Animal Economic: The Evidence from China

S. Tan*, M. Feng, J. Xu†

2024.11 New!

Abstract

The study systematically organizes data on pet-related trends from the past five years, along with over 20 influencing factors, and employs these datasets to construct forecasts for the subsequent three years. That is possible for solving these question which restate in the Chapter 2.1.

In Chapter 1 and Chapter 2, we clarified the background and explored the significance of the pet economics. Then the workflow is shown in the form of a diagram. To address the research questions, Chapter 3 develops the necessary models, which are further elaborated in Chapter 4 with detailed methodologies and solution approaches. Chapter 5 synthesizes and presents the findings derived from these analyses.

Chapters 3 and 4 constitute the core of this study. To address the research objectives, multiple models were developed and applied. From the perspective of parameter selection, the Ordinary Least Squares (OLS) method was utilized to identify significant variables based on their statistical significance. For predictive purposes, the Grey Forecasting Model exhibited robust performance in handling small sample sizes ranging from 4 to 10 observations. Furthermore, the Random Forest model offered a viable solution for pet-related predictions, leveraging the bootstrap methodology to enable effective model training and enhance predictive accuracy.

The findings indicate that over the next three years, the population of cats is expected to continue growing, while the population of dogs will likely transition from a decline to a growth trend. These results, however, are not considered optimal, as income effect and substitution effect models could provide a more robust framework; yet, the substitution effect model could not be fully utilized due to data limitations. To address this, the approach involved using OLS to identify effective factors, applying a Grey Forecasting Model to predict their values over the next three years, and training a Random Forest model with bootstrap methods on five years of historical data for robustness. The Grey Forecasting results were then used as inputs for the Random Forest to generate final predictions, with the same methodology applied to Questions 2 and 3. To address Question 4, tariff models and elasticity models were employed to provide explanations and insights.

Keywords: Animal Economic, Grey Prediction Model, Random Forest, Income and Substitution Effect

*Corresponding author: sepinetam@gmail.com

†Shanghai University, Shanghai, China

1 Introduction

1.1 Background

In recent years, with the improvement in living standards and the acceleration of social pace, the "popularity" of the pet industry has steadily risen, with the purchase of pets continuing to increase. An increasing number of young people are choosing to adopt cats, dogs, and other pets, primarily to alleviate the pressures of daily life and fulfill emotional companionship needs. Meanwhile, related products and services in the pet industry have rapidly developed, ranging from pet food and toys to pet care and healthcare services, all of which continuously meet the growing market demand, driving the vigorous development of the pet economy.

- **Domestic and International Data Changes:** According to the given data, the number of pet cats in China showed continuous growth from 2019 to 2023. Meanwhile, the number of pet dogs experienced a noticeable decline during the COVID-19 pandemic in 2020, followed by irregular fluctuations in subsequent years.In the United States, the population of pet cats hit its lowest point in 2020, rebounded in 2021, but stabilized at a relatively low level in 2022 and 2023. In European countries such as France, the number of pet cats steadily increased year by year from 2019 to 2023. In Germany, however, the population of pet cats declined in 2021 before beginning a slight recovery in 2022.

- **Factors Influencing Change:** The fluctuations in pet populations are closely tied to the development of related industries, such as pet food, healthcare, and services, with varying degrees of influence. The price, variety, and availability of pet food significantly shape consumer decisions, thereby exerting a considerable impact on changes in pet numbers. Additionally, market prices for pet healthcare and grooming services play a crucial role in influencing consumer preferences. Furthermore, the substitution effect between cats and dogs in terms of the utility perceived by pet owners also affects their choices.

- **Market Trends:** Based on the development of consumption patterns in the pet industry and data projections, along with the shifts in social demographics and the influences of Covid-19, it is possible to preliminarily forecast the future trends in the number of cats and dogs both domestically and globally. Additionally, changes in demand, supply, and price elasticity for pet food across various countries and globally can be used to infer and predict the export share of China's pet food market.

1.2 Research Significance

This study establishes a mathematical model to predict the changes in the population of cats and dogs both domestically and internationally, as well as the trends in the production and export of China's pet food industry. Additionally, it discusses the development strategies of China's pet food industry in response to the implementation of relevant economic policies abroad. This research holds significant theoretical importance for promoting the development of the global pet economy.

- **Developing strategies for enterprises:** By analyzing the trend of changes in the number of cats and dogs, we provide data support and market insights for the future development of the pet industry. Enterprises will be able to better understand consumer demand dynamics, thereby formulating more precise product positioning and marketing strategies. It also contributes to the healthy development of related industry chains.

- **Suggestions for Optimizing Government Management:** We provide theoretical support for the government in areas such as pet management, animal protection, and public health. With the increasing of the pets' market, data-driven predictions on how to reasonably regulate pet populations, enhance pet welfare, and reduce the impact of pets on society can help formulate more effective policies and laws for the harmonious coexistence of pets and the community.

- **Guidance for Multinational Enterprises:** The development of the global pet economy is closely tied to the economic, cultural, and policy frameworks of various countries. Predictions based on the population of common pet types, such as cats and dogs, help reveal the differences and trends in pet economy consumption across different countries and regions, offering multinational enterprises opportunities to grasp the scale of foreign markets.

1.3 Data Collection and Processing

Official statistics reveal that the populations of cats and dogs demonstrate distinct annual variation trends, driven by a range of influencing factors. A comprehensive analysis was conducted to identify and select key indicators that potentially affect these fluctuations. At the national level, the temporal trends of these indicators were systematically analyzed and correlated with the observed changes in cat and dog populations to uncover underlying relationships.

To broaden the scope of the analysis, the study was extended to a global context, enabling comparative evaluations of region-specific influencing factors and variation patterns. This global perspective provided deeper insights into the dynamics of pet ownership trends across diverse regions. Based on these analytical outcomes, projections were made regarding the development trajectory of the pet food industry in China and globally, with particular emphasis on production scale, market demand, and export volume dynamics.

The collected data revealed the presence of missing values, which were addressed through the application of an interpolation algorithm. Specifically, when both preceding and succeeding data points were available, the missing values were estimated using their arithmetic mean. In cases where only one adjacent data point was available, an arithmetic progression method was applied to complete the data.

To prevent the loss of validity in significance testing caused by excessively large data values, normalization was performed to eliminate the influence of dimensionality. This process ensured that the data values were on a comparable scale, thereby enhancing the robustness of subsequent statistical analyses.

In the subsequent parameter selection process, parameters with significant results from the OLS regression were chosen for future predictions. This approach ensures that only statistically robust variables contribute to the predictive analysis.

