Project 3

By Ben Bronoski

Introduction

Since the introduction of the assembly line, car production and sales have skyrocketed. Due to this, it has become critical for companies to be able to accurately forecast any selling trends. There are many things that can affect trends in car sales, such as a global pandemic. Even considering outlandish situations like this, it becomes apparent whether there are consistent fluctuations in sales over the course of many years. This project set out to attempt to forecast the sales for the Toyota Camry, Toyota Corolla, and Honda Accord based on data pulled from 2005 to 2022. The data was all collected from goodcarbadcar.net which tracks the total amount of cars sold by make and model.

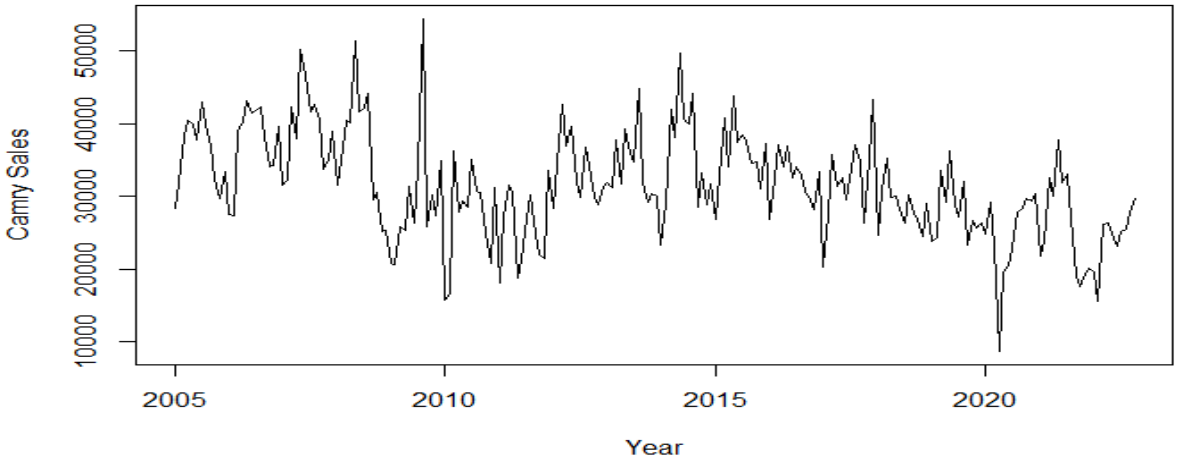
Model	**Mean Sales**	**SD of Sales**
Toyota Camry	31590	7213
Toyota Corolla	25752	6800
Honda Accord	25932	7832

