Team Number:	apmcm24209163
Problem Chosen:	C

2024 APMCM summary sheet

This paper comprehensively explores various issues concerning the development of the pet industry in China and the world, especially the **pet food industry**.

For the first question, the development of China's pet industry over the past five years by pet type was analyzed through data visualization. Meanwhile, the **Multiple Linear Regression** model is constructed with per capita disposable income, urbanization rate, and elderly population as regressors and the total number of pets as response variables, which is statistically significant. Then, the **ARIMA** model is applied to forecast the total pet numbers for 2024, 2025, and 2026 as 12809, 13376, and 13930 respectively, representing the industry's growth trend.

For the second question, a visual analysis of the changing trends in the global pet industry, mainly reflected by the United States, Germany, and France, is performed. The following two indicators are selected to reflect global pet food demand: total pet population and pet food market size. Then, conduct **ARIMA** model to predict the total number of pets for the next three years as 32292, 32199, and 32199 with corresponding market sizes of USD 156.198, 1616.11, and 176.673 billion, which indicates that the pet population is close to saturation and the pet food market has great growth potential.

For the third question, China's pet food industry development is analyzed respectively through production and export perspectives. Based on **Multiple Linear Regression** and **Holt's Linear Trend Method**, with pet market size and household penetration as factors, future pet food production is projected at 304.71, 378.11, and 528.30 billion CNY. At that point, using the **Prophet Combined Usage Model**, export values are estimated at 3.983, 3.714, and 4.790 billion USD, demonstrating steady growth.

For the fourth question, the global development of China's pet food industry is assessed using the total export value of retail-packaged dog and cat food as an indicator. After that, the **Difference-in-Differences** model evaluated the impact of five foreign economic policies. Last, synthesizing insights from all four questions and the data analysis, four strategic recommendations were proposed to promote the sustainable growth of China's pet food sector.

Finally, a sensitivity analysis is conducted to highlight our models' strengths and weaknesses. Meanwhile, we conclude the whole paper and offer some potential areas for improvement.

Keywords: Pet Industry, Pet Food Industry, Multiple Linear Regression, ARIMA, Holt's Linear Trend Method, Prophet, Difference-in-Differences, Python, Stata

Contents

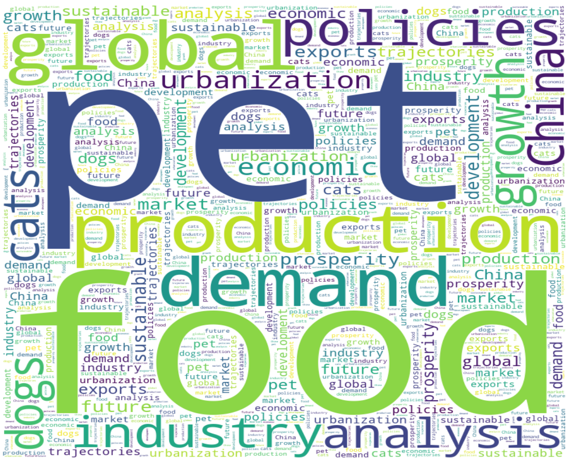
I. Introduction

1.1 Background

The pet industry has emerged as a rapidly growing sector globally, driven by economic growth and rising per capita income. In China, the industry's expansion can be traced back to the establishment of the China Small Animal Protection Association in 1992, followed by the entry of international pet brands such as Royal Canin and Mars in 1993. Over the past few decades, increasing urbanization, shifting consumer philosophies, and the rise of "pet companionship" have fueled demand for pet-related services and products, including pet food, veterinary clinics, pet supplies, and grooming services.

The worldwide pet industry has also experienced significant growth, particularly in regions like Europe and North America. Economic prosperity, demographic changes, and the cultural normalization of pet ownership have played pivotal roles in this trend. The industry exhibits strong interconnections with other economic sectors, such as agriculture, manufacturing, and trade.

China's pet food sector, a critical subset of the pet industry, has witnessed substantial growth in production and exports, capitalizing on domestic and international market demand. However, evolving global economic policies, such as tariffs, pose challenges to sustainable development. This study seeks to analyze the development trends and forecast future trajectories of the pet industry and its food sectors, providing strategic recommendations for continued growth in global and domestic challenges.



(a) Word Cloud

(b) Lovely Pets [1]

1.2 Problem Restatement

Considering the contextual background and the defined constraints outlined in the problem statement, it is essential to address the following key questions.

- **Problem 1**
  Analyze the development of China's pet industry by pet type over the past five years. Build a mathematical model to analyze its factors and forecast its growth over the next three years.
- **Problem 2**
  Examine the global pet industry trends by pet type. Develop a model to predict the global demand for pet food in the next three years.
- **Problem 3**
  Evaluate China's pet food production and export trends. Create a model to project their growth over the next three years.
- **Problem 4**
  Quantitatively assess the impact of new foreign economic policies on China's pet food industry using a mathematical model. Formulate sustainable development strategies based on the calculation.

II. Assumptions and Justifications

- **Assumption 1:** Assume the retrieved data demonstrates a satisfactory degree of credibility and reliability.
- **Assumption 2:** Assume stable economic policies for problem 3, no extreme events such as major trade barriers or global economic crises, and a certain correlation between global demand and China's export volume.
- **Assumption 3:** Assume the exchange rate between USD and CNY is 7.2.
- **Assumption 4:** Assume that all pet food exported from China is retail-packaged dog or cat food.
- **Assumption 5:** Assume the pet in our paper is represented by dogs and cats because the number of dogs and cats takes a large proportion of the whole pet type.
- **Assumption 6:** Assume the policies that influenced China's pet food export volume are selected after 2018 in consideration of the timeliness of policy impact.

III. Notations

Symbol	Description	Unit
Y_{Total}	The total number of pets (dogs and cats)	$10,000,000s$
X_{Inc}	The disposal income per capita for the nth year	CNY
X_{Urb}	The urbanization rate for the nth year	%
X_{Eld}	The number of elderly population for the nth year	$10,000s$
LnP	Log-transformed production for the nth year	10^8 CNY
LnM	Log-transformed market size for the nth year	10^8 CNY
LnR	Log-transformed market penetration rate for the nth year	%

IV. What Drives China's Pet Industry Growth and What Is the Future Trends?

4.1 Data Description

4.1.1 *Data Collection*

To investigate the key factors influencing the development of China's pet industry, it is essential to gather relevant data sets. Table 1 below presents the sources of the data sets used in this chapter, along with detailed explanations of the variables.

Disposable income per capita	CSMAR [2]
Urbanization rate	National Bureau of Statistics [3]
Number of elderly population	CSMAR [2]
The number of pet cats and dogs in 2017 and 2018	Pet Industry Research Report [4]

Table 1 Data Source

4.1.2 *Data Preprocessing*

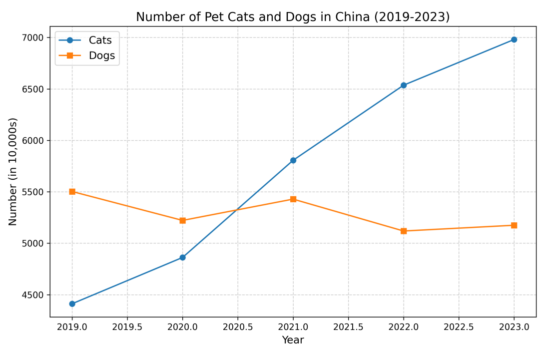
To address the issue of differing dimensions among various data sets and facilitate the construction of a multivariate regression model, normalization of variables is essential. Considering the data's relatively stable distribution, with no significant outliers in the maximum and minimum values, this study employs min-max normalization as the data processing method. The normalization procedure is represented by the following formula (1):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

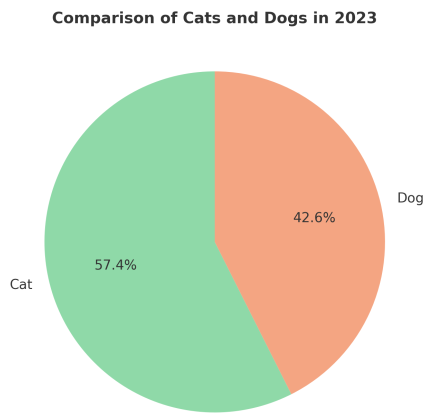
The variable x represents a specific index, where the normalized value $x' \in [0, 1]$. $\max(x)$ and $\min(x)$ denote the maximum and minimum values of data, respectively. Notably, data normalization is only applied in this section for the 3 independent variables, and other sections will specify the data processing methods.

