Problem Chosen	2023 MCM/ICM Summary Sheet	Team Control Number
Y		2333586

Formulating Second-Hand Sailboat Price as a Function of Boat Specifications and Regional Effect

Summary

Like many luxury goods, sailboats vary in value as they age and market conditions change. And to find out what specific factors that affect the price. Second-hand sailboat price is important for manufacturers. To better understand price fluctuation, we figure out a formula to describe how a factor affects the price.

To explain the listing price of each sailboat, we consider additional sailboat information–**beam**, and quantify the sailboat characteristics into beam, length, age, and area effects into climate characteristics and area purchasing power. Ridge regression was applied to obtain the coefficients of the different variables, and the R^2 values of the ridge regression were **0.8** and **0.756** (in both types of boats). To make our model **more explanatory**, so we chose the most frequent of those discrete variables as the basic group so that these coefficients characterize the relative price change brought about when the independent variables change, and each coefficient has a practical meaning.

In the second problem, we follow the results of the model of the first question and explore in depth the significance of the model's coefficient, while using a variable control approach in order to highlight the influence of regional characteristics. Ignoring the factors of boat characteristics, a primary curve is finally obtained, and the significance of slope and intercept is explored on this basis. Namely, the influence of purchasing power and climate. This has some variability in the two sailboat types.

In the Hong Kong (SAR) market, we did not find much data on the listed prices of every sailing boat, for which we gave an explanation. Considering the scarcity of the data, we intend to switch our thinking and choose all the sailboat variants of those Make that sell sailing boats in Hong Kong as a subset of the full set given to fit the results (the R^2 values of the ridge regression were **0.809** and **0.72**), while using the sailboat prices we found in Hong Kong as a test, and performing error analysis.

As for interesting features, We found that **regional historical** factors and **tourism and development** can have an impact on prices.

Keywords: Ridge Regression;Covariance Analysis;Region Effect Quantification

Contents

1 Introduction

The prices of used sailboats are often strongly correlated with region. Previous studies[1] have tended to focus on the correlation between region and price without giving quantitative explanations. In addition, there may be a problem of covariance between different influencing factors, leading to abnormally large regression coefficients.

To address the quantitative description of the region effect and the potential covariance between the independent variables, this paper proposes a Sailboat Asset Pricing Model (SAPM).

The following are the details of the article. The **section 2** of the article enumerates the main assumptions and notations of the model. In **section 3**, we propose the Sailboat Asset Pricing Model (**SAPM**), which is implemented by Ridge Regression and One-Hot code.

Based on the obtained regression equations, we propose a hull characteristic difference effect and a geographic effect to explain the price variation, where the hull characteristic effect consists of brand, age, and underlying hardware characteristics, and the geographic effect can be decomposed into a regional income effect and a climate effect. To the best of our knowledge, our model is the most explanatory of the used sailboat pricing models and is able to explain well the effects that differences in geography and hull characteristics bring to sailboat prices.

In **section 4**, we specifically analyze the geographic effect using the control variables approach, and the results show that based on the benchmark group, our model can better explain the price impact of the climate and income effects among the geographic differences.

In **section 5**, we apply the model to the Hong Kong (SAR) market for empirical analysis to analyze and explain the model fitting effect. We fully explored the sailing market and found the effect of human history and regional development on prices in **section 6**.

And we discuss the strengths and weaknesses of our model in **section 7**. Finally, we provide our empirical findings and suggest effective marketing strategies to sailing intermediaries in Hong Kong in **section 8**.

2 Assumptions and Notation

2.1 Assumptions

To simplify the problem and make it solvable, we propose the assumption below:

Assumptions 1: The effect of a region is reflected in the climate and local purchasing power of that region, that is, these two factors can replace the region factor and play a role in the price.

Assumptions 2: The type of sailboat is reflected in the two features of beam and length, because different types of sailboats have different beams and lengths, so we assume that these two factors can replace the type of sailboat in the price.

Assumptions 3: The price of sailboats is affected by the manufacturer, that is, there is a brand effect.

Assumptions 4: The local purchasing power of a state is replaced by the city with the highest purchasing power in that state

Assumptions 5: If some information of sailboats is not found, then give up the whole sailboat sample directly.

2.2 Notations

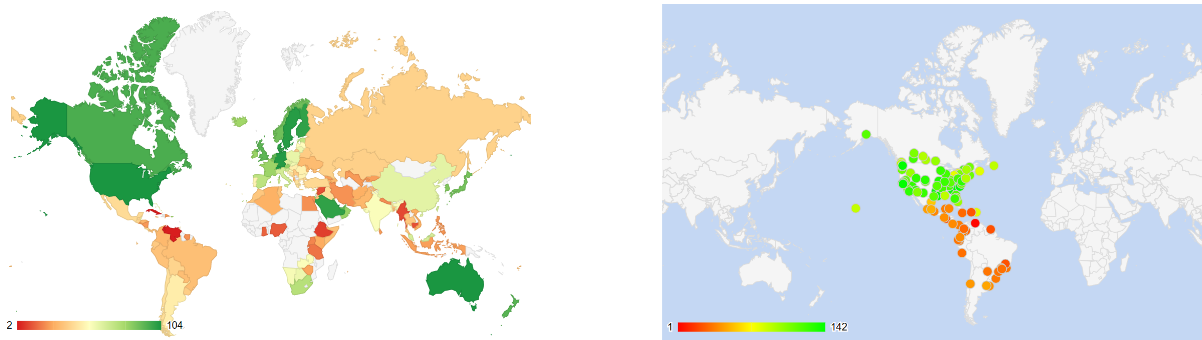
Symbol	Explanation
j	The number of classes in a discrete data
b	Intercept of the regression equation
K	Coefficient of the penalty term
u_i	The error term for the i^{th} sample
M_{ij}	A one-hot code represents the i^{th} maker ,a vector
C_{ij}	A one-hot code represents different climates,a vector
α_j	The coefficient of j^{th} class for maker
W_m	The coefficient vector for maker
β_j	The coefficient of j^{th} class for climate
W_c	The coefficient vector for climate
γ	The coefficient of local purchase power
S_i	Local purchase power for the i^{th} sample
P_i	The price of the sailboat
X_{i1}	The i^{th} sample's beam
X_{i2}	The i^{th} sample's length
$\lambda_{1,2,3}$	Coefficient of beam, length, year
$\hat{P}_i$	Model prediction price

3 Sailboat Asset Pricing Model Using Ridge Regression

3.1 Data Preparation

Because the number of data type provided in the 2023_MCM_Problem_Y_Boats.xlsx file is relatively small, we needed to find additional data to put into the analysis, but since some of the sailboat information could not be found on the website, and, even worse, as the extended information became more and more available, it became more difficult to find all of the information for each sample, taking into account this information.

