Intervention Analysis to Analyze the Effect of COVID-19 on Used Car Sales and Forecasting in the US

Marc Edwards

Time Series Analysis and Forecasting

Spring 2023

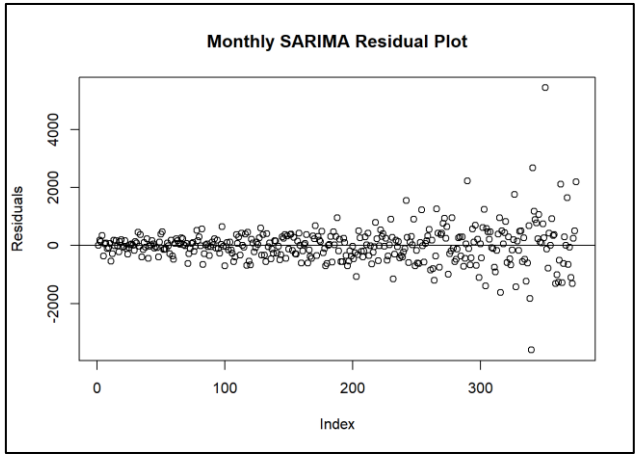
Problem Statement

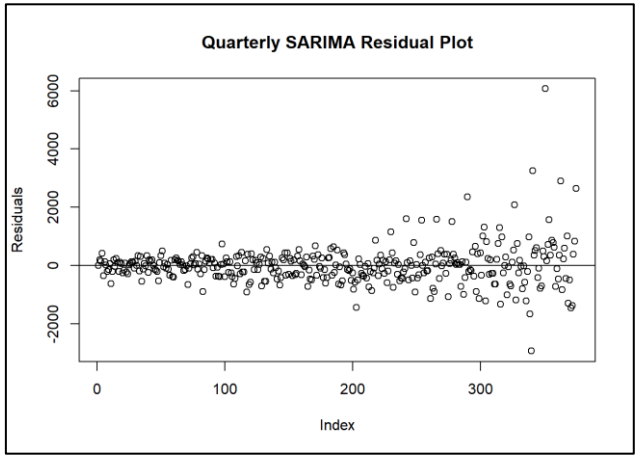
- Historical trends in used car prices can provide valuable insights into the current market.
- Prices for used cars have been on the rise due to the increasing demand for used cars and the limited supply.
- The economic downturn caused by the COVID-19 pandemic led to a decrease in the production of new cars, which further increased the demand for used cars.

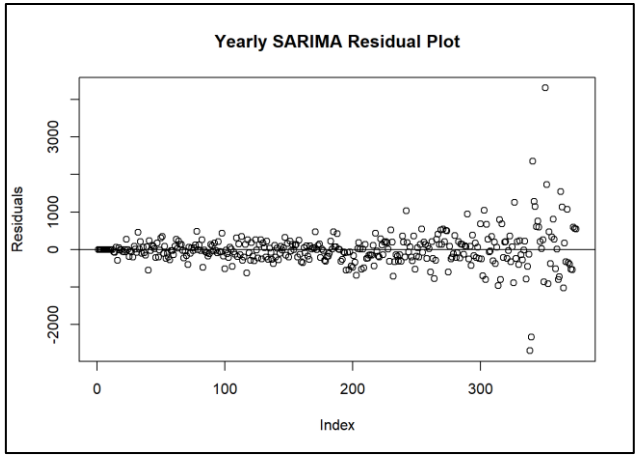
How did the COVID-19 pandemic affect the used car market in the US?
How do different methods and models forecast future sales?

Assumptions & Hypotheses

1. **Seasonal Patterns**: The hypothesis is that the time series data exhibits recurring seasonal patterns or cycles. These patterns could be monthly, quarterly, or annual cycles. The assumption is that the patterns repeat consistently across multiple cycles.
2. **Additive or Multiplicative Seasonality**: The assumption is that the seasonal component can be modeled as either an additive effect, where the amplitude of the seasonal pattern remains relatively constant, or a multiplicative effect, where the amplitude of the seasonal pattern varies with the level of the time series.
3. **Seasonal Stationarity**: The assumption is that the seasonal component of the data is stationary, meaning that the statistical properties of the seasonal patterns do not change over time. This assumption is necessary for modeling and forecasting the seasonal component accurately.
4. **Independence and Identically Distributed (IID) Residuals**: The hypothesis is that the errors or residuals of the time series model are independent and identically distributed. In the data, the residuals' distribution begin to change after the 300th index. That is the start of the COVID-19 pandemic; that intervention will be tested later.
5. **Homoscedasticity**: The assumption of homoscedasticity implies that the variance of the errors remains constant across different time points and does not exhibit any systematic patterns. In the data, the spread of the residuals begins to increase in a heteroskedastic manner after the 300th index. That is the start of the COVID-19 pandemic; we will test the impact of that intervention later.
6. **Absence of Outliers**: The assumption is that the time series data is free from outliers or influential observations that can significantly impact the model's performance and accuracy. Outliers should be identified and handled appropriately before modeling. COVID-19 presents some outliers later in the data.



