



Contents List available at VOLKSON PRESS  
**World Symposium on Mechanical and Control  
 Engineering (WSMCE)**



## A STUDY OF OWNERSHIP FORECASTING BASED ON REGRESSION-COMBINATION OF CIVILIAN VEHICLES IN XI'AN

Sun Jian<sup>1\*</sup>, Shan Yan<sup>1</sup> · Xu Chao-Ran<sup>1</sup>, Sun Ya-Ru<sup>2</sup>, Song Jin<sup>1</sup>, Chen Nuo<sup>1</sup>

<sup>1</sup>Department of Automobile, Chang'an University, Art Road, Xi'an, china

<sup>2</sup>Department of Transportation, Shandong University of Science and Technology, Qian wan gang Road, Qingdao, china

\*Corresponding Author email: mrsunjian@outlook.com

### ARTICLE DETAILS

#### Article History:

Received 02 october 2017

Accepted 06 october 2017

Available online 11 november 2017

#### Keywords

Traffic Engineering, Car  
 Ownership, GM (1, 1) Modeling,  
 Exponential Smoothing, Linear  
 Regression

### ABSTRACT

To analysis the forecasting method of car ownership, this paper describes the recent research results of domestic scholars, then to forecast the ownership of civilian vehicles in Xi'an, GM (1, 1) model, exponential smoothing and linear regression method were established and analysis the pros and cons of the prediction method (the average relative error). The results show that: the analysis of ownership yearly changes if only rely on time-series data is more sensitive to the data itself and forecasting methods, different results of different forecasting methods; combination forecasting method can compensate for the lack of a single method, the quality of the text input variable combinations of the model will affect accuracy.

### 1. Introduction

With the improvement of people's living standards, people's dependence on travel tools increase obviously. Traffic is the link in social and economic activities, and in addition to traffic generated a large number of passenger flow, but also produce a very significant traffic flow. The operation of the vehicle optimizes the allocation of resources and car ownership and socio-economic development is also closely related. The forecast of car ownership is the basis for the development of road network planning and public transport planning in urban traffic and there are many ways to forecast car ownership. The combination of qualitative analysis and quantitative analysis is often highly persuasive. When the factors affecting the car ownership is not understood, the forecasting method based time series has become the universal forecasting method. In this paper, we first describe some forecasting methods proposed by scholars in recent years, and combined with regression analysis and use the gray forecasting model and the exponential smoothing model with the statistics data of Shaanxi province for the past 24 years to carry out case prediction, we propose a combination forecast model.

### 2. LITERATURE REVIEW

In order to forecast the amount of car ownership, a lot of analysis and discussion have been done in this paper, and the new forecasting method and combination forecasting method have been gradually applied from the basic forecasting method and a series of fruitful results have been obtained.

A researcher has points out the importance of car ownership in the relationship between transport, land use planning, energy consumption, the environment and health and based on the willingness of households to own vehicles, the system of accreditation, household income structure, household employment, the provisions of the company selected the vehicle and purchase and data of cost and other data, the econometric model is used to forecast the car ownership of 2031, and the social demographic characteristics of the UK are introduced in the model [1]. Based on the previous literature, based on the longitudinal data from 1995 to 2001, a researcher also analyzed the factors of Irish family car ownership using dynamic models. A Nolan believes that family travel behavior is a key factor, and income, family structure and other factors are

significant [2]. A group of researcher argue that the dependence of cars restricts the sustainable development of traffic and after analyzed the relevant studies in Latin America, Central America and some South American countries, they pointed out that the disaggregated model (individual income and way of life) is a more reliable method to explore such questions [3].

