Received November 3, 2018, accepted November 27, 2018, date of publication December 10, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2884877

Comparison of Econometric Models and Artificial Neural Networks Algorithms for the Prediction of Baltic Dry Index

XIN ZHANG, TIANYUAN XUE, AND H. EUGENE STANLEY
[1]College of Communication and Transport, Shanghai Maritime University, Shanghai 201306, China
[2]Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215, USA

Corresponding author: Xin Zhang (zhangxin@shmtu.edu.cn)

This work was supported in part by the National Science Foundation of China under Project 71601112, in part by the Shanghai Pujiang Program under Grant 15PJC061, in part by NSF at the Boston University Center for Polymer Studies under Grant PHY-1505000, Grant CMMI-1125290, and Grant CHE-1213217, and in part by DTRA Grant HDTRA1-14-1-0017.

ABSTRACT The shipping market, a major component of the global economy, is characterized by high risk and volatility. The Baltic dry index is an influential indicator in the world shipping market and international trade. Several studies have used a variety of techniques to generate Baltic dry index predictions. The most prominent techniques utilize either econometric or artificial intelligence computing. We compare the forecasting accuracy of two typical univariant econometric models and three artificial neural networks (ANNs)-based algorithms. We find that when using daily data, econometric forecasting models produce better one-step-ahead predictions than ANN-based algorithms. When forecasting weekly and monthly data, ANN-based algorithms produce fewer errors and a higher direction matching rate than econometric models. We also compare the predictive power of a number of different models when applied to the 2008 financial crisis and find that the generalized autoregressive conditional heteroskedasticity model and the back propagation neural network algorithm produce the best one-step-ahead and seven-steps ahead predictions, respectively.

INDEX TERMS Baltic dry index prediction, ARIMA,GARCH, artificial neural networks(ANN), BP neural network, RBFNN, ELM.

I. INTRODUCTION

More than 90% of current international trade utilizes shipping, an industry that is extremely dynamic and volatile [1]. Volatility measurements generated by the Baltic Dry Bulk Index (BDI) indicate that it is significantly higher in shipping markets (>79%) than in commodity (>50%) and equity markets (e.g.,the Standard & Poor's 500 Index (S&P500) > 20%) [2]–[4].

The Baltic dry index is regarded as a ‘‘barometer’’ to evaluate the shipping industry, international trade, and the global economy [5]–[7]. Investors, speculators, and researchers have long found these indices to be useful, theoretically challenging, and relevant when projecting future profits. Because many managerial decisions are based on future prospects, forecasting accuracy is essential in large- scale organizations and companies. Recent advances in both analytical and computational methods have produced a number of new ways of mining freight index time-series data.

The analytic methodology of dry bulk freight index forecasting falls into two categories. The first includes univariant and multivariant econometric methods, such as the auto-regressive integrated moving average (ARIMA), vector auto-regression (VAR), generalized autoregressive conditional heteroskedasticity (GARCH), and the vector error correction (VEC) models.

Cullinane *et al.* [8] was the pioneer in developing a BDI index analyzing method using the ARIMA model. Kavussanos and Alizadeh [9] then created a seasonal ARIMA model of a single variable and a VAR model to study the seasonal characteristics of the dry bulk shipping market. Batchelor *et al.* [10] compared the ARIMA, VAR, and VECM models in predicting spot and forward freight rates. Chen *et al.* [11] used the ARIMA and VAR models to predict the freight rates of several dry bulk routes, and they found that VAR performs better in out-of-sample forecasts than ARIMA. Tsioumas *et al.* [12] developed a multivariate vector autoregressive model with exogenous variables (VARX) to improve the forecasting accuracy of the BDI, and they found that the VARX model outperforms the ARIMA. Adland *et al.* [13] presented cointegrated time series models in continuous and discrete time to analyze the dynamics of regional ocean freight rates.

Stopford [14] indicates that maritime forecasting has a poor reputation because it is difficult for traditional econometric and statistical methods to capture the nonlinear characteristics hidden in dry bulk freight indices [15]. Thus a second category of analytic methodologies now utilizes a number of non-linear and artificial intelligence (AI) methods, such as ANN, support vector machines (SVM), and non-linear regression.

Leonov and Nikolov [16] use a model based on wavelets and neural networks to predict dry bulk freight rates. Bulut *et al.* [17] apply a fuzzy vector autoregressive integrated logical model to forecast time charter rates. Duru *et al.* [18] propose a fuzzy-DELPHI adjustment method of increasing accuracy when statistically forecasting dry bulk shipping indices. Han *et al.* [19] use wavelet transform to denoise the BDI data series and combine wavelet transform and a support vector machine to forecast BDI. Zeng *et al.* [20] use empirical mode decomposition (EMD) and artificial neural networks (ANN) to forecast the BDI. Chou and Lin [21] propose an integrated fuzzy neural network combined with technical financial market indicators to predict BDI and find that forecasts using an integrated fuzzy neural network are more accurate that those using the traditional approach. Şahin *et al.* [22] compare the BDI forecasting accuracy of three ANN models and find that their performances are similar, but that the most consistent is an ANN using BDI input data from the two most-recent weekly observations.

The ANN-based prediction model produces good nonlinear approximations, but its structure is difficult to determine. Neural networks are susceptible to either insufficient or excessive training, and this can induce shortages—such as trapping in a local minimum—caused by its sensitivity to initial values.

Although much research on BDI prediction has been conducted using various techniques, we still do not understand the applicability, superiority, and deficiencies of different forecasting techniques. Approximations of econometric models when applied to complex nonlinear problems are inadequate, and using ANN-based algrithoms to model linear problems also produces misleading results. For example, Markham and Rakes [23] used simulated data and found that the performance of ANNs is strongly affected by sample size and noise level. Because it is difficult to know all the characteristics of real-world datasets, the blind application of ANNs is unwise.

There are few studies that compare the performance of these two techniques for different time-scale datasets and for single and multi-period ahead forecasting. We here compare the sensitivity and predictive accuracy of the typical univariant econometric models with commonly-used ANN-based algrithoms when applied to the BDI at different time-scales and in various prediction senarios. The goal of this comparative study is to determine the suitable model for forecasting future shipping market trends, the results of which would be useful in academic research and industrial practice.

Section 2 of this paper describes the principles and forecasting procedures of econometric models and ANN-based algrithoms. Section 3 reviews BDI shipping freight market data, detects stationarity and votility, and proposes several metrics for quantifying predictive performance. Section 4 compares the performance of several empirical results. Section 5 presents conclusions and recommendations for future studies.

II. METHODOLOGY

A. ECONOMETRIC FORECASTING MODEL

The most common methods of forecasting trends and seasonal components are time series analysis. Here we focus on the (i) ARIMA and (ii) ARIMA-GARCH models.

1) ARIMA

ARIMA models have dominated time series forecasting for over 50 years. Auto-regressive (AR) models were first introduced by Yule and later generalized by Walker, and moving average (MA) models were first introduced by Chatfield [24]. Box and Jenkins combined the two to form the auto-regressive moving average (ARMA) method. In an ARMA (p, q) model, the future value of a variable is assumed to be a linear function of several past observations and random errors, i.e.,

$$\begin{aligned} y_t &= c + \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} \ldots - \theta_q \varepsilon_{t-q}, \varepsilon_t \sim N(0, 1), \end{aligned} \tag{1}$$

where y_t and ε_t are the actual value and random error at time period t, respectively, φ_i $(i = 1, 2, \ldots, p)$ and θ_j $(j = 0, 1, 2, \ldots, q)$ are model parameters, and p and q are integers often referred to as orders of the model. Random errors ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

When a variable is not stationary, a common solution is to use different variable values and the integrated part converts an ARMA(p, q) model into an ARIMA (q, d, q)model. Here d is the number of differences needed for stationarity. Hence an ARIMA model can be expressed

$$\begin{aligned} \Delta y_t &= c + \varphi_1 \Delta y_{t-1} + \ldots + \varphi_p \Delta y_{t-p} \\ &\quad + \varepsilon_t - \theta_1 \varepsilon_{t-1} \ldots - \theta_q \varepsilon_{t-q}, \varepsilon_t \sim N(0, 1), \end{aligned} \tag{2}$$

where Δy_t is the value of y_t after d number of different values. Here when $d = 1$ then $\Delta y_t = y_t - y_{t-1}$. Equation (2) entails several special cases of the ARIMA family of models. When $q = 0$, Eq. (2) becomes an AR model of order p. When $p = 0$, the model reduces to an MA model of order q. A central task of ARIMA model building is to determine the appropriate model parameters (p, d, q).