4.2 The Development of China's Pet Industry in the Past 5 Years

Based on the data in Attachment 1, we can see the development of the number of cats and dogs from 2019 to 2023 as Figure 2a. The number of cats shows a continuous upward trend, while the number of dogs fluctuates but remains relatively stable. In 2021, the number of pet cats surpassed pet dogs, becoming the "nation's favorite pet". As Figure 2b, the ratio of cats to dogs is approximately 6:4 in 2023, which also indicates that pet owners show a stronger preference for cats.



(a) Development from 2019-2023



(b) Comparison in 2023

Figure 2 The Situation of Pets Dog and Cat in China

4.3 Model of Influencing Factors of Pet Industry in China

4.3.1 *Influencing Factors and Index Setting*

In this context, the number of pets can serve as an important indicator reflecting the development of the pet market. Therefore, we will use it as the dependent variable in the subsequent multiple linear regression analysis. Since cats and dogs account for a significant proportion of the pet structure[5], we use their total number to represent the total pet population (Y_{Total}).

Combined with practical considerations and according to relevant research, the main factors affecting the development of the pet industry include per capita disposable income (X_{Inc}), urbanization rate (X_{Urb}), and the number of elderly population (X_{Eld}).

4.3.2 *Multiple Linear Regression Model*

Based on the analysis, the multiple linear regression model for the total number of pets is defined as equation (2):

$$Y_{\text{Total}} = \beta_0 + \beta_1 X_{\text{Inc}} + \beta_2 X_{\text{Urb}} + \beta_3 X_{\text{Eld}} + \epsilon \tag{2}$$

4.3.3 *Results*

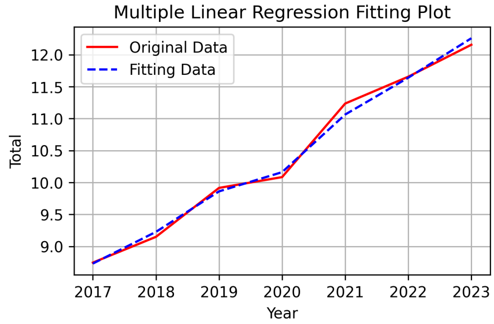
Using Stata's regress function, a multiple linear regression model is constructed, and we get the regression coefficient β. Then the regression model can be presented as (3):

$$Y_{\text{Total}} = 8.731 + 4.708 \cdot X_{\text{Inc}} - 1.864 \cdot X_{\text{Urb}} + 0.681 \cdot X_{\text{Eld}} \tag{3}$$

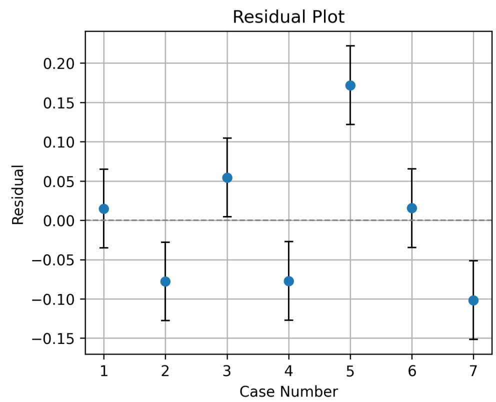
R-squared	**F-test**	**p-value**	**Root MSE**
0.9945	251.98	0.0004	0.135

Table 2 Model Summary

As Table 2, the model summary indicates an R-square value of 0.9945, with a high F-statistic and a p-value of 0.0004 (less than 0.05), demonstrating that the model is statistically significant overall and the regression coefficients are reliable. The model exhibits a strong overall fit. The multiple linear regression fitting plot and residual plot are presented below (Figure 3):



(a) Fitting Plot



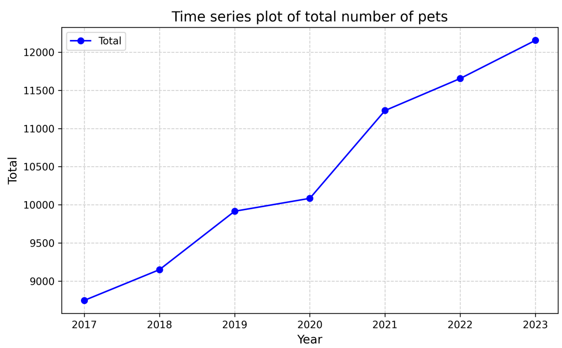
(b) Residual Plot

Figure 3 Multiple Linear Regression Fitting Plot and Residual Plot

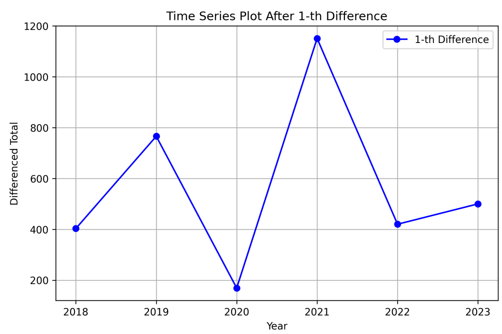
4.4 The Forecasting Model For China's Pet Industry Future

4.4.1 *Data Checking*

In this part, we will still use the data before, which is the total number of pets (the unit is altered to 10,000s). In order to predict the development of China's pet industry in the next 3 years, we will set the ARIMA model in the following section. Firstly, let's check if the data is suitable for the model. From Figure 4a, we can see that the upward trend in the data is not stable. Therefore, we need to use differential processing to transform a non-stationary time series into a stationary one for better model fitting. Conduct the Augmented Dickey-Fuller (ADF) test after each differencing step. If the p-value is less than 0.05, the test is considered successful. Otherwise, proceed with further differencing until the p-value drops below 0.05.



(a) Time Series Plot of Total Number of Pets



(b) Time Series Plot of Total Number of Pets after the 1st Difference

Figure 4 Comparison of Time Series Plots

Figure 4b shows the time series plot for the total number of pets after the 1st differential. Moreover, Table 3 presents the ADF statistics and p-value before and after the ADF test.

	total number of pets	
	ADF statistics	p-value
before differencing	-0.0899	0.9504
after 1-th differencing	-4.5247	0.0002

Table 3 ADF Statistics and P-value

Finally, the data is ready for the model.

4.4.2 *ARIMA time series forecasting*

The ARIMA (Autoregressive Integrated Moving Average) model is a commonly applied technique in time series analysis, used for predicting future values by analyzing patterns in historical data. The mathematical expression of an ARIMA(p, d, q) model can be written as (4):

$$\phi(B)(1 - B)^d y_t = \theta(B)\epsilon_t \tag{4}$$

In the ARIMA model, p represents the autoregressive order, d is the differencing order, and q is the moving average order. The backshift operator B is used to denote lagged values ($B^k y_t = y_{t-k}$), $(1 - B)^d$ is the differencing operator to ensure stationarity, and ϵ_t represents the white noise error term. In 4.4.1, we have known $d = 1$, it is necessary to solve p and q.

Figure 5 represents the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots, respectively, which are used to select the parameters p and q. From the ACF plot, lag 2 appears to be the first significantly correlated lag, suggesting that $q \approx 2$. Similarly, from the PACF plot, lag 1 is the first significantly correlated lag, indicating that $p = 1$. Thus, our ARIMA model is ARIMA(1, 1, 2).

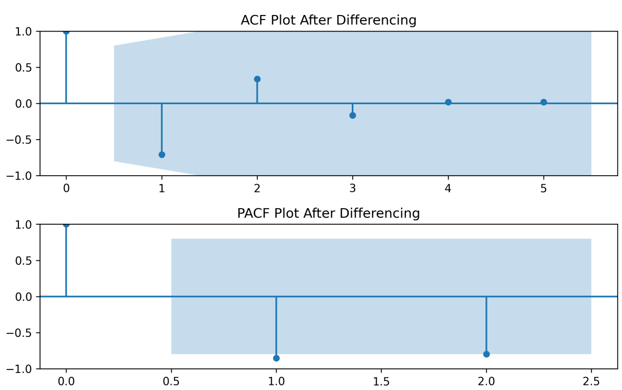


Figure 5 ACF and PACF

4.4.3 *Results*

ARIMA(1,1,2) is used to fit and forecast the total number of pets in the next 3 years. Table 4 manifests the accurate value of prediction.

Based on the projected data, we can see that the number of pets will increase continuously, and the growth rate is about 4%. Figure 6 shows the trend for the history and future. Therefore, it means the development of China's pet industry is strong.

