

# Time charter rate forecasting by Parsimonious Intelligent Support Vector regression Search Engine

Xuefei Song<sup>1</sup> and Zhong Shuo Chen<sup>2\*</sup> 

<sup>1</sup> Jilin Provincial Experimental School, Changchun 130022, China

<sup>2</sup> School of Intelligent Finance and Business, Xi'an Jiaotong-Liverpool University, Suzhou 215400, China

\* Corresponding author, E-mail: [zhongshuo.chen@xjtlu.edu.cn](mailto:zhongshuo.chen@xjtlu.edu.cn)

## Abstract

Reliable and accurate forecasts of time charter rates are crucial for shipowners navigating the highly volatile global shipping market. This paper introduces a novel framework, the Parsimonious Intelligent Support Vector Regression Search Engine (SVRIMSE), designed to enhance the forecasting of time charter rates. Unlike traditional models that rely on subjective assumptions regarding causality, time lags, and training set sizes, which can compromise accuracy and pose challenges for non-experts, SVRIMSE offers a more robust solution. It not only delivers precise forecasts but also autonomously identifies significant explanatory variables without requiring prior knowledge or assumptions about the shipping market. Our comparative analysis demonstrates that SVRIMSE outperforms several baseline models across three error metrics. Notably, the results highlight Fleet Development (Number) and Scrap Value as the most influential variables in forecasting time charter rates. These findings provide both accurate forecasts and valuable insights into the factors impacting the freight market, offering a significant contribution to the field.

**Keywords:** Maritime transport; Shipping market; Machine learning; Forecasting; Time charter rate

**Citation:** Song X, Chen ZS. 2025. Time charter rate forecasting by Parsimonious Intelligent Support Vector regression Search Engine. *Digital Transportation and Safety* 4(3): 188–194 <https://doi.org/10.48130/dts-0025-0017>

## Introduction

Anticipating future changes, which helps make opportune decisions, is significant in the global shipping market. The global shipping market is highly dynamic and volatile due to trade wars, economic recessions, rigid market disequilibria, trends, and seasonality. This huge volatility poses critical difficulties for shipping companies, ship owners, shipyards, and brokers in making opportune decisions<sup>[1,2]</sup>. Therefore, forecasting the future changes of the global shipping market accurately is significant because of its effect on managerial decisions. Many researchers put much effort into understanding the mechanism of the seaborne freight market. The pioneer analysis of the balance between supply and demand of the seaborne freight market was conducted by Rothbarth<sup>[3]</sup>. Zannetos<sup>[4]</sup> investigated the relationship between the period and the spot shipping market and proposed seminal theoretical models. All the studies showed that the freight market is highly volatile and difficult to model. However, most literature analyzing the seaborne freight market is based on professional and rational postulations about the economic predictors of the freight market.

Since the freight market is highly volatile, some literature focuses on forecasting charter rates using statistical models or machine learning models<sup>[5–11]</sup>. For example, Li & Parsons<sup>[7]</sup> implemented neural networks to forecast the freight rates of oil tankers for the first time and compared two neural networks with different inputs. However, most of the researchers designed the model's structure arbitrarily, and did not consider optimizing the training sample size, and selected the explanatory variables based on causality. Such relaxation may deteriorate the forecasting accuracy because of the non-objective design of the models.

This work contributes to seaborne crude oil transportation by proposing a novel and objective framework of Support Vector Regression (SVR) to forecast the time charter rates. To the best

knowledge of the authors, SVR is utilized to forecast the time charter rates for the first time. In addition, the existing literature deals with time charter rate forecasting based on subjective assumptions of the model's structure and explanatory variables. However, our method which establishes the model based on the accuracy of the validation set, is automatic and objective. Moreover, most of the related literature analyzed the freight market from the perspective of supply and demand, however, our data-driven method can provide insights into the significant explanatory variables based on forecasting accuracy.

The rest of this paper is organized as follows. A review of the studies analysing the freight market and literature about the forecasting methods is presented. The models compared in this study are then described briefly in the next section. Then, the source and pre-processing steps for the data are presented. After introducing the data, we present the comparison results on the time charter rates in terms of three error metrics. The prediction curves on the test set and hyper-parameter settings of the model are presented. Then, a detailed analysis of the models' forecasting accuracy and explanatory variables' contribution is provided. Finally, a conclusion is drawn, and limitations and future work are briefly described.

## Literature review

Many studies focus on investigating the mechanism of the freight market and the predictors that affect the freight rates. The pioneer analysis of the seaborne freight market considering the balance between supply and demand was conducted by Rothbarth<sup>[3]</sup>. Zannetos<sup>[4]</sup> proposed seminal theoretical models based on the relationship between the spot and period shipping market. Hale & Vanags<sup>[12]</sup> and Glen et al.<sup>[13]</sup> proposed frameworks to figure out the relationship between the spot and the long-term markets.

Strandenes<sup>[14]</sup> investigated the effect of expectations of the spot market on the period market.

