Information Systems and Economics (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2025.060217
ISSN 2523-6407 Vol. 6 Num. 2

Grasping the Trend of the Shipping Market: a Baltic Dry Index Prediction Method Based on Deep Learning

Haoxiang Zhang[1,*]

[1]Cushing Academy, 39 School St, Ashburnham, MA 01430, United States
[*]Corresponding author: hazhang26@cushing.org

Keywords: BDI Index, Artificial Intelligence, Deep Learning, Gated Recurrent Unit Neural Network

Abstract: The Baltic Dry Index (BDI) is an international benchmark used to gauge the level of shipping freight rates for dry bulk goods and is an important measure of both the generalised economic consequences relative to the international shipping industry. The non-stationarity and strong non-linearity involved creates a huge forecasting challenge. The present study is based on a new forecasting framework based on a Gated Recurrent Unit (GRU) deep learning forecasting model based on a gating mechanism that provides for long run non-linear dependence in time series data. A data set exists comprising a total of 90 variables from (1988-2024): e.g. Supply Related (fleet capacity), Demand Related (prices for various raw materials) and Macro-Economic. The relevant model was trained and eventually compared with a number of current econometric and classical machine models, e.g., SVR, Random Forests etc., comparison being effected over different horizons using Mean Squared Error (MSE), Mean Absolute Error (MAE) etc., as being relevant for forecasting performance. The results show that the model presented consistently and significantly outperformed the existing benchmark models. Hence the study shows the capability and validity of the deep learning concept. Its ability to extract features hierarchically leads to latent features that exist, and which are not easily detectable by mainstream statistical models. Thus, it is a very powerful tool for conducting the analysis of the shipping market.

1. Introduction

The Baltic Dry Index (BDI) is a well-known barometer of the status of the overall world economy and a continual source of information regarding international trade and the maritime transportation business [1]. The BDI is an important leading indicator of world economic activity because it measures spot freight rates of dry bulk trading commodities [2]. However, due to the complexity and difficulty in forecasting this particular index, it will not be attempted herein. There are highly non-stationary, non-linear and volatile characteristics in the index [3]. These result from a complicated dynamical system, which is influenced by various interrelated factors, such as commodity demand, fleet supply, and macro-economic variables. Because of this fact, accurate forecasts are very important for maritime transportation enterprises and institutional investors.

Ordinary time series methods of forecasting have been widely used in the forecasting of the BDI, and classical econometric models [4]. Although the moving averages and simple linear regression type of methods have given some insights, they are not found efficient in practice [5]. The simple architecture of the above methods does not lend itself well to treating the deep and complicated serial relationships, and possibly hidden regularities in the noisy BDI figures [6]. Therefore, these models do not adequately capture the sudden changes and complex cyclical behavior of the index. This raises the need for a stronger approach to forecasting procedures [7].

In view of the above factors, this paper presents a forecasting method based on deep learning procedures. The Gated Recurrent Unit (GRU) artificial neural network method of forecasting is employed herein as it is an advanced structure specially designed to handle sequential data and capture long range dependencies effectively. This research makes two main contributions to the existing literature; (1) the forecasting model is developed from a comprehensive heterogeneous database of 90 different factors affecting supply side, demand side and macro-economic forecasting. (2) Empirical comparisons to show that the GRU method give higher forecast accuracy for different time horizons than classical machine learning models. This thus provides a more robust and effective tool for the analysis of the shipping market. Since 20000 separate BDI forecast values can be calculated in a reasonable time, we show full forecasts for one period using the forecasting model.

2. Methodology

2.1. Data

After analyzing the potential influencing factors of the BDI index, the data used in this study mainly includes three categories: the supply side of transportation capacity, the demand side of transportation capacity, and macroeconomic factors.

On the supply side of shipping capacity, factors related to global shipping fleet capacity are mainly considered [8]. To describe the supply of transportation capacity in the global market, this study mainly considers the number of global dry bulk fleets, global dry bulk fleet deadweight tonnage, global container fleet deadweight tonnage, global dry bulk ship order quantity, global container ship order quantity, global dry bulk cargo ship contract quantity, global container ship contract quantity, global dry bulk ship delivery quantity, global container ship delivery quantity, global dry bulk ship dismantling quantity, and global container ship dismantling quantity. In addition, the price of ship construction or transfer, whether it is a new or second-hand ship, will also affect the supply of transportation capacity. For this purpose, factors such as 5-year ship age, 10-year ship age, price index of newly built ships, and total sales of second-hand ships in the data source were also included. In addition, considering the internal correlation of the shipping market, the factors collected also include the freight rate index of container ships. In summary, the relevant data on the supply side of these transport capacities all come from the Clarksons database, a professional shipping database.