2 Restatement of the Problem

2.1 Research Problem

To ensure a more structured and precise understanding of the task, we restate and frame the problem as follows:

1) We aim to develop a model to identify the effective factors which affect the number of pets significantly. And we construct a framework based on it which is capable of accurately predicting the data for these factors over the next three years.

2) Up to now, we have identified and predicted the effective factors. Building on this foundation, we aim to construct a parameterized model to predict pet populations. This model will serve as a tool for projecting future trends in pet ownership with a focus on methodological rigor and accuracy.

3) However, the populations of cats and dogs appear to exhibit a certain degree of correlation. We seek to develop a model to explain this phenomenon, providing a rigorous framework for analyzing the underlying relationship between their quantities.

4) Then, we could build a model to construct the future trend of global pet markets. We attempt to use elasticity concepts from economics to analyze the global market.

2.2 Brief Work Flow

To make the process more intuitive, workflow is constructed as illustrated in Figure 1.

We collected data on potential explanatory variables and their associated measurements.Subsequently, the dataset was systematically organized and preprocessed to enhance its quality and suitability for further analysis.

To identify the most relevant factors, we conducted validity tests on the parameters and evaluated the statistical significance of regression coefficients using the Ordinary Least Squares (OLS) method. This rigorous analysis enabled us to determine the effective parameters for subsequent modeling.

Following the identification of effective parameters, a grey prediction model—a methodology specifically designed for time series forecasting under uncertainty—was employed to project these parameters over a 3-year horizon. The forecasted parameters were then integrated into the random forest model for advanced predictive analyses.

Simultaneously, a random forest model was trained using the historical data of the effective parameters. This model was subsequently utilized to predict future

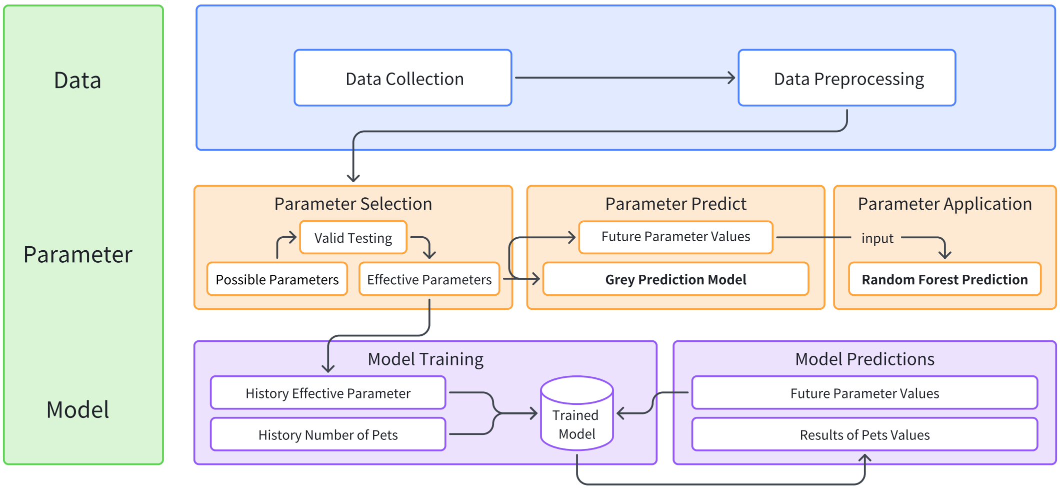

Figure 1: Work Flow

population dynamics of cats and dogs, offering insights into their expected trends over time.

3 Model Building and Explaining

3.1 Basic Model Overview

3.1.1 Grey Prediction Model

Deng (1982) pioneered the grey system theory designed to analyze systems characterized by both known and uncertain parameters. Grey Prediction Model (GM) stands out as particularly efficacious, which performs particularly well in small sample and fuzzy systems.

In this problem, there are manifold factors which could affect the number of pets, but since the sample size is only five years, which is a limited data samples, we consider using the GM Model to predict the value of these factors in the next three years.

3.1.2 OLS

OLS (Ordinary Least Squares) serves as a cornerstone methodology in statistical analysis for parameter estimation in linear regression models (Woodridge, 2016). This approach derives optimal regression coefficients through the minimization of squared residuals between model predictions and empirical observations.

Nevertheless, to maintain statistical validity and ensure asymptotic properties, OLS analysis conventionally necessitates a minimum threshold of 30 observations.

And at the same time, the value of parameters would affect the number of pets; however, the number of pets also affects the value of these parameters, which is described as reverse causality. Generally, for time series data, reverse causality is addressed by using lagged variables in the regression framework. However, in this study, the limited sample size restricts our ability to incorporate lagged observations effectively, leaving the issue unresolved.

3.1.3 Random Forest Algorithm

Random Forest algorithm (Breiman, 2001) is a robust ensemble learning methodology designed to integrate multiple decision trees through bootstrapping and random feature selection. Random Forest stands out as particularly advantageous, demonstrating remarkable effectiveness in handling high-dimensional data and mitigating overfitting issues.

Caused by the unavoidable limitations in historical data coverage, as the available records only extend across a five-year period, which naturally precludes the implementation of linear regression analysis (OLS), we adopted the Random Forest methodology to forecast pet population trends.

Although we cannot directly employ OLS for population prediction due to sample size constraints, the OLS analysis remains valuable for identifying statistically significant factors. These identified key variables will subsequently be incorporated into the Random Forest model, thereby mitigating potential overfitting bias through focused feature selection.

Given that our current sample size falls below the optimal threshold required for Random Forest implementation, we employed bootstrap resampling techniques to augment our dataset to meet the requisite training standards.

3.1.4 Bootstrap Method

The bootstrap methodology, introduced in the seminal work of Efron (1979), represents a revolutionary advancement in statistical resampling techniques. This robust approach circumvents the limitations of traditional parametric methods by generating multiple resampled datasets through random sampling with replacement, thereby enabling reliable statistical inference even with modest original sample sizes. The method's particular strength lies in its ability to estimate sampling distributions without imposing restrictive distributional assumptions, while maintaining the fundamental statistical properties of the original dataset.

In our context, given the temporal limitations of our dataset, we employed the bootstrap resampling approach to address the challenge of insufficient sample size, enabling effective training of the Random Forest model while preserving the statistical characteristics of the original data.

3.2 Symbols and Assumptions

3.2.1 Symbols with Definitions

Here are some definition of the symbols from Table 1.