Since making opportune decisions in the freight market is important for making profits, some researchers try to anticipate the future by using forecasting models. The pioneering work carried out by Beenstock & Vergottis<sup>[15,16]</sup> utilized the econometric model to forecast the annual freights and period charter rates. Glen & Martin<sup>[17]</sup>, and Veenstra<sup>[18]</sup> implemented the Vector Autoregression (VAR) model and provided insights into the term structure of freight rates. The authors concluded that the spread between freight and charter rates did not contribute to forecasting. The volatility in the spot and time charter markets was studied by the extension of the Autoregressive Conditional Heteroskedasticity (ARCH) class of models by Kavussanos<sup>[19]</sup>.

Instead of econometric and statistical forecasting models, intelligent algorithms were touched on in some literature. Bulut et al.<sup>[20]</sup> proposed a novel vector autoregressive fuzzy integrated logical model that was robust to noise to forecast time charter rates of two kinds of bulk carriers and compared the performance with several famous fuzzy time series models. Besides fuzzy time series, neural networks show a strong ability to discover patterns and learn knowledge from historical data. The pioneering work using neural networks to forecast oil tankers' spot freight rates was conducted by Li & Parsons<sup>[7]</sup>. They showed that the neural networks outperformed the conventional forecasting method, Autoregressive Integrated Moving Average (ARIMA). Lyridis et al.<sup>[8]</sup> implemented the neural networks to forecast Very Large Crude Carrier (VLCC) spot freight rates, but the results were not compared with baseline models. Eslami et al.<sup>[5]</sup> combined neural networks with a Genetic Algorithm (GA) to forecast the tanker freight rates with parsimonious variables. The neural networks used in the above literature are feed-forward structures and such networks are called Multi-Layer Perceptron (MLP). Santos et al.<sup>[9]</sup> showed that the MLP and Radial Basis Function (RBF) network outperformed ARIMA on the period charter rates of VLCC tankers.

Intelligent forecasting algorithms have been proven to outperform conventional econometric and statistical models in various datasets<sup>[21–26]</sup>. The modern forecasting model aims at learning a function from the historical data and extracting valuable temporal information. Besides the literature mentioned in the above paragraph, many researchers have developed and implemented intelligent algorithms to forecast various time series from the shipping market. Syriopoulos et al.<sup>[27]</sup> applied the Support Vector Machine (SVM) to forecast ship prices and the results showed that SVM outperformed ARIMA. Zeng et al.<sup>[28]</sup> utilized the Empirical Mode Decomposition (EMD) to decompose the Baltic Dry Index (BDI) first and feed the sub-series into neural networks to generate predictions. Many researchers put much effort into forecasting the BDI series, such as the fuzzy time series<sup>[29]</sup>, SVM<sup>[30,31]</sup>, neural networks<sup>[32–34]</sup>, and hybrid models<sup>[35]</sup>.

## Methodology

### Brief introduction of baseline models

First, suppose the input multivariate data is  $\gamma = \mathbb{R}^{N \times m}$  and the target time series are  $\phi_t$ . Then, a concise overview of the baseline forecasting models is presented.

**Naïve forecast:** The naïve forecast is the most common benchmark for forecasting tasks. The naïve forecast uses the most recent value as its forecast<sup>[1]</sup>. Naïve forecast is essentially a no-rule forecasting scheme, but a forecasting algorithm must outperform it.

**ARIMA:** The ARIMA model which aims at describing the autocorrelation of time series is another popular baseline in the forecasting

field. ARIMA succeeds in various forecasting tasks because it is an integration of differencing, autoregressive, and moving average models. The differencing helps stabilize the time series. The whole model ARIMA ( $p, d, q$ ) is defined as<sup>[1,2,36]</sup>:

$$\phi_t = c + a_1 \phi_{t-1} + \dots + a_p \phi_{t-p} + b_1 \epsilon_t + \dots + b_q \epsilon_q + \epsilon_t, \quad (1)$$

where,  $\phi_t$  represents the time series after  $d$  times differencing,  $\epsilon$  is the residual, and parameters  $a$ ,  $b$ , and  $c$  need optimization.

**SVR:** The key idea of SVR is to transform the linear data into high-dimensional space with a suitable non-linear transformation and solve the problem in that space in a linear fashion<sup>[1]</sup>. First, the SVR transforms the input variables  $\gamma$  into high-dimensional space by non-linear function  $\psi(\gamma)$ . Then the SVR learns a function which minimizes the  $\epsilon$ -loss function<sup>[21]</sup>:

$$loss = C \sum_{i=1}^n (\xi_i + \xi_i^*) + \frac{1}{2} \|\omega\|^2 \quad (2)$$

with the constraints defined as,

$$\phi_i - ((\omega \psi(\gamma_i))) \leq \epsilon + \xi_i \quad (3)$$

$$((\omega \psi(\gamma_i))) - \phi_i \leq \epsilon + \xi_i^* \quad (4)$$

where,  $\xi_i$  and  $\xi_i^* \geq 0$ ,  $i = 1, 2, 3, \dots, n$ . The  $\xi_i$  and  $\xi_i^*$  are defined as the slack variables and  $C$  is the regularization parameter.  $\epsilon$  denotes the region where no penalty is counted. A small  $\epsilon$  means the SVR puts more effort into fitting the data than the SVR with a large  $\epsilon$ , but a small  $\epsilon$  causes over-fitting. By introducing the Lagrange multipliers, this problem is transformed into:

$$f(\gamma, a_i, a_i^*) = \sum_{i=1}^n (a_i - a_i^*) K(\gamma, \gamma_i) + b \quad (5)$$

where, the kernel function  $K(\gamma, \gamma_i) = \psi(\gamma) \psi(\gamma_i)$ . The Lagrange multipliers  $a_i$  and  $a_i^*$  can be calculated by solving the dual formulation:

$$\min \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*) (a_j - a_j^*) K(\gamma_i, \gamma_j) - \sum_{i=1}^n \phi_i (a_i - a_i^*) + \epsilon \sum_{i=1}^n (a_i + a_i^*) \quad (6)$$

subject to:

$$\sum_{i=1}^n (a_i - a_i^*) = 0, \quad 0 \leq a_i, \quad a_i^* \leq C, \quad i = 1, 2, \dots, n \quad (7)$$

**MLP:** MLP is a stacked neural network with a feed-forward structure<sup>[1]</sup>. The hidden layer learns a meaningful representation of its input based on the task. MLP's success is essential because of its strong ability to extract the features in an end-to-end fashion. Different from the hand-crafted or unsupervised features, the hidden nodes learn the features which are suitable for the given tasks. Each hidden node has its non-linear activation function which transforms the input data into a different space. Once the information propagates from the input layer to the output layer, the loss can be computed. Then the famous back-propagation algorithm can be applied to optimize the weights of all layers<sup>[2]</sup>. In addition, many improved versions of learning algorithms are proposed.

**Decision Tree:** Decision Tree (DT) is a data-driven model with a tree structure. Each node in the tree functions the partitioning of the input space<sup>[37]</sup>. As a result, a learned decision tree divides the whole input space into many areas, and then the regression error is optimized for each area.

### Parsimonious Intelligent SVR Search Engine

Inspired by Bekiroglu et al.<sup>[38]</sup>, we propose a novel model, called Parsimonious Intelligent SVR Search Engine (SVRIMSE). Bekiroglu et al. proposed the Intelligent Model Search Engine (RIMSE) to forecast the crude oil price and the RIMSE model is a automatic and object forecasting model<sup>[38]</sup>. RIMSE assumes no prior knowledge about the data and model's structure and establishes the forecasting model based on accuracy of out-of-sample data. Although RIMSE achieved great success in crude oil price forecasting and outperformed the famous forecasting models, it is essentially a linear system which

cannot learn the complex and non-linear relationship among the various explanatory variables. Therefore, we propose to use SVR to replace the linear regression (Ridge regression) to enhance the learning of non-linear patterns.

In addition, most of the existing works focus on a dense collection of time lags, which may cause overfitting and introduce irrelevant inputs. To handle these shortcomings, we propose to optimize the time lags in a parsimonious fashion. In other words, instead of selecting a successive sequence of time lags, the combinations of time lags are considered. Finally, to the best knowledge of the authors, there is little SVR-related forecasting literature considering optimizing the training sample size for SVR.

First, we transform all the time series into the first difference. Then, the validation and test set are split from the whole dataset. After the split, the following steps are implemented.

Step 1. The shrinking window size is chosen, which defines the size of the training set. Such a partitioning scheme is shown in Fig. 1. When the shrinking window becomes larger, the training set size becomes smaller. A larger shrinking window offers less data to train the SVR.

Step 2. One set of explanatory variables among a large set of variables is chosen. Although it is well-known that more explanatory variables can help the model learn more complex knowledge from the data, irrelevant variables will cause a burden on the model's learning and degrade the performance.

Step 3. One combination of time lags is chosen, which serves as the input for the learning candidate. Different from the existing works, various combinations of time lags are considered in our model. For example, a forecasting model with time lag three is usually trained by feeding data at  $t-1$ ,  $t-2$ , and  $t-3$ , but our model can utilize combinations of the lags, such as  $t-1$  and  $t-3$ .

Step 4. Then, the SVR models with different hyper-parameters using the corresponding combinations of time lags are trained based on the chosen shrinking window size.

Step 5. After training, the model is applied to the validation set, and the corresponding error metric is calculated.

Step 6. All errors on the validation sets are compared and the model with the minimum error is the final SVR.

Step 7. The final SVR is re-trained with the selected SVR's hyper-parameters, shrinking window size, a combination of explanatory variables, and time lags.

Step 8. Finally, the retrained SVR is applied to the test set.

