Research on the Prediction of Retail Car Sales in California

Shuyi Yang*

College of Fac of Science & Engineering, University of Liverpool, Liverpool, L69 3BX, United Kindom

*Corresponding author: sgsyan17@liverpool.ac.uk

Abstract. Retail car sales are important because it is relevant to all the public including the government, market and consumer. With the influence of COVID-19 and many other factors, the change in retail car sales requires further research. Therefore, the prediction is in need of all sides to refer. Within all the models, the ARIMA model is more widely used, and the data satisfy the stationary required by the model, so it is chosen to be used in this research. The results of the forecast show that the retail car sales are going to decrease after the increase within the upper half of this year. However, whether the impact of the pandemic will last or it is a temporary factor remains a problem to solve. Conclusively, with the help of previous essays, the paper is expected to offer a forecast for retail car sales for the public to make decisions about retail car sales.

Keywords: Retail car sales; California; ARIMA model.

1. Introduction

Prediction for retail car sales is significant and useful. According to Nunnari et al., the prediction for sales is of paramount significance for companies, so huge manual and financial resources are devoted to perform this task in a careful and reliable way, and an efficient sales predicting model can assist the management process of supply chain and the ascending profits [1]. The factors that can influence retail car sales are sophisticated. According to Sivak et al., there are many important predictors of retail vehicle sales including the size of population, the price of gasoline, and the unemployment rate. Specifically, higher retail sales are the results of both more population and higher gas prices, and lower retail sales result from higher unemployment rates [2]. In recent years, COVID-19 has become another significant impact factor for retail car sales. The COVID-19 pandemic had a profound impact on the automotive industry, leading to significant disruption in car sales around the world. According to the data, early estimates indicated a dramatic drop ranging from -15% to -30% globally, with specific regions like Europe, China, and the United States experiencing sharp declines [3]. Additionally, compared to the general condition, focusing on one area to research is more reasonable. According to Andersen, the public is commonly interested in the prediction of the general conditions for business, while management is more interested in the conditions in the specific regions with specific purposes, so area analysis is important because the variance among regional markets may be significant, which can range from declining trends to ascending trends [4]. Since California is one of the biggest typical economic complexes in the US and globally, and to keep the variable factors less compared to the whole country and the globe, the deeper factors influence the rate can be noticed and adjusted. Thus, this article will focus on the retail car sales in California using R to analyze and interpret.

The topic of the essay is predicting future retail car sales in California. There is some existing research in the car industry. Pierdzioch et al. studied the car sales prediction in the United States and car registrations prediction in Japan [5]. Becker et al. explored the forecasts of electric vehicles in the United States to 2030 [6]. The characterizing of early adopter heterogeneity and forecasting of market diffusion for electric vehicle buyers in California was discussed by Lee et al. [7]. Cozzi and Petropoulos did research on the global SUV sales [8]. Guo and Zhou went deeply into analyzing the residual value of plug-in vehicles in the United States is [9]. The prediction for used cars' prices using random forest was explored by Pal et al. [10]. These researches vary in the method used, the type of car, and research area et al., and already have some conclusions about the car industry.

In summary, to predict the retail sales for general cars in a particular area, this article will use the time series model and the ARIMA model to do analysis and make interpretations. This research aims to help the car companies decide their selling plan for and the related organizations to make advanced policies to develop the car industry.

2. Methods

2.1. Data Sources

In the research, the data is collected from the Fred website to display the monthly amount of retail sales for motor vehicles and parts dealers in California in the unit of thousands of persons, and the data is seasonally adjusted from January 1990 to September 2023.

2.2. Variable Selection

Since the amount of retail car sales is a long-term statistic to follow, the whole statistic is concluded in the research for different aspects to research. The complete and whole statistics are there for the research using time series and segmented time series, finding the main rules of how these trends were changing, which are plotted below in Fig 1.

