Research on the prediction of the development trend of the pet industry based on multi-dimensional analysis

Bohan Zeng[1], Yanqi Jiang[2]

[1]School of Automation, Nanjing University of Science and Technology, [2]School of Computer Science and Engineering, Nanjing University of Science and Technology

Abstract

This paper focuses on the development of the pet industry in a multi-dimensional analysis, aiming to promote the sustainable development of the domestic pet industry through data collection, model building and question answering.

Question 1 uses the relevant data of the past five years, including per capita disposable income, the number of marriage registrations, the number of people aged 65 and above, the urbanization rate, etc., and reveals the change trend and influencing factors of the number of pet cats and dogs and the output value of the pet industry through visual analysis and correlation analysis. The grey prediction model GM (1,1) was used to predict the number of pet cats and dogs and the output value of the pet industry in the next three years.

Question 2 analyzes the pet cat and dog population data from China, the United States, France, and Germany over the past eleven years. Based on the different trend characteristics of the data, an appropriate exponential smoothing model (single, double, or triple exponential smoothing model) will be selected to predict the pet cat and dog quantities in these four countries for the next three years. And these predicted data reflect the global market demand for pet food.

Question 3 analyzes the development of China's pet food industry based on the global pet demand trend and the actual situation in China. A multiple linear regression model was established, and the significant variables were screened through the stepwise regression method, and the key factors affecting the production and export of pet food were finally determined, and the production and export volume of pet food in the next three years was predicted.

Question 4 builds upon the previous questions to explore the potential impact of the foreign economic policies of European and American countries on the Chinese pet food industry. A corresponding model has been established for analysis, proposing feasible strategies for the sustainable development of the Chinese pet food industry.

Keywords

Grey prediction model GM (1,1), Exponential smoothing model, Multiple linear regression model, Stepwise regression method, Pet food industry

Introduction

Question Background

The rapid economic development has resulted in an increase in per capita income and consumption capacity. Simultaneously, changes in the family structure, such as aging population

and declining birth rates, have given rise to people's emotional demand for companionship. "Pet Companionship" has fulfilled the sense of belonging and happiness of countless individuals. Since the early 20th century, the pet industry has gradually gained traction as a nascent industry globally, featuring markets in areas such as food, clinics, supplies, and care. It has also matured in all aspects of the supply chain. This highlights the social and economic value of the pet industry's growth. Therefore, this article aims to promote the development of the domestic pet industry through a multidimensional analysis of global and domestic trends, market demands, and economic environments.

Question Restatement

Question1：Based on the data in Attachment 1 and other data collected by your team, analyze the development of China's pet industry by pet type over the past five years. It also further explores the factors driving the development of China's pet industry in order to construct a suitable mathematical model to predict the development trend of China's pet industry in the next three years.

Question2：In recent years, the overseas pet market, especially in Europe and the United States, has shown a booming trend. Please use the data provided in Attachment 2 and the additional information gathered by our team to provide an in-depth analysis of the growth trends of the global pet industry by different pet species. On this basis, a scientific and reasonable mathematical model is constructed to predict the changes in global pet food demand in the next three years.

Question3：Referring to the production and export data of China's pet food in Attachment 3, please analyze the growth trend of China's pet food industry and predict how China's pet food production and export will change in the next three years, regardless of economic policy changes, based on the global pet food market demand trend and China's own development.

Question4：China's pet food industry is bound to be affected by new foreign trade policies in Europe and the United States, such as tariff adjustments. In order to quantify this impact, an appropriate mathematical model needs to be constructed, incorporating the data provided in the Attachment, self-collected supplementary data, and the calculations of the aforementioned problems. Based on the results of the model analysis, please plan an effective strategy to promote the sustainable development of China's pet food industry.

Assumption and Justification

It is assumed that the relationship between the variables (e.g., per capita income and the number of pets, the number of pets and the output value of the pet industry, etc.) will remain stable and will not change abruptly due to external factors.

It is assumed that China's economic policies (especially those related to the pet industry) and the new foreign trade policies of Europe and the United States will not change significantly during the forecast period, so as to focus on the analysis of market factors when making forecasts for the next three years.

It is assumed that all data collected, including those in Annex 1 and those collected through other sources, is accurate and complete, with no omissions or errors.

It is assumed that the demand of the pet market will remain continuous in the next three years,

and there will be no sharp decline in market demand due to emergencies (such as epidemics, natural disasters, etc.).

When analyzing the impact of the new foreign trade policies in Europe and the United States on China's pet food industry, it is assumed that the impact of these policies is quantifiable and can be accurately assessed through models.

It is assumed that there is no lag in the market in question 4.

Description of the symbol

Due to the large number of variables, the symbols of the variables will be defined below, and only the representative symbols of the variables that are frequently used and whose variable names are too long are given here.

Symbols	Description
PCDI	personal disposable income
MR	the number of registered marriages
P>=65	the number of individuals aged 25 and above
UR	the urbanization rate
PIV	the output value of the pet industry

Problem Analysis

For Question 1

To undertake an in-depth analysis of the development of the pet industry in China, aside from the data offered in Appendix 1, we have also gathered data such as personal disposable income (PCDI), the number of registered marriages (MR), the number of individuals aged 25 and above (P>=65), the urbanization rate (UR), and the output value of the pet industry (PIV) from channels such as the National Bureau of Statistics and pet industry associations in the past five years. And separate visual analyses of the relevant data for cats and dogs will be conducted to disclose the development trends of the pet industry in the past five years. After grasping the trends of the pet industry through visual analysis, we need to undertake further correlation analyses on the factors influencing the development of the pet industry in order to explore the relationships among the data. Finally, the most appropriate model is chosen based on the outcomes of data visualization and correlation analysis to forecast the development of the pet industry in China in the next three years.

For Question 2

In addition to the number of pet cats and dogs in China, the United States, France, and Germany from 2019 to 2023 given in Attachment 2, we also collected data for each country from 2013 to 2018. We analyze the development of the global pet industry by pet species, represented by these four countries. Based on the data of the number of cats and dogs in these 11 years, we can find a suitable method - exponential smoothing method to predict the number of cats and dogs in the next three years, and then multiply the predicted number with the amount of cat and dog food consumed

by each cat and dog in a year to calculate the demand of the pet industry in the next three years.

