Problem Chosen	2023	Team Control Number
Y	MCM/ICM Summary Sheet	2331037

Using Variable Coefficient Model to Estimated Sailing Price

Summary

Nowadays, with the increasing development of the world economy, sailing has been used by more and more people as a new way of entertainment, which makes the second-hand sailing market expand rapidly. To explore the pricing reasons for second-hand sailing boats in the market, we model the pricing of second-hand sailing boats.

Firstly, we get the factors that affect the price of sailboats, including sailboats' parameters and regional influences. Firstly, correlation analysis is carried out among various parameters. Finally, the LWL, draft, water, S.A., fuel, GDP, the number of harbor, travel, and temperature index is selected as the index of model establishment, and regression analysis is carried out. It is obtained that LWL, draft, water, GDP, labor, and travel are roughly linear with it, and fuel, S.A., and temperature have a complex nonlinear relationship with price. Then, using the version parameter model, the model formula of price and parameter is obtained, namely

$$\begin{aligned}\hat{Y}_t =&0.88496715X_{1t} + 0.07983221X_{2t} + 0.15883855X_{3t} + 0.921148X_{4t} + 0.89238497X_{5t} \\ &+ 0.89291231X_{6t} + \hat{\alpha}_1 (U_t) + \hat{\alpha}_2 (U_t) Z_{1t} + \hat{\alpha}_3 (U_t) Z_{2t}\end{aligned}$$

In the second question, we choose four indicators that can portray the influence of region on ship value and simulate them following the model of the first question to derive the differences in the influence of different regions on ship prices and select representative examples for analysis. In the third question, we collected data on sailboat transactions in Hong Kong and tested the model using the second question, and found that the correction coefficients derived in the second question did not apply to prices in Hong Kong. By adjusting the correction coefficients, we obtain more accurate prediction results and based on this, we discuss the difference between Monohulled Sailboats and Catamarans regarding the regional impact.

Finally, we give some interesting facts and advise Hong Kong brokers about the sailing boat market.

Keywords: variable coefficient model; price estimating;region influence.

Contents

1 Introduction

1.1 Problem Background

As a non-essential means of transportation in daily life, sailing boats are mainly used for leisure and entertainment. That makes sailing boats have many luxury characteristics, such as their value changing with age and market conditions. Studying the relationship between the sailing boats' parameters and values is necessary. On the one hand, it can price sailing boats out of the factory. At the same time, it can also guide the second-hand sailing boats market. Using data provided by a boating enthusiast, we can obtain reliable data for different sailboat models, albeit with some errors.

1.2 Restatement of the Problem

Sailing boats are usually traded by brokers. A sailing boats broker in Hong Kong(SAR), China, would like to obtain a report on the pricing of second-hand sailboats. You can complete the broker's request with the information provided by sailing enthusiasts and data searched by yourself, which includes:

- Build a mathematical model to explain the sailing boats' prices given by the table. You can use the data provided by the question or the data searched (describe and identity the sources) as a basis. You also should pay attention to discussing the precise precision of the sailing boats' price.
- Determine whether different regions have an impact on the listing price of sailing boats and consider whether the effect is the same for all types of sailing boats.
- The model is applied to the Hong Kong market, then select some types of sailboats from the available types and divided into two types —- Monohulled Sailboats and Catamarans. According to the regional characteristics of Hong Kong, the price range of each type of sailboat is determined, and whether the Catamarans and monohulled sailboats are the same.
- Identify other useful, meaningful inferences or conclusions
- A one to two-page report including ICONS is provided for sailing brokers in Hong Kong to facilitate their understanding of the conclusions and models.

1.3 Literature Review

The question is mainly related to the second-hand recovery price estimation algorithm. In recent years, there are many related types of research on second-hand recycling, most of which are multiple linear regression, current market value method, replacement cost method, support vector machine, and neural network algorithm, etc. The multiple regression model is mainly used for linear problems, but it ignores the interaction effect of nonlinear relations and variables; The forward market price method is based on the active market, but sailboat liquidity is poor, so it is not suitable for this problem; The replacement cost method involves a lot of reference data, but the data provided by the question and the open data on the Internet all inaccurate and incomplete; However, support vector machine and neural network have the disadvantages of large computation and poor interpretation. In recent years, the semi-parametric model combines the advantages of the above models, so we will use the value evaluation model based on the version parameter model in this case.

- The semi-parametric partially additive model is the extension of the semi-parametric model, which contains both linear and nonlinear relations between dependent variables and explanatory variables. It has strong modeling ability and flexible interpretation ability; and can avoid the "dimensional disaster" problem of non-parametric regression surface. Therefore, the model has become a powerful statistical inference tool for high-dimensional data analysis. Scholars in many fields have conducted systematic research on semi-parametric models, and deeply studied the estimation theory, model recognition, and variable selection of semi-parametric models.
  For example, Hastie et al.(1990))[1] introduced the additive modeling; Li(2000)[2] came up with a method for estimating the parametric of the linear model; Fan et al. (1998)[3] raise a method of kernel estimating for part linear mode; Liang et al.(2008)[4] use attenuation correction to make the progressive nature of the parameters.
- In the same way, important semi-parametric models have been developed rapidly in recent years, so that the models can not only explain the linear and nonlinear relations; but also reflect the interaction between variables. Zhang et al.(2002)[5]came up with a theory about partial polynomial estimation method and asymptotic theory; Xia et al.(2004)[6]A semi-local least squares estimation proposed method and model selection method; Fan & Huang(2005)[7]The least square estimation method of the section is proposed and the section similarity is established natural ratio statistics and Wald statistics and their asymptotic distribution;

1.4 Our Work

The problem requires us to understand different parameters' influence on sailing boats' prices. Our work mainly includes the following:

1. Building a semi-parametric model and analyzing the influence on price;
2. Dividing all regions into three parts and all types of boats into three main classical types. Then analyze the region's influence on a different type of boat.
3. Based on the above data and results, we formulate the sailing boats' price interval.

In order to avoid complicated descriptions, and intuitively reflect our work process, the flow chart is shown in Figure 3:

2 Assumptions and Explanations

Considering those practical problems always contain many complex factors, first of all, we need to make reasonable assumptions to simplify the model, and each hypothesis is closely followed by its corresponding explanation:

Assumption 1: **Only the influence of those parameters on the price is considered, and other factors are ignored.**
Explanation:We should use the most representative and related data and we select different indexes by reading the relative paper.