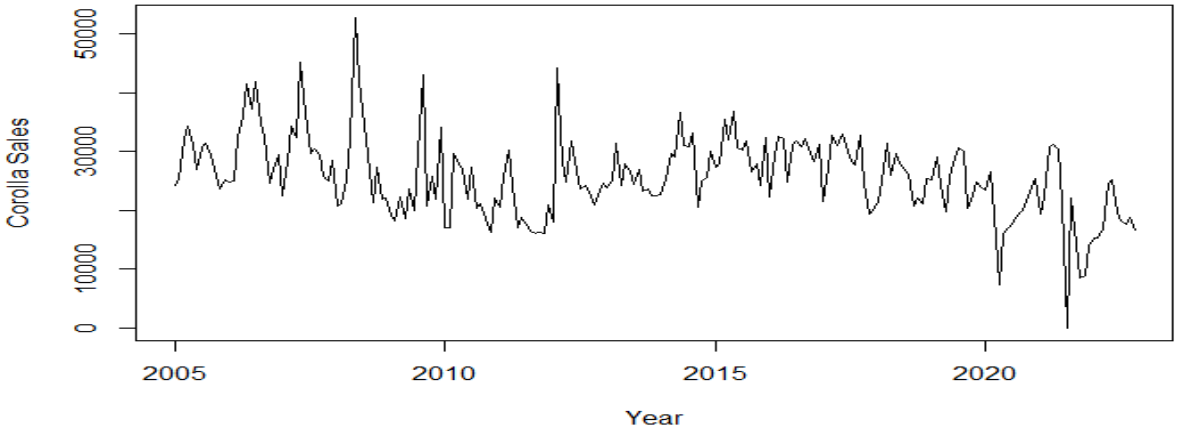
Methods

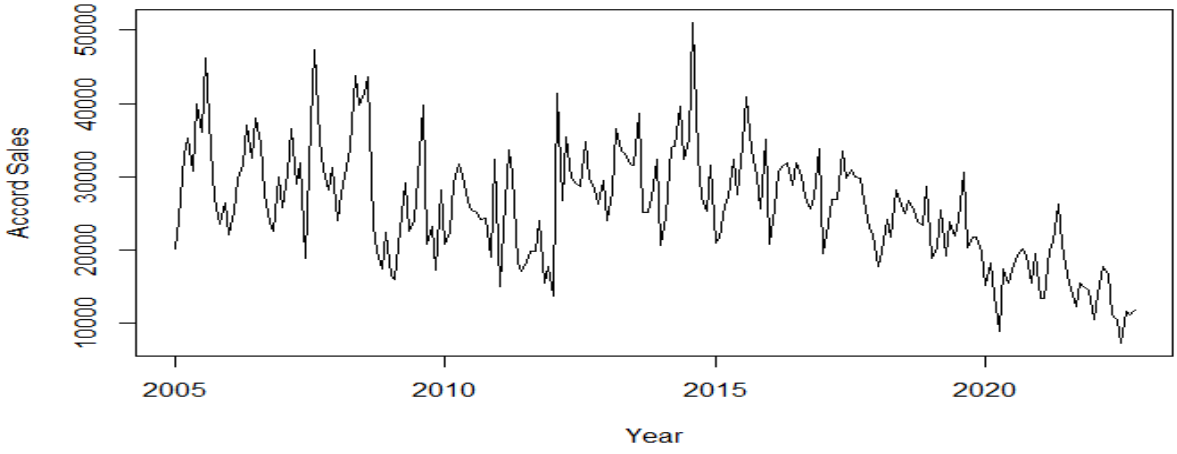
All of the statistical analysis for this research was done using R version 4.2.1. After formatting the data correctly, time series were created for each of the makes and models. These time series were split into training and test sets, with the test sets consisting of the last 6 months of data, and the training sets being all of the data from January of 2005, through to the beginning of the test sets. All makes and models were tested with different time series methods and the most accurate method was chosen to be the predictor. This method was chosen as the one that could most accurately forecast the 6 missing months from the training data compared to the actual data, or the test sets. This model selection process was repeated to forecast 1 year and 10 years into the future to depict accuracy over time for each model.

Results

All models show that there is not a reliable trend to follow over time. They do all show the effect of the pandemic on sales as there is a sizable dip from mid 2020 through the end of 2021.



