Research on the Chinese Pet Market Based on Multivariate Prediction Model

Wenbaiyang Wu *

School of Electronic Information, Hangzhou Dianzi University, Hangzhou, China, 310000

* Corresponding Author Email: fuchenrx@gmail.com

Abstract. China's pet industry is experiencing robust growth, prompting this research to develop a series of innovative forecasting models that not only predict the future number of pets and the size of the pet food market in China and globally but also offer practical suggestions for industry expansion. A key innovation lies in the integration of the Grey Model GM(1,1) to forecast the future population size of pets in China with high precision, coupled with correlation analysis to pinpoint the critical factors driving industry development. Furthermore, the study employs the ARIMA model to predict global pet food demand trends and estimates China's pet food production, demonstrating a holistic approach. An additional novelty is the use of a multiple linear regression model to forecast China's pet food export volume, considering various domestic and international factors, thereby providing a comprehensive understanding and prediction of the pet industry's future trajectory.

Keywords: Grey Model, ARIMA, Multiple Linear Regression, Pet Market.

1. Introduction

Driven by a combination of robust economic development and evolving consumer preferences, China's pet industry has witnessed unprecedented growth in recent years, particularly in the realms of pet food and healthcare. This surge has been further fueled by the active promotion of pet welfare by organizations like the China Small Animal Protection Association and the influx of international brands into the market. Existing literature on the subject has primarily focused on quantifying the industry's growth and identifying key market players, yet often neglects to account for the intricate interplay of external factors and regional nuances that significantly influence consumer behavior and market dynamics [1-2]. To bridge this gap, our study undertakes a comprehensive literature review that integrates insights from diverse sources, including academic research, industry reports, and consumer surveys. By employing advanced modeling techniques that incorporate these multifaceted considerations, our analysis offers a more nuanced and accurate forecast of market shifts over the next three years. This approach not only assesses the internal drivers of growth within the Chinese pet industry but also evaluates global pet trends and future food demand, providing a holistic view of China's industry status. Through a meticulous demand analysis that takes into account both domestic and international markets, our study predicts production and export trends, offering valuable insights for stakeholders seeking to navigate and capitalize on the rapid expansion of this dynamic and ever-evolving sector.

2. Forecast and Correlation Analysis of Pet Population in China

2.1. Pet Population Forecast Based on Grey Model

The white paper on the Chinese pet industry and other content provides data references for research in this direction. After analyzing the existing pet data, this study concludes that the overall pet population in China remained stable from 2019 to 2023, with slight fluctuations. The pet population data for these five years meet the requirements for the time series prediction. Furthermore, considering the issues of small and incomplete data, this study employs a grey model to forecast the pet population in China for the next three years. Given that dogs and cats account for the majority of pets, they are selected as research objects.

$GM(1.1)$The discrete form of the prediction formula is as follows [3]:

$$\Delta^{(1)}\left(x^{(1)}(k+1)+a(k+1)\right)=u, \tag{1}$$

$$\hat{x}^{(1)}(k+1)=\left[x^{(1)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k}+\frac{\hat{u}}{\hat{a}}, \tag{2}$$

Based on the derivative definition of the discrete form prediction formula, when the time change is sufficiently small, the derivation result is as follows:

$$x(t+1)-x(t)=\frac{\Delta x}{\Delta t}, \tag{3}$$

$$\frac{\Delta x}{\Delta t}=x(k+1)-x(k)=\Delta^{(1)}(x(k+1)), \tag{4}$$

After organizing the above conclusions, the final derivation result is as follows:

$$(k+1)=a\left[-\frac{1}{2}\left(x^{(1)}(k)\right)+x^{(1)}(k+1)\right]+u, \tag{5}$$

By substituting the values into the formula for calculation, the study obtain the forecast results for the number of pet dogs and cats in China over the next three years in Figure 1 [4].

