI(-2), YUD(-1), DE(-1), WTI(-1), ONI(-1), YED(-2), CO(-2), IO(-2), CI(-2)	const, BDI(-1), EPU(-1), ONI(-1)
9	const, BDI(-1), IO(-1), CI(-1)	const, BDI(-1), BDI(-2), YUD(-1), DE(-1), YED(-1), EPU(-1), WTI(-1), CO(-1), IO(-1), ONI(-1)
10	const, BDI(-1), BDI(-2)	const, BDI(-1)
11	const, BDI(-1)	const, BDI(-1)
12	const, BDI(-1)	const, BDI(-1)

Notes: This table presents the selected variables across the forecasting horizons. ADL(1st) and ADL(2nd) indicate the best and second best models, respectively, in terms of the RMSE. const. is the intercept term.

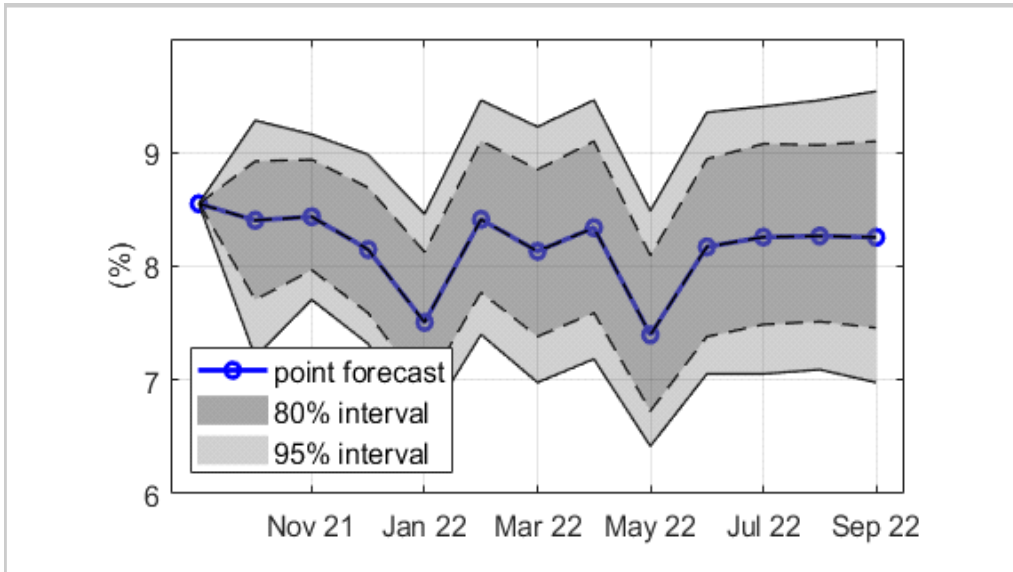
것으로 보아 BDI 지수에 영향을 미치는 다양한 정보변수를 고려한 모형이 오히려 BDI 지수 예측력 향상에 더 도움이 된다는 것을 알 수 있다. Bae et al.(2021)의 연구에서도 BDI 지수 예측에 있어 단일변수 시계열 모형의 예측력이 다양한 정보변수를 고려한 모형보다 낮은 것으로 나타났다. 또한 Kim et al.(2019)은 단일변수 시계열 모형을 이용하여 BDI 지수를 예측하는 것은 모형 특성상 용이성으로 인해 실무 적용이 쉽다는 장점이 존재하나 해운시장의 추세와 글로벌 금융위기, 국제 유가급등, 코로나 팬데믹 등 외부적 충격을 반영하지 못하는 한계가 있다고 지적하였다.