The SVRIMSE effectively captures both temporal and cross-variable dependencies through its integrated components, particularly the Intelligent Model Search Engine. This component is pivotal in identifying critical variables and the appropriate time lags that optimize predictive performance, assessed rigorously on a validation dataset. In the process of model development, the Intelligent Model

Search Engine first conducts a systematic search across potential predictors and their corresponding time lags. This search is not arbitrary; it is finely tuned to discern which variables and their historical values (time lags) most significantly influence the target variable, thereby elucidating cross-variable dependencies. The identified variables are those that consistently show a predictive relationship with the outcome across various scenarios within the validation set. Simultaneously, by determining the optimal time lags for these variables, the Intelligent Model Search Engine captures the temporal dependencies crucial for time-series forecasting. These time lags reflect how past values of one or more variables influence future values of the target variable, which is essential for understanding the dynamics within the data. Following this variable and lag selection phase, the SVR component of the model is trained. SVR is particularly well-suited for this task as it excels in modelling non-linear relationships. By training the SVR on the variables and time lags selected by the Intelligent Model Search Engine, the model not only learns the direct relationships but also the intricate non-linear interactions among the variables over time. This comprehensive approach ensures that SVRIMSE not only captures, but also leverages these dependencies to enhance forecasting accuracy.

### Ship size

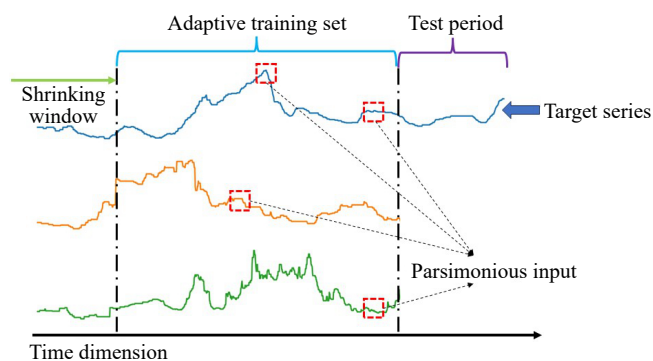
There are three types of tanker vessels in the data collected in this study. Table 1 summarizes the key features. Aframax ships refer to medium-sized crude oil tankers with a deadweight tonnage ranging between 80,000–120,000 metric tons. Tankers of this size have a cargo-carrying capacity between 70,000–100,000 metric tons, with an average cargo-carrying capacity of approximately 750,000 barrels. Due to their advantageous size, Aframax tankers are ideal for short to medium-haul oil trades and are primarily used in areas that do not have very large ports to accommodate bigger crude oil tankers or very large crude oil tankers. Vessels falling within this range are also referred to as the 'workhorses' of the world tanker fleet, as they carry a large number of oil products from many producing regions and can serve most of the ports in the world.

Suezmax ships are the largest marine vessels that can meet the restrictions of Suez, and transit the Suez Canal in a laden condition. Suezmax ships have a deadweight tonnage ranging between 120,000–200,000 metric tons. Ships from this size category are larger than the Aframax ships, and they can carry about 800,000 to more than 1,000,000 barrels of crude oil. After the expansion of the Suez Canal from 18 to 20.1 m in 2009, it is possible for a Suezmax ship up to 200,000 deadweight tonnage to pass through it.

Very Large Crude Carriers (VLCC) have a size ranging between 180,000–320,000 deadweight tonnage and they are capable of passing through the Suez Canal in Egypt. As a result, VLCC vessels are largely deployed around the North Sea, Mediterranean, and West Africa. Compared to the other smaller crude oil tankers, VLCCs are the larger vessels that provide better economies of scale for crude oil shipment. Specifically, VLCC ships are capable of carrying between 1.9 million and 2.2 million barrels of crude oil, and offer good flexibility for operating in ports with some depth limitations.

**Table 1.** Three types of tankers analyzed in this study.

Ship type	Deadweight tonnage (DWT) (MT)	Key features
Aframax	80,000–120,000	Short to medium-haul oil trades in regions with port limitations
Suezmax	120,000–200,000	Largest ships able to transit the Suez Canal fully laden
VLCC	180,000–320,000	Larger vessels with better economies of scale for long-haul routes



**Fig. 1** Partitioning scheme of SVRIMSE.

## Data description and pre-processing

The data investigated in this study come from various databases including Bloomberg Inc., Lloyd's List Maritime Intelligence, Tradewinds, and Clarkson Shipping Intelligence. A suitable and correct data pre-processing approach helps the machine learning model generate accurate outputs. First, the first differences of all time series are calculated to remove the linear trend. Computing the first difference of time series is a common approach in the forecasting literature. After obtaining all the first differences, the first differences are normalized for the implementation of machine learning models. The max-min normalization is implemented in this study. We assume that the maximum and minimum of the training set are  $x_{\max}$  and  $x_{\min}$ , respectively. The data are scaled into the range [0,1] using the following Eq (8):

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

where,  $x_{\text{normalized}}$  and  $x$  represent the normalized and original time series, respectively.

