Modeling and solving used sailboat market strategy and pricing method

Junhao Zhang *, Zishu Zhang

Shandong University of Finance and Economics, Jinan, China

* Corresponding Author Email: 18677066480@163.com

Abstract. The value of a sailboat changes with age and market conditions, and the price of a sailboat is also affected by many factors such as manufacturer, variant, length, etc. Based on this, the purpose of this article modeling is to study the pricing of used sailboats so that brokers can sell them better. This paper explains the market situation and appropriate pricing of sailboats through the following methods. Firstly, data cleaning and processing are performed to construct an influence factor map for the value of second-hand sailboats, establish an influence factor matrix and network, and utilize a decision tree model to explain the pricing. Additionally, other data resources are incorporated to enhance the accuracy of the model, and statistical software is used for modeling and validation. A multiple regression model is established to convert geographic area variables into dummy variables, assess the impact of geographic regions on prices through multiple linear regression, and explain the practical and statistical significance. The regression model is then employed to predict prices in the Hong Kong market, with comparisons to actual market data to demonstrate practical and statistical significance. Lastly, a cluster analysis algorithm is implemented to classify second-hand sailboats and improve sales effectiveness. In summary, the team prepared a report through modeling and solving, which explained the market for sailing boats and reasonable quotations for second-hand sailboats.

Keywords: Used sailboats; Decision tree algorithm; regression models; Cluster analysis algorithm; Correlation Analysis.

1. Background

Sailing ships originated as a means of water transportation for ancient people who lived in the Haile River area. As the backbone of a vessel, the hull needs proper maintenance and upkeep. But the method of maintenance depends on the material from which the boat is made. When moving a ship ashore, be sure to concentrate on repairing even minor issues. Start with one or two layers of anti-fouling paint to prevent dirt underwater. All parts of the rigging should be kept in good condition so that the effectiveness and safety of the voyage can be guaranteed. A thorough inspection of masts, fixed rigging, movable rigging and sails is important to know if there are cracks or areas of wear and tear in the sections.

However, the financial crisis of 2008 dealt an almost devastating blow to the entire shipping and shipbuilding industry, and the BDI fell from a high of 11,793 in 2008 to around 2,000 points now. And the price of sailboats, whether new or used, is also up and down in this stormy market. The price of sailing ships is an important indicator reflecting the prosperity of the shipping market, and it is also affected by the shipping market and has strong volatility [1]. Figure 1 shows the price ratio and second-hand ship price index of newbuilding in sailing from 2000 to 2018.



Figure 1. Used ship price ratio and used ship price index

2. Literature Review

There are many pricing methods for second-hand sailing ships, and the main applications include the replacement cost method[2], the present value of earnings method, the market comparison method[3], and the risk method[4]. , among which the replacement cost method is divided into 2 types, restore replacement cost and update replacement cost [5]. Xiao Qijun (2013) proposed the premise of different methods and their respective advantages and disadvantages by comparing the application and comparison of replacement cost method, income method and market method in the price of sailing ships[6]. He believes that the replacement cost method and the present value of earnings method are only applicable to personalized vessels such as LNG, while the market comparison law is more applicable to the sailing market with active sailing transactions and large market changes, but when the market is poor, it is prone to inaccuracies and underestimates. Liu Hong and Wu Si (2013) established a multi-dimensional data model of second-hand ship price by using multi-dimensional association rule data mining technology, and compared the strength and weakness relationship of various factors, and proved its applicability in second-hand ship investment decision-making[7]. Yin Zhilin (2011) carried out the development research of web-based rapid cost estimation system for sailboats, which made up for the shortcomings of manual valuation, and explained the development principles, system structure and functions of the system The systematic development of valuation has deep reference significance [8].

In terms of second-hand ship price prediction, Zhong Ming (2007) calculated the price of second-hand ships by linear regression, eliminated the influence of variable autocorrelation and multilinearity through the correlation analysis of various influencing factors, and finally selected the newbuilding price, borrowing rate, one-year term charter rate, ship breaking price and new ship order volume as independent explanatory variables, and compared the calculation results with the actual transaction price. Good results were obtained [9]. Luo Fucai (2011) established 5 BP neural network models for the study of second-hand ships and end-of-life ships, but because there is no detailed classification of the types of sailboats, and each model explanatory variable does not exceed 2 It is difficult to accurately explain the prices of used and end-of-life ships[10]. Liu Yang and Gao Jin (2012) conducted a study on the value assessment of second-hand ships from the perspective of the influencing factors of second-hand ship financial leasing[11].