After analyzed of the various stages of China's economic development and taking into account the economic, population, car ownership growth rate and the distribution of per capita GDP in different regions and other factors and used the trend extrapolation, elasticity coefficient, domestic and foreign scenarios Analogy and transportation workload calculation and other methods, a group of scientists made a comprehensive forecast on China's future car ownership estimated that by 2030, China's car ownership will reach 2.3 to 250 million, about 1.7 to 1.9 times of 2020, 4.6 times in 2010 [4]. Based on dual BP neural network, From the perspective of the entire automotive industry and standing in the context of national economic development, in other study the researchers used per capita GDP and car ownership of the Gompertz curve model predicted the development trend of China's car ownership in the next few decades [5]. The article draws on the developed countries and Singapore and other emerging developing countries' development process which have achieved the car ownership. According to their experience, the results show that China's car ownership has been moving along the path of the Gompertz curve and has entered a period of high-speed development, and the trend of rapid growth will continue until about 2037 [5]. Based on the traditional Logistic model and using the gray theory, beside that generated the data processing method and the difference format and the prediction parameters are estimated [6]. The model can avoid the shortcomings of the predictive parameters and predictive environment normalization in the traditional forecasting model, so that the model can objectively and dynamically reflect the future trend of China's automobile ownership. Finally, combined with the historical data of China's civilian vehicles, he forecasted China's civilian car ownership of next decade [6]. Other research first use the gray relational analysis method to filter out the main indicators related to car ownership (gross national income, per capita GDP, total imports and exports, urban residents per capita disposable income, steel production, road passenger traffic and total retail sales of social consumer goods). And then the deterministic factor is used as the forecast index of automobile ownership to establish the vehicle ownership

forecasting model based on BP neural network. The results show that the BP neural network model has high accuracy, the maximum relative error is 2.2%, and the average relative error is 1.5% [7].

### 3. THEORETICAL REVIEW

#### 3.1 Exponential smoothing method

The exponential smoothing method is proposed by Robert G. Brown, and Brown argues that the time series is stable or regular, so the time series can be reasonably proclaimed. The exponential smoothing method is carried out as follows:

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

$S_t$  is the smooth value of time  $t$ ;  $x_t$  is the actual value of time  $t$ ;  $S_{t-1}$  is the smoothing value of time  $t - 1$ ;  $\alpha$  is the smoothing constant and the value range is  $[0, 1]$ . This article takes it as 0.5.

#### 3.2 Fundamentals of GM (1, 1) Prediction Model

The gray forecasting method is to establish a GM model from the past to the future according to the past and present known or not known information, so as to determine the trend of the system in the future development and provide the basis for planning decision. In the gray prediction model, the quantitative prediction is made by the time series, and the randomness is weakened, and the certainty is enhanced. In this case, the generating function is obtained at the generating level and the sequence prediction of the request sequence is set. The prediction model is a first order differential equation which is equivalent to a gray model with only one variable, denoted as GM (1, 1) model. Gray GM (1, 1) prediction model is mainly based on matrix in the calculation process. The combination of it and MATLAB effectively solves the problem of gray system theory in matrix calculation and provides a new application for gray system theory Methods.

#### 3.3 Multiple Linear Regression Analysis

For some common problems, the variable (the factor) of the dependent variable is more than one, but more than one, so that the regression of one of the dependent variables with multiple independent variables is multiple regression. When the relationship between multiple variables is linear, it is called multiple linear regression.

Let the dependent variable be  $y$  and the number of independent variables  $x_1, x_2, \dots, x_k$  is  $k$ . The equation describing how the dependent variable  $y$  depends on the independent variable and the error term  $\varepsilon$  is called the multiple linear regression model. The general form can be expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (6)$$

$x_1, x_2, \dots, x_k$  is called the explanatory variable or the independent variable,  $\beta_0, \beta_1, \dots, \beta_k$  is the model parameter;  $y$  is the explanatory variable or the dependent variable; The error term  $\varepsilon$  explains the part of the change in the dependent variable that can't be completely interpreted by the independent variable. Other random factors are not belong to linear relationship influence on  $y$ . First, the linear regression model has the following assumptions: 1) The parameters of the regression model are linear; 2) Random error terms are not related to explanatory variables; 3) The mean is zero; 4) The variance homogeneity is equivalent to the same variance; 5) Independence; 6) There is an incomplete collinearity between explanatory variables; 7) Normality.