For Question 3

Question 3 requires us to predict the production and export volume of pet food in the next three years, considering that these two are determined by a number of factors, and these factors have been specifically analyzed by the first and second questions, for example, in the first question, we get that the number of pets has a strong linear correlation with the output value of the pet industry, so we consider using a multivariate stepwise linear regression model, and dynamically remove variables during the model establishment process to optimize the model performance. This flexibility allows the model to more accurately reflect the complex relationships in the actual data, so as to more comprehensively capture the changes in the total production and export value of the pet industry. By solving the coefficients of the linear regression equation and bringing in the relevant predictions from the first two questions, the forecast of pet food production and export in the next three years can be completed

For Question 4

Question 4 requires us to predict the future development of China's pet food industry by quantifying the impact of the new foreign economic policies of European and American countries and incorporating this impact into the established model. In order to determine the extent of the impact of foreign economic policies on the pet food industry, we collected tariff, global average food prices, pet industry expenses, and revenue share of the US pet market, which can reflect foreign economic policies to a certain extent. In order to quantitatively analyze the impact of these data on China's pet food industry, we use a stepwise linear regression equation to screen and determine the impact of these data on China's pet food industry based on the third question. Finally, based on the regression equation and how the data we collect will change in the future, we can develop a feasible strategy that can make China's pet food industry sustainable.

Model Establishment and Problem Solving for Question1

Visualization Analysis

Firstly, conduct a visualization analysis of the quantity of cats and dogs.

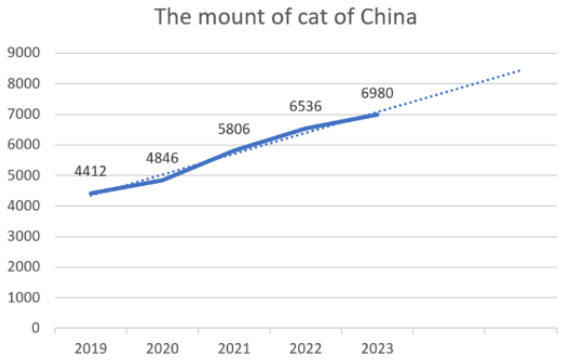


Figure 1 The mount of China

Figure 2 The mount of dog

It can be discerned that during the recent five years: for pet cats, their quantity has been escalating year by year, and the ascending trend is largely in conformity with a linear trend; for pet dogs, the change in their quantity has fluctuated but is generally on a downward trend. This initially indicates that due to the influence of multiple factors, people in China now have a greater preference for keeping cats.

Analysis of Regulatory Role

Conduct descriptive statistical analyses on the data in Appendix 1 and the data we have collected, obtaining statistical parameters such as the minimum value, maximum value, mean, standard deviation, skewness, kurtosis, and standard error.

Table 1 Descriptive statistical table

Statistics	PCDI	MR	>=65	UR	Cat	Dog	PIV
N	5	5	5	5	5	5	5
Minimum	30733.0	681.8	17767.0	62.7	4412.0	5119.0	2212.0
Maximum	39218.0	1009.11	21676.0	66.16	6980.0	5503.0	5928.0
Average	34830.2	837.7	19908.2	64.5	5716.0	5289.6	4194.2
Deviation	3439	129.5	1548.2	1.31	1088.0	167.1	1548.0
Skewness	0.061	0.271	-0.412	-0.337	-0.098	0.504	-0.411
Kurtosis	-1.578	-1.174	-0.908	-0.209	-2.230	-2.411	-2.022

Through descriptive statistical analysis, it can be known that the number of pet cats is greater than that of dogs, indicating that people are more willing to keep cats. The standard deviation of per capita disposable income is the largest, indicating its greater fluctuation. The standard deviation of urbanization rate is the smallest, indicating its relatively smaller fluctuation and more stable data volume. Among the above indicators, the skewness value of Dog is the largest and greater than 0, indicating that the data distribution is relatively shifted to the right.

Correlation analysis

Draw scatter plots for each pair of all variables to observe the relationships among them.

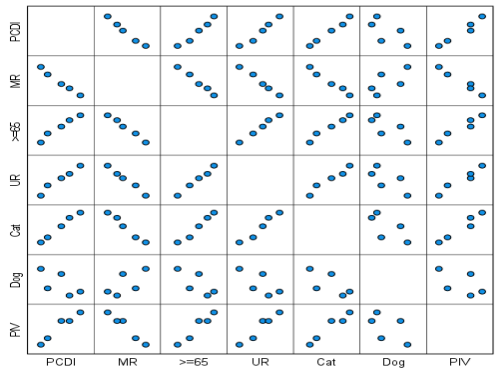


Figure 3 Scatter plot

It can be observed that the majority of the variables exhibit a linear relationship, so we consider employing the Pearson correlation coefficient to analyze the relationships among the data.

Pearson correlation analysis

The first step is that the Pearson correlation coefficient demands that the sample data satisfy the requirements of normal distribution. Hence, we conduct a normality test on the data first, using the SPSS software. The obtained results are as follows:

	Shapiro-Wilk		
	Statistic	df	Sig.
PCDI	.969	5	.871
MR	.977	5	.919
>=65	.977	5	.915
UR	.992	5	.985
Cat	.944	5	.697
Dog	.896	5	.390
PIV	.908	5	.455

Figure 4 Tests of Normality

We set the null hypothesis H0 as the data not conforming to the normal distribution. It was discovered that the p-value of the Shapiro - Wilk (W test) is greater than the 0.05 level (Shapiro - Wilk = 0.969, 0.977, $P = 0.871, 0.919 > 0.05$). Thus, the null hypothesis was rejected, and it was considered that the sample met the requirements of the normal distribution.

The second step is that the Pearson correlation coefficient is employed to gauge whether two data sets lie on a single line, it means to measure the linear relationship between interval variables. When both variables are normally distributed continuous variables and a linear relationship exists between them, the Pearson correlation coefficient is frequently selected to describe the degree of correlation between the two. The specific calculation formula is presented as follows:

$$r = \frac{\Sigma_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})} \cdot \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})}}$$

In this formula, r represents the correlation coefficient, n is sample size, and Xi and Yi respectively represent the two attribute values of the i-th sample.

When $r = 1$, X and Y are referred to as being completely correlated, at which point there exists a linear functional relationship between them. When $r > 0.8$, it is termed highly correlated; when $r < 0.3$, it is regarded as lowly correlated, and in other instances, it is considered moderately correlated.

