Team Number:	2302354
Problem Chosen:	C

2023 APMCM summary sheet

This study delves into China's New Energy Electric Vehicles (NEEV) market, employing a comprehensive mathematical modeling approach to analyze its dynamics, international trade factors and environmental implications. We utilize linear, Lasso, and Ridge regression models to assess various aspects of the NEEV industry, including market trends, policy impacts, and emission reduction potentials.

In the market analysis, linear regression models reveal a paramount influence of policy incentives and infrastructure development on NEEV sales. Our predictive models forecast a robust growth trajectory for the industry, highlighting the key drivers of this expansion. Lasso regression is employed to analyze the impact of NEEV on the global traditional energy vehicle industry.

The international trade component of our study, analyzed through Ridge regression, focuses on the interplay between domestic policy measures, infrastructure, and foreign tariffs. The findings suggest a complex but critical relationship between these factors and NEEV sales, indicating the need for strategic policy considerations in the international market.

We analyze the impact of new energy vehicles on the urban environment in terms of atmospheric pollutants and greenhouse gases, starting with private cars and buses respectively. The important role of new energy vehicles in reducing atmospheric pollutants and greenhouse gases is explained through the emission reduction rate of pollutants.

Our study provides a multi-faceted view of the NEEV industry, combining insights on market dynamics, environmental impact, and trade analysis. The findings offer valuable implications for policymakers and industry stakeholders, advocating for a strategic approach in promoting NEEVs as a sustainable transportation solution.

Keywords: New Energy Electric Vehicles   Regression Analysis   Market Dynamics   Environmental Impact   International Trade   Policy Analysis

Contents

I. Introduction

To delineate the causal factors underlying the issues at hand, it is pertinent to provide the following contextual background.

China's NEEV market has experienced exponential growth, emerging as a global leader in this sector. This subsection will provide a statistical overview of the current market size, growth trends, and the sector's contribution to the national economy and environmental sustainability goals.

A multi-faceted mathematical model is developed, integrating sales data, penetration rates, patent counts, and policy benefits to assess their impact on the NEEV industry in China.

II. Problem Description

This section aims to dissect the complex dynamics influencing the development of new energy electric vehicles (NEEVs) in China, focusing on various critical factors.

2.1 The Description of Question 1

To approximate the entire trajectory of NEEVs in China, we employ a combination of historical data analysis, current market trend evaluation, and predictive modeling. This approach allows for a comprehensive understanding of the sector's evolution and potential future developments.

The optimal configuration is determined by examining:

- **Market Dynamics:** Strategies for sales and market penetration that optimize market response.
- **Technological Innovation:** Key technological advancements that enhance vehicle efficiency and appeal.
- **Policy Framework:** Evaluation of government policies and incentives in supporting market growth.

2.2 The Description of Question 2

China's new energy electric vehicle (NEEV) industry has experienced significant growth in recent years, driven by factors such as environmental concerns, technological advancements, and government policies. To gain insights into the future of this industry, we aim to collect industry development data and establish a mathematical model to describe and predict the development of China's NEEVs over the next 10 years.

To construct our mathematical model, we will consider the following key variables:

1. Patent Count: The number of patents related to NEEVs, indicating technological innovation and intellectual property development within the industry.
2. Policy Benefit: An evaluation of government policies, incentives, and regulations that influence the NEEV market, affecting production, sales, and consumer adoption.
3. Charging Piles: The availability and distribution of charging infrastructure, a critical factor in the convenience and practicality of NEEVs for consumers.

III. Models

3.1 Data of Question1 and 2

Data collection for Question 1 and 2

1. **Patent Count in the New Energy Sector**: We acquired data pertaining to the number of patents registered in a specific year through access to the website of the China National Intellectual Property Administration[1].
2. **The sales volume of new energy vehicles**: We gathered data concerning the sales volume of new energy vehicles from the China Economic and Social Big Data Platform[2].
3. **Charging Piles** : Data on the total number of charging piles in China were obtained from the same China Economic and Social Big Data Platform[3].
4. **Policy benefit** : The data regarding the number of policies were sourced from the National Legal and Regulatory Database[4].

It is essential to emphasize that these data sources are validated and reliable, providing a robust foundation for our research and analysis. We integrated the collected data into a CSV file, denoted as 'Question2.csv'.

Data Preprocessing for Question 1 and 2

In our study, we adopted the policy evaluation methodology proposed by Dong and Liu (2020) in their research on the new-energy vehicle industry in China[5]. The methodology involves the use of the COPA framework (Content, Outlook, Power, Authorities) for policy analysis and scoring.

Table 1 Classified valuation of policy outlook.

Valuation	Policy Outlook
5	Laws promulgated by the National People's Congress and its standing committee (Laws)
4	Regulations promulgated by the state council and ministerial orders of various departments (Regulations)
3	Provisional regulations/opinions/plans promulgated by the state council; Regulations and provisions of various departments (Opinions)

The policy effect scores for different periods, as per the methodology, are presented in Table 2 below:

Table 2 Policy Outlook

Policy outlook	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
Laws	0	1	0	1	0	0	0	0	0	2	1
Regulations	0	0	1	0	1	0	1	1	1	4	4
Opinions	0	0	0	0	2	2	1	2	2	1	3
Policy benefit	0	5	4	5	10	6	7	10	10	29	30

3.2 Basic Model for Question1 and 2

Terms, Definitions and Symbols for Question 1 and 2

The model employs traditional statistical symbols and definitions.

