ISSN:2790-1661 Volume-6-(2023)

Construction of A Used Sailboat Listing Price Model Based on Big Data

Shaoyi Niu [1], Weizhen Lin [2], Guanghua Lu [1], Li Guo [3], Tao Liu [2, 4, a], Ruofeng Qiu [5], Yunfei Qi [5]

[1] School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China;

[2] School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China;

[3] School of Resources and Materials, Northeastern University at Qinhuangdao, Qinhuangdao, 066004, China;

[4] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798, Singapore;

[5] Eighth Geological Brigade of Hebei Bureau of Geology and Mineral Resources Exploration, Qinhuangdao, 066000, China.

[a] liutao@neuq.edu.cn; tao.liu@ntu.edu.sg; math.taoliu@gmail.com

Abstract. With the continuous advancement of economic globalization, the second-hand sailboat market is also booming. However, many problems have arisen during the development process. To standardize the second-hand market, provide reasonable second-hand sailboat pricing, and stabilize the development of the second-hand sailboat market, this article constructs a relevant data model and provides several suggestions. To build the model, this article partially visualized the monohulled sailboat data and conducted chi-square analysis on factors such as region and sailboat age. It also drew a correlation matrix between various factors to obtain their correlation. Combined with data, factor analysis was used to reduce the dimensions of the considered factors. Finally, multiple regression was used to construct a model for the listing price of second-hand sailboats. The relevant data was obtained through a crawler program from SailboatData.Com.

Keywords: Monohulled Sailboats, Catamaran, Market, Sailboat Age, Pricing Strategy.

1. Introduction

With the economic development, the process of globalization continues to advance, the international environment is stable, and the global shipbuilding industry gradually recovers [1-5]. The number of new ship orders gradually decreases, making the supply and demand of the ship market more balanced.

With the booming development of the second-hand sailboat market, various problems also gradually emerge: the market transaction process is not transparent, the information asymmetry between buyers and sellers, the lack of channels for consumers to buy second-hand sailboats, the lack of unified evaluation standards, etc. These problems limit the development of the second-hand sailboat market [6-10]. With the popularization and development of the Internet, relevant second-hand sailboat information can be found in various regions, but at this stage various platforms and policies are not perfect, and there is no unified evaluation standard [11-14].

Due to the particularity of sailboat assets, the factors affecting the price of second-hand sailboats include not only the age of the sailboat, the equipment on board and other self-loss factors, but also related policies of different countries and regions, the freight rate of the sailboat market, and the scope of sailboat operation [15-18]. In fact, in addition to being consistent with the value evaluation of the sailboat itself, sailboat price evaluation is also constrained by the collectable data and information asymmetry. Therefore, it is of great significance to establish a comprehensive evaluation method for second-hand sailboats of different types.

2. General Assumptions and Notations

The following basic assumptions are made to simplify problems.

(1) The market supply and demand are balanced, the market information is transparent and sufficient, consumers fully understand the characteristics of used sailboats, and there are almost no transaction costs.

(2) All used sailboat price research is placed in the same market, where consumers can fully possess information and freely choose any used sailboat without constraints.

(3) There is no significant difference between the same variant produced by different manufacturers and the same variant produced in different years.

(4) Ignore the impact of factors such as the new shipbuilding market and the shipbreaking market on the used sailboat market [19].

Additional assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate locations.

Table 1. Symbol and description

Symbol	Description	Unit
p_1	Estimated listing price after standardization of monohull sailboats	
p_2	Estimated listing price after standardization of catamaran sailboats	
k_i	Coefficient	
l_i	Coefficient	
a	Age of sailboats after standardization	
E	GDP per capita after standardization	
f_1	Performance index scores for monohull sailboats	
f_2	Size index score for monohull sailing boats	
g_1	Catamaran width and performance index scores	
g_2	Catamaran length and draft index scores	
P_1	Monohull sailboat listing price	USD
P_2	Catamaran sailboat listing price	USD

3. Model building and solving

3.1 Data visualization for monohulled sailboats

Four variants with more comprehensive data were selected for analysis.