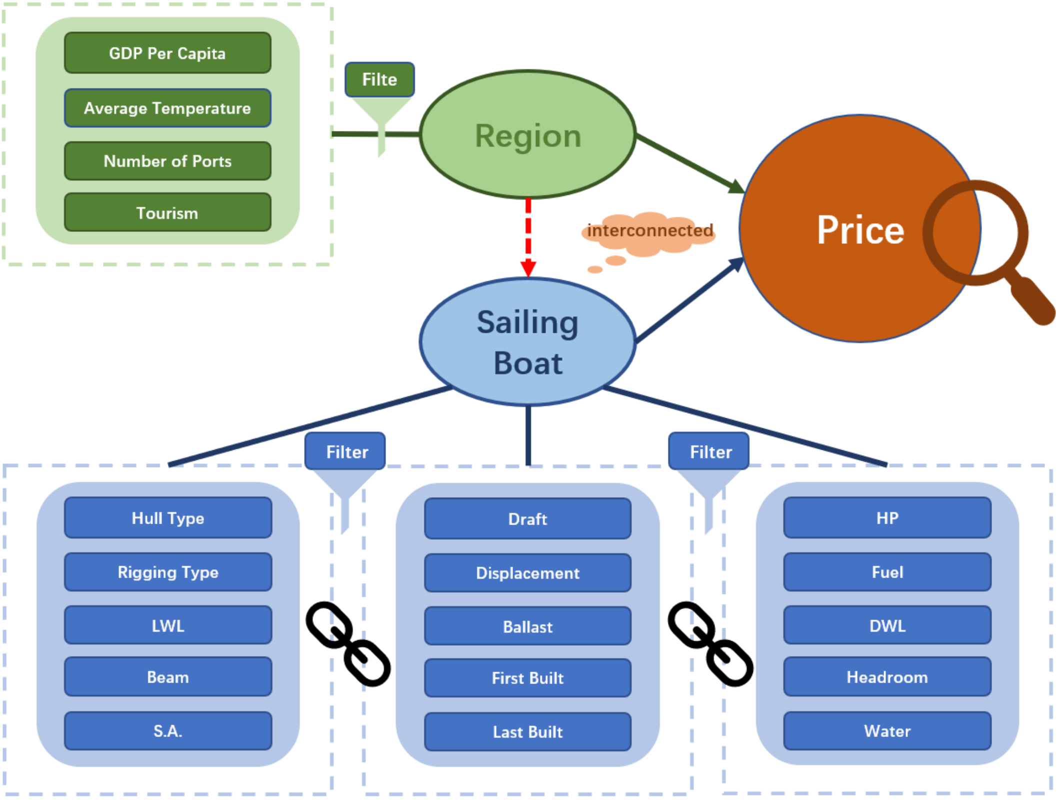


Figure 1: Flow Chart of Our Work

Assumption 2: **We assume all regions' people have the same passion for sailing.** *Explanation:*Though in some regions sailing boats are not popular, their price only depends on the GDP and whether near the sea.

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

3 Notations

Some important mathematical notations used in this paper are listed in Table 1.

4 Model Preparation

4.1 Data Overview

The question did not provide us with data accurate and comprehensive, so we need to pretreat with the data provided as well as consider which data to collect in the model building. Through the analysis of the problem, we need to collect the relevant information of sailing boat parameters such as **hull type, rigging type, LWL, beam, S.A. (reported), draft, Displacement, Ballast S.A./Disp., Bal./Disp., Disp./Len., First Built, Last Built, HP Fuel, DWL Headroom, Water** and so on.

Table 1: Notations used in this paper

Symbol	Description
X_1	The sailing boats' Water storage capacity vector
X_2	The sailing boats' draft capacity vector
X_3	The sailing boats' waterline length
X_4	The average GDP of the different region
X_5	The number of ports of the different region
X_6	The tourism industry of different regions
U	The temperature of the different region
Z_1	The fuel consumption of sailing boats vector
Z_2	The sail area vector

*Those notations are used in chapter **Appilication**.

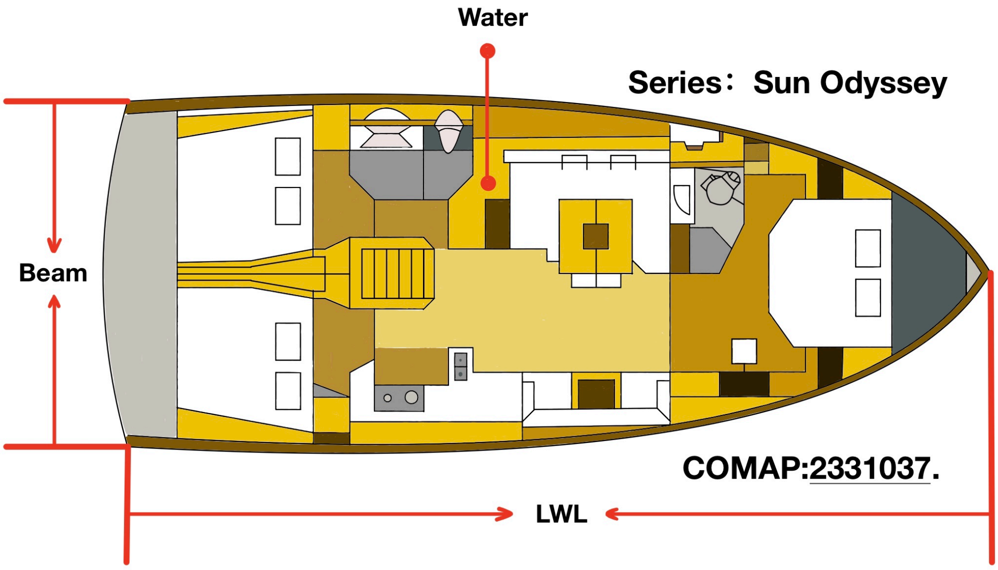


Figure 2: Literature Review Framework

4.1.1 Data Collection

The website of sailing boats was queried and lots of data about sailing boats' parameters were obtained. And other data sources are shown in Table x.

Table 2: Data and Database Websites

Database Names	Database Websites
The different regions' GDP and	https://xueqiu.com/
The sailing boats' parameters	https://sailboatdata.com/sailboat/sun-odyssey-42-cc-jeanneau

4.1.2 Pretreating data

1. Because the sailboat data is incomplete, so we use the mean of the all data for filling in the data
2. There are different prices in different regions for the same type of sailboat, the average of all regional prices for the same type of sailboat is taken as the price for that type of sailboat;
3. Regarding the depreciation problem, we considered the effect of time on the value of a sailboat, especially considering the depreciation of a sailboat due to the reduced life span and other issues. We referred to the depreciation model for used cars and determined a depreciation rate parameter of about 0.9 per year by analyzing the information on the prices of ships produced in different years given in the question. Based on this depreciation rate parameter, we uniformly transformed the data for the same type of ship produced in different years to the value of the latest year in which that type appears in the question data. In this way, we eliminated the effect of time on the price of the ship and thus were able to better analyze the effect of other indicators on the price of the ship.
4. We conducted a data search in different regions, mainly including the United States and Europe. It is not difficult to find that the sailing data in the United States accounts for more than half of the sailing data, so we separated the United States from other regions, that is, the United States is taken as the unit of states, and other regions are taken as the unit of countries.

4.1.3 Data Screening

To better understand the distribution of sailing ships intuitively, we made a heat map, of course, the heat map data is taken from the data provided by the given statistics.

1. The analysis of locations of hovering VHF/UHF radio-repeater drones for fires can be more accurate if we have more complete data;
2. The assumption that the "boots-on-the-ground" forward teams can be approximated as near the fire site is a bit idealized. If the trajectory of the team is taken into account, a more practical model and results can be obtained.
3. Some approximate analysis methods are applied to model other places, which may lead to a situation that not to be the most optimal.