We looked up additional beam[2-9] factors for sailboats, and, in the process of looking up, we found that the hull material of sailboats is basically fiberglass, so we omitted this factor in the use of the lookup. Also, since regional and economic factors can be considered, we also looked up the climate type of the region or country and the local purchasing power of the country in 2020.



(a) country purchase power figure in 2020 (b) states' purchase power of USA in 2020

Figure 1: Purchasing power distribution heat map

Also, according to assumption 5, we directly discarded some of the samples for which we did not find data. Therefore, the number of data that can participate in the model calculation is 759 for Catamarans and 1724 for Monohulled Sailboats.

3.2 Data Pre-processing

3.2.1 Use one-hot code for discrete data

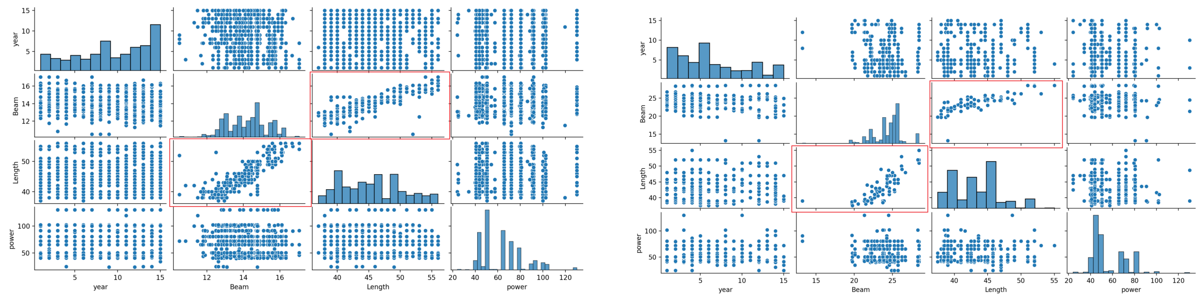
For those discrete data, such as regions, and makers and climate categories, they cannot be directly involved in the calculation. So we use one-hot coding to convert these discrete data into computable quantities, and then put them into the ridge regression model for fitting.

3.2.2 Transform data

In order to keep our coefficient value at a relatively normal size, we divide the price of the sailboat by 10,000, and then the unit is 10k. Also, considering the influence of the year of production, the earlier the boat is produced, the easier it is to depreciate. So we subtract the year of production of each boat from 2020 and name it year_value, so that the earlier the boat is, the larger the value of this item will be, which means This is a negative data, in order to make the coefficient of year_value of our final calculation negative, that is, this has a negative impact on the price of the ship.

3.2.3 Data Characteristics

We draw figures to observe the image of the price distribution of the sailboat, which can be observed to be approximately satisfied with a normal distribution, but there are sudden changes in the middle of the state.

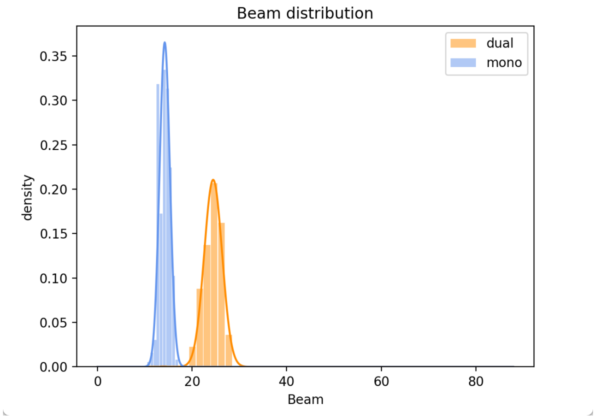


(a) Scatter plot between quantitative data in Monohulled Sailboats.

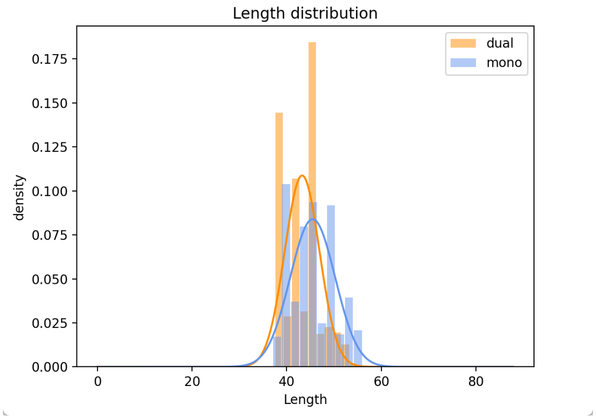
(b) Scatter plot between quantitative data in Catamarans