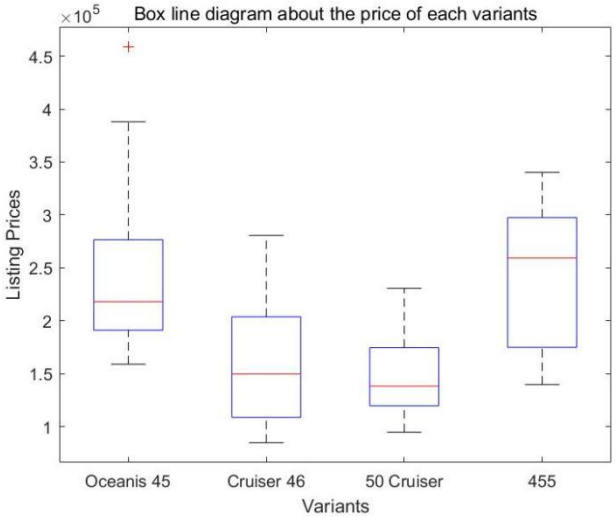

Fig. 1 Price box plot of four monohulled sailboat variants

From Figure 1, we can see that most of the data vary steadily within a certain range, but there are some outliers, which we remove by data preprocessing.

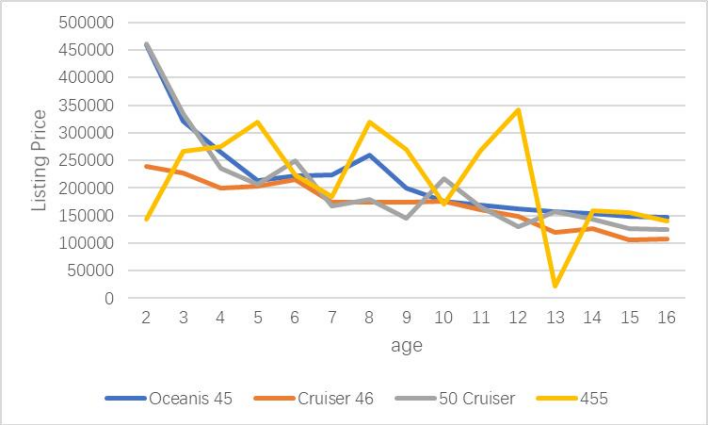

Fig. 2 Price Change Line Chart

The horizontal axis of the graph is the length of time between the production date of the monohulled sailboat and December 2020, which we will refer to here as the age of the sailboat, and the vertical axis is the listing price for the corresponding year, from which it can be seen that the overall price tag decreases as the age of the sailboat increases.

3.2 Data analysis of monohulled sailboats

Table 2. Inter-subject effect test1
Dependent variable: Listing Price (USD)

	Type III Sums of Squares	Degree of freedom	Mean Square	F	Significance
Modified model	907321683466.237a	6	151220280577.706	58.905	<0.001
intercept distance	6051966430376.205	1	6051966430376.205	2357.443	<0.001
Length	47687012542.407	1	47687012542.407	18.576	<0.001
Variants	27804247592.063	1	27804247592.063	10.831	0.001
Length * Variants	0.000	0			

Table 3. Inter-subject effect test 2
Dependent variable: Listing Price (USD)

	Type III Sums of Squares	Degree of freedom	Mean Square	F	Significance
Modified model	1239070856016.242	98	12643580163.431	7.088	<0.001
intercept distance	2920306170240.811	1	2920306170240.811	1637.109	<0.001
Country/Region/State	150743146038.772	30	5024771534.626	2.817	<0.001
Year	506660165475.589	14	36190011819.685	20.288	<0.001
Country/Region/State * Year	122238926440.920	54	2263683822.980	1.269	0.142

We selected the data related to 9 informative variants for analysis, Table 2 shows the ANOVA of listing price on Length and variants, and Table 3 shows the ANOVA of marker price on region and

age of sailboat, the significance of each data from the table is less than 0.001, and the significance of Country/Region/State × Year is less than 0.5. It can be seen that there is no interaction utility between the two, and the sum of class III squares of Length × Variants is approximately equal to 0. It can be seen that Length varies with Variants, and the degree of variation of Length in Variants is not significant.