## Results and discussion

### Hyper-parameter optimization

Each model has several hyper-parameters that define the specific structure and affect the performance dramatically. To determine the suitable hyper-parameters for each model, the last-block cross-validation is adopted<sup>[39]</sup>. The time series are split into training, validation, and test sets. The data from 2016 and 2017 are validation and test set, respectively. The hyper-parameters that achieve the minimum error on the validation set are selected as the winner.

### Comparative analysis of all models

The analysis of the comparative results consists of two perspectives. First, the accuracy demonstrates the success of the proposed model in forecasting time charter rates. Second, the selected explanatory variables indicate their significant contribution to accurate forecasts of time charter rates. Three forecasting error metrics are utilized to evaluate the accuracy of these models. The first error metric is the classical Root Mean Square Error (RMSE) whose definition is:

$$RMSE = \sqrt{\frac{1}{L} \sum_{j=1}^L (\hat{x}_j - x_j)^2} \quad (9)$$

where,  $L$  is the size of the test set,  $x_j$  and  $\hat{x}_j$  are the raw data and predictions. Another error metric implemented in the paper is the

Mean Absolute Scaled Error (MASE)<sup>[40]</sup>. The definition of MASE is given in Eq (10):

$$MASE = \text{mean} \left( \frac{|\hat{x}_j - x_j|}{\frac{1}{T-1} \sum_{t=2}^T |x_t - x_{t-1}|} \right) \quad (10)$$

where,  $T$  represents the size of the training set. The denominator of MASE is the mean absolute error of the in-sample naive forecast. The third error metric is the Mean Absolute Percentage Error (MAPE) whose definition is described by Eq (11):

$$MAPE = \frac{1}{n} \sum_{j=1}^L \left| \frac{\hat{x}_j - x_j}{x_j} \right| \quad (11)$$

The comparative results of these three time charter rates in terms of RMSE, MAPE, and MASE are presented in Tables 2–4. According to the comparative results, we can claim that the proposed SVRIMSE outperforms the other models in terms of RMSE, MAPE, and MASE. The SVRIMSE achieves better performance than the SVR, which emphasizes the significance of training set size optimization and feature selection. The RIMSE also achieves great success in time charter rate forecasting because its structure is determined under the same framework as SVRIMSE. However, the learning candidate of RIMSE is linear, it cannot learn the complex non-linear patterns as SVRIMSE does. The models with the input of all variables cannot outperform the Naïve forecast, because the redundant variables hinder the learning of valuable relationships. The ridge regression performs much worse and its MASE is the lowest on all datasets because of its linear structure and lack of input optimization. As a result, we claim that the SVRIMSE can learn the knowledge of the freight market efficiently. The comparison of forecasts on the test set is visualized in Figs 2–4. All models predict the trend of Aframax time charter rates accurately according to Fig. 2. According to Fig. 3, there is a huge deviation from ridge regression's forecasts and raw data. The SVRIMSE forecasts the trend and values of these three-time charter rates accurately.

The selected variables for forecasting time charter rates are summarized in Table 5. The selection of the corresponding variable indicates their significant contribution to forecasting time charter rates. The Fleet Development (Number) and Scrap Value are selected for these three-time charter rates, which demonstrates their contribution to forecasting time charter rates most. Newbuilding prices are used for Aframax and Suezmax, which indicates the newbuilding prices contribute less than the Fleet Development (Number) and Scrap Value. In addition, the Fleet Development

**Table 2.** Comparative results of Aframax time charter rates.

	Naïve	ARIMA	SVR	MLP	DT	Ridge	RIMSE	SVRIMSE
RMSE	433.321	491.109	425.050	470.196	536.133	534.771	276.776	223.468
MAPE	0.023	0.026	0.023	0.024	0.028	0.029	0.015	0.010
MASE	0.363	0.424	0.373	0.383	0.460	0.460	0.238	0.166

**Table 3.** Comparative results of Suezmax time charter rates.

	Naïve	ARIMA	SVR	MLP	DT	Ridge	RIMSE	SVRIMSE
RMSE	758.099	728.449	1050.105	1273.288	742.331	1825.631	514.996	228.246
MAPE	0.022	0.029	0.042	0.057	0.037	0.078	0.014	0.010
MASE	0.266	0.345	0.489	0.665	0.431	0.915	0.175	0.122

**Table 4.** Comparative results of VLCC time charter rates.

	Naïve	ARIMA	SVR	MLP	DT	Ridge	RIMSE	SVRIMSE
RMSE	1129.470	1284.526	977.207	1302.169	1744.310	1764.768	886.920	601.970
MAPE	0.033	0.039	0.033	0.044	0.046	0.057	0.026	0.018
MASE	0.326	0.387	0.319	0.430	0.443	0.558	0.257	0.180

(Million DWT), Orderbook (DWT), and Orderbook (Number) are used once because of their general influence on the time charter rates.

