Highlights

Newbuilding ship price forecasting by parsimonious intelligent model search engine

Ruobin Gao, Jiahui Liu, Qin Zhou, Okan Duru, Kum Fai Yuen

- A novel intelligent model search engine algorithm is proposed for newbuilding ship prices.
- The proposed algorithm ensures the parsimony of the lag structure.
- The selected lags and explanatory variables are highly interpretable.

Newbuilding ship price forecasting by parsimonious intelligent model search engine

Ruobin Gao[a,*], Jiahui Liu[a], Qin Zhou[a], Okan Duru[b], Kum Fai Yuen[a]

[a]*School of Civil and Environmental Engineering, Nanyang Technological University*
[b]*Ocean Dynamex*

Abstract

Asset prices play a significant role in the financial survival and profitability of ship-owning firms. In a highly volatile shipping market, prices of newbuilding ships must be predicted to detect security shortfalls as well as opportunities for temporal arbitration (gaining on high-low pricing). Accordingly, this paper proposes an improved version of the intelligent model search engine (IMSE) by asynchronous time lag selection. The parsimonious IMSE algorithm comprises the essential components such as input and training data size selection by a grid search procedure. In the initial IMSE algorithm, time-lag (memory size) selection is designed such that a serial cluster of memory groups is assigned synchronously for all inputs. By relaxation of lag structures selection, the proposed algorithm estimates unique lead-lag relations for the input of the intended problem set. An extensive benchmark study with several baseline models and the persistence forecast (Naïve I) is performed to observe the out-of-sample accuracy of the proposed approach. The empirical

*Corresponding author

Email addresses: `gaor0009@e.ntu.edu.sg` (Ruobin Gao), `LIUJ0094@e.ntu.edu.sg` (Jiahui Liu), `zh0031in@e.ntu.edu.sg` (Qin Zhou), `okan.duru@oceandynamex.com` (Okan Duru), `kumfai.yuen@ntu.edu.sg` (Kum Fai Yuen)

results indicate that second-hand ship prices, scrap values, and orderbook (no. of orders) have predictive features and are selected by the search engine for two ship sizes. Different lag structures are estimated for each input with asynchronous time-lag selection improvement.
Keywords:
Forecasting, shipping market, grid search algorithm, machine learning.

1. Introduction

The significant impact of the shipping market on global economies is well acknowledged by economists Stopford (2008). The global shipping market is exceptionally volatile for various reasons, such as trade wars, economic recessions, severe market disequilibrium, and seasonality. As a result, such volatility poses challenges for shipping companies to survive in a dynamic environment. The market value of a shipping company largely depends on its assets and its dynamic decisions to make profits. Although the transactions of newbuilding ships can generate massive potential income, the volatile and dynamic prices raise business risks Alizadeh and Nomikos (2007). The volatile fluctuation of the newbuilding price is mainly affected by the supply and demand of seaborne trades; however, the global economic conditions and other factors also affect the newbuilding prices. Accurate forecasts of newbuilding prices serve as an essential reference for shipowners, investors, and shipyards to make sensible decisions. Moreover, investigating the factors that contribute to the forecasting accuracy most is prominent, although many factors indicate the shipping market situation. Therefore, improving the forecasting accuracy of newbuilding ship prices via machine learning algorithms that mine the knowledge and interactions among various variables is crucial.

The dynamics of ship prices have been studied by various studies for theoretical investigations Adland et al. (2006); Alizadeh and Nomikos (2007). On the other hand, most studies do not test the predictive value and accuracy of proposed theoretical findings. Such a theoretical perspective grounded on rational postulations about the economic predictors of ship prices is expected to shed light on the investing decisions by a predictive feature. This paper aims to extend and complement existing research by (1) introducing advanced prediction models to improve the accuracy, (2) employing variable selection techniques to identify the most significant factors affecting future newbuilding prices, (3) defining the suitable time window of historical data which contributes most to forecasting the future, and (4) considering parsimony in the selection of time lags (a.k.a. regularization).

Time series forecasting helps investigate the relationships of variables and temporal evolution by learning a function of past values and has been proven to succeed in various fields. However, conventional forecasting algorithms fail to capture the complex relationships among the high-dimensional input and the temporal patterns. Hence, intelligent algorithms are much superior and have demonstrated their success in capturing these complex relationships Ahmed et al. (2010); Arora et al. (2021); Bulut et al. (2012b); Bulut (2014); Cao and Tay (2003); Duru et al. (2021); Gao et al. (2020, 2021b,a); Liu et al. (2019); Qiu et al. (2017). In the context of the newbuilding market, ship price is characterized by a volatile trend, and there are many explanatory variables that may contribute to the forecasting accuracy, rendering the determination of a model's structure more difficult. In the review of previous studies, the modelling of newbuilding prices is largely driven by the expert's judgment to determine the explanatory variables and model's structure. However, inconsistency or lack of knowledge may impair the prediction performance. The subjective intervention of experts is also likely to introduce a high level of uncertainty and biases in determining inputs and establishing model's structure. As a result, there is an urgent need to implement an independent, automatic forecasting model which eliminates experts' intervention and subjective design of models, which can, in turn, improve the forecasting performance and enhance the understanding of newbuilding ship price.

The structure includes the memory size (time lag), the input variables, the training set size, and the model's hyper-parameters in a forecasting model. The model cannot learn long-term information with small memory size, but a large memory size may cause over-fitting problems. Similarly, redundant input variables may introduce noise and confuse the model. In addition, an extensive training set would allow obsolete data to be included for forecasting, which worsens accuracy when the time series are evolving with time. Therefore, establishing a forecasting model requires optimizing all of these parameters based on accuracy. To address the above-mentioned issues, the intelligent model search engine (IMSE) algorithm is proposed, and it can select the input variables and determines model's structure in a data-driven manner Bekiroglu et al. (2018); Duru et al. (2021). However, the IMSE algorithm only considers optimizing a dense collection of time lags, including noisy and redundant lags that deteriorate performance. As a result, we propose a parsimonious IMSE (PIMSE) which also considers the various combinations of time lags and preserves the parsimony of input according to the performance of out-of-sample data (i.e., the validation set). The parsimony helps the model focus on specific time lags instead of a long input vector which may input redundant information. Besides, such parsimony offers insights into the temporal characteristics of the newbuilding market by selecting a parsimonious collection of discontinuous time lags.

