Modeling Rear-End Car Crash Fatalities

Elizabeth Luft

A Master's Project Submitted in Partial Fulfillment of the Requirements for the Degree of MASTER OF SCIENCE Department of Mathematics
ILLINOIS STATE UNIVERISTY 2014

TABLE OF CONTENTS

ABSTRACT

In this paper, an overview of crash-avoidance technology is given, as well as an introduction to what the future holds in terms of autonomous vehicles. The current features, and those that may present themselves in the coming years, could provide an issue for automobile insurance companies and their finances. Of these safety features, ones that help prevent rear-end collisions are most effective at reducing accidents so far. Data containing the number of fatalities from rear-end collisions from 2000 to 2012 was taken from the National Highway Traffic Safety Administration. They have a database called the Fatality Analysis Reporting System Encyclopedia. A model was fitted to this data using R, and it was then used to predict when the number of fatal accidents of this specific type decreased by 50%, 65%, and 80%.

INTRODUCTION

Crash-avoidance technology in automobiles is becoming more and more prevalent as technology improves. Some features available in many vehicles today include a back-up camera, autonomous braking system, lane departure systems, adaptive headlights, and even parking assist. Some companies, such as Google, have even begun work on a completely driverless car. As more and more vehicles with these features begin to inhabit the roads, people are hopeful that technology will help dramatically reduce some of the accidents that are caused solely by human error. This poses a big question for insurance companies: If these safety features begin to drastically reduce the number of accidents, how will customers react? Surely they will feel like they don't need the insurance coverage if the technology is going to make sure that they don't get in the accident in the first place! For companies that draw large percentages of premium from policies for automobile insurance, this could ultimately become a financial problem.

In an article posted by the Highway Loss Data Institute in July of 2012, it is shown that the only safety features that have thus far showed significant benefits are the autonomous braking systems and adaptive headlights. They found that the autonomous braking systems caused a reduction in property damage liability (PDL) claim frequency, implying that this technology is in fact doing what it set out to do. In fact, the article states, "PDL frequencies for Acura and Mercedes models were 14 percent lower when the vehicles were equipped with forward collision warning with autonomous braking than when they weren't." (HLDI News, 2012). Although the adaptive headlights also showed benefits, they were not as great as the ones found by the autonomous braking systems.

There are many different aspects to consider when an insurance company is trying to figure out what to do in the future. It is hard to predict the rate at which technology will improve. Therefore, it is hard to predict what kind of safety features will be in cars ten years down the road. Another thing to consider is the rate at which cars without these safety features are leaving the road, and the rate at which new cars equipped with this technology are entering the road. This technology is still fairly new. In an article written by David Zuby, he provides a chart showing the percentage of new vehicle series with forward collision warning by model year, and another chart showing the actual percentage of registered vehicles with forward collision warning by calendar year. It is amazing because in 2012, more than 20% of new vehicles were equipped with this technology, but less than 5% were actually registered. This shows that the introduction of new cars onto the road is happening at a very slow pace.

Another factor to take into consideration is the legal aspect. There are multiple scenarios to look at in this situation. One is that the government comes out with new mandates and safety laws, and the introduction of such items is completely unpredictable. The other aspect has to do with the liability of crashes that happen even with the safety features intact. An article was published in the Santa Clara Law Review by Gary Marchant and Rachel Lindor. In it, they address the fact that technology is not perfected. Situations will still arise in which the cars will still crash. The question then begins to fall on who is liable for the crash if the vehicle is equipped with a specific feature that is guaranteed to prevent it. In this article, they state that the liability requires an analysis of the following three factors: who is liable, what weight will the court's finder of fact give to the overall comparative safety of autonomous vehicles when determining whether those involved in a crash should be held liable, and will a vehicle "defect" that creates potential manufacturer liability be found in a higher percentage of crashes than with conventional vehicle crashes where driver error is usually attributed to the cause. (Lindor, Marchant 2012). This is a very important issue that also needs to be considered when looking at the impact of autonomous vehicles.

One paper written by Tyler Guzzino as part of a project with The Warranty Group does a wonderful job of suggesting the idea of a combined insurance product that would help account for some of the arising issues of liability and warranty. They suggest the creation of a possible link between insurance and warranty groups. They make the point that, "…warranty is separate from insurance…", (Guzzino, 2014), and state that the lack of strict warranty regulation would be solved by the proposed link. Since insurance is much more highly regulated, warranty would also need to be regulated in such a strict manner in order for the link to be successful. Overall, they state that this type of combined product would have a rough start, but that it is a possible option.

A society which is filled with perfectly working completely driverless cars would seem far-fetched and definitely not possible for many years, but this does not mean that insurance companies don't need to worry about the implications of autonomous vehicles. Although vehicles are not completely driverless, technology improvements have already shown to be impacting claim frequency in accidents. Talk of government mandates for certain safety features in vehicles would greatly increase the number of vehicles on the road with safety features intact. It is only a matter of time before people begin lowering the coverage with the guaranteed safety provided by the technology.