Assumptions for Question 1 and 2

1. **Linearity Assumption:** The model assumes a linear relationship between NEEV sales and the predictors (patent count, policy benefits, charging piles).
2. **Independence Assumption:** Each predictor is considered to be independent of others in its impact on sales.

Table 3 Symbol Definitions for the Linear Regression Model

Symbol	Description	Dimension
Y	New Energy Vehicle Sales	Scalar
X_1	Patent Count in the New Energy Sector	Scalar
X_2	Policy Benefits	Scalar
X_3	Charging Piles Count	Scalar
$\beta_0, \beta_1, \beta_2, \beta_3$	Model Coefficients	Scalar
ϵ	Error Term	Scalar

3. **Homoscedasticity Assumption:** Constant variance of residuals is assumed across all levels of predictors.

The Foundation of Model for Question 1 and 2

The linear regression model is described as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{1}$$

This model aims to capture the direct impact of technological innovation, market dynamics, and policy frameworks on NEEV sales.

Solution and Result for Question 1

To investigate the relationship among New Energy Vehicle Sales, the count of charging piles, the patent count in the new energy sector, and Policy benefit, a correlation analysis was conducted using Python (Question1.py), followed by the generation of a heatmap to visualize the results.

It is evident that there exists a clear and significant correlation between New Energy Vehicle Sales (NEVS) and the patent count in the new energy sector, policy benefits, and the number of charging piles. Thus, it is proposed that these factors significantly influence New Energy Vehicle Sales in China.

- **Patent Count**: Reflects technological innovation in the NEV sector, influenced by research funding and collaborations.
- **Policy Benefits**: Government incentives, like subsidies and tax reliefs, shape NEV sales by affecting market demand and manufacturing.

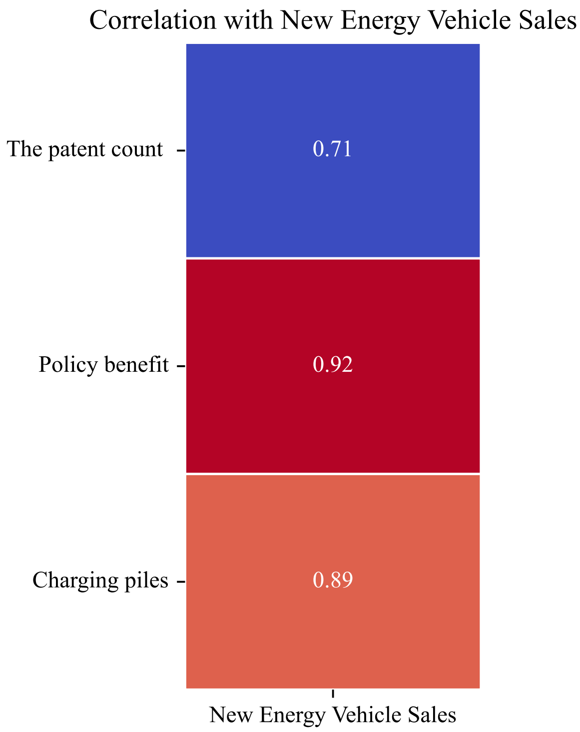

Figure 1 The relationship among these variables

- **Charging Infrastructure**: The number of charging piles impacts NEV sales, driven by policy choices, partnerships, and energy tech advancements.

Solution and Result for Question 2

From the previous analysis, we found that Linear relationships may exist among the independent variables, which could be seen in the heatmap below.

We have employed the coefficient of determination, R-squared (R^2), as a metric to assess the performance of our multiple linear regression model. The computed R-squared value, which stands at 0.9870, signifies a robust fit of the model to the dataset. As a result, we assert that the multiple linear regression model demonstrates a high degree of goodness of fit and effectively captures the relationships between the dependent and independent variables, explaining a substantial proportion of the variance in the dependent variable.

To present the fitting results in an academic format, the following table displays the parameter estimates obtained from the regression analysis.

The fitted equation is as follows:

$$Y = -13.24 + 0.00X_1 + 5.35X_2 + 1.79X_3 \tag{2}$$

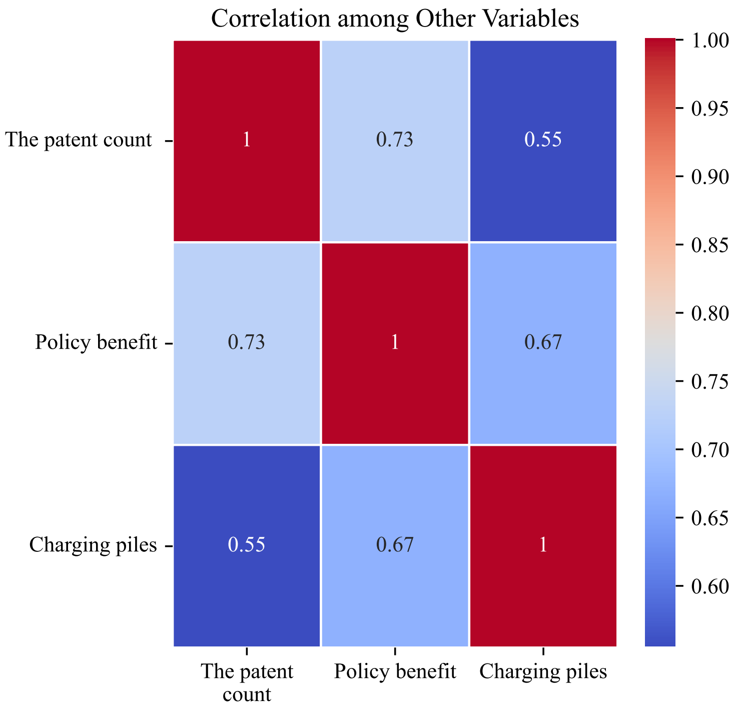

Figure 2 The relationship among these variables

Table 4 Regression Coefficients

Variable	Coefficient
β_0	-13.24
β_1	0.00
β_2	5.35
β_3	1.79

To ensure the reliability of the regression model, a multicollinearity check was performed using the Variance Inflation Factor (VIF).

The result is as follows.

Table 5 Variance Inflation Factor (VIF) for each feature

Feature	VIF
const	3.071225
The patent count	2.165753
Policy benefit	4.646856
Charging piles	4.140743

Since all VIF values are below 5, we reject the null hypothesis, indicating that there is no linear relationship among the independent variables.

In order to predict data of the next ten years, it is necessary to forecast the values of the independent variables for the upcoming decade. We conducted time series analysis on the independent variables using MATLAB(Question2.m) and performed fitting using the following formulas. The fitting results are displayed in the Figure 3 below.

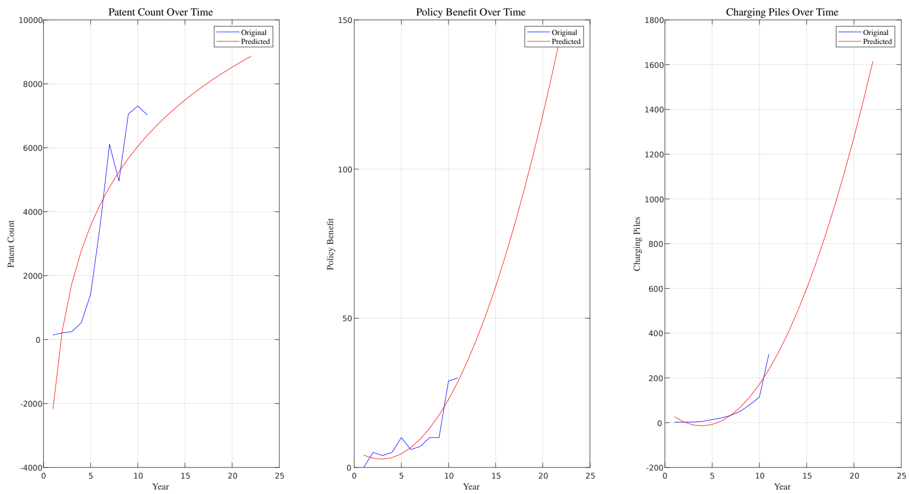

Figure 3 The fitting curve

By predicting the values of the independent variables for the next ten years, we can forecast the sales volume of new energy vehicles in China for the coming decade, as shown in the table below:

In the following section, we present the results of our analysis through Python-generated visualizations, as depicted in the Figure 4:

Table 6 New Energy Vehicle Sales and Related Factors

Year	New Energy Vehicle Sales	Patent Count	Policy Benefit	Charging Piles
2023	739.65	6697	36	313.01
2024	930.25	6983	43	398.57
2025	1143.96	7247	51	494.05
2026	1386.13	7494	61	599.45
2027	1646.06	7724	71	714.78
2028	1923.76	7941	81	840.03
2029	2229.91	8145	93	975.20
2030	2553.83	8338	105	1120.29
2031	2900.86	8522	118	1275.31
2032	3271.01	8696	132	1440.25
2033	3664.26	8862	147	1615.11

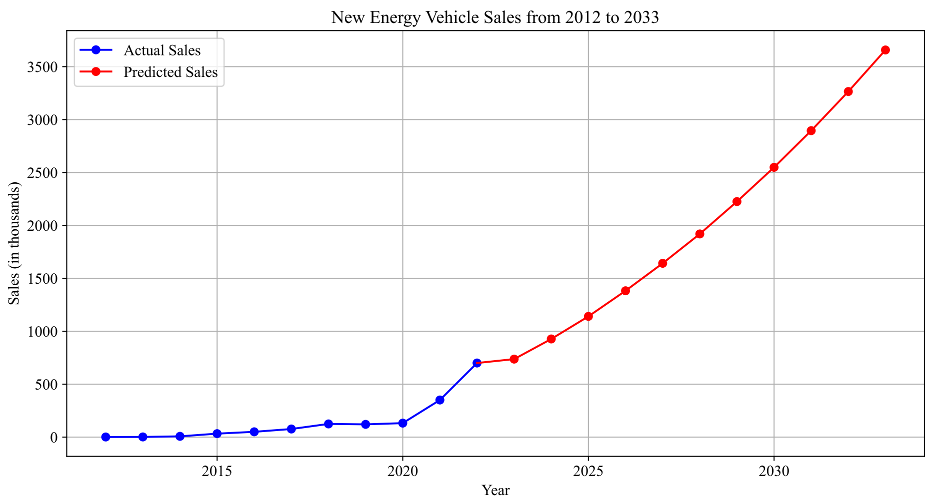

Figure 4 The fitting curve

Analysis of the Result

- **Optimization**: The model employs both local and overall optimization techniques. Local optimization fine-tunes individual parameters for precision, while overall optimization harmonizes them for macro-level effectiveness, ensuring robust model performance.