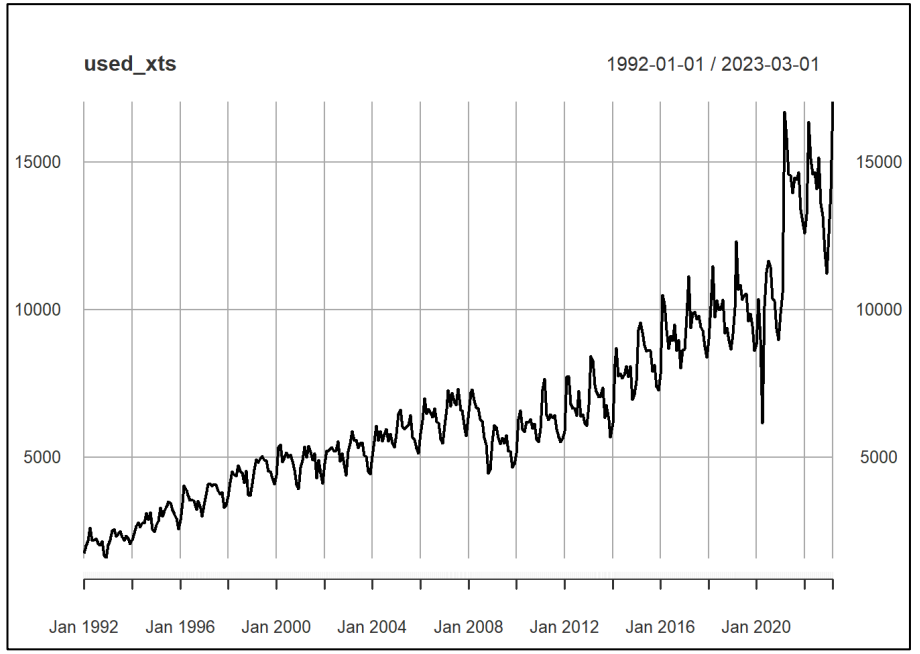


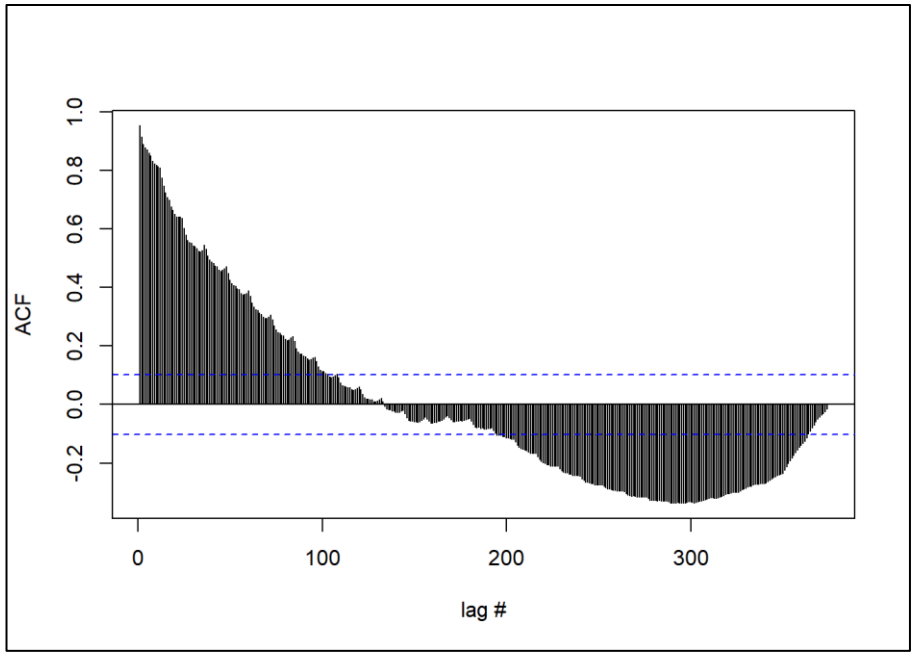


Data Properties

- The ACF plot shows the ACF dying down very slowly as the lag increases.
  - This gives us the evidence that the **time series is non-stationary.**
- The ADF test has a p-value of 0.4946, we do not reject the null hypothesis (at a 5% level).
  - We have significant evidence that the **time series is non-stationary**.
- The KPSS Test has a p-value of 0.01, we reject the null hypothesis (at a 5% level).
  - We have significant evidence that the **time series is non-stationary**.