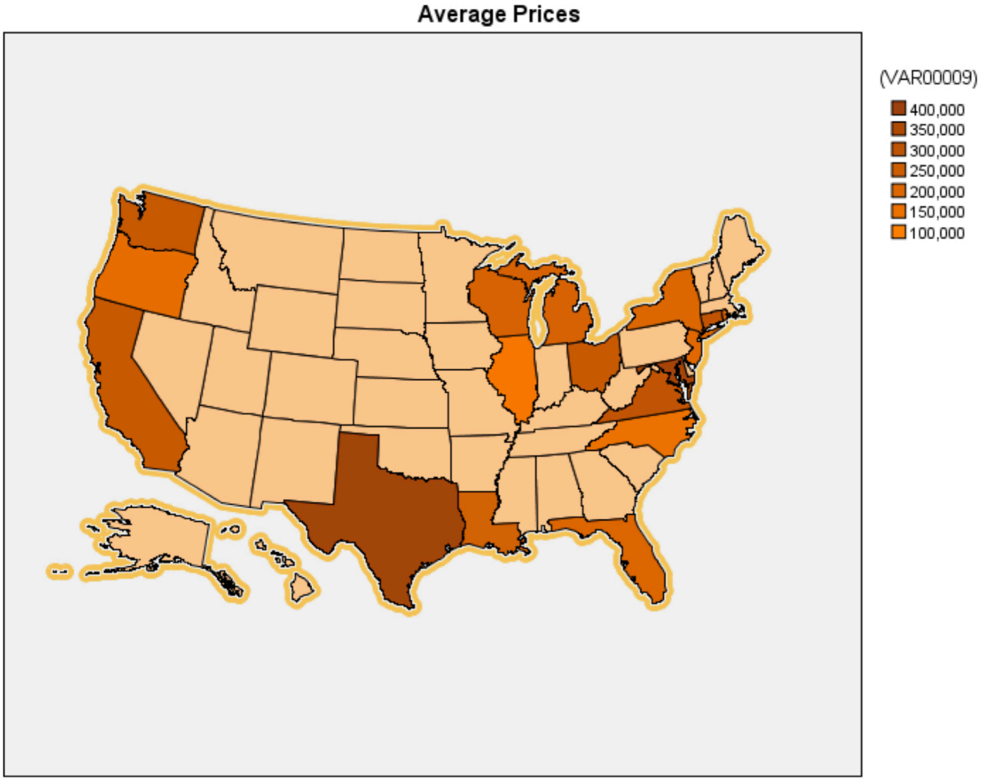


Figure 3: Literature Review Framework

5 Related models and algorithms

5.1 Linear regression model

We assume$Y_t$are dependent variables and $X_t = (X_{t1}, \cdots, X_{tp})^T$is a p-dimension covariable, then a linear model could be presented as

$$Y_t = \beta_0 + X_{t1}\beta_1 + \cdots + X_{tp}\beta_p + \varepsilon_t, t = 1, \cdots, n,$$

we also can use the matrix to rewrite it, thus

$$Y = X\beta + \epsilon,$$

which$X = \left(X_1{}^T, X_2{}^T, \cdots, X_{\text{n}}{}^T\right)^T$is a matrix cobined by n vector. The error in this model is $\epsilon = Y - X\beta$. We want to make ϵ reach the minimum number, so we have

$$Q(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T\varepsilon = (Y - X\beta)^T(Y - X\beta) = Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta$$

by taking the derivative with respect toβ, we get

$$\frac{\partial Q(\beta)}{\partial \beta} = -2X^TY + 2X^TX\beta = 0,$$

then we have

$$X^TX\beta = X^TY.$$

The solution is not difficult to obtain, which is

$$\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_p\right)^T = \left(X^T X\right)^{-1} X^T Y$$

We plug this solution into the linear equation and call

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_0$$

logistic regression equation.

5.2 Nonlinear regression model

We also need to think about the nonlinear relationship between the concomitant variable and the dependent viable. Assume the concomitant variable and dependent variables have a nonlinear relationship, then we can present it as

$$Y_t = m\left(X_t\right) + \varepsilon_t, t = 1, \cdots, n$$

which $\epsilon_t$is error($E(\epsilon_t) = 0$)and $m(\cdot)$is a unknown and smooth function.

5.3 additive model

This model has the formula as

$$Y_t = \alpha + \sum^{q} \phi_i\left(Z_{it}\right) + \varepsilon_t, t = 1, \cdots, n,$$

which $\phi_i(\cdot)$is a single-variable and smooth function, ϵ_t independent to Z_{it}. Meanwhile, we assume $E\epsilon_t = 0, Var(\epsilon_t) = \sigma^2$.

So if the additive model is correct, then for all k we have

$$E\left(Y_t - \alpha - \sum_{i=1, i \neq k}^{q} \phi_i\left(Z_{it}\right) \mid Z_{kt}\right) = \phi_k\left(Z_{kt}\right).$$

5.4 semi-parametric partially additive model

This model can be presented as

$$Y_t = X_t^T \beta + \sum_{i=1}^{q} \phi_i\left(Z_{it}\right) + \varepsilon_t, t = 1, \cdots, n,$$

which $Y_t$is dependent variable, $X_t = \left(X_{1t}, \cdots, X_{pt}\right)^T$is a p-dimension concomitant variable vector, $\beta = \left(\beta_1, \cdots, \beta_p\right)^T$is a p-dimension parameter vector, $\phi_1(\cdot), \cdots, \phi_q(\cdot)$is a q-dimension smooth function vector and $Z_t = \left(Z_{1t}, \ldots, Z_{qt}\right)^T$is q-dimension observable vector.

We use the backwards fitting Method to get the solution. For all fixed β, the semi-parametric partially additive model can be written as

$$Y_t - X_t^T \beta = \phi_1\left(Z_{1t}\right) + \cdots + \phi_q\left(Z_{qt}\right) + \varepsilon_t, t = 1, 2, \cdots, n.$$

Assume

$$Y=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n \end{pmatrix}, X=\begin{pmatrix} x_1^T\\ {x_2}^T\\ \vdots\\ {x_n}^T \end{pmatrix}, \phi_i\left(Z_{it}\right)=\begin{pmatrix} \phi_i\left(Z_{i1}\right)\\ \phi_i\left(Z_{i2}\right)\\ \vdots\\ \phi_i\left(Z_{in}\right) \end{pmatrix}, D_{Z_i}^i=\begin{pmatrix} 1 & Z_{i1}-Z_i\\ 1 & Z_{i2}-Z_i\\ \vdots & \vdots\\ 1 & Z_{in}-Z_i \end{pmatrix},$$

local linear regression of the collaborator's variable vector for the smooth matrix

$$S_i=\begin{pmatrix} e_1^T\left(D_{Z_{i1}}^{iT}K_{Z_{i1}}D_{Z_{i1}}^i\right)^{-1}D_{Z_{i1}}^{iT}K_{Z_{i1}}\\ e_1^T\left(D_{Z_{i2}}^{iT}K_{Z_{i2}}D_{Z_{i2}}^i\right)^{-1}D_{Z_{l2}}^{iT}K_{Z_{l2}}\\ \vdots\\ e_1^T\left(D_{Z_{in}}^{iT}K_{Z_{in}}D_{Z_{in}}^i\right)^{-1}D_{Z_{in}}^{iT}K_{Z_{in}} \end{pmatrix}, i=1,2,\cdots,q,$$

which $e_1=(1,0)^T, K_{Z_i}=\operatorname{diag}\left\{K_{h_i}\left(Z_{i1}-Z_i\right), K_{h_i}\left(Z_{i2}-Z_i\right),\cdots,K_{h_i}\left(Z_{in}-Z_i\right)\right\}$, $K_{h_i}(\cdot)=\frac{K(\cdot/h_i)}{h_i}$, $K(\cdot)$ kernel function, h_i is bandwidth. Let $W_M=\sum_{i=1}^q W_i, W_i=E_iS^{-1}C$, Which E_i is a $n\times nq$matrix.