Figure 2: Scatter plots

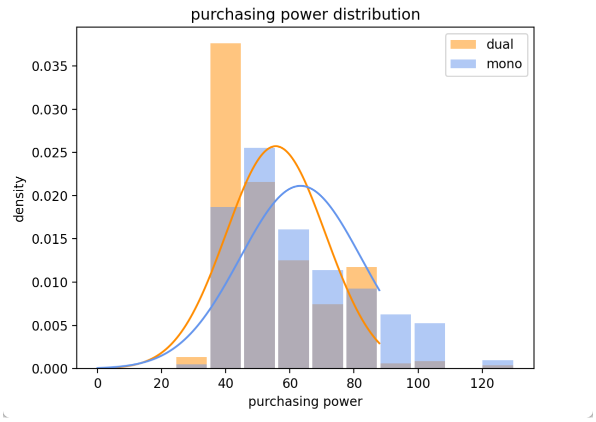
In the following we plot the distribution of the three factors beam, length, and regional purchasing power of sailboats. It is found that length and beam have a clear linear correlation, so we try to find the correlation coefficient of these variables



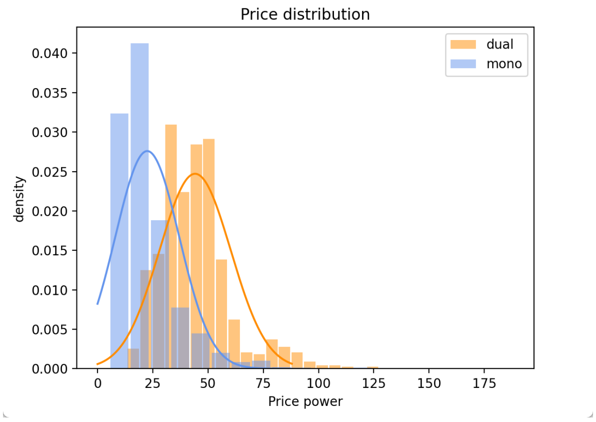
(a) Distribution of beam



(b) Distribution of length



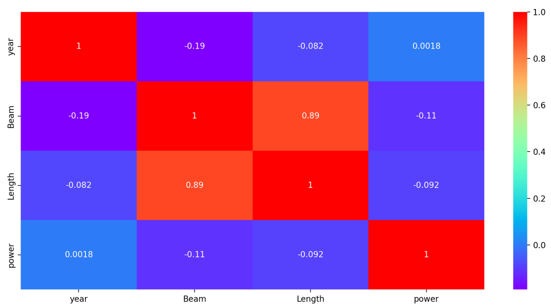
(c) Distribution of purchasing power



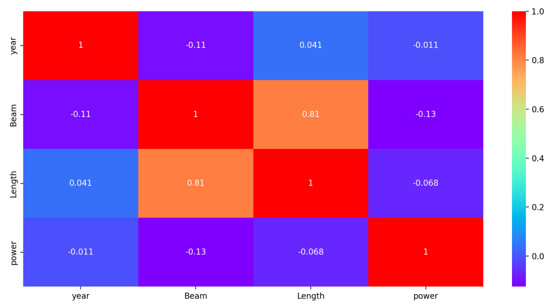
(d) Distribution of price

Figure 3: distribution figures

So we tried to find the correlation coefficients of these variables, in order to further prove the correlation between our data



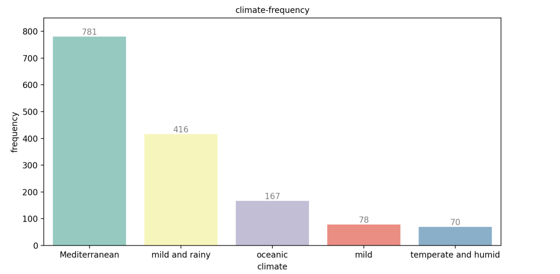
(a) correlation coefficients of the variables in Monohulled Sailboats.



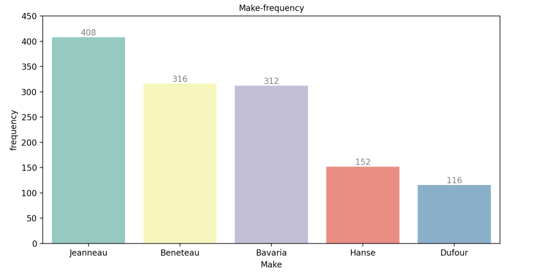
(b) correlation coefficients of the variables in Catamarans

Figure 4: correlation coefficients heap maps

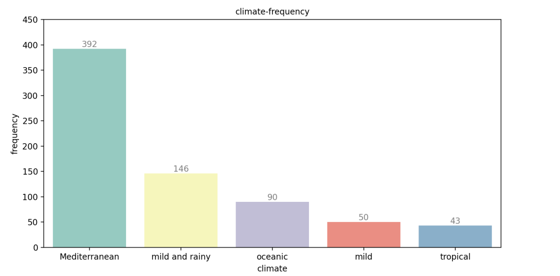
The heat map above illustrates very well the great correlation between our beam and length, which is not difficult to understand, generally the larger the length of the sailboat, the larger its beam, otherwise the boat will be very narrow and not stable enough



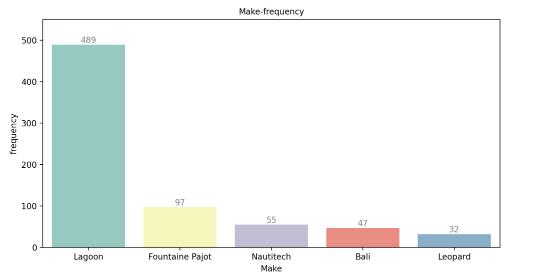
(a) region climate frequency in data sheet Monohulled Sailboats.



(b) maker frequency in data sheet Monohulled Sailboats.



(c) region climate frequency in data sheet Catamarans



(d) maker frequency in data sheet Catamarans

Figure 5: frequency figures

3.3 Build Model

3.3.1 Problem analysis

Our regression model has 2 sets of categorical variables, where Make has a p-dimension and climate has a q-dimension. Take Monohulled Sailboats as an instance. To avoid full linearity of the dummy variables, only $m - 1$ dummy variables need to be introduced for categorical data with m dimensions, where we specify a set of categorical data without specifying dummy variables as the base group for categorical data (as in the regression equation of Monohulled Sailboats specifying **Jeanneau** as the base group for Make and **Mediterranean** as the base group for climate), the intercept of the regression equation means the intrinsic price you have to pay for a Jeanneau brand sailboat in Mediterranean climate (the price you have to pay if you want to buy a sailboat, or Intentional payment). The coefficients of the dummy variables (grade intercept coefficients) indicate the price you are willing to pay more or less for compared to the base group.