To select d, we first check the autocorrelation by plotting the autocorrelation function (ACF) and partial autocorrelation function (PACF). Taking into consideration the lag of ACF and PACF, we apply first order or higher order differences to the original BDI to achieve stationarity after d order differencings, which is confirmed using the augmented Dickey-Fuller Test.

To select the best fitting ARIMA model, we investigate several ARIMA models with p ranges from 1 to 10 and q ranges from 1 to 10. This produces 100 different ARIMA models with different p, q, and the best fitting model is the one that exhibits the lowest Akaike information criteria (AIC).

2) ARIMA-GARCH

Over the three decades since its introduction, GARCH model (Bollerslev, 1986; Engle, 1982) and numerous variants have improved volatility estimation and time series prediction [25], [26]. Some studies [27]–[30] find that GARCH models outperform such traditional techniques as moving average, exponential smoothing, and linear regression when modeling and forecasting conditional time series.

Because a GARCH model is able to deterministically describe volatility at time t using the information available at time $t-1$, it can model and forecast its value when the time series of each component q is heteroskedastic. Pham and Yang [31] estimate the ARMA(p, q)–GARCH (r, s) model for a univariate process using

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \varepsilon_t \sim N(0, \sigma_t^2), \quad (3)$$

where y_t and ε_t are the actual value and random error at time period t, respectively, and φ_i ($i = 1, 2, \ldots, p$) and θ_j ($j = 0, 1, 2, \ldots, q$) are model parameters. This differs from the ARIMA model in that random errors ε_t are not assumed to be independently and identically distributed with a constant variance of σ^2. Otherwise, $\varepsilon_t = \sigma_t Z_t$, where $Z_t \sim iidN(0, 1)$. Then ε_t follows a GARCH (r, s) model,

$$\sigma_t^2 = a + \sum_{g=1}^{r} \beta_g \sigma_{t-g}^2 + \sum_{k=1}^{s} \omega_k \varepsilon_{t-k}^2, \quad (4)$$

where σ_t^2 is the conditional variance of y_t, and a, ω_k, and β_g are the constant and estimated parameters of ARCH and GARCH, respectively. After selecting the appropriate ARIMA, we use the Lagrange-multiplier (L-M) test to detect ARCH effects. If we find GARCH effects, we employ ARIMA-GARCH models with r ranges from 0 to 4 and s ranges from 0 to 4. As in the ARIMA model, the model with the lowest AIC value is the best fitting model.

B. ARTIFICIAL NEURAL NETWORKS ALGORITHMS

A series of ANN algorithms for univariable time series forecasting were recently developed, and they have proven to be superior to traditional forecasting models [32], [33].

There are three steps in the time series forecasting method that uses ANN:

(i) Prepare the training and test datasets.
Assuming that the original sample set has N observation values, we use the sliding window method with window length L and sliding step $l = 1$ to divide the original BDI dataset into $M = [N - L + l]$ data samples. Each sample contains L observation values of the BDI $\{X_t, X_{t-1}, \cdots, X_{t-L}\}$. For each sample, we select the last s observation values as the output set and the previous $L - s$ observation values as the input ANN set. Here s is the number of steps in the multi-periods-ahead prediction. In one-step-ahead predicting $s = 1$ and $L = 5$. In seven-steps-ahead predicting $s = 7$ and $L = 35$. We then divide the sample dataset into two parts, a training set and test set. The training set here is the first $M - 1$ samples in the sample dataset and the test set is the last or sample M. We also divide the training dataset into a sub-training set and a sub-validation set to generate the trained neural network. Using a previous study [34], we set the ratio of the sub-training set and the sub-validation dataset at 8 : 2.

(ii) Data normalization and denormalization.
To use the ANNs algorithm we first minimize the difference between the threshold and actual data. The training data are usually normalized before being input,

$$X_t' = \frac{X_t - X_{min}}{X_{max} - X_{min}}, \quad (5)$$

where X_t' is the data after normalization at time t, and X_{min} and X_{max} are the minimum and maximum of original data X_t in our study, which are daily, weekly, or monthly BDIs. After processing we antinormalize the output,

$$\widehat{Y}_t = Y_t'(Y_{max} + Y_{min}) + Y_{min} \quad (6)$$

where Y_t' is the predicted data after anti-normalization at time t, and Y_{min} and Y_{max} are the minimum and maximum of output Y_t'.

(iii) Train the neural network and simulate the prediction results.
Finally we use the normalized sub-training dataset and sub-validation dataset to train the neural network, which we regard an in-sample prediction similar to that using an econometric forecasting model. We set all ANN-based algorithms to use a two-layer artificial neural network with one hidden layer and one output layer. Here the number of units in the hidden layer can vary from model to model. We then use k-fold cross validation to select the ANN architecture with the fewest errors. We randomly partition the original training dataset into k equal sized sub-samples, here $k = 5$. Of the five sub-samples, a single sub-sample is retained as sub-validation data to test the model. The remaining four sub-samples are used as sub-training data. We then repeat the cross-validation process five

times, using each of the five sub-samples once as validation data [35]. After determining the best ANN, we input the $L - s$ values of the test set into the neural network and output the s values as the forecasting result, which is the out-sample prediction. For each prediction, we train the neural network and simulate the predicted value ten times and use the average of all predictions as the final forecasting result, which we then compare with the actual value of the testing dataset.

We here focus on three widely-used models, (i) the back propagation neural network (BPNN), (ii) the radial basis function neural network (RBFNN), (iii) and the extreme learning machine (ELM).

Although BPNN is a widely-used neural network models, it requires a long training period. The RBFNN uses local transformations that improve BPNN by enabling much more rapid training. Because RBFNN with one hidden layer can also approximate any function, we select it to be the second ANN-based model in our comparison. ELM is an emerging learning algorithm for forward-feed neural network training. Unlike BPNN and RBFNN, the internal parameters of ELM are not iteratively adjusted, but are set to random values, usually uniformly distributed, which simplifies the requirements for determining the output hidden unit weights. Thus the computational burden of ELM is significantly less, but its performance is comparable to that of BPNN and RBFNN in different applications.

1) BACK PROPAGATION NEURAL NETWORK

The BPNN model is one of the most widely used ANN-based algorithms for classification and prediction [36]–[38]. This technique is an advanced multiple regression analysis that deals with responses that are more complex and non-linear than those of standard regression analysis. The basic formula of the BP algorithm is

$$W(n) = W(n-1) - \Delta W(n), \tag{7}$$

where

$$\Delta W(n) = \eta \frac{\partial E}{\partial W}(n-1) + \gamma \Delta W(n-1), \tag{8}$$

where W is the weight, η the learning rate, E the gradient of error function, and $\gamma \Delta W(n-1)$ the incremental weight.

Using BPNN we select the hyperbolic tangent sigmoid transfer function to be the activation function, set the maximum number of training epochs to 1000, and set the learning rate to 0.01. The number of nodes in the hidden layer ranges from 20 to 100 with an interval of 10, and we use k-fold cross-validation to determine the number of nodes.

