The Impact of Covid-19 on Used Car Sales

Jingying Pan
Computing ID: jp9pbd
Statistical Consulting STAT7995
Advised by: Prof. Karen Kafadar

November 21, 2020

Abstract

Since March 2020, the Covid-19 spread very quickly in US. This project analysed the used car sales since January 1992 to September 2020. The used car sales before and after the Great Recession fitted the same model. So the Great Recession doesn't change the pattern of the used car sales. It just move the sales number to a lower level. With the assumption that the sales data after Covid-19 Recession would also fit our model before Covid-19, we added the mean residuals of the last four months to the model prediction of the future 12 months. In this way we get our final prediction between Oct.2020 and Sep.2021.

1 Introduction

Since December 2019, the Coronavirus(Covid-19) spread in central China. Since March 2020, it spread very quickly in US. There are 11,792,715 positive cases and 255,428 deaths till Nov.18. Most states locked down in March and after that. In June, some states reopened the schools, restaurants and it caused a quick increasing on the number of positive cases. This fatal pandemic has changed how we work, learn and interact as social distancing guidelines have led to a more virtual existence, both personally and professionally. People avoid mass transportation and are more sensitive to auto cost in the recession, so the used car sales increased during the pandemic. By April and May, with a lack of new cars from auto plants able to hit dealer lots, and consumers more cautious about spending on big items, used car sales boomed.

In history, we also can find some similar period that used car sales are affected by market environment, such as the great recession in 2008. The great recession started in the U.S. officially in December 2007 and lasted until June 2009 [1]. It had a significant economic and political impact on the United States.

In this project, we plan to use the used car sales data since Jan. 1992 to Sep. 2020 and compare the impact of great recession and the covid-19 recession on used car sales. The

objectives of this project are: 1. construct models to fit the used car sales data before the great recession in 2008 and after that; 2. analyse the impact of the recession by comparing the models and investigate the impact of Covid-19; 3. predict the used car sales of 2021.

2 Data

The original dataset is from U.S. Census Bureau, Retail Sales [2], retrieved from FRED, Federal Reserve Bank of St. Louis. It contains 2 variables: Date, Sales. The 345 observations ranges from Jan. 1992 to Sep. 2020.

3 Data Processing

3.1 Recession Periods

In this project, the two recession periods are:

- *Great Recession*: Dec. 2007 to Jun. 2009
- *Covid-19 Recession*: started Mar. 2020

3.2 Plots Explanation

In this project, the colors in the plots represent:

- **black**: true values used to construct models
- **red**: prediction by the models
- **blue**: true values before the recession used to test the models
- **green**: true values during the recession
- **yellow**: true values after the recession used to analyse the impact of the recession
- **purple**: our final prediction after analysing the models and the recession impact

3.3 Seasonal Decomposition

Before analysing the data, we visualized the structure of the data, as shown in ***Figure*1**. From the figure we can see that the data shows an overall increasing trend and a strong seasonal component. In 2008, used car sales had a big slump during the recession. And during Covid-19 recession, the used car sales decreased at first and then rose up quickly. Due to the lockdown, the auto factory are suspended and people's requirements for isolate transportation increased, which we believe are the major reasons for the change of used car sales during Covid-19.

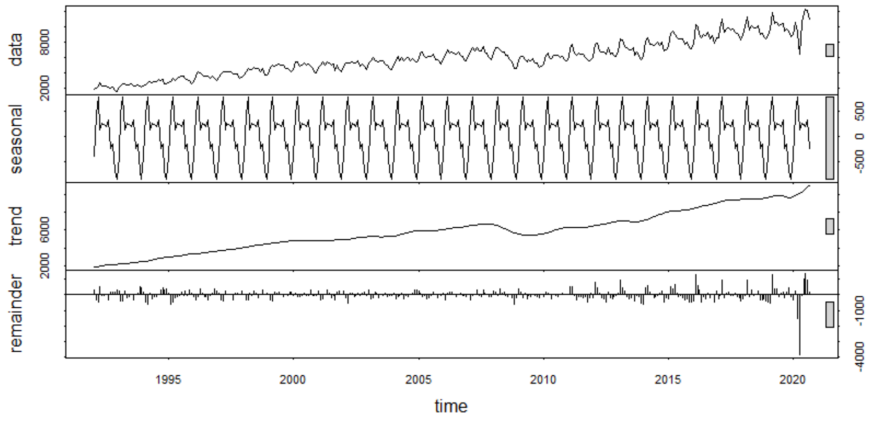

Figure 1: Seasonal Decomposition of Used Car Sales

3.4 Construct Models and Impact Analysis

Classical regression is often insufficient for explaining all of the interesting dynamics of a time series. Based on the features of the data, we constructed SARIMA models to fit the data.