For this paper, it was important to try to find a time frame for when the number of accidents drastically decreases from where they are at now. This would provide a nice time frame for insurance companies to figure out how they may handle the situation when a large percentage of their money is not coming in like it used to.

DATA SELECTION

Data was selected for this project based on the results from the Highway Loss Data Institute that state that autonomous braking systems and adaptive headlights have been the most successful at decreasing the number of accidents that they were created to deter. The autonomous braking system is most beneficial at reducing rear end collisions, and thus data was pulled only concerning those car crashes that the autonomous braking systems help to avoid.

Another factor in the selection of this particular data was that the National Highway Traffic Safety Administration only had record of fatalities from car crashes. Therefore, accidents that did not result in a fatality were not considered for this project. Below is the graph of the number of crashes that resulted in fatalities caused by rear end collisions.

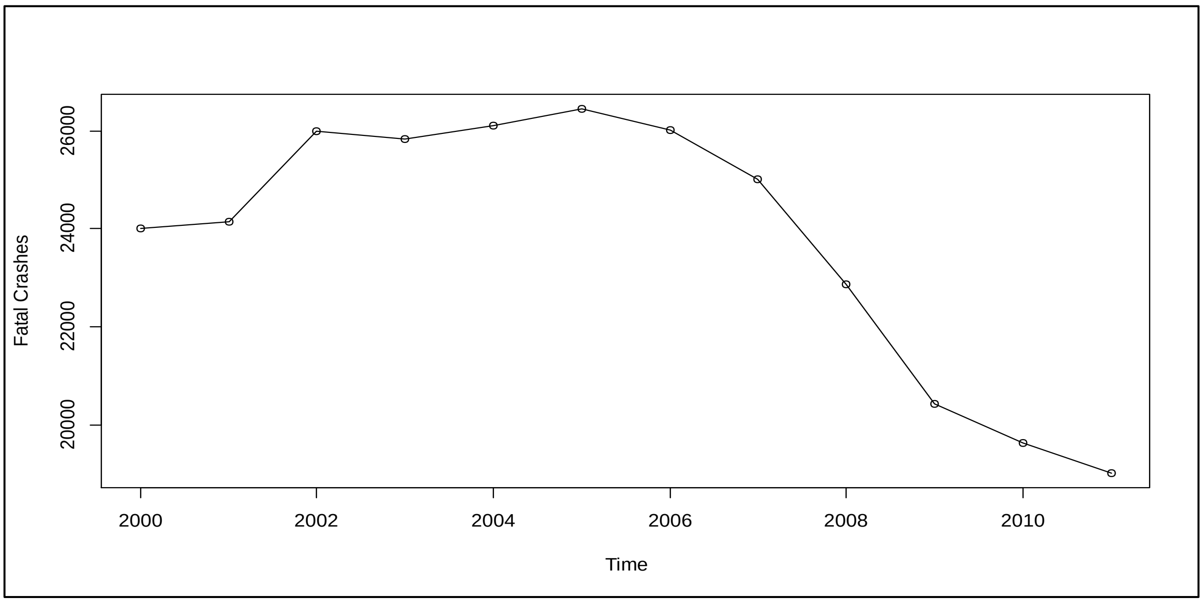

Figure 1: Number of fatal crashes from 2000 to 2012

The data ranges from 2000 to 2012. Data for 2013 was unavailable. Based on the lack of visual trend for the yearly data, it was decided that we used the monthly data from January of 2000 to December of 2012 for the full data set, pictured below in Figure 2. This provided more data points and also a more visual trend. During the model selection process, the last seven points were excluded during the fitting process in order to allow for comparison during the forecasting process.