On the demand side of transportation capacity, the influencing factors should mainly reflect the transportation demand for global dry bulk cargo [9]. This analysis mainly considers several major commodities such as iron ore, coal, and grains, among which grains mainly include corn, wheat, and soybeans. The futures prices of these commodities will have a significant impact on transportation demand. Specifically, futures prices, import and export volumes, and other data for commodities such as iron ore, thermal coal, corn, wheat, and soybeans are used.

Macroeconomic factors are important factors affecting the trend of the BDI index [10]. This study collected factors such as global GDP, US dollar index, US guaranteed overnight financing rate, London Interbank Offered Rate, and fuel prices [11].

Table 1 Influencing factors of BDI index.

No.	Influencing Factor	No.	Influencing Factor	No.	Influencing Factor
1	Futures closing price: CBOT corn	31	China: Iron Ore Spot Price Index (CSI): Imported lump ore	61	Total_Orderbook_All_Bulk_num
2	Futures closing price: iron ore	32	China: Import Quantity: Soybeans	62	Total_Orderbook_Containerships_num
3	Futures settlement price: coking coal	33	China: Import Quantity: Coal and lignite	63	Total_Orderbook_Bulkers_dwt
4	dollar index	34	China: Import quantity: coking coal	64	Total_Orderbook_All_Bulk_dwt
5	SOFR	35	China: Import quantity: thermal coal	65	Total_Orderbook_Containerships_dwt
6	SOFR term interest rate: 1 month	36	China: Import quantity: Coal	66	TotalDeliveries_Bulkers_num
7	SOFR term interest rate: 6 month	37	China: Import quantity: Corn	67	TotalDeliveries_All_Bulk_num
8	SOFR term interest rate: 3 month	38	China: Import quantity: Refined copper	68	TotalDeliveries_Containerships_num
9	SOFR term interest rate: 12 month	39	China: Import quantity: Wheat	69	TotalDeliveries_Bulkers_dwt
10	Henry Hub (CMOTC)	40	China: Import quantity: coking coal	70	TotalDeliveries_All_Bulk_dwt
11	Henry Hub (CME)	41	China: Import quantity: iron ore	71	TotalDeliveries_Containerships_dwt
12	Closing price: Coal Index	42	Shanghai Import Container Freight Index: Corporate Index	72	Total_Demolition_Bulkers_num
13	China GDP	43	China Import Container Freight Index: Composite Index	73	Total_Demolition_All_Bulk_num
14	United States GDP	44	LIBOR:3 months	74	Total_Demolition_Containerships_num
15	Eurozone GDP	45	LIBOR:12 months	75	Total_Demolition_Bulkers_dwt
16	Japan GDP	46	LIBOR:1 month	76	Total_Demolition_All_Bulk_dwt
17	Real GDP growth rate: Global	47	LIBOR:6 months	77	Total_Demolition_Containerships_dwt
18	China: Average unit price of imports: Soybeans	48	Total_World_Fleet_Bulkers_num	78	Earnings
19	Thermal Coal Price Index (TPI): Import 5870	49	Total_World_Fleet_All_Bulk_num	79	4tc_t
20	Thermal Coal Price Index (TPI): Import 4770	50	Total_World_Fleet_Containerships_num	80	5tc_t
21	China: Average unit price of imports: Coal	51	Total_World_Fleet_Bulkers_dwt	81	10tc
22	China: Average unit price of imports: Coking coal	52	Total_World_Fleet_All_Bulk_dwt	82	10tc_k
23	China: Average unit price of imports: Thermal coal	53	Total_World_Fleet_Containerships_dwt	83	Panamax(yrs
24	China: Import Price Index: J04: Corn	54	Total_Orderbook_Bulkers_num	84	Handysize 10 yrs
25	China: Landed dutiable price: Imported corn	55	Total_Orderbook_All_Bulk_num	85	Supramax 5yrs
26	China: Import Price Index: J04: Corn	56	Total_Orderbook_Containerships_num	86	Supramax 10yrs
27	China: Landed dutiable price: Imported corn	57	Total_Orderbook_Bulkers_dwt	87	Supramax 5yrs
28	China: Average unit price of imports: Wheat	58	Total_Orderbook_All_Bulk_dwt	88	Panamax 10yrs
29	China: Average unit price of imports: Copper ore and its concentrates	59	Total_Orderbook_Containerships_dwt	89	Capesize 5yrs
30	China: Iron Ore Spot Price Index (CSI): Imported Fines Ore	60	Total_Contracting_Bulkers_num	90	Capesize 10yrs