3.3 Monohulled sailboat model construction

Based on the above analysis, we drew a tree diagram of the influencing factors, as the figure 3 shows:

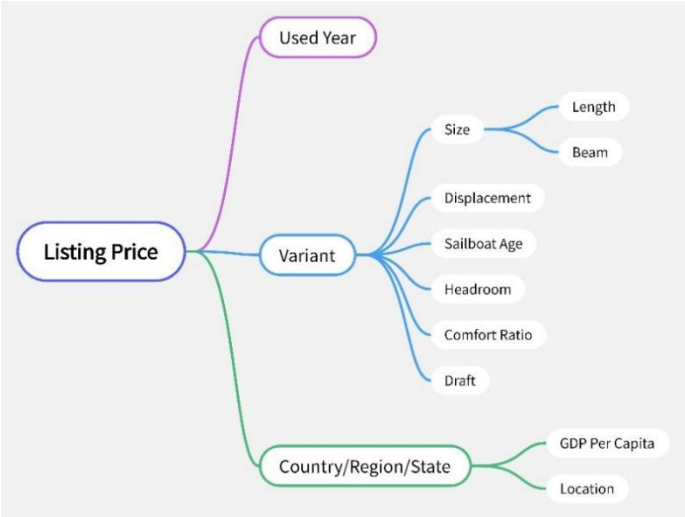

Fig. 3 Factors influencing the listing price

Under variant, we consider the effect of Size, Displacement, Sail Aera, Headroom, Sailboat Age, Comfort Ratio, Draft on the price of a sailboat, and use these indicators to represent the model of the sailboat. Under Country/Region/State we have considered the effect of GDP per capita and the geographical location of each region on the price. Since most of the given data have waterfront, we ignore the effect of waterfront factor on the results in the modeling process, and only use GDP to represent the nature of the region and the local economy.

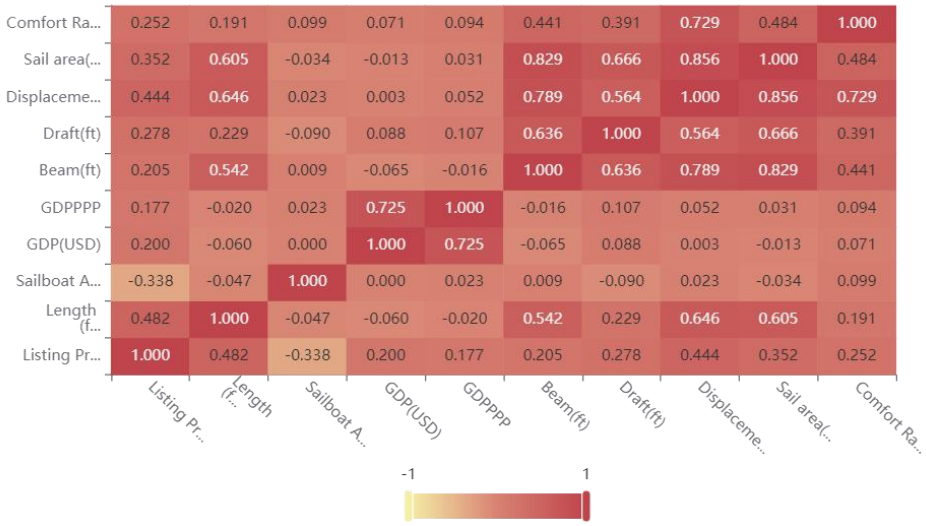

Fig. 4 Heat map of the correlation coefficients of the factors selected for the monohulled sailboat

Based on the data in SailboatData.Com, we plotted Figure 4. Figure 4 shows the heat map of the correlation coefficients of the factors selected for the monohull sailboats, and there is a certain correlation between the bid price and each factor, among which the correlation with Displacement, Length is higher and the correlation with Headroom and GDP per capita is lower. We used KMO test to find KMO=0.736, so we used factor analysis for dimensionality reduction and rotated to obtain Table 4 component matrix.