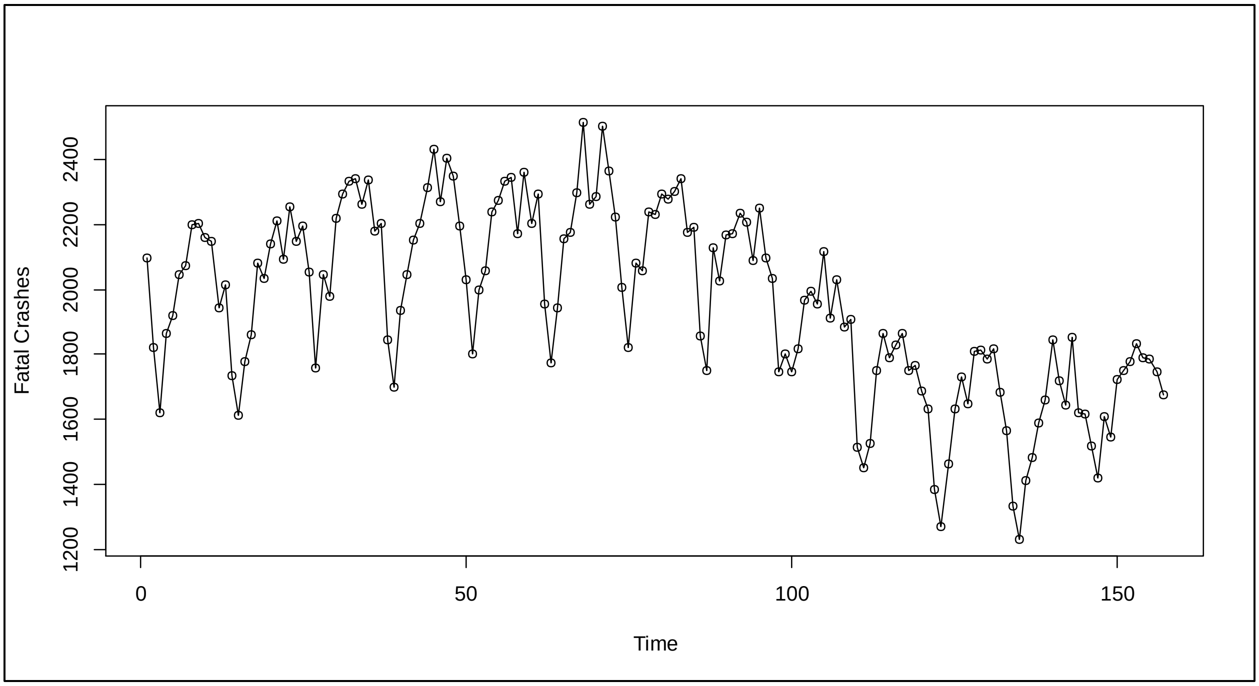

Figure 2: Monthly number of fatal crashes from January 2000 to December 2012

Based on Figure 2, the monthly data has a clear seasonal trend. Using time series analysis and R, a model was able to be fitted to this data.

ANALYSIS TOOLS

All data analysis and model fitting was done in the program, R. The time series package was used in order to create and analyze the time series data. Models considered were seasonal ARIMA models in the form of ARIMA(p,d,q)x(P,D,Q), where p = the number of auto-regressive terms, d = the number of nonseasonal differences needed for stationarity, and q = the number of lagged forecast errors in the prediction equation. For the seasonal part of the ARIMA model, P = the number off seasonal autogregressive terms, D = the number of seasonal differences, and Q = the number of seasonal moving average terms. There are three assumptions that need to be taken into consideration in order to use an ARIMA model. The residuals need to be normally distributed and independent, and the data needs to be stationary. In order for data to be stationary, the mean, variance, and autocorrelation are constant over time. If your original data is not stationary, you can perform mathematical transformations, such as the first difference, to help correct the issues.

Many tests were used during model selection, the first of which is the augmented Dickey-Fuller test. This test takes a given time series $Y_1, Y_2, \ldots, Y_N$ and uses autoregressive equations in order to detect the presence of a unit root. The presence of a unit root would indicate that the data is not stationary. Three differential-form autoregressive equations are used to detect a unit root.

$$(1) \qquad \Delta Y_t = \gamma Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t$$

$$(2) \qquad \Delta Y_t = \alpha + \gamma Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t$$

$$(3) \qquad \Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{j=1}^{p} \left(\delta_j \Delta Y_{t-j} \right) + e_t$$

In these equations, t is the time index, $\propto$ is the drift, β is the coefficient on a time trend, γ is the coefficient presenting process root, p is the lag order of the first-differences autoregressive process, and e_t is an independent identically distributed residual term. The test focuses on testing to see whether $\gamma = 0$ or not. If $\gamma = 0$, then the process is a random walk. The three hypotheses for the three tests are listed below.

$$(h1) \qquad \begin{array}{l} H_0 : Y_t \textit{ is random walk } \; OR \quad \gamma = 0 \\ H_1 : Y_t \textit{ is stationary process } \; OR \quad \gamma < 0 \end{array}$$

$$(h2) \qquad \begin{array}{l} H_0 : Y_t \textit{ is random walk around a drift } \; OR \; \{ \gamma = 0, \alpha \neq 0 \} \\ H_1 : Y_t \textit{ is level stationary process } \; OR \; \{ \gamma < 0, \alpha \neq 0 \} \end{array}$$

$$(h3) \qquad \begin{array}{l} H_0 : Y_t \textit{ is random walk around a trend } \; OR \; \{ \gamma = 0, \beta \neq 0 \} \\ H_1 : Y_t \textit{ is trend stationary process } \; OR \; \{ \gamma < 0, \; \beta \neq 0 \} \end{array}$$

In R, the Dickey-Fuller tests for constant, non-constant, and linear trends gives us a p-value. If that p-value is significant (0.01 or below), then we can reject the null hypothesis. If the p-value is greater than 0.01, then we cannot reject the null hypothesis, and therefore we conclude that our data is not stationary.