Test	Test Statistic	P-Value
Augmented Dickey-Fuller Test	-2.1953	0.4946
KPSS Test for Level Stationarity	5.3905	0.01





Feature Engineering & Transformations

- The original dataset was transformed into ‘xts’ and ‘ts’ objects for its respective tasks during modelling.
- The data was clean, had no missing data, and had consistent observation times (1st of every month).
- No additional feature engineering or data transformations were required for analysis.

DATE	Revenue, Millions of $USD
1992-01-01	1744
1992-02-01	1990
1992-03-01	2177
1992-04-01	2601
…	…
2022-12-01	11219
2023-01-01	12570
2023-02-01	14100
2023-03-01	17039

Proposed Approaches

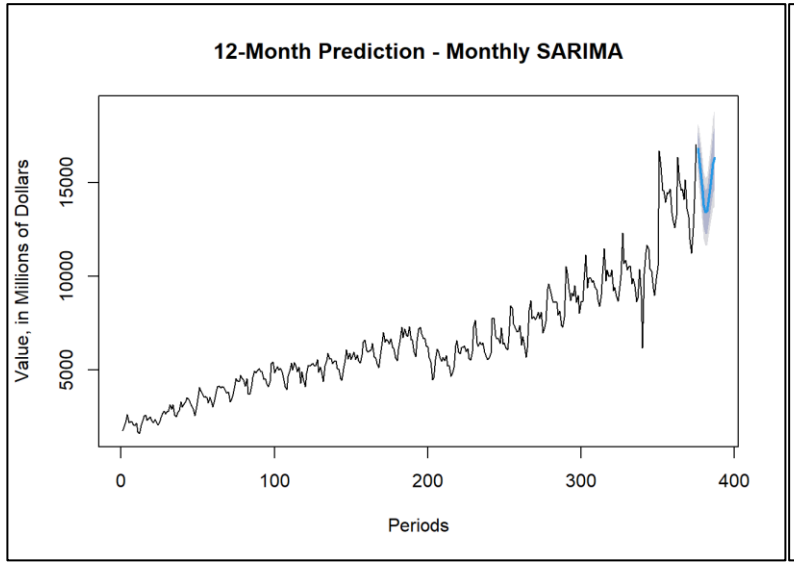
1. Monthly, Quarterly, and Yearly SARIMA Models for Forecasting

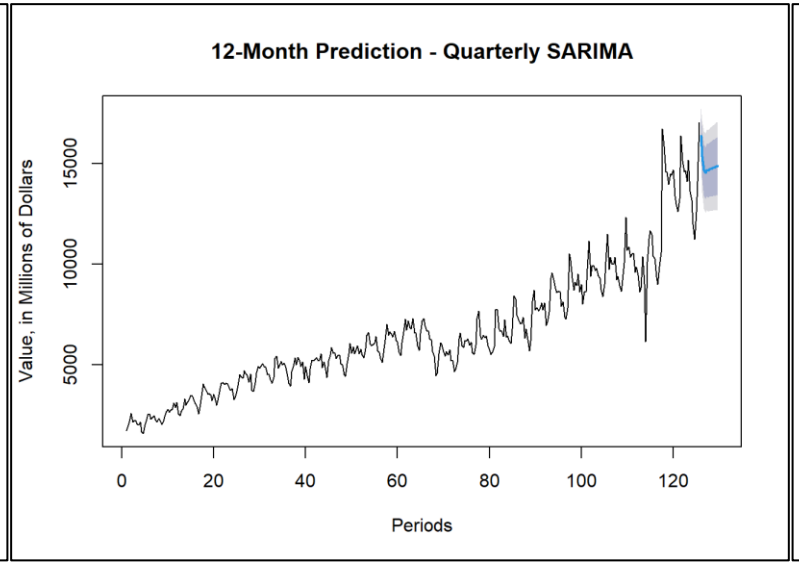
- SARIMA models extend the ARIMA framework to handle time series data with seasonal patterns.
- SARIMA models include additional seasonal components to capture the seasonality present in the data.
- SARIMA models can effectively capture and forecast time series data that exhibit seasonal patterns, such as monthly, quarterly, or yearly cycles.