2) RADIAL BASIS FUNCTION NEURAL NETWORK

RBFNN is an artificial neural network that uses radial basis functions as activation functions. Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function, and a linear output layer [39]–[41]. The input can be modeled as a vector of real numbers $x \in \mathbb{R}^r$, and the prototype of the input vectors $B_i \in \mathbb{R}^r$. The output of each RBF unit is

$$R_i(X) = R_i(\| X - B_i \|), i = 1, 2, \ldots, u, \tag{9}$$

where $\| \quad \|$ is the Euclidean norm on the input space. Because it can be factored, the Gaussian function is the preferred radial basis function. Thus

$$R_i(X) = exp[-\frac{\| X - B_i \|^2}{\sigma_i^2}] \tag{10}$$

where σ_i is the width of RBF unit i. The output $Y_j(X)$ of unit j of an RBFNN is

$$Y_j(X) = \sum_{i=1}^{u} R_i(X) * W(j, i), \tag{11}$$

where $R_0 = 1$, $W(j, i)$ is the weight or strength of receptive field i to the output j, and $W(j, 0)$ is the bias of output j. Geometrically, an RBFNN partitions the input space into several hypersphere subspaces. The parameters of the RBF networks are the center, the influence field of the radial function, and the output weight (between the intermediate layer neurons and those of the output layer). The training process can obtain these parameters.

For RBF we set the radial basis function to the Gaussian function. The number of nodes in the hidden layer varies from 90 to 200, increasing each step by ten in its architectural optimization.

3) THE EXTREME LEARNING MACHINE

ELM was originally applied to single hidden-layer feed-forward neural networks and then extended to generalized feed-forward networks [42], [43]. For a set of training samples $(X_j, C_j)_{j=1}^N$ with N samples and C classes, the single hidden layer feed-forward neural network with h hidden nodes and activation function $f(x)$ is

$$\sum_{i=1}^{h} \beta_i f_i(X_j) = \sum_{i=1}^{h} \beta_i f(W_i * X_j + b_i) = Y_j, j = 1, 2, \ldots N, \tag{12}$$

where $X_j = [x_{j1}, x_{j2}, \cdots, x_{jm}]^T$, $C_j = [c_{j1}, c_{j2}, \cdots, c_{jm}]^T$, $W_j = [w_{j1}, w_{j2}, \cdots, w_{jm}]^T$, and b_i are the input, its corresponding output, the connecting weights of hidden neuron i to input neurons, and the bias of hidden node i, respectively. Here $\beta_j = [\beta_{j1}, \beta_{j2}, \cdots, \beta_{jm}]^T$ are the connecting weights of hidden neuron i to output neurons, and Y_j the actual network output with respect to input X_j. Because the hidden parameters W_i, b_i can be randomly generated during the training period without tuning, ELM solves a compact model that minimizes the error between C_j and Y_j, i.e., $\min\| H\beta - C \|_F$

Here H is the hidden layer output matrix and β the output weight matrix. The merit of ELM is that only the output weights are needed when randomly selecting the hidden node parameters (the input weights and bias). We set the number of ELM nodes in the hidden layer in a range between 10 to 50, increasing each interval by 10.

III. DATA DESRCIPTION AND PREDICTION PERFORMANCE METRICS

A. DATA

To service the participants in the dry bulk shipping market, the London Baltic Exchange began publishing the Baltic Freight Index (BFI) in 1985. On 1 November 1999 it was replaced by the BDI, which is now widely used by industry practitioners and considered the ‘‘barometer’’ of the dry bulk shipping market. The BDI is an index that measures 26 shipping routes in terms of time-charter and voyage. We here compare the forecasting of the daily, weekly, and monthly BDI data from 1 November 1999 to 30 February 2018. All of the data is from the world's leading shipping database, Clarkson Sin (https://sin.clarksons.net/)

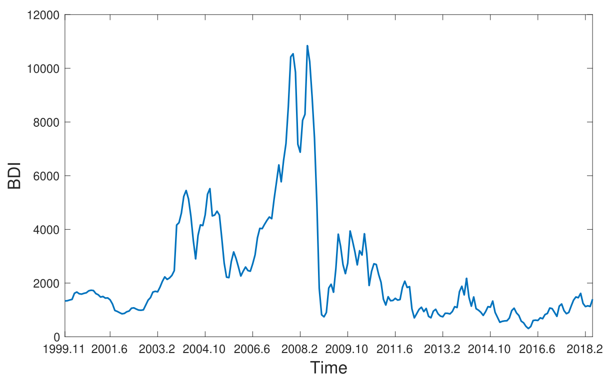

FIGURE 1. Index values of BDI from November 1999 to February 2018.

Figure 1 shows that the fluctuations of the BDI have a large amplitude, a high frequency, and are irregular. From November 1999 to early 2003, the BDI value gently fluctuates, but beginning in 2003 market volatility increases, the BDI reaches 11793 on 20 May 2008, and then drops sharply to approximately 700 points in the following six months.

TABLE 1. Descriptive statistics of daily value and rate of change of BDI.

Statistic	Daily BDI Value	Rate of change for BDI
Mean	2402.9	0.0002
Min	290	-0.1137
Median	1576.5	0.0000
Max	11793	0.1463
Skewness	1.9801	0.2894
Kurtosis	7.0340	7.9160
ADF	-0.6139	-24.2159
[p-value]	[0.4271]	[0.001]
J-B	6148.8	4713.7
[p-value]	[0.001]	[0.001]
Observations	4618	4617

Table 1 provides descriptive statistics of the daily BDI and its rate of change. The rate of change of the daily BDI value is $R_{y_t} = \frac{y_t - y_{t-1}}{y_{t-1}}$.

Note that the volatile daily movements of the BDI value range from -11.37% to 14.63%. The Jarque and Bera (1980) test rejects the null hypothesis of normality at the 5% significance level for both BDI daily values and returns. Finally, the values of the augmented Dickey and Fuller (ADF) unit root test suggest that the BDI value is unstationary, and that a first order difference is needed if predictions are to be based on ARIMA and GARCH models.

B. ACCURACY METRICS

To measure the forecasting accuracy of these proposed methods we use the mean absolute percentage error (MAPE) and the root mean square error (RMSE) [44], [45], which are defined

$$MAPE = \frac{1}{N}\sum_{t=1}^{N} \left| \frac{X(t) - \hat{X}(t)}{X(t)} \right|, \tag{13}$$

and

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\hat{X}(t) - X(t))^2}{N}}, \tag{14}$$

where $\hat{X}(t)$ and $X(t)$ are the predicted and real values at time t, respectively, and N is the size of the dataset being tested. The MAPE technique measures the mean absolute relative error of the prediction models, and the RMSE technique measures their standard deviation.

In using these error criteria we find that the smaller the MAPE and RMSE values the greater the accuracy of the model.

For any given prediction, the actual outcomes above and below the predicted outcome are treated asymmetrically when using the MAPE and RMSE [46] techniques. Thus the directional tendencies of the data fluctuations—whether upward, stable or downward—are important. We measure them using the direction matching rate Dsta, which is defined

$$\text{Dsta} = \frac{1}{N}\sum_{t=1}^{N} a(t), \tag{15}$$

$$a(t) = \begin{cases} 1, & (X(t+1) - X(t)) * (\hat{X}(t+1) - X(t)) \geq 0 \\ 0, & otherwise. \end{cases} \tag{16}$$

The closer the Dsta value is to 1, the higher the accuracy of the directional prediction of the models, and the closer the Dsta value is to 0, the lower the accuracy of their directional predictions.

In addition to errors and directional matching rates, we must also evaluate whether the proposed forecasting models outperform simple banchmark forecasting methods. We use two simple banchmark methods, (i) the naïve method based on the most recent observation, and (ii) the historical mean method using data up to the most recent observation.

To compare the predictive accuracy of our proposed models with that of the naïve method, we use the mean absolute scaled error (MASE) proposed in 2005 by Hyndman and Koehler [47] and defined

$$MASE = \frac{MAE}{MAE_{sn}}, \tag{17}$$

where MAE is the out-sample mean absolute error,

$$MAE = \frac{\sum_{t=M+1}^{M+N} |\hat{X}(t) - X(t)|}{N}. \quad (18)$$

Here MAE_{sn} is the MAE of the one-step or multi-step ahead naïve method for in-sample prediction, which is defined

$$MAE_{sn} = \frac{\sum_{t=1+n}^{M} |X(t) - X(t-n)|}{M-n}, \quad (19)$$

where $\hat{X}(t)$ and $X(t)$ are the predicted and real values at time t, respectively, N is the size of the dataset being tested, M is the size of the training dataset, and n is the length of the period-ahead prediction.