Year	ARIMA(1,1,2)
2024	12809
2025	13376
2026	13930

Table 4 Predicted Total Number of Pets (in 10,000s)

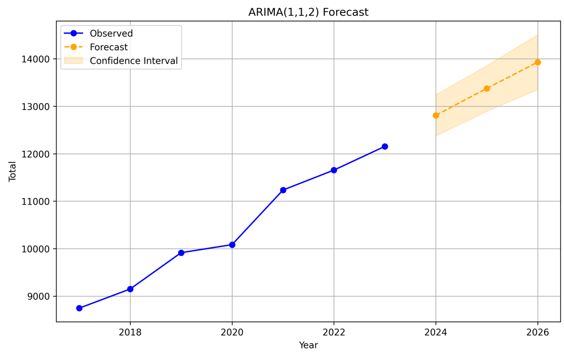


Figure 6 ARIMA Model Forecast for Total Number of Pets

V. The Current State of the Global Pet Industry and Future Demand for Pet Food

5.1 Data Description

5.1.1 *Data Collection*

For this question, we need to shift our perspective from China to the global stage and collect new data accordingly. But at the same time, we followed the approach from the previous question, using the total number of mainstream pets (dogs and cats) as an indicator to study the global demand for pet food in 5.3. Moreover, we will use another indicator to reflect the global demand for pet food: the pet food market size (in 10^8 USD), which comes from industry research reports [6] [7].

5.1.2 *Data Preprocessing*

- Firstly, from the perspective of pet numbers, we mainly divide the global pet number into 3 parts. China represents Asia, the United States represents the Americas, and Germany and France represent Europe. Figure 7 shows the 4 countries' geographical position in the world, and Figure 8 indicates the total number of pet cats and dogs from 2019 to 2023.

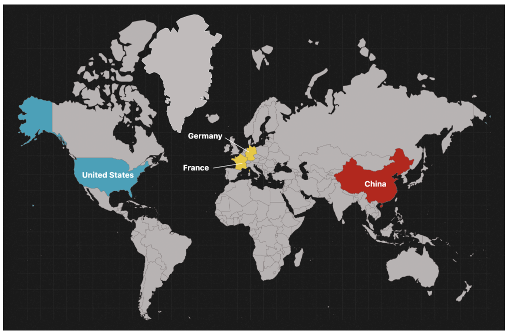


Figure 7 Geographical Position [8]

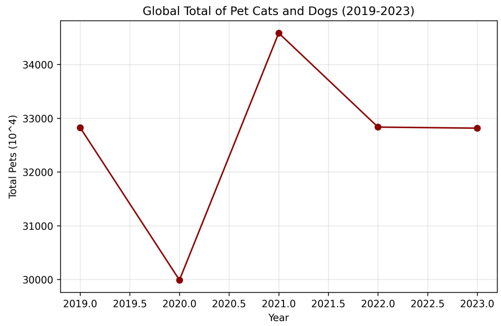


Figure 8 Time Series of Global Number of Pet Dogs and Cats

- Secondly, Figure 9 unfolds the value and growth rate of the global pet food market size from 2014 to 2023, and Figure 10 shows the time series.

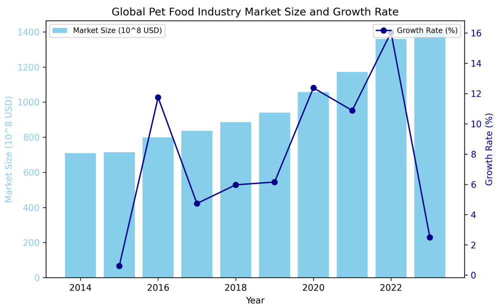


Figure 9 Global Pet Food Market Size

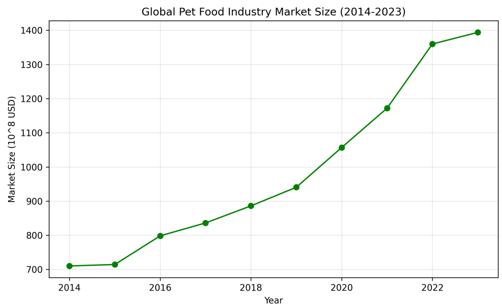


Figure 10 Time Series of Global Pet Food Market Size

- Finally, because the global pet food market size shows exponential growth, we use logarithmic scaling to stabilize the trend. The log-transformed chart is shown in Figure 11. After the logarithmic process, the data is prepared for the 5.3 forecast.

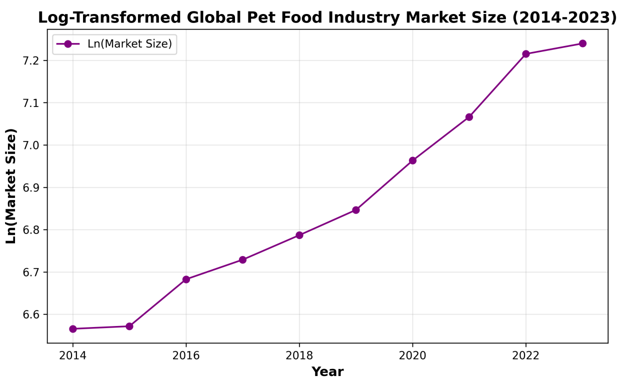
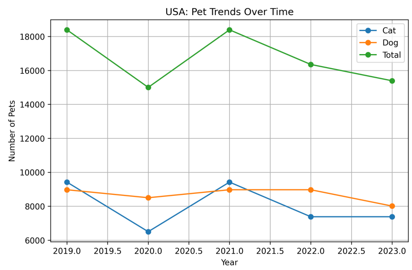


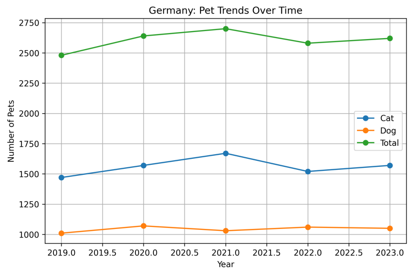
Figure 11 Log-Transformed Time Series of Global Pet Food Market Size

5.2 The Development of Global Pet Industry in the past 5 years by the pet type

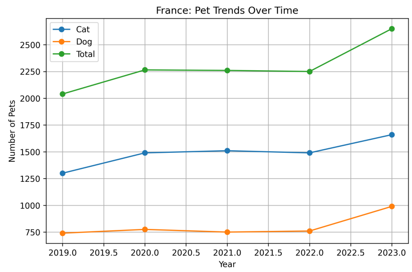
From the view of the number of pets in Figure 12, we can see the 3 countries have different but similar trends of the pet numbers in the past 5 years. For America, the number of dogs is greater than cats in the most time but the number of cats is always larger than dogs for the other two countries. Meanwhile, the 3 countries' variation trends are relatively stable.



(a) USA



(b) Germany



(c) France

Figure 12 Comparison among 3 Countries

Figure 13 and 14 show the comparison among the 3 countries about the number of pet dogs and cats. We can see that America has more pet cats and dogs. This is easy to understand because America holds more population (approximately 334.9 million) comparing France and Germany only own about 68.17 million and 84.48 million respectively.

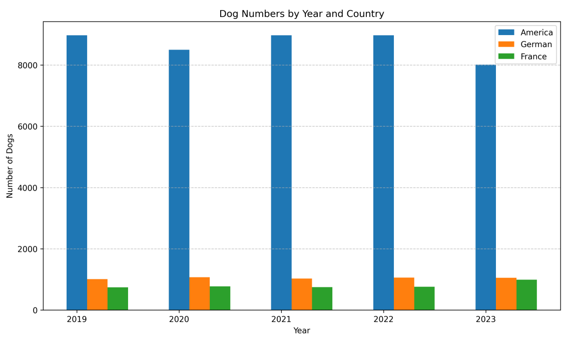


Figure 13 Comparison of Dogs

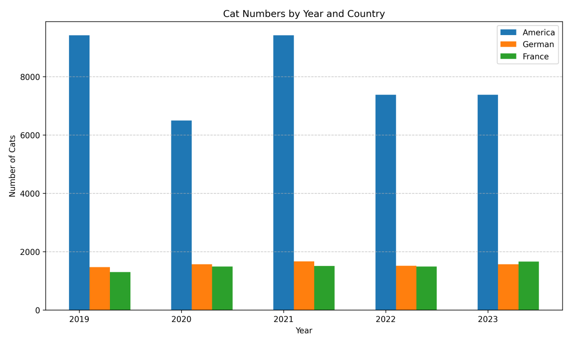


Figure 14 Comparison of Cats

Hence, we also compare the ownership ratio by country and years. As Figure 15 and 16, in recent years, the dog ownership ratio in the United States has been nearly twice that of Germany and France, whereas the cat ownership ratio among the three countries has been converging.