In addition, the model needs to estimate the parameters and the target values (dependent variables) with known data which is often based on the least squares method, the estimated multiple linear regression equation is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad (7)$$

$\hat{y}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$  is the estimate of  $y, \beta_0, \beta_1, \beta_2, \dots, \beta_k$ . Define  $\hat{\beta}_k$  as the partial regression coefficient. Represents the magnitude of the effect of each unit on A when the argument is invariant. The magnitude of the effect

is usually referenced by the average amount of change. Define  $y_i - \hat{y}$  for the residual term denoted as  $e_i = y_i - \hat{y}$ , often called  $y_i = \hat{y}_i + e_i$  for the random sample regression function. The residual term  $e_i$  can be used as an estimate of  $\varepsilon_i$ .

### 4 BUILDING THE COMBINATION MODEL

In order to overcome the shortcomings of the above forecasting methods, this paper combines the predictive data of the two methods according to the regression analysis method. Using the two methods to establish a new model to obtain the prediction sequence (The data calculated from the input data). The model was established as follows:

$$y_c = \alpha x_g + \beta x_b + \varepsilon \quad (9)$$

In the above equation,  $y_c$  is the dependent variable (car ownership),  $\alpha, \beta$  is the independent variable coefficient,  $x_g$  and  $x_b$  are the gray prediction value and the exponential smoothing method predictor, and  $\varepsilon$  is the constant which the independent variable can't explain.

### 5 CASE ANALYSIS

#### 5.1 Application steps

1) Enter the original data; 2) The exponential smoothing method and the GM (1, 1) gray model are used to predict and get the initial prediction sequence. 3) The regression equations were obtained by regression analysis using the predicted sequences obtained by the two prediction methods as the independent variables and the original civilian vehicle ownership data as the target variables. 4) By inputting the expected annual data getting from the gray model and the exponential smoothing method into the regression equation, we can get the prediction with considerable accuracy.

#### 5.2 Data prediction

##### 5.2.1 Gray prediction and BP neural network prediction

The time series data are brought into the GM (1, 1) gray model and the first exponential smoothing prediction model. The results are shown in Table 1.

**Table 1:** Original data sequence and single prediction data sequence from 1991 to 2011

Years	Statistical value / (vehicle)	GM (1,1) prediction sequence / (vehicle)	Exponential smoothing prediction sequence / (vehicle)
1991	51584	51584	61720
1992	59993	13600.98577	56652
...	...	...	...
2011	1174874	443884.8348	798483.2

##### 5.2.2 Linear combination regression prediction

The original time series data, GM (1, 1) gray model sequence data and exponential smoothing sequence data are introduced into SPSS statistical analysis software. The related indicators of the regression model are shown in Table 2.

**Table 2:** Variance analysis of the model table

Model	sum of square	Degrees of freedom	Mean square	F	Significance
Return	1992848801897.590	2	996424400948.795	2279.264	.000
1 Residual	7869047175.647	18	437169287.536		
Total	2000717849073.238	20			

**Table 3:** Estimates and tests of model parameters

Model	Non - standardized coefficient		Standard coefficient	t	Significance
	B	Standard error	Beta		
1 (Constant)	-29283.013	8339.260		-3.511	0.002

GM (1, 1) data	-0.013	0.413	-0.005	-0.031	0.975
Once the index is smooth	1.520	0.248	1.003	6.132	0.000