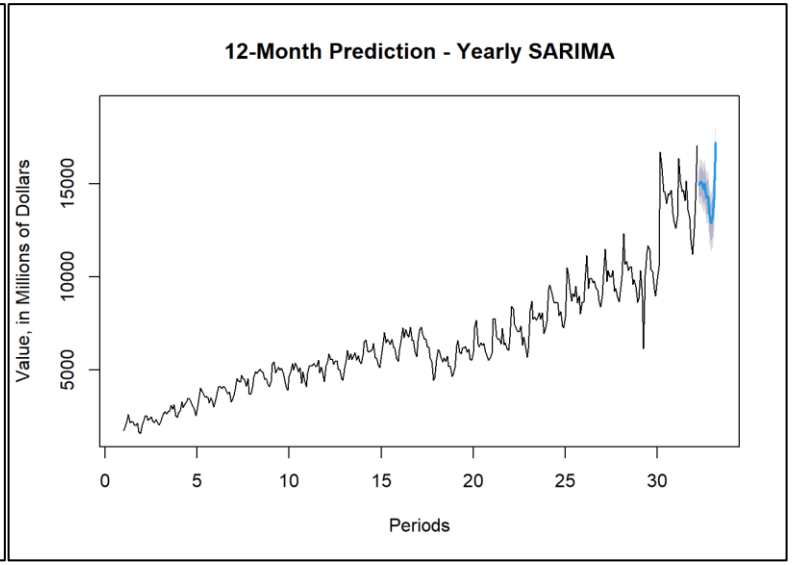
2. Pulse and Step Intervention Analysis (ARIMAX Prediction)

- Analyze and quantify the impact of the COVID-19 pandemic on used car sales.
- It helps to understand whether and how COVID-19 caused a significant change or disruption in the observed time series.
- The goal of intervention analysis is to assess the causal effect of COVID-19 on the time series and provide insights into its impact.

Results - SARIMA







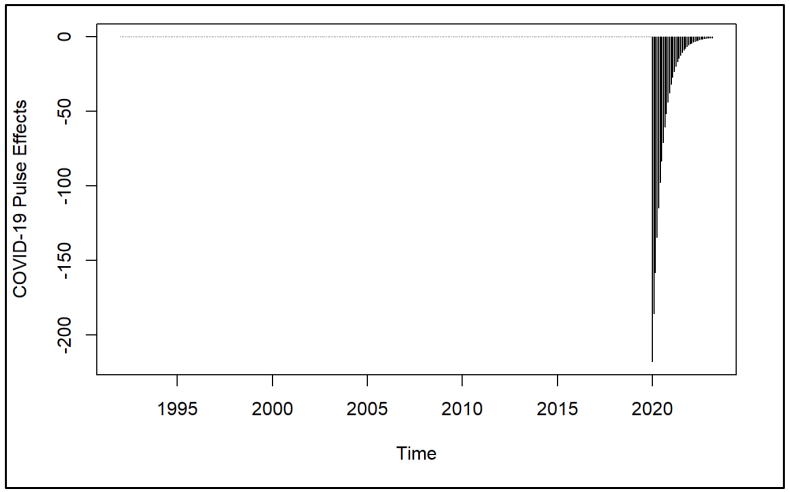
Model Order	(2,1,3)
Durbin-Watson (Autocorrelation)	1.948138
Box-Ljung P-Value (Autocorrelation)	0.8370336
Result	Not Autocorrelated

Model Order	(2,1,2)(0,0,1)
Durbin-Watson (Autocorrelation)	1.97895
Box-Ljung P-Value (Autocorrelation)	0.8532571
Result	Not Autocorrelated

Model Order	(3,0,2)(0,1,2)
Durbin-Watson (Autocorrelation)	1.935789
Box-Ljung P-Value (Autocorrelation)	0.5553718
Result	Not Autocorrelated