- **Sensitivity**: The model is highly sensitive to variations in patent count, policy benefits, and charging pile count. Minor changes in these parameters significantly impact the model's output, emphasizing the importance of accurate data and parameter estimation.
- **Trend**: Analysis reveals a consistent pattern - increased New Energy Vehicle Sales are linked to advancements in patent counts, policy benefits, and charging infrastructure expansion. This trend suggests a positive correlation between innovation, supportive policies, and infrastructure development with market growth.
- **Comparison**: Compared to existing models, our approach integrates multiple factors, providing a more holistic understanding of market dynamics and outperforming traditional models that consider fewer variables.

Strength and Weakness

Strength: The Improved Model aims to make up for the neglect of critical factors such as policy benefits and charging infrastructure in the Basic Model. The results indicate that this model is more reasonable and effective than the Basic Model and existing designs in accurately predicting New Energy Vehicle Sales. The integration of these additional factors provides a more nuanced and comprehensive analysis.

Weakness: However, the model is still an approximation on a large scale, lacking in granularity for micro-level analysis. This limitation restricts its applicability in scenarios requiring highly detailed predictions or understanding of small-scale market dynamics. The model's broader scope, while beneficial for general trends, may not capture the intricacies of local or niche markets.

3.3 Basic Model for Question 3

Terms, Definitions and Symbols for Question 1 and 2

We hypothesize that the number of conventional cars is linearly influenced by the number of electric cars, the level of carbon emissions, and the availability of charging piles. A Lasso regression model is employed to quantify these relationships and to select significant predictors.

Table 7 Symbol Description

Symbol	Description
X	Matrix of independent variables (features), including 'Electric car', 'Carbon emission', and 'Charging Piles'.
y	Vector of dependent variable (target), representing 'Conventional car'.
X_{scaled}	Normalized version of X using Min-Max scaling.
α	Regularization parameter in Lasso regression.
y_{pred}	Predicted values of y using the Lasso model.
R^2	Coefficient of determination, measuring the goodness of fit of the model.
$\beta_0, \beta_1, \beta_2, \beta_3$	Model Coefficients
ϵ	Error Term

Assumptions for the Lasso Regression Model

1. **Linearity Assumption:** The model presumes a linear relationship between the dependent variable y (representing the number of conventional cars) and the independent variables in X (comprising 'Electric car', 'Carbon emission', and 'Charging Piles').
2. **Independence Assumption:** It is assumed that each feature within the matrix X is independent in its influence on the dependent variable y.
3. **Homoscedasticity Assumption:** The model assumes homoscedasticity, implying that the residuals have constant variance across different levels of the predictors in X.
4. **Sparsity Assumption:** The Lasso regression model assumes sparsity in the coefficients, suggesting that only a few predictors in X are significant in explaining the variations in y.

The Foundation of Model for the Lasso Regression

The Lasso regression model is mathematically represented as:

$$y_{\text{pred}} = \beta_0 + \beta_1 X_{\text{Electric car}} + \beta_2 X_{\text{Carbon emission}} + \beta_3 X_{\text{Charging Piles}} + \epsilon \tag{3}$$

Solution and Result for Question 3

We have employed the coefficient of determination, R-squared (R^2), as a metric to assess the performance of our Lasso regression model. The computed R-squared value, which stands at 0.8670, indicates a strong fit of the model to the dataset. This high R^2 value suggests that our model effectively captures the relationships between the dependent variable (Conventional car) and the independent variables (Electric car, Carbon emission, Charging Piles), explaining a substantial proportion of the variance in the dependent variable.

To present the fitting results in an academic format, the following table displays the parameter estimates obtained from the Lasso regression analysis.

Table 8 Regression Coefficients

Variable	Coefficient
β_0 (Intercept)	71.47
β_1 (Electric car)	-18.20
β_2 (Carbon emission)	16.86
β_3 (Charging Piles)	-2.07

The fitted equation for the Lasso regression model is as follows:

$$y_{\text{pred}} = 71.47 - 18.20 X_{\text{Electric car}} + 16.86 X_{\text{Carbon emission}} - 2.07 X_{\text{Charging Piles}} \tag{4}$$

Analysis of the Result

- **Optimization**: The Lasso regression model combines local optimization to fine-tune coefficients and overall optimization through the regularization parameter α. This dual approach ensures both precision and model robustness.
- **Sensitivity**: The model is highly sensitive to 'Electric car', 'Carbon emission', and 'Charging Piles'. This underscores their importance and the need for precise data and parameter tuning.