Table 1: Symbols Decisionion

Symbols	Description
Cat_t	the number of cats at year t
Dog_t	the number of dogs at year t
$X_{i,t}$	the value of the i-th factor at year t
β_i	the regression coefficient of the i-th factor
α	the fixed effects in OLS
ϵ	random error term
$Interaction_{i,j}$	the interaction of j on i
Pay_i	the willingness to pay for pet i
$Food_i$	the cost on food for pet i
Med_i	the medical expenses for pet i
$Serve_i$	the serve fee for pet i

3.2.2 Assumptions

To maintain logical coherence with the empirical evidence, we based our analysis on the following assumptions.

The primary factors influencing pet population dynamics can be categorized into four main areas: willingness to pay, cost of food, medical expenses, and pet service fees (iResearch, 2024). Therefore, these four aspects serve as the primary considerations.

Despite the fluctuations in exchange rates caused by the Sino-US trade friction in recent years, we still define the exchange rate of CNY to the USD as 7.2, because these fluctuations will not have a particularly significant impact on our forecast results and statistics.

3.3 Random Forest Prediction Based on GM

3.3.1 OLS Test main parameters

As we said, we could not apply OLS to directly predict the number of pets due to the sample sizes, but we could detect the main parameters by OLS. Therefore, we set the OLS model as following:

$$Cat_t = \beta_0 + \sum_i \beta_i \cdot X_{i,t} + Interaction_{cat,dog} \cdot Dog_t + \alpha_{Cat} + \epsilon$$

$$Dog_t = \beta_0 + \sum_i \beta_i \cdot X_{i,t} + Interaction_{dog,cat} \cdot Cat_t + \alpha_{Dog} + \epsilon$$

Then we could capture the effective parameters of the OLS model, focusing on those coefficients that provide meaningful insights, where "effective parameters" refer to those achieving statistical significance at the 1% level, as determined by standard regression significance tests.

3.3.2 GM Prediction

Having identified the effective parameters, we could use GM to forecast their trajectories over a three-year horizon. The core mechanism of GM lies in its ability to transform an original data series into a cumulative sequence, from which it establishes a differential equation to describe the underlying trends.

In this study, we implement a GM Model that initializes data by creating a cumulative sum sequence. The parameters of the model, denoted as a and b, are estimated using the least squares method applied to the generated background sequence. These parameters capture the intrinsic dynamics of the system, allowing the model to predict future values by solving the established differential equation iteratively. The implementation also includes visualization tools for comparing the original data and extended predictions, ensuring interpretability of the results.

The route for the GM model is as follows:

1) Data Initialization and Accumulation

To begin, initialize the original data sequence $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}$ and generate the accumulated sequence $X^{(1)}$ to smooth out variations and highlight the underlying trend:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n.$$

2) Differencing and Reconstruction

Next, construct a first-order grey differential equation based on $X^{(1)}$:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b.$$

To discretize this equation, a background sequence $z^{(1)}(k)$ is introduced:

$$z^{(1)}(k) = 0.5 \left(x^{(1)}(k) + x^{(1)}(k-1) \right), \quad k = 2, 3, \ldots, n.$$

This step provides a means to approximate the system behavior for subsequent analysis.

3) Parameter Estimation

Using the least squares method, the parameters a and b are calculated by solving the discrete form of the grey differential equation:

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \ldots, n.$$

In matrix form, the solution is expressed as:

$$\begin{bmatrix} a \\ b \end{bmatrix} = (B^\top B)^{-1} B^\top Y,$$

where B and Y represent the background matrix and original sequence, respectively.

4) Prediction

Finally, use the estimated parameters to forecast future values. The cumulative sequence $X^{(1)}$ is predicted using the time response function:

$$\hat{x}^{(1)}(k) = \left(x^{(1)}(1) - \frac{b}{a}\right) e^{-a(k-1)} + \frac{b}{a}.$$

Restore the original sequence using the inverse operation:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1), \quad k = 2, 3, \ldots.$$

This streamlined process facilitates accurate forecasting of future data points, utilizing the GM model's robustness in handling small datasets and uncertain environments.

In our analysis, the GM model should effectively capture the trends of the identified effective parameters, delivering reliable predictions within the specified temporal horizon. These forecasts provide valuable insights into potential future trajectories and support more informed decision-making in the context of this research.

3.3.3 Random Forest Model Predictions

Based on the GM model, we have packaged the future trends for these effective parameters, which would be implemented to predict the number of pets.

To predict the future number of pets, we trained a random forest model using the effective parameters as input features and the current number of pets as the target variable y_t. Mathematically, the training dataset is represented as:

$$\{(\mathbf{X}_t, y_t) : t = 1, 2, \ldots, n\},$$

where $\mathbf{X}_t = \{x_{1,t}, x_{2,t}, \ldots, x_{p,t}\}$ denotes the effective parameters at time t. The random forest model consists of an ensemble of M decision trees, each trained on a bootstrapped subset of the data. The prediction of a single tree T_m for y_t is given by:

$$\hat{y}_t^{(m)} = T_m(\mathbf{X}_t).$$

The final output of the random forest is obtained by aggregating the predictions of all decision trees:

$$\hat{y}_t = \frac{1}{M} \sum_{m=1}^{M} \hat{y}_t^{(m)}.$$

Random forest effectively captures nonlinear relationships and interactions among the effective parameters, providing robust performance even in the presence of complex dynamics.

Using the GM-predicted values of the effective parameters $\hat{\mathbf{X}}_{t+k}$, we forecasted the future number of pets $\hat{y}_{t+k}$ over a forecast horizon h. Specifically, the GM model predicts the values of the effective parameters at future time steps $t+k$ as:

$$\hat{\mathbf{X}}_{t+k} = \mathbf{f}_{\mathrm{GM}}(\mathbf{X}_t, k), \quad k = 1, 2, \ldots, h,$$

where $\mathbf{f}_{\mathrm{GM}}(\cdot)$ represents the GM time response function. These predicted values $\hat{\mathbf{X}}_{t+k}$ are then used as input features for the random forest model to forecast the future number of pets:

$$\hat{y}_{t+k} = \frac{1}{M}\sum_{m=1}^{M} T_m(\hat{\mathbf{X}}_{t+k}), \quad k = 1, 2, \ldots, h.$$

To enhance the reliability of these predictions, we incorporated a time series perspective by analyzing historical trends in pet numbers. Assuming that the number of pets follows an autoregressive process of order p (AR(p)), the historical trend is modeled as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where ϕ_i are the autoregressive coefficients and ϵ_t is the error term. By comparing the random forest predictions with this time series analysis, we validated the consistency of the projected pet numbers with observed historical patterns.