We employed SPSS to compute the Pearson correlation coefficients among the variables and constructed a heat map.

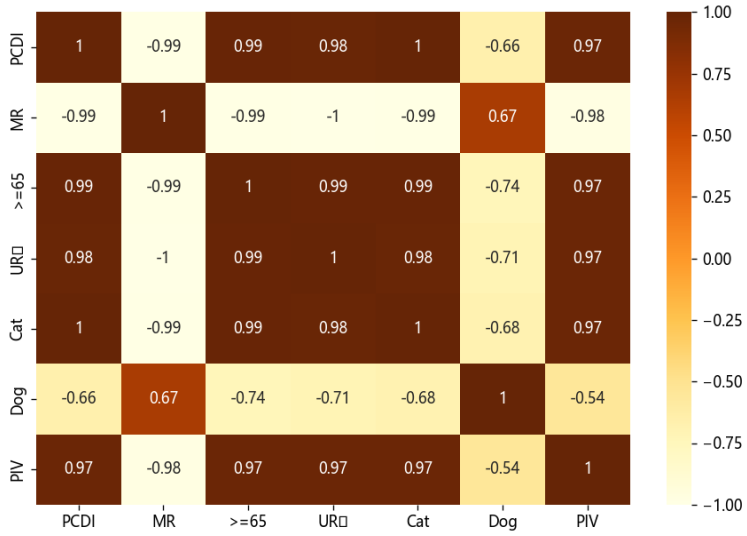


Figure 5 Person Heat Map

From the above figure, it can be observed that Per capita disposable income, Seniors over 65 years old, and Urbanization rate all have an extremely strong positive correlation with Cat and Pet industry value, while Marriage registration shows a negative correlation with Cat and Pet industry value. Dog is completely opposite and has only a moderate correlation.

In the third step, a hypothesis test is performed on the obtained Pearson correlation coefficient, and we propose the null hypothesis H_0 and the alternative hypothesis H_1. For a Pearson correlation coefficient r calculated by us, to test whether it is significantly different from 0, the null hypothesis and the alternative hypothesis are set as follows: $H_0: r = 0, H_1: r \neq 0$.

Under the condition that the null hypothesis holds, a statistic conforming to a certain distribution is constructed by using the quantity to be tested.

For Pearson's correlation coefficient r, under certain conditions, we can construct a statistic:

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

It can be proved that t follows a t-distribution with $n - 2$ degrees of freedom.

After calculating t, the p-value is computed by referring to the t-distribution table. For a two-sided test, the calculated t-value needs to be compared with the critical value of the t-distribution, and then the corresponding p-value is found. In a two-sided test, the probability of the rejection region needs to be multiplied by 2. We employed MATLAB programming to calculate the P-values among various variables and produced the following table:

Table 2 P-Values

	PCDI	**MR**	**>=65**	**UR**	**Cat**	**Dog**	**PIV**
PCDI	1.0000	0.0005***	0.0020***	0.0024***	0.0004***	0.2299*	0.0049***
MR	0.0005***	1.0000	0.0004***	0.0004***	0.0012***	0.2161*	0.0023***
>=65	0.0020***	0.0004	1.0000***	0.0005***	0.0018***	0.1542*	0.0075***
UR	0.0024***	0.0004***	0.0005***	1.0000	0.0045***	0.1791*	0.0063***
Cat	0.0004***	0.0012***	0.0018***	0.0045***	1.0000	0.2105*	0.0064***
Dog	0.2299*	0.2161*	0.1542*	0.1791*	0.2105*	1.0000	0.3463*
PIV	0.0049***	0.0023***	0.0075***	0.0063***	0.0064***	0.3463*	1.0000

***. P Value is significant at the 0.01 level.

**. P Value is significant at the 0.1 level.

*. P Value is significant at the 0.5 level.

Evidently, all variables except for the quantity of dogs can reject the null hypothesis H0 at a 99% confidence level, and all variables can reject the null hypothesis H0 at a 95% confidence level. Hence, all r are significantly different from 0, and there is a strong linear correlation between these variables.

Model Establishment: Grey prediction model GM(1,1)

Due to the very limited amount of data, establishing a multiple linear regression model with the number of cats or dogs as the dependent variable and other variables as the independent variables would generate considerable errors and have no predictive capability for the development of the pet industry in the next three years. However, the grey prediction method is mainly applicable to circumstances of "small data and poor information", which does not require large sample sizes or specific distributions of data. Thus, in this case, the grey prediction model is adopted to predict the data for the next three years.

The modeling procedures are presented as follows with the prediction of the quantity of pet cats as an instance:

Data Preparation: Denote $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$ as the data of the number of pet cats in China provided in Problem One (in ten thousands), where n represents the number of data points.

Data accumulation generation: The original data sequence is subjected to a one-time cumulative generation to obtain a new sequence: $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$, where$x^{(1)}(k) = \sum_{i=1}^{k} x^{(1)}(i)$。

Since the essence of the GM(1,1) model is a conditional exponential fitting, a smoothness test

is conducted on the sequence $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\}$. We use Python programming to obtain the test result for the number of cats:

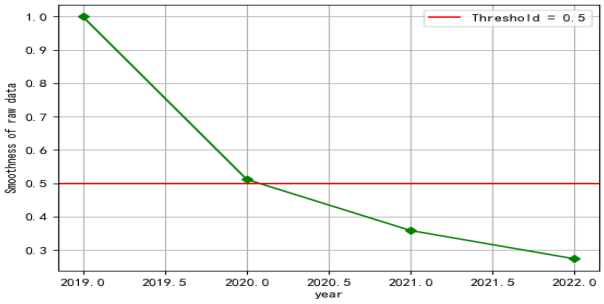


Figure 6 Data smoothness ratio

Indicator 1: The proportion of data with a smoothness ratio less than 0.5 is 50%.

Indicator 2: Excluding the first two periods, the proportion of data with a smoothness ratio less than 0.5 is 100%.

Hence, we consider that the sequence is smooth.