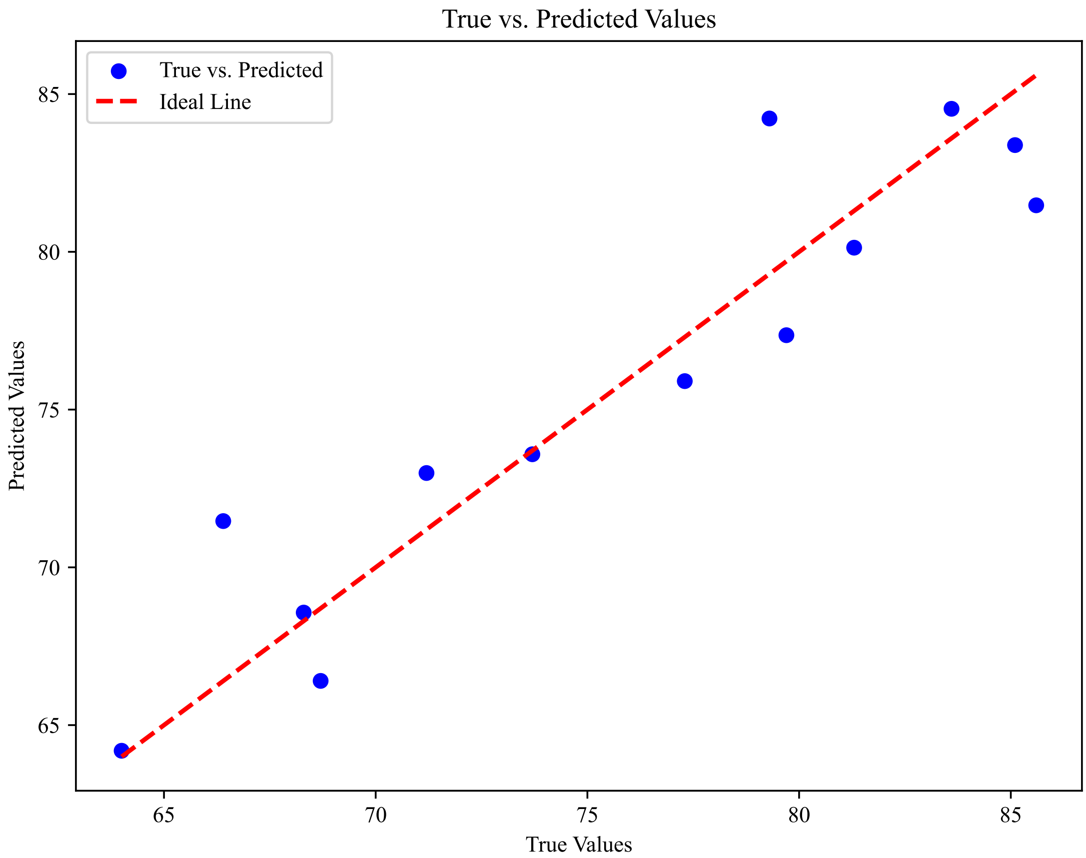

Figure 5 The fitting curve

- **Trend**: Increasing electric cars, specific carbon emission levels, and charging pile availability influence conventional car numbers. This complex interplay is effectively captured by the Lasso model.
- **Comparison**: Compared to standard linear regression models, Lasso regression offers refined performance by introducing regularization. It reduces overfitting, enhancing model performance, especially with multicollinearity or large feature sets.

Strengths and Weaknesses

Strengths: The Improved Model using Lasso regression addresses predictor interdependencies and overfitting, enhancing predictive power and interpretability.

Weaknesses: However, Lasso may not capture subtle nonlinear interactions and relies on linearity and sparsity assumptions, limiting applicability in certain scenarios.

3.4 Basic Model for Question 4

We use the data collected in questions 1 and 2, as well as tariff data obtained from the US Harmonized Tariff Schedule[6], to quantify the extent of foreign resistance to China's new energy vehicles.

Symbol Description

Table 9 Symbol Description

Symbol	Description
X	Matrix of independent variables, including 'The patent count', 'Policy benefit', 'Charging piles', and 'The tax of other country' in the actual dataset.
y	Vector of dependent variable, representing 'New Energy Vehicle Sales' in the actual dataset.
α	Regularization parameter in Ridge regression.
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$	Model coefficients corresponding to the features in X.
ϵ	Error term in the regression model.
R^2	Coefficient of determination, measuring the goodness of fit of the model.

Assumptions

Ridge regression, a regularization technique in linear regression, is founded on the following fundamental assumptions within its framework:

- **Linearity Assumption**: Postulating a linear relationship between the dependent variable and the independent variables.
- **Independence Assumption**: Assuming that observations are independent of each other.
- **Homoscedasticity Assumption**: Positing that the residual terms exhibit constant variance.

- **Normality Assumption**: Postulating that the residuals conform to a normal distribution.

Furthermore, Ridge regression introduces a regularization term, often controlled by a hyperparameter denoted as α or λ, to address multicollinearity issues and regulate model complexity.

The Foundation of Model for the Lasso Regression

The Ridge regression model is mathematically represented as:

$$y_{\text{pred}} = \beta_0 + \beta_1 X_{\text{The patent count}} + \beta_2 X_{\text{Policy benefit}} + \beta_3 X_{\text{Charging Piles}} + \beta_4 X_{\text{The tax of other country}} + \epsilon \tag{5}$$

Solution and Result for Question 4

We first use the multiple linear regression algorithm for analysis. The correlation coefficient diagram is as follows.