When MASE is less than one it produces a better forecast than the average one-step naïve forecast computed in-sample. Conversely, when it is greater than one the forecast is worse than the average one-step naïve forecast computed in-sample. When computing multi-step forecasts, we can scale using the in-sample MAE derived from multi-step naïve forecasts.

To compare the predictive accuracy of our proposed models with the mean of the training dataset, we apply the Nash–Sutcliffe model efficiency coefficient (NSE) [48], which is defined

$$NSE = 1 - \frac{\sum_{t=1}^{N}(\hat{X}(t) - X(t))^2}{\sum_{t=1}^{N}(X(t) - \bar{X})^2}, \quad (20)$$

where $\bar{X}$ is the mean of the training dataset. NSE can range from ∞ to 1. When NSE $= 0$, the model predictions are as accurate as the mean of the observed data, but when $NSE < 0$ the observed mean is a better predictor than the forecasting model. The closer the NSE model is to 1, the higher its accuracy. There are threshold values producing sufficient level of accuracy between $NSE > 0.5$.

IV. EMPRICAL RESULTS

Because previous studies have found that predictive models are sensitive to the time scale of the data, we compare the accuracy of the econometric models with ANN-based algorithms using daily, weekly, and monthly BDI.

Because the shipping market is cyclical, we divide our dataset into three long-term cycles. Each cycle contains at least one BDI peak and contraction. In our BDI dataset the three cycle periods are (1999-2006), (2007-2010), and (2011-2018).

To compare how econometric and ANN-based models behave in different market cycles, we set the length of each cycle period to be the same as that of the sample data. Thus for daily, weekly, and monthly datasets, we select three stages when we compare forecasts.

A. DAILY BDI PREDICTIONS

When using econometric predictive models, for each stage we randomly select the beginning time point, make 200 consecutive observations to obtain the best fitting econometric model, and use the best fitting model to conduct out-sample one-step-ahead and seven-steps-ahead forecasts. As in ANN-based forecasting, we use the same observations as the training data and also predict the one-day-ahead and seven-days-ahead values. We compare the out-sample predictive accuracy of the econometric models with the ANN-based algorithms.

TABLE 2. Predictive performance of one-step head forecasting (daily data).

Model	MAPE	RMSE	MASE	Dsta	NSE
Stage I: 1999.11.10-2000.8.28					
ARIMA(6,1,2)	0.00173	2.82	1.89	0	0.9994
GARCH(1,1)*	0.00003	0.05	0.05	1	1.0000
BPNN(5,30,1)	0.00050	0.81	0.81	1	0.9461
RBFNN(5,90,1)	0.00052	0.84	0.84	1	0.9423
ELM(5,20,1)	0.00036	0.59	0.59	1	0.9714
Stage II: 2004.8.17-2005.6.7					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(2,1,2)	0.00987	28.94	0.45	1	0.9997
GARCH(2,1)*	0.00882	25.87	0.40	1	0.9997
BPNN(5,20,1)	0.01038	30.42	0.48	1	0.9996
RBFNN(5,120,1)	0.04292	32.58	0.60	1	0.9999
ELM(5,10,1)	0.05706	36.72	0.76	1	0.9999
Stage III: 2012.8.21-2013.6.11					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(3,2,5)	0.01490	12.62	0.57	1	0.5199
GARCH(1,1)*	0.01345	11.39	0.52	1	0.6876
BPNN(5,40,1)	0.01750	14.82	0.67	1	0.2458
RBFNN(5,90,1)	0.01448	12.26	0.56	1	0.8063
ELM(5,10,1)	0.01650	13.97	0.64	1	0.6500

*: The best predictive model in the testing stage (similarly hereinafter)

Table 2 compares the one-step-ahead performance of econometric models with that of ANN-based algorithms. Note that the values of MASE and NSE indicate that all models perform better than the naïve method and the mean value of historical data. The GARCH model outperforms the ARIMA model in all three forecasting stages with lower MAPE and RMSE. The GARCH model also exhibits lower MAPE and RMSE outcomes than the ANN-based algorithms in all stages.

Table 3 shows the seven-steps-ahead predictive performances of econometric models and ANN-based algorithms. We find that GARCH outperforms ARIMA among the econometric forecasting models. Among the ANN-based models, no single method performs the best in all stages. For example, in the second stage ELM performs better than the other ANN-based models, but the RBFNN is the most accurate of the ANN-based models in the first and third stages. This indicates that the predictive power of ANN-based models is sensitive to the characteristics of the training data. The difference in errors between the GARCH model and the ANN-based algorithms is slight but if we compare them with Dsta the predictions of ANN-based models produce a better directional matching than the econometric models. Note that in some stages, simple baseline forecasting methods produce more accurate forecasts than the proposed models. For example, in the third stage the simple naïve method and mean value of historical data both outperform BPNN.

B. WEEKLY BDI PREDICTIONS

To make weekly data predictions, we apply the same procedure to obtain a sample dataset and make out-sample

TABLE 3. Predictive performance of seven-step head forecasting (daily data).

Model	MAPE	RMSE	MASE	Dsta	NSE
Stage I: 1999.11.10-2000.8.28					
ARIMA(6,1,2)	0.00823	17.19	1.13	0.28	0.9793
GARCH(1,1)	0.00215	4.19	0.03	0.43	0.9988
BPNN(35,60,7)	0.00443	10.19	0.65	0.29	0.9900
RBFNN(35,150,7)*	0.00212	4.30	0.29	0.57	0.9900
ELM(35,50,7)	0.00591	12.75	1.01	0.29	0.9800
Stage II: 2004.8.17-2005.6.7					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(2,1,2)	0.04220	132.13	0.43	0.29	0.9945
GARCH(2,1)	0.03809	119.74	0.54	0.43	0.9955
BPNN(35,70,7)	0.10583	335.91	1.05	0.14	0.9644
RBFNN(35,120,7)	0.04979	235.66	0.47	1.00	0.9825
ELM(35,40,7)*	0.03174	98.79	0.40	1.00	0.9969
Stage III: 2012.8.21-2013.6.11					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(3,2,5)	0.06429	70.63	1.01	0.14	0.5199
GARCH(1,1)	0.05990	65.46	0.94	0.14	0.5876
BPNN(35,80,7)	0.07981	88.52	1.29	0.34	0.2458
RBFNN(35,150,7)*	0.04540	44.86	0.70	0.54	0.8063
ELM(35,40,7)	0.09040	100.59	1.48	0.44	0.0262

TABLE 4. Predictive performance of one step head forecasting (weekly data).

Model	MAPE	RMSE	MASE	Dsta	NSE
Stage I: 1999.11.5-2003.9.12					
ARIMA(2,1,4)	0.02861	69.12	0.92	1	0.9949
GARCH(2,1)	0.02943	71.10	0.95	1	0.9946
BPNN(5,30,1)*	0.01486	35.91	0.48	1	0.9986
RBFNN(5,100,1)	0.01771	42.79	0.57	1	0.9980
ELM(5,20,1)	0.01817	43.90	0.59	1	0.9979
Stage II: 2008.7.25-2012.6.8					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(4,1,4)	0.02703	24.41	0.89	1	0.9997
GARCH(1,1)	0.03679	33.22	1.22	0	0.9994
BPNN(5,60,1)*	0.00891	8.04	0.29	1	1.0000
RBFNN(5,100,1)	0.07697	69.51	2.54	0	0.9975
ELM(5,30,1)	0.06977	63.00	2.31	0	0.9980
Stage III: 2011.12.16-2015.10.16					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(8,1,1)	0.00310	2.38	0.17	1	0.9999
GARCH(0,1)	0.00582	4.48	0.32	1	0.9997
BPNN(5,20,1)*	0.00083	0.64	0.04	1	1.0000
RBFNN(5,90,1)	0.03931	30.26	1.13	1	0.9851
ELM(5,20,1)	0.00397	3.06	0.22	1	0.9998

predictions for one-step-ahead and seven-steps ahead-values using econometric models and ANN-based algorithms. Table 4 compares the one-step-ahead performance of the econometric models with the ANN-based algorithms for the weekly BDI data.