Constructing the background value sequence: To construct the gray differential equation, it is necessary to generate the background value sequence $Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\}$, where $Z^{(1)}(k) = \frac{1}{2}[x^{(1)}(k) + x^{(1)}(k-1)]$。

Establishing the grey differential equation: Based on the accumulated generation sequence and the background value sequence, the following grey differential equation $x^{(0)}(k) + aZ^{(1)}(k) = b$ can be established, where a and b are parameters determined according to different data.

Establish a grey prediction model: Transform the grey differential equation into a differential equation for prediction $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$。

Problem Solving

According to the above steps, the visualization of the gray prediction models for the quantity of pet cats, the number of pet dogs, and the output value of the pet industry, along with the fitting to the original data and the prediction of future data, is presented as follows.

(1) Fitting and predicting curves of the number of pet cats and dogs are as follows:

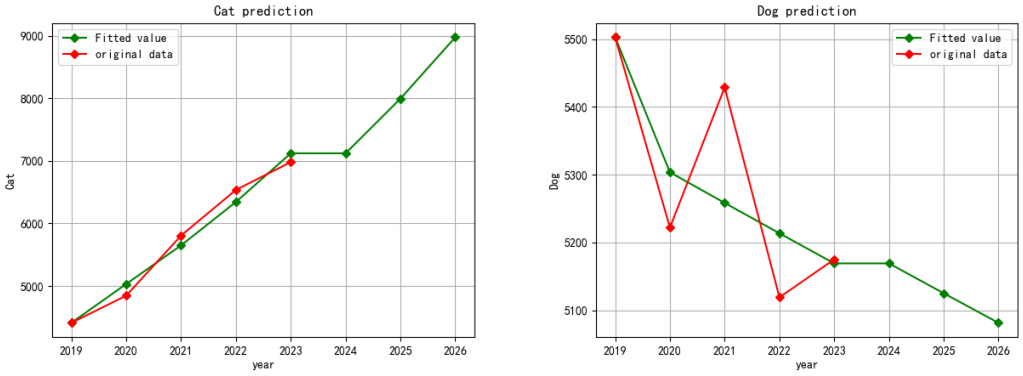


Figure 7 Fitting and predicting curves (The left is cat, the right is dog)

From this, it can be observed that the number of pet cats is increasing annually in the future. This might be because pet cats, as relatively independent and easy-to-tend pets, are gradually being accepted by a greater number of people. The number of pet dogs is gradually and slowly decreasing year by year in the future. This could be attributed to the accelerating pace of modern life and the escalating work pressure, where many people do not have sufficient time and energy to look after pet dogs. In contrast to pet cats, raising a dog demands more companionship, training, and outdoor activities, which poses a significant burden for the busy modern individuals.

The forecasts for the next three years are: The predicted quantities of pet cats for the years 2024 - 2026 are 71.21 million, 79.95 million, and 89.75 million. The predicted quantities of pet dogs for the years 2024 - 2026 are 51.69 million, 51.25 million, and 50.81 million.

(2) Fitting and predicting curves of values of the Pet industry is as follows:

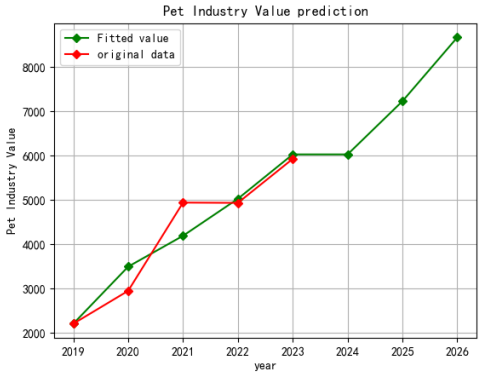
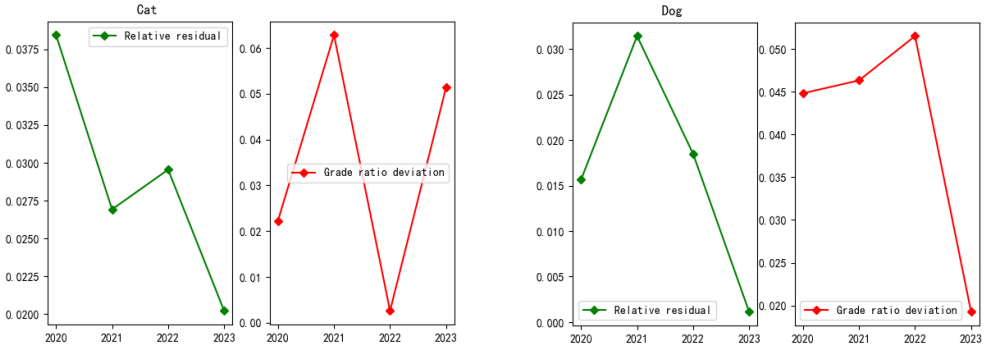


Figure 8 Fitting and predicting curves of values of the Pet industry

The Pet industry value is anticipated to increase year on year in the future. This might be attributed to the sustained and stable economic growth in China and the continuous improvement of residents' income levels, which offer a solid economic foundation for pet consumption. As disposable income rises, people are more inclined to purchase high-quality food, supplies, and services for their pets.

The forecast for the next three years is: The projected values of the Pet industry for the years 2024 - 2026 are 7121, 7995, and 8975 (in 100 million yuan).

In order to examine the fitting degree of the $GM(1,1)$ model to the original data, we utilized Python to depict the relative residuals and the order ratio deviation of each group of data.



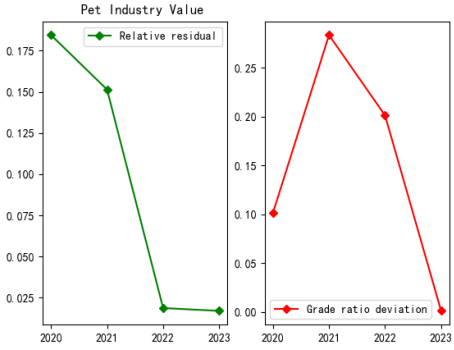


Figure 8 Relative residual and level ratio deviation

It can be observed that the relative residuals of all the data are less than 0.04, and the order ratio deviations are less than 0.3. Therefore, it is considered that the fitting degree of this model to the original data is relatively good.

Model Establishment and Problem Solving for Question 2

Visualization Analysis

We start by doing a visual analysis to grasp the overall distribution of the data.