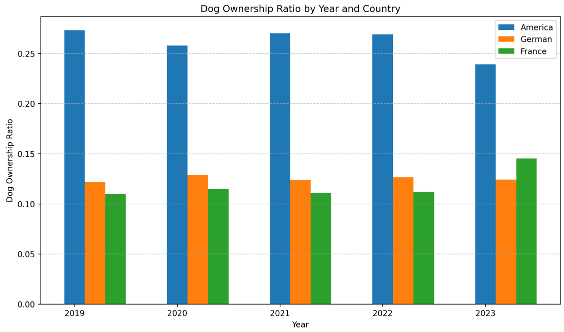


Figure 15 Ownership Ratio of Dogs

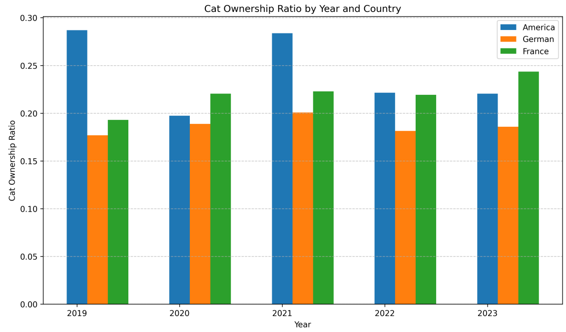


Figure 16 Ownership Ratio of Cats

5.3 Model to Forecast the Global Demand for Pet Food in the Future

5.3.1 *ARIMA Time Series Forecasting*

In this place, we still use the ARIMA model but from two indexes. One is the total number of pets, and the other is the global pet food market size. Because we have shown the equation of ARIMA before (4), now let's find the parameters p, d, q.

- Differential processing and ADF test

  Because we have introduced the differential processing and ADF test before in 4.4.1, we will just use the method to get parameter d. Let's perform the ADF test directly and see if there is a need to make a difference. Until the p-value is less than 0.05, we record the result in Table 5. Figure 17 and 18 shows the time series after the 1st difference.

	Total Pets		Ln(Pet Food Market Size)	
	ADF statistics	p-value	ADF statistics	p-value
Initial ADF Test	-2.4008	0.1415	0.5835	0.9872
First-order differencing	-3.3663	0.0122	-3.6227	0.0053

Table 5 ADF statistics and p-value

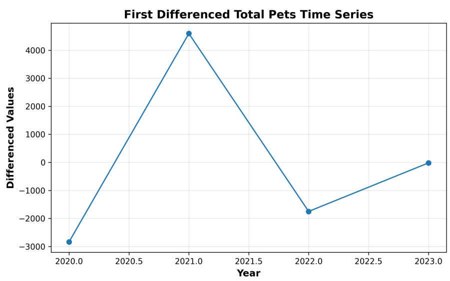


Figure 17 Total number of pets after 1st Difference

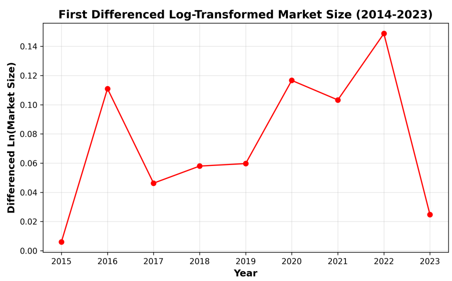


Figure 18 Log-Transformed Market Size after 1st Difference

- Determination of p and q

  In 4.4.2, we use ACF and PACF figures to confirm p and q. So in this Model, we want to use another tool, AIC and BIC. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are statistical tools used to compare and select models. AIC seeks a balance between a model's goodness of fit and its complexity, aiming to find an optimal trade-off. A lower AIC value indicates a better model. It is calculated as $\text{AIC} = 2k - 2\ln(\hat{L})$, where k represents the number of parameters in the model, and $\hat{L}$ is the maximum likelihood of the model.

  Similarly, BIC also supports model selection but places a stronger emphasis on simplicity by penalizing model complexity more heavily. It favors simpler models, and lower BIC values indicate better models. The formula for BIC is $\text{BIC} = -2\ln(\hat{L}) + k\ln(n)$, where n is the sample size.

  This paper examines various combinations of p and q values, evaluating their respective AIC and BIC metrics. Ultimately, it identifies the specific p and q combinations that achieve the lowest AIC or BIC score.

  For the total number of pets, when p=0 and q=2, the AIC and BIC have the lowest value. For global pet food market size, when $p = 2$ and $q = 0$, the AIC and BIC have the lowest value. Finally, we decided to utilize ARIMA(0,1,2) and ARIMA(2,1,0) in the following results prediction.

5.3.2 *Results*

Table 6 below shows the forecasting of global demand for pet food in the next 3 years. Meanwhile, the global pet food market size data is inversely transformed from the logarithm form.

Year	Total Number of Pets (10^4s)	Global Pet Food Market Size (10^8 USD)
2024	32292	1561.98
2025	32199	1616.11
2026	32199	1766.73

Table 6 Forecast Results for Global Pet Food Market and Example Variable

Based on the objected data, as Figure 19 and 20, we can see that the growth of the total number of pets globally comes to a standstill, which is due to the total number of the population becomes stable. However, the growth of the global pet food market size is rapid and strong, which is similar to other scholar's research. In summary, the global demand for pet food is evident, and the potential of the food industry is significant.

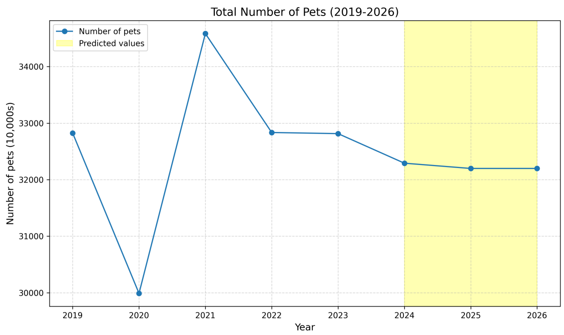


Figure 19 Prediction of the Total Number of Pets

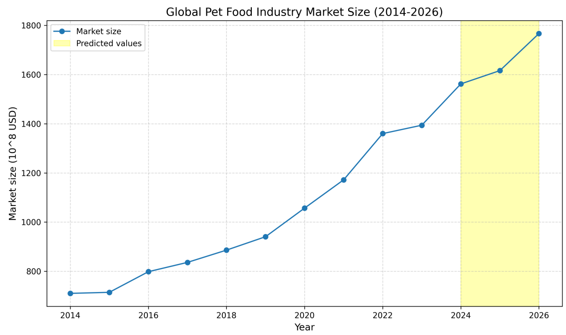


Figure 20 Prediction of the Global Pet Food Market Size

VI. History and Future: Production and Export of China's Pet Food Industry

6.1 Data Description

6.1.1 *Data Collection*

Besides Attachment 3, we find other data that may influence China's production and export of pet food. For production, we will use China's pet food market size [9] [10] and pet household penetration rate [11] [12], which comes from industry research reports. For export, based on Attachment 3, we further searched America and Thailand's pet food export from the UN Comtrade Database [13]. Moreover, we will use the global pet food market size as a factor to reflect the global demand for pet food.

6.1.2 *Data Preprocessing*

To better analyze the production and export of China's pet food in 6.2, we converted the export data from USD to CNY to allow for direct comparison. Since the export data in 2019 is already in CNY from Attachment 3, we transformed the export data for 2020 to 2023 into CNY to maintain consistency. Secondly, we will apply a logarithmic transformation (log-log [14]) to the data used for predicting the total value of China's production (pet food market size and pet household penetration rate), as well as the total value itself. These transformed variables will be utilized in the multiple linear regression model constructed in Section 6.3.1 to better fit the model.

6.2 The Development of China's Pet Food Industry

Under the data of Attachment 3, we will analyze the development of China's pet food industry. As Figure 21 and 22, the production shows significant fluctuations in growth, particularly in 2021 and 2023, but the growth momentum remains unchanged. For exports, the growth rates highlight a steep decline in 2020, followed by a rapid recovery in 2022 and 2023. We speculate that the decrease in export volume is likely due to the impact of the COVID-19 pandemic.