3. Notations

Table 1. Symbol and Description

Symbol	Description
v_i	Residual error
u_{jl}	Factor initial load
b_j^o	Unit feature vector
P_{ij}	Joint probability
$g(\bullet)$	The transfer function of the output layer
$\phi(x)$	Linear mapping
ξ_i^*	Slack variables
v_k^i	The velocity of the particle at the iteration step i k
w_k	Inertia weights
v_i	Residual error

Suppose that all the data given in the question are reasonable.

Suppose that sudden second-hand sailing price intervention does not exist.

It is assumed that there will be no explosive changes in future used sailing price predictions.

It is assumed that the encoding of the data in the model does not affect the reflection of the original information.

Suppose that the data published by different sailing factories and different regions have the same statistical principles.

Exclude small probability events in life (e.g. black swan events, abnormal situations)

4. Data Preprocessing

In this question, the focus is on mining and processing abnormal data, as well as handling missing values in the dataset. Data mining is used to explore and model statistical data to reveal hidden patterns and laws. The first step is to identify outliers, which can be done through methods like the 3-sigma rule, cluster analysis, and regression analysis.

In the case of this question, the selected index data was analyzed using Stata software, revealing that improper handling of the beam, draft, and drainage indicators could introduce significant errors. Outliers are typically tested using methods such as t-tests, Lajda criterion, Lagerbus criterion, and Dixon criterion. In this case, the 3-sigma criterion is used, which involves determining whether to exclude suspicious data based on their distance from the sample center. The choice of using 3 times or 2 times the standard deviation depends on the significance level, indicating the probability of test errors or the confidence level. A significance level of 0.01 corresponds to a test confidence level of 1-, while a significance level of 0.05 corresponds to a test confidence level of 2.

The dataset also contains a significant amount of missing data, which can have a substantial impact on China's second-hand sailing ship price index data. Handling missing values is therefore crucial. Methods for handling missing values include deleting tuples, data completion, or leaving them as is. If the missing values cannot be filled based on relevant fields, the K nearest neighbor method can be used. This involves determining the K samples closest to the sample with missing data, based on the Euclidean distance, and using their values to complete the missing data.

Overall, the analysis involves mining abnormal data using various methods, such as outlier detection, and handling missing values through approaches like data completion and the K nearest neighbor method.

5. Correlation Analysis

Correlation analysis describes the closeness of the relationship between two variables, mainly represented by the value of the correlation coefficient, when the absolute value of the correlation coefficient is closer to r 1, it means that the correlation between the two variables is more significant. The main indicators of bivariate coefficient measurement are chi-square measurement, Spearman correlation coefficient, Pearson correlation coefficient, etc., because the data is fixed distance data, the correlation test between the two is judged by Pearson correlation coefficient, and its formula is:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

The Pearson simple correlation coefficient test statistic is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

where t the statistic follows the distribution of $n-2$ degrees of freedom t.

5.1. Correlation between variables and coefficient method *Pearson*

General|r|>0.95, with significant correlation; |r|<0.3 The relationship is very weak and considered irrelevant. 0.5≤|r|≤0.8 is moderately correlated, and 0.3≤|r|≤0.5 is considered low.

Pearson Coefficient method: Calculation of data from a fixed distance variable. The formula is

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3}$$

(where is the correlation coefficient; , are r $\bar{x}$ $\bar{y}$ the means of variables x and y, respectively; , is the first observation of the variable, respectively x_i y_i x y i).

Using SPSS19.0, distance correlation analysis was performed to investigate the degree of proximity between similar or dissimilar measurements.

SPSS 19.0 was used to analyze the correlation between the two, and the indicators with large correlation (n 2≤≤n 10) were selected as the correlation indicators $f_k(x_i)$

$$x_i, x_j, x_m \text{.....}$$

The regression equation is established as follows:

$$\begin{cases} f_1 = f\left(x_i, x_j, x_m \text{.....}\right) \\ f_2 = f\left(x_k, x_p, x_z \text{.....}\right) \\ f_3 = f\left(x_i, x_p, x_l \text{.....}\right) \\ f_2 = f\left(x_i, x_j, x_m \text{.....}\right) \end{cases} \tag{4}$$

5.2. Model Solving

Table 2 shows descriptive statistics.