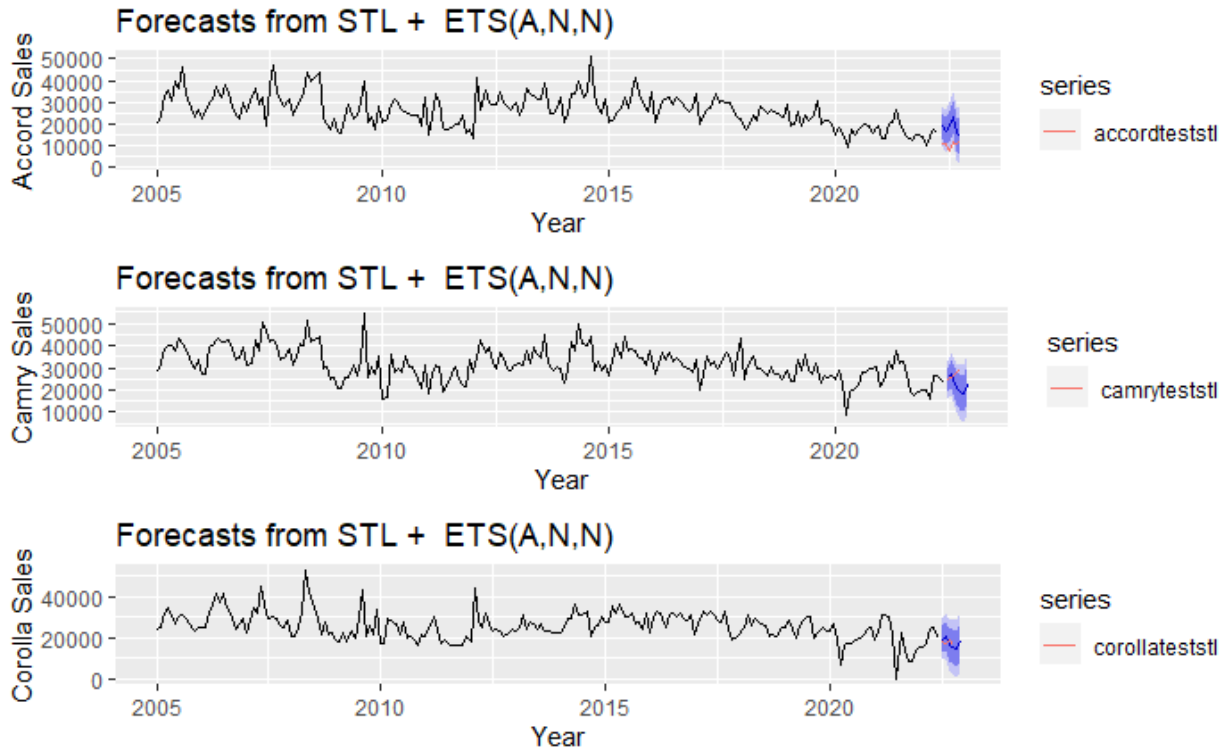




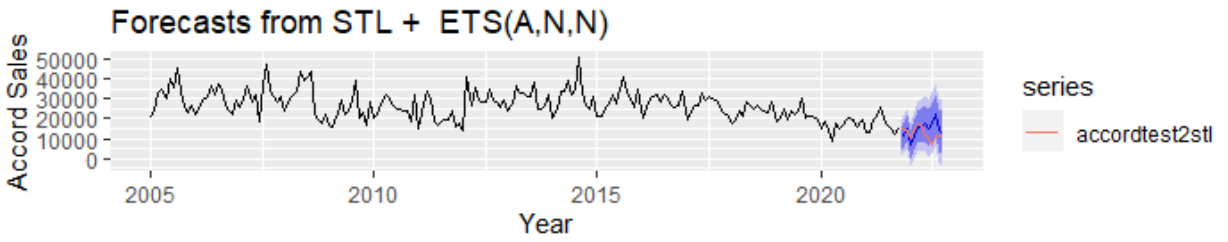
As we can see, the data seems incredibly sporadic with the only consistency being that the sales drop dramatically after the beginning of the COVID pandemic.

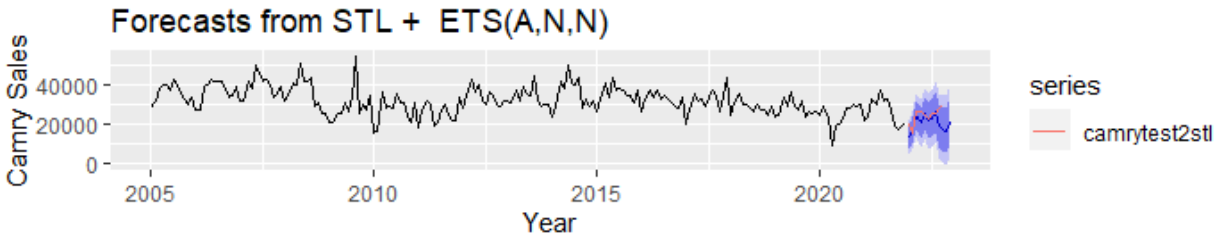
Although it may be difficult to predict, we can come within a reasonable range of what the actual numbers are.

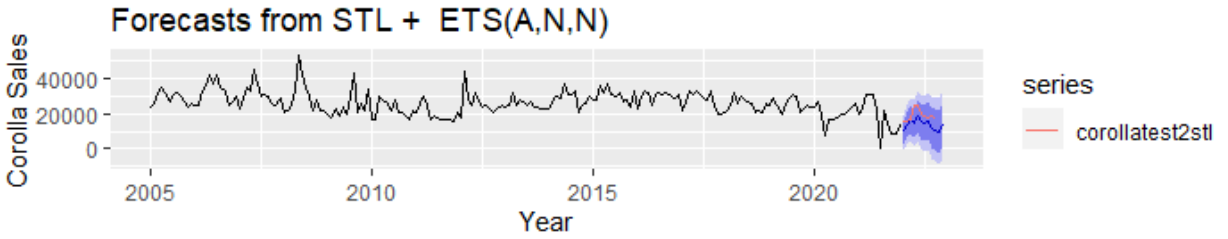
When we create a model to forecast our test data from the training data, we find it is difficult to *accurately* forecast due to the random nature of the data. An STL model seemed to fit the data the best and all of the prediction ranges include the actual data line as shown below.