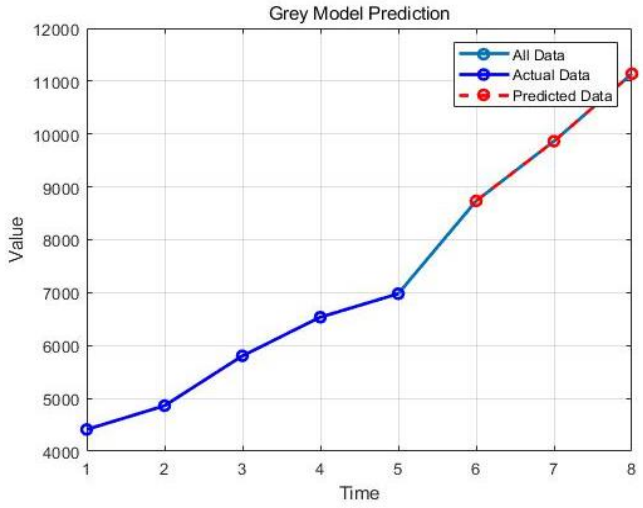

Figure 1. Grey model prediction

To further verify the accuracy of the Grey Model, the study employ the posterior error test method. The results are then subjected to cumulative subtraction transformation. The formulas for calculating the residuals and the posterior error ratio are as follows:

$$e(k)=x^{(0)}(k)-\hat{x}^{(0)}(k), \tag{6}$$

$$S_1^2=\frac{1}{n}\sum_{k=1}^{n}[x^{(0)}(k)-\overline{x}]^2, \tag{7}$$

$$S_2^2=\frac{1}{n}\sum_{k=1}^{n}[e(k)-\overline{e}]^2, \tag{8}$$

$$C=S_2/S_1, \tag{9}$$

The posterior error ratio reflects the accuracy of the model. For the forecast results of the number of pet dogs and cats in China over the next three years, the study obtained a mean squared error ratio of less than 0.35, indicating that the model has good accuracy.

2.2. Correlation Analysis of the Development of China's Pet Industry

Based on the analysis of China's pet population from 2019 to 2023, this study examines the development of China's pet industry over the past five years, incorporating real-world scenarios and statistical data on pet ownership rates in Chinese households from relevant institutions. To explore the relationship between these data and the pet population, this study employs correlation analysis.

Firstly, a normality test is conducted using the Shapiro-Wilk Test. The results of the test indicate that the relevant data from the past five years follow a normal distribution, allowing for further calculations, as shown in Table.1.

Table.1. Shapiro-Wilk test results

Shapiro-Wilk Test Results	Cats	Dogs
W Statistic	0.94	0.71
P Statistic	0.89	0.39

Correlation test for correlation analysis, the study obtains the Pearson correlation coefficient [5-6]. The results indicate that there is a significant positive correlation between the size of China's pet population and the pet ownership rate in China, as shown in Table.2.

Table.2. Pearson correlation of cats and dogs

	Cats	Dogs
Pearson Correlation	0.873	0.944

Based on the above results, this study identifies the following characteristics of the development of China's pet industry:

(1) The number of pet cats is showing steady growth. Due to the relatively low difficulty in caring for pet cats and their weaker territorial instincts, they are favored by more people, especially young people. As a result, related industries are gradually becoming mainstream.

(2) The number of pet dogs remains stable. This is because pet dogs are more difficult to care for and have stronger territorial instincts, leading to a stable pet dog population.

(3) The pet ownership rate in Chinese households is relatively low, with an overall rate below 20%, indicating significant room for growth.

3. Pet Food Demand Forecasting Based on ARIMA

After analyzing existing global data on pet dogs and cats, this study concludes that the overall number of pets worldwide remained stable from 2019 to 2023, with some variations across different countries. Some data showed significant fluctuations around the COVID-19 pandemic in 2020. Excluding the impact of sporadic events, this study believes that the pet population over the past five years meets the requirements for time series forecasting. Based on the Autoregressive Integrated Moving Average (ARIMA) model, this study establishes a pet population forecasting model to predict global pet food demand for the next three years [7-9]. The formula for ARIMA is as follows:

$$y_t = \mu + \sum_{i=1}^{p} r_i y_{t-i} + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}, \tag{10}$$

Considering the requirements of the ARIMA model, the study first conduct a stationarity test to ensure that the mean and variance of the data remain constant over time. We calculate the mean and variance of the global pet population over the past five years and perform an autocorrelation analysis to assess stationarity. On this basis, we use the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The results of the ADF test and the KPSS test are obtained, as shown in Table.3.

Table.3. ADF inspection and KPSS inspection

	Cats	Dogs
Stability	Stationary Sequence	Stationary Sequence

The test results indicate that the global pet population has shown a stable trend over the past five years. Therefore, the differencing order d in the time series is set to 0.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used to compare the model fitting performance of different combinations of p and q. By plotting a heatmap

of AIC and BIC values for various p and q combinations, the study can facilitate the selection of the optimal model parameters and obtain the final results.The relevant results are shown in Figure 2.