Note that the forecasting errors are lowest in the BPNN model in all stages. In one-step-ahead predictions, BPNN forecasts the best among the ANN-based algorithms in all three stages, but the forecasts of the other ANN-based models are not better than those of the econometric models. The accuracies of the ARIMA and GARCH models are similar, but the accuracies of the ANN-based models vary dramatically across different stages.

Table 5 compares the seven-steps-ahead predictive performances. In the longer forecasting horizon of weekly data, the errors of the econometric models rapidly increase over those made using daily data, but the ANN-based algorithms maintain their predictive power, and the MAPE is lower than 0.09 for all forecasting results. BPNN also outperforms all other ANN-based models in all three stages. In addition, most ANN-base methods perform better than simple naïve models and the econometric models perform worse than the simple baseline models in some stages.

TABLE 5. Predictive performance of seven-step head forecasting (weekly data).

Model	MAPE	RMSE	MASE	Dsta	NSE
Stage I: 1999.11.5-2003.9.12					
ARIMA(2,1,4)	0.29500	1454.76	2.82	0.14	0.5801
GARCH(2,1)	0.28200	1410.15	2.26	0.29	0.5693
BPNN(35,150,7)*	0.00960	21.49	0.94	0.45	0.6138
RBFNN(35,200,7)	0.01710	32.83	1.23	1.00	0.8031
ELM(35,50,7)	0.01830	37.49	1.47	0.71	0.6671
Stage II: 2008.7.25-2012.6.8					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(2,1,4)	0.13326	160.52	0.51	0.43	0.9842
GARCH(2,1)	0.15885	187.58	0.27	0.29	0.9784
BPNN(35,140,7)*	0.06888	96.64	0.16	0.57	0.9943
RBFNN(35,80,7)	0.08770	96.45	0.22	0.14	0.9809
ELM(35,80,7)	0.08425	94.87	0.11	0.14	0.9945
Stage III: 2011.12.16-2015.10.16					
Model	MAPE	RMSE	MASE	Dsta	NSE
ARIMA(8,1,1)	0.21100	114.22	1.35	0.57	0.3558
GARCH(0,1)	0.23800	134.62	1.83	0.57	0.4853
BPNN(35,120,7)*	0.06500	45.87	0.83	0.71	0.4968
RBFNN(35,180,7)	0.08502	43.30	0.96	0.71	0.8999
ELM(35,40,7)	0.08915	44.42	0.85	0.71	0.8060

C. MONTHLY BDI PREDICTIONS

Because the time horizon of the monthly BDI data is relatively small, we increase the observation length to 100 and make out-sample predictions of one-step-ahead and seven-steps-ahead values using econometric models. We use the same sample data as the training data for ANN-based algorithms, and we predict the one-step-ahead and seven-steps-ahead values. Table 6 shows that in one-step-ahead prediction the ELM model outperforms the others in the first and third stages, and that RBFNN is the best in

TABLE 6. Predictive performance of one step head forecasting (monthly data).

Model	MAPE	RMSE	MASE	Dsta	NSE
Stage I: 1999.12-2008.3					
AMRIMA(2,1,7)	0.14697	103.56	0.29	1.00	0.9420
GARCH(0,1)	0.12091	974.85	0.25	1.00	0.9608
BPNN(5,40,1)	0.12357	996.28	0.27	1.00	0.9590
RBFNN(5,110,1)	0.26674	153.26	0.43	1.00	0.8091
ELM(5,10,1)*	0.03690	31.29	0.08	1.00	0.9963
Stage II: 2003.8-2011.11					
Model	MAPE	RMSE	MASE	Dsta	NSE
AMRIMA(9,1,4)	0.01205	22.51	0.04	1.00	0.9999
GARCH(0,1)	0.04274	79.88	0.12	1.00	0.9985
BPNN(5,80,1)	0.00516	9.64	0.02	1.00	1.0000
RBFNN(5,120,1)*	0.00415	7.76	0.01	1.00	1.0000
ELM(5,30,1)	0.06814	127.34	0.21	0.00	0.9962
Stage III: 2008.3-2016.6					
Model	MAPE	RMSE	MASE	Dsta	NSE
AMRIMA(8,1,8)	0.13838	129.72	0.79	1.00	0.9979
GARCH(1,4)	0.13792	129.29	0.79	1.00	0.9979
BPNN(5,40,1)	0.18868	164.36	0.78	0.00	0.9835
RBFNN(5,130,1)	0.14111	134.16	0.84	1.00	0.9987
ELM(5,40,1)*	0.08160	76.50	0.23	1.00	0.9993

the second stage. All proposed models outperform simple baseline methods, even though the MAPE value is higher than 0.2.

Comparing the predictive performances of seven-steps-ahead forecasting, Table 7 shows that the errors of both econometric models and ANN-based algorithms increase, and that in the second testing stage the MAPE values in all models are > 0.2, indicating that no model produces a usable prediction.

TABLE 7. **Predictive performance of seven-step head forecasting (monthly data).**

Stage I: 1999.12-2008.3					
Model	MAPE	RMSE	MASE	Dsta	NSE
AMRIMA(2,1,7)	0.27290	1291.80	1.23	0.57	0.8427
GARCH(0,1)	0.19540	991.10	1.75	0.71	0.8887
BPNN(35,120,7)*	0.11140	644.14	0.92	1.00	0.8864
RBFNN(35,200,7)	0.12700	137.98	2.01	0.71	0.8234
ELM(35,50,7)	0.11610	111.34	1.32	0.86	0.8231
Stage II: 2003.8-2011.11					
Model	MAPE	RMSE	MASE	Dsta	NSE
AMRIMA(9,1,4)	0.36600	266.96	0.33	0.71	0.9509
GARCH(0,1)	0.23290	410.24	0.41	0.29	0.9252
BPNN(35,150,7)	0.29490	409.47	0.79	0.71	0.6981
RBFNN(35,180,7)*	0.22900	389.89	0.23	0.71	0.8172
ELM(35,40,7)	0.36560	441.53	0.43	0.71	0.8973
Stage III: 2008.3-2016.6					
Model	MAPE	RMSE	MASE	Dsta	NSE
AMRIMA(8,1,8)	0.33320	373.49	0.28	0.57	0.8830
GARCH(1,4)	0.25370	237.13	0.29	0.57	0.8495
BPNN(35,110,7)	0.14800	137.37	0.13	0.57	0.9759
RBFNN(35,170,7)	0.10830	114.24	0.11	0.71	0.9572
ELM(35,30,7)*	0.10290	104.83	0.08	0.71	0.9565

Previous studies show that when the time scale of the datasets is longer, the accuracy of the forecasting model decreases. Thus the effectiveness and robustness of a forecasting model can be evaluated using the relationship between the predictive accuracy and the time scale of the dataset and the error changes between single-step and multi-step predictions. Figures 2 and 3 compare the values of the average MAPE and RSME of five forecasting models for daily, weekly, and monthly datasets when making single-step and multi-step predictions.