1.1. Innovation and impact

This article contributes to the literature from several perspectives.

1. Implementing a grid search to identify a parsimonious structure of the forecasting model. Such parsimony cannot only enhance the generalization ability, but also reveals the most significant time lags.

2. The grid search algorithm helps select the explanatory variables from a large set of them based on out-of-sample accuracy. The selection of corresponding variables indicates their prominent contribution to the future of newbuilding market.

3. A comparison with many popular machine learning models demonstrates the superiority of the proposed algorithm. The linear model with parsimonious input outperforms the complex forecasting model on the newbuilding ship prices in terms of three metrics.

The rest of this paper is organized as follows. In section 2, a review of the studies about ship prices and forecasting methods is presented. The proposed model and benchmark forecasting models are introduced in section

3. Then, the introduction of data, hyper-parameter optimization and simulation results are presented in section 4. Finally, conclusion is drawn, and limitations are described briefly in section 5.

2. Literature Review

Prices forecasting plays a significant role in the daily routine of players, brokers, commodity traders, and any decision-makers from the global shipping market. Anticipating the future is essential from the perspective of the academy and industry. As a result, the researchers from forecasting and shipping devote themselves to investigating the shipping market's mechanism and propose novel models to boost forecasting accuracy. This section reviews the related literature from two perspectives: first, the works about investigating the newbuilding market are summarized, and the second parts present the literature which utilizes intelligent algorithms for the predictive analytics of the shipping market.

Many researchers have dedicated themselves to understanding the newbuilding market and investigating explanatory variables that significantly impact newbuilding prices. The shipbuilding market depends on the freight volumes according to the conclusions drawn in Tinbergen (1931). In Hawdon (1978), the author estimates the tanker newbuilding prices using steel price and freight rate. The freight rates and steel prices affect the newbuilding market positively. However, the influence caused by fleet size is negative. The authors of Charemza and Gronicki (1981) show that freight and activity rates affect the ship prices via analyzing the supply and demand in the newbuilding and global shipping market. However, the supply and demand model is not practical in estimating newbuilding prices because of the longevity of a ship's life-cycle Beenstock (1985). Moreover, secondhand prices are found to have a strong influence on the newbuilding prices, and they evolve similarly along with time Beenstock and Vergottis (1993). An asset pricing model is developed to study the interaction between secondhand ship prices and the newbuilding ship prices Beenstock and Vergottis (1993). Oil prices and secondhand ship prices show their significant contribution to the newbuilding demand in Jin (1993). Although much work has been done to explain the factors which influence the newbuilding market, none is data-driven in an end-to-end fashion. In this circumstance, our alternative approach aims to establish the forecasting model and select the input variables according to the forecasting accuracy using a validation set. Through this approach, selected variables would highlight their distinguished contribution to predicting newbuilding prices over those that are not selected.

The modern time series forecasting model based on intelligent algorithms outperform the econometric methods in various studies Basu et al. (2018); Bekiroglu et al. (2018); Gao et al. (2020, 2021a); Cao and Tay (2003); Deb et al. (2017). The fundamental process of modern time series forecasting is to learn a function from the historical data based on the assumption that all valuable knowledge is included in the data. Inspired by the advanced development of modern time series forecasting, shipping scholars utilize the intelligent algorithm to various data from the global shipping market. Although little work has been done to forecast newbuilding ship prices via machine learning, a considerable number of studies have been conducted to examine the various aspects of the shipping market. For example, many researchers have applied intelligent algorithms to forecast the freight market Bulut et al. (2012a); Eslami et al. (2017); Goulielmos and Psifia (2009); Santos et al. (2014); Yang et al. (2019), sales and purchase market, and second-hand market Syriopoulos et al. (2021). The multi-layer perceptron (MLP) and radial basis function network are developed to forecast the period charter rates of VLCC tankers Santos et al. (2014). In Lyridis et al. (2004), the authors find that the explanatory variables can improve MLP's forecasting accuracy for VLCC spot freight rates. The authors of Yang and Mehmed (2019) show that the forward freight agreement information can boost ANN's performance of forecasting shipping freight rates. A combination of ANN and genetic algorithm is applied to forecast the tanker freight rates with parsimonious variables in Eslami et al. (2017). The support vector regression is recently implemented to forecast the newbuilding market and is compared with ARIMA Syriopoulos et al. (2021). The authors of Bekiroglu et al. (2018) are the pioneers who consider the influence of training sample size, and the same idea is applied to forecast the dry bulk market in Duru et al. (2021).

This paper argues that most of the above literature starts with certain prior assumptions about the explanatory variables, model structure, and memory size. The arbitrary relaxation may cause the deterioration of forecasting performance Bekiroglu et al. (2018); Duru et al. (2021). Although the IMSE is proposed to find the optimal model's structure and combinations of explanatory variables, the discontinuous collection of time lags is not investigated yet. Most forecasting literature feeds a dense collection of time lags into the model. However, this paper argues that a discontinuous and parsimonious collection of lags carries more important information and less noise than the dense collection. Taking monthly data as an example, the data at $t-1$ and $t-12$ carry the most important temporal information without the redundancy located in the other lags.

3. METHODOLOGY

This section first presents the proposed algorithm and then describes the benchmark forecasting models briefly.