Table 4. Component matrix after rotation

	1	2
Length(ft)	0.032	0.946
Beam(ft)	0.606	0.657
Draft(ft)	0.743	0.255
Displacement(lb)	0.715	0.629
Sail area(ft^2)	0.629	0.693
Comfort Ratio	0.853	0.041

We obtained two factors with a cumulative variance contribution of 85.926%, and set the two factors as f_1, f_2, which we interpreted as follows: f_1 represents the performance index of the boat, and f_2 represents the overall scale, size of the sailboat. We adopt multiple linear regression to construct the following model:

$$p_1 = k_1 f_1 + k_2 f_2 + k_3 a + k_4 \mathrm{E} + k_5 \#(1)$$

where E is the GDP per capita, a is the age of the sailboat, $k_i(\mathrm{i} = 1,2,\ldots,5)$ represents the corresponding coefficient, and p_1 is the listing price of the monohulled sailboat.

3.4 Model solving for the listing price of a single sailboat

The filtered data were normalized and then solved using multiple linear regression to obtain:$p_1 = 0.1838 f_1 + 0.3892 f_2 - 0.3286 a + 0.1855\mathrm{E} - 0.0013\#(1)$
Significance level α=0.05, the test statistic F=182.4187>2.37 was obtained and passed the test.

3.4.1 Catamaran sailboat model construction

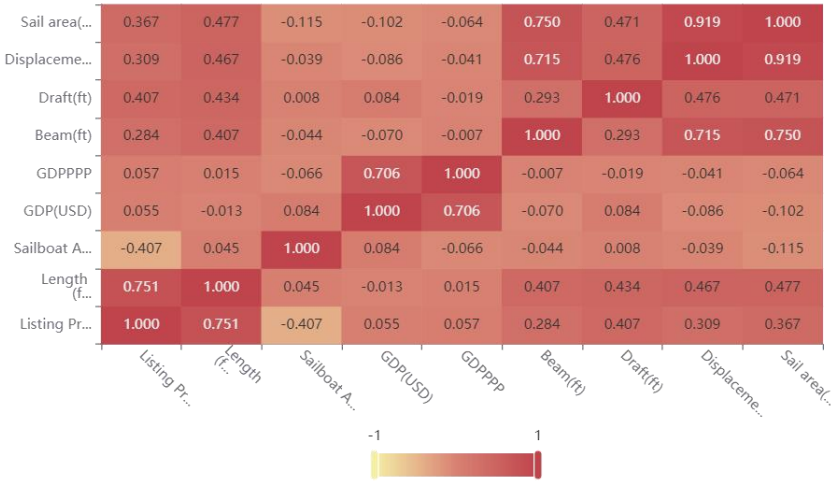

Fig. 5 Heat map of correlation coefficients of factors selected for catamaran sailing

In the same way as above, Figure 5 shows the heat map of the correlation coefficients of the selected factors for the catamaran sailing boat, which is similar to the case of the monohulled sailing boat, but the length has a higher degree of influence on the catamaran sailing boat than the monohulled sailing boat. We used the same method, and after calculation, we obtained KMO=0.786, so we used factor analysis for dimensionality reduction and rotation to obtain the Table 5 component matrix.