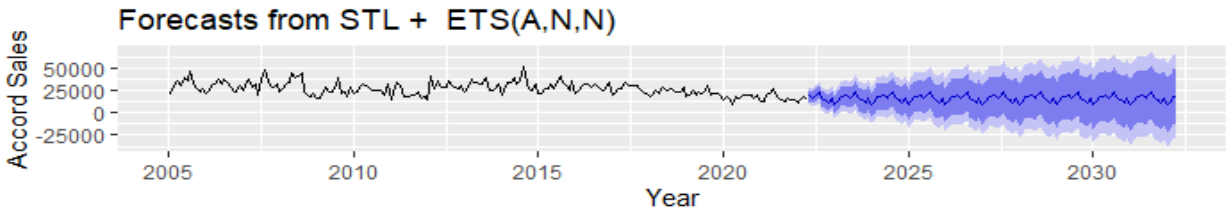
The only line that does not fall near the center of our prediction is the test data for the Honda Accord. For some reason, the sales numbers for this car have dropped dramatically in the past few years and have yet to recover from the pandemic. We can conclude, however, that it is reasonable to use this method to forecast about 6 months into the future, even with data that seems to act randomly. This brings up the question of how far into the future can we accurately forecast? A test was run to see if our model could predict one year of data.

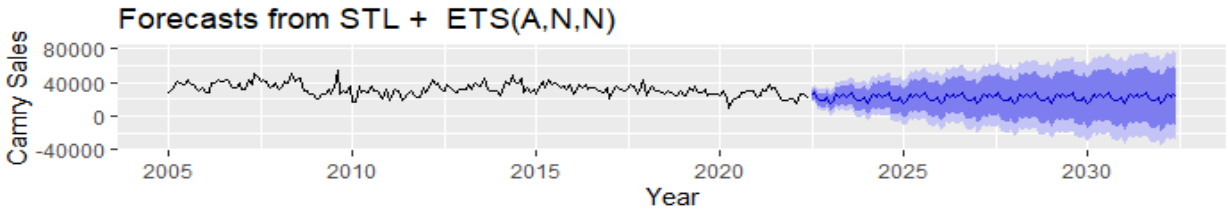


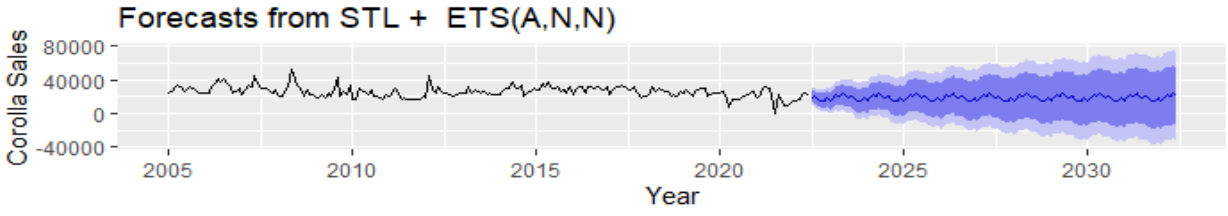




Once again, it seems like our prediction range is rather accurate. There are discrepancies in the Corolla and Accord graphs, but the Camry prediction is very close to the real data. Unfortunately, it is very hard to forecast much farther into the future than this. If a test was run to predict up to 10 years into the future, what you would find is that the range of predicted values for every make and model is over 70,000. This means that there could be as few as 10,000 cars sold, or as many as 80,000+ sold. Overall, our model can predict sales quite accurately up to one year as proven above. Once we go beyond this, the prediction interval becomes too large for comfort and does not show reliable sales numbers.







As we can see, there is a cone shape that begins to form and even begins to show negative numbers as possible. This is of course completely incorrect and should immediately signal that this forecast is unreliable.

This project demonstrates the incredible forecasting ability we have in the short term. We can capture past trends and push them as far as a couple of years into the future without too much of a penalty. Unfortunately, it is also clear that there is a limitation to this ability. The exact cutoff of the accuracy would be impossible to determine as it would be on a case-to-case basis, but in our scenario, anything over a year or two would be out of the question.