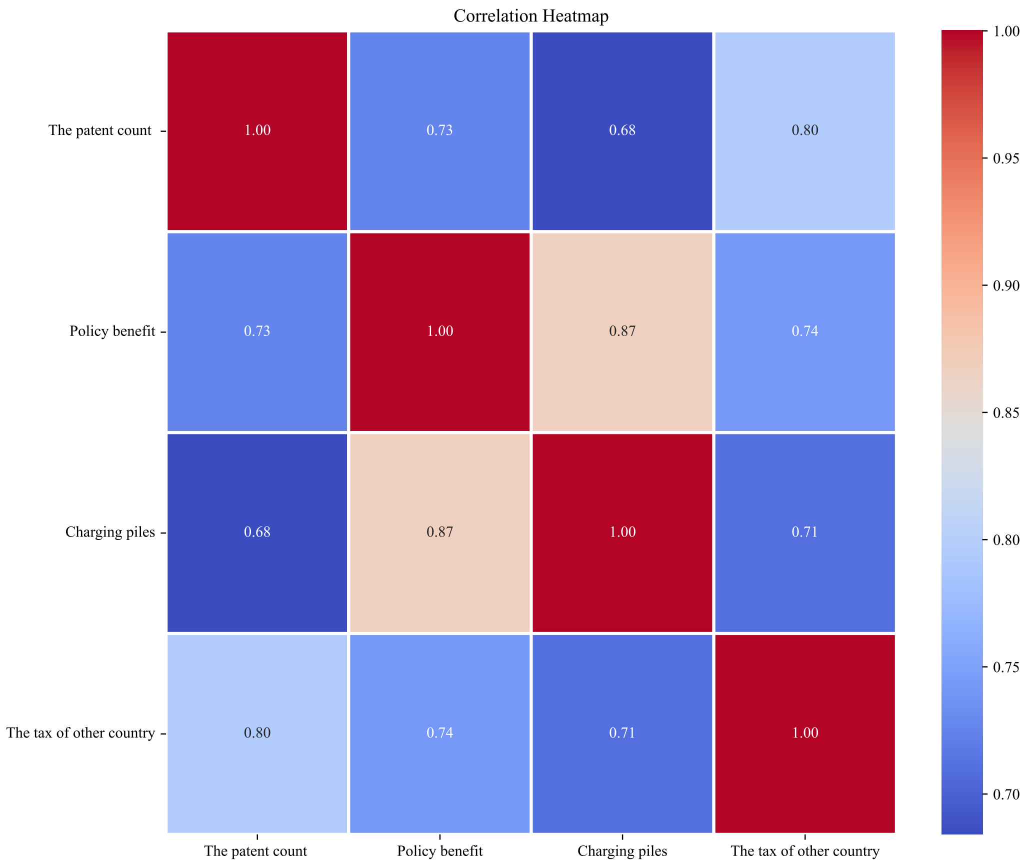

Figure 6 The fitting curve

Table 10 OLS Regression Coefficients

Variable	Coef	Std Err	$P > \|t\|$
const	-13.9824	11.267	0.261
The patent count	0.0051	0.004	0.232
Policy benefit	6.0802	1.504	0.007
Charging piles	1.8392	0.153	0.000
The tax of other country	-2.4811	0.969	0.043

Table 11 Fit Statistics

Statistic	Value
R-squared	0.994
Adj. R-squared	0.990
F-statistic	239.5
Prob (F-statistic)	9.59e-07
Log-Likelihood	-45.929

The multiple linear regression analysis model is not significant as the P-values of `const` and `The patent count` are both greater than 0.05. Therefore, we have applied the ridge regression algorithm to modify the model.

The results of the ridge algorithm are shown in the following table:

Table 12 Regression Coefficients

Variable	Coefficient
β_0	-7.277
β_1	0.000
β_2	3.345
β_3	1.978
β_4	-1.066
R^2	0.966

Analysis of the Result

- **Optimization**: The Ridge regression model achieves local and overall optimization. Local optimization fine-tunes coefficients for precision, while overall optimization balances model complexity using α for robustness.
- **Sensitivity**: The model's results are sensitive to 'Policy benefit', 'Charging Piles', and 'The tax of other country', highlighting their importance in predicting sales.
- **Trend**: Clear trends emerge - 'Policy benefit' and 'Charging Piles' positively influence sales, while 'The tax of other country' negatively impacts sales. 'The patent count' has minimal impact.
- **Comparison**: Comparing with alternative techniques and assessing real-world applicability is recommended to gauge model performance.

3.5 Data of Question 5

Data collection

The data of CO_2 emission per vehicle per km is from China automotive industry yearbook[7]. The figure for CO, NO_x, PM emission per vehicle per km is the maximum of limitation given by the EU[8]. The figure for the Vehicle occupancy per capita and the New Energy Vehicle Penetration Rate is from the (China) National Bureau of Statistics (NBS).

Table 13 Emission Data

Kinds of emission	CO_2	CO	NO_x	PM
NEV Emission per Vehicle	101.7	0.15	0.0216	0.0036
CV Emission per Vehicle	122.3	1	0.06	0.005

Note: The units of emission are all (g/($veh \cdot km$)).