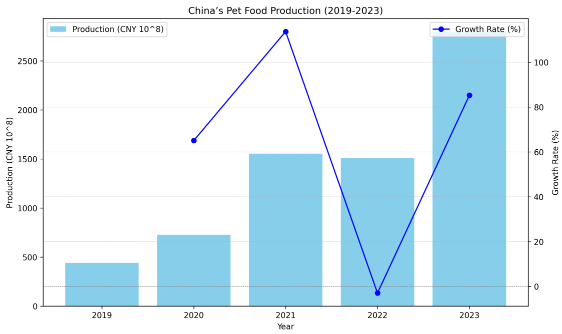


Figure 21 Production of China's Pet Food

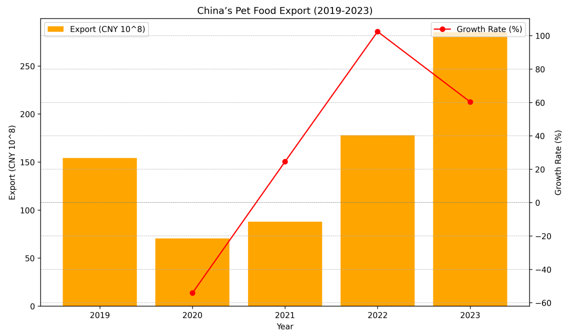


Figure 22 Export of China's Pet Food

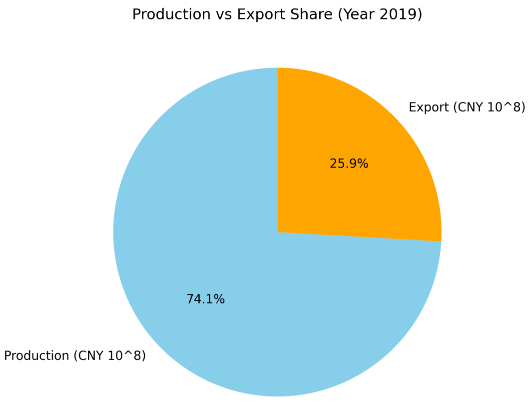


Figure 23 Production vs Export in 2019

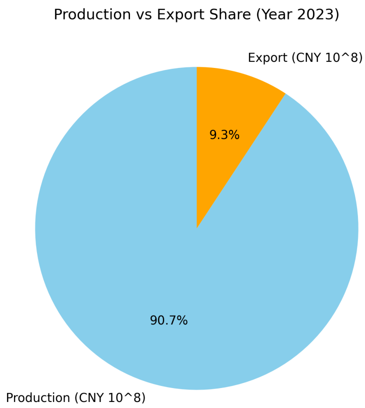


Figure 24 Production vs Export in 2023

Moreover, from Figure 23 and 24 we can observe the change in the ratio of production to export volume. The production-to-export ratio shifted from 76:24 in 2019 to approximately 90:10 by 2023. This indicates the growing demand from Chinese pet food consumers, while also reflecting an increase in pet food production in other countries and the diversity of imports. Naturally, potential underlying reasons for this outcome may include the impact of the COVID-19 pandemic.

6.3 Forecast Production and Export of Pet Food

6.3.1 *Production—Multiple Linear Regression and Holt's Linear Trend Method*

- Firstly, let's set the index of the data in 6.1.2. We use LnP to represent the logarithm-transformed total value of China's pet food production, LnM to represent the logarithm-transformed market size of China's pet food, and LnR to represent the logarithm-transformed pet household penetration rate in China. Next, we establish the multiple linear regression equation (5) as follows:

$$\mathrm{LnP} = \beta_0 + \beta_1 \mathrm{LnM} + \beta_2 \mathrm{LnR} + \epsilon \tag{5}$$

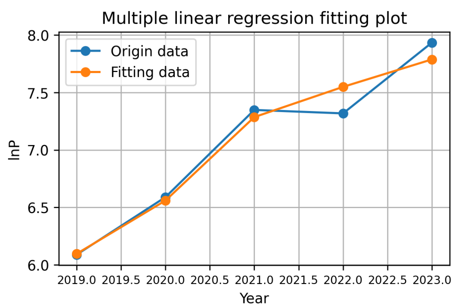
To construct a multiple linear regression model in Stata using the `regress` function, we can determine the estimated values of the regression coefficients β. The resulting regression model can be expressed as equation (6):

$$\mathrm{LnP} = 34.321 - 2.754 \cdot \mathrm{LnM} + 4.274 \cdot \mathrm{LnR} \tag{6}$$

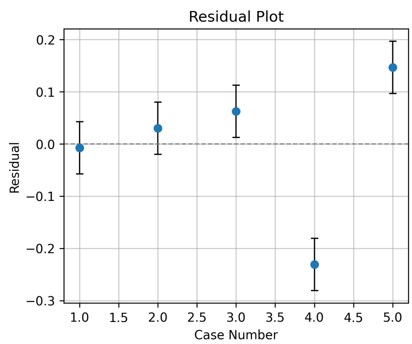
R-squared	F-test	p-value	Root MSE
0.9617	64.21	0.0153	0.1995

Table 7 Model Summary

As Table 7, the model summary reports an R-square value of 0.9617, accompanied by a high F-statistic and a p-value of 0.0153 (below 0.05). This indicates that the model is statistically significant, and the regression coefficients are robust. Overall, the model demonstrates a strong fit. The multiple linear regression fit plots and the residual plot are displayed below as Figure 25:



(a) Fitting Plot



(b) Residual Plot

Figure 25 Multiple Linear Regression Fitting Plot and Residual Plot of 6.3.1

- Secondly, let's use Holt's Linear Trend Method to predict the value of LnM and LnR in the next 3 years. The equations are below (7) (8) (9):

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{7}$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{8}$$

$$F_{t+h} = L_t + hT_t \tag{9}$$

where L_t represents the level, indicating the current state of the time series. T_t denotes the trend, showing the rate of increase or decrease in the series. F_{t+h} is the forecasted value for time $t+h$, while Y_t is the actual observed value at time t. The parameters α and β are smoothing coefficients for the level and trend, respectively, with values between 0 and 1. Lastly, h represents the forecast horizon, indicating steps into the future.

- Now, we need to find the most optimized parameter (α and β). The parameter searching method we use is grid search, which can systematically explore a predefined set of hyperparameter values in order to identify the combination that results in the best performance of a model. It ensures an exhaustive search within the specified parameter space, allowing the optimization of the model's performance metrics, such as minimizing error or maximizing accuracy. After that, we get α=0.9 and β=0.6 for LnM and α=0.7 and β=0.7 for LnR. By fitting the data into the model, we obtain as Table 8:

Year	**LnM_Predicted**	**LnR_Predicted**
2024	7.319402	-1.436926
2025	7.327301	-1.381342
2026	7.334733	-1.298290

Table 8 Predicted Values for LnM and LnR

- Finally, we take the predicted values into the regression Equation 6 and get the predicted LnP. The predicted value of LnP and its values after taking the antilogarithm are below as Table 9 and Figure 26 shows the expected trend in the next 3 years.
- From the prediction, we know that the production trend of China's pet food industry is continuously upward, and we can infer that the demand for pet food will be more and more substantial.

Year	LnP Predicted	P Predicted
2024	8.021945	3047.098691
2025	8.237757	3781.051173
2026	8.572254	5283.023567

Table 9 Predicted LnP and P-values

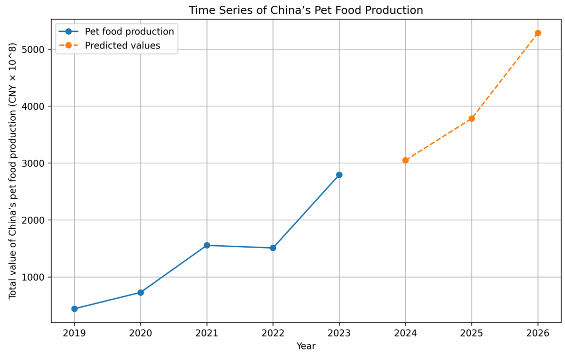


Figure 26 Time Series of Predicted Values of Production

6.3.2 *Export—Prophet Combined Usage Model*

- Initially, we need to predict the value of American pet food export, Thailand's pet food export, and the global pet food market size in the next 3 years. We use the package of prophet in Python and get the expected value below as Table 10 :

Year	Thai Export	USA Export	Global Market Size
2024	23.3446	26.2524	1520.8442
2025	29.2831	30.4037	1664.1022
2026	29.8885	32.3100	1785.1946

Table 10 Predicted Export and Global Market Size Data

- Then we will build a prophet model with multiple external variables and substitute the predicted data into the model to obtain the forecast value for China's pet food export. This approach better incorporates the interference of external factors, making the prediction more accurate. The forecasting value and future trend are below as Table 11 and Figure 27.