Opsomer& Ruppert(1999) build a estimation about backward fitting connecting the β and $M\left(Z_{it}\right)=\sum_{i=1}^q\phi_i\left(Z_{it}\right)$. That is

$$\hat{\beta}=\left[X^T\left(I_n-W_M\right)^T\left(I_n-W_M\right)X\right]^{-1}X^T\left(I_n-W_M\right)^T\left(I_n-W_M\right)Y,$$
$$\hat{M}\left(Z_{it}\right)=W_M(Y-X\hat{\beta}),$$

5.5 Variable coefficient model

We assume that Y is dependent variable, $X_1,X_2,\cdots,X_p$ and U are independent variable. If Y, $X_1,X_2,\cdots,X_p$ and U satisfy

$$Y=\alpha_1(U)X_1+\alpha_2(U)X_2+\cdots+\alpha_p(U)X_p+\varepsilon=\sum_{j=1}^p\alpha_j(U)X_j+\varepsilon,$$

then we call this Variable coefficient regression model.

Assuming $\left(Y_t,X_{t1},\cdots,X_{tp},U_t\right)(t=1,2,\cdots,n)$ be the independent and observation data, then the model turns to

$$Y_t=\sum_{j=1}^p\alpha_j\left(U_j\right)X_{tj}+\varepsilon_t, t=1,\cdots,n.$$

Assuming $\alpha_j(u)(j=1,2,\cdots,p)$ have constant derivation respect to u , $\mho$ is Value range for u. For all fixed $u_0\in\mho$, using Taylor formula, then we have u_0 in neighborhood

$$\alpha_j(u)\approx\alpha_j\left(u_0\right)+\alpha_j'\left(u_0\right), j=1,2,\cdots,p.$$

$K(t)$ is a given kernel function, $K_h(t)=K_h(t/h)/h$, then Variable coefficient model's local linear select $\alpha_j\left(u_0\right)\ \ \alpha_j'\left(u_0\right), j=1,2,\cdots,p$. That makes

$$\sum_{t=1}^n\left\{Y_t-\sum_{j=1}^q\left[\alpha_j\left(u_0\right)+\alpha_j'\left(u_0\right)\left(U_t-u_0\right)\right]X_{tj}\right\}^2K_h\left(U_t-u_0\right).$$

reaching the minimum.

Let

$$X(u_0)=\begin{bmatrix} X_{11} & \cdots & X_{1p} & X_{11}(U_1-u_0) & \cdots & X_{1p}(U_1-u_0) \\ X_{21} & \cdots & X_{2p} & X_{21}(U_2-u_0) & \cdots & X_{2p}(U_2-u_0) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n1} & \cdots & X_{np} & X_{n1}(U_n-u_0) & \cdots & X_{np}(U_n-u_0) \end{bmatrix}$$

$$Y=(Y_1,Y_2,\cdots,Y_n)^T,$$

$$W(u_0)=\operatorname{diag}(K_h(U_1-u_0),K_h(U_2-u_0),\cdots,K_h(U_n-u_0))$$

$$\alpha(u_0)=\left(\alpha_1(u_0),\alpha_2(u_0),\cdots,\alpha_p(u_0),\alpha_1'(u_0),\alpha_2{}'(u_0),\cdots,\alpha_p'(u_0)\right)^T.$$

then making the minimum in the above problems in the solution can be expressed as

$$\begin{aligned}\hat{\alpha}(u_0)&=\left(\hat{\alpha}_1(u_0),\hat{\alpha}_2(u_0),\cdots,\hat{\alpha}_p(u_0),\hat{\alpha}_1'(u_0),\hat{\alpha}_2'(u_0),\cdots,\hat{\alpha}_p'(u_0)\right)^T\\&=\left(X^T(u_0)W(u_0)X(u_0)\right)^{-1}X^T(u_0)W(u_0)Y.\end{aligned}$$

On the type, the variable coefficient vector $\alpha(u)=(\alpha_1(u),\alpha_2(u),\cdots,\alpha_p(u))^T$ in the u_0 place local linear estimation is

$$\begin{aligned}\hat{\alpha}(u_0)&=(\hat{\alpha}_1(u_0),\hat{\alpha}_2(u_0),\cdots,\hat{\alpha}_p(u_0))^T\\&=(I_p,0_p)\left(X^T(u_0)W(u_0)X(u_0)\right)^{-1}X^T(u_0)W(u_0)Y.\end{aligned}$$

5.6 A variable coefficient model and profile least-squares estimation method

In the variable coefficient model, if there is some constant coefficient, and the other part of the coefficient of function change with U, have the following half a variable coefficient model:

$$Y_t=X_t^T\beta+Z_t^T\alpha(U_t)+\varepsilon_t,$$

which Y_t dependent variable, $X_t=(X_{t1},\cdots,X_{tp})^T,Z_t=(Z_{t1},\cdots,Z_{tq})^T$ is interpretative vector, $\beta=(\beta_1,\cdots,\beta_p)^T$ is parametric vector and $\alpha(\cdot)=(\alpha_1(\cdot),\cdots,\alpha_q(\cdot))^T$ is q-dimension unknown function vector.

Assuming $\beta=(\beta_1,\beta_2,\cdots,\beta_p)^T$ is known, then we can transport the model to

$$\left(Y_i-X_i^T\beta\right)=\alpha_1(U_i)Z_{i1}+\cdots+\alpha_q(U_i)Z_{iq}+\varepsilon_i.$$

Rewriting in the matrix, we have

$$Y-X\beta=M+\varepsilon,$$

which $\mathrm{M}=\left[Z_1^T\alpha(U_1),Z_2^T\alpha(U_2),\cdots,Z_n^T\alpha(U_n)\right]^T$. And next we will use variable coefficient local linear estimation functionαU. Assuming the random variable have boundary support and $\{\alpha_j(\cdot),j=1,\cdots,q\}$ have constant second derivative, then in some neighborhood with U_0 we have

$$\alpha_j(U)\approx\alpha_j(U_0)+\alpha_j'(U_0)(U-U_0)\equiv\alpha_j+b_j(U-U_0),j=1,\cdots,q$$

which $\alpha_j'(U_0)=\left.\frac{\partial\alpha_j(U)}{\partial U}\right|_{U=U_0}$. By using the local weighted least squares method, we get

$$\hat{\beta}=\left(\tilde{X}^TW_U\tilde{X}\right)^{-1}\tilde{X}^TW_U\tilde{Y}$$

and

$$\hat{\alpha}(U)=(I_q0_{q\times q})\left(D_U^TW_UD_U\right)^{-1}D_U^TW_U(Y-X\hat{\beta}).$$

6 Application

6.1 estimation price of sailing boat

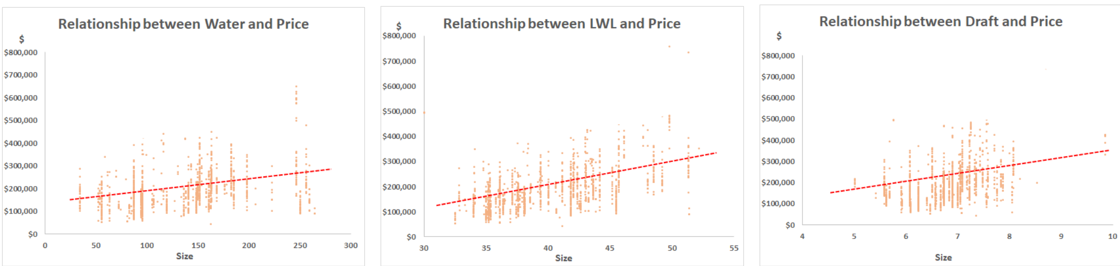
First, we explored the correlation coefficient matrix between variables when the significance level was 0.01, as shown in Figure x.