We apply ridge regression to data sets containing both discrete and quantitative We abbreviate our model as **SAPM** (Sailboat Asset Pricing Model)

From the above frequency distribution image(fig.4), we choose the category with the highest frequency among these types as the base group, Then we can formulate our idea(n_f in the equation is 3)

$$Price = base + ship\_featureprice + region\_price \tag{1}$$

$$Y_i = b + Y_{ship\_featureprice} + Y_{region\_price} + u_i \tag{2}$$

$$Y_{ship\_featureprice} = \sum_{j=1}^{j=n_f} \lambda_j X_{ij} + \sum_{j=1}^{j=n_m-1} \alpha_j M_{ij} \tag{3}$$

$$Y_{region\_price} = \gamma S_i + \sum_{j=1}^{j=n_c-1} \beta_j C_{ij} \tag{4}$$

3.3.2 Theological Explanation

Here we use y to represent the dependent variable, and the $\hat{y}$ represent the prediction value of y. Variant x can represent each variant, and it may contain more than one attribute, e.g. When there are m attributes: $x_i = (x_{i1}, x_{i2}, ..., x_{im})$ and we delete base group (assume the m^th value belongs to base group), so x_i is $(x_{i1}, x_{i2}, ..., x_{i(m-1)})$

$$g(X_i) = \begin{bmatrix} x_{i1} & x_{i2} & x_{i3} & ... & x_{i(m-1)} & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_{m-1} \\ b \end{bmatrix} \tag{5}$$

which can be written in a more simple form:

$$g(x_i) = X_i^T \cdot W_i \tag{6}$$

So for each X_i we can find a weight value or weight vector W_i. Then X is a set (vector) which consists of all the X_i, and W is a set(vector) which consists of all the W_i

$$X \cdot W - Y = \begin{bmatrix} \hat{y}_1 - y_1 \\ \hat{y}_2 - y_2 \\ . \\ . \\ . \\ \hat{y}_{n-1} - y_{n-1} \\ \hat{y}_n - y_n \end{bmatrix} \tag{7}$$

For the ridge regression model, we need to determine an optimized objective function. It is used to measure the difference between the output of the model and the true value, and we use here the mean square error,that is, for $\hat{P}*_i = f(x_i)$ the error is:

$$\mathcal{L}(W) = ||X \cdot W - Y||_2^2 + K||W||_2^2 \tag{8}$$

and the "||" symbol means:

$$||x||_2^2 = \sum_{i=1}^{n} x_i^2 \tag{9}$$

In this model we add a L_2 penalty term in the objective function. And when the model parameters are large the model may change drastically, i.e., overfitting may occur. Let us now see why adding a linear regression with a L_2 penalty term will reduce the overfitting phenomenon. Because there is a two-parameter squared weight in the loss function, when the weight is too large, the loss of the model will be larger, but the model needs to reduce the loss, then it needs to reduce the value of the weight, once the value of the weight is low, the possibility of sudden changes will become smaller, so to a certain extent can suppress the phenomenon of overfitting.

3.4 Results

3.4.1 Coefficient values

After using the spss software, we get the result shown below:

k = 0.156	Non-standardized coefficient		standardized coefficient	p value
	B	Standard Error	Beta	
b	-32.673	2.054	-	0.000***
λ_2	0.923	0.03	0.303	0.000***
λ_3	-1.152	0.034	-0.332	0.000***
γ	0.077	0.009	0.101	0.000***
λ_1	1.219	0.132	0.092	0.000***

Table 1: coefficient of all quantitative data(Monohulled Sailboats)

K=0.148	Non-standardized coefficient		standardized coefficient	p value
	B	Standard Error	Beta	
b	-75.522	4.082	-	0.000***
λ_2	1.896	0.072	0.431	0.000***
λ_3	-1.885	0.068	-0.463	0.000***
γ	0.035	0.019	0.034	0.069*
λ_1	1.893	0.135	0.222	0.000***

Table 2: coefficient of all quantitative data(Catamarans)

So all in all, we can use an equation to sum up the variants:

$$\begin{aligned} P(x)_1 = &-32.673 + 0.923 \times X_2 - 1.152 \times X_3 + 0.077 \times S \\ &+ 1.219 \times X_1 + W_{m1} \cdot M_1^T + W_{c1} \cdot C_1^T \end{aligned} \tag{10}$$

$$\begin{aligned} P(x)_2 = &-75.522 + 1.896 \times X_2 - 1.885 \times X_3 + 0.035 \times S \\ &+ 1.893 \times X_1 + W_{m2} \cdot M_2^T + W_{c2} \cdot C_2^T \end{aligned} \tag{11}$$

Equation 10 is the fitted function for Monohulled Sailboats prices and equation 11 is the fitted function for Catamarans prices

3.4.2 Contribution of Each Factor

We consider that the contribution of each factor is determined by the magnitude of the standardized coefficients calculated by the model, so the magnitude of the standardized coefficients of each dummy variable for the variables CLIMATE and MAKE of the fixed class is treated in absolute terms, that CLIMATE for example, the magnitude of each CLIMATE standardized coefficient represents the team is relative to the BASE group (Mediterranean), the price change due to climate differences. And the average of these is calculated as the contribution. Finally, a pie chart of the contribution of different factors is drawn