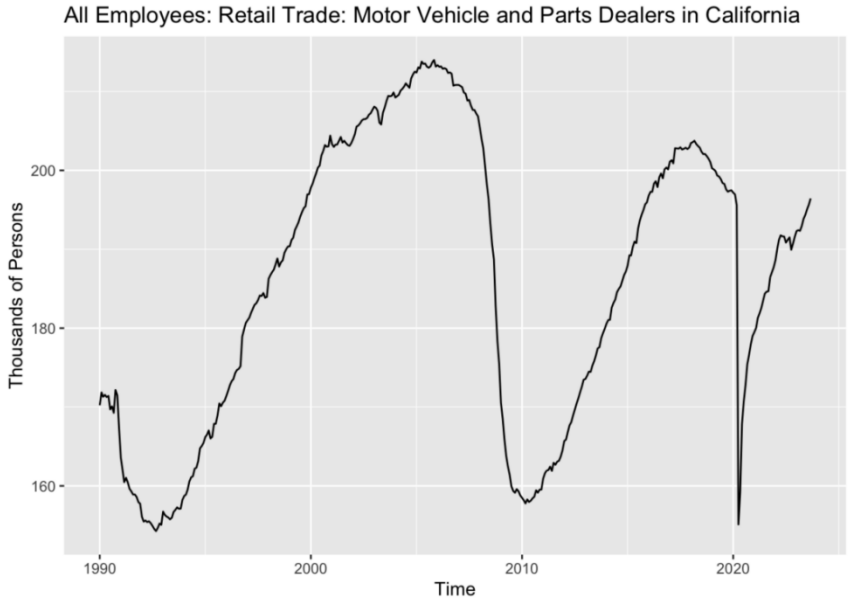


Fig. 1 All Employees: Retail Trade: Motor Vehicle and Parts Dealers in California, Thousands of Persons, Monthly, Seasonally Adjusted

Observation of the monthly time series plot of the amount of retail car sales reveals some simple conclusions. First of all, the retail car sales amount has a cyclical variation, but the cycle is not fixed. Furthermore, the level of the unemployment rate may have a decreasing trend over the time frame of the data.

2.3. Data Processing

The ACF plot in Fig 2 shows the data does not have a strong auto-correlation. The PACF plot in Fig 3 shows that although the first strip exceeds the blue line, the others are under the blue line. Through the result of the ADF test, the p-value is 0.6843, so it is larger than 0.05. So the data satisfies stationary.