Table 2. Descriptive statistics

	average value	standard deviation	Number of cases
Decoration	2.14	.945	58
type	1.69	.467	58
color	2.69	1.810	54
weathering	1.59	.497	58

Table 3 shows the correlations.

Table 3. Correlation analysis

Decoration	Decoration	type	color	Surface weathering
type	1	.099	.194	.049
color	.099	1	.424**	.344**
Surface	.194	.424**	1	-.115
weathering	.049	.344**	-.115	1

**. At level 0.01 (double-tailed), the correlation is significant.

The above table shows that the surface weathering of glass artifacts has the highest correlation with glass type.

The correlation between the dimensions is shown in Figure 2.

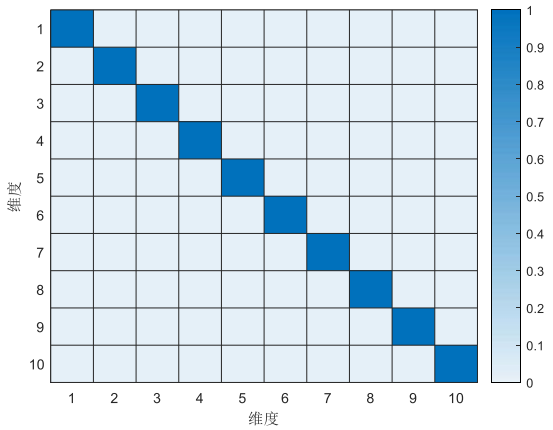

Figure 2. Correlation between dimensions

The following is a regression analysis, which is as follows:

Table 3. Model summary

model	R	R-square	Adjusted R side	Error in standard estimates	Durbin Watson
1	. 678[a]	.104	. 072	. 386	0.860

Party R in the table is 0 678, greater than 50%, indicating that the model prediction is accurate and the study of the model is meaningful.

Table 4. ANOVAa

	model	Sum of squares	degree of freedom	mean square	F	Salience
1	regression	2.414	3	.805	3.685	.018b
	Residuals	10.919	50	.218		
	total	13.333	53			

This table considers whether the regression equation makes sense, significance 0.005<0.05, so the equation makes sense.

Table 5. VIF diagnosis

model		Unstandardized coefficients		t	Salience	Collinearity statistics	
		B	standard error			Tolerance	VIF
1	(constant).	.930	.264		3.515	.001	
	Decoration	.040	.070	.074	.568	.573	.062
	type	.466	.149	.442	3.128	.003	.316
	color	-.088	.040	-.317	-2.215	.031	-.115

The VIF values in the table are not greater than 5, indicating that there is no multicollinearity between the independent variables, which reflects the accuracy and reliability of the regression model.

Based on all the above analysis, the regression equation for the independent and dependent variables is:

$$y = 0.04x1 + 0.466x2 - 0.088x3 + 0.930$$

The residual plot is shown in Figure 3.

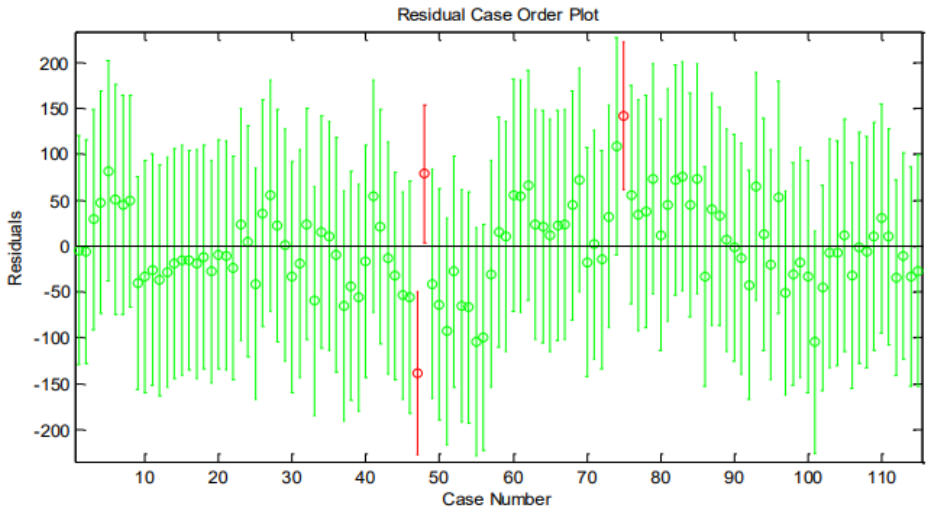

Figure 3. Residual diagram

6. Cluster analysis

Fuzzy C-means clustering algorithm is a clustering method that blurs the definition of classical division and uses the degree of membership to determine the degree of belonging to a certain cluster. Among them, there are two important parameters, namely the number of clusters c; Fuzzy weighted index. The algorithm divides vectors into groups $m\ n\ x_k \subset R^5\ c\,();k=1,2,\cdots,n$ s is the dimension of the vector and finds the clustering center for each group. x_k