Firstly, we split the data into two parts: before(including) the Great Recession and after the Great Recession, more specifically, before Jun. 2009(obs. 1 to 210) and after Jul. 2009(obs.211 to 345). Then, we use the data before(including) Nov. 2006(obs. 1 to 179) to construct a SARIMA model(we call it ***ModelI***), and use the data between Dec.2006 and Nov.2007(obs.180 to 191) to test ***ModelI***.

After first order differencing and seasonal differencing(we use 12 as frequency because it is a monthly dataset), the data looks roughly stationary. So we use d=1 and D=1 in our SARIMA model.

After trying several parameters according to ACF and PACF plot we finally choose SARIMA$(1,1,1)\times(0,1,1)_{12}$ as the fitting model. The model diagnosis is shown in ***Figure2***. There is no clear pattern in the residuals with constant variance. ACF shows no significance for all lags. Residuals fall closely to the diagonal line in the QQ-plot. Although in the Ljung-Box p-values of lag 5 fall under the blue line, the model can predict the sales very well(compared to SARIMA$(4,1,1)\times(0,1,1)_{24}$, which qualifies all the criteria but two of the true values between Dec.2006 and Nov.2007 are outside the prediction interval).

After that we use ***ModelI*** to predict the used car sales between Dec.2007 to Jun.2010(obs. 192 to 222). As shown in ***Figure3***: the red line represent the model predictions; the blue line represent the true values true values between Dec.2006 and Nov.2007, which is 12 months before the Great Recession; the green line represent the true values between Dec.2007 and Jun.2009, which is during the Great Recession; the yellow line represent the true values between Jul.2009 and Jun.2010, which is 12 months after the Great Recession. By comparing the predicted value with the true value we can see that almost all of the true values between

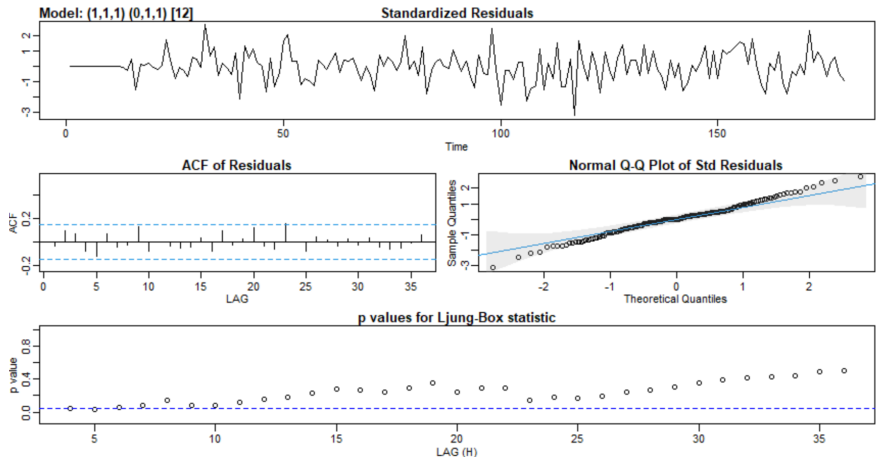

Figure 2: Diagnosis of Model I, SARIMA$(1,1,1) \times (0,1,1)_{12}$

Dec.2006 and Nov.2007(blue) are located in the 95% prediction interval. The errors in the prediction is acceptable because the historical seasonal pattern are not exactly same, there must be some unexpected factor that can affect the used car sales. As expected, the predicted values during the Great Recession deviate from the true values. After the recession, the used car sales didn't go back to the prediction, it followed the sales in the end of the recession, while, the pattern looks similar to the prediction.

There are 43 prediction values in ***Figure3***. We use a scatter plot to analyse the deviation in detail, as shown in ***Figure4***. The blue points represent the ***ModelI*** prediction residuals between Dec.2006 and Nov.2007, which is 12 months before the Great Recession. The green points represent the residuals between Dec.2007 to Jun.2009, which is during the Great Recession. The yellow points represent the residuals between Jul.2009 and Jun.2010, which is 12 months after the Great Recession. In the 12 months before the Great Recession(even during the first 5 months of the recession), the residuals are lower than 0 but roughly "stable". The predictions are lower than true values by a relatively table value. During the Great Recession, the residuals rose up quickly. In the 12 months after the Great Recession(plus the end of the recession), the residuals are "stable" again. This verifies our assumption that the used car sales after the Great Recession followed the sales in the end of the Great Recession with the pattern of the model prediction. Also, the impact seems terminated before the Great Recession finally disappeared.

