Effects of Vision Zero on Pedestrian Fatality Rates: A Time Series Analysis on Pedestrian Crash Rates in New York

Sherien Hassan

May 2020

Introduction

Auto Accidents are the 3rd leading cause of fatal injuries and deaths among New York City Residents as stated by the *Center for Disease Control and* Prevention. Despite the many measures put out by the Mayor Bill De Blasio to ensure that drivers on the road are considerate of the safety of those surrounding them, death rates continue to rise. One specific measure that Mayor De Blasio set forth was *Vision* Zero, a citywide effort to eliminate all traffic related deaths and injuries, executed in 2014. The program assess components such as engineering, education, enforcement and legislation in order to revise the city's current state of pedestrian safety. For instance, their engineering efforts consisted of creating better