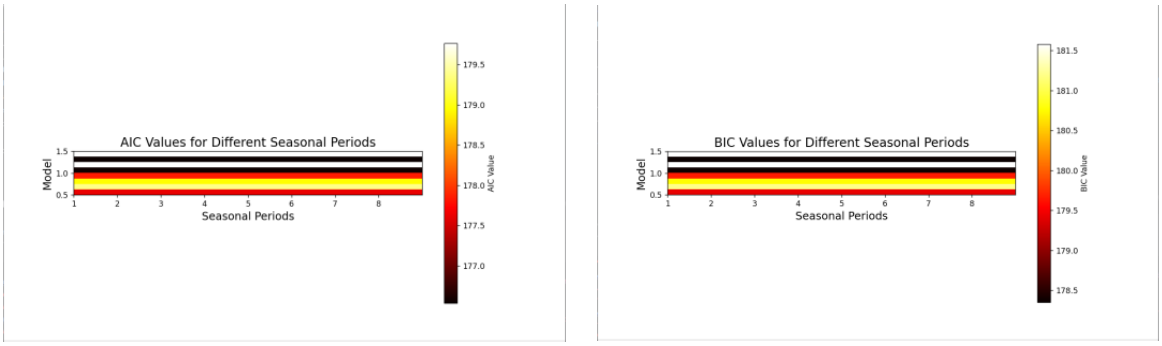

Figure 2. Heat map of AIC and BIC values

Based on the analysis of the heatmap results, the study has determined that the number of autoregressive terms p=1 and the number of moving average terms q=1.

Combining the existing conclusions, this study conducts a time series forecast for the global number of pet dogs and cats, resulting in the predicted numbers for the next three years, as shown in Figure.3.

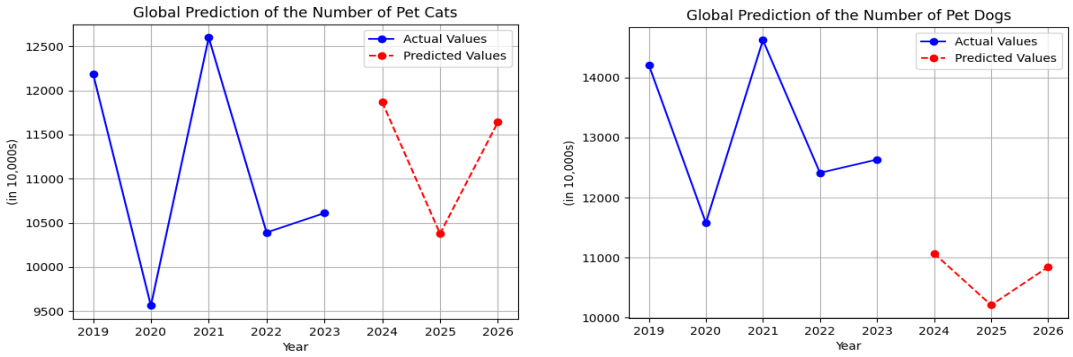

Figure 3. Global prediction of the number of pet cats and dogs

Further analyzing the global demand for pet dog and cat food over the next three years, it is generally assumed that the quantity of pet food consumed is directly proportional to the number of pet dogs and cats. The proportional relationship can be expressed as follows:

Where a and b represent the quantity of pet dog and cat food and the number of pet dogs and cats, respectively. This indicates that the ARIMA model conditions determined for the global number of pet dogs and cats are also applicable to the quantity of pet dog and cat food.

Based on the past five years of global pet dog and cat food consumption data, this study conducts a time series forecast for global pet dog and cat food consumption with constant values of p, q, and d, ultimately obtaining the forecasted results for the next three years in Figure 4.

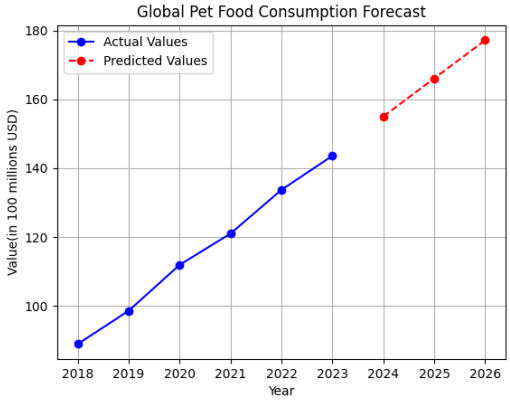

Figure 4. Global pet food consumption forecast

4. Prediction of China's Pet Food Production and Export Volume

4.1. ARIMA Forecast for China's Pet Food Production

After analyzing the existing data on China's pet food production, this study concludes that the overall production of pet food in China from 2019 to 2023 has been in a significant growth state. The production levels in 2021 and 2022 tended to stabilize before experiencing explosive growth. Excluding the impact of sporadic events, this study believes that the pet food production data for these five years in China meets the requirements for time series forecasting. Therefore, the ARIMA model is continued to be used to establish a forecasting model for China's pet food production [10].