Once we have been able to conclude that our data is stationary, it is important to complete residual analysis. In this paper, the normality of the residuals was tested using a QQ Plot and the Shapiro-Wilk test. The Quantile-Quantile Plot, or Q-Q Plot plots the theoretical expected values for each point and plots them on a graph. If these points lie in a straight line, then the data can be said to be normally distributed. The Shapiro-Wilk test is also a test for normality. For this test, a statistic W is calculated in order to test whether a random sample comes from a normal distribution. W can be calculated using the formula below.

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

In this formula, $x_{(i)}$ are the ordered sample values, a_i are constants generated from the means, variances, and covariances of the order statistics of a sample size n from a normal distribution. In R, a p-value is calculated for this test. If the p-value is significant, we can reject the null hypothesis that states that the data comes from a normal distribution. If the p-value is large, then we cannot reject the null hypothesis, and we assume that our data is from a normal distribution. For this paper, we are testing the normality of the residuals, not the data itself.

The third thing that needs to be tested is whether or not the residuals are independent. For this paper, the p-value plot and the runs test were used in order to assume independence. The p-value plot shows whether the calculated p-values are significant by also plotting a significance line. If any of the points lie above the line, then we can reject the null hypothesis the states that the residuals are independent. In order to assume independent residuals, all points must be below the line. The p-values are calculated up to a certain number of lags which is determined by the researcher. The values themselves are calculated for a Portmanteau goodness-of-fit test.

The runs test is a more quantitative test that can give us a p-value in order to reject or not reject our null hypothesis, which is that our residuals come from a random process. In the runs test, a run is defined as a series of consecutive positive or negative numbers. The null hypothesis for the runs test states that the sequence was produced randomly, and therefore the alternative hypothesis states that it was not produced randomly. The test statistic for the runs test is given below.

$$Z = \frac{R - \bar{R}}{s_R}$$

In this test statistic, R = number of observed runs, $\bar{R}$ = the expected number of runs, and s_R = the standard deviation of the number of runs. The expected number of runs and the standard deviation of the number of runs can be found by using the following formulas.

$$\bar{R} = \frac{2n_1 n_2}{n_1 + n_2}$$

$$s_R^2 = \frac{2n_1 n_2(2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Where n_1 and n_2 denote the number of positive and negative values in the series. The runs test in R gives us a p-value. If this value is significant, we can reject our null hypothesis that our residuals come from a random process and are therefore independent. If our p-value is larger than 0.01, then we can assume independence for our residuals.

Another standard that was used was the Aikake Information Criterion value. This value is given by the formula, $-2L_m + 2m$. In this formula, L_m is the maximized log-likelihood and m is the number of parameters in the model. Lower values of the AIC indicate the preferred models.

After fitting a model and getting the estimates, the statistical significance for the estimate is calculated by dividing the estimate by the standard error of the estimate. An estimate is considered significant if this absolute value is greater than 1.

After an adequate model was found, R was used to detect any outliers, and then the model was tested on its forecasting ability. The model was fitted leaving seven of the last data points out, and these were used to compare how good the model was at forecasting. To test how well the model performed, I ran it on the first three data points, and all seven. After both of these sessions, the Mean Absolute Deviation was calculated in order to find the average distance between each predicted data value and the mean. It is preferred that this number be very small in order to conclude good performance. Another method of testing the model's forecasting ability was to calculate the mean square error (MSE) of the predicted values versus the actual values. We would also like this value to be as small as possible when comparing models.

MODEL SELECTION

The modeling process began by reading the data into R, and making sure it was in time series format. In order to choose a model, three assumptions about the data must be made. The data must be stationary, and the residuals need to be both normally distributed and independent. Multiple tests for stationarity were then performed on the data series. Results yielded that the Augmented Dickey-Fuller test had a p-value of 0.01. This was significant enough to conclude from this test that the data was stationary. However, it is best to also test for stationarity with a non-constant, linear, and constant trend in your data. The p-values for these augmented Dickey-Fuller tests were 0.4497, 0.08079, and 0.01 respectively, indicating that the original data set was not stationary, and we therefore needed to correct this problem in order for us to move on with the model selection process.

The first difference of the time series data was taken, and once again tested for stationarity. All four versions of the augmented Dickey-Fuller test produced p-values less than 0.01 which allowed us to reject the null hypothesis and conclude the first difference of the time series data to be stationary. Since the data was visually seasonal, the seasonal difference was also taken, and the data remained stationary.