By summarizing the influencing factors from the above three aspects, this work ultimately identified 90 factors that may affect the trend of the BDI index as features of its predictive model. For all features, data were collected between October 19, 1988, and August 13, 2024, with a sample size of 8988. Table 1 shows the potential influencing factors identified in this study.

2.2. Gated Recurrent Unit Neural Network Method

Recurrent Neural Networks (RNNs) are a class of neural networks specifically designed to process sequential data, capable of capturing temporal dependencies within such data. However, traditional RNNs often suffer from vanishing or exploding gradient problems when handling long sequences, which limits their ability to model long-term dependencies. To address this issue, researchers have developed various improved architectures, among which Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are the most representative gated recurrent variants.

The gated recurrent neural network was proposed to more effectively capture dependencies between time steps that are far apart in time series. Its core idea is to introduce a gating mechanism that regulates the flow of information through learnable parameters. The GRU, as a widely used gated recurrent unit, modulates the hidden state update process via a reset gate and an update gate, thereby enhancing the model's ability to handle long-term contextual information.

Specifically, assuming the number of hidden units is h, given the input at time step t, $X_t \in \mathbb{R}^{n\times d}$ (where n is the number of samples and d is the input dimension), and the hidden state from the previous time step $H_{t-1} \in \mathbb{R}^{n\times h}$, the computations for the two gates in the GRU are as follows:

Reset Gate:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \tag{1}$$

This controls the extent to which the previous hidden state influences the current candidate hidden state.

Update Gate:

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \tag{2}$$

This balances the information ratio between the previous hidden state and the current candidate hidden state.

The candidate hidden state $\tilde{H}_t$ and the final hidden state H_t are computed as:

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h) \tag{3}$$

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t \tag{4}$$

Here, σ denotes the sigmoid activation function, $\odot$ represents element-wise multiplication, and W and b are learnable parameters.

The main advantages of GRU lie in its simple structure, fewer parameters, and high training efficiency, while achieving performance comparable or even superior to LSTM in many sequence modeling tasks. It is particularly practical in scenarios with limited computational resources. However, in some highly complex sequence modeling problems, due to its simplified gating mechanism, the expressive power of GRU may be slightly inferior to that of LSTM.

To more intuitively illustrate the structure and information flow of the GRU, Figure 1 provides a detailed computational graph of the GRU unit:

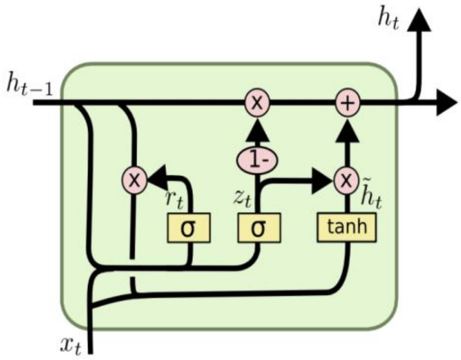


Figure 1 Schematic diagram of the GRU unit structure, showing the update gate, reset gate, and state update process.

Additionally, the following code snippet demonstrates an example of building a GRU layer using Python and TensorFlow/Keras, which can serve as a reference for high school readers to understand practical implementation:

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import GRU, Dense
model = Sequential()
model.add(GRU(units=50, return_sequences=True, input_shape=(time_steps, input_dim)))
model.add(GRU(units=50))
model.add(Dense(1, activation=sigmoid))
model.compile(optimizer=adam, loss=binary_crossentropy, metrics=[accuracy)
```

In conclusion, due to its effectiveness and efficiency, the GRU is a deep learning model widely used for time series prediction applications, such as predicting the BDI index. The relative simplicity and efficiency of the GRU model make it an accessible yet powerful tool for time-series forecasting tasks.

3. Experimental Design

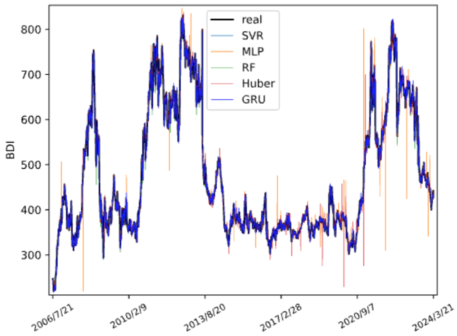
To rigorously test the efficacy of the proposed Gated Recurrent Unit (GRU) model, a comprehensive forecasting experiment was undertaken. The historical data set comprising 8,988 daily observations for the period October 19, 1988, to August 13, 2024, was divided into two equal data sets of 4,494 samples each, the first for training of the models and the second, 4,494 samples, for the out of sample test data for validation. Since the shipping market operates on a five-day trading week, a lag of five days was used so that each prediction for a given day was based upon that day’s data for the preceding five days.

To provide a comprehensive benchmark, the performance of the proposed GRU model was compared to that of five standard regression techniques; these were Support Vector Regression (SVR), a normal multilayer perceptron MLP, an ensemble Random Forest (RF) method, a robust Huber regression model and a standard ordinary least squares regression (OLS) method.

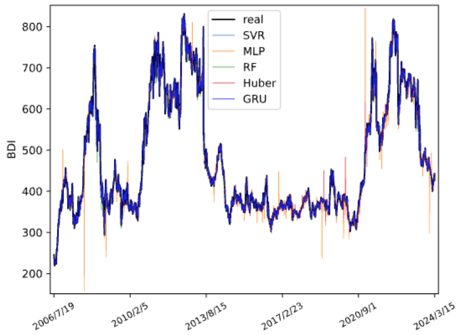
To quantify model performance, a set of four standard measures of performance was applied: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). These measures provide a good over-all view of prediction accuracy, since they give a reasonable balance between the effect of being sensitive to large errors and the need to know the size of those errors. All experiments were performed under Python 3.9 using scikit-learn and TensorFlow/Keras libraries.

4. Results

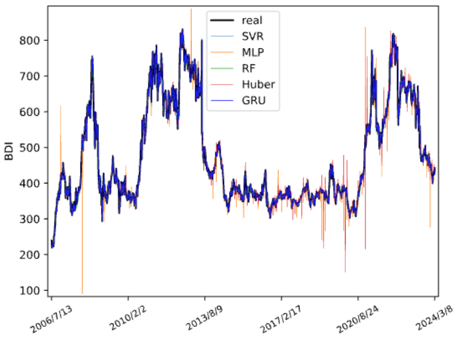
Based on the trained model, the prediction results of the deep learning model were compared with five other benchmark machine learning models on the test set. In order to more effectively demonstrate the effectiveness of deep learning models for BDI prediction and eliminate randomness, their performance was compared in short-term prediction, medium-term prediction, and long-term prediction scenarios, namely, with a horizon of 1 day (h=1), 1 week (h=5), 2 weeks (h=10), 3 weeks (h=15), and 1 month (h=20), as shown in Figure 2.



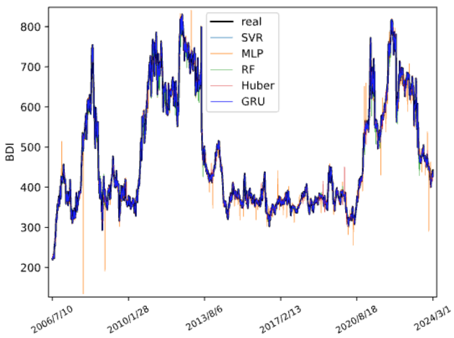
(a) $h = 1$



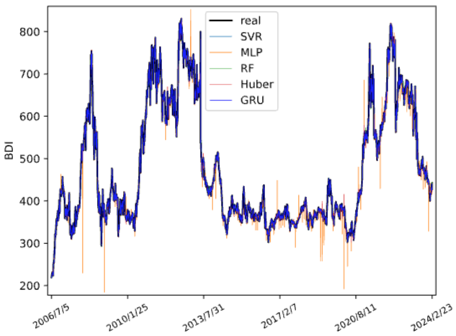
(b) $h = 5$



(c) $h = 10$



(d) $h = 15$



(e) $h = 20$

Figure 2 Comparison of Prediction Performance.

From Figure 2, it can be seen that both deep learning models and the compared SVR, MLP, RF, and Huber models can predict the trend of BDI well overall, indicating that: 1.The model was successfully trained on the training set; 2.The experimental results are relatively fair, that is, there is no deliberate use of models with particularly poor performance to highlight the performance of deep learning models. However, it is difficult to distinguish to what extent the deep learning model outperforms other benchmark models in the figure. Therefore, this experiment introduced four indicators, MSE, RMSE, MAE, and MAPE, to evaluate the predictive performance of different models. The evaluation results are shown in Table 2.

Table 2 Comparison of Predicted Performance under Various Evaluation Indicators.

Model	MSE	RMSE	MAE	MAPE (%)
Panel A: h=1				
SVR	113.44738	10.651168	6.45789702	1.32963776
MLP	235.669816	15.3515412	7.48900692	1.5501392
RF	119.37542	10.9259059	6.67636473	1.36624474
Huber	150.044829	12.2492787	6.94168615	1.43949039
OLS	2.29E+26	1.5145E+13	8.4908E+11	1.8569E+11
GRU	112.601504	10.6113856	6.39389521	1.31712046