Repeating the ARIMA derivation process, the study obtained the values of p=1, q=1, and d=1. The results indicate that there has been some fluctuation in China's pet food production data over the past five years, and a stationary series can be obtained through first-order differencing. With the known values of p, q, and d, time series forecasting was conducted on China's pet food production data, ultimately yielding the forecasted results for China's pet food production over the next three years in Figure 5.

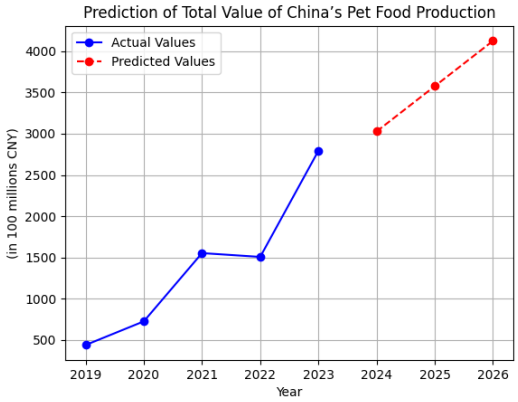

Figure 5. Prediction of total value of China's pet food production

4.2. Prediction Model for China's Pet Food Export Volume

Although the ARIMA model can effectively achieve time series forecasting for pet food production, the forecasting of pet food export volume cannot be directly solved using the same model as pet food production. This is because pet food export volume is not only related to its previous export volumes but is also influenced by other factors such as the current year's pet food production, global market demand, and global trade tariffs. Therefore, this study further employs multiple linear regression to predict the scale of China's pet food exports [11].

Multiple linear regression incorporates multiple independent variables to more comprehensively reflect the relationships within the system. The ideal formula for multiple linear regression is as follows:

$$y = a^T x + b, \tag{11}$$

Where a represents the feature weights and b represents the model bias.

For the pet food export data, this study assumes that there are three variables: pet food production, global market demand, and global trade tariffs. Thus, the formula for pet food export data is as follows:

$$y = \beta_1 x + \beta_2 y + \beta_3 z + b, \tag{12}$$

Introduce Mean Squared Error (MSE) to refine the model's performance. The final result is obtained when the MSE is minimized. The formula for MSE is as follows:

$$E(\hat{k}^*) = arg_{(\hat{k})}\min (y - X\hat{k})^T (y - X\hat{k}), \tag{13}$$

By applying the method of least squares and deriving the mean squared error function, the study obtain the optimized solution formula as follows:

$$\frac{\delta}{\delta \hat{k}} E(\hat{k}) = 2X^T(X\hat{k} - y) = 0, \tag{14}$$

Since it is known that the feature weights are definitely non-zero, the corresponding multiple linear regression matrix must be a full-rank matrix [12]. For a full-rank matrix, the multiple linear regression formula is derived as follows:

$$f(\hat{y}_i) = \hat{x}_i(X^T X)^{-1} X^T y \tag{15}$$

Taking into account market growth, based on the global pet food market sales data from 2019 to 2024, the average annual growth rate in recent years is 5.23%. Therefore, the study can set a market growth constraint.

Considering economic policies, the current tariff on pet food in China is 4%, allowing us to establish an economic policy constraint.

Combining these relevant constraints, the study ultimately obtains the forecasted results for China's pet food exports over the next three years in Figure 6 and Table.4.