3.1. Parsimonious IMSE

The power of data offers new lives for various industries based on anticipating the future accurately. Although the implementation of advanced algorithms is easy with software development, the practitioners still need to determine the model's structure and inputs. The birth of the IMSE algorithm solves such a problem tIt is not necessary to have many prior assumptions or theoretical knowledge of the explanatory variables from a specific field. The PIMSE is proposed to solve the complex forecasting problems from the perspective of forecasting without any user-related judgment.o some extent. However, its dense collection of time lags may cause an overfitting problem and degenerate the performance. To enhance the performance further, we propose the parsimonious IMSE (PIMSE) algorithm, which considers a parsimonious collection of time lags. The following innovations are born following the PIMSE algorithm:

1. It is unnecessary to have many prior assumptions or theoretical knowledge of the explanatory variables from a specific field. The PIMSE is

proposed to solve the complex forecasting problems from the perspective of forecasting without any user-related judgment.

2. PIMSE aims at searching for the optimal parsimonious combinations of explanatory variables and time lags and simultaneously defining the training sample size according to the out-of-sample accuracy.

3. A shrinking window is defined to fine-tune the starting point based on the out-of-sample accuracy. The shrinking window eliminates the ancient data and irregularities and retains the meaningful patterns.

4. The dense collection of time lags is likely to retain noise and degrade the accuracy. For example, the lag of $t-12$ usually reflects the annual seasonality of monthly data. It is widespread to input all values from $t-1$ to $t-12$ into the model, but the redundant time lags may deteriorate the accuracy. Instead, the PIMSE focuses on a parsimonious and discontinuous collection of time lags which helps the model focus on the significant lag and alleviate over-fitting.

The above innovations and contributions offer a unique identifier for the PIMSE in the literature about the tanker market and time series forecasting. For the completeness of this paper, the steps of the PIMSE are described below.

Given d historical explanatory variables to forecast newbuilding tanker prices, the aim is to learn a function f

$$y_{t+1} = f(\mathbf{y}_t, \mathbf{u}_t) = \beta_1^T \mathbf{y}_t + \sum_{i=1}^{d} \beta_{2,i} \mathbf{u}_{t,i} + \beta_0 + \epsilon_t, \tag{1}$$

where $y_{t+1} \in \mathbb{R}$ represents the future values which the model tries to forecast for time $t+1$, $y_t \in \mathbb{R}^n$ and $u_{t,i} \in \mathbb{R}^n$ are the vectors with a parsimonious collection of n time lags and ϵ is the additive noise. Finally, the model's coefficients can be learned by minimizing any classical loss function, such as mean squared error. Different from the IMSE in Bekiroglu et al. (2018); Duru et al. (2021) where the y_t and $u_{t,i}$ consist of the values from a dense collection of time lags, the PIMSE considers a discontinuous collection of time lags. In IMSE, when the memory size is fixed, all the values within the memory size are included as input. However, some redundant values may be included and worsen the forecasting accuracy. Taking monthly data as an example, a memory size twelve means all the values from the last year are utilized, but the most significant time lags maybe $t-1$ and $t-12$, which captures the most recent information and the annual pattern, respectively, and the others may not boost the accuracy. Therefore, the PIMSE is proposed to handle such limitations. Unlike the IMSE, the PIMSE considers different combinations of time lags instead of a dense collection of them. Before presenting the pseudo-codes of PIMSE, the compact version of our model is given

$$\mathbf{Y} = \mathbf{A}\mathcal{X} + \epsilon, \tag{2}$$

where lag_{max} represents the maximum time lag among the collection of time lags, $Y = [y_{lag_{max}+1}, y_{lag_{max}+2}, ..., y_{lag_{max}+N}]$ and $\epsilon = [\epsilon_{lag_{max}+1}, \epsilon_{lag_{max}+2}, ..., \epsilon_{lag_{max}+N}]$ for any real N. The ith row of $\mathbf{A} \in \mathbb{R}^{N-lag_{max}\times(dn+n+1)}$ is $[1, y^T_{lag_{max}+i-1}, u^T_{lag_{max}+i-1,1}, ...,$ $u^T_{lag_{max}+i-1,d}]$ for each $i = 1, 2, 3, ..., lag_{max}$. Finally, the coefficients vector $\mathcal{X} \in \mathbb{R}^{dn+n+1}$ is $[\beta_0, \beta_1, \beta_{2,1}, \beta_{2,2} ..., \beta_{2,d}]$.

The PIMSE aims at solving the following problem

$$\begin{aligned} \min \quad & \phi(\epsilon) \\ s.t. \quad & \epsilon = \mathbf{Y} - \mathbf{A}\mathcal{X} \quad , \end{aligned} \tag{3}$$

where the loss function ϕ can be specified according to the residuals ϵ. In Bekiroglu et al. (2018); Duru et al. (2021), the l_2 is chosen and the performance is outstanding. Therefore, we utilize the same loss function in this paper.

To select the best model according to the validation performance, the mean absolute scaled error (MASE) from the family of scaled errors is utilized as in Bekiroglu et al. (2018); Duru et al. (2021). The MASE is proposed in Hyndman and Koehler (2006) and the definition is

$$MASE = mean(\frac{\hat{x_j} - x_j}{\frac{1}{T-1}\sum_{t=2}^{T}|x_t - x_{t-1}|}), \tag{4}$$

where T represents the size of training set. The denominator of MASE is the mean absolute error of the in-sample naive forecast which offers a stable measure of the scale of the data. Next, we describe the main steps of PIMSE.

1. The first step refers to the data splitting step. The data for estimation and validation are split from the whole dataset according to the shrinking window size ξ_e and the validation set's size ξ_v.

2. Given the data for estimation and validation, the model's coefficients can be computed by solving the problem in 3 for all possible combinations of explanatory variables and time lags.

3. Apply each model from the previous step to forecast the validation set (ξ_v) and compute $MASE$ for them.

4. Select the model whose $MASE$ is the minimum as M_{final}.

5. Update the coefficients of M_{final} by including validation set ξ_v.

6. Apply the updated M_{final} to forecast the test set (ξ_t).

3.2. Benchmark forecasting models

In addition to the grid search model, a detailed comparison with benchmark forecasting models is conducted based on out-of-sample accuracy. These famous models are the persistence model, ARIMA, decision tree, MLP, SVR, Long short-term memory network (LSTM), and ridge regression.