The basic steps are:

(1) Meet the constraints of equation (5).

$$\sum_{i=1}^{c} u_{ik} = 1 u_{ik} \in (0,1) \tag{5}$$

(2) Calculate the cluster center V according to equation (6).

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m x_k}{\sum_{k=1}^{n} u_{ik}^m}, i = 1,2,\cdots,c \tag{6}$$

(3) Calculate the objective function according to equation (7).

$$J\left(U,v_1,\cdots,v_c\right)=\sum_{i=1}^{c}J_i=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}d_{ik}^{2} \tag{7}$$

$$u_{ik}=1/\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}},i=1,2,\cdots,c \tag{8}$$

$$k=1,2,\cdots,n$$

In equations (5) and (6), is the Euclidean distance between the first fuzzy group and the first cluster center; is the fuzzy weighted index, $d_{ik}=\sqrt{\sum_{q=1}^{s}\left(x_{kq}-v_{iq}\right)^2}$ $k\ i\ m\ \mathrm{m}\in(1,+\infty)$, 2≤c≤n. I

It is calculated from the final membership matrix $U=\left\{u_{ik}\right\}_{cxn}$

$$i=\arg\max_{1\le i\le c}u_{ik},k=1,2,\cdots,n \tag{9}$$

It can be seen that after learning, the sample is gathered into the class. $x_k\ i\ i=1,2,\cdots,c$

According to the above principle analysis, we will collect data samples and feature dimension collections. The characteristic dimensions are beam, draft, displacement, rigging, sail area and other information.

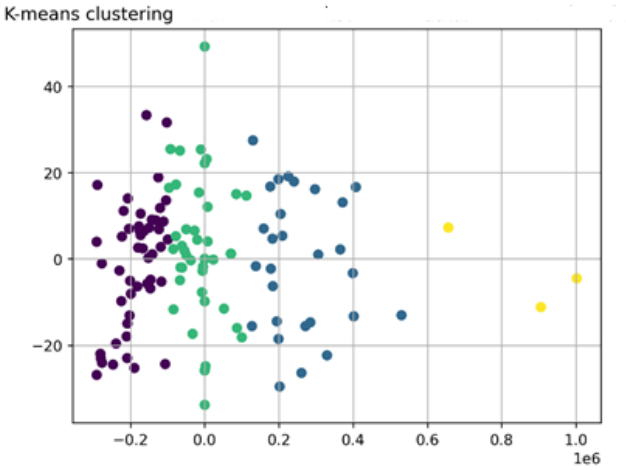

Figure 4. K-means clustering results

The results obtained by the training of the K-means clustering algorithm are shown in Figure 4, and it can be seen that the dataset is divided into four categories, and the classification effect is better.

7. Conclusion

Due to the strong complexity and volatility of the sailing market, the characteristics of huge amount, high risk and long payback period of sailing investment, enterprises investing in sailing ships have strong uncertainty. Assuming that the company has stable capacity demand, it currently uses chartering to meet its own capacity demand, and is bullish on the future charter market. From the perspective of controlling costs, companies began to consider whether to buy sailboats. Since newbuilding requires a certain cycle time, consider a second-hand vessel that is immediately available, such as this one that is currently chartered, or a similar boat in the market that is interested in selling. Then the timing of investment, the price and rental of second-hand ships have become important factors that enterprises need to consider when formulating investment plans. First of all, interest rate levels, investment costs, etc. have certain variables, so the timing of investment is very important.

Secondly, if an enterprise chooses to sublease to buy, it will bring two consequences: first, the price of second-hand ships will rise higher than the increase in expenses such as rent in the future, so buying second-hand ships at a low price has an appreciation effect; Second, in the future, the price of second-hand ships will fall more than rents and other expenses, so buying a sailboat at a high price will become a failed investment decision. Therefore, when making investment decisions, enterprises should fully compare the relationship between the cash flow of future expenditure and the price of second-hand ships, and choose the appropriate investment time from the perspective of long-term leasing and sublease to buy the expected expenditure, and make the most favorable investment decision.