Year	Predicted Export
2024	39.8336
2025	37.1402
2026	47.8995

Table 11 Predicted Value of China's Pet Food Export

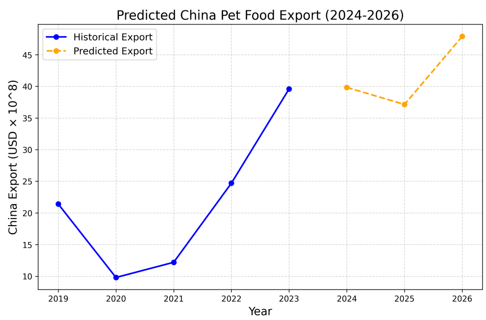


Figure 27 Time Series of Predicted Value of Export

- According to the prediction, we get that the exports will not grow constantly like production. The potential influence of the global economic environment should be considered.

VII. How Do Foreign Economic Policies Impact China's Pet Food Industry Strategies

7.1 Data and Information Description

7.1.1 *Data Collection*

In this section, we obtained the total export value of retail-packaged dog or cat food feed from January 2015 to October 2024 from the General Administration of Customs of China [15] to represent the global development of pet food industry. This data allows for a more accurate analysis of how economic policies in other countries impact China's pet food industry. As Figure 28, we can see the trend of retail-packaged dog or cat food feed.

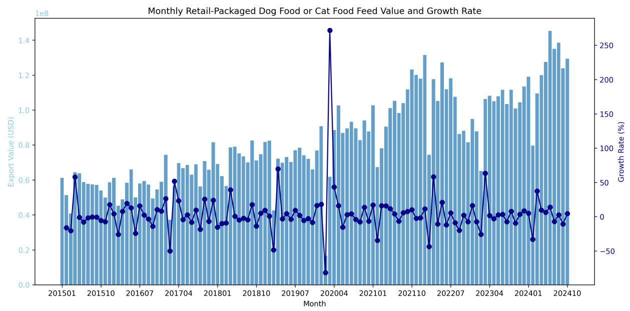


Figure 28 Monthly Retail-Packaged Dog Food or Cat Food Feed Value and Growth Rate

Figure 29 presents the variation of China's monthly retail-packaged pet food feed export value.

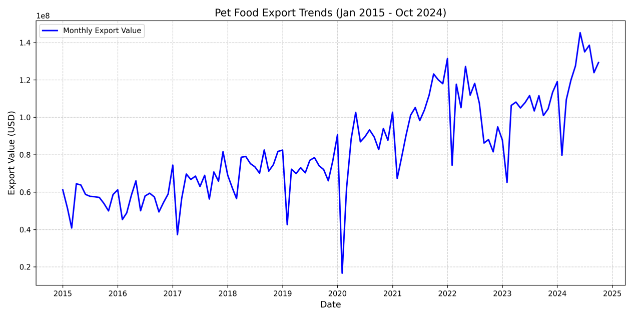


Figure 29 Time Series of Monthly Retail-Packaged Dog Food or Cat Food Feed Value

7.1.2 *Foreign Economic Policies*

Different policies can lead to different changes, which are specifically reflected in the Retail-Packaged Dog Food or Cat Food Feed Value. We have collected foreign economic policies from June 2017 to February 2023.

- **Policy 1 (July 2018):**
  In 2018, trade tensions between the United States and China escalated, leading the U.S. to impose tariffs on a large number of goods imported from China, including some pet foods. Specifically, in July 2018, the U.S. government imposed a 25% tariff on pet foods, toys, and other related products exported from China.
- **Policy 2 (May 2019):**
  On May 9, 2019, the U.S. government announced that starting May 10, 2019, tariffs on $200 billion worth of goods imported from China would increase from 10% to 25%, marking an official escalation in U.S.-China trade tensions.
- **Policy 3 (December 2020):**
  In December 2020, China and the EU signed the "China-EU Comprehensive Investment Agreement," which simplified trade for certain Chinese products, including pet food. While the focus was on investment, it also improved market access for Chinese exports and reduced administrative barriers.
- **Policy 4 (January 2021):**
  In January 2021, the United Kingdom officially exited the European Union and began implementing new tariff policies for non-EU countries, which could affect the import of pet food from China.
- **Policy 5 (January 2024):**
  The UK will introduce new import and tax regulations in 2024 under the Target Operating Model (TOM), digitizing trade processes and enhancing health protections. Pet food imports from non-UK countries, including the EU, will face new classifications and fees.

7.2 Difference-in-Differences Model

Difference-in-Differences (DID) is a method that quantifies the causal impact of policy interventions on China's pet food market export volume by constructing appropriate control and treatment groups and using time dummy variables before and after the policy. This provides a basis for future policy adjustments. Following is the procedure to construct the DID model:

- **Determining the treatment group and control group:**
  Treatment group: This group is affected by the policy intervention, and the DID model

compares the difference in the export volume before and after the policy, relative to the control group.

Control group: This group assumes no policy intervention, and any observed export volume changes are attributed to natural trends (such as seasonal effects).

- **Mathematical Representation of the DID Model:**

  Because there are 5 policies issued from January 2018 to October 2024, it is reasonable to use the Multi-policy DID model to consider the total impact of these policies on export volume, as the following equation (10).

$$Y_{it} = \beta_0 + \beta_1 \text{Policy}_{it} + \beta_2 \text{TimeTrend}_t + \beta_3 \text{LagExport}_{i(t-1)} + \sum_{k=1}^{12} \gamma_k \text{Month}_k + \epsilon_{it} \tag{10}$$

where Policy_{it} is a binary variable, determined based on the policy's implementation time. If the policy is effective after time t, then $\text{Policy}_{it} = 1$, otherwise $\text{Policy}_{it} = 0$.

β_1 is the key parameter, representing the net effect of the policy on the export volume.

By incorporating the TimeTrend_t variable, the DID model can control for changes in export volume over time, avoiding biases caused by temporal variations.

Meanwhile, $\text{LagExport}_{i(t-1)}$ helps control for the fact that past export performance might influence future performance.

Additionally, seasonal dummy variable γ_k is introduced to account for impacts caused by seasonal factors.

Therefore, DID estimates the net effect of the policy by measuring the changes within each group before and after the policy implementation (first difference) and comparing the changes between the treatment and control groups to eliminate common external influences (second difference), thus isolating the effect of the policy independent of other factors.

7.3 Results

The monthly export volume-policy diagram used in the DID model is below as Figure 30.

The goodness-of-fit metrics and the policy impact results derived from the DID model we applied are summarized as Table 12.

DID Model			Policies	
R-squared	**Adj.R-squared**	**Prob (F-statistic)**	**Coef**	**P>|t|**
0.93322	0.8887	0.017257	9.9587e+06	0.08607

Table 12 Model Summary

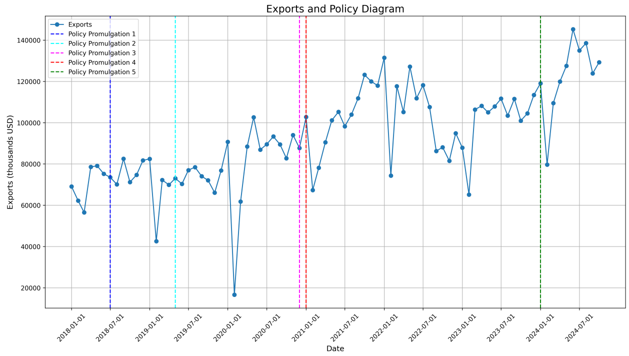


Figure 30 Exports-Policy Diagram

The results of the DID model demonstrate a strong fit to the data, with an R-squared of 0.933 and an adjusted R-squared of 0.889. This indicates that 93.3% of the variation in the dependent variable is explained by the predictors included in the model. The model's overall significance is confirmed by a Prob (F-statistic) of 0.017257, highlighting that the independent variables collectively have a significant influence on the outcome. Furthermore, the coefficient of the policy variable is 9.9587e+06, and its p-value of 0.086 falls within the 10% significance level. This suggests that the policies likely have a notable effect on the outcome variable, even though the impact is not statistically significant at the 5% level.

7.4 Strategies

By integrating the results from the above 4 questions, China's pet food industry can focus on technological innovation, market diversification, data-driven strategies, and policy adaptation to formulate comprehensive strategies for the sustainable development of China's pet food industry:

- **Based on Question 1 (Development and Forecast of China's Pet Industry)**
  It is suitable to conduct some strategies to enhance domestic production and technological capabilities:
  1) Optimize supply chains (improve the efficiency of raw material procurement and production processes) to reduce costs and meet the growth in pet ownership. 2) Increase cat food production due to the increment of the proportion of cats to dogs. 3) Promote technological innovation and develop high-value-added products to meet market demands.
- **Based on Question 2 (Development and Forecast of the Global Pet Industry)**
  It is reasonable to take certain strategies to address overseas market demand and competition:
  1) Reduce reliance on a single market and explore diversified international markets. 2) Efforts can be made to develop customized products, driving competitiveness through quality. 3) Increase brand recognition overseas by leveraging cross-border e-commerce, social media

marketing, and other channels to expand market share.