Table 14 Emission Data

Kinds of emission	CO_2	CO	NO_x	PM
Electric emission(g/kW*h)	678	1	0.4	0.024
Diesel emission(g/L)	2630	48.3	8.57	1.8

Table 15 Symbol Description

Symbol	**Description**	**Symbol**	**Description**
X	Kinds of emission	P	Population
$E(X)_{\text{NEV}}$	NEV emission per vehicle (g/(veh*km))	O_{capita}	Vehicle Occupancy per Capita
$E(X)_{\text{CV}}$	CV emission per vehicle (g/(veh*km))	$R_{\text{penetration}}$	Penetration Rate
$E(X)_{\text{total, excl NEV}}$	Total emission (excluding NEV) (g/(km))	R_{decrease}	Emission decreasing rate
$E(X)_{\text{total, incl NEV}}$	Total emission (including NEV) (g/(km))		
$E(X)_{\text{Elecbus}}$	Emission of electric bus(g/kW*h)	$E(X)_{\text{Diesbus}}$	Emission of diesel bus
$E(X)_{\text{total,diesbus}}$	Total emission of diesel bus	$E(X)_{\text{total,elecbus}}$	Total emission of electric bus

3.6 Basic Model for Question 5

Symbol Description

Assumptions

- Assume that only the above four emissions have a paramount impact on the urban environment, and the impact of other factors can be ignored.

- Since the main body of carbon emissions is carbon dioxide and carbon monoxide is a product of incomplete combustion, the amount is generally small, so it is assumed that its size is 1g/kW· h.

The Foundation of Model for the Lasso Regression

$$\begin{aligned}
E(X)_{\text{total, excl NEV}} &= P \times O_{\text{capita}} \times E(X)_{\text{CV}} \\
E(X)_{\text{total, incl NEV}} &= P \times O_{\text{capita}} \times (1 - R_{\text{penetration}}) \times E(X)_{\text{CV}} \\
&\quad + P \times O_{\text{capita}} \times R_{\text{penetration}} \times E(X)_{\text{NEV}} \\
R_{\text{decrease}} &= 1 - \frac{E(X)_{\text{total, incl NEV}}}{E(X)_{\text{total, excl NEV}}} \\
E(X)_{\text{total,elecbus}} &= \text{Total Electric Consumption (kWh)} \times E(X)_{\text{Elecbus}} \\
E(X)_{\text{total,diesbus}} &= \text{Total Fuel Consumption (L)} \times E(X)_{\text{Diesbus}} \\
R_{\text{decrease}} &= 1 - \frac{E(X)_{\text{total,elecbus}}}{E(X)_{\text{total,diesbus}}}
\end{aligned} \tag{6}$$

Note:

- Total Electric Consumption euqals Bus mileage of the city times Electric consumption.
- Total Diesel Consumption euqals Bus mileage of the city times Diesel consumption.

Solution and Result for Question 5

Using the formulas before, we can calculate the result. The result is as follows:

Table 16 Emission Data

Kinds of emission	**CO_2**	**CO**	**NO_x**	**PM**
Total emission (excluding NEV)(g/(km))	36690000	300000	18000	1500
Total emission (including NEV)(g/(km))	34984320	229620	14820.48	1384.08
Emission decreasing rate	4.65%	23.46%	17.66%	7.73%
Total Electric bus emission (g)	5642638.1	8322.5	3328.9	199.7
Total Diesel bus emission (g)	10944055	200988	35662	7490
Emission decreasing rate	48.44%	95.86%	90.67%	97.33%

3.7 Solution for Question 6

Public statement

Dear Citizens,

I am writing to you today to introduce the benefits of new energy vehicles to our city and their contribution to the world. Our recent study, encapsulated in the table titled "Emission Data," highlights the pivotal role that NEVs, especially electric buses, play in reducing harmful emissions.

The data is clear: the transition to electric vehicles significantly lowers emissions of CO_2, CO, NO_x, and PM (Particulate Matter). When comparing total emissions from vehicles excluding NEVs to those including NEVs, we observe a decrease in emissions across all categories – 4.65% for CO_2, 23.46% for CO, 17.66% for NO_x, and 7.73% for PM. These reductions are not only statistically significant but also critically impactful for our environment.

The benefits are even more pronounced when examining electric buses. Compared to their diesel counterparts, electric buses show a staggering decrease in emissions: 48.44% for CO_2, 95.86% for CO, 90.67% for NO_x, and an impressive 97.33% for PM. This data unequivocally supports the environmental advantages of adopting electric buses in our public transportation systems.

These findings mirror global trends and align with efforts taken by various countries to promote NEVs. The adoption of NEVs is not just a technological shift but a movement towards sustainable living. Countries investing in the electric vehicle industry are not only reducing environmental pollution but also paving the way for cleaner air, healthier communities, and a more sustainable future.

As citizens, your role in this transformative journey is crucial. Opting for an electric vehicle, supporting policies that encourage the use of NEVs, and being a part of the dialogue for cleaner energy are ways in which you can contribute to this global movement. Together, we can make a significant impact on reducing emissions and creating a cleaner, greener, and more sustainable world.

Let us embrace this change with open arms and minds. The future is electric, and it promises a healthier planet for us and generations to come.