This hybrid approach combines the GM model's ability to forecast parameter dynamics with the random forest's capability to model complex relationships. By using the GM-predicted effective parameters as inputs for the random forest, this framework ensures that both historical trends and parameter-driven influences are incorporated into the predictive process.

3.4 Income Effect and Substitution Effect Model

3.4.1 Inevitable

It looks like an economic model (Mankiw, 2016), and under the framework of economic modeling, adherence to the fundamental principles of supply and demand, as emphasized in microeconomic theory, is essential. To construct a model that is analytically tractable and theoretically coherent, we introduce the following assumptions:

We assume that each individual derives varying levels of utility from different pets. Generally, the overall utility for a specific type of pet follows a normal distribution.

3.4.2 Budget Constraints

From an economic sight, people can generally be categorized into two groups: those who are constrained by a budget and those without strict budget constraints. Typically, it is estimated that individuals who own both cats and dogs constitute only a quarter of those who own either a cat or a dog (iResearch, 2024). Therefore, budget constraints can be considered one of the primary factors influencing this choice, reflecting the income effect.

Considering that the disposable income of individuals without budget constraints is increasing annually, along with their growing population (Yi and Zhou, 2018), we simplify the model by focusing solely on individuals with budget constraints when analyzing the income effect (budget constraints). This approach ensures both clarity and professionalism in the model.

For individuals without budget constraints, this issue can be effectively analyzed from the perspective of utility. In this framework, we assume that each individual derives different levels of utility from different pets, which, at the aggregate level, follow a normal distribution. For a given individual, the difference in utility derived from owning a cat versus a dog may be substantial. Consequently, we posit that the choices of individuals constrained by budgets exhibit a high degree of subjectivity. However, at the population level, these preferences still conform to a normal distribution.

3.4.3 Pet Substitution Effect

For the same individuals with budget constraints, there exists a subset whose utility derived from different pets is relatively equal. In such cases, their primary decision variable becomes cost. When the cost of owning a dog is higher, they are more likely to choose a cat; inversely, if the situation is reversed, they will opt for a dog.

The underlying mechanism is that each consumer's objective function is

$$i^* = arg \max_{i \in \{\text{Cat, Dog}\}} U_i - \text{Cost}_i$$

i.e.

$$\Delta = (U_{cat} - Cost_{cat}) - (U_{dog} - Cost_{dog})$$

What consumers should consider is the sign of Δ when deciding whether to adopt a cat or a dog. For individuals whose utility from cats and dogs is similar, cost becomes the primary consideration, meaning they are more inclined to choose the pet with the lower associated cost.

3.5 International Market

3.5.1 Tariff Model

An attempt is made to develop a tariff model that approximates the real world as closely as possible. This model involves three key agents: exporters, importers, and the government of the importing country. The objective of both importers and exporters is to maximize their own profits, while the government's primary goals are economic growth and full employment.

We attempt to describe this framework within a game-theoretic context and will compute the Nash equilibrium of the model in subsequent sections.

The objective of both importers and exporters is to maximize their own profits, while the government's primary goals are economic growth and full employment. The strategic space of exporters includes deciding whether to export and selecting an appropriate export price. The strategic space of importers involves deciding whether to accept the offered price. The strategic space of the government lies in setting the tariff T.

3.5.2 Elasticity Measure Model

Grounded in economic principles, this study aims to calculate import and export volumes through the estimation of price elasticity of demand.

We can define price elasticity of demand as the responsiveness of the quantity demanded to changes in price, expressed mathematically as:

$$\begin{aligned} E_p &= \frac{\Delta Q/Q}{\Delta P/P} \\ &= \frac{(Q_2 - Q_1)/Q_1}{(P_2 - P_1)/P_1} \\ &= \frac{(Q_2 - Q_1) \cdot P_1}{(P_2 - P_1) \cdot Q_1} \end{aligned}$$

Typically, the prices and quantities of goods in supply and demand exhibit relatively stable changes. However, the emergence of Covid-19 caused a significant disruption to international trade during 2020–2022. This exogenous shock provides an opportunity to calculate the price elasticity of demand for various goods in international markets. Therefore, future price could be predicted based on potential market demand.

4 Solution and Result

4.1 Elaboration

Future variation tendency was revealed by the terms of "overall volume". Cats and dogs were divided by means of subsitution effects in future growth. Figure 2 illustrates a consistent annual increase in the total number of pets, accompanied by notable shifts in the proportion of cats to dogs. The growth in the overall pet population can be effectively explained through the income effect model, which posits that rising income levels alleviate budget constraints, allowing more households to afford pet ownership. Meanwhile, the observed changes in the ratio of cats to dogs are analyzed within the framework of the substitution effect model.

This approach seems promising, given the evident overall growth trend. However, even the most insightful ideas are inevitably bound by the constraints of available data. Therefore, the workflow presented in Figure 1 was selected as the primary framework for conducting the analysis.

4.2 OLS Result, GM Prediction, Random Forest Prediction

4.2.1 OLS Test Results

The results of our baseline OLS regression are summarized in Table 2. The regression coefficients and their respective significance levels were calculated using robust standard errors. These highly significant parameters, with narrow confidence intervals, were identified as the effective factors driving the observed outcomes.