(1) *Persistence model*: The persistent model is the most common benchmark for forecasting tasks. It uses the most recent value as its forecast, which is essentially a no-rule scheme, but a forecasting algorithm must outperform it. The actual value observed at $t-1$ is the forecast for time step t Makridakis et al. (2008).

(2) *ARIMA*: ARIMA succeeds in various forecasting tasks because it is an integration of differencing, autoregressive and moving average model. The differencing helps stabilize the time series. An ARIMA with order (p, d, q) is defined as

$$\psi_t = s + a_1\psi_{t-1} + ... + a_p\psi_{t-p} + b_1\epsilon_1 + ... + b_q\epsilon_q + \epsilon_t, \tag{5}$$

where ψ_t represents the time series after d times differencing, ϵ is the residual and $s, a_1, a_2, ..., a_p, b_1, b_2, ..., b_q$ are the coefficients.

Algorithm 1: Parsimonious IMSE

Input: ξ_e, ξ_v, ξ_t,

the target variable $\mathbf{Y} = [y_1, y_2, ..., y_N]$,

the i^{th} explanatory variable $\mathbf{u_i} = [u_{1,i}, u_{2,i}, ..., u_{N,i}]$

Output: $model_{final}$,

selected combination of time lags,

selected combination of input variables and

forecasts for ξ_t

Data: Target and explanatory variables

foreach ξ_e **do**

  $\mathbf{Y_e} = [y_{t-\xi_e}, y_{t-\xi_e+1}, ..., y_t]$ $\mathbf{u_{e,i}} = [u_{t-\xi_e,i}, u_{t-\xi_e+1,i}, ..., u_{t,i}]$

  \\\\ Estimation Part

  $\mathbf{Y_v} = [y_t, y_{t+1}, ..., y_{t+\xi_v}]$ $\mathbf{u_{v,i}} = [u_{t,i}, u_{t+1,i}, ..., u_{t+\xi_v,i}]$

  \\\\ Validation Part

  foreach c^{th} *combination of lags* **do**

    foreach k^{th} *combination of variables* **do**

      Solve Equation 3 and collect the result, $M_{\xi_e,c,k}$

    end

    Forecast the validation period (ξ_v)

    Compute the $MASE_{\xi_e,c,k}$ of each $M_{\xi_e,c,k}$

  end

end

Find the corresponding M_{final} whose $MASE_{\xi_e,c,k}$ is the minimum.

Update the coefficients by including ξ_v

Forecast the text period (ξ_t) using M_{final}, selected combination of time lags and input variables.

(3) *Decision tree*:Decision tree is a data-driven model with a tree structure. Each node in the tree partitions the input space. 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.

(4) *MLP*: MLP is a stacked neural network with a feed-forward structure. The hidden layer learns a meaningful representation of its input based on the task. MLP's success is due to its strong ability to extract the features in an end-to-end fashion. Unlike the hand-crafted or unsupervised features, the hidden nodes learn the features 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 backpropagation algorithm can be applied to optimize the weights of all layers. Besides, many improved versions of learning algorithms are proposed by existing research.

(5) *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. First, the SVR transforms the input variables γ into high-dimensional space by non-linear function $\varphi(\gamma)$. The the SVR tries to minimize the $\epsilon-$loss function

$$loss = C\sum_{i=1}^{n}(\xi_i + \xi_i^*) + \frac{1}{2}||\omega||^2 \tag{6}$$

with the following constraints

$$\psi_i - ((\omega * \varphi(\gamma_i))) \leq \epsilon + \xi_i \tag{7}$$

$$((\omega * \varphi(\gamma_i))) - \psi_i \leq \epsilon + \xi_i^* \tag{8}$$

where the slack variables, ξ_i and ξ_i^*, are positive and $i = 1, 2, 3, ..., N$, C is the regularization parameter, ϵ denotes the region where no penalty is counted. By introducing the Lagrange multipliers, this problem is transformed into:

$$f(\gamma, a_i, a_i^*) = \sum_{i=1}^{n} (a_i - a_i^*) K(\gamma, \gamma_i) + b, \tag{9}$$

where the kernel function $K(\gamma, \gamma_i) = \varphi(\gamma)\ \varphi(\gamma_i)$ The Lagrange multipliers a_i and a_i^* can be calculated via minimizing the following function

$$\begin{aligned} &\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i - a_i^*) \left(a_j - a_j^*\right) K(\gamma_i, \gamma_j) - \\ &\qquad \sum_{i=1}^{n} \phi_i (a_i - a_i^*) + \epsilon \sum_{i=1}^{N} (a_i + a_i^*) \end{aligned} \tag{10}$$

with the constraints

$$\sum_{i=1}^{N} (a_i - a_i^*) = 0, \quad 0 \leq a_i, a_i^* \leq C,\ i = 1, 2, \ldots,\ N. \tag{11}$$

(6) LSTM: The LSTM is the most popular variant of the family of recurrent neural networks (RNNs). When utilizing RNN in the context of forecasting, the historical data are processed one by one sequentially. The recurrent architecture offers the powerful ability to deal with sequential data because the valuable information about the history is memorized in the hidden state h. The functionalities of LSTM cell are summarized in the following equations,

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{12}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_i) \tag{13}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{14}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_i) \tag{15}$$

$$h_t = o_t tanh(c_t), \tag{16}$$

where σ represents the sigmoid non-linear transformation, and i, f, o and c are input gate, forget gate, output gate and the cell vectors Hochreiter and Schmidhuber (1997).

(7) Ridge regression: Ridge regression is the regularized version of the conventional linear regression. By imposing a l_2 norm regularization to the weights, the generalization ability of linear regression is enhanced significantly.

$$loss = ||Y - XW||^2 + \lambda||W||^2 \tag{17}$$

$$W_{ridge} = (X^T X + \lambda I)^{-1} x^T Y \tag{18}$$

where λ is the regularization parameter.