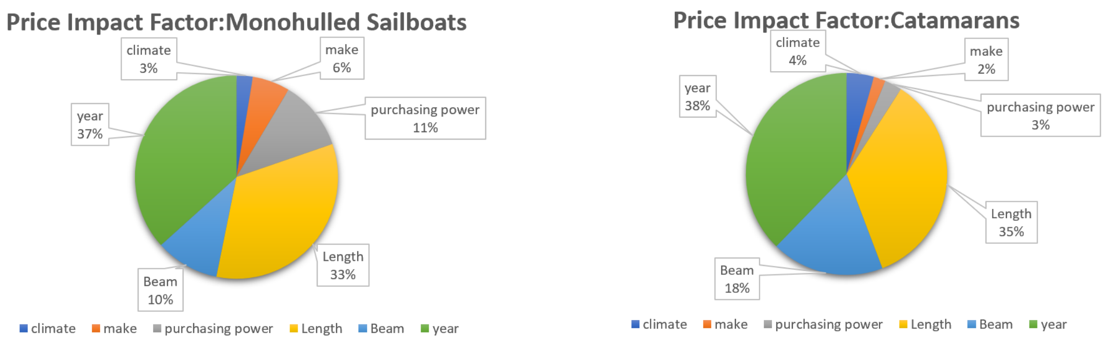


Figure 6: Contribution on two types

From the image above, the year has the largest contribution, followed by beam and length, which are hull characteristics, indicating that the later the year is, in other words, the newer and larger the boat is, the higher the price it will sell for.

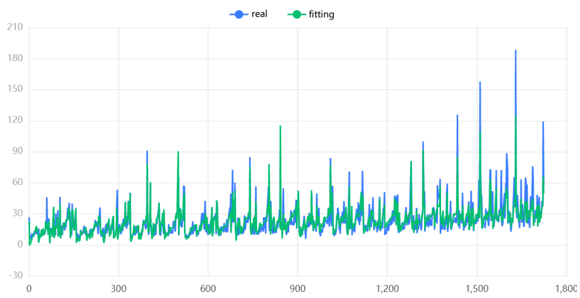
3.5 Validating the Model

We calculated the fitted R^2-values of our model with and without the participation of the region variable at the same time, and found that there was basically no R^2-value basically no change, and, we didn't remove the base group, here are the results:

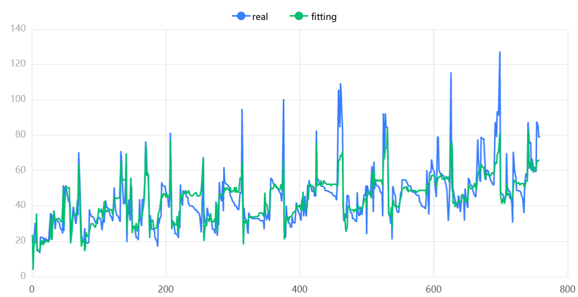
Type	R^2	Adjusted R^2	F
Monohulled Sailboats with region	0.806	0.797	86.457(0.000***)
Monohulled Sailboats without region	0.806	0.797	90.264(0.000***)
Catamarans with region	0.756	0.744	63.88(0.000***)
Catamarans without region	0.756	0.746	70.455(0.000***)

Table 3: Contrast R^2 & Adjusted R^2 & F values

so we removed the variable of region name, which echoed our assumption 1 at the same time. And the results below are based on data which base group was removed.



(a) fitting curve for Monohulled Sailboats.



(b) fitting curve for Catamarans.

Figure 7: fitting curves

Also, to quantify the fit of our model, we calculated the R^2-values of our model for the two datasets as follows:

Type	R^2	Adjusted R^2	F
Monohulled Sailboats	0.8	0.791	89.308(0.000***)
Catamarans	0.756	0.745	75.004(0.000***)

Table 4: R^2 & Adjusted R^2 & F values of our model

4 Explain the Region Effect on Price

4.1 Problem Analysis

The ridge regression model above enables the calculation of specific coefficients that have practical significance, which also better explains the magnitude of the effect of different variables on prices. Also this question requires us to explain the effect of region on price through our model, so we focus on our coefficients and hope to uncover possible effects from them

4.2 Meaning of Model Coefficients

- α_j and β_j : The coefficients of the dummy variables α_j and β_j are step intercepts. α_j measures the brand premium that consumers are willing to pay for the j^{th} Make produced sailboat, and β_j measures the climate premium that consumers are willing to pay in the j^{th} climate compared to the base climate

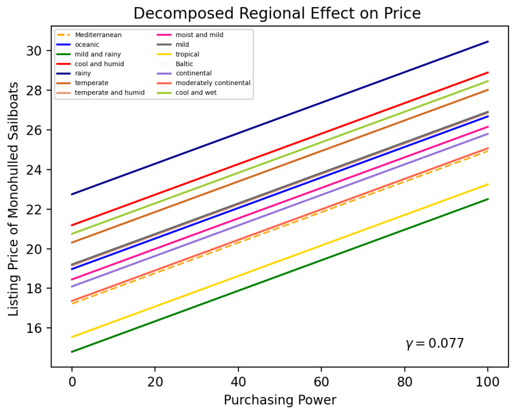
- γ : The price people are willing to pay for a sailboat per unit of Purchasing power per consumer compared to the base purchasing power

- λ_j : The price consumers are willing to pay for the j^{th} hull feature per unit compared to the base groups

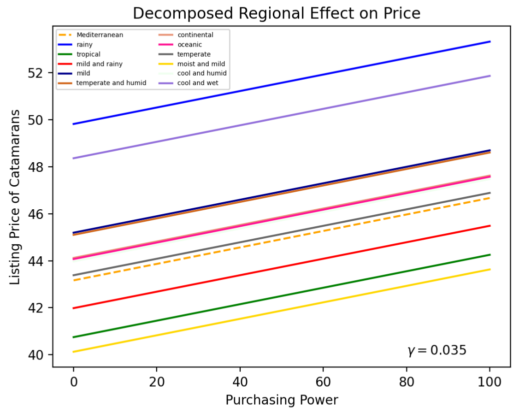
4.3 Our Method for Explanation

Since the specific model can be determined by the Make and hull characteristics, the problem of consistency between geographic effects and specific models can be transformed into the problem of discriminating the consistency of geographic effects for different sizes of boats produced by different makes. Then we take the base group in Make, which is Jeanneau, which means that other dummy variables are 0, and the beam and length are the average of the sailboats made by Jeanneau. This way we can control the other variables from changing and focus only on climate and local purchasing power, that is, regional effects. For Catamarans part, we choose Lagoon as the base group, and the beam and length are the average of the sailboats make by Lagoon.