	Length	LWL	Beam	S.A. (repo...	Draft	Displaceme...	HP	Fuel	DWL	Water
Water	-0.281	-0.291	-0.232	-0.616	-0.346	0.069	-0.361	0.797	-0.501	1.000
DWL	0.828	0.840	0.885	0.858	0.670	0.710	0.851	0.033	1.000	-0.501
Fuel	0.225	0.185	0.330	-0.203	-0.141	0.631	0.216	1.000	0.033	0.797
HP	0.911	0.886	0.961	0.780	0.253	0.871	1.000	0.216	0.851	-0.361
Displaceme...	0.858	0.814	0.923	0.577	0.248	1.000	0.871	0.631	0.710	0.069
Draft	0.439	0.482	0.412	0.642	1.000	0.248	0.253	-0.141	0.670	-0.346
S.A. (repo...	0.892	0.907	0.818	1.000	0.642	0.577	0.780	-0.203	0.858	-0.616
Beam	0.965	0.948	1.000	0.818	0.412	0.923	0.961	0.330	0.885	-0.232
LWL	0.983	1.000	0.948	0.907	0.482	0.814	0.886	0.185	0.840	-0.291
Length	1.000	0.983	0.965	0.892	0.439	0.858	0.911	0.225	0.828	-0.281

Figure 4: Literature Review Framework

It is not difficult to find that the correlation coefficient of length, LWL, beam, HP, and DWL is high. S.A., the draft also presents a high correlation. It is known from consulting the data that these variables can be transformed into each other through mapping, so water, LWL, draft, fuel, and S.A. is taken as the representative.

From Figures 5 to 6 the scatter plot and fitting regression relationship between sailing prices and write variables are shown respectively. From figure 7 the regression



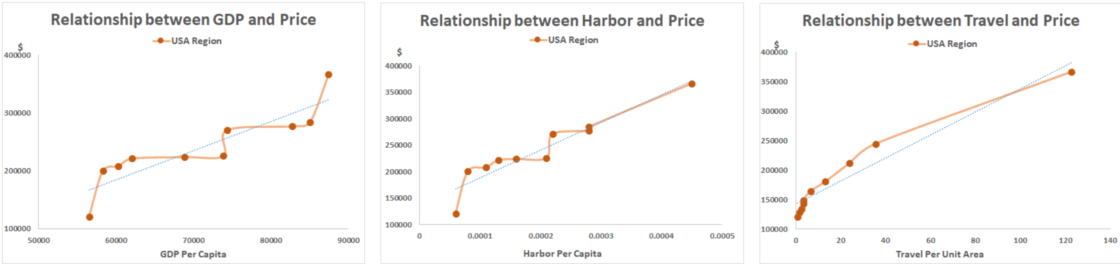
(a) Data Screening(left) (b) Data Screening(right) (c) Data Screening(right)

Figure 5: Data Screening

relationship between a scatter plot and fit can be seen, fuel, S.A., and temperature show a nonlinear relationship. According to the scatter diagram in figure xxx and figure xxx fitting relationship can be seen that the response variables and covariable $X_1, X_2, X_3, X_4, X_5, X_5$, a linear relationship, and covariable Z_1, Z_2, Z_3 relationship is nonlinear. Use half a variable coefficient model in this chapter

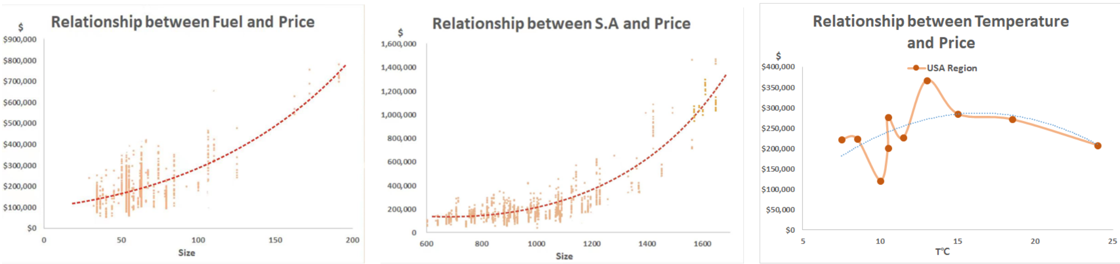
Then we can build the half-coefficient model

$$Y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + X_5\beta_5 + X_6\beta_6 + \alpha_1(U) + Z_1\alpha_2(U) + Z_2\alpha_3(U) + \varepsilon$$



(a) Data Screening(left) (b) Data Screening(right) (c) Data Screening(right)

Figure 6: Data Screening



(a) Data Screening(left) (b) Data Screening(right) (c) Data Screening(right)

Figure 7: Data Screening

We rewrite it by matrix and get

$$Y = X^T\beta + Z^T\alpha(U) + \varepsilon$$

which$X = (X_1, X_2, X_3, X_4, X_5, X_6)^T$, $Z = (I, Z_1, Z_2)^T$, ε is random error, $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = \sigma^2$.

For all β, we then rewrite it to

$$Y^* = Y - \sum_{j=1}^{6} \beta_j X_j = Z^T\alpha(U) + \varepsilon.$$

Thus we can get the optimal solution is

$$\begin{aligned}
&\left(\hat{\alpha}_1(U), \hat{\alpha}_2(U), \hat{\alpha}_3(U), \hat{\alpha}_4(U), \hat{\alpha}_5(U)h, \hat{\alpha}_6(U)h, \hat{b}_1(U), h\hat{b}_2(U), h\hat{b}_3(U)\right)^T \\
&= \left(D_U^T W_U D_U\right)^{-1} D_U^T W_U (Y - X\beta),
\end{aligned}$$

Therefore, the least squares of the section in Variable coefficient function$\alpha(U)$ is

$$\hat{\alpha}(U) = (\hat{\alpha}_1(U), \hat{\alpha}_2(U), \hat{\alpha}_3(U), \hat{\alpha}_4(U), \hat{\alpha}_5(U), hat\alpha_6(U))^T = (I_6, 0_6)^T \left(D_U^T W_U D_U\right)^{-1} D_U^T W_U (Y - X\beta).$$

In this paragraph, we use the Gaussian kernel and bindwidth$h = 1.06283$ to get the minimum of the least squares. We select 80% data for training and 20% for testing. Then we get the parametric evolution $\hat{\beta} = \left(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5\right)^T$ and Variable coefficient function's fitting curve including $\alpha_1(U)$ $\alpha_2(U)\alpha_3(U), \alpha_4(U), \alpha_5(U), \alpha_6(U)$, which is $\hat{\alpha}_1(U), \hat{\alpha}_2(U)$ and $\hat{\alpha}_3(U)$ respectively.

Finally, we get the price and parameter relationship is

$$\hat{Y}_t = 0.88496715X_{1t} + 0.07983221X_{2t} + 0.15883855X_{3t} + 0.921148X_{4t} + 0.89238497X_{5t} + 0.89291231X_{6t} + \hat{\alpha}_1(U_t) + \hat{\alpha}_2(U_t)Z_{1t} + \hat{\alpha}_3(U_t)Z_{2t}(*)$$

Meanwhile, we get the RMSE(Root Mean Squared Error)

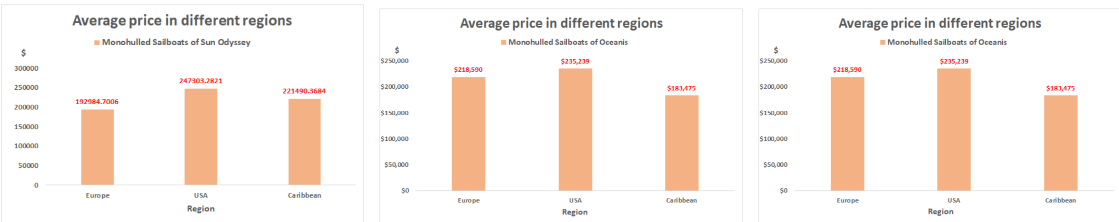
$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} = 0.233647, n = 147$$

so we have the error interval $[x - 1364, x + 1364]$.