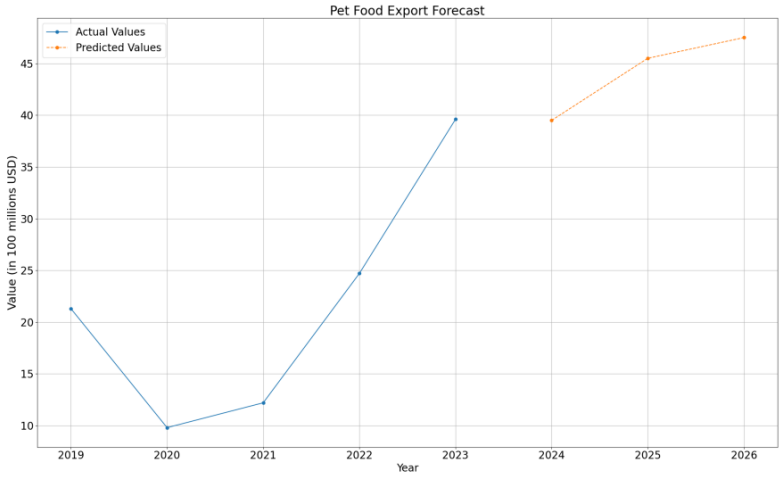

Figure 6. Pet food export forecast

Table.4. Pet Food Export Forecast Value（in 100 million USD）

2024	2025	2026
35.51	41.56	43.53

5. Conclusions

This study employs both the Grey Model GM(1,1) and the ARIMA model to predict the development trends of China's pet industry, covering aspects such as pet population, pet food demand, and pet food production. The findings indicate a steady growth in China's pet population over the next three years, along with significant increases in global pet food demand and China's pet food production. Additionally, the study uses a multiple linear regression model to assess China's pet food export volume for the coming three years. These models are renowned for their stability and efficiency in analyzing limited datasets, providing valuable insights for the related industries and offering guidance for stakeholders to plan for future industry expansion. Looking forward, as the pet industry continues to evolve, it will be crucial to refine and adapt these models to incorporate new data sources and emerging trends. By leveraging advanced analytics and real-time data, stakeholders can gain a more comprehensive understanding of the market dynamics and make more informed decisions to drive sustainable growth and innovation in China's pet industry. This ongoing refinement and adaptation will be key to ensuring that the industry remains resilient and continues to thrive in the face of changing consumer preferences and global market conditions.

References

[1] Kwak, M. K., Cha, S. S. A study on the selection attributes affecting pet food purchase: After COVID-19 pandemic[J]. International Journal of Food Properties, 2021, 24(1): 291-303.

[2] Liu, Z. Unlocking consumer choices in the digital economy: Exploring factors influencing online and offline purchases in the emerging pet food market[J]. Journal of the Knowledge Economy, 2023, 15(3): 10174-10199.

[3] Bi, Z., Du, Y., Hao, R. Predict the price change over time based on GM (1,1) model[J]. Financial Engineering and Risk Management, 2022, 5(2).

[4] Li, Y. S., Zhang, L. G., Song, T. A study on forecasting the popularity of Chinese national brands using a grey-weighted Markov model[J]. Electronic Commerce Research, 2024, [prepublish]: 1-17.

[5] Yu, H., Hutson, A. D. Inferential procedures based on the weighted Pearson correlation coefficient test statistic[J]. Journal of Applied Statistics, 2024, 51(3): 481-496.

[6] Liu, P., Wang, S., Zhao, P. Robust estimation and test for Pearson's correlation coefficient[J]. Random Matrices: Theory and Applications, 2024.

[7] Kushnir, M., Tokarieva, K. A generalization of the ARIMA model to the nonlinear and continuous cases[J]. Cybernetics and Systems Analysis, 2023, 59(6): 900-909.

[8] Li, X., Liu, X., Zhong, Y. Research on sales and pricing optimization of vegetable category based on ARIMA model and exponential smoothing method[J]. Information Systems and Economics, 2024, 5(3).

[9] Ondieki, R., Waititu, H., Nyabwanga, R. N. Modeling and forecasting high frequency currency exchange rates: A comparative study of ARIMA, ANN, and hybrid ARIMA-ANN models[J]. Asian Journal of Probability and Statistics, 2024, 26(10): 32-45.

[10] Quang, P. D., Oanh, N. T., Hao, L. T. P., Duong, P. H., Linh, L. K., Ngan, N. T. K. Estimating and forecasting bitcoin daily prices using ARIMA-GARCH models[J]. Business Analyst Journal, 2024, 45(1): 11-23.

[11] Yan, Z. Real estate price model based on multiple linear regression and analysis of influencing factors of regional economic vitality development[J]. Academic Journal of Business & Management, 2021, 3(4).

[12] Peijiang, C., Xueyin, Y. Study on the evaluation method of blended learning effect based on multiple linear regression analysis[J]. International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), 2023, 18(1): 1-15.