The dashed line in the figure below represents the base group, and then the upper and lower positions of the intercept of the line can represent the price needed to buy a sailboat whose make is Jeanneau(Lagoon) under the average beam and length without considering the purchasing power, but only considering the influence of climate factors in the regional effects.



(a) Monohulled Sailboats



(b) Catamarans

Figure 8: climate effect on different kind of boats

At the same time, we found that the magnitude of climate change on price in Monohulled Sailboats is relatively small compared to that in Catamarans. Meanwhile, the slope in Monohulled Sailboats is larger than that in Catamarans. This means that climate has a relatively large impact on price in Catamarans, but purchasing power has a relatively large impact on Monohulled Sailboats.

This is also reflected in the pie chart in fig.6, where the regional purchasing power contribution is 11% for Monohulled Sailboats, but only 3% for Catamarans.

Our guess is that the price of sailing boats in Catamarans is generally higher, and on this basis, the price fluctuation will not be very big, which is in a certain state of saturation, but the price of sailing boats in Monohulled Sailboats is generally lower, so the price fluctuation will change a bit more with the fluctuation of purchasing power, because the regions with large purchasing power will spend more money to buy boats, and it is easier to raise the price.

5 Apply SPAM to Hong Kong (SAR) market

5.1 Data Preparation

We looked up a very small selection of sailboat types for sale in Hong Kong on the sailboat for sale website [3], with the following results:

Make	Variant	Price	Year	Climate	Purchasing Power
Beneteau	Sense 43	220,000	2012	subtropical	65.3
Beneteau	Oceanis 51.1	543,375	2018		
Beneteau	Oceanis 38	165,000	2014		
X-Yachts	X-55	380,362	2009		
Lagoon	450	685,000	2017		
Lagoon	450	560,000	2016		
Nautor	54	1,770,000	2019		

Table 5: price data for that subset from the Hong Kong (SAR) market[2]

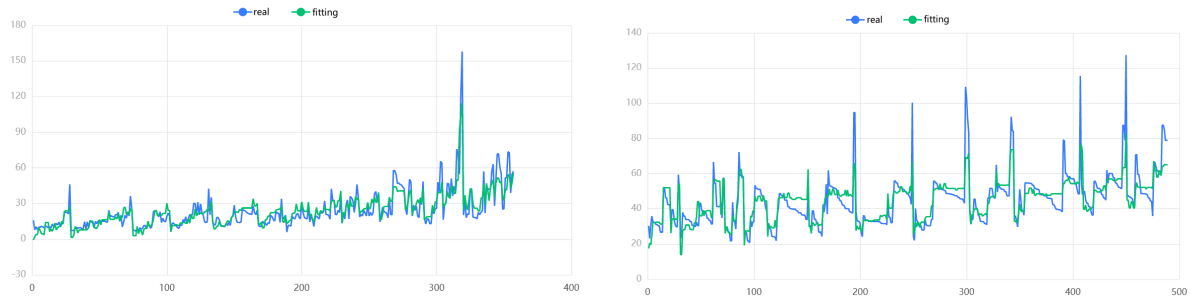
5.2 Problem Analysis

Considering that Hong Kong is only one city, it has a limited number of types of sailboats purchased, so we will use only these 7 samples as the test data for our model, while we select all sailboats with the same manufacturer information as in the above table as a subset in the full set given. As before, we do not consider the ship variant here either, because according to our hypothesis 1, the model variant of a sailboat can be replaced by its length and beam.

Also since We had a small amount of data, so we switched our thinking anyway, using our hypothesis 2 and replacing specific regions with climate and local purchasing power, so that the results of the model equations fitted in other regions could be applied to the Hong Kong region. And the small amount of sailing boat prices we found in Hong Kong was used to test the goodness or badness of our fitting effect.

5.3 Fitting

In this problem, because our data set has changed from a full set to a subset, we need to perform a new fit, taking into account the limited types of sailboats sold in Hong Kong, for example, the Catamarans type is only Lagoon, so in the process of our new fit we need to ignore the Make factor, which is the same, and the climate type and the region type will be somewhat reduced because we selected a subset of the climate and region types involved in a subset. Under these conditions, the results of our fit are as follows:



(a) Monohulled Sailboats (b) Catamarans

Figure 9: fitting curves

5.4 Result and Error Analysis

In this fit, the R^2 of our fit results reached 0.809(Monohulled),0.72(Catamarans) respectively, which is not much different compared to the previous fits on the full set.

Type	R^2	Adjusted R^2	F
Monohulled Sailboats	0.809	0.796	64.144(0.000***)
Catamarans	0.72	0.711	86.889(0.000***)

Table 6: R^2 & Adjusted R^2 & F values of our model for subset

As in the second problem, the regional effect in Hong Kong is reflected in the climate and purchasing power. The coefficients of these two factors fitted by the model, taking the coefficient of purchasing power as an example, have a coefficient size of 0.093 and 0.007 on the two data types Monohulled, Catamarans, similar to the effect in the second problem. The magnitude of the coefficient of the former is larger than that of the latter, indicating that among the vessels sold in the Hong Kong market (all vessel types in our subset), the regional purchasing power contributes more in Monohulled type