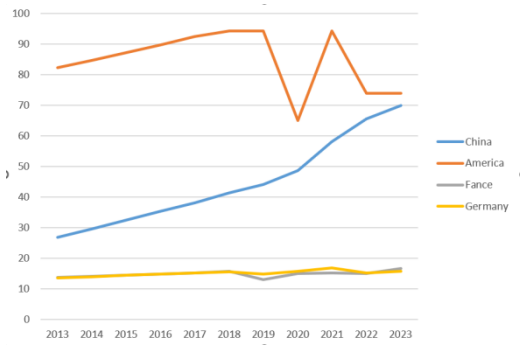


Figure 9 The number of cats in past eleven years

Regarding the number of cats (in millions): China's data continued to rise during this period, especially after 2020, with the growth rate accelerating significantly, reaching its peak in 2023. Data for the U.S. rose steadily between 2013 and 2019, but declined significantly in 2020, then rebounded and then fell again, stabilizing in 2023. Indicators in France remained generally flat, with only slight fluctuations between 2020 and 2021. Germany's performance is similar to that of France, with relatively stable indicators and slight fluctuations.

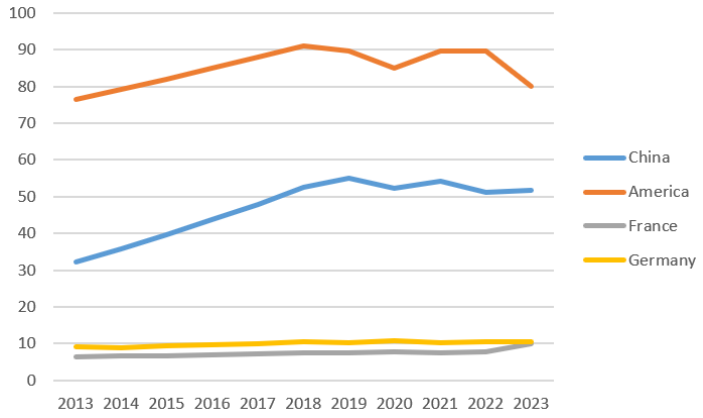


Figure 10 The number of dogs in past eleven years

Regarding the number of dogs (in millions): The data for the United States continued to rise between 2013 and 2017, peaked in 2018 and then declined slightly, and remained relatively stable between 2020 and 2022, before declining slightly again in 2023. China's data showed a trend of first increasing and then stabilizing, rising from 2013 to 2018, and then declining slightly and remaining stable after 2018, with a slight movement during the period. In contrast, data for France and Germany remained relatively stable throughout the time period, with minor changes and only slight fluctuations.

Model Establishment: Time series models

The data we want to predict is time series data. Time series data is arranged in chronological order, and there is often a temporal correlation between the data. Time series models can take full advantage of this temporal correlation and predict future values by capturing the changes in data over time. Based on the above visual analysis of the data, it can be seen that the data are all time series data with stationary or trend changes and only need to be predicted in the short term, so this problem is very suitable for using the exponential smoothing method in the time series to fit the existing data and predict the data in the next three years.

The modeling steps are as follows:

Selecting the appropriate type of exponential smoothing: primary exponential smoothing is used for stationary time series (e.g., the number of cats and dogs in France and Germany), quadratic exponential smoothing is used for time series with linear trends (e.g., the number of cats in China), and cubic exponential smoothing is used for time series with linear trends and seasonal changes (e.g., the number of dogs in China and the United States, and the number of cats in the United States).

Establishing the formula for calculating the smoothed value and the predicted value:

For a primary exponential smoothing:

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1}$$

Among them, S_t is the smoothing value of period t, y_t is the actual observed value of period t, S_{t-1} is the actual observed value of period $t-1$, and α is the smoothing coefficient ($0 \leq \alpha \leq 1$).

For quadratic exponential smoothing:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha) S_{t-1}^{(2)}$$

Among them, $S_t^{(2)}$ is the secondary smoothing value of the t period, $S_t^{(1)}$ is the primary smoothing value of the period t, $S_{t-1}^{(2)}$ is the secondary smoothing value of the period $t-1$, and α is the smoothing coefficient ($0 \leq \alpha \leq 1$).

For cubic exponential smoothing:

$$S_t^{(3)} = \alpha S_t^{(2)} + (1 - \alpha) S_{t-1}^{(3)}$$

Among them, $S_t^{(3)}$ is the cubic smoothing value of the t period, $S_t^{(2)}$ is the secondary smoothing value of the period t, $S_{t-1}^{(3)}$ is the third smoothing value of the period $t-1$, and α is the smoothing coefficient ($0 \leq \alpha \leq 1$).

Determine the smoothing factor: The smoothing coefficient is a key parameter in the exponential smoothing method, and its value will directly affect the prediction results. Through the analysis of rank correlation coefficient results, the number of pets in each country showed a significant change trend and fluctuated greatly, so the α value chose a large range value (between 0.5~1). We set different smoothing coefficients for analysis and measurement, and select the most appropriate smoothing coefficient value, that is, the smoothing coefficient value that minimizes the prediction error.

Problem Solving

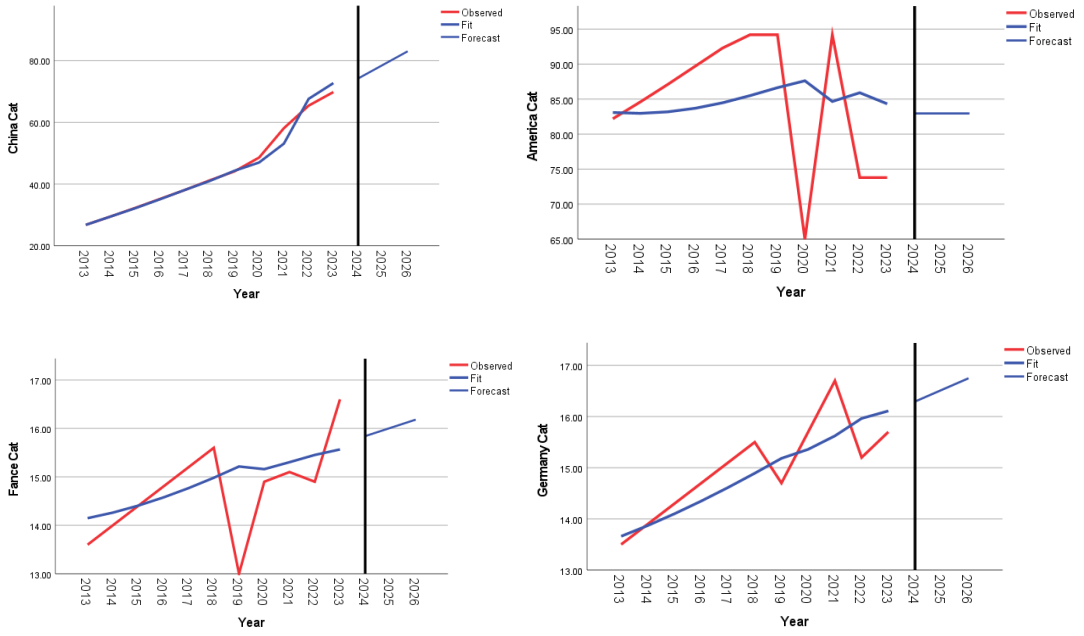
Set Initial Value: Set the observed value of the first sample as the initial smoothed value.