Panel B: h=5				
SVR	113.631478	10.6598067	6.46559127	1.33211857
MLP	220.579597	14.8519223	7.41049825	1.54128778
RF	116.512776	10.7941084	6.51588306	1.33984726
Huber	129.517741	11.3805862	6.7708711	1.40247857
OLS	4.28E+27	6.541E+13	3.5237E+12	7.8954E+11
GRU	112.718892	10.6169154	6.40159035	1.31810414
Panel C: h=10				
SVR	113.396606	10.6487842	6.44469051	1.32726453
MLP	238.481518	15.4428468	7.47886863	1.55800554
RF	117.823339	10.8546459	6.58135983	1.35176242
Huber	193.458606	13.9089398	7.21343197	1.49787285
OLS	1.16E+29	3.3991E+14	1.4152E+13	3.0913E+12
GRU	112.708417	10.616422	6.4018546	1.31837445
Panel D: h=15				
SVR	113.059552	10.6329465	6.43944449	1.32448323
MLP	203.993219	14.2826195	7.56736515	1.56600894
RF	127.299105	11.2826905	6.74152115	1.37350585
Huber	119.326198	10.9236531	6.61890981	1.36736285
OLS	4.82E+27	6.9417E+13	3.0387E+12	6.6419E+11
GRU	112.781641	10.6198701	6.40223519	1.31720838
Panel E: h=20				
SVR	113.542767	10.6556448	6.45441803	1.32844083
MLP	201.812505	14.2060728	7.4716045	1.55707086
RF	113.215751	10.6402891	6.44958435	1.32463536
Huber	116.009316	10.7707621	6.52930476	1.34772446
OLS	2.41E+28	1.5537E+14	6.7742E+12	1.5316E+12
GRU	112.801233	10.6207925	6.40145446	1.31710003

From the analysis of Table 2, the following results can be derived: (1) The ordinary least squares (OLS) model performed the worst across all evaluation metrics and forecasting horizons (1 day, 1 week, 2 weeks, 3 weeks, and 1 month). This discovery indicates that linear models are unsuited for the task of characterising and learning the non-linear, non-periodic and non-stationary habits of the BDI. (2) The deep learning models are the best of the various models proposed according to the various measures of prediction capability used. In short-term forecasting, medium-term forecasting, and long-term forecasting in a forecasting lead time of one day, one week, two weeks, three weeks, and one month, respectively, deep learning models consistently delivered superior performance. This discovery indicates that deep learning models are more fitted for the task of forecasting the

BDI time series data by non-linear, non-periodic and non-stationary habits.

5. Discussion

A major conclusion from these results is that the fact that the results for the GRU model is the best for these results shows that it is because this model is able to incorporate the temporal relationships and dependencies built into it from the processed data. The BDI is not a series of serially independent random variables but is in fact the end product of the memory of the last state of the economy with regard to the inclusion of the past conditions present in the economy with regard to the past supply chains of goods being moved and subject to the various economic cycles. The GRU because of its deep learning nature and not a nose continuing the memory of the past as in the benchmark models that are not as good at extracting the serial dependent variable time structures embedded in it, the temporal correlations it is able to cache the information of the long run dependent variables. Therefore, this constructive structure of the model is able to retrieve the raw information that requires this hierarchical approach. This indicates that the model is able to uncover factors that are hidden and are not observed which influence the market dynamics which extend beyond simple asset valuation. These indications of speculative motives are ones that are hidden and subtle from the past, and these will help as dynamic and vitality in forecasting the BDI.

As well as giving the practitioner the correct model to work with, this paper also indicates the strategic value of full model feature engineering. The results show improved explanations of the findings in relation to the fluctuations of the BDI when the point of view is expressed, as it should be, from the combined supply side, demand side and macroeconomic and causal variables. The high accuracy level in this study indicates the necessity for full feature engineering. When going from a full feature model to a full feature model, the model is found to work better. This shows that this increase in dimensionality should resulted in greater accuracy for the model, suggesting that with further research on the data that it will be found that the accuracy can still be improved.