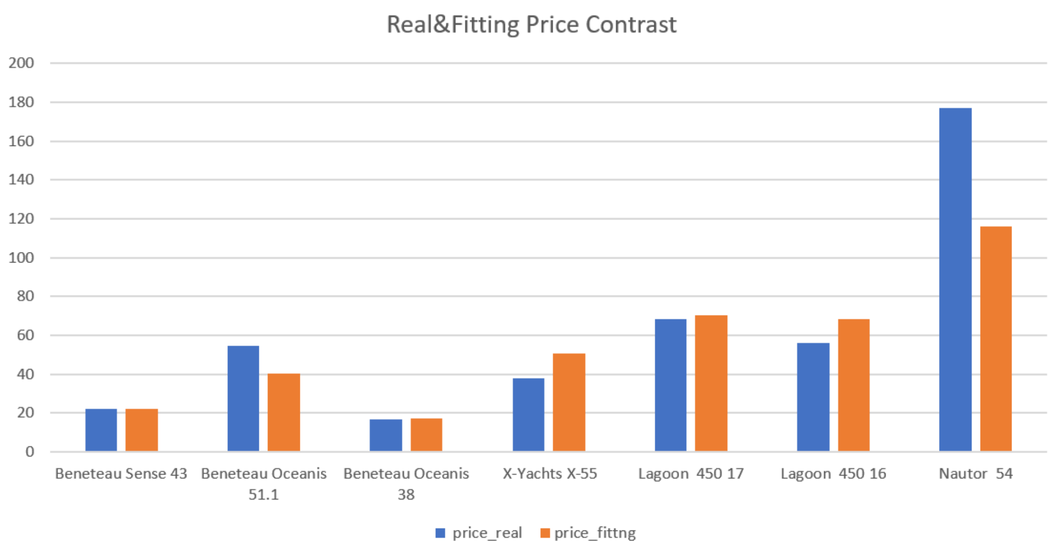


Figure 10: Contrast

We have given the following possible reasons for the error in the model:

1 Error in the source of information, the price of sailboats from the website [10] may be artificially high due to the seller.

2 Hong Kong is a global financial center with a large number of rich people and a high demand for sailboats, compared to the low supply of used sailboats in Hong Kong, resulting in a seller's market situation.

3 The number of used sailboats for sale on the website[10] is not more than 10. The model fits a small sample with random error, which leads to a poor fit given the small sample size.

4 The limited area of Hong Kong's waters and the tight berths result in high mooring fees, which are also reflected in the price of used sailboats, a variable not taken into account in our model.

5 Hong Kong is an international shipping center and has a well-developed maritime service industry, which may lead to higher prices for used sailing boats due to their better maintenance condition.

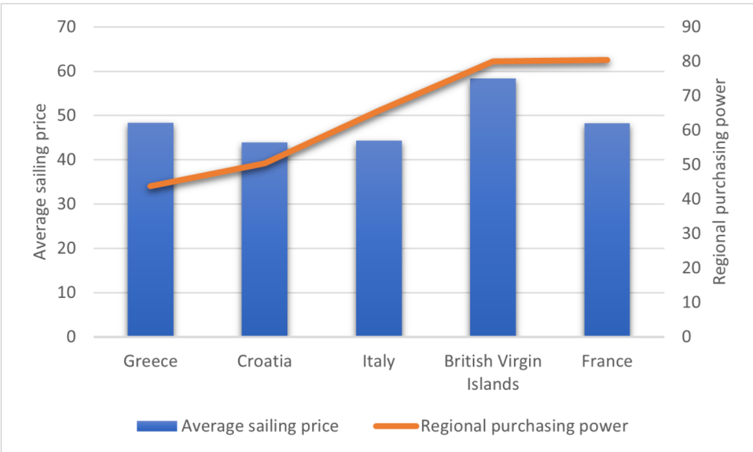
6 Interesting Features

6.1 Tourism and Development Factors

According to the above standardized coefficients: moist and mild, tropic climate has the largest price increase, while oceanic has the smallest. Moist and mild, tropic climate regions generally have warm weather and beautiful beaches, so the local tourism industry is developed and the sailing development level is high. However, oceanic maritime climate regions have many cloudy days and rainy days, and there are fewer days suitable for sailing in a year. From the above, it can be seen that the price is cheaper in areas with a high level of sailing development and higher in underdeveloped areas.

6.2 History Effect

On the other hand, the model shows that the price of sails is positively correlated with local purchasing power. However, when we extract the average price of a Lagoon 450, we find that the sailboat prices in Greece and the United Kingdom conflict with the above relationship. After searching, we learned that Greece and the United Kingdom have a long history of sailing, and now Greece is the world's largest ship-owning country and the United Kingdom has the largest sailing club and racing area. Therefore, from a historical and cultural perspective and modern development, sailboat prices in Greece and the United Kingdom will be higher than other regions.



7 Strengths and weaknesses

7.1 Strengths

- **Feature transforming**
  In our model building, we search for local purchasing power and climate data from various regions, and replace the "region" factor into the above two more intuitive and easy-to-quantify characteristics that affect the price. Which also is our assumption 1

- **Accuracy**
  Our model uses the ridge regression method and achieves an R^2 value of

0.8 in data sheet Monohulled Sailboats,and 0.756 in data sheet Catamarans, indicating a good fit. Meanwhile, The p-value test for the coefficients of the quantitative data is basically at 0.000(***) indicating a significance of 1%, indicating that the results of our fitted coefficients are reliable

- **Interpretability**
  Because our model gives a coefficient for each variable, each coefficient has a certain realistic and statistical significance, and can clearly derive the price fluctuations brought about by changes in the variables, so the interpretation is very strong

7.2 Weaknesses

- **Lack of other potentially relevant factors**
  Different models of sailboats have other characteristics, such as draft depth, sailboat material, warranty date, etc., but we did not take these factors into account for research. Because our initial consideration was that if we needed to find more data, we might lose the sample size because some of the sailing data could not be found, so we chose to find only a portion of the data in comparison, so the results have some limitations

- **Manufacturing time**
  We believe that the newer the ship is, the more expensive it is, But a sailboat is a kind of luxury items, and it is possible that the earlier the production date, the more collectible the sailboat is, and this may bring an increase in price, but we ignore the collection value of old ships.