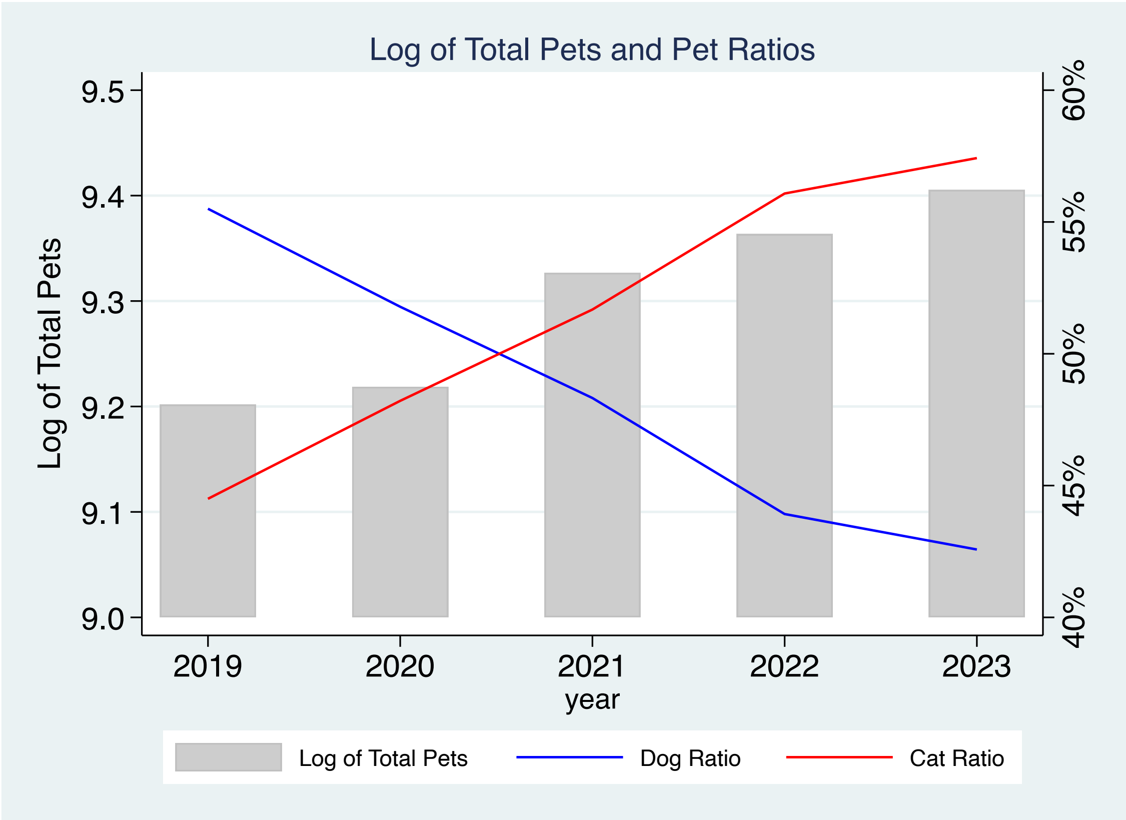

Figure 2: Logarithmic Bar Chart of Total Pets & Line Chart of Cat-Dog Proportions

Table 2: OLS Regression Results

	(1) cat	(2) dog	(3) cat	(4) dog	(5) cat	(6) dog	(7) cat	(8) dog
Pay_c	-0.41	1.03***	-0.57***		0.62***			0.62***
	(.)	(0.00)	(0.00)		(0.00)			(0.00)
$Food_c$	0.59	6.89***	-7.64***		7.48***			7.48***
	(.)	(0.00)	(0.00)		(0.00)			(0.00)
Med_c	-0.03	-7.89***	-1.38***		-7.93***			-7.93***
	(.)	(0.00)	(0.00)		(0.00)			(0.00)
$Serve_c$	0.00	0.00	10.57***		0.00			0.00
	(.)	(.)	(0.00)		(.)			(.)
Pay_d	0.84	-0.84***		-0.46***		0.11***	0.05	
	(.)	(0.00)		(0.11)		(0.00)	(.)	
$Food_d$	0.00	0.00		7.87***		10.45***	4.92	
	(.)	(.)		(0.49)		(0.00)	(.)	
Med_d	0.00	0.00		-8.33***		-9.30***	-4.37	
	(.)	(.)		(0.43)		(0.00)	(.)	
$Serve_d$	0.00	0.00		0.00		0.00	0.00	
	(.)	(.)		(.)		(.)	(.)	
dog					-1.00***		-0.47	
					(0.00)		(.)	
cat						-2.12***		-1.00***
						(0.00)		(0.00)
Constant	-0.00	1.00***	-0.00	1.06***	1.00***	1.00***	0.47	1.00***
	(.)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(.)	(0.00)
Observations	17	17	17	17	17	17	17	17

Robust standard errors used, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

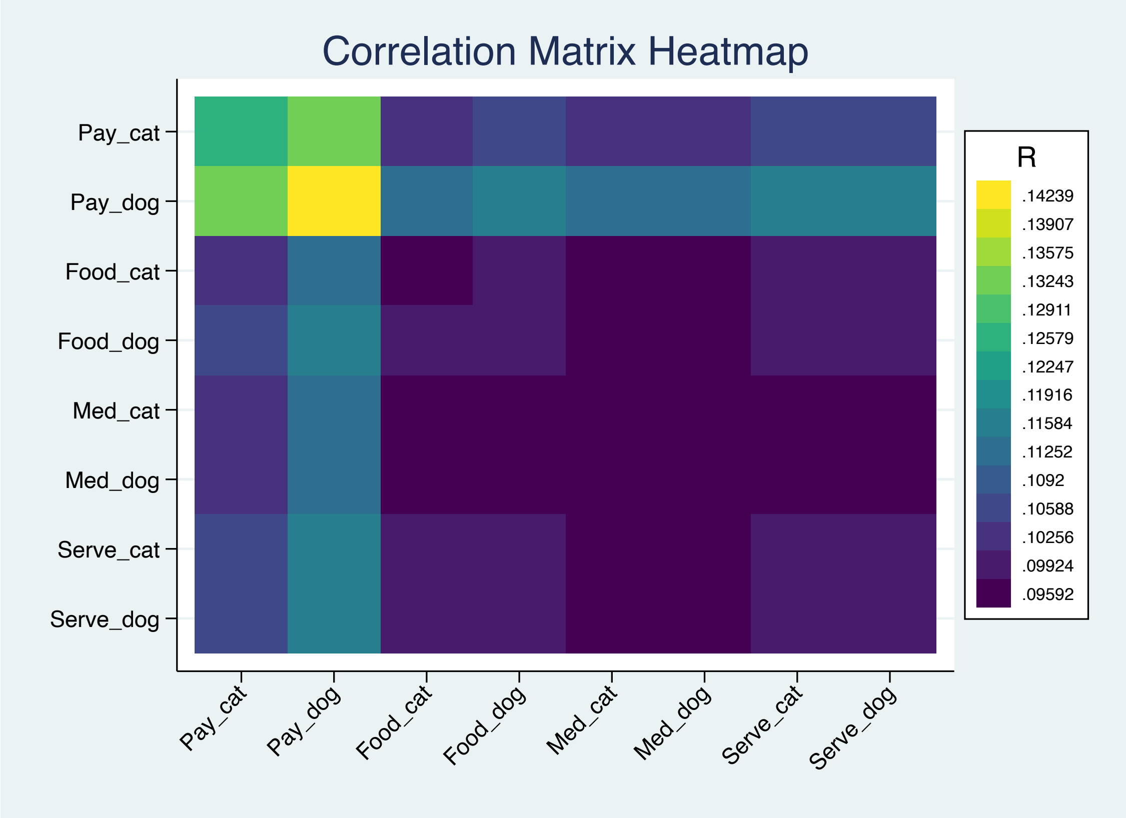

Figure 3: The Correlation Matrix Heatmap between the 8 Factors

The OLS results reveal a significant correlation between cats and dogs. Building upon the baseline OLS estimation, fixed effects were incorporated into the model to control for unobserved heterogeneity, thereby enhancing the robustness of the findings. This refinement continues to underscore the significant correlation between cats and dogs. Additionally, both cats and dogs demonstrate significant associations with their respective or each other's Pay, Food, Medicine, and Services, indicating that these factors are influential. However, when the factors associated with the other type of pet are included in the model, the results uniformly lose significance. This outcome raises the possibility of multicollinearity among these factors. To investigate further, correlation coefficients were computed for the eight factors. The resulting heatmap of these coefficients is presented in Figure 3.