- **Based on Question 3 (Forecast of Production and Export)**

  It is necessary to leverage data for development:

  1) Regularly monitor deviations between actual export data and forecasted values and dynamically adjust export strategies in response to market changes. 2) Leverage big data to optimize export processes, inventory management, and market forecasting, improving overall operational efficiency.

- **Based on Question 4 (Impact of Foreign Economic Policies)**

  It is significant to address the impact of external economic policies:

  1) Strengthen partnerships with local importers in policy-sensitive regions (e.g., the U.S. and Europe) and establish local manufacturing facilities to bypass high tariff barriers.

  2) Create a flexible export market layout and expand into markets with lower or zero tariffs.

  3) Enhance monitoring and responsiveness to international trade policies and actively participate in bilateral or multilateral trade negotiations to secure favorable trade conditions.

VIII. Sensitivity Analysis and Error Analysis

When performing a multiple linear regression model in Problem 1, to assess the model's sensitivity, an 85%-115% confidence interval should be provided for the influential factor.

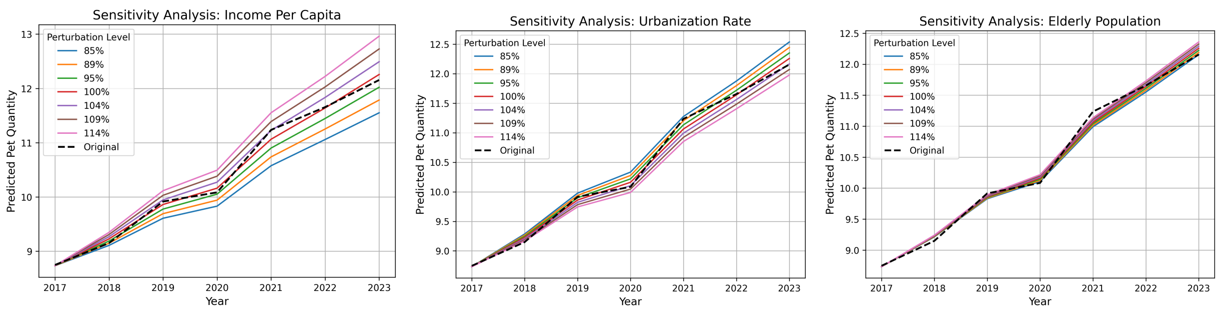


Figure 31 Sensitivity Analysis for Problem 1

Figure 31 shows the sensitivity analysis for income per capita, urbanization rate and elderly population. The analysis demonstrates that while per capita disposable income exerts the most significant influence on pet quantity predictions, the overall model exhibits low sensitivity to input variations and maintains robust predictive stability.

Moreover, in Problem 2, we also have a multiple linear regression and need to assess model's sensitivity. Like before, an 85%-115% confidence interval will be provided for the influential factor.

From Figure 32, the sensitivity analysis reveals that both the pet household penetration rate

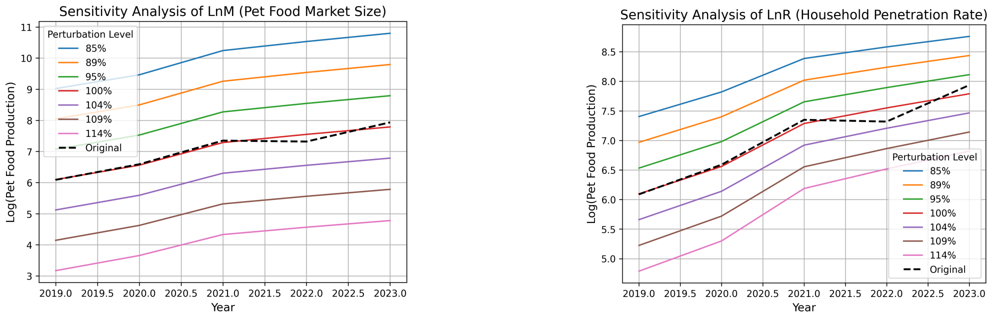


Figure 32 Sensitivity Analysis for Problem 2

(LnR) and the pet food market size (LnM) significantly influence pet food production (LnP). LnM exhibits a slightly higher sensitivity than LnR, indicating its dominant role in driving LnP predictions. While the model demonstrates strong responsiveness to variations in both variables within the 85%-115% range, the overall trends remain consistent, showcasing the model's robustness and predictive stability.

IX. Model Evaluation and Further Discussion

9.1 Model Evaluation

9.1.1 *Advantages of the Model*

- For questions 1 and 2, the multiple linear regression model is simple, interpretable, and suitable for continuous variables like disposal income per capita. Moreover, the ARIMA model excels in time series forecasting, effectively capturing trends, handling non-stationary data through differencing, and offering various extensions, which is good for predicting pet industry trends.
- For question 3, the combination of multiple linear regression and Holt's linear trend method allows for the simultaneous handling of the effects of multiple independent variables and the trend changes in time series, thereby enhancing the model's forecasting ability and flexibility. Additionally, the two-step Prophet modeling approach first separately models the time series of two influencing factors, using their predicted outcomes as inputs to predict the final variable. This method improves forecasting accuracy, captures complex temporal dependencies, and enhances the model's flexibility, interpretability, and robustness.
- For question 4, the DID model estimates the causal effect of policies or interventions by comparing the differences between treatment and control groups, controlling for time and group effects. It is straightforward to implement and well-suited for estimating the impact of policy changes on export volumes.

9.1.2 *Shortcomings of the Model*

- The amount of data collected is limited, resulting in low model accuracy.
- The relatively simple and incomplete consideration of indicators and influencing factors may lead to model distortion.
- The policies (in Problem 4) taken into account are limited, which may cause some deviation to our outcomes.

9.2 Promotion of the Model

- We can seek more data and set additional indicators and influencing factors to build a more comprehensive and complete mathematical model, enabling better analysis and forecasting of the development of the pet industry in China and globally.
- We need to explore more foreign economic policies to explain the trends in export volume changes and build a more accurate DID model to achieve better-fitting results.

X. Conclusion

In this study, we primarily used regression models and time series models to analyze and forecast the development and changes in the pet industry, especially the pet food sector, both in China and globally. Through our analysis and forecasting, we gained insights into the rapid growth and market potential of the pet industry in China, as well as the emerging trends in the global pet industry.

Our research findings can provide support and assistance to researchers and scholars in the pet industry. The predicted results can, to some extent, serve as a reference for those looking to invest in the pet industry and the pet food sector.

In addition, we have proposed suggestions that contribute to the sustainable development of China's pet food industry. These recommendations will provide practical guidance for the government and market regulatory authorities, helping them formulate reasonable and effective policies and strategies to promote growth in production and exports.

XI. References

[1] Nipic, Image and Design Resource Platform, `https://www.nipic.com`, visit time (2024, November, 24).

[2] CSMAR, China Stock Market & Accounting Research Database, `https://data.csmar.com/`, visit time (2024, November, 24).

[3] National Bureau of Statistics of China, National Statistical Data, https://www.stats.gov.cn/, visit time (2024, November, 24).

[4] Kezhen, Zhang, Industry Analysis Report, https://www.chyxx.com/industry/201910/800030.html, visit time (2024, November, 24).

[5] ITJUZI, 2020 China pet consumption market analysis report, https://pdf.dfcfw.com/pdf/H3_AP202101151450898873_1.pdf?1610718668000.pdf, visit time (2024, November, 24).

[6] Insight and Info, Research Report on the Development Status and Investment Prospects of China's Pet Food Industry (2023-2030), https://www.chinabaogao.com/detail/646061.html, visit time (2024, November, 24).

[7] Sohu, Pet food market analysis, https://www.sohu.com/a/826984187_120488747, visit time (2024, November, 24).

[8] AmCharts, Pixel Map Visualization Tool, https://pixelmap.amcharts.com/, visit time (2024, November, 24).

[9] Insight and Info, Research Report on the Development Status and Investment Prospects of China's Pet Food Industry (2023-2030), https://www.chinabaogao.com/detail/636201.html, visit time (2024, November, 24).

[10] Flush Finance, "'2024 China Pet Food Consumption Report' released, where is the next 'blue ocean' of industry competition?", https://news.10jqka.com.cn/20240703/c659439045.shtml, visit time (2024, November, 24).