According to comparative results, the SVRIMSE outperforms the other models, which demonstrates its success in forecasting time charter rates. The SVRIMSE achieves better performance than RIMSE because the SVR can mine non-linear and complex relationships among various variables. The primary factor that contributes to the superior performance of SVRIMSE over RIMSE is its ability to capture and leverage non-linear information inherent in complex datasets. Specifically, RIMSE is constrained by its reliance on the Ridge regression model, which, although robust in handling multicollinearity, is fundamentally limited to linear interactions. This linear approach restricts RIMSE from effectively modelling the intricate and dynamic non-linear patterns that are prevalent in volatile markets such as shipping. In contrast, SVRIMSE incorporates SVR, which is inherently designed to handle non-linear relationships. SVR achieves this through the use of kernel functions, which transform the data into a higher-dimensional space where non-linear relationships can be linearized and thus more effectively modelled. This capability allows SVRIMSE to adapt to the complex and non-linear dynamics that

Table 5. Selected variables for SVR.

	Aframax	Suezmax	VLCC
Time charter rate	*	*	*
Newbuilding prices	*	*	
Secondhand prices			
Fleet development (Number)	*	*	*
Fleet development (Million DWT)		*	
Scrap value	*	*	*
Orderbook (Number)			*
Orderbook (CGT)			
Orderbook (DWT)			*
Fleet growth			

characterize the shipping market's multivariate time series data. Moreover, the shipping market's volatility introduces rapid changes in trends and anomalies, which are better captured by the non-linear modelling approach of SVR. The SVR's framework, equipped with features like regularization and kernel-based learning, is more adept at managing unpredictability and ensuring robustness against overfitting compared to the linear methods used in RIMSE. The combination of SVR's advanced non-linear modelling capabilities and the iterative feature selection process of RIMSE within SVRIMSE further enhances its performance. By selectively incorporating significant variables and their respective time lags that best capture the temporal and cross-variable dependencies, SVRIMSE not only maintains model simplicity and computational efficiency, but also significantly improves forecasting accuracy. The other machine learning models fail to forecast the time charter rates accurately, because of the redundant explanatory variables. Although machine learning algorithms are good at extracting complex patterns based on more variables, the redundant variables convey much noise and degrade the performance. The success of the SVRIMSE demonstrates the importance of the objective design of the model and the information within the explanatory variables is non-linear.

Conclusions

This paper proposes a new framework for establishing the SVRIMSE to forecast the time charter rates of three categories of tankers, the Aframax, Suezmax, and VLCC. The forecasting process optimizes the SVR's hyper-parameter, training set size, memory size, and input variables. A comprehensive comparison of the SVRIMSE and other forecasting models is conducted, and three forecasting error metrics are computed to evaluate the performance. In addition, the SVRIMSE outperforms the RIMSE, which confirms that there are complex and non-linear relationships among these explanatory variables and time charter rates. The SVRIMSE selects more variables compared with the RIMSE because the linear structure of RIMSE cannot extract complex information from those variables. This paper also reveals the explanatory variables which contribute to the accurate forecasts of time charter rates significantly. The Fleet Development (Number) and Scrap Value are the most significant explanatory variables because they are selected to forecast time charter rates of all ship sizes. The Fleet Development (Number) indicates the current number of ships in operation all over the world and influences the time charter rates significantly based on domain knowledge. Scrap values also have the potential to affect charter rates by indicating the possibility of the deconstruction of ships. When the scrap values are high, the shipowners may prefer to dismantle the older ships instead of managing them.

The proposed forecasting framework provides practical implications for decision-making in the maritime industry. From the industry players' perspective (e.g., shipowners, charterers, and operators),

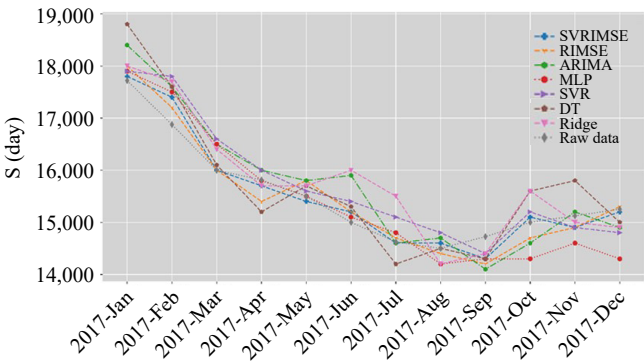


Fig. 2 Comparison of forecasts on Aframax time charter rates.

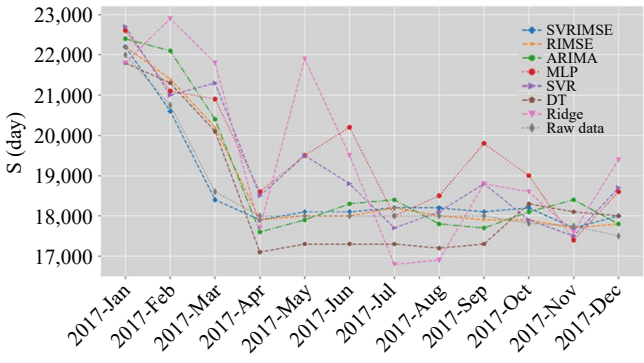


Fig. 3 Comparison of forecasts on Suezmax time charter rates.