Then we constructed a SARIMA model for the used car sales after the Great Recession(we call it ***ModelII***). Similarly, we use the data between Jul.2009 and Feb.2019 to construct ***ModelII***, and use the data between Mar.2019 and Feb.2020 to test ***ModelII***. The diagnosis of ***ModelII*** is in ***Figure5***. There is no clear pattern in the residuals with constant variance. ACF shows no significance for all lags. Residuals fall closely to the diagonal line in the QQ-plot. In the Ljung-Box, all the p-values fall above the blue line. Although there are two true values between Mar.2019 and Feb.2020 fall outside the 95% prediction intervals, it doesn't deviate too much. So, after checking the diagnosis and the prediction between Jul.2009 and Feb.2019 we choose SARIMA$(1,1,1) \times (0,1,1)_{12}$ as our ***ModelII***. By

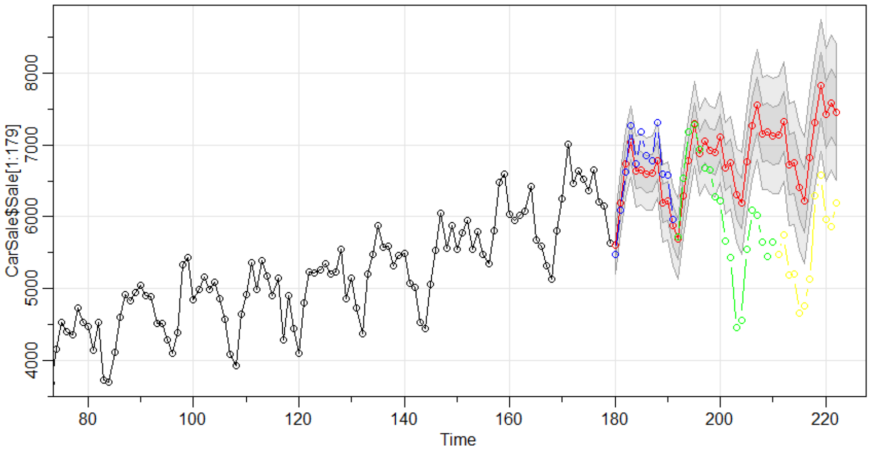

Figure 3: Model I Prediction and True Values(**black**: true values between Jan.1992 and Nov.2006, used to construct Model I; **red**: model predictions; **blue**: true values between Dec.2006 and Nov.2007,12 months before the Great Recession; **green**:true values between Dec.2007 to Jun.2009, during the Great Recession; **yellow**: true values between Jul.2009 and Jun.2010, 12 months after the Great Recession)

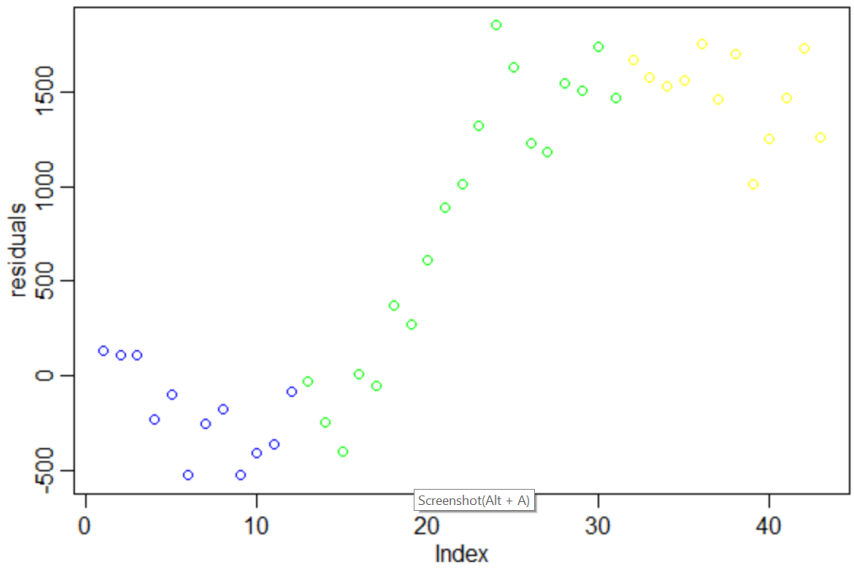

Figure 4: Model I Residuals(**blue**: between Dec.2006 and Nov.2007, 12 months before the Great Recession; **green**: between Dec.2007 to Jun.2009, during the Great Recession; **yellow**: between Jul.2009 and Jun.2010, 12 months after the Great Recession)