4. Empirical Results

4.1. Data and its nature

The data investigated come from various databases, including Bloomberg Inc., Lloyd's List Maritime Intelligence, Tradewinds and Clarkson Shipping Intelligence. First, we describe the ships briefly. Then, the statistics of the data are presented.

Aframax ships refer to medium-sized crude oil tankers with a deadweight tonnage ranging between 80,000-120,000 tons. The 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 extensive 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 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 18m to 20.1m in the year 2009, a Suezmax ship with up to 200,000 deadweight tonnage can 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 Suze 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, VLCC is the larger vessel that provides better economies of scale for crude shipment. Specifically, VLCC ships can carry between 1.9 million and 2.2 million barrels of crude oil and offer good flexibility for operating in ports with some depth limitations.

This paper focuses on PIMSE's ability to forecast newbuilding ship prices and the influence of other explanatory variables from January 1999 to December 2017. There are 228 data points in total. PIMSE's performance over other baseline forecasting models is also compared. The variables investigated are newbuilding prices (NP), secondhand prices (SP), scrap values (SV), orderbook (OB) expressed in number, CGT and DWT, fleet growth (FG) of three tankers, and labor interest rates (LIR). detailed statistics of all data are summarized in Table 1.

Table 1: Descriptive statistics of the datasets.

	Kurtosis	Max	Mean	Median	Min	Skewness	Std
VNP	-0.27	162.00	97.98	95.50	62.50	0.70	24.50
VSP	0.53	165.00	81.32	74.00	53.00	1.13	26.28
VSV	-0.66	28.39	13.87	14.88	3.82	-0.08	5.82
VOB_{Number}	-0.41	257.00	116.66	90.00	55.00	0.93	51.66
VOB_{CGT}	-0.40	11443181.30	5196075.29	3995538.41	2467449.96	0.94	2294931.15
VOB_{DWT}	-0.42	79755103.00	36083709.54	27669491.50	17337447.00	0.93	16099722.72
VFG	-0.06	9.68	3.10	3.39	-6.07	-0.52	3.13
SNP	-0.03	100.00	62.35	61.25	42.50	0.70	13.76
SSP	0.09	105.00	56.10	50.00	35.00	0.98	17.40
SSV	-0.22	18.80	8.49	9.01	2.76	0.28	3.55
SOB_{Number}	-0.98	174.00	88.01	78.00	34.00	0.52	36.70
SOB_{CGT}	-0.98	5311636.61	2692314.98	2391557.82	1025878.66	0.51	1117762.88
SOB_{DWt}	-0.96	27462502.00	13844923.34	12309854.00	5176953.00	0.51	5801141.73
SFG	-0.79	10.35	4.51	4.86	-4.43	-0.24	3.48
ANP	-0.34	82.50	51.12	51.00	33.00	0.44	11.71
ASP	-0.43	79.00	42.55	39.50	24.00	0.83	14.14
ASV	-0.45	12.35	5.92	6.38	1.96	0.06	2.33
AOB_{Number}	-0.06	306.00	146.28	139.50	31.00	0.68	69.81
AOB_{CGT}	-0.06	7825420.80	3754618.85	3613652.23	819496.10	0.69	1775662.27
AOB_{DWT}	-0.09	33731052.00	16119785.38	15635385.50	3281778.00	0.66	7737138.89
LIR	-0.77	0.07	0.02	0.01	0.00	0.73	0.02

For simplicity, we use V, S and A to represent three ship sizes, VLCC, Suezmax and Aframax, respectively. The abbreviation for specific variables consist of the abbreviation of ship size and variable's name. For example, VNP, VSP, VSV, VOB and VFG represent VLCC tanker's newbuilding prices, secondhand prices, scrap value, orderbook and fleet growth, respectively. LIR represents Libor interest rates.

4.2. Data pre-processing

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, they are normalized for the implementation of machine learning models. The max-min normalization is implemented in this paper. We assume that the maximum and minimum of the training set are x_{max} and x_{min}, respectively. The data are scaled into range [0,1] using the following equation:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{19}$$

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

4.3. Hyper-parameter optimization

The datasets are split into three sets, the training, validation, and test set, to adopt the last-block validation technique. The data of 2014 and 2015, 2016 and 2017 are validation and test set, respectively. The hyper-parameters of all the machine learning models are optimized by cross-validation Bergmeir and Benítez (2012). All variables are used as input for the baseline models, and the hyper-parameters which achieve the best performance on the validation set are optimal. Finally, all models are re-trained using the optimal hyper-parameters, including the validation set, and forecast the test set. The neural network models are simulated using the popular python library Pytorch. For the baseline models, the data from the previous year are fed into the model. For the neural networks, the hidden node varies from 2 to 12 with a step size of 2. For the MLP network, sigmoid activation is used. The MLP is trained with a learning rate of 0.0001 and 2000 epochs. For the decision tree model, the ratio of the samples per split and leaf varies from 0.1 to 1 with a step size of 0.1. The shrinking window size of PIMSE and IMSE varies from 0 to 36 with a step size of 4, which equals one season. There are eight variables in total, including one target variable and seven explanatory variables. The memory size of IMSE starts from one to twelve, which represents the previous year. It is impossible and super time-consuming to search for all the combinations of lags from $t-1$ to $t-12$, because the number of combinations of explanatory variables is also large. The lag $t-1$ is always included because it offers the most recent information, and IMSE definitely includes it. We relax such search space by only searching for the combinations whose number of lags is not more than three and from $t-1$ to $t-6$ and $t-12$. In total, 866,140 models are estimated and evaluated on the validation set.