6.2 Exploring regional influence on prices

From (*) we get the price and parameter's relationship, so we only need to control the irrelevant variable and change the relevant variable to see the differences.

According to our model, the impact of region on ship price is reflected in the following four indicators: the impact of per capita GDP on ship price, the impact of average annual temperature on ship price, the impact of the port number, and the ratio of population in the corresponding area, and the impact of total tourism and the ratio of population in the corresponding area. Sailboats are divided into three different types, the application model is analyzed, and the results shown in the figure below are obtained Figure 8. We will analyze these charts in **7.2**



(a) Oceanis (b) Cruiser (c) Sun odyssey

Figure 8: original

7 Results

7.1 Problem 1

Through the analysis of question 1, we get the influence of parameters on the price of sailboats.

- The results show that LWL, water, draft, Habor, travel, and GDP have a positive linear effect on the sailing price, while S.A., Fuel, and temperature have a positive correlation with sailing price.
  This means that if other conditions remain the same, the higher the draft, the higher the price of sailboats generally. It is not difficult to figure out the draft and

size of sailboats, and the larger the size of sailboats is more high-end, so the price of sailboats is higher.
Similarly, the higher the water storage capacity, the higher the price of sailing boats, because the water storage capacity represents the sailing distance of sailing boats, and whether the resources are abundant during the voyage, so it is highly valued by many people.

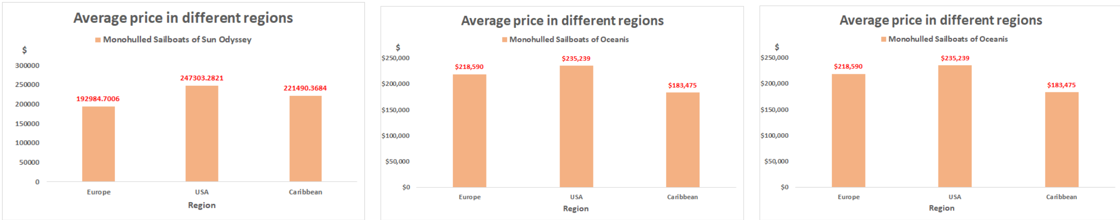
- It is interesting to note that prices rise and then fall as temperatures rise. We analyze this phenomenon because high temperatures and cold temperatures are not fitted to not conducive to sailing.

7.2 Problem 2

Our model considers that the influence of region on ship prices is reflected in the following four indicators: the influence of GDP per capita on ship prices, the influence of average annual temperature on ship prices, the influence of the ratio of the number of ports to the population of the corresponding region on ship prices, and the influence of the ratio of total tourism to the population of the corresponding region on ship prices. The reasons for their respective effects are as follows:

1. GDP per capita
GDP per capita is an important indicator used to measure the economic strength of a country. The reason we choose this indicator is that higher GDP per capita means higher national consumption power. In turn, higher national consumption power means higher demand for imported goods, and such demand can drive the import/export industry, thus affecting ship prices.
2. Average annual temperature
As a basic and important climate indicator, the annual average temperature can better reflect the local climate conditions. In particular, import and export commodities such as food, machinery, and biology also have high requirements on shipping temperature, so we decided to use temperature as an indicator to show the climate and thus consider the influence of climate on ship prices.
3. The ratio of the number of ports to the population of the corresponding area
Port density is the most direct indicator of a region's maritime strength, however, in the process of considering this indicator, we cannot simply consider the number of local ports but should consider "density", otherwise it will not be able to correctly portray the degree of maritime development.
4. The ratio of total tourism to the population of the corresponding area
Tourism is the best indicator of the size of the maritime market in a region, however, similar to the port perspective above, instead of considering the total amount of tourism, we choose to use the unitization of the population of the corresponding region to reflect the intensity and realization of the local desire for tourism.

By analyzing the data from different regions for different models of prices and from different regions for different versions of the same model, we find that the regional impact on the ship variant is not consistent across different models and different versions of the same model. Firstly, analyzing the price impact of different regions on different models, we selected the Sun Odyssey series, Cruiser series, and Oceanis series with more sufficient amount of data, which can be found in the following visualized table:



(a) Sun Odyssey (b) Oceanis (c) Cruiser

Figure 9: Comparison

USA is always number one in the average price for boats, but the margin of exceeding other regions is different, and Europe and the Caribbean are not simply one higher and one lower for different types of boats, so it can be concluded that the impact of different regions on boat model variants is inconsistent. Continuing with the Sun Odyssey series, Cruiser series, and Oceanis series study, after compiling the data, we found that the USA ranked first in all three regions in terms of sailboat prices for each sailboat category. Therefore, we have singled out the USA to explain the reasons for its high sailboat prices and related phenomena. For Europe and the Caribbean, we will discuss them under each sailboat category. USA: The USA is a high-income, high-expense country. At the same time, due to the weakness of the manufacturing industry and the high cost of manufacturing sailboats, they generally sell at higher prices. Also given the regional culture of Americans, etc., more luxurious and wider sailboats are more popular and have higher pricing in the USA.

SUN ODYSSEY The demand for sailing in Europe and the Caribbean is different,

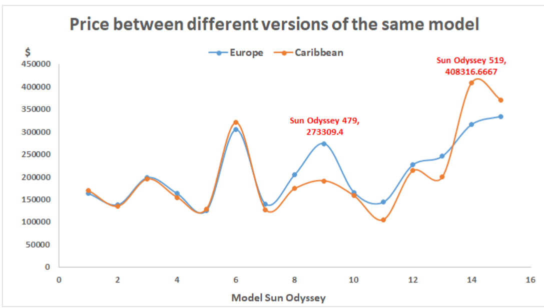


Figure 10: Sun Odyssey-Comparison

and therefore there is some difference in the price of sailing between the two. Overall, the average price in the Caribbean is slightly higher than in Europe, which is related to Sun Odyssey's focus on the theme of boating impressions and experiences Sailboat buyers in the Caribbean are more likely to purchase sailing boats for tourism purposes, and buying for high enjoyment fits right in with Sun Odyssey's design philosophy. European boat buyers, because of their focus on the economy, also favor smaller boats, for example, the Sun Odyssey 479 is priced much higher in Europe than in the Caribbean. And in the Caribbean, the Sun Odyssey 519 sells for a much higher price than in Europe.