Table 5. Component matrix after rotation

	1	2
Length (ft)	0.300	0.746
Beam(ft)	0.894	0.134
Draft(ft)	0.176	0.869
Displacement(lb)	0.872	0.355
Sail area(ft^2)	0.890	0.347

Two factors are obtained, and the cumulative variance contribution rate is 80.946%. Let the two factors be g_1 , g_2 , and we interpret these two factors as follows: g_1 represents the width and part of the performance of the sailboat, and g_2 represents the length and draft of the sailboat, and the following model is constructed by adopting multiple linear regression as follows:

$$p_2 = l_1 g_1 + l_2 g_2 + l_3 a + l_4 E + l_5 \#(1)$$

Where E is the GDP per capita, a is the age of the sailboat, $l_i(\mathrm{i} = 1,2,\ldots,5)$ represents the corresponding coefficient, and p_2 is the markup of the sailboat.

3.4.2 Model solving for the listing price of a catamaran sailboat

The filtered data were normalized and then solved using multiple linear regression to obtain:$p_2 = 0.1574 g_1 + 0.6234 g_2 - 0.4284 a + 0.0427 E - 2.3494 \times 10^{-7} \#(1)$

Significance level α=0.05, the statistic F=199.9259>2.37 was obtained and passed the test.

4. Summary

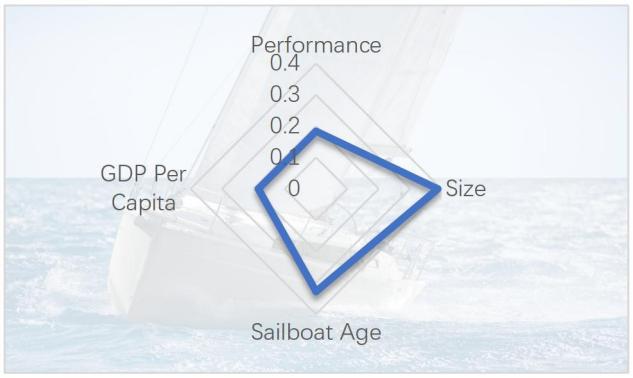

Fig. 6 Radar chart of monohulled sailboats

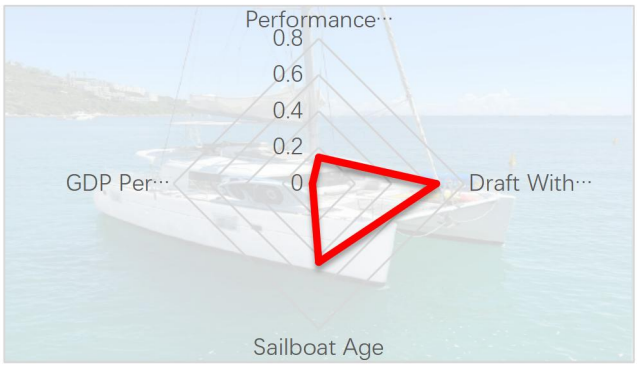

Fig. 7

Figures 6 and 7 show radar plots of the degree of influence of each factor on the listing price of monohulled or catamaran sailboats, respectively. For monohulled sailboats, boat size and age have a

greater degree of influence. For catamarans, boat length, draft and age are more influential and less influenced by local economic conditions.

We give the following explanations for the above-mentioned cases with different degrees of impact:

a. For the impact of economic development, in general, the selling price of sailboats is higher than the GDP per capita, i.e. it is more difficult for families in general economic condition to buy sailboats, which leads to the improvement of economic development within a certain range does not have an obvious impact on the purchase of sailboats, i.e. there is no obvious change in the market supply and demand, so the price of sailboats is affected less. The price of a catamaran is generally higher than that of a monohulled for the same performance, which results in the price of a catamaran being less influenced by economic conditions than that of a monohulled

b. For the size of the sailboat, generally speaking, the larger the sailboat, the more materials are needed to build it, and the more difficult it is to build, requiring better technology and better-quality materials, so the size of the sailboat becomes the main factor affecting the price of the sailboat. The size of the sailboat becomes a major factor in the price of a sailboat. And the size of a catamaran is generally larger than that of a monohulled sailboat, which results in the price of a catamaran being affected by the size of the sailboat.