4.4. Results

The comparison of these ten models has two perspectives. First, the results reflect the superiority of PIMSE's forecasting ability on newbuilding prices. Second, the forecasting accuracy demonstrates the influence of several variables on the newbuilding prices. Third, the parsimonious memory, training sample size and selection of features are presented. 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

$$RMSE = \sqrt{\frac{1}{L}\sum_{j=1}^{L}(\hat{x_j} - x_j)^2}, \tag{20}$$

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) Hyndman and Koehler (2006). The definition of MASE is

$$MASE = mean(\frac{\hat{x_j} - x_j}{\frac{1}{T-1}\sum_{t=2}^{T}|x_t - x_{t-1}|}), \tag{21}$$

where T represents the size of 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

$$MAPE = \frac{1}{L}\sum_{j=1}^{L}|\frac{\hat{x_j} - x_j}{x_j}|. \tag{22}$$

The performance on the test sets is summarized in Table 5. According to Table 5, we can find that the IMSE and PIMSE are the best two models. Their predictive performances on Suezmax newbuilding tanker prices are similar, but the PIMSE outperforms IMSE on the other two kinds of tankers. Especially, the PIMSE outperforms the other models significantly on Aframax newbuilding tanker prices. The machine learning models do not show their superiority necessarily. In addition, the IMSE and PIMSE outperform the ridge regression, which introduces l_2 regularization to overcome overfitting. Therefore, we claim that the regularization by grid search offers more accurate results compared with the classical regularization. The estimation results of PIMSE on VNB, SNB and ANB are shown in Tables 2, 3 and 4. The searching results of PIMSE and IMSE for NP forecasting are

Table 2: Estimation results of VNB forecasting.

	β	SE	t	P
β_0	-0.205	0.109	-1.875	0.062
VNB_{t-12}	0.048	0.065	0.736	0.463
VSP_{t-12}	-0.140	0.106	-1.324	0.187
VSV_{t-12}	0.253	0.107	2.362	0.019
VOB_{t-12}	-0.183	1.585	-0.116	0.908
$VOB_{DWT_{t-12}}$	0.190	1.608	0.118	0.906
VNB_{t-1}	0.432	0.065	6.672	0.000
VSP_{t-1}	0.382	0.104	3.686	0.000
VSV_{t-1}	0.116	0.103	1.121	0.264
VOB_{t-1}	-0.344	1.684	-0.204	0.838
$VOB_{DWT_{t-1}}$	0.371	1.713	0.217	0.829
Diagnostics				
AIC	-372.6		Log-Likelihood	197.280
BIC	-337.0		F-statistic	12.040
R-squared:	0.406			

Table 3: Estimation results of SNB forecasting.

	β	SE	t	P
β_0	0.238	0.079	3.019	0.003
SNB_{t-3}	0.170	0.076	2.224	0.027
SFG_{t-3}	0.020	0.073	0.277	0.782
SNB_{t-2}	-0.059	0.086	-0.682	0.496
SFG_{t-2}	0.089	0.073	1.222	0.223
SNB_{t-1}	0.506	0.076	6.628	0.000
SFG_{t-1}	-0.066	0.073	-0.912	0.363
Diagnostics				
AIC	-301.6		Log-Likelihood:	157.780
BIC	-279.3		F-statistic	12.590
R-squared:	0.305			

Table 4: Estimation results of ANB forecasting.

	β	SE	t	P
β_0	0.034	0.108	0.320	0.750
ANB_{t-12}	0.039	0.064	0.616	0.539
ASP_{t-12}	-0.174	0.087	-1.992	0.048
ASV_{t-12}	0.150	0.098	1.521	0.130
AOB_{t-12}	0.053	0.055	0.967	0.335
ANB_{t-1}	0.421	0.064	6.609	0.000
ASP_{t-1}	0.394	0.084	4.669	0.000
ASV_{t-1}	0.053	0.096	0.552	0.581
AOB_{t-1}	0.070	0.054	1.301	0.195
Diagnostics				
AIC	-355.5		F-statistic:	15.300
BIC	-326.6		Log-Likelihood:	186.760
R-squared:	0.413			

shown in Table 6. We can find that the variable selection shares some similarities. For VLCC tankers, SP, SV,OB, and OB_{CGT} are selected by both PIMSE and IMSE. For Aframax tankers, SV and OB are selected by both PIMSE and IMSE. The overlapping selection of the corresponding variables demonstrates their strong contribution in terms of forecasting. The LIR is not chosen, which indicates its less contribution to forecasting NB. The results in Table 6 further demonstrate the superiority of PIMSE is because of the parsimony of the input. According to Table 5, the PIMSE outperforms IMSE dramatically on Aframax NP, because there are only two time lags for PIMSE whereas IMSE has seven time lags as shown in Table 6. The PIMSE selects the two most significant time lags, $t-1$, which measures the short-term pattern, and $t-12$, which captures the annual seasonality. However, the IMSE selects a dense collection of time lags from $t-1$ to $t-7$ because it does not consider a parsimonious combination of time lags. According to Table 6, the same phenomenon happens for VLCC tankers. The PIMSE and IMSE select the same set of explanatory variables and training sample size, but PIMSE only utilizes $t-1$ and $t-12$ which offer more valuable information compared with $t-1$, $t-2$ and $t-3$ selected by IMSE. The time lags $t-1$, $t-2$ and $t-3$ only capture the short-term temporal patterns, but omits the annual seasonality carried by $t-12$. The combination of discontinuous lags of PIMSE helps capture fruitful features with fewer lags compared with the IMSE. The PIMSE can easily capture both the short-term and long-term temporal information with two lags, $t-1$ and $t-12$. However, if parsimony and discontinuity are not considered, it is necessary but redundant to input all data points from $t-1$ to $t-12$, which deteriorates the performance.

Table 5: Performance on test sets of three newbuilding tankers' prices.