From Table 2 to Table 3, we can see that the model has a high significance and the p test value of GM (1, 1) data is greater than 0.05. This paper takes into account the characteristics of the model independent variables and retains them. The resulting regression model is as follows:

$$y_r = 1.52x_e - 0.013x_g - 29283.013 \tag{9}$$

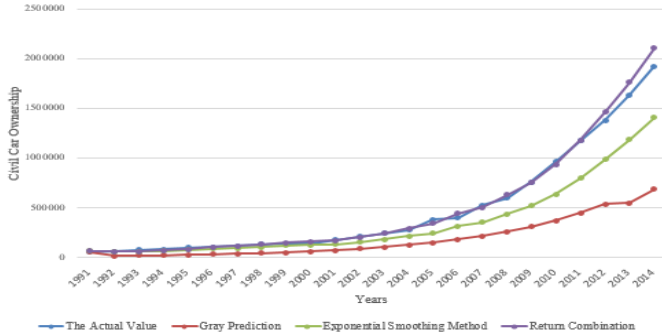


Figure 1: Raw data and forecast data curve

The data obtained from the predicted data are given to the data of Table 4 and Fig. 1 and their mean relative error and mean square error are analyzed. As can be seen from Table 4, the data obtained by the linear combination are minimized in the average relative error.

Table 4: Forecast data analysis

Time	Actual value	GM (1,1) ash prediction	BP neural network	Linear combination regression model
2012	1380125	533259.65	986678.6	1463536.081
2013	1634885	547737.79	1183402	1762367.13
2014	1924402	682541.96	1409143	2103741.909
Average relative error		0.6413	0.2763	0.0772

6 CONCLUSION

With the deep understanding of the nature of the planning, people are increasingly aware of the demand forecast (relying on car ownership). Based on the analysis of this paper, we get the following conclusions:

- a) Depending only on the time series data, the forecast of year-on-year changes in car ownership is more sensitive to the data itself and the forecasting method and the results of the different prediction methods are different.
- b) It can be seen from the predicted image that the predictive data obtained by the exponential smoothing method is smoother and the prediction of the data is more conservative. The GM (1, 1) gray prediction method has some fluctuation for the data obtained from the nonlinear model.
- c) The method of regression combination prediction can compensate for the deficiency of single method and the data quality of input variable of combination model will affect the accuracy. The result of regression combination forecasting method is reliable and has certain value.

REFERENCES

[1] Whelan, G. 2007. Modelling car ownership in Great Britain. *Transportation Research Part A Policy and Practice*, 41, 205-219. doi: 10.1016/j.tra.2006.09.013\*

[2] Nolan, A. 2010. A dynamic analysis of household car ownership. *Transportation Research Part A Policy and Practice*, 44, 446-455. doi: 10.1016/j.tra.2010.03.018\*

[3] Roque, D., Masoumi, H.E. 2016. An analysis of Car Ownership in Latin American Cities: a Perspective for Future Research. *Periodica Polytechnica Transportation Engineering*, 44, 5-12. doi:10.3311/PPtr.8307\*

[4] Yan-Li, M., Yue-E, G. 2007. Forecast of Automobile Population in China. *Journal of Highway and Transportation Research and Development*, 01, 121-125. Doi: 10.3969/j.issn.1002-0268.2007.01.029\*

[5] Hong-Mei, Z. 2012. The Medium and Long-Term Forecast of China's Vehicle Stock per 1000 Person Based on the Gompertz Model. *Industrial Technology and Economy*, 31, 7-23. doi: 10.3969/j.issn.1004-910X.2012.07.001\*

[6] Yu-Long, R., Rong, C., Yue-Feng, S. 2011. Prediction of Civil Vehicles' Possession Based on Combined Logistic Model. *Industrial Technology and Economy*, 30, 90-97. doi: 10.3969/j.issn.1004-910X.2011.08.016\*

[7] Ji-Bao, G., Rong, Y., Jian-Lin, W. 2010. A Combined forecasting of Chinese civilian vehicle quantity *Industrial Technology Economy*. *Industrial Technology and Economy*, 29, 123-127. doi: 10.3969/j.issn.1004-910X.2010.04.032\*