The substitution effect model provides a framework for explaining the significance of regression coefficients and correlation results. However, the primary focus of this study lies in forecasting the number of pets over the next three years, rather than delving into these explanatory mechanisms.

4.2.2 Forecasting the Values of Factors

Subsequently, a grey forecasting model is employed to conduct an initial prediction based on eight parameters. The results are presented in Table 3. This approach enables the prediction of pet populations for the next three years.

Table 3: GM Prediction

Year	Pay_c	Pay_d	$Food_c$	$Food_d$	Med_c	Med_d	$Serve_c$	$Serve_d$
2024	1506	2199	4571	3459	1438	2156	905	1358
2025	1787	2688	5789	4234	1800	2701	1141	1712
2026	2077	3220	7189	5079	2210	3315	1411	2116

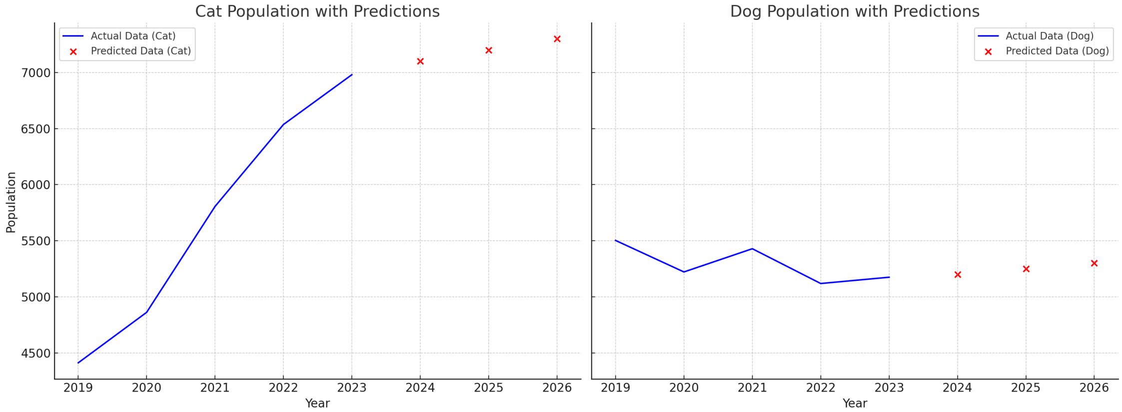

Figure 4: The Trend of Cats and Dogs in China (in 10,000s)

4.2.3 Random Forest Prediction

To address the issue of insufficient sample size that might prevent the training of a random forest model, the bootstrap method was effectively employed. This approach allowed for an expansion of the limited sample size, enabling the model to be trained successfully. Therefore, the factors predicted by the GM model serve as inputs for the trained random forest model, enabling the forecasting of pet populations. This integration leverages the predictive strengths of both models, ensuring a methodologically robust approach to estimation.

From Figure 4, it is evident that the population of cats continues to exhibit an upward trend, while the population of dogs has transitioned from a declining trajectory to an improving one. Given that the model's predictions are primarily trend-based, the final numerical results are not presented. Instead, a trend chart, as depicted in Figure 4, is provided to illustrate the projected scenarios for cats and dogs over the next three years based on the model's forecasts.

Following the forecast of domestic pet market development, the same methodological framework was utilized to predict future trends in the global pet market. This approach enables the projection of parameter changes on a global scale, facilitating a comprehensive analysis of anticipated trends in pet population growth and dynamics.

4.3 Mechanical Repeating

Although the structural prediction based on the aforementioned model is not what we desire, since we have many excellent models that could be designed, we are

Figure 5: The Trend of Cats and Dogs in Global (in 10,000s)

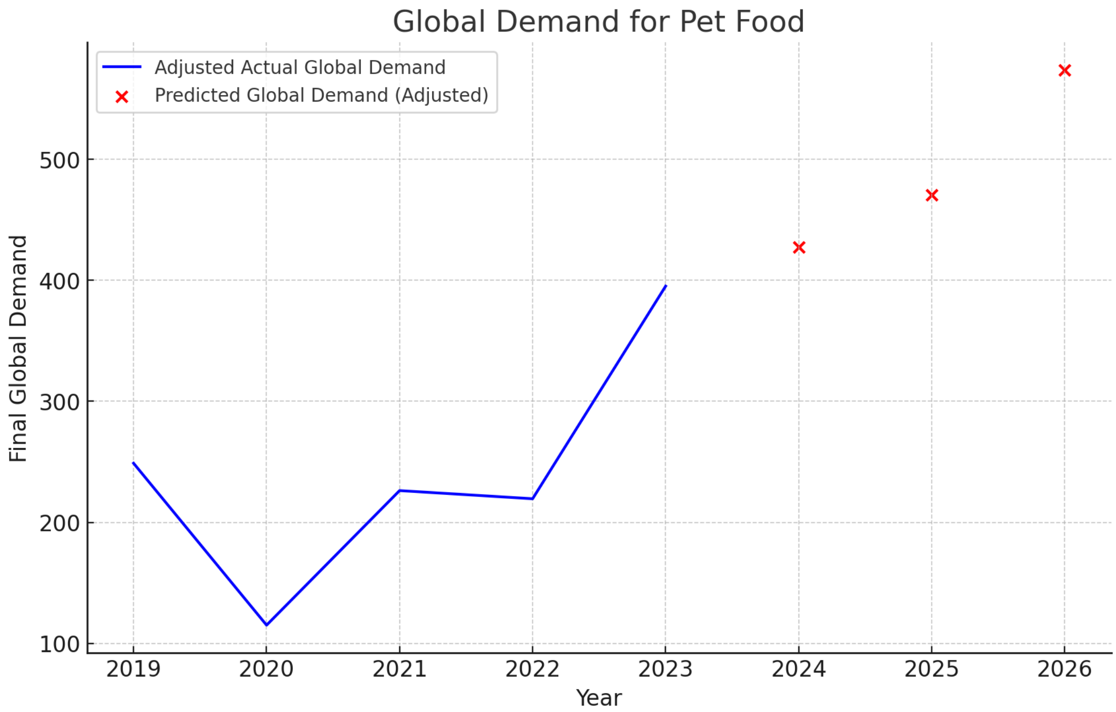

Figure 6: Global Demand for Pet Food (in trillions of USD)

compelled to do so due to the limitations of the data.