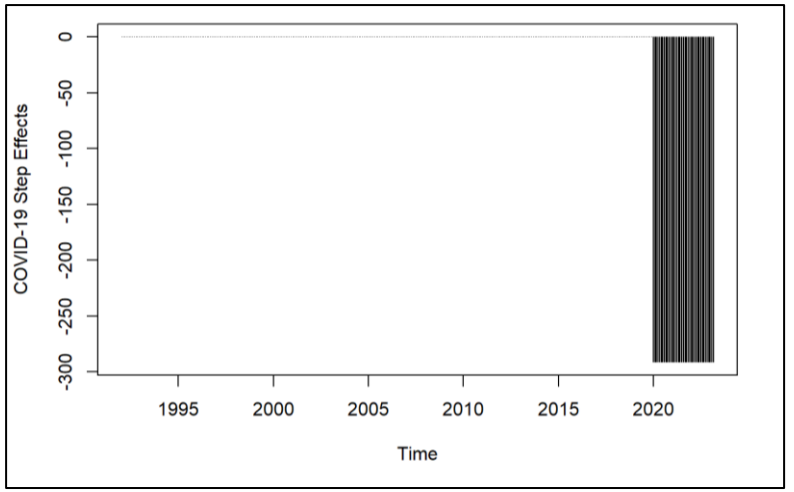
Results – Intervention Analysis

Pulse Effect



- Negative temporary change in the sales of used cars in the US.
- Very quickly recovers to original level.
- Cannot determine what happens to time series after returning to initial level.

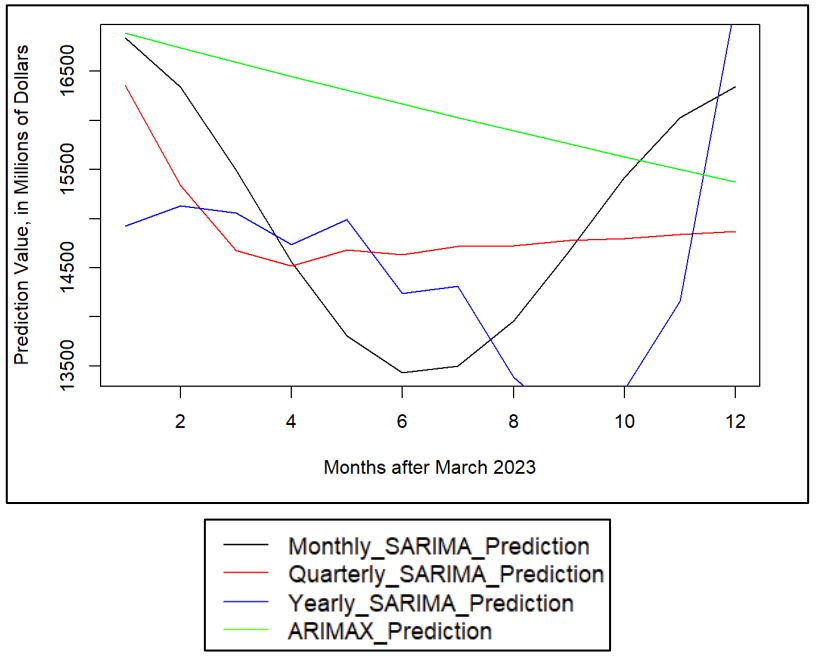
Step Effect



- Downward shift in the sales of used cars in the US until return to its initial level.
- Cannot determine what happens to time series after returning to initial level.

Conclusions

- The SARIMA forecasts follow the same relative pattern.
- The ARIMAX prediction from the intervention analysis is linearly decreasing.
- Cannot test the effectiveness and accuracies of the forecasts and predictions without future data.



Future work

1. **Model Evaluation and Refinement:** Evaluate the performance of the SARIMA models and intervention analysis by comparing the forecasted values with actual values after the observed data is collected post March 2024. Then, Assess the accuracy of predictions with root mean square error (RMSE) or mean absolute error (MAE). Fine-tune the models by experimenting with different (p, d, q) (P, D, Q) orders, seasonality, and intervention points.
2. **Ensemble Methods:** Combine multiple forecasting models such as model averaging, weighted averaging, and/or stacking to potentially improve prediction accuracy by leveraging the strengths of different models (TBATS, Holt-Winters, etc.). Additional models can also capture complex patterns and dependencies in the data that SARIMA may not capture effectively.
3. **Explore Feature Engineering Techniques:** Experiment with different lag variables, rolling statistics, or domain-specific features that may enhance the predictive power of the models.
4. **Incorporate Additional External Factors:** Not only consider COVID-19 as an external factor but also look at the effects that the new car market and global supply chain had on the used car market at the onset of the COVID-19 pandemic.

Thank you

References

- https://fred.stlouisfed.org/series/MRTSSM44112USN
- https://www.linkedin.com/pulse/breaking-down-current-trends-used-car-prices-chaiz/
- TS8 Lecture Notes, Arnab Bose, M.Sc. Analytics, University of Chicago