Many different models were fit to the first and seasonal difference of the original time series data. Also, multiple differences of seasonality were also considered in attempt to produce models that yielded better results. Tables for each model considered are shown in the next section and include the estimates given by the models, the standard error of the estimates, and the AIC value. Residual analysis information is given after each table.

MODEL 1			**AIC**
ARIMA (1,1,0)X(1,0,0)$_{12}$	Ar1	Sar1	1798.11
Estimates	-0.4291	0.7986	
Standard Error	0.0761	0.0456	
Significance Level	-5.639	17.513	

Model 1 provided significant estimates, and therefore needed to be tested for normality and independence of the residuals. To test for normality of residuals, two methods were used. The visual Q-Q Plot showed that the residuals were pretty close to normal, but the Shapiro-Wilk test helps give us a better idea with a p-value of 0.7566. This was not a significant enough value to reject the null hypothesis, and therefore it was concluded that the residuals were normal. To test for independence, the runs test was used. This produced a p-value of 0.461, which was not significant enough to reject the null hypothesis, and therefore the residuals were treated as independent based on this test. Another procedure was to look at the p-value plots, and since all of the p-values were below the significant line, this confirmed the runs test and the residuals were concluded to be independent.

Since all assumptions held, the over-fitting process for this model was necessary. The first over-fitting model is shown below.

MODEL 1.1				**AIC**
ARIMA (1,1,0)X(2,0,0)$_{12}$	Ar1	Sar1	Sar2	1776.18
Estimates	-0.4573	0.4678	0.4113	
Standard Error	0.0742	0.0749	0.0783	
Significance Level	-7.722	6.246	5.253	

All estimates remained the same, so a residual analysis was done on Model 1.1. What the Shapiro-Wilk test p-value of 0.1882 and the Q-Q Plot showing fairly normal residuals, it was concluded that we could not reject the null hypothesis which states that the residuals are from a normal distribution. Also, from the p-value plots and the runs test p-value of 0.579, we were not able to reject the null hypothesis which states that the residuals from this model are independent.

Since both Model 1 and 1.1 produced significant estimates and normal and independent residuals, I chose to move forward from Model 1.1 since it had the lower Akaike Information Criterion value. From here, I needed to perform over-fitting on Model 1.1, which is illustrated in Model 1.2.

MODEL 1.2					**AIC**
ARIMA (2,1,0)X(2,0,0)$_{12}$	Ar1	Ar2	Sar1	Sar2	1769.09
Estimates	-0.5839	-0.2523	0.4805	0.4251	
Standard Error	0.0823	0.0817	0.0753	0.0786	
Significance Level	-7.095	-3.088	6.381	5.408	

All estimates remained significant, and a residual analysis was done on Model 1.2. Both the Shapiro-Wilk test (p-value 0.1819), and the Q-Q Plot indicated that our residuals were normally distributed. Also, the p-value plot as well as the runs test (0.888) indicated that our residuals for Model 1.2 were independent. Also, the AIC value for Model 1.2 was slightly lower than the AIC value for Model 1.1, so the over-fitting process was continued, shown in Model 1.3.

MODEL 1.3						**AIC**
ARIMA (2,1,0)X(3,0,0)$_{12}$	Ar1	Ar2	Sar1	Sar2	Sar3	1766.92
Estimates	-0.5983	-0.2805	0.3949	0.3411	0.1892	
Standard Error	0.0185	0.0822	0.0862	0.0889	0.0913	
Significance Level	-32.341	-3.412	4.581	3.837	2.077	

As the table suggests, all estimates remained significant in Model 1.3. The AIC value was also the smallest one out of all of the models thus far. In addition, residual analysis concluded that the residuals were normally distributed and independent. The Q-Q Plot looked good, and the Shapiro-Wilk test gave a p-value of 0.1725. The p-value plot looked good, and the runs test gave a p-value of 0.888. This model was kept for over-fitting, but the addition of any estimates made one or more of the estimates from Model 1.3 insignificant, and therefore Model 1.3 was kept as one of the final models to compare. A search for outliers was conducted on Model 1.3, but none were found.

The next model considered involved the first difference of the seasonal aspect. This is shown as Model 2.

MODEL 2						**AIC**
ARIMA (2,1,0)X(3,1,0)$_{12}$	Ar1	Ar2	Sar1	Sar2	Sar3	1599.58
Estimates	-0.6046	-0.3091	-0.6822	-0.4496	-0.3530	
Standard Error	0.0827	0.0833	0.0878	0.1017	0.0882	
Significance Level	-7.311	-3.711	-7.770	-4.421	-4.002	

Model 2 produced significant estimates, and also and AIC value lower than any one that was observed up to this point in the process. Although it had the lowest AIC value, the Q-Q Plot and Shapiro-Wilk test (p-value 0.02249) did not provide results as expected. Similarly, the p-value plot and the runs test (p-value 0.0315) had worse results as well. Although both of these results were not as strong as the previous models, I decided to try over-fitting this model because it had such a low AIC value, and although our p-values were small, they were not significant enough to reject the null hypotheses.