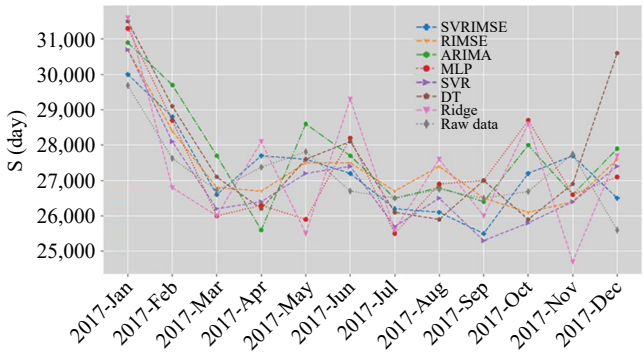


Fig. 4 Comparison of forecasts on VLCC time charter rates.

the proposed framework offers a reliable tool for forecasting time charter rates, enabling better planning, and optimization of fleet deployment. By identifying key explanatory variables that influence charter rates, stakeholders can make informed decisions regarding vessel allocation, contract negotiations, and risk management. For example, shipowners can use the model to determine optimal charter durations or adjust their strategies in response to market fluctuations, while charterers can leverage the forecasts to secure cost-effective contracts. From the related investors' perspective, the model provides valuable insights into the dynamics of the freight market, aiding in the assessment of investment opportunities and risks. Investors can use the forecasts to evaluate the profitability of investing in specific vessel types or market segments, as well as to anticipate market trends and adjust their portfolios accordingly. Additionally, the model's ability to objectively identify significant variables can help investors understand the underlying drivers of charter rate volatility, enhancing their ability to make data-driven investment decisions. By bridging the gap between theoretical forecasting and practical decision-making, our framework not only advances the academic understanding of freight market dynamics, but also provides actionable insights for industry stakeholders and investors, ultimately contributing to more efficient and informed operations in the maritime industry.

The limitation of this study includes the application of ship types and data availability. The proposed framework was tested exclusively on three types of tankers. Its applicability to other vessel types, for instance, bulk carriers, remains unverified. Additionally, the exclusion of variables such as regulations and fuel price volatility may restrict the model's ability to capture emerging market dynamics. If more data is available, the proposed forecasting model could be further verified regarding accuracy.

## Author contributions

The authors confirm their contributions to the paper as follows: study conception and design, data collection, analysis and interpretation of results, draft manuscript preparation: Song X, Chen ZS. Both authors reviewed the results and approved the final version of the manuscript.

## Data availability

The datasets analyzed during the current study are not publicly available due to confidential requirements.

## Acknowledgments

This research was supported by the Xi'an Jiaotong-Liverpool University (Grant No. RDF-24-01-026 XJTLU) Research Development Fund.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Dates

Received 30 January 2025; Revised 19 March 2025; Accepted 14 April 2025; Published online 28 September 2025