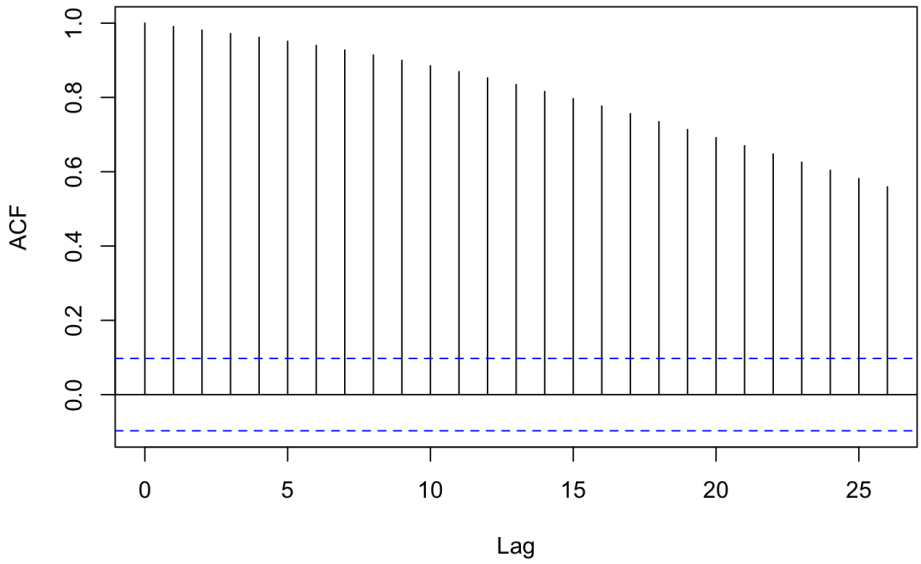


Fig. 2 The ACF plot of the retail car sales

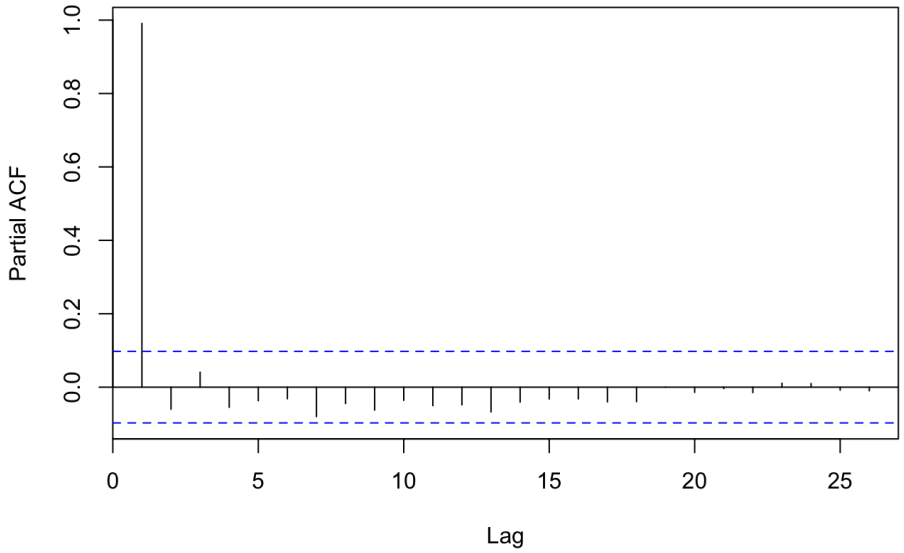


Fig. 3 The PACF plot of the retail car sales

2.4. Model Selection

Auto-regressive Integrated Moving Average (ARIMA) model, as a statistical model, is commonly used for time series analysis and prediction. It is a useful tool for capturing patterns and trends in data that evolve over time. Since the data satisfies the stationary required by the ARIMA model, it is chosen to establish a forecasting model for California's retail car sales. The ARIMA model has three key components: auto-regressive (AR), integrated (I), and moving average (MA).

2.5. Model Evaluation

Assess the model's performance by comparing the fitted values to the actual values. Since AIC balances model performance and complexity and helps prevent over-fitting, this assessment is chosen. According to the AIC minimum criterion, to find the best model, different ARIMA models with different parameters should be tested, and the one with the least AIC value can be found. Since the original data satisfies the required stationary, the following eight ARIMA models in Table 1 are chosen and their AIC values are checked. Through comparison, ARIMA(1,0,1) has the lowest AIC value which is 1827.09, so this is the optimal model and further prediction will be based on this model.

Table 1. Model Evaluation

Model	AIC
ARIMA(1,0,0)	1828.35
ARIMA(0,0,1)	3017.23
ARIMA(2,0,0)	1827.45
ARIMA(0,0,2)	2669.2
ARIMA(1,0,2)	1827.89
ARIMA(2,0,1)	1827.59
ARIMA(1,0,1)	1827.09
ARIMA(2,0,2)	1829.47

3. Results and Discussion

3.1. Forecasting

The main result is that: the forecasting shows that the retail car sales will decrease in the future, and the exact sales are also predicted and recorded in Table 2 and shown as a blue decreasing line in Fig 4.

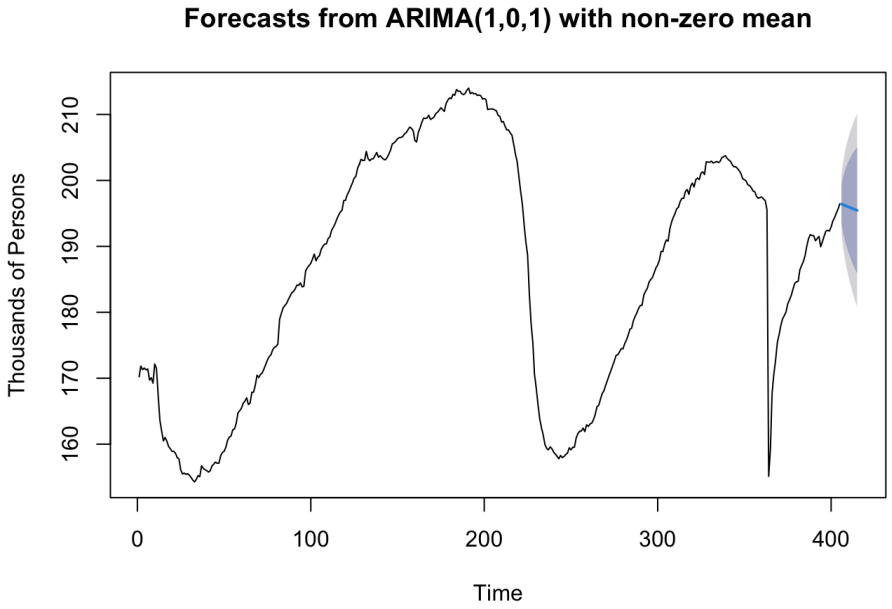


Fig. 4 Forecasting with ARIMA(1,0,1)

Table 2. Predicted value with ARIMA(1,0,1)