MODEL 2.1							**AIC**
ARIMA (2,1,0)X(4,0,0)$_{12}$	Ar1	Ar2	Sar1	Sar2	Sar3	Sar4	1597.69
Estimates	-0.5666	-0.2986	-0.7583	-0.5460	-0.4994	-0.199	
Standard Error	0.0845	0.0833	0.0936	0.1128	0.1111	0.0983	
Significance Level	-6.705	-3.585	-8.101	-4.84	-4.495	-2.024	

The over-fitting of Model 2 to Model 2.1 resulted in significant estimates, and therefore residual analysis was completed for this model. The Q-Q Plot showed that the residuals were very close to normal, with the tail ends deviating just a bit. The Shapiro-Wilk test provided a p-value of 0.1755, which was high enough to not reject the null hypothesis. The p-value plot had all points below the significance line, and the runs test provided a p-value of 0.124, indicating that we could not reject the statement that the residuals were independent. Since both independence and normality held, outliers were then searched for, but none were found. This model was kept as one for final analysis.

MODEL FORECASTING

Models 1.3 and 2.1 were both used to forecast. The final model was chosen based on which one had the better forecasting results. Since the model was fitted to only the first one hundred fifty data points, this left seven points of the original data that we could use to test how well each model forecasted. Each model was tested to see how well it forecasted three months ahead as well as seven months ahead.

Model 1.3 produced a calculated MAD value of 0.0223363, meaning that the predicted values for three months ahead were very good. The MSE for three months ahead for Model 1.3 was 661.5077, which is also a good value. When predicting seven months ahead, the MAD value did not change too much. It was 0.02712555. However, the MSE value changed drastically to 25593. This tells us that Model 1.3 is better at predicting only a few months into the future.

Model 2.1 was used to forecast the same amounts. For three months in the future, the MAD value was 0.008507709, and the MSE value was 9.845261. This told us that Model 2.1 was very good at predicting three months into the future. In Figure 4, you can clearly see that the predicted values are very close to the actual values in the first three months. When we tested its ability to predict seven months into the future, the MAD value was 0.02355451, and the MSE value was 24735.31. We can see in figure four that the predicted values deviate much more from months four to seven than they did for the first three months. Therefore, it performed about the same as Model 1.3 for seven months ahead, but much better for predicting three months into the future. Therefore, I chose Model 2.1 as my final model to use.

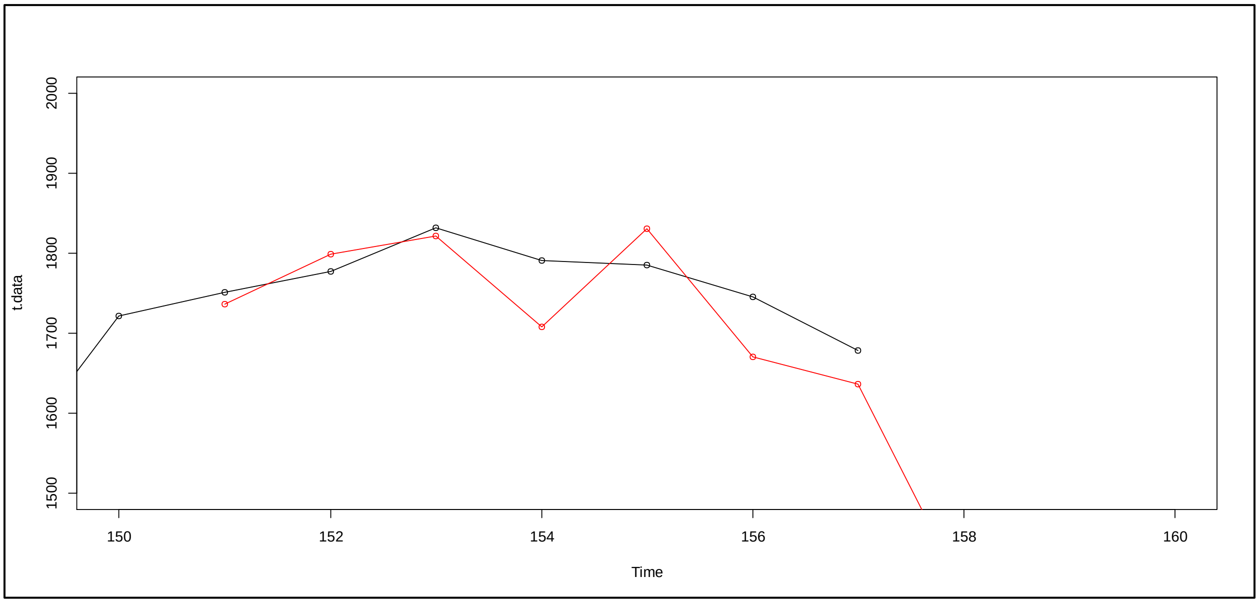

Figure 4: Predicted values for Model 2.1 in red versus actual values in black.