OCEANIS

The Oceanis' design is characterized by simple lines, good balance, low operating difficulty, and large space. The Oceanis 38.1 and Oceanis 48 are priced much higher in the Caribbean than in Europe, while the Oceanis 41.1, Oceanis 46, and Oceanis 54 are priced the other way around. We believe the reason for this difference is that Oceanis 38.1 and Oceanis 48 have fewer and slimmer windows in the interior cabins to protect

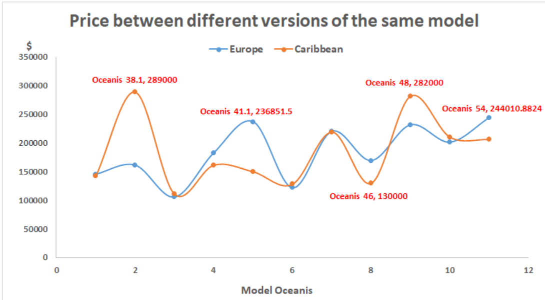


Figure 11: Oceanis-Comparison

the privacy of visitors and better suit the desires of Caribbean visitors, while people in Europe have higher requirements for light because they are located at mid to high latitudes and prefer more sunlight, while Oceanis 41.1, Oceanis 46 and Oceanis 54 have higher requirements for light. Oceanis 41.1, Oceanis 46, and Oceanis 54 have wide skylights, allowing passengers to enjoy sunbathing in the interior cabin, so the price in Europe is higher.
CRUISER
The Cruiser series of boats are known for its wide space and is more casual. We can

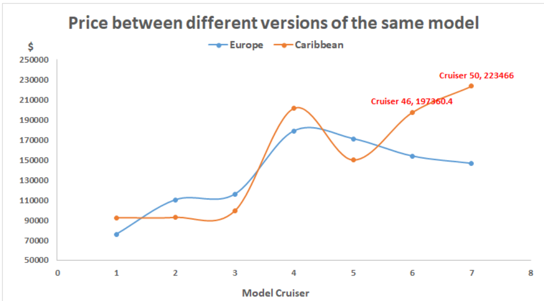


Figure 12: Cruiser-Comparison

find that for the lower length Cruiser, the prices in the Caribbean are lower than in Europe, because the smaller but wider boats fit the economic mentality of Europeans when buying a boat, and after the Cruiser 41, the prices in Europe are slowly decreasing due to the slight redundancy of the larger boats for private use. However, as far as tourism is concerned, a wider boat means a more comfortable travel experience with the same amount of people. Therefore, it can be observed that the price of large cruisers exceeds the price in Europe in the Caribbean region.

7.3 Problem 3

The reason why our model of region-to-ship prices is well applicable to Hong Kong is that the four indicators we have chosen are all highly universal and can be applied to most regions of the world. At the same time, these indicators can also highlight the special characteristics of Hong Kong as a city. sailing boats' prices. Our work mainly includes the following:

1. Real GDP per capita
   As a city not rich in natural resources but coastal and well-connected, Hong Kong, as one of the international financial centers, has a huge flow of high-net-worth individuals in the world. The demand for the maritime service industry is high and the

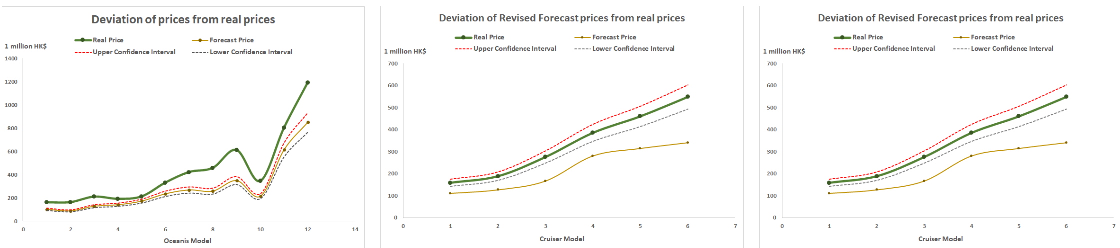
development requirement is high, which can pull the shipping market. Therefore, the real GDP per capita can be used as an indicator to portray the influence of the Hong Kong area on ship prices.

2. Average annual temperature
Hong Kong is located at low and middle latitudes and its climate type is subtropical monsoon climate. As a result, its harbor does not freeze all year round, so Hong Kong's harbor is suitable for maritime recreational activities all year round. Therefore temperature, as an important indicator affecting Hong Kong's maritime tertiary industry, is included in our indicator of the impact of the Hong Kong region on ship prices.

3. The impact of the ratio of the number of ports to the population of the corresponding region on ship prices
Various factors such as the well-developed tourism in the Hong Kong region and the bridge connecting China with the world have put high demands on Hong Kong's port service industry. Therefore, we think it is reasonable to use the ratio of the number of ports to the population of the corresponding region as an indicator to portray the influence of the Hong Kong region on ship prices.

4. The influence of the ratio of total tourism to the population of the corresponding area on ship prices
Hong Kong has an extremely developed maritime tertiary industry, which greatly boosts the demand for the sailing ship market. A large part of the total tourism in Hong Kong comes from the maritime tertiary industry in Hong Kong. Therefore, we think it is reasonable to use the ratio of total tourism to the population of the corresponding area as an indicator of the influence of the Hong Kong area on ship prices.

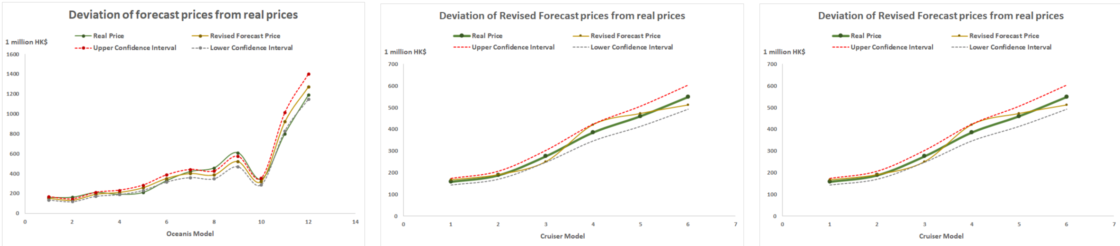
We use the model from the previous question to substitute data for two subsets of sailing boats in the Hong Kong region, and the results are as follows:
Monohulled Sailboats:



(a) Oceanis (b) Cruiser (c) Sun odyssey

Figure 13: original

When we adjust the correction factor of the monohulled sailboat model to (for times the original), the results are as follows: It can be seen that the model predicts better when the correction factor of a monohulled sailboat is adjusted to (for times of the original). According to the discussion, the reason for our belief is that Hong Kong has a small land area and an underdeveloped manufacturing industry, so if we need to purchase sailboats we must import them. However, unlike Europe and the United States, sailing ships in Europe and the United States only need to cross the Atlantic Ocean and have low transportation costs, while imports in Hong Kong need to cross



(a) Oceanis (b) Cruiser (c) Sun odyssey

Figure 14: corrected

the Pacific Ocean and have higher transportation costs. Also, land prices are higher in the Hong Kong region, and the cost of anchorage required for the preservation of sailing ships is higher. Combining the above-mentioned and other factors, we believe that the cost of sailing in the Hong Kong region is higher than the remaining three regions, and thus has a higher correction factor. Catamarans: When we adjust the correction factor for the Catamarans model to (for times the original), the results are as follows: It can be seen that the model does not predict well when the correction factor of the

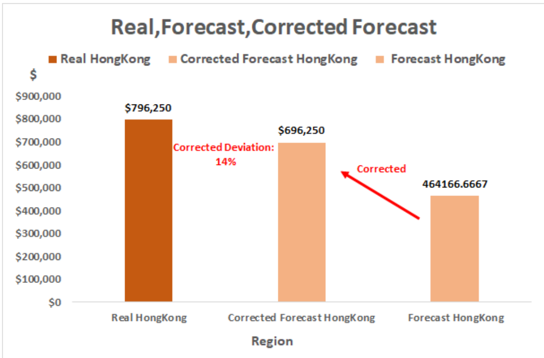


Figure 15: Catamarans

monohulled sailboat is adjusted to (for times the original). After discussion, we believe that this result is due to the difference in sailboat types. monohulled sailboats are more focused on the driving experience and their audiences have relatively stable groups as well as cultures; Catamarans are better in terms of leisure and comfort, and their price fluctuations are more unstable compared to others because they are becoming more popular, and the model does not set an indicator for the influence of the popularity factor. Also, due to insufficient data given, the correction coefficients in the model are more influenced by chance factors, resulting in poor prediction of the model. As far as the price prediction results of our model for Hong Kong are concerned, the effect of Monohulled Sailboats is better. The reason for this is that the effect of the Hong Kong region on Monohulled Sailboats prices is found to be better similar to the remaining three regions. In the case of Catamarans, however, our price prediction results are poorer, suggesting that the Hong Kong region factor has a different degree of influence on the price of both sailboats. We believe that the reason for this is that Catamarans are slowly becoming trendy, which is why the actual prices are more expensive than predicted.