Model training: Use existing data to train the exponential smoothing model to determine the model parameters.

Make predictions: Predict future data with a definite model prediction formula.

Results:

The visualization results of cat and dog population fitting and prediction in China, the United States, France and Germany are as follows.



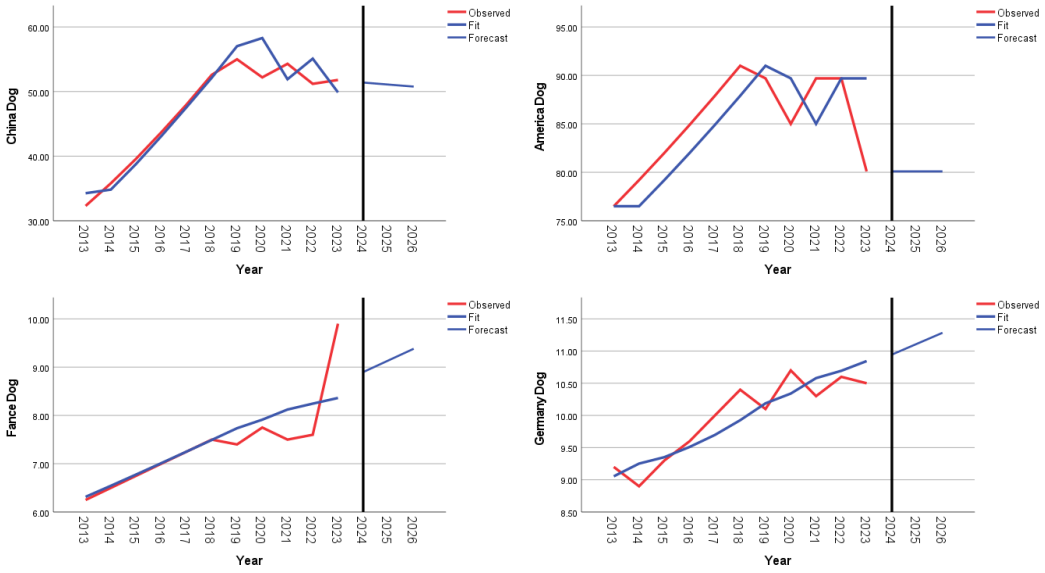


Figure 11 Fitting curves of pet numbers in various countries

The fitting of these images shows some variation. Some images, such as China-Cat, Germany-Cat, China-Dog, America-Dog, France-Dog, Germany-Dog, show a relatively high fit that follows the observed values well most of the time. However, there are also some images that are relatively poorly fitted, such as America-Cat, France-Cat, where there is a large deviation between the predicted value and the observed value, especially at certain time points (such as after 2021), where the bias and fluctuations are more pronounced.

The forecast results of the number of cats and dogs in China, the United States, France and Germany in the next three years are as follows.

Table 3 Prediction of the number of cats(in millions)

Year	China	America	France	Germany
2024	74.20	82.96	15.84	16.29
2025	78.60	82.96	16.01	16.52
2026	83.00	82.96	16.18	16.75

The number of cats is expected to continue to grow, from 74.2 in 2024 to 83.0 in 2026, in the United States at around 829,600, and in France and Germany, albeit slightly, but to a lesser extent. Therefore, we can expect a steady increase in global demand for cat food over the next three years.

Table 4 Prediction of the number of dogs(in millions)

Year	China	America	France	Germany
2024	51.40	80.10	8.90	10.95
2025	51.09	80.10	9.14	11.11

2026	50.77	80.10	9.38	11.28

As for the number of dogs, China showed a slight downward trend, from 51.4 in 2024 to 50.7 in 2026, while the United States also remained stable at around 80.1, with France and Germany increasing slightly. We can expect to see a decrease in domestic demand for dog food in China in the future, but globally, the demand for dog videos should remain stable.

Through the above analysis and forecast, we can know that each country has different preferences for different pets, some countries have more cats in recent years, and some countries have more dogs, resulting in changes in global pet food demand.

Model Establishment and Problem Solving for Question 3

Model Establishment

Through the analysis of supply and demand, China's pet food production depends on the world's pet food demand and domestic pet food demand, so we established a multiple linear regression model to determine the relationship between the number of pets in each country and China's pet food production.

We set Total Value of China’s Pet Food Production to y_1, Total Value of China’s Pet Food Exports to y_2, and the number of pets in China, the United States, Germany, and France to x_1, x_2, x_3, x_4.The multiple linear regression equation is expressed as follows:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_0$$

And $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are coefficients to be determined.

We used the Stepwise Stepwise regression method to remove variables from the multiple linear regression model to determine the final regression model based on their statistical significance. The steps are as follows.

Fitting the initial model: Start by building a model that contains all the independent variables but does not have intercept terms.

Variable deletion: If the p-value of any variable in the model is greater than the exit tolerance (i.e., if there is a small probability that the assumption of a zero coefficient can be rejected), the variable with the maximum p-value is removed. Repeat this step until all remaining independent variables in the model are statistically significant.

End condition: When no insignificant explanatory variables are eliminated from the regression equation, the stepwise regression process ends, and the undetermined coefficient values in the standard multiple linear regression equation are obtained.

The coefficients of the y_1 and y_2 regression models are shown in the table below.