I used Model 2.1 to predict 300 months into the future. These predicted values are in red in Figure 5 below. Looking at this graph, we can give an approximate time frame for when fatal rear end crashes will be decreased by a certain percentage.

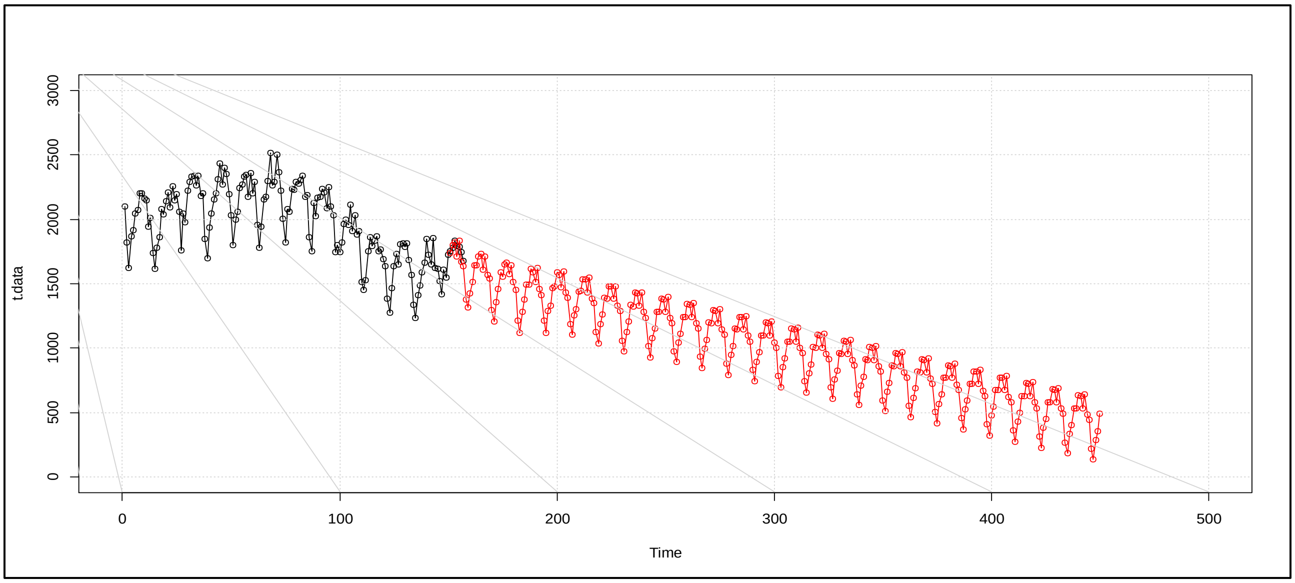

Figure 5: Original data in black, predicted points in red. Number of fatal rear end crashes on a monthly basis beginning January 2012.

The model predicts a clear decrease in the number of fatal crashes caused by rear end collisions in the coming years. This indicates that insurance companies should definitely be looking in the near future to see when this may become a concern to them.

To find an approximate time range, I took the average monthly accidents for 2012. This was 1681.417. I then wanted to know how long, approximately, it would take for the number of accidents to decrease by 50%. This would mean that we would have about 840 accidents. I also wanted to consider a decrease of 65% and also 80%. These numbers were approximately 590 and 340 respectively. Below is the graph with these lines plotted on them.