Accordingly, in order to make the listing price of sailboats more reasonable, we give the following suggestions for the pricing of second-hand sailboats. When pricing used monohull sailing boats, the main basis for pricing is the size and aging of the boat, and the listing price should be increased according to the current economic condition of the listing area and the relevant performance index of the boat. When pricing the used catamarans, the main basis for pricing is the length, draft and aging of the boat, and the listing price will be slightly increased based on the current economic condition of the listing area and the relevant performance indexes of the boat.

References

[1] Xavier J R, Ramesh B. A study on the effect of multifunctional tantalum carbide nanofillers incorporated graphene oxide structure in the epoxy resin for the applications in the shipbuilding industry. Materials Science and Engineering: B, 2023, 289: 116234.

[2] Lazim H M, Abdullah J. Malaysia shipbuilding industry: A review on sustainability and technology success factors. Journal of Advanced Research in Applied Sciences and Engineering Technology, 2022, 28(3): 154-164.

[3] Liu T. Porosity reconstruction based on Biot elastic model of porous media by homotopy perturbation method. Chaos, Solitons & Fractals, 2022, 158: 112007.

[4] Nam H S, De Alwis N, D'agostini E. Determining factors affecting second-hand ship value: Linkages and implications for the shipbuilding industry. WMU Journal of Maritime Affairs, 2022, 21(4): 493-517.

[5] Emelianov V, Zhilenkov A, Chernyi S, et al. Application of artificial intelligence technologies in metallographic analysis for quality assessment in the shipbuilding industry. Heliyon, 2022, 8(8): e10002.

[6] Kou Y, Liu L, Luo M. Lead–lag relationship between new-building and second-hand ship prices. Maritime Policy & Management, 2014, 41(4): 303-327.

[7] Peng W H, Adland R, Yip T L. Investor domicile and second-hand ship sale prices. Maritime Policy & Management, 2021, 48(8): 1109-1123.

[8] Liu T. Parameter estimation with the multigrid-homotopy method for a nonlinear diffusion equation. Journal of Computational and Applied Mathematics, 2022, 413: 114393.

[9] Lee C, Park K. Deep learning-based modeling of second-hand ship prices in South Korea. IAES International Journal of Artificial Intelligence, 2022, 11(3): 886.

[10] Lim S S, Lee K H, Yang H J, et al. Panamax second-hand vessel valuation model. Journal of Navigation and Port Research, 2019, 43(1): 72-78.

[11] McNamara E. Sailboats. Windsor Review, 2019, 52(1): 175-349.

[12] Liu T, Yu J, Zheng Y, et al. A nonlinear multigrid method for the parameter identification problem of partial differential equations with constraints. Mathematics, 2022, 10(16): 2938.

[13] Fahrnholz S F, Caprace J D. A machine learning approach to improve sailboat resistance prediction. Ocean Engineering, 2022, 257: 111642.

[14] Morge A, Pelle V, Wan J, et al. Experimental studies of autonomous sailing with a radio controlled sailboat. IEEE Access, 2022, 10: 134164-134171.

[15] Balmat J F, Lafont F, Maifret R, et al. MAritime RISk Assessment (MARISA), a fuzzy approach to define an individual ship risk factor. Ocean Engineering, 2009, 36(15-16): 1278-1286.

[16] Liu T, Xia K, Zheng Y, et al. A homotopy method for the constrained inverse problem in the multiphase porous media flow. Processes, 2022, 10(6): 1143.

[17] Wibisono Y, Sulistyono N, Putra I G N G M. Implementation of life safety equipment on board a ship operating at the pontoon pier at Gilimanuk Crossing Port of Bali Province. IWJ: Inland Waterways Journal, 2021, 3(1): 49-55.

[18] Michail N A, Melas K D. Quantifying the relationship between seaborne trade and shipping freight rates: A Bayesian vector autoregressive approach. Maritime Transport Research, 2020, 1: 100001.

[19] Liu H, Wu S. Application of multidimensional association rule data mining in the analysis of ship price influencing factors. Journal of Shanghai Maritime University, 2013, 34(4): 31-37.