Table 5 Regression coefficient of y_1

The number of iterations	Regression coefficients				
	α_0	α_1	α_2	α_3	α_4
1	-10853.1	40.7	11.6	302.6	-42.6
2	-11735.3	40.6	11.4	294.5	--

Table 6 Regression coefficient of y_2

The number of iterations	Regression coefficients				
	α_0	α_1	α_2	α_3	α_4
1	1010.4	--	1.61	47.2	-85.0

Through stepwise regression, the pet quantity variable in Germany"x_4"was excluded from the regression of pet food production. In the regression of pet food exports, the pet quantity variable in China"x_1"was excluded. The fitting results after excluding non-significant variables are visualized below.

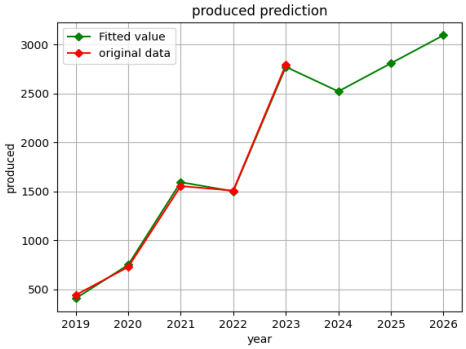


Figure 12 Fitting results of y_1

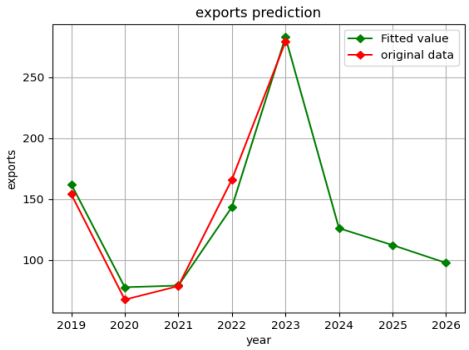


Figure 13 Fitting results of y_2

These two plots show that the fitting results are very good. Thus, the multiple linear regression equations for y_1 and y_2 are obtained.

$$y_1 = -11735.3 + 40.6\,x_1 + 11.4\,x_2 + 294.5\,x_3$$

$$\boldsymbol{y_2 = 1010.4 + 1.61\,x_2 + 47.2\,x_3 - 85\,x_4}$$

Problem Solving

By bringing the number of pets in each country predicted in question 2 in the next three years into the two multiple linear regression models established, the production value and export value of China's pet industry can be predicted. The results are as follows.

Table 7 The results of prediction

Year	y_1	y_2
2024	2521.9	126.1
2025	2808.8	112.3
2026	3095.3	97.7

The forecast results show that in the next three years, the Total Value of China’s Pet Food Production will maintain a steady increase, but the Total Value of China’s Pet Food Exports will decrease. This phenomenon may be due to the rapid increase in the number of pets in China in recent years, which has led to an increase in domestic demand for pet food in China.

Model Establishment and Problem Solving for Question 4

Model Establishment

We started with descriptive statistics on the collected data.

Table 8 Statistics table

Statistics	tariff	Global average food prices	Pet industry expenses	Revenue share of the US pet market
N	5	5	5	5
Minimum	10.00	2.26	97.10	8.70
Maximum	25.00	2.79	147.00	14.60
Mean	22.0000	2.5400	121.9800	11.7800
Deviation	6.70820	0.21575	21.57364	2.28298
Skewness	-2.236	-0.206	-0.056	-0.195
Kurtosis	5.000	-1.601	-2.429	-0.673

Through the calculation, it can be seen that the standard deviation of Pet industry expenses is the largest, indicating that its volatility is larger, the standard deviation of Global average food prices is the smallest, indicating that its volatility is relatively smaller, and the data volume is relatively stable in the above indicators, and the skewness value of tariff is the smallest and less than 0, indicating that the data distribution is relatively shifted to the left.

Let's look the equation in question 3:

$$y_2 = 1010.4 + 1.61\,x_2 + 47.2\,x_3 - 85.0\,x_4$$

The variables x_2, x_3, x_4 and x_5 (tariff), x_6 (global average food prices), x_7 (pet industry expenses), and x_8 (Revenue share of the US pet market) were selected as independent variables to construct multiple linear regression equations. The equation is as follows.

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \alpha_7 x_7 + \alpha_0$$

We use the same method as the third question to obtain the parameters of the multiple linear regression model. The following table shows the results of parameter iteration.

Table 9 Equation coefficient iteration

The number	Regression coefficients

of iterations	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1	-1.57	17.76	-4.87	-12.95	0	3.4	0
2	-1.57	17.76	-4.87	-12.95	--	3.4	0
3	--	-9.39	1.73	-21.13	--	-9.28	162.69
4	--	--	-10.78	-17.34	--	-6.48	122.58
5	--	--	-18.51	-11.80	--	--	51.83
6	--	--	--	-14.38	--	--	54.40

Through stepwise regression, x_6, x_2, x_3, x_7, and x_4 were excluded from the regression of pet food exports. So the final multiple linear regression equation is expressed as follows.

$$y = -175.40 - 14.38\,x_5 + 54.40\,x_8$$

The fitted renderings are shown below.

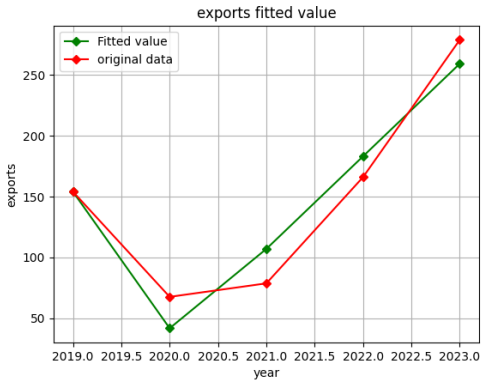


Figure 14 Fitting effect of multiple linear regression equations

As can be seen from the above table, the fitting results of the multiple linear regression model are good.

The test statistics that reflect the effect of regression are shown in the table below.

Table 10 Regression test statistics

Statistics	R squared	P	F	RMSE
y	0.9253	0.0747	12.40	32.78

From the table above, we can see that R squared is close to 1, indicating a good fit, and the P value of the y is less than 0.1, F is greater greatly than 0, and the RMSE is low.

In summary, we believe that the resulting multiple linear regression equation is acceptable.

Problem Solving

According to the multiple linear regression model established in 8.1, we can see that the two variables of x_5 (tariff) and x_8 (Revenue share of the US pet market) have a significant impact on China's pet food exports. In addition, an increase in tariffs will lead to a decrease in exports, and an increase in revenue share of the US pet market will lead to an increase in exports. It can be seen that if European and American countries implement new tariff policies, it will directly affect the export cost of China's pet food products, resulting in an increase in export prices and a decrease in competitiveness. Therefore, the change of tariff policy is a key factor that needs to be paid close

attention to in the development of the industry. The decline in revenue share of the US pet market will also lead to a decline in the demand side, which will directly affect the overseas sales of Chinese pet food products.

To sum up, we will give a feasible sustainable development strategy for China's pet food industry.

The overall strategy is that China should increase domestic demand for pet food and promote the pet food industry to reduce its dependence on the international market.

Here are some specific measures:

Deepen the development of the domestic market: In response to the increasing concern of domestic pet owners about pet health and nutrition, we will launch more high-end, natural and organic pet foods that meet the market demand to meet consumers' demand for high-quality products.

Strengthen brand building: Enhance brand awareness and reputation through online and offline multi-channel marketing, and enhance consumers' trust and recognition of domestic pet food.

Expand sales channels: In addition to traditional offline pet stores and supermarkets, actively use emerging sales channels such as e-commerce platforms and live streaming to expand market coverage.

Optimize policies: Continuously optimize the policies and regulations for raising pet cats and pet dogs, so that the number of pet cats and dogs in China increases, so as to promote the domestic demand for pet food.

Model Evaluation and Improvement

Evaluation and improvement of the model for question 1

Evaluation:

Data Adaptability: The grey prediction model GM (1,1) performed well in handling small sample data and nonlinear relationships. In our case, due to the scarcity of data related to the pet industry, the model effectively solves the problem of insufficient data volume.

Forecast Accuracy: By comparing the predicted value of the model with the actual value, we found that the relative residual and grade deviation were small, indicating that the model fit was better.

Limitations: Grey predictive models, while capable of handling small sample data, may not perform as well as other statistical models when dealing with large data sets.

Improvement:

Model Fusion: Consider fusing grey prediction models with other statistical models (e.g., ARIMA, neural networks, etc.) to improve prediction accuracy and applicability. Through model fusion, you can take full advantage of the advantages of different models and improve the overall prediction performance.

Introducing external variables: Analyze and introduce external variables that affect the development of the pet industry, such as policy changes, consumer behavior, etc., to improve the model structure. By building a multivariate grey prediction model, the complexity of the development of the pet industry can be more comprehensively reflected.

Evaluation and improvement of the model for question 2

Evaluation:

Data Adaptability: Exponential smoothing is ideal for forecasting time series data, especially when there is a trend or seasonality in the data. In this problem, the pet population data of various countries around the world has obvious time series characteristics, so the model is well chosen.

Forecast Accuracy: Through the visualization results, it can be seen that the model can fit the actual data well for most of the time period, but there are large deviations in some time periods (such as the prediction of the number of pet cats in the United States after 2021).

Limitations: The exponential smoothing method is sensitive to emergencies and outliers, which may lead to large deviations in the prediction results.

Improvement:

Model Parameter Optimization: Through trial and error, the optimal smoothing coefficient value is found to improve the prediction accuracy of the model.

Combined Models: Consider combining exponential smoothing with other time series models (e.g., ARIMA models, SARIMA models, etc.) to take advantage of the benefits of different models.

Evaluation and improvement of the model for question 3

Evaluation:

Model Applicability: The multivariate stepwise linear regression model can handle the relationship between multiple independent and dependent variables, which is suitable for the case where the complex relationship between pet populations in multiple countries and pet food production and exports from China needs to be analyzed. Through the stepwise regression method, the model can dynamically select the independent variables that have a significant impact on the dependent variables and exclude the non-significant variables, so as to improve the prediction accuracy and explanatory ability of the model.

Forecast Accuracy: From the prediction results, the model can better capture the change trend of China's pet food production and export, but there may be a certain deviation in the specific value, which may be related to the simplified assumptions of the model, data quality and other factors.

Limitations: The model does not take into account other factors that may affect the production and export of pet food, such as raw material prices, production technology, market competition, etc.

Improvement:

Introduce more variables: Consider introducing more factors that may affect China's pet food production and export, such as raw material prices, production technology, market competition, consumer preferences, etc. By adding variables, the model can be made to reflect the actual situation of the market more comprehensively and improve the accuracy of the forecast.3

Model Fusion: Consider combining a nonlinear regression model or a machine learning model to explore the nonlinear relationship between the independent and dependent variables.

Evaluation and improvement of the model for question 4

Evaluation:

Data limitations: The model relies on data that comes primarily from statistical analysis of past trends, which may not fully capture the uncertainties caused by future policy changes. In particular, changes in foreign and economic policies are often sudden and complex, and difficult to accurately predict through historical data.

Market response lag: The model assumes that the market response is immediate, i.e., policy changes are immediately reflected in the production and export of pet food. However, in real situations, there is often a certain lag in market reactions, which may lead to deviations between the model prediction results and the actual situation.

Improvement:

Introduce more relevant variables: In addition to the number of pets at home and abroad, more variables related to the pet food industry can also be introduced, such as raw material prices, production costs, consumer preferences, etc. These variables can provide a more comprehensive picture of the industry and improve the prediction accuracy of the model.

Consider the lag of market reactions: A time lag effect can be introduced into the model, i.e., a time delay that takes into account the impact of policy changes on the market.

References

Liu, H. , Chu, P. , Meng, Y. , Ding, M. , Lin, P. , & Ding, R. , et al. (2024). The predictability limit of oceanic mesoscale eddy tracks in the south china sea. *Advances in Atmospheric Sciences*(9).

Lin, L. , Zhengkun, L. I. , Ruikun, L. I. , Xuliang, L. I. , & Gao, J. . (2024). Diffusion models for time-series applications: a survey. *Frontiers of Information Technology and Electronic Engineering, 25*(1), 19-41.

Wang, Y. , Li, W. , Liu, J. , Chen, G. , & Wang, J. . (2021). Color prediction model of gray hybrid multifilament fabric. *Journal of Shanghai Jiaotong University (Science)*(3).