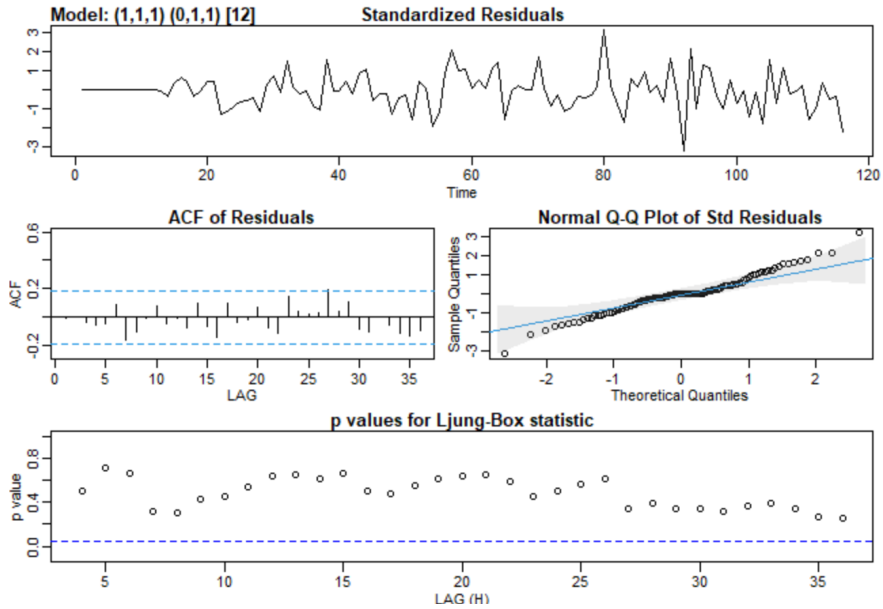

Figure 5: Diagnosis of Model II, SARIMA$(1,1,1) \times (0,1,1)_{12}$

comparing it with ***ModelI*** we can see that they are the same model. This (again) verifies our assumption that the Great Recession doesn't change the pattern of the used car sales, it just move the sales number to a lower level. So we assume that the Covid-19 Recession would have similar impact on the used car sales. It changes the sales number but doesn't change the pattern.

Figure6 shows the ***ModelII*** prediction and the true values during and before Covid-19 Recession. The red line represent the prediction values by ***ModelII***. The blue line represent the true values of 12 months before Covid-19 Recession. The purple line is our final prediction for 12 months(we will explain how we get it later).

To make it more clear, we use a scatter plot of the residuals for ***ModelII***, shown in ***Figure7***. The blue points represent the residuals of ***ModelII*** prediction between Mar.2019 and Feb.2020, which is 12 months before the Covid-19 Recession. The green points represent the residuals of ***ModelII*** prediction between Mar.2020 to Sep.2020, which is during the Covid-19 Recession. In the 12 months before the Covid-19 Recession, the residuals looks quite "stable" even though they are below 0(which means we underestimated the used car sales). Then, the Covid-19 affected the sales to deviate from the model. While, the last four points(Jun.2020 to Sep.2020) look very similar. So we assume that the impact of Covid-19 lasts for three month (Mar.2020 to May 2020). It changes the sales value to the sales of Jun.2020. With this assumption, the residuals after Jun.2020 would be also "stable". Then we calculate the mean residual of the last four months(Jun.2020 to Sep.2020). As we have discussed before, the Great Recession doesn't change the pattern of used car sales. We assume that Covid-19 won't change it as well. So we add the mean residual to ***ModelII*** prediction between Oct.2020 and Sep.2021(12 months after the original dataset) and thus we get our final prediction.