## References

1. Song X, Chen ZS. 2024. Shipping market time series forecasting via an Ensemble Deep Dual-Projection Echo State Network. *Computers and Electrical Engineering* 117:109218
2. Song X, Chen ZS. 2024. Enhancing financial time series forecasting in the shipping market: a hybrid approach with Light Gradient Boosting Machine. *Engineering Applications of Artificial Intelligence* 136:108942
3. Rothbarth E. 1939. Tanker freight rates and tankship building. *The Economic Journal* 49(196):760–62
4. Zannetos ZS. 1959. *The theory of oil tankship rates*. Doctoral dissertation. Massachusetts Institute of Technology, U.S. pp. 1–299
5. Eslami P, Jung K, Lee D, Tjolleng A. 2017. Predicting tanker freight rates using parsimonious variables and a hybrid artificial neural network with an adaptive genetic algorithm. *Maritime Economics & Logistics* 19(3):538–50
6. Goulielmos AM, Psifia ME. 2009. Forecasting weekly freight rates for one-year time charter 65 000 dwt bulk carrier, 1989–2008, using nonlinear methods. *Maritime Policy & Management* 36(5):411–36
7. Li J, Parsons MG. 1997. Forecasting tanker freight rate using neural networks. *Maritime Policy & Management* 24(1):9–30
8. Lyridis DV, Zacharioudakis P, Mitrou P, Mylonas A. 2004. Forecasting tanker market using artificial neural networks. *Maritime Economics & Logistics* 6(2):93–108
9. Santos AAP, Junkes LN, Pires FCM Jr. 2014. Forecasting period charter rates of VLCC tankers through neural networks: A comparison of alternative approaches. *Maritime Economics & Logistics* 16(1):72–91
10. Yang Z, Mehmed EE. 2019. Artificial neural networks in freight rate forecasting. *Maritime Economics & Logistics* 21(3):390–414
11. Zhang J, Zeng Q, Zhao X. 2014. Forecasting spot freight rates based on forward freight agreement and time charter contract. *Applied Economics* 46(29):3639–48
12. Hale C, Vanags A. 1989. Spot and period rates in the dry bulk market: Some tests for the period 1980–1986. *Journal of Transport Economics and Policy* 23(3):281–91
13. Glen D, Owen M, Van der Meer R. 1981. Spot and time charter rates for tankers, 1970–77. *Journal of Transport Economics and Policy* 15(1):45–58
14. Stranden SP. 1984. *Price determination in the time charter and second hand markets*. Working Paper MU 6:15. Norway: Center for Applied Research, Norwegian School of Economics and Business Administration
15. Beenstock M, Vergottis A. 1989. An econometric model of the world tanker market. *Journal of Transport Economics and Policy* 23:263–80
16. Beenstock M, Vergottis A. 1993. *Econometric modelling of world shipping*. First Edition, ed. Homa MS. Dordrecht, Netherlands: Springer Science & Business Media
17. Glen D, Martin B. 2005. A survey of the modelling of dry bulk and tanker markets. *Research in Transportation Economics* 12(1):19–64
18. Veenstra AW. 1999. *Quantitative analysis of shipping markets*. Thesis. Erasmus Universiteit Rotterdam, Netherlands. pp. 1–266
19. Kavussanos MG. 1996. Comparisons of volatility in the dry-cargo ship sector: Spot versus time charters, and smaller versus larger vessels. *Journal of Transport Economics and Policy* 30(1):67–82
20. Bulut E, Duru O, Yoshida S. 2012. 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 Economics & Logistics* 14(3):300–18
21. Chen ZS, Lam JSL, Xiao Z. 2023. Prediction of harbour vessel fuel consumption based on machine learning approach. *Ocean Engineering* 278:114483
22. Chen ZS, Lam JSL, Xiao Z. 2024. Prediction of harbour vessel emissions based on machine learning approach. *Transportation Research Part D: Transport and Environment* 131:104214
23. Duru O, Bulut E. 2014. A non-linear clustering method for fuzzy time series: Histogram damping partition under the optimized cluster paradox. *Applied Soft Computing* 24:742–48
24. Gao R, Du L, Duru O, Yuen KF. 2021. Time series forecasting based on echo state network and empirical wavelet transformation. *Applied Soft Computing* 102:107111
25. Gao R, Du L, Yuen KF. 2020. Robust empirical wavelet fuzzy cognitive map for time series forecasting. *Engineering Applications of Artificial Intelligence* 96:103978

26. Qiu X, Ren Y, Suganthan PN, Amaratunga GAJ. 2017. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing* 54:246–55
27. Syriopoulos T, Tsatsaronis M, Karamanos I. 2021. Support vector machine algorithms: an application to ship price forecasting. *Computational Economics* 57(1):55–87
28. Zeng Q, Qu C, Ng AKY, Zhao X. 2016. A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks. *Maritime Economics & Logistics* 18(2):192–210
29. Duru O. 2010. A fuzzy integrated logical forecasting model for dry bulk shipping index forecasting: an improved fuzzy time series approach. *Expert Systems with Applications* 37(7):5372–80
30. Bao J, Pan L, Xie Y. 2016. A new BDI forecasting model based on support vector machine. 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, Chongqing, China, 20–22 May 2016. USA: IEEE. pp. 65–69 doi: [10.1109/ITNEC.2016.7560320](https://doi.org/10.1109/ITNEC.2016.7560320)
31. Han Q, Yan B, Ning G, Yu B. 2014. Forecasting dry bulk freight index with improved SVM. *Mathematical Problems in Engineering* 2014:460684
32. Chou CC, Lin KS. 2019. A fuzzy neural network combined with technical indicators and its application to Baltic Dry Index forecasting. *Journal of Marine Engineering & Technology* 18(2):82–91
33. Kamal IM, Bae H, Sunghyun S, Yun H. 2020. DERN: Deep ensemble learning model for short-and long-term prediction of Baltic dry index. *Applied Sciences* 10(4):1504
34. Şahin B, Gürgeç S, Ünver B, Altın I. 2018. Forecasting the Baltic Dry Index by using an artificial neural network approach. *Turkish Journal of Electrical Engineering & Computer Sciences* 26(3):1673–84
35. Wong HL. 2014. BDI forecasting based on fuzzy set theory, grey system and ARIMA. *Proc. XXVII International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Kaohsiung, Taiwan, China, 2014*. 27. Springer International Publishing. pp. 140–49 doi: [10.1007/978-3-319-07455-9](https://doi.org/10.1007/978-3-319-07455-9)
36. Zhang GP. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50:159–175
37. Liu C, Hu Z, Li Y, Liu S. 2017. Forecasting copper prices by decision tree learning. *Resources Policy* 52:427–34
38. Bekiroglu K, Duru O, Gulay E, Su R, Lagoa C. 2018. Predictive analytics of crude oil prices by utilizing the intelligent model search engine. *Applied Energy* 228:2387–97
39. Bergmeir C, Benítez JM. 2012. On the use of cross-validation for time series predictor evaluation. *Information Sciences* 191:192–213
40. Hyndman RJ, Koehler AB. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22(4):679–88



Copyright: © 2025 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.