〈Fig. 2〉와 〈Fig. 3〉은 예측시계별로 ADL(1st) 모형으로부터 추정된 표본 외 예측결과를 그림으로 나타낸 것이며 파란색 점선은 사후 예측 분포의 평균, 빨간색 실선은 실제 데이터, 점선은 90% 신용구간을 나타낸다. 〈Table 2〉의

ADL(1st) 모형 예측결과에서 단기 예측력이 중기 및 장기 예측력에 비해 월등히 우수한 것을 〈Fig. 2〉와 〈Fig. 3〉의 비교를 통해 확인할 수 있다. 비록 중기 예측력은 기타 예측시계에 비해 상대적으로 저하된 것으로 나타났지만 실제 값이 90%의 신용구간을 크게 벗어나지 않아 분포예측에서는 불확실성이 적절히 예측된 것으로 판단된다. 특히 단기와 장기 예측에서는 2019년 초반부터 급격하게 상승하여 중반부터 급격하게 하락한 BDI 지수를 정확하게 예측하였음을 확인하였다. 또한 장기 예측결과가 중기 예측결과에 비해 보다 정확한 것을 확인할 수 있는데, 이러한 현상은 예측시계별로 선택된 예측변수의 수가 다르며 BDI 지수에 대한 정보가 누적되어 예측변수 및 모형의 가중치가 베이지안 업데이트를 통해 학습되면서 예측력이 향상된 것으로 판단된다.

〈Table 3〉는 ADL 모형을 추정된 결과, 각

Fig. 4. Posterior Predictive Distribution: ADL(1st)



Notes: This figure plots the H-quarter-ahead posterior predictive distributions for  $H=1,2,\dots,$  and 12. The blue solid line is the posterior mean, the black dashed and solid lines are 80% and 95% credibility intervals, respectively.

예측시계별로 선택된 예측변수를 나타낸다. 본 연구에서는 베이지안 변수선택 기법을 적용하여 BDI 지수의 분포예측을 시행할 뿐만 아니라 각 예측시계에 따라 어떠한 예측변수들이 선택되었는지 해석할 수 있게 하였다. <Table 3>에 나타나듯이 예측시계에 따라서 선택된 예측변수가 상이하고 일정하지 않으며 상대적으로 단기와 장기 예측에서 적은 수의 예측변수가 선택되는 반면, 중기 예측에서는 보다 많은 예측변수가 선택되는 것으로 나타났다. 특히 단기 예측에서는 전기의 BDI 지수가 중요하고 중기 예측에서는 BDI 지수, 철광석 가격, 석탄 가격, 밀 가격, 국제유가, 클락슨 지수, 리보금리가 중요한 예측변수로 선택되었다. 장기 예측에서는 BDI 지수, 클락슨 지수, 철광석 가격이 BDI 지수 예측에 중요한 예측변수로 작용하였다. 결과적으로 단기와 장기 예측에서는 주로 전기의 BDI 지수가 중요한 예측변수로 선택된 반면, 중기 예측에서는 원자재 가격에서의 예측변수가 모두 선택된 것으로 보아 BDI 지수의 중기 예측에서는 원자재 가격의 영향을 많이 받는

것으로 판단할 수 있다.