Regarding global pet food production, an elasticity analysis (Espey et al., 1997) was initially considered. However, due to the limitations of available databases, sufficient data to calculate price elasticity of demand for prediction was not obtainable. Consequently, the random forest method was employed to forecast global demand for pet food. Thus, the same methodology was applied to train a new random forest model. In contrast to the previous approach, the model incorporated the population of cats and dogs as additional features during the training process which are presented in Figure 6.

To address the question of analyzing the development trajectory of China's pet food industry and forecasting its production and export volumes over the next three years, we constructed Figure 7 and Figure 8.

4.4 Game Theory

While I do not have the data to test tariff model, I will try my best to give a solution for the question 4.

The game considered is a three-party dynamic interaction in which the government first sets the tariff level T. Following this, the exporter decides whether to export and, if so, determines the optimal price P. Finally, the importer decides whether to accept the import price $P + T$.

Backward induction is applied to solve this dynamic game, analyzing the decisions of each agent sequentially from the end of the game to the beginning.

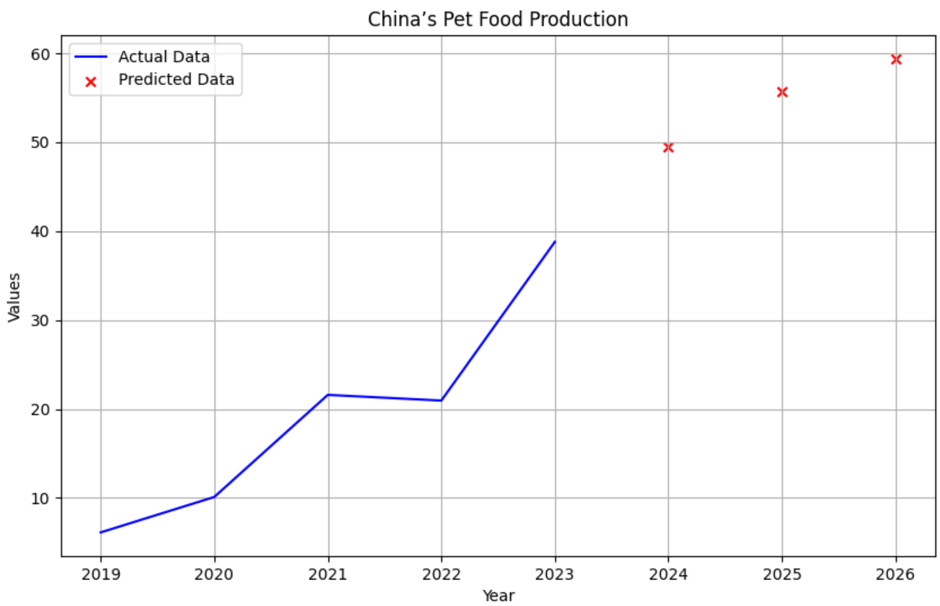

Figure 7: China’s Pet Food Production (in billions USD)

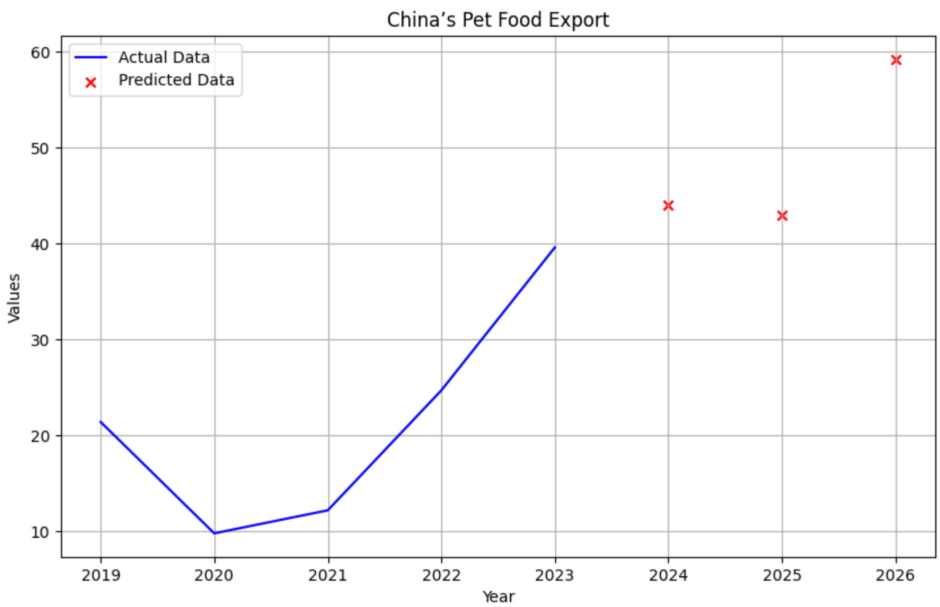

Figure 8: China’s Pet Food Export (in billions USD)

4.4.1 Importer's Decision

The importer's decision depends on whether the price $P + T$ is acceptable. The utility function of the importer is defined as:

$$U_{\text{importer}} = V(Q) - (P + T)Q,$$

where $V(Q)$ represents the total benefit derived from importing quantity Q, and $(P + T)Q$ is the total cost incurred.

The condition for the importer to accept the price is that the utility must be non-negative:

$$U_{\text{importer}} \geq 0 \implies V(Q) \geq (P + T)Q.$$

This condition imposes an upper bound on the price:

$$P + T \leq \frac{V(Q)}{Q}.$$

4.4.2 Exporter's Decision

Given the tariff T imposed by the government, the exporter selects the price P to maximize its profit. The exporter's profit function is:

$$\Pi_{\text{exporter}} = P \cdot Q(P + T),$$

where $Q(P+T)$ is the demand function of the importer, which depends on the total cost $P + T$.