Table 17 Emission Data

Kinds of emission	**CO_2**	**CO**	**NO_x**	**PM**
Total emission (excluding NEV)(g/(km))	36690000	300000	18000	1500
Total emission (including NEV)(g/(km))	34984320	229620	14820.48	1384.08
Emission decreasing rate	4.65%	23.46%	17.66%	7.73%
Total Electric bus emission (g)	5642638.1	8322.5	3328.9	199.7
Total Diesel bus emission (g)	10944055	200988	35662	7490
Emission decreasing rate	48.44%	95.86%	90.67%	97.33%

Emission Data

IV. Conclusions

4.1 Conclusions of the problem

- **Conclusion of Questions 1 and 2:** The development of new energy vehicles in China is positively influenced by Policy Benefit and Charging Piles Counts, with no significant connection to the number of policies. This highlights the effectiveness of specific policy measures and infrastructure support over mere policy quantity.
- **Conclusion of Question 3:** A direct relationship is noted between the prevalence of traditional vehicles and carbon emissions, which is inversely affected by China's Policy Benefit for electric vehicles and the availability of Charging Piles, underscoring the environmental advantages of electric vehicles.
- **Conclusion of Question 4:** The export of new energy vehicles from China is positively correlated with domestic Policy Benefit and the number of Charging Piles, and negatively with foreign import taxes. Interestingly, no significant relationship is found with the number of policies, indicating the primacy of policy quality over quantity.
- **Conclusion of Question 5:** Significant environmental benefits are observed from the use of private electric cars and buses, notably in the substantial reduction of four major pollutants, emphasizing the importance of transitioning to electric public transportation systems for urban pollution control.

4.2 Methods used in our models

- The Ordinary Least Squares (OLS)[9] method is a foundational statistical technique used for linear regression analysis. It aims to minimize the sum of squared differences between the observed values and the values predicted by the linear model. The equation representing OLS is given by:

$$\hat{\beta}_{\text{OLS}} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \right\} \tag{7}$$

- Where $\hat{\beta}_{\text{OLS}}$ is the estimated parameter vector, y_i represents the observed values, x_i is the vector of predictor variables for the i-th observation, β is the parameter vector, and n is the number of observations.

- LASSO[10] (Least Absolute Shrinkage and Selection Operator) extends the OLS method by introducing a regularization term. This term penalizes the absolute size of the regression coefficients and thus helps in feature selection by shrinking some coefficients to zero. This is particularly useful in models with a large number of predictors, where it helps in selecting only the most significant variables. The LASSO equation is:

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_1 \right\} \tag{8}$$

- Here, λ is the regularization parameter, controlling the strength of the penalty. The term $\|\beta\|_1$ represents the L1 norm of the parameter vector, which is the sum of the absolute values of the coefficients.

- Ridge[11] Regression is similar to OLS but it includes a penalty on the size of coefficients. Unlike LASSO, Ridge Regression does not set coefficients to zero but shrinks them towards zero, which is advantageous in cases where many variables are correlated. It helps to prevent overfitting by adding a degree of bias to the regression estimates. The Ridge Regression formula is as follows:

$$\hat{\beta}_{\text{ridge}} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_2^2 \right\} \tag{9}$$

- In this equation, $\|\beta\|_2^2$ denotes the L2 norm of the parameter vector, which is the sum of the squares of the coefficients. The regularization parameter λ influences the extent of shrinkage applied to the coefficients.

V. Future Work

5.1 Another model

The limitations of queuing theory

In light of the temporal limitations inherent to our research project, the acquisition of additional data was regrettably unfeasible. This constraint on data collection notably curtails our capacity to employ more robust and intricate econometric models. Under ideal circumstances, with access to a more expansive dataset, it would be feasible to apply sophisticated mathematical techniques, such as ordinary or partial differential equations. These methods hold the potential to delineate with greater precision the intricate relationships between our target variable and various independent variables.

In such a scenario, these advanced mathematical frameworks would enable us to not only explore but also to elucidate the dynamic interplay between these variables. Particularly, the use of differential equations could offer a more nuanced understanding of how changes in the independent variables influence the target variable over time or across different conditions. This would allow for the derivation of precise analytical solutions, tailored to specific values of the independent variables, thereby significantly enhancing the accuracy and reliability of our predictive insights.

Moreover, this approach would permit a more granular examination of the causal mechanisms at play, potentially uncovering hidden patterns and correlations within the data. Such depth of analysis would invariably contribute to a more comprehensive and empirically grounded understanding of the phenomena under investigation, thereby enriching the overall quality and applicability of our research findings.

VI. References

[1] `https://pss-system.cponline.cnipa.gov.cn/conventionalSearch`.

[2] `https://data.cnki.net/`.

[3] `https://data.cnki.net/`.

[4] `https://flk.npc.gov.cn/`.

[5] Feng Dong, Yajie Liu, Policy evolution and effect evaluation of new-energy vehicle industry in China, Resources Policy, Volume 67, 2020, 101655,

[6] https://hts.usitc.gov/

[7] CHINA AUTOMOTIVE INDUSTRY YEARBOOK,2022.

[8] https://www.autoexpress.co.uk/tips-advice/106870/what-are-euro-7-emissions-standards

[9] Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Models (Fourth Edition) [M]. Beijing: Higher Education Press, 2011.

[10] Ma Li. MATLAB Mathematical Experiments and Modeling [M]. Beijing: Tsinghua University Press, January 2011.

[11] Si Shoukui, Sun Zhaoliang. Mathematical Modeling Algorithms and Applications [M]. 2nd Edition. Beijing: National Defense Industry Press, 2016.

VII. Appendix

7.1 Data sheet

Table 18 Data sheet

Question	File name
1	Question1and2.csv
2	Question1and2.csv
3	Question3.xlsx
4	Question4.xlsx
5	Bus emission.xlsx, Vehicle emission.xlsx

7.2 Code sheet

Table 19 Code sheet

Question	File name
1	`Question1.py`
2	`Question2_linear.py`, `Question2_predict.m`
3	`Question3_lasso.py`
4	`Question4_linear.py`, `Question4_Ridge.py`