7.4 Problem 4

1. Sun Odyssey is a line of ships, the DS series has higher pricing. After researching, DS stands for "deck saloon", which is characterized by the fact that the saloon is located above the deck, not below it, as is the case with other boats, and the DS series has rooms such as bedrooms below the deck. This design separates the entertainment area from the private area and has a high price in all areas.

2. Cruiser, known for its spaciousness, is priced surprisingly low in Europe, and interestingly, the boat's manufacturing company, Bavaria, is a French company. Products inside Europe are priced higher and more popular outside Europe.

3.On March 10th of this year, German media revealed that a CruIser 50 was rented by suspicious people and was presumed to be used for the Nord Stream pipeline terror attack.

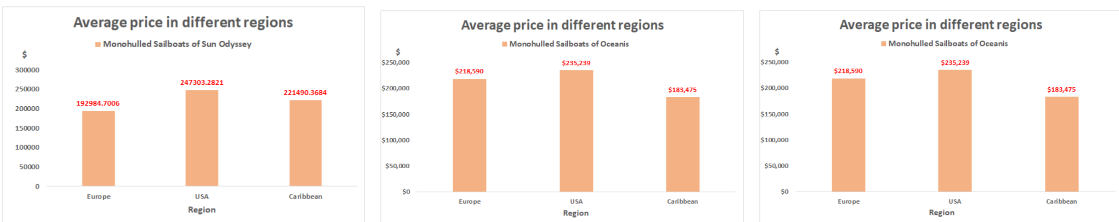
Figure 16: terror attack

7.5 Problem 5

By studying and comparing the sailing market in Europe, the United States, and the Caribbean, we have come to the following conclusions. First of all, due to the high tariffs, high income, and high expenditure of national living standards in the US region, tourism, and maritime culture go hand in hand, sailing is popularly spread as an aristocratic sport, and the prices of sailing boats in the US are generally high, except for one case where the price of Catamarans is lower than in Europe and the Caribbean, and the prices of all other models are higher than in Europe and the Caribbean. From the survey results, Americans prefer larger sailing boats, so the pricing of larger sailing boats in the U.S. market is also a bit higher. Europe, as a region where marine culture is prevalent, has a higher popularity of sailing. On average, one out of 6-7 people in countries like Norway, Finland, and Switzerland owns a yacht. Meanwhile, according to the British Marine Federation, nearly a quarter of British adults participate in marine sports. Therefore, the British yacht market is more practical, family-oriented, and safe, and boats under 40 feet are more popular. In particular, British people in the middle and high latitudes prefer sailing boats with good lighting. In these categories, the market price in Europe is higher than in the Caribbean. Caribbean sailing buyers are willing to pay higher prices for a more comfortable, luxurious experience due to the growing needs of the maritime tourism industry. the market is more aggressive above 40 feet. Unlike Europe, the Caribbean, located in the lower and middle dimensions, is not willing to pay higher prices for a boat with an interior cabin that is exposed to the sun. Instead, they are more willing to pay for a ship with more privacy. After

Figure 17: sun bath

Figure 18: privacy



(a) Luxury boat (b) Utility boat (c) Wide boat

Figure 19: Comparison

referring to the data from the three regions mentioned above and adjusting and processing them, a reference price for the total category of sailing boats in Hong Kong was derived. However, we believe that it is not reasonable to compare Hong Kong with the above mentioned regions to determine the sailing market in Hong Kong. You can make adjustments to the reference prices given in our model that are more in line with the reality of Hong Kong. Due to the developed tourism industry, we think that Hong Kong region can refer to the Caribbean region in terms of boat size and can set higher prices for sailing boats with a higher sense of experience and luxury. Since maritime culture is not popular in the Hong Kong region, we do not recommend borrowing pricing rules from the European region, and we do not believe that the Hong Kong sailing market should set high prices for practical, family-oriented boats. Finally, considering that Hong Kong's economic situation is closer to that of the United States, etc., we believe that Hong Kong could set higher prices for larger boats.

(a) (b) (c)

Figure 20: Prediction V.S. Factor

7.6 Advantage and Disadvantage

Model advantages and disadvantages. 1. model I does not show well the effect of nonlinear factors on price, and some factors affect price in a complex nonlinear way making the fit of the local linear estimation not as good as expected. 2. The selection of factors is not comprehensive enough, and at the same time, due to the presence of a large number of missing data, only methods such as mean padding can be used, making a large amount of data simply assimilated to the mean, losing some of the information that would otherwise be available. 3. Our model can only roughly predict the price of a particular type of vessel based on the average price of that type of vessel in the market, and because the popularity of each type is not consistent in different geographic regions, the model's price estimates for particular vessels will vary widely. 4. The modified model fits better for monohulled sailboats in Hong Kong, but there is a larger variation in the fit for Catamarans. This is due to the small amount of data for Catamarans in Hong Kong and the fact that the demand for Catamarans in Hong Kong has only become popular in recent years, so there will be a higher premium, but it cannot be reflected in the model. 5. The correction coefficients obtained from data for some countries or regions are not necessarily applicable to other regions, and the correction coefficient needs to be recalculated when estimating boat prices for a new region.

References

[1] GLOBAL FOREST WHATCH OF AUSTRALIA https://www.globalforestwatch.org/topics/fires/?topic=fires#footer

[2] HU Teng, LIU Zhanjun, LIU Yang, et al. 3D reconnaissance path planning of multiple UAVs. Journal of Systems Engineering and Electronics, 2019, 41(7): 1551-1559.

[3] BASBOUS B. 2D UAV path planning with radar threatening areas using simulated an-nealing algorithm for event detection. The 2018 International Conference on Artificial Intelligence and Data Processing. Malatya, Turkey: IEEE,2018: 1-7.

[4] WANG W F, WU Y C, ZHANG X. Research of the unit decomposing traversal method based on grid method of the mobile robot. Techniques of Automation and Applications, 2013, 32(11): 34-38.

[5] XU Jian, ZHOU Deyun, HUANG He. Multi UAV path planning based on improved ge-netic algorithm. Aeronautical Computing Technique, 2009, 39(4): 43-46.

[6] Zhang W Y,Lee S Y,Song X Y.Local polynomial fitting in semi-varying coefficient model.Journal of Multivariate Analysis,2002,82:166-188.

[7] Xia Y C,Zhang W Y,Tong H.Efficient estimation for semi-varying coefficient models.Biometrika,2004,91(3):661-681.

[8] Fan J Q,Huang T.Profile likelihood inferences on semi-parametric varying-coefficient partially linear models.Bernoulli,2005,11:1031-1057.