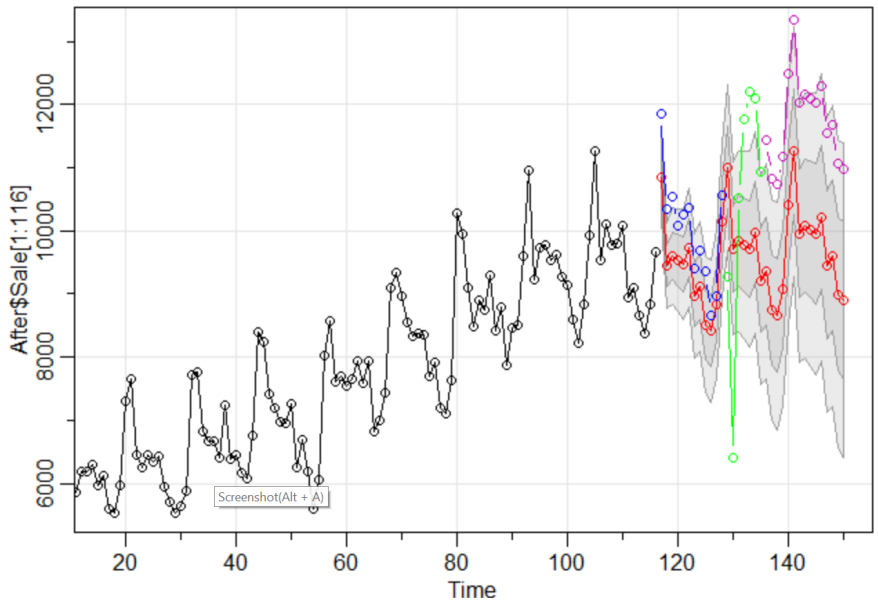

Figure 6: Model II Prediction and True Values(**black**: true values between Jul.2009 and Feb.2019, used to construct Model II; **red**: model predictions; **blue**: true values between Mar.2019 and Feb.2020; **green**:true values between Mar.2020 to Sep.2020 - the Covid-19 Recession; **purple**: our final prediction between Oct.2020 and Sep.2021)

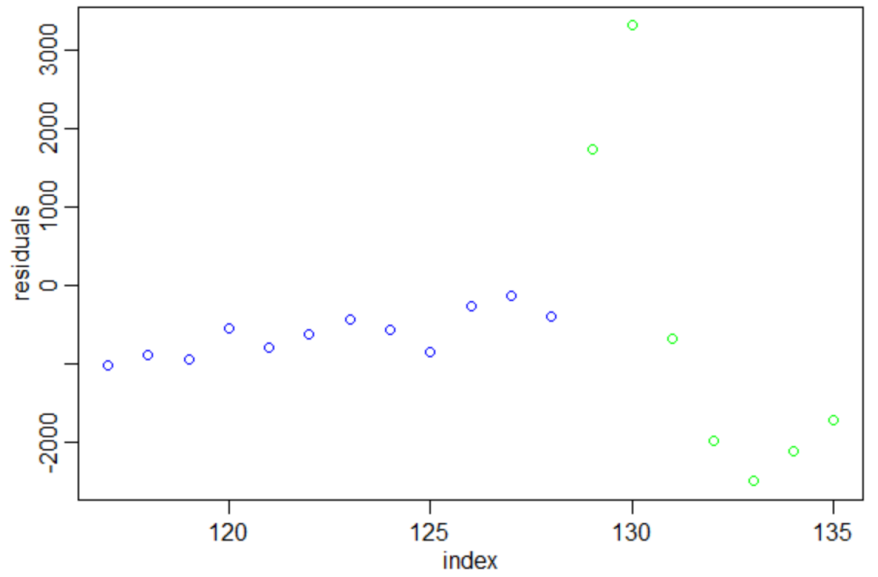

Figure 7: Model II Residuals (**blue**: between Mar.2019 and Feb.2020, 12 months before Covid-19 Recession; **green**: between Mar.2020 to Sep.2020, during Covid-19 Recession.)

4 Conclusions

The used car sales before the Great Recession fits our model SARIMA$(1,1,1)\times(0,1,1)_{12}$, the sales after the Great Recession also fits the same model. So the Great Recession doesn't change the pattern of the used car sales. It just move the sales number to a lower level. So we assume that the Covid-19 Recession would have similar impact on the used car sales, it would change the sales number to a different level but not change the pattern, which means that the sales data after Covid-19 Recession would fit our model before Covid-19, SARIMA$(1,1,1)\times(0,1,1)_{12}$. The Covid-19 Recession has changed the used car sales to around 11761(Jun.2020). By adding the mean residuals of the last four months to the prediction of the future 12 months, we get our final prediction between Oct.2020 and Sep.2021.

5 Limitations

Our model didn't correctly predict all the values for 12 months before the recession. There must be other factors that can affect the used car sales. We might be able to analyse these factors if we investigate the used car market more carefully.

Further more, as of the writing date of this project, we didn't get the updated sales data for the time later than Sep.2020. In our assumption, the mean residuals of Jun.2020 to Sep.2020 are stable. However, we are not sure whether we have seen the end of the Covid-19 impact on the used car sales(with analysis of the Great Recession we guess that the impact might end before the recession finally ends). Pfizer and Moderna announced very promising trial data on the Covid-19 vaccines this month. Hopefully the vaccines can bring us back to normal life though the Covid-19 Recession is still far from the end.

References

[1] *Great Recession*, available at https://en.wikipedia.org/wiki/Great_Recession

[2] https://fred.stlouisfed.org/series/MRTSSM44112USN