Ship size	Metric	Naive	ARIMA	Ridge	DT	SVR	MLP	LSTM	IMSE	PIMSE
VLCC	RMSE	0.97895	0.76946	0.89899	1.04714	0.97891	0.83140	0.91487	0.77728	**0.70711**
	MAPE	0.00733	0.00615	0.00794	0.00841	0.00733	0.00721	0.00739	0.00587	**0.00583**
	MASE	0.48798	0.41203	0.52896	0.56003	0.48798	0.47947	0.49164	**0.39038**	**0.39038**
Aframax	RMSE	0.68084	0.60410	0.63540	0.73161	0.68082	0.70839	0.65201	0.70156	**0.54247**
	MAPE	0.00896	0.00921	0.01039	0.01086	0.00896	0.01042	0.00950	0.01168	**0.00657**
	MASE	0.56202	0.57263	0.64709	0.67854	0.56203	0.65881	0.59686	0.71657	**0.40746**
Suezmax	RMSE	0.72529	0.62670	0.62925	0.77605	0.71722	0.78938	0.77086	**0.59730**	**0.59730**
	MAPE	0.00880	0.00873	0.00841	0.01013	0.00882	0.01021	0.01017	**0.00830**	0.00832
	MASE	0.59012	0.57901	0.56290	0.67817	0.59103	0.68616	0.68012	**0.55323**	**0.55323**

Table 6: Hyper-parameter optimization and variable selection for NP forecasting.

Model	Ship size	Shrinking window	Time lags	NP	SP	SV	OB	OB_{CGT}	OB	FG	LIR
PIMSE	VLCC	4	$t-1, t-12$	*	*	*	*	*			
PIMSE	Suezmax	12	$t-1, t-2, t-3$	*						*	
PIMSE	Aframax	8	$t-1, t-12$	*	*	*	*				
Model	Ship size	Shrinking window	Memory size	NP	SP	SV	OB	OB_{CGT}	OB	FG	LIR
IMSE	VLCC	4	1	*	*	*	*	*			
IMSE	Suezmax	0	4	*	*	*					
IMSE	Aframax	8	7	*		*	*		*		

For the simplicity of visualization, only the comparison among raw data, IMSE and PIMSE's forecasts are visualized in Figures 1, 2 and 3. It is straightforward to find that both PIMSE and IMSE can accurately anticipate the characteristics of the three newbuilding prices. All these three prices decrease from the beginning of 2016 and remain relatively stable in the year 2017. The forecasts of PIMSE and IMSE both follow such decreasing trends and the following stable pattern.

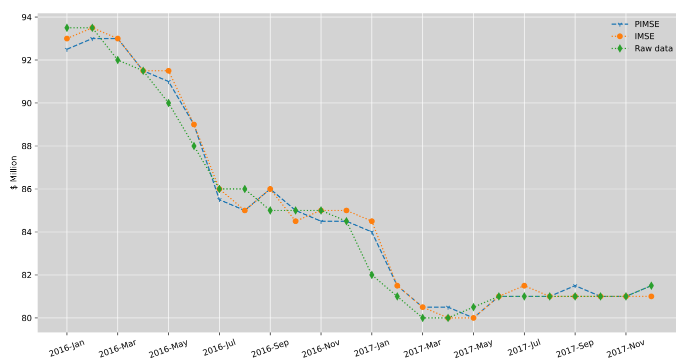

Figure 1: Comparison between raw data and the forecasts of VNP.

5. Conclusion

This paper proposes an improved PIMSE to forecast the newbuilding ship prices of three tanker types, the Aframax, Suezmax, and VLCC. The PIMSE accounts for the hyper-parameters of the learning candidate, memory length, input variables, and training set size. The novelty of our paper is that the algorithm's structure is determined objectively (i.e., data-driven) according to the performance of the validation set. In addition, we compare the PIMSE

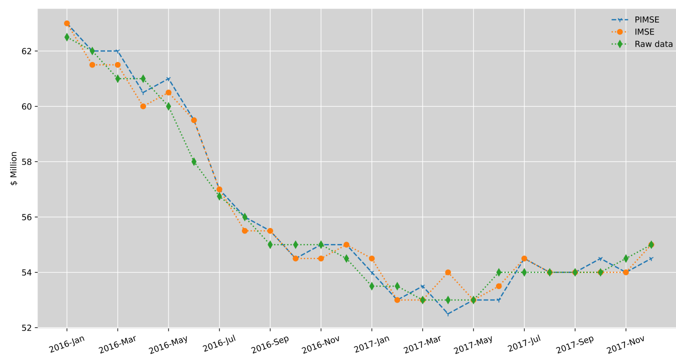

Figure 2: Comparison between raw data and the forecasts of SNP.

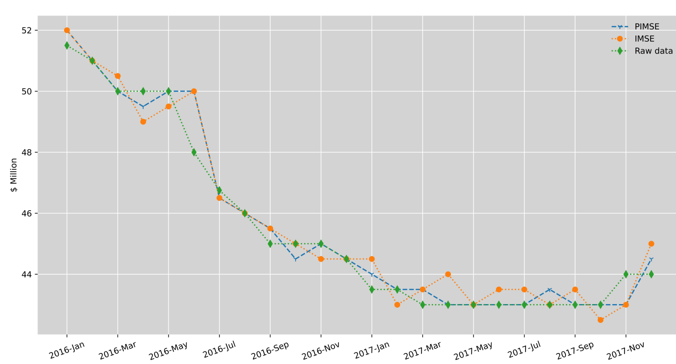

Figure 3: Comparison between raw data and the forecasts of ANP.

with several baseline models, including machine learning models, ARIMA, and the persistence model. The comparative results confirm the superiority of the PIMSE on this forecasting task. The PIMSE outperforms ridge regression which further highlights the importance of training set size and feature selection. This paper also contributes to the literature by providing insights into the interaction among various variables of the shipping market. The success of PIMSE also indicates that a linear model with objective optimization is sufficient to forecast new shipbuilding prices.