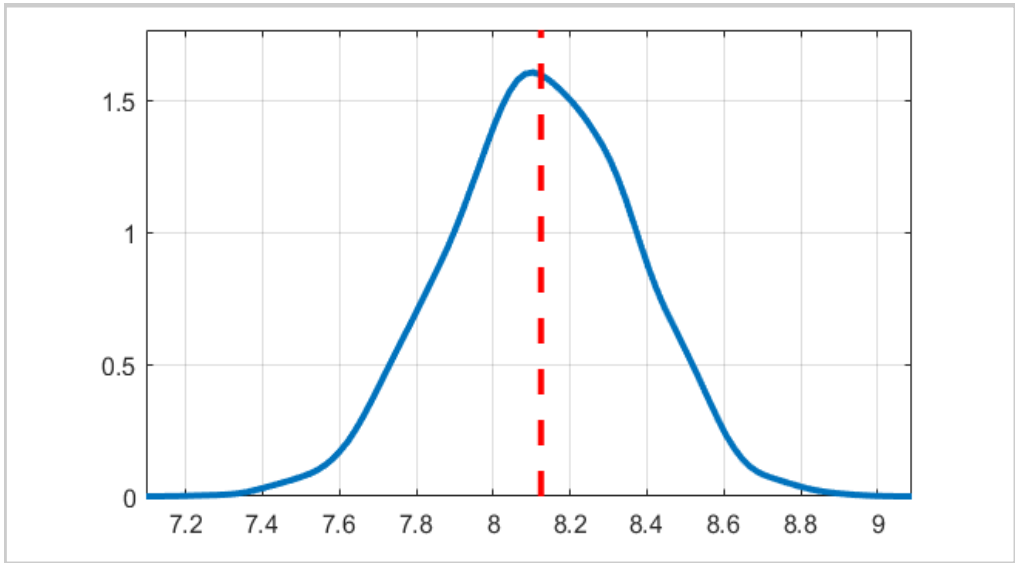
## 2. 실제 예측결과

<Fig. 4>는 2021년 9월 시점에서 향후 1년 동안 BDI 지수의 사후 예측분포를 그림으로 나타낸 것으로 파란색 실선은 사후예측분포의 평균, 점선은 80% 신용구간, 실선은 95% 신용구간을 나타내며 BDI 지수의 실제 예측결과를 요약하면 다음과 같다.

첫째, BDI 지수는 2022년 9월까지 8% 내외의 수준을 유지할 것으로 예측되었다. 다만 2022년 초반까지 차츰 감소하여 2022년 1월은 7%대로 하락하지만 2월은 8%대로 상승하면서 4월까지 유지되고 5월에 다시 7%대로 하락하고 그 이후에는 8%대로 상승할 것으로 예상된다. 따라서 최근 BDI 지수의 급격한 상승세는 일시적인 현상으로 추정되고 향후 1년 동안 8% 내외의 수준을 유지할 것으로 예상된다.

둘째, 예측시계가 커질수록 신용구간의 폭이 점차 넓어지면서 불확실성도 동시에 증가한다

Fig. 5. Medium-term BDI Forecast



Notes: This figure plots the posterior predictive distribution of the average log BDI over the period, 2022:M4 to 2022:M9.

는 점에서 반드시 유의할 필요가 있다. 신용구간은 예측치의 신뢰도를 나타내며 동시에 불확실성의 크기로 사용된다. 이는 실현치가 신용구간의 범위에 위치할 가능성을 반영하기 때문이다. 예측시계에 따라 다르게 나타났는데, 2021년 후반까지 95% 신용구간의 폭은 2%p가 안 되지만 2022년 초반부터 2%p를 초과한다. 특히, 2022년 9월의 예측치는 8%를 조금 상회하지만 95% 신용구간으로부터 최대 9% 이상으로 상회할 가능성도 배제해서는 안 된다. 따라서 BDI 지수의 불확실성을 과소평가하게 되면 해운금융시장뿐만 아니라 여러 경제주체 이해관계자의 의사결정과 경영성과에 상당한 영향을 미친다. 예를 들어, 해상운임 변동 위험에 대비하여 국적선사가 BDI 지수를 예측하여 파생금융 상품인 해상운임선도거래를 활용한다면 자체 예측역량에 대한 과도한 신뢰로 인해 운임가격을 미래 운임가격보다 낮은 가격으로 화주와 계약을 성사시켜 선사의 입장에서는 결국 손실이 발생하게 된다.

(Fig. 5)는 ADL(1st) 모형의 실제 예측결과를 통해 중기 시계 BDI 지수의 사후분포를 그

림으로 나타낸 것이며 현재 시점에서 1년 후 2022년 4월부터 2022년 9월까지 6개월 동안 평균 BDI 지수의 사후분포를 산출하였다. 사후 평균은 대략 8.1 수준이고 90% 신용구간은 [7.7, 8.5]로 나타났다. 따라서 해운금융시장 및 여러 경제주체 이해관계자는 BDI 지수의 예측력 향상을 위해 예측치뿐만 아니라 실현 가능한 범위를 산출하여 현재 시점에서 보다 정확한 의사결정을 할 수 있기를 기대한다.

#### IV. 결론

본 연구에서는 해운시장의 경기상황을 나타내는 중요한 지표인 BDI 지수의 예측에 목적을 두고 있다. 따라서 본 연구에서는 BDI 지수의 예측을 위해 베이지안 변수선택 기법을 활용하여 BDI 지수 분포예측을 시행하고 분석하였으며 예측력을 비교 및 평가하였다. BDI 지수의 월별 자료를 이용하여 예측을 시행한 결과를 요약하면 다음과 같다.