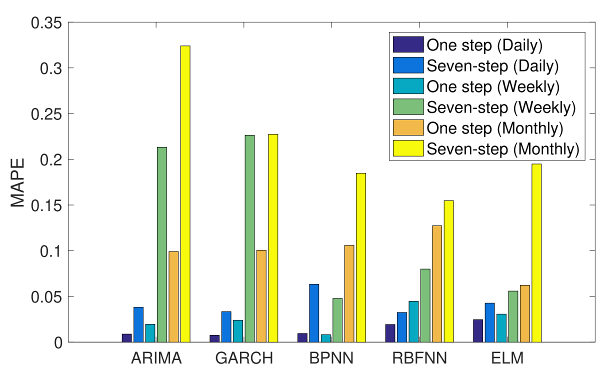

FIGURE 2. **Average MAPE comparison of different time scale dataset in one step and seven-step ahead predictions.**

Figures 2 and 3 show that the behavior of econometric models and ANN-based algorithms differs between

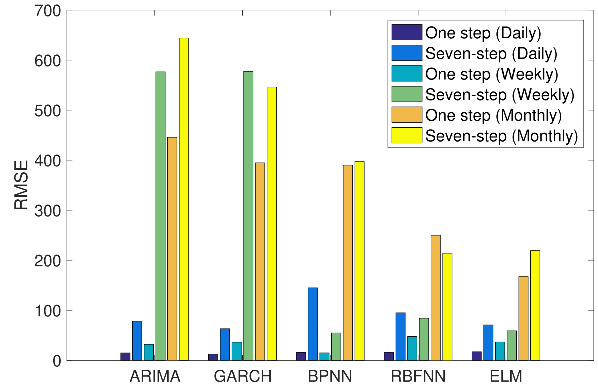

FIGURE 3. **Average RMSE comparison of different time scale dataset in one step and seven-step ahead predictions .**

the various times-ahead forecasts with different time scale datasets. Unlike daily BDI predictions, MAPE and RMSE both increase when applying econometric models to predict weekly and monthly BDI, and the MAPE values of ANN-based algorithms slowly increases under the same conditions. In addition, the MAPE and RSME values of ANN-based algorithms also increase more slowly than econometric models when the periods of out-sample forecasting are increased. This indicates that ANN-based models are more effective when applied to longer time-scale datasets and multi-period predictions.

D. CHALLENGING SITUATION FORECAST COMPARISON

In order to compare the robustness of prediction performance in the challenging situation, we use different models to forecast the extreme fluctuation of BDI during world financial crisis in 2008. We apply moving sample data from the starting period of November 1999 to November 2007, and make one-step ahead and seven-step ahead forecasts respectively based on daily, weekly and monthly BDI data. The out-sample test data covers May 2008 to December 2008 when BDI plunged sharply from the peak that is close to 12000 to around 700 points.

Fig.4, Fig.5 and Fig 6 compare the one step ahead and seven-step ahead forecasted results with the actual value of BDI in the out-sample prediction period from May 2008 to December 2008 based on daily, weekly and monthly data respectively.

As we can observe, for daily data prediction, all five models predict the trend of BDI well while econometric models demonstrate lower errors than ANN models in one step ahead forecasting case. GRACH model outperforms others. But in seven-step ahead forecasting case, ANN models present much powerful prediction than econometric models.

This result is same in weekly data prediction shown in Fig.3, in which econometric models have advantages compared to ANN models in one step forecasting, however ANN models superior to econometric approaches.

For monthly data prediction shown in Fig.4, all ANN models can reasonably well reproduce the trend of BDI

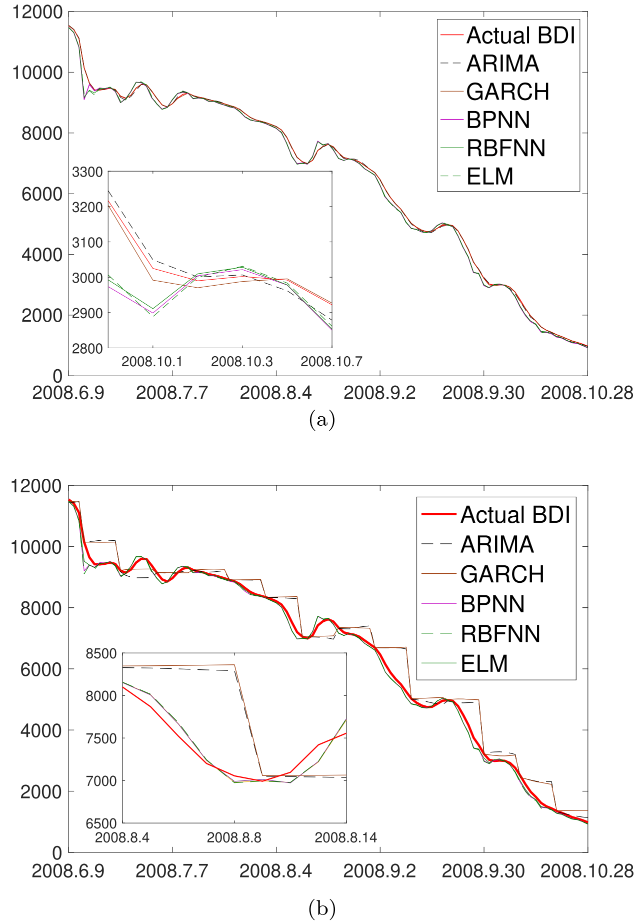

FIGURE 4. Forecasts of two econometric models and three ANN algorithms with actual value of daily BDI. (a) One step ahead forecasting results vs actual value. (a) Seven-step ahead forecasting results vs actual value.

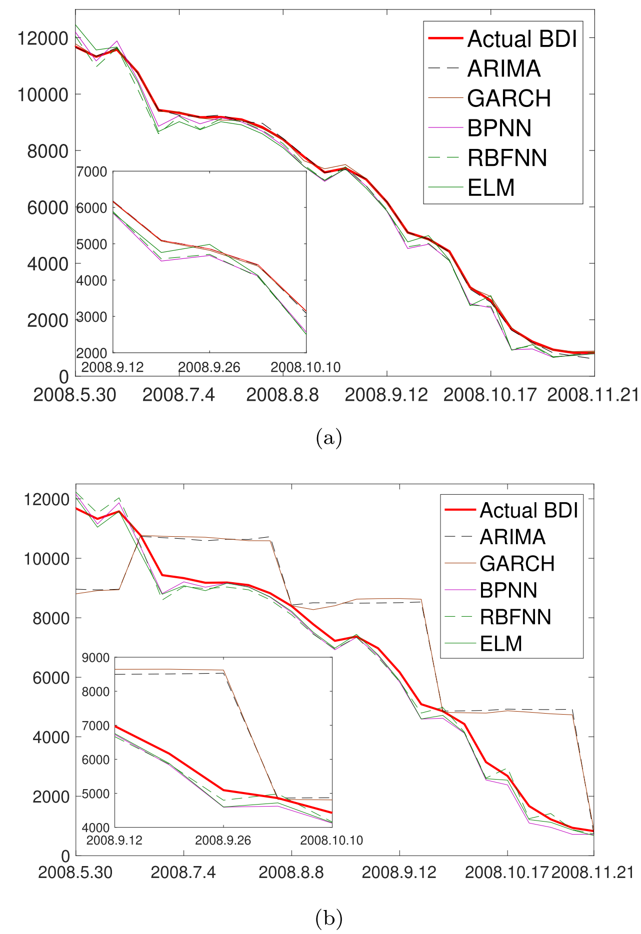

FIGURE 5. Forecasts of two econometric models and three ANN algorithms with actual value of weekly BDI . (a) One step ahead forecasting results vs actual value. (a) Seven-steps ahead forecasting results vs actual value.

both in one step and seven-step ahead forecasting but with the exception of RBFNN model. The prediction of RBFNN deviated from actual value dramatically while BPNN and ELM still present well predictive power. Similar to daily and weekly data, econometric models outperform in one step ahead forecasting while BPNN and ELM excel in seven-step ahead case.

Through the comparison, the BPNN model can be regarded as the most robust one for both short term and long term prediction as well as in data with various time scales.

V. DIEBOLD AND MARIANO TEST

In order to evaluate whether there is any statistical significant difference between the models we aim to compare, we employ Diebold and Mariano test (hereafter, the DM test).