- **Other impacts of regional effects**
  When analyzing the regional effects of sailboats, But this region is influenced by not only the purchasing power of the region and the climate, but probably also the number of ports, whether coastal or not, and GDP per person,To simplify the problem, we did not take them into account

8 Letter

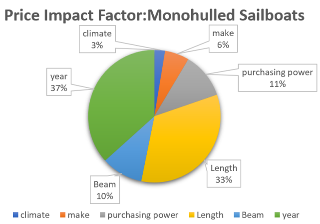
Dear Hong Kong
sailboat broker
Our team is very interested in sailboats, especially their prices. Therefore, we analyzed COMAP's data and obtained a well-fitting model. We hope this model can be of help to you in terms of pricing.

There are many factors that affect the price of sailboats. We have classified these factors into brand and model effects, regional effects, and other effects.

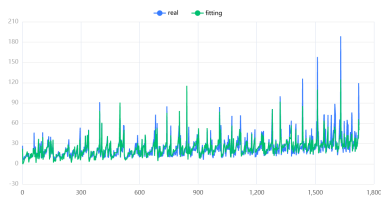
In addition, we also searched for the necessary data on various authoritative websites to improve the accuracy of the model.

In our modeling, we found that the model and year of manufacture of the boat had the most significant impact on the price, while the local climate had the most negligible impact. As shown in the subfig.a. In this case, we suggest that when determining the price of a boat, you should focus on the impact of the model and year.

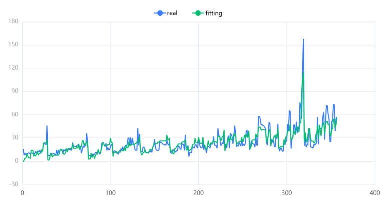
In the end, as can be seen from the subfig.b., our model achieved excellent fitting results. In addition, we looked for existing boats in Hong Kong and other influencing factors such as local climate and purchasing power. We also extracted the prices of the same model boats in other regions from the total data and fitted them again into the model. As can be seen from the subfig.c, the fitting results also showed that our model has very high accuracy. The figures are all fitting curves for monohulled sailboats.



(a) contribution of each factor



(b) fitting curve for Monohulled Sailboats



(c) fitting curve for Monohulled Sailboats on subset

Additionally, we also found that historical and tourism factors make boats in Greece, the UK and underdeveloped tourist areas more expensive. Therefore, we do not recommend buying the same model of boat in these areas as in other regions. Thanks for taking the time out of your busy schedule to read my letter, and we hope our suggestions can be of help to you.

Sincerely yours

MCM 2023 Team 2333586

References

[1] Kim, Harrison & Thurston, Deborah & Kwak, Minjung. (2012). Formulating Second-Hand Market Value as a Function of Product Specifications, Age, and Conditions. Journal of Mechanical Design. 134. 10.1115/1.4005858].

[2] http://www.yachtworld.com/

[3] http://www.annoncesbateau.com/

[4] http://www.cruisingworld.com/

[5] http://www.racing-yachts.com/

[6] http://boat-specs.com

[7] http://jeanneau.com

[8] http://x-yachts.com

[9] http://sailboatdata.com

[10] http://luxboating.com

Appendices

Appendix A Values in coefficient vector

A.1 Monohulled Sailboats

Make	Alubat	Bavaria	Beneteau	Catalina	Comar
value	9.976	-2.638	-0.39	0.158	1.596
Dehler	Delphia	Dufour	Elan	Grand Soleil	Hallberg-Rassy
3.505	1.431	-0.702	-1.365	6.035	25.315
Hanse	Harmony	Hunter	Hylas	Island Packet	Malo
0.209	-4.706	0.004	21.174	14.469	17.733
Moody	Najad	Nauticat	Salona	Sunbeam	Sweden Yachts
17.804	24.404	24.893	1.669	11.18	17.306
Tartan	Van De Stadt	Wauquiez	X-Yachts	Alliage	Nordship
13.426	6.198	10.956	17.067	20.466	13.424
Sabre	Zuanelli	Allures	Amel	Archambault	Bestevaer
7.846	12.169	13.569	20.979	2.482	47.313
Cabo Rico	Contest	J boats	J Boats	Nautor	Oyster
15.468	26.2	1.436	7.957	70.187	45.94
Poncin	Rustler	Sly	Morris	B-Yachts	RM Yachts
-9.763	20.091	6.703	30.195	17.717	7.272
Solaris	Southerly	Triplast	Discovery	Vismara	Azuree
20.131	29.282	2.567	80.698	24.562	7.308
Tofinou	Boreal	Passport	Maxi		
12.621	31.229	51.577	6.758		

Table 7: values in W_m

climate	oceanic	mild and rainy	cool and humid	rainy
value	1.742	-2.434	3.955	5.517
climate	temperate	temperate and humid	moist and mild	mild
value	3.08	1.92	1.218	1.972
tropical	Baltic	continental	moderately continental	cool and wet
-1.689	0.08	0.862	0.133	3.519

Table 8: values in W_c

A.2 Catamarans

Make	Fountaine Pajot	Leopard	Nautitech	Broadblue	Maine Cat
value	5.057	-3.232	4.752	6.518	8.632
Make	Privilege	Catana	Seawind	Alliaura	Chris White
value	6.453	15.139	6.634	12.172	9.223
Make	Outremer	Voyage Yachts	Dolphin Ocema	Bali	Dix Harvey
value	28.165	5.372	-12.245	0.218	5.917

Table 9: values in W_m

climate	rainy	tropical	mild and rainy	temperate	moist and mild
value	6.649	-2.414	-1.181	0.218	-3.036
mild	temperate and humid	continental	oceanic	cool and humid	cool and wet
2.027	1.935	0.954	0.906	0.807	5.192

Table 10: values in W_c