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
196.4080	193.4930	199.3230	191.9499	200.8661
196.2979	191.9937	200.6021	189.7152	202.8806
196.1889	190.8642	201.5136	188.0455	204.3323
196.0810	189.9181	202.2438	186.6557	205.5062
195.9742	189.0883	202.8600	185.4432	206.5051
195.8684	188.3412	203.3956	184.3566	207.3802
195.7638	187.6572	203.8703	183.3659	208.1616
195.6602	187.0236	204.2967	182.4517	208.8686
195.5576	186.4315	204.6837	181.6004	209.5148
195.4561	185.8745	205.0376	180.8024	210.1098

3.2. Critical Thinking

Although the result seems to be figured out, there are still some points that need to be considered. One is that COVID-19 must have a great impact influence on retail car sales in California because in Fig 1, though the general change that happened around 2020 has some similarities with that happened from 1992 to 2000 and from 2007 to 2015, there is a very sharp decrease and a very sharp increase in 2020 when the pandemic is in serious condition, and these changes are obviously sharper than the changes before. So even if the prediction with the preceding data is produced, whether the impact of the pandemic will last or what hidden long-lasting impact it will have on retail car sales in California is not clear. Additionally, how to assess and solve the sharp change remains a problem that needs to be solved.

4. Conclusion

This research used the classical time series model ARIMA to analyze the retail car sales data in California, and this model is able to provide a clear view of the trends of retail car sales in recent times and conclude reasonably. From the data and the time series model analysis, the final conclusion is that under this situation, retail car sales have a trend of decreasing after the recent increase, even if the pandemic might take more roles in the prediction, the result of the prediction can contribute to the research on retail car sales in California.

There are also some drawbacks to the research. The data used include the data collected between 2019-2021 when the pandemic was a serious factor that can impact retail car sales. It can be seen from the time series plot of the whole data that the change in retail car sales during the pandemic time is obviously sharp. However, the influence of the pandemic is not assessed and it can not be determined whether it is a temporary factor or has a long-lasting impact, so whether extra test on the data collected during this time is needed is still a problem that needs to be solved. It would be hard yet necessary to analyze these in further research.

Looking into the future with the data and analysis obtained, there are some suggestions or ideal policies to help the automobile retail industry recover from this recession or continue with the stable trend that the industry might be in in the future.

To think of suggestions, it is important to further research the specific cause of the decrease and implement appropriate policies to counteract the causes. The government can provide subsides and reduce taxes for the buyers to stimulate the consumption. For the car retail seller, it is required to create new selling mode such as developing online selling platform and virtue car exhibition. Under the energy risk, the car industry should do research on electric car to defend people's concern about rising oil price and lack of affordable oil. And there are other factors such as retail car production and retail car purchase which are closely-related to retail car sales. The government should do more research on these fields at the same time to better revitalizing the retail car industry. If the problem cannot be solved quickly, the government should also think about ways to make new balance between car production and sales to avoid the problems resulting from mismatching. It is crucial to make actions to quickly balance the retail automobile market.

References

[1] Nunnari G, Nunnari V. Forecasting Monthly Sales Retail Time Series: A Case Study. IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, Greece, 2017, 1-6.

[2] Sivak M, Schoettle B. Economic Factors and Retail Sales of New Vehicles in the United States. The University of Michigan Sustainable Worldwide Transportation, 2017.

[3] Wyman O. Car Demand in times of COVID-19. Working paper, 2020.

[4] Andersen T A. Application of Four Major Methodologies in Projecting New Car Sales in California Illustrates the Problems of-Regional Industry Forecasting. California Management Review, 1966, 8(3) : 51-56.

[5] Pierdzioch C, Rülke J C, Stadtmann G. Forecasting US car sales and car registrations in Japan: Rationality, accuracy and herding. Japan and the World Economy, 2011, 23(4): 253-258.

[6] Becker T A, Sidhu I, Tenderich B. Electric vehicles in the United States: a new model with forecasts to 2030. Center for Entrepreneurship and Technology, University of California, Berkeley, 2009, 24.

[7] Lee J H, Hardman S J, Tal G. Who is buying electric vehicles in California? Characterizing early adopter heterogeneity and forecasting market diffusion. Energy Research & Social Science, 2019, 55: 218-226.

[8] Cozzi L, Petropoulos A. Global SUV sales set another record in 2021, setting back efforts to reduce emissions. Working paper, 2021.

[9] Guo Z, Zhou Y. Residual value analysis of plug-in vehicles in the United States. Energy Policy, 2019, 125: 445-455.

[10] Pal N, Arora P, Kohli P, Sundararaman D, Palakurthy S S. How much is my car worth? A methodology for predicting used cars' prices using random forest. In Advances in Information and Communication Networks: Proceedings of the 2018 Future of Information and Communication Conference (FICC), 2019, 413-422.