첫째, 베이지안 변수선택 기법을 적용한 ADL(1st) 모형의 예측력이 모든 예측 시계에서 AR 모형에 비해 우수한 것으로 나타났다. 특히 단기 예측력이 증기 및 장기 예측력에 비해 월등히 우수한 것으로 나타났다. 이러한 연구 결과는 BDI 지수 예측에 영향을 미치는 다양한 정보변수들이 존재하며 예측 시계별로 예측에 중요한 정보변수를 선택하는 것이 예측력 향상에 매우 중요하다는 것을 알 수 있다. 둘째, ADL(1st) 모형의 각 예측시계별로 선택된 예측변수를 살펴보면 단기 예측에서는 전기의 BDI 지수가 중요한 변수로 선택되었고 증기 예측에서는 BDI 지수, 철광석 가격, 석탄 가격, 밀 가격, 국제유가, 클락슨 지수, 리보금리가 중요한 예측변수로 선택되었으며 장기 예측에서는 BDI 지수, 철광석 가격, 클락슨 지수가 BDI 지수 예측에 중요한 예측변수로 작용하는 것으로 나타났다. 특히 단기와 장기 예측에서는 BDI 지수가 핵심 변수로 선택되었지만, 증기 예측에서 원자재 가격에서의 예측변수들이 모두 선택된 것으로 보아 BDI 지수의 증기 예측에서는 원자재 가격의 영향을 많이 받는 것으로 판단된다. 셋째, BDI 지수는 현재 시점에서 향후 1년 동안 8% 내외의 수준을 유지하고 2022년 1월과 5월은 7%대로 하락할 것으로 예측되었다. 하지만 현재 시점에서 과거의 BDI 지수를 살펴보면 2022년 1월부터 5월은 급격하게 상승한 구간으로 나타나 본 연구에서의 실제 예측결과와 다르게 나타났다. 이는 금융위기와 COVID-19 이

후 BDI 지수의 급격한 변화는 BDI 지수 예측에 있어 구조변화 가능성을 고려하는 것이 중요하지만 본 연구에서는 이를 명시적으로 모형 내에 도입하지 않았기 때문이라고 생각된다. 앞으로 보다 많은 자료가 지속적으로 업데이트되면서 구조변화 도입을 통해 BDI 지수의 예측정확도를 개선할 수 있을 것으로 기대한다.

본 연구에서 제시한 모형 특성상 여러 한계점이 존재한다. 베이지안 변수선택 기법을 적용한 모형에 예측변수의 선택 여부를 선형적으로 결정하는 데 따르는 문제들을 완화하는 점에서 바람직한 추정 방법이 될 수 있다. 또한 예측력 면에서도 가장 우수하다는 점에서 추정의 정확성을 개선시키는 데 도움이 된다는 것을 알 수 있다. 하지만 특정 예측변수가 추정과정에 선택되더라도 BDI 지수와 인과관계는 존재하지만, 구체적으로 BDI 지수에 어떤 영향을 미치는지 해석할 수 없으며 구체적인 근거를 파악할 수 없다. 또한 높은 변동성을 지닌 BDI 지수의 특성상 본 연구를 해당 실무에 적용하기 위해서는 세밀한 작업이 추가로 필요하다. 추후 연구과제로서 BDI 지수 예측에 도움이 되는 정보변수를 찾고 동시에 세밀한 변수 선택을 적용하여 더욱 깊이 있는 연구가 이루어질 필요가 있다. 앞으로 BDI 지수 예측에 도움이 되는 시계열 자료가 확보된다면 향후 지속적인 후속 연구를 통해 앞서 언급한 한계점을 보완하는 과정에서 BDI 지수의 예측력을 향상시킬 수 있을 것으로 기대한다.