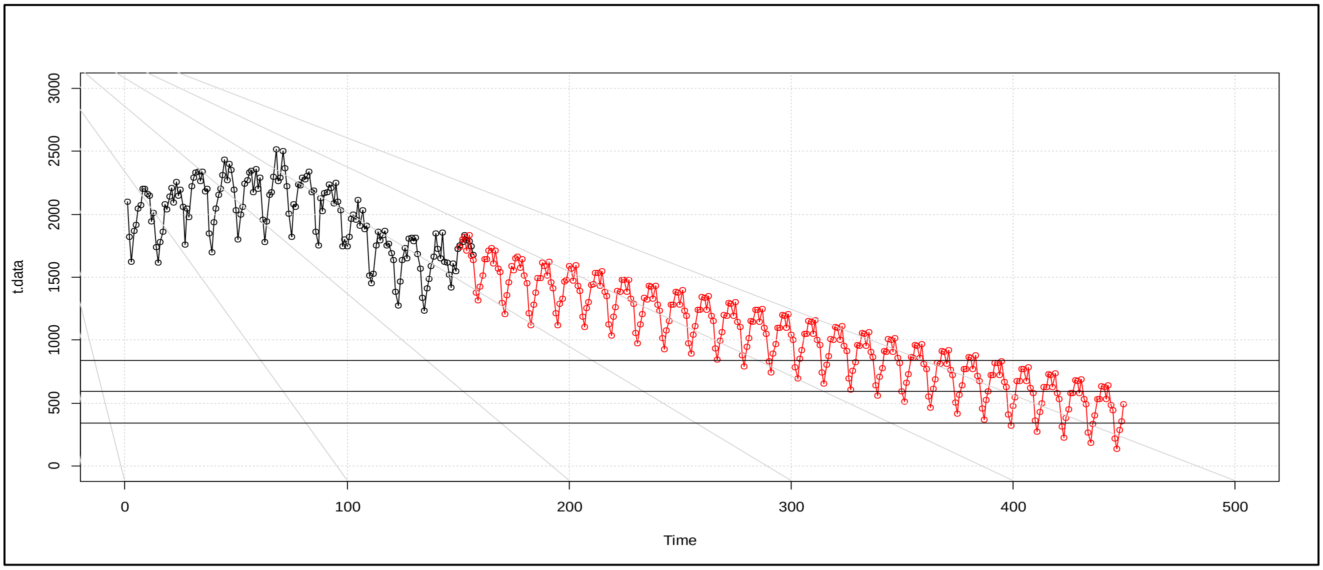

Figure 6: Predicted values in red, actual values in black. Lines at 840, 590, and 340 representing 50%, 65%, an 20% decreases.

From the printed predictions, we hit our first month of the number of accidents being about 840 in the beginning of 2022. The first time we hit the number of accidents being 590 is in 2027, and the first time that we hit the number of accidents being about 340 is the beginning of 2036. If this model is correct, this means that we are looking at a very significant decrease in the number of fatal accidents per year caused by rear end collisions in the next twenty-five years.

CONCLUSIONS

Model 2.1 ended up being the best model that I could find for the given data. I was able to make the data stationary, find a model with significant estimates and normally distributed independent residuals. The model also performed wonderfully predicting three months into the future. It also showed a trend in the future of the number of accidents declining, which supports the idea that as cars become equipped with safety features, the number of accidents will decrease. Based on this model, it was concluded that insurance companies should at least begin to look at the effects autonomous vehicles may have on the industry.

A set of data that would have been helpful would have been what percentage of the accidents was caused by a car equipped with autonomous braking. This process could be furthered by also considering accidents that could have been prevented using adaptive headlights. Once the data for 2013 comes out, it would also be a good idea to tweak the model to make sure it still works for last year. Technology changes so quickly that one year could make a big difference in the model.

Since it has been concluded that insurance agencies could definitely see some financial differences with claim rates decreasing and possible insurance policies lowering, I think it is important that they begin considering how they are going to handle the technological steps being implemented towards creating an autonomous vehicle. They should begin looking at possible product changes, other venues in which they might be able to draw more income. Before they panic too much, though, it is important that they check to see whether the decrease in claims will make up the difference if premiums go down. I do not believe that there will be a direct correlation, but it is another avenue which should be addressed. The automobile world is changing, and that includes not only the vehicles, but the laws regulating the vehicles, the warranty, the liability, and most importantly for this project, the insurance industry.

REFERENCES

Engineering Statistics Handbook. (2014). Runs Test for Detecting Non-randomness. 20 November 2014. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35d.htm .

Guzzino, Tyler. (2014). Autonomous Vehicles: Legal Implications and Impacts on the Insurance And Warranty Industries.

Lindor, Rachel and Marchant, Gary. (2012). The Coming Collision Between Autonomous Vehicles and the Liability System. *Santa Clara Law Review.*

Zuby, David. (2014). The Promise and Reality of Crash Avoidance Technology: Driverless Cars, Cars that "Talk" to Each Other, Forward Collision and More. *Presentation for Insurance Institute for Highway Safety Highway Loss Data Institute.*

National Highway Traffic Safety Administration. (2014). Fatality Analysis Reporting System Encyclopedia. 27 November 2014. http://www-fars.nhtsa.dot.gov/QueryTool/QuerySection/Report.aspx.

Nau, Robert. (2014). General Seasonal Arima Models. 27 November 2014. http://people.duke.edu/~rnau/seasarim.htm .

RTMath. (2014). Augmented Dickey-Fuller (ADF) Test. 27 November 2014. https://www.rtmath.net/help/html/93a7b7b9-e3c3-4f19-8a57-49c3938d607d.htm .

Zuby, David. (2014). The Promise and Reality of Crash Avoidance Technology: Driverless Cars, Cars that "Talk" to Each Other, Forward Collision and More. *Presentation for Insurance Institute for Highway Safety Highway Loss Data Institute.*