The profit-maximizing price is determined by solving the following optimization problem:

$$\max_{P} \Pi_{\text{exporter}} = P \cdot Q(P + T).$$

The first-order condition (FOC) for this problem is:

$$\frac{\partial \Pi_{\text{exporter}}}{\partial P} = Q(P + T) + P \cdot \frac{\partial Q(P + T)}{\partial P} = 0.$$

Rewriting, the exporter's optimal pricing rule is:

$$Q(P + T) + P \cdot Q'(P + T) = 0.$$

This condition determines the exporter's optimal price, denoted as $P^*(T)$, as a function of the tariff T.

4.4.3 Government's Decision

The government selects the tariff T to maximize overall welfare. The social welfare function typically comprises three components:

1. **Tariff revenue:** $T \cdot Q(P + T)$.

2. **Importer's net utility:** $U_{\text{importer}} = V(Q) - (P + T)Q$.

3. **Exporter's profit:** $\Pi_{\text{exporter}} = P \cdot Q(P + T)$.

The total welfare function is expressed as:

$$W = T \cdot Q(P + T) + [V(Q) - (P + T)Q] + P \cdot Q(P + T).$$

The optimal tariff T^* is determined by solving:

$$\frac{\partial W}{\partial T} = \frac{\partial}{\partial T} [T \cdot Q(P + T) + V(Q) - (P + T)Q + P \cdot Q(P + T)] = 0.$$

By substituting the exporter's pricing response $P^*(T)$ and the importer's demand response $Q(P^*(T) + T)$ into this equation, the optimal tariff can be derived.

4.4.4 Nash Equilibrium

The Nash equilibrium of this dynamic game is a set of strategies $(T^*, P^*(T^*), Q(P^*(T^*)+T^*))$ satisfying the following conditions:

- The government's choice of T^* maximizes the social welfare function W.
- The exporter's choice of $P^*(T^*)$ maximizes the profit function Π_{exporter}, given the tariff T^*.
- The importer's choice of $Q(P^*(T^*) + T^*)$ satisfies the utility-maximizing condition.

This equilibrium ensures that the strategies of all parties are mutually optimal, considering the decisions of the other two participants.

4.5 Judgement

Strength1: The innovation of this model lies in its consideration of future variation tendency in indicators from the perspective of "overall volume". A new model was developed based on income and substitution effects, capturing the impact of income levels and the substitutive relationship between cats and dogs in terms of utility on decision-making.

Strength2: The bootstrap method rigorously evaluates the interdependence among final sample values, thereby substantially addressing potential biases and errors arising from limited sample data.

Strength3: The analysis incorporates domestic and international pet food prices, relative demand elasticity, and market-based projections of export volume changes. This approach, which has not been addressed in prior models, constitutes a significant innovation in the proposed framework.

Weakness1: It is with regret that we did not utilize the substitution effect to divide future growth into cats and dogs in our actual solution. Due to the limitations of the dataset, it is not possible to provide deeper insights into consumer behavior and preferences regarding the choice between cats and dogs.

Weakness2: Another limitation is that we used the interpolation algorithm when selecting the optimal parameters in OLS, but considering the possibility of seasonal changes in pet numbers, it may lead to data that cannot accurately describe the trend of the parameters.

Weakness3: Due to the insufficient sample size, it is challenging to accurately estimate the slope of the annual substitution effect as well as the slope of the projected future substitution effect.

5 Conclusions

5.1 Peroration

5.1.1 Question 1 & 2

As the results depict developmental trends, graphical representations are considered the most suitable approach to address these two questions. Specifically, Figure 4 responds to Question 1, while Figures 5 and 6 together provide answers to Question 2.

5.1.2 Question 3

And for question 3, it was be addressed by the Figure 7 and Figure 8.

5.1.3 Question 4

Reluctantly, I must admit that the data available to me is highly limited. As a result, this section cannot empirically test the validity of our model or accurately predict future scenarios. Nevertheless, the previous section have offerd a detailed explanation of the tariff model, ensuring clarity and rigor in its theoretical foundation.

5.2 Feasibility

The proposed model is feasible for the following reasons:

Economics Background: The models constructed are grounded in solid economic reasoning, derived from the principle of maximization to solve the equations systematically. This approach ensures that the theoretical foundation aligns with established economic methodologies.

Baptism of History: Numerous methods and algorithms have been thoroughly validated, including the bootstrap method, a well-established statistical resampling technique designed to expand sample size and improve the reliability of estimates. This approach is widely acknowledged in both theoretical and practical applications, offering robust support for model analysis and inference.

5.3 Emphasize

For Question 1, the domestic population of cats and dogs was forecasted over the next three years, with the number of cats expected to remain stable at over 70 million and the number of dogs projected to show a modest upward trend, ranging between 50 and 55 million. Additionally, data from the United States, France, and Germany were analyzed to predict global cat and dog populations, revealing trends similar to the domestic forecasts, though limited sample data impacted the accuracy of the results. Based on global pet ownership trends, the demand for pet food is projected to increase annually, surpassing 400 trillion USD by 2024. Regarding Question 4, the impact of foreign economic policies indicates that China's pet food production will grow steadily, reaching approximately 60 billion USD by 2026, with pet food exports also expected to follow an overall upward trajectory.

References

[1] Deng Julong. Grey dynamic model (gm) and its application in long-term grain prediction. Nature Exploration, (03):37–43, 1984.

[2] Deng Julong. Gray fuzzy prediction and control of grain. Journal of Huazhong University of Science and Technology, (02):1–8, 1983.

[3] Dua'a B. Telfah, Aiman Q. Jaradat, Rabah Ismail, Examining the Long-Run and Short-Run Relationship between Water Demand and Socio-Economic Explanatory Variables: Evidence from Amman, Sustainability, 10.3390/su16062315, 16, 6, (2315), (2024).

[4] Jeffrey M Wooldridge, Mokhtarul Wadud, and Jenny Lye. Introductory econometrics: Asia pacific edition with online study tools 12 months. Cengage AU, 2016.

[5] N Gregory Mankiw. Principles of economics. Cengage Learning, 2016.

[6] Yang Hualong, Liu Jinxia, and Zheng Bin. Improvement and application of grey prediction gm(1,1) model. Mathematics Practice and Understanding, 41(23):39–46, 2011.

[7] Yi Xingjian and Zhou Li. Does the development of digital inclusive finance significantly affect household consumption? micro evidence from chinese households. Financial Research, (11):47–67, 2018.3