However the conclusions have to be borne in mind and seen in reference to the limits of this study, which gives some goals for future research. This is a large feature set, but there are others. It may also worth while improving further by research on the predictive models as they are run, by building in smaller feature variables such as such geo-political risk factors, or satellite information regarding the velocity of goods and relevant commodities that are traded. Another line of fruitful research will be to build further an ensemble model of this model, involving by using deep learning to combine it with others, thus improving predictive accuracy and stability. Finally another line which no doubt leads fruitful research would show a more systematic control of hyperparameters involved in this model, perhaps by using as is suggested in literature today the AutoML, that should lead to improvement in the control of accuracy of feature engineering of this forecasting framework rather than being heuristically as it is at the moment, based only on the forecaster's intuitive response to determining the setting of this crucial aspect of the model. This hyperparameter tuning, therefore needs in the future to take place in a systematic study by which this suggested paper or framework can help with the prediction of the future using some of the models used in the past.

6. Conclusion

This study successfully addressed the classic problem inherent in forecasting the highly volatile and notoriously unpredictable BDI. To this end, a GRU deep learning model was employed fed by a broad, multi-variate data set. The results show a strong alternative to conventional forecasting methods. The empirical evidence yields a distinct and consistent performance advantage for the proposed method over various time horizons. Ultimately, this paper shows that the growing application of deep learning in this area is an effective means of improving economic analysis and, in particular, provides an enhanced and more robust method of examining and forecasting the behavioural patterns of the intensely complex global shipping market.

References

[1] Bakshi, G., Panayotov, G. and Skoulakis, G. (2011) The Baltic Dry Index as a Predictor of Global Stock Returns, Commodity Returns, and Global Economic Activity. Social Science Electronic Publishing.

[2] Zhang X , Xia Z , He F ,et al.Forecasting crude oil prices with alternative data and a deep learning approach[J].Annals of Operations Research, 2025, 345(2).DOI:10.1007/s10479-024-06056-8.

[3] Angelopoulos, J. (2017) Time–Frequency Analysis of the Baltic Dry Index. Maritime Economics & Logs, 19, 1-23.

[4] Si, S. (2016) Forecasting Approaches of the Baltic Dry Index: A Literature Review. Social Science Research Network.

[5] Papailias, F. and Thomakos, D.D. (2013) The Baltic Dry Index: Cyclical Analysis and Forecasting. SSRN Electronic Journal.

[6] Z., H. and L.I., X. (2009) Volatility of Baltic Dry Index Using GARCH Type Models with Different Distributions. Journal of Shanghai Maritime University.

[7] Katris, C. and Kavussanos, M.G. (2021) Time Series Forecasting Methods for the Baltic Dry Index (BDI). Journal of Forecasting.

[8] Ren, J. (2020) Analysis of Correlation and Influence Degree between BDI Index and Fluctuation of Transport Capacity Index. Francis Academic Press.

[9] Geman, H., Smith, et al. (2012) Shipping Markets and Freight Rates: An Analysis of the Baltic Dry Index. Journal of Alternative Investments.

[10] Lin, F. and Sim, N.C.S. (2013) Trade, Income and the Baltic Dry Index. European Economic Review, 59, 1-18.

[11] Kuo, P.L., Chiu, C.L., Chen, C.S., et al. (2020) The Dynamic Relationships between the Baltic Dry Index and the BRICS Stock Markets: A Wavelet Analysis. Asian Economic and Financial Review, 10, 340-351.