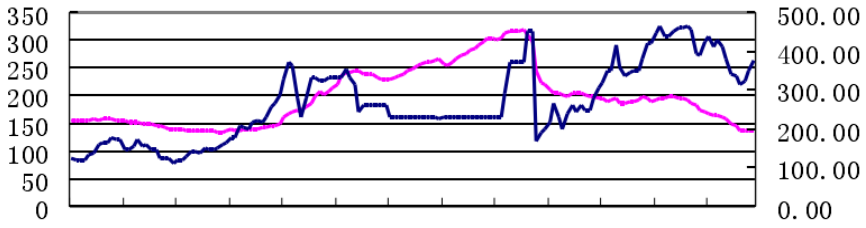

Figure 5. Used Sailing Price Index and Price Correlation Chart

The following advice is given for investors who sublease sailboats to buy:

First of all, fully consider the uncertainty of the sailing investment market and avoid investment risks. The sailing investment market is inextricably linked with the macro environment such as the world economy and trade, as well as the small environment such as the supply and demand relationship of sailing capacity, and the psychology of ship owners, which is complex and changeable. When considering the investment decision of subleasing to buy, enterprises should fully consider their risk tolerance and should not make blind decisions.

Second, seize the opportunity to invest. When enterprises choose to sublease as an investment opportunity, to look at the entire shipping market, the shipping market is cyclical, there are certain market laws, in different periods of the market then the size of the parameters are also different, investors should clearly understand their own market environment, based on effective and accurate market estimates, grasp the inflection point of the price of second-hand sailing ships, cautiously shoot, and obtain a larger investment report.

Finally, actively participate in the negotiation of second-hand ship transactions for sailboats. In the analysis of this article, one of the important parameters is the commission of the transaction, the commission, etc. Many other factors, such as discount rate, residual value, etc., are market-based factors and are difficult to be influenced by artificially. However, for investors, actively participating in the negotiation of second-hand ship transactions can grasp greater initiative, reduce other costs incurred by transactions, and be more conducive to investment decisions.

References

[1] Hussa, Kim Yong Quan, At SWISS. Research on marketing strategy of sailing into campus project [J]. Stationery & Technology, 2022(24): 184-186.

[2] Han Qi, Zhang Yunchen. The establishment and evaluation of the monopoly system of the Manila galleon trade [J]. International Journal of Sinology, 2021(04): 28-43+2+198-199. DOI: 10.19326/j.cnki.2095-9257.2021.04.004

[3] Ge Fengmin Research on sustainable development of sailing tourism in Xiamen [J]. Contemporary Sports Science and Technology, 2021, 11(18): 165-167. DOI: 10.16655/j.cnki.2095-2813.2010-1579-2152

[4] Huang Suzhen The current situation, problems and countermeasures of the development of sailing in China [J] Journal of Physical Education Adult Education, 2020, 36(04): 37-40. DOI: 10.16419/j.cnki.42-1684/g8.2020.04.008

[5] Yang Lihong, Zhu Zhengye. The wooden sailing industry developed amid challenges and opportunities——centered on the modern Huaihe River Basin [J] Journal of Huaibei Coal Normal University (Philosophy and Social Sciences Edition), 2009, 30(04): 40-43

[6] Wang Huiyong Discussion on ship asset valuation methods [J] Tianjin Navigation, 2015, 36(04): 41-43

[7] Chen Li, Wang Xuefeng, Zheng Shiyuan Risk measurement of dry bulk second-hand ship trading based on value-at-risk model [J] Journal of Shanghai Maritime University, 2012, 33(04): 57-63

[8] Gao Wei, Deng Zhaowen, Yu Yang Improvement and evaluation system design of used car replacement cost method [J]. Tractors and Agricultural Transport Vehicles, 2015, 42(3): 32-35

[9] Xiao Qijun. Research on ship price evaluation method based on BP neural network [D] Shanghai Jiao Tong University, 2013

[10] Liu Hong, Wu Si Application of multi-dimensional association rule data mining in the analysis of influencing factors of ship prices [J]. Journal of Shanghai Maritime University, 2013, 34(04): 31-37

[11] Yin Zhilin Research on the development of Web-based ship rapid cost estimation system [J]. Naval Engineering, 2011, 33(S1): 187-190.