[11] Topsperity Securities, Pet Industry In-depth Report Series (1), https://pdf.dfcfw.com/pdf/H3_AP202408011638808311_1.pdf?1722521041000.pdf, visit time (2024, November, 24).

[12] Frost China, China Pet Products Consumption Trend Report 2023-2024, https://img.frostchina.com/attachment/17066304/8G4Xb43ah6Kbh86cbF188D.pdf, visit time (2024, November, 24).

[13] UN Comtrade Database, International Trade Statistics, https://comtradeplus.un.org/, visit time (2024, November, 24).

[14] Ken Benoit, Log-Log Model Explanation, https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf, visit time (2024, November, 24).

[15] General Administration of Customs of the People's Republic of China (GACC), Customs Data and Analysis, http://xian.customs.gov.cn/customs/syx/index.html, visit time (2024, November, 24).

XII. Appendix

Listing 1: The Stata code for problem 1 (4.3.2)

```
* Min-Max Normalized Data
summ Income_per_captia_Yuan
gen z_income_per_capita = (Income_per_captia_Yuan - r(min)) / (r(max) - r(min))
summ Urbanization_rate
gen z_urbanization_rate = (Urbanization_rate - r(min)) / (r(max) - r(min))
summ people_older_than_65_10_4_ren
gen z_people_older_than_65 = (people_older_than_65_10_4_ren - r(min)) / (r(max) -
    r(min))

* Perform Regression
regress total10_7 z_income_per_capita z_urbanization_rate z_people_older_than_65,
    vce(robust)
```

Listing 2: The Python code for problem 1 (4.4.2)

```
# Difference and ADF Test to get stable time series data and parameter d
# Draw ACF and PACF diagram to get parameter p and q
# ARIMA Model Training
model = ARIMA(df['Total'], order=(1, 1, 2))
fitted_model = model.fit()
# Predict future 3 years value
forecast = fitted_model.get_forecast(steps=3)
forecast_index = [2024, 2025, 2026]
forecast_values = forecast.predicted_mean
forecast_conf_int = forecast.conf_int()
print("Predicted Values:")
print(forecast_values)
```

Listing 3: The Python code for problem 2 (5.3.1)

```
# Difference and ADF Test to get stable time series data and parameter d
# Using AIC and BIC to get parameter p and q
# ARIMA Model Training
model = ARIMA(df["Value"], order=(0, 1, 2)).fit()
```

```
# Predict future 3 years value
forecast = model.forecast(steps=3)
forecast_years = [2024, 2025, 2026]
```

Listing 4: The Stata code for problem 3 (6.3.1)

```
* Generate log-transformed variables
gen ln_production_value = ln(production_value)
gen ln_market_size = ln(market_size)
gen ln_penetration_rate = ln(penetration_rate)
* Perform Regression
regress ln_production_value ln_market_size ln_penetration_rate, vce(robust)
```

Listing 5: The Python code for problem 3 (6.3.1)

```
# Using grid research to find parameter
# Parameters for Holt's method
alpha_lnM, beta_lnM = 0.9, 0.6 # for lnM
alpha_lnR, beta_lnR = 0.7, 0.7 # for lnR
# Initialize levels and trends for lnM and lnR
L_lnM, T_lnM = df["lnM"].iloc[-1], df["lnM"].iloc[-1] - df["lnM"].iloc[-2]
L_lnR, T_lnR = df["lnR"].iloc[-1], df["lnR"].iloc[-1] - df["lnR"].iloc[-2]
# Forecast future values for 2024, 2025, 2026
future_years = [2024, 2025, 2026]
lnM_forecast = []
lnR_forecast = []
for h in range(1, 4): # h = 1, 2, 3
    # Update level and trend for lnM
    L_lnM = alpha_lnM * L_lnM + (1 - alpha_lnM) * (L_lnM + T_lnM)
    T_lnM = beta_lnM * (L_lnM - df["lnM"].iloc[-1]) + (1 - beta_lnM) * T_lnM
    lnM_forecast.append(L_lnM + h * T_lnM)

    # Update level and trend for lnR
    L_lnR = alpha_lnR * L_lnR + (1 - alpha_lnR) * (L_lnR + T_lnR)
    T_lnR = beta_lnR * (L_lnR - df["lnR"].iloc[-1]) + (1 - beta_lnR) * T_lnR
    lnR_forecast.append(L_lnR + h * T_lnR)
# Combine the results
forecast_df = pd.DataFrame({
```

```python
    "Year": future_years,
    "LnM_Predicted": lnM_forecast,
    "LnR_Predicted": lnR_forecast,
})
forecast_df
```

Listing 6: The Python code for problem 3 (6.3.2)

```python
# Use Prophet to predict influencing factors individually
def predict_with_prophet(series, future_years):
    """Use Prophet to predict a single time series"""
    df = pd.DataFrame({'ds': data['Year'], 'y': series})
    model = Prophet()
    model.fit(df)
    future = pd.DataFrame({'ds': future_years})
    forecast = model.predict(future)
    return forecast[['ds', 'yhat']].rename(columns={'yhat': 'Predicted'})

future_years = [2024, 2025, 2026]

# Predict Thailand export values
thai_forecast = predict_with_prophet(data['Thai Export'], future_years)
thai_forecast = thai_forecast.rename(columns={'Predicted': 'Thai Export'})
# Predict USA export values
usa_forecast = predict_with_prophet(data['USA Export'], future_years)
usa_forecast = usa_forecast.rename(columns={'Predicted': 'USA Export'})
# Predict global market size
global_market_forecast = predict_with_prophet(data['Global Market Size'],
    future_years)
global_market_forecast = global_market_forecast.rename(columns={'Predicted':
    'Global Market Size'})
# Combine predictions for future influencing factors
future_factors = pd.concat([thai_forecast, usa_forecast['USA Export'],
    global_market_forecast['Global Market Size']], axis=1)
# Use predicted influencing factors to forecast China's export values
data_prophet = data[['Year', 'China Export']].rename(columns={'Year': 'ds',
    'China Export': 'y'})
data_prophet['Thai Export'] = data['Thai Export']
```

```
data_prophet['USA Export'] = data['USA Export']
data_prophet['Global Market Size'] = data['Global Market Size']
# Create Prophet model
model = Prophet()
model.add_regressor('Thai Export')
model.add_regressor('USA Export')
model.add_regressor('Global Market Size')
# Train the model
model.fit(data_prophet)
# Use future influencing factors to forecast China's export values
future_factors_prophet = future_factors.rename(columns={'ds': 'ds'})
forecast = model.predict(future_factors_prophet)
```

Listing 7: The Python code for problem 4 (7.2)

```
# Define policy implementation dates
policy_dates = ['2018-07', '2019-05', '2020-12', '2021-01', '2024-01']
policy_dates = [pd.to_datetime(date) for date in policy_dates]
# Add a policy variable (combine all policy dates into one variable)
data['Policy'] = data['Year and Month'].apply(lambda x: 1 if any(x >= date for
    date in policy_dates) else 0)
# Create a time trend variable
data['Time_Trend'] = (data['Year and Month'] - data['Year and
    Month'].min()).dt.days
# Add seasonal variables (month dummy variables)
data['Month'] = data['Year and Month'].dt.month
seasonal_dummies = pd.get_dummies(data['Month'], prefix='Month')
# Add a lagged variable for exports
data['Lag_Export'] = data['USD'].shift(1)
# Prepare the cleaned data for regression
data_clean = pd.concat([data, seasonal_dummies], axis=1).dropna()
data_clean['Intercept'] = 1
# Convert data types
data_clean['Policy'] = data_clean['Policy'].astype(int)
data_clean['Lag_Export'] = data_clean['Lag_Export'].astype(float)
for col in seasonal_dummies.columns:
    data_clean[col] = data_clean[col].astype(int)
# Set regression model variables
```

```
X = data_clean[['Intercept', 'Policy', 'Time_Trend', 'Lag_Export'] +
list(seasonal_dummies.columns)]
y = data_clean['USD']
# Apply the OLS regression model
model = sm.OLS(y, X).fit()
# Extract results
summary_table = {
    "R-squared": [model.rsquared],
    "Adj. R-squared": [model.rsquared_adj],
    "Prob (F-statistic)": [model.f_pvalue],
    "Coef (Policy)": [model.params['Policy']],
    "P>|t| (Policy)": [model.pvalues['Policy']]
}
# Convert results into a dataframe for display
result_df = pd.DataFrame(summary_table)
# Display results
print("DID model results:")
print(result_df)
```