Diebold and Mariano [49], [50] introduced a statistical test for the null hypothesis of equal forecast accuracy between two competing models. Here the loss function is set to be the absolute error (AE), defined as $AE = |\hat{X(t)} - X(t)|$, here where $\hat{X}(t)$ and $X(t)$ are the predicted and real values at time t. The null hypothesis is that the AE value of the tested model equaling to that of the alternative model. The DM statistic is defined

$$DM = \frac{\overline{D}}{\sqrt{V_{\overline{D}}/N}}, \tag{21}$$

where $\overline{D} = \frac{1}{N}\sum_{t=1}^{N} d(t)$, $d(t) = e_{test}(t) - e_{alternative}(t)$, in which $e(t)$ is the AE value at time t, $V_{\overline{D}} = \gamma_0 + 2\sum_{\infty}^{q=1}\gamma_q$, $\gamma_q = cov(d_t, d_{t-q})$

In using the DM test to statistically compare the accuracy of econometric and ANN-based forecasting models, the DM value of Eq.(21) and the p-value are used to measure how much the difference is between test model over the alternative model.

Table 8 lists the DM test results for daily, weekly and monthly BDI datasets respectively. If the p-value is less than 0.05, we reject the null hypothesis that there's no difference between two proposed forecasting models. It is noted that for all three datasets ANN-based model have significant different accuraies with econometric models. For daily dataset, ARIMA and GARCH present equal forecasting accuraies while different ANN-based models also exhibit no statistically difference in accuraies. However, for weekly datasets ARIMA and GARCH perform differently and

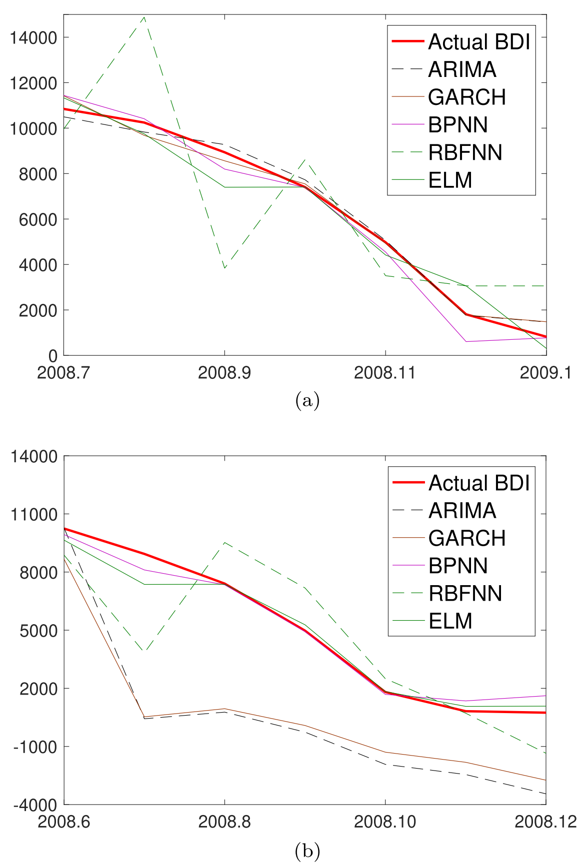

FIGURE 6. Forecasts of two econometric models and three ANN algorithms with actual value of monthly BDI . (a) One step ahead forecasting results vs actual value. (a) Seven-step ahead forecasting results vs actual value.

TABLE 8. DM test results for econometric and ANN-based models.

Data type	Tested model	Reference model			
		GARCH	BPNN	RBFNN	ELM
Daily	ARIMA	-1.6328 (0.1182)	2.4308* (0.0246)	2.4594* (0.0231)	2.4221* (0.0251)
	GARCH		2.2701* (0.0344)	2.2909* (0.033)	2.2627* (0.0349)
	BPNN			0.3971 (0.6955)	-1.7523 (0.095)
	RBFNN				-1.2322 (0.2322)
Weekly	ARIMA	-2.8086* (0.0148)	2.2093* (0.0457)	2.191* (0.0473)	2.2276* (0.0442)
	GARCH		2.224* (0.0445)	2.2061* (0.046)	2.2461* (0.0427)
	BPNN			-1.6596 (0.1126)	-2.0614 (0.0525)
	RBFNN				-1.9883 (0.0606)
Monthly	ARIMA	-4.5004* (0.0041)	-5.4579* (0.0016)	-5.5169* (0.0015)	-5.3391* (0.0018)
	GARCH		-5.4389* (0.0016)	-5.4983* (0.0015)	-5.3198* (0.0018)
	BPNN			-7.8986* (0.0002)	11.7279* (0.00002)
	RBFNN				10.485* (0.00004)

*: Reject the null hypothesis of no difference between two compared predictive models at the significance level 0.05

ANN-based models still present equal forecasting accuraies. When the time-scale of the dataset is longer such as for monthly dataset, all accuraries of five forecasting models are statistically different. This indicates that econometric methods and ANN-based model are two distict methodology caterigories. And within the caterigories, the longer the time scale of the data to predict, the obvious the difference among various forecasting models.

VI. CONCLUSION

We have studied and compared the performance of two forecasting techniques: the econometric approach and the computational approach of artificial neural network (ANN) algorithms. We use BDI data from 1999 to 2018 to test the predictive power of the proposed models.

(i) We find that the ANN approach predicts the most accurate weekly and monthly BDI values. Econometric models produce good one-step-ahead daily values, but the ANN models produce the best seven-step ahead forecasting values. Thus ARIMA and GARCH produce short-term daily BDI predictions that are better than those of the ANN models.

(ii) We also find that ANN models are sensitive to input data and forecasting horizon. In different sets of training sample data, ANN model predictions vary greatly, indicating that no particular model is the best in all situations.

(iii) Econometric models and ANN algorithms have both advantages and disadvantages. The long-term forecasting based on monthly data of both econometric models and ANN algorithms is weak. Thus a possible course of future study would be to develop a hybrid forecasting method that combines econometrics and the computational approach of ANN.

REFERENCES

[1] M. Stopford, "How shipping has changed the world and the social impact of shipping," in *Proc. Global Maritime Environ. Congr.*, Hamburg, Germany, vol. 7, 2010, pp. 1–10.

[2] C. Chistè and G. van Vuuren, "Investigating the cyclical behaviour of the dry bulk shipping market," *Maritime Policy Manage.*, vol. 41, no. 1, pp. 1–19, 2014.

[3] X. Zhang, B. Podobnik, D. Y. Kenett, and H. E. Stanley, "Systemic risk and causality dynamics of the world international shipping market," *Phys. A, Stat. Mech. Appl.*, vol. 415, pp. 43–53, Dec. 2014.

[4] G. De Monie, J.-P. Rodrigue, and T. Notteboom, "Economic cycles in maritime shipping and ports: The path to the crisis of 2008," in *Proc. Int. Workshop Integrating Maritime Transport Value Chains*, Montréal, QC, Canada, 2009, pp. 9–12.

[5] D. S. Jacks and K. Pendakur, "Global trade and the maritime transport revolution," *Rev. Econ. Statist.*, vol. 92, no. 4, pp. 745–755, 2010.

[6] F. Lin and N. C. Sim, "Trade, income and the baltic dry index," *Eur. Econ. Rev.*, vol. 59, pp. 1–18, Apr. 2013.

[7] L. Pascali, "The wind of change: Maritime technology, trade, and economic development," *Amer. Econ. Rev.*, vol. 107, no. 9, pp. 2821–2854, 2017.

[8] K. Cullinane, K. Mason, and M. Cape, "A comparison of models for forecasting the baltic freight index: Box-Jenkins revisited," *Int. J. Maritime Econ.*, vol. 1, no. 2, pp. 15–39, 1999.

[9] M. G. Kavussanos and A. H. Alizadeh-M, "Seasonality patterns in dry bulk shipping spot and time charter freight rates," *Transp. Res. E, Logistics Transp. Rev.*, vol. 37, no. 6, pp. 443–467, 2001.

[10] R. Batchelor, A. Alizadeh, and I. Visvikis, "Forecasting spot and forward prices in the international freight market," *Int. J. Forecasting*, vol. 23, no. 1, pp. 101–114, 2007.