## References

- Bae, Sung-Hoon, Ha, Young-Mok, Park, Keun-Sik (2018). "An Empirical Study on the Effect of the Factors Influencing on the Dry Bulk Freight Rate." *Korea Logistics Review*, 28(5), 117-132.
- Bae, Sung-Hoon, Lee, Gun-Woo Park, Keun-Sik (2021). "A Baltic Dry Index Prediction using Deep Learning Models." *Journal of Korea Trade*, 25(4), 17-36.
- Baumeister, C., D. Korobilis and T. K. Lee (2020). "Energy markets and global economic conditions." *The Review of Economics and Statistics*, 1-45.
- Bildirici, M. E., F. Kayıkçı and I. Ş. Onat (2015). "Baltic Dry Index as a major economic policy indicator: the relationship with economic growth." *Procedia-Social and Behavioral Sciences*, 210, 416-424.



- Cullinane, K. P. B., K. J. Mason and M. Cape (1999). "A comparison of models for forecasting the Baltic freight index: Box-Jenkins revisited." *International journal of maritime economics*, 1(2), 15-39.
- George, E. I. and R. E. McCulloch (1997). "Approaches for Bayesian variable selection." *Statistica sinica*, 339-373.
- Kang, Kyu-Ho, Kim, Jung-Sung and Shin Se-Rim (2021). "Forecasting Korean CPI Inflation." *Economic Analysis*, 27(4), 1-42.
- Kim, Bu-Kwon, Kim, Dong-Yoon and Choi, Ki-Hong (2019). "Analysis of dependency structure between international freight rate index and crude oil price." *Journal of Korea Port Economic Association*, 35(4), 107-120.
- Kim, Chang-Beom (2011). "The Effects of International Finance Market Shocks and Chinese Import Volatility on the Dry Bulk Shipping Market." *Journal of Korea Port Economic Association*, 27(1), 263-280.
- Kim, Hyun-Sok and Chang, Myung-Hee (2014). "Bayesian VAR Analysis of Dynamic Relationships among Shipping Industry, Foreign Exchange Rate and Industrial Production" *Journal of Korea Port Economic Association*, 30(2), 77-92.
- Kim, Hyeong-Jun, Ryu, Doo-Jin and Cho-Hoon (2019). "Short-term Forecasts of the Baltic Dry Index (BDI) Using Time-series Factor Decomposition." *Korean Management Review*, 48(3), 715-731.
- Kim, Young-Min and Lee, Seo-Jin (2020). "Exchange rate predictability: A variable selection perspective." *International Review of Economics & Finance*, 70, 117-134.
- Koo, Byung-Soo (2020). "Estimation of the Korean Yield Curve via Bayesian Variable Selection." *Economic Analysis*, 26(1), 84-132.
- Lee, Chang-Hoon, Kang, Kyu-Ho, Ann, Ji-Hee (2020). "A Bayesian Variable Selection Method for Seoul Apartment Price Index Prediction." *The Korean Journal of Economic Studies*, 68(1), 153-190.
- Lee, Sung-Yhun and Ahn, Ki-Myung (2018). "Study on the Forecasting and Effecting Factor of BDI by VECM." *Korean Institute of Navigation and Port Research*, 42(6), 546-554.
- Lin, A. J., H. Y. Chang and J. L. Hsiao (2019). "Does the Baltic Dry Index drive volatility spillovers in the commodities, currency, or stock markets?." *Transportation Research Part E: Logistics and Transportation Review*, 127, 265-283.
- Mo, Soo-Won (2010). "Forecasts of the BDI in 2010-Using the ARIMA-Type Models and HP Filtering" *Journal of Korea Port Economic Association*, 26(1), 222-233.
- Ruan, Q., Y. Wang, X. Lu, and J. Qin, (2016). "Cross-correlations between Baltic Dry Index and crude oil prices." *Physica A: Statistical Mechanics and its Applications*, 453, 278-289.
- Tsioumas, V, S. Papadimitriou, Y. Smirlis and S. Z. Zahran (2017). "A novel approach to forecasting the bulk freight market." *The Asian Journal of Shipping and Logistics*, 33(1), 33-41.
- Xu, J. J., T. L. Yip and P. B. Marlow (2011). "The dynamics between freight volatility and fleet size growth in dry bulk shipping markets." *Transportation research part E: logistics and transportation review*, 47(6), 983-991.
- Zeng, Q., C. Qu, A. K. Ng and X. Zhao (2016). "A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks." *Maritime Economics & Logistics*, 18(2), 192-210.