This study has several significant contributions to the literature about shipping markets. First, this paper proposes the PIMSE to achieve accurate forecasts of the newbuilding ship prices. The accurate forecasts of newbuilding ship prices are crucial for the shipping companies to survive in the global shipping market. Second, the PIMSE outperforms the ARIMA, famous machine learning models, and the persistence model based on the comparative results. Third, the algorithm generates accurate forecasts and selects significant dependent variables, which offers new academic and managerial insights. The survival of the corresponding variables from competing with other variables indicates their significance to the newbuilding ship prices. Fourth, the parsimonious collections of discontinuous time lag enhance the generalization ability.

Although the forecasting accuracy and interpretability of the results are high, the PIMSE cannot learn non-linear patterns from the data. However, the exhaustive optimization of non-linear machine learning models can cause too much burden on the computation because of the large number of hyperparameters. One potential solution is to utilize Bayesian optimization or evolutionary optimization, enabling exploration in a more extensive search space.

References

Adland, R., Jia, H., Strandenes, S., 2006. Asset bubbles in shipping? an analysis of recent history in the drybulk market. Maritime Economics & Logistics 8, 223–233.

Ahmed, N.K., Atiya, A.F., Gayar, N.E., El-Shishiny, H., 2010. An empirical comparison of machine learning models for time series forecasting. Econometric Reviews 29, 594–621.

Alizadeh, A.H., Nomikos, N.K., 2007. Investment timing and trading strategies in the sale and purchase market for ships. Transportation Research Part B: Methodological 41, 126–143.

Arora, P., Khosravi, A., Panigrahi, B., Suganthan, P., 2021. Remodelling state-space prediction with deep neural networks for probabilistic load forecasting. IEEE Transactions on Emerging Topics in Computational Intelligence .

Basu, S., Roy, S., DasBit, S., 2018. A post-disaster demand forecasting system using principal component regression analysis and case-based reasoning over smartphone-based dtn. IEEE Transactions on Engineering Management 66, 224–239.

Beenstock, M., 1985. A theory of ship prices. Maritime Policy and Management 12, 215–225.

Beenstock, M., Vergottis, A., 1993. Econometric modelling of world shipping. Springer Science & Business Media.

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, 2387–2397.

Bergmeir, C., Benítez, J.M., 2012. On the use of cross-validation for time series predictor evaluation. Information Sciences 191, 192–213.

Bulut, E., 2014. Modeling seasonality using the fuzzy integrated logical forecasting (filf) approach. Expert systems with applications 41, 1806–1812.

Bulut, E., Duru, O., Yoshida, S., 2012a. 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, 300–318.

Bulut, E., Duru, O., Yoshida, S., 2012b. A fuzzy time series forecasting model for multi-variate forecasting analysis with fuzzy c-means clustering. International Journal of Computer and Information Engineering 6, 390–396.

Cao, L.J., Tay, F.E.H., 2003. Support vector machine with adaptive parameters in financial time series forecasting. IEEE Transactions on neural networks 14, 1506–1518.

Charemza, W., Gronicki, M., 1981. An econometric model of world shipping and shipbuilding. Maritime Policy & Management 8, 21–30.

Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W., 2017. A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews 74, 902–924.

Duru, O., Gulay, E., Bekiroglu, K., 2021. Predictability of the physical shipping market by freight derivatives. IEEE Transactions on Engineering Management .

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, 538–550.

Gao, R., Du, L., Duru, O., Yuen, K.F., 2021a. Time series forecasting based on echo state network and empirical wavelet transformation. Applied Soft Computing 102, 107111.

Gao, R., Du, L., Yuen, K.F., 2020. Robust empirical wavelet fuzzy cognitive map for time series forecasting. Engineering Applications of Artificial Intelligence 96, 103978.

Gao, R., Duru, O., Yuen, K.F., 2021b. High-dimensional lag structure optimization of fuzzy time series. Expert Systems with Applications 173, 114698.

Goulielmos, A.M., Psifia, M.E., 2009. Forecasting weekly freight rates for one-year time charter 65 000 dwt bulk carrier, 1989–2008, using nonlinear methods. Maritime Policy & Management 36, 411–436.

Hawdon, D., 1978. Tanker freight rates in the short and long run. Applied Economics 10, 203–218.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation 9, 1735–1780.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International journal of forecasting 22, 679–688.

Jin, D., 1993. Supply and demand of new oil tankers. Maritime Policy and Management 20, 215–227.

Liu, S., Ji, H., Wang, M.C., 2019. Nonpooling convolutional neural network forecasting for seasonal time series with trends. IEEE transactions on neural networks and learning systems 31, 2879–2888.

Lyridis, D., Zacharioudakis, P., Mitrou, P., Mylonas, A., 2004. Forecasting tanker market using artificial neural networks. Maritime Economics & Logistics 6, 93–108.

Makridakis, S., Wheelwright, S.C., Hyndman, R.J., 2008. Forecasting methods and applications. John wiley & sons.

Qiu, X., Zhang, L., Suganthan, P.N., Amaratunga, G.A., 2017. Oblique random forest ensemble via least square estimation for time series forecasting. Information Sciences 420, 249–262.

Santos, A.A., Junkes, L.N., Pires Jr, F.C., 2014. Forecasting period charter rates of vlcc tankers through neural networks: A comparison of alternative approaches. Maritime Economics & Logistics 16, 72–91.

Stopford, M., 2008. Maritime economics 3e. Routledge.

Syriopoulos, T., Tsatsaronis, M., Karamanos, I., 2021. Support vector machine algorithms: An application to ship price forecasting. Computational Economics 57, 55–87.

Tinbergen, J., 1931. Ein schiffbauzyklus? Weltwirtschaftliches Archiv , 152–164.

Yang, D., Jiang, L., Notteboom, T., 2019. Innovative solutions for shipping market turmoil: The search for profitability, sustainability and resilience. Transport Policy 82, 75–76.

Yang, Z., Mehmed, E.E., 2019. Artificial neural networks in freight rate forecasting. Maritime Economics & Logistics 21, 390–414.