[11] S. Chen, H. Meersman, and E. van de Voorde, "Forecasting spot rates at main routes in the dry bulk market," *Maritime Econ. Logistics*, vol. 14, no. 4, pp. 498–537, 2012.

[12] V. Tsioumas, S. Papadimitriou, Y. Smirlis, and S. Z. Zahran, "A novel approach to forecasting the bulk freight market," *Asian J. Shipping Logistics*, vol. 33, no. 1, pp. 33–41, 2017.
[13] R. Adland, F. E. Benth, and S. Koekebakker, "Multivariate modeling and analysis of regional ocean freight rates," *Transp. Res. E, Logistics Transp. Rev.*, vol. 113, pp. 194–221, May 2018.
[14] M. Stopford, *Maritime Economics*. Evanston, IL, USA: Routledge, 2013.
[15] S. Dai, Y. Zeng, and F. Chen, "The scaling behavior of bulk freight rate volatility," *Int. J. Transport Econ.*, vol. 43, nos. 1–2, pp. 85–104, 2016.
[16] Y. Leonov and V. Nikolov, "A wavelet and neural network model for the prediction of dry bulk shipping indices," *Maritime Econ., Logistics*, vol. 14, no. 3, pp. 319–333, 2012.
[17] E. Bulut, O. Duru, and S. Yoshida, "A fuzzy integrated logical forecasting (FILF) model of time charter rates in dry bulk shipping: A vector autoregressive design of fuzzy time series with fuzzy c-means clustering," *Maritime Econ., Logistics*, vol. 14, no. 3, pp. 300–318, 2012.
[18] O. Duru, E. Bulut, and S. Yoshida, "A fuzzy extended DELPHI method for adjustment of statistical time series prediction: An empirical study on dry bulk freight market case," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 840–848, 2012.
[19] Q. Han, B. Yan, G. Ning, and B. Yu, "Forecasting dry bulk freight index with improved SVM," *Math. Problems Eng.*, vol. 2014, Jun. 2014, Art. no. 460684.
[20] Q. Zeng, C. Qu, A. K. Y. Ng, and X. Zhao, "A new approach for baltic dry index forecasting based on empirical mode decomposition and neural networks," *Maritime Econ., Logistics*, vol. 18, no. 2, pp. 192–210, 2016.
[21] C.-C. Chou and K.-S. Lin, "A fuzzy neural network combined with technical indicators and its application to baltic dry index forecasting," *J. Marine Eng., Technol.*, vol. 17, no. 2, pp. 2–9, Jul. 2018.
[22] B. Şahin, S. Gürgen, B. Ünver, and İ. Altin, "Forecasting the baltic dry index by using an artificial neural network approach," *Turkish J. Electr. Eng., Comput. Sci.*, vol. 26, no. 3, pp. 1673–1684, 2018.
[23] I. S. Markham and T. R. Rakes, "The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression," *Comput., Oper. Res.*, vol. 25, no. 4, pp. 251–263, 1998.
[24] C. Chatfield, *The Analysis of Time Series: An Introduction*. Boca Raton, FL, USA: CRC Press, 2016.
[25] T. Bollerslev, "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model," *Rev. Econ. Statist.*, vol. 72, no. 3, pp. 498–505, 1990.
[26] R. F. Engle and K. Sheppard, "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," Nat. Bureau Econ. Res., Cambridge, MA, USA, Tech. Rep. w8554, 2001.
[27] D. Lien, Y. K. Tse, and A. K. Tsui, "Evaluating the hedging performance of the constant-correlation GARCH model," *Appl. Financial Econ.*, vol. 12, no. 11, pp. 791–798, 2002.
[28] R. C. Garcia, J. Contreras, M. V. Akkeren, and J. B. C. Garcia, "A GARCH forecasting model to predict day-ahead electricity prices," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 867–874, May 2005.
[29] S. Kim, "Forecasting Internet traffic by using seasonal GARCH models," *J. Commun. Netw.*, vol. 13, no. 6, pp. 621–624, 2011.
[30] A. Wilhelmsson, "GARCH forecasting performance under different distribution assumptions," *J. Forecasting*, vol. 25, no. 8, pp. 561–578, 2006.
[31] H. T. Pham and B.-S. Yang, "Estimation and forecasting of machine health condition using ARMA/GARCH model," *Mech. Syst. Signal Process.*, vol. 24, no. 2, pp. 546–558, Feb. 2010.
[32] Z. Ghahramani, "Probabilistic machine learning and artificial intelligence," *Nature*, vol. 521, no. 7553, p. 452, 2015.
[33] M. Wang *et al.*, "A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms," *Appl. Energy*, vol. 220, pp. 480–495, Jun. 2018.
[34] A. R. Carvalho, F. M. Ramos, and A. A. Chaves, "Metaheuristics for the feedforward artificial neural network (ANN) architecture optimization problem," *Neural Comput. Appl.*, vol. 20, no. 8, pp. 1273–1284, 2011.
[35] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, Mar. 2010.
[36] A. Goh, "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.*, vol. 9, no. 3, pp. 143–151, 1995.
[37] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Syst. Appl.*, vol. 42, no. 2, pp. 855–863, 2015.
[38] R. Law, "Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting," *Tourism Manage.*, vol. 21, no. 4, pp. 331–340, 2000.
[39] J.-S. R. Jang and C.-T. Sun, "Functional equivalence between radial basis function networks and fuzzy inference systems," *IEEE Trans. Neural Netw.*, vol. 4, no. 1, pp. 156–159, Jan. 1993.
[40] E. S. Chng, S. Chen, and B. Mulgrew, "Gradient radial basis function networks for nonlinear and nonstationary time series prediction," *IEEE Trans. Neural Netw.*, vol. 7, no. 1, pp. 190–194, Jan. 1996.
[41] R. Zemouri, D. Racoceanu, and N. Zerhouni, "Recurrent radial basis function network for time-series prediction," *Eng. Appl. Artif. Intell.*, vol. 16, nos. 5–6, pp. 453–463, 2003.
[42] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[43] X. Wang and M. Han, "Online sequential extreme learning machine with kernels for nonstationary time series prediction," *Neurocomputing*, vol. 145, pp. 90–97, Dec. 2014.
[44] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38–48, 2016.
[45] E. W. Steyerberg *et al.*, "Assessing the performance of prediction models: A framework for some traditional and novel measures," *Epidemiology*, vol. 21, no. 1, p. 128, 2010.
[46] J. McKenzie, "Mean absolute percentage error and bias in economic forecasting," *Economics Lett.*, vol. 113, no. 3, pp. 259–262, 2011.
[47] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006.
[48] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu, "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series," *J. Hydrol.*, vol. 374, nos. 3–4, pp. 294–306, 2009.
[49] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus., Econ. Statist.*, vol. 20, no. 1, pp. 134–144, 2002.
[50] F. X. Diebold, "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests," *J. Bus., Econ. Statist.*, vol. 33, no. 1, p. 1, 2015.

XIN ZHANG received the master's and Ph.D. degrees in system science. In 2013 and 2018, she was a Research Fellow with Boston University. She is currently an Associate Professor with the College of Communication and Transport, Shanghai Maritime University, China. Since 2010, she has been directing 6 research projects, including the National Natural Science Foundation of China, the Social Science Foundation of Ministration of Education, the Pujiang High-level Talent Project of Shanghai, and the R&D Award of ITOPF. Her research interests include complex system modeling, data mining, and their applications to financial engineering and risk management.

TIANYUAN XUE received the bachelor's degree from the College of Communication and Transport, Shanghai Maritime University, China, in 2016, where he is currently pursuing the master's degree. He is currently focusing on a project on shipping market data. His research interests are machine learning and data mining.

H. EUGENE STANLEY received the Ph.D. degree in physics from Harvard University in 1967. He is an American Physicist and currently a University Professor with Boston University, USA. He has made fundamental contributions to complex systems and is one of the founding fathers of econophysics. His current research interests include complexity science and econometrics. He was elected to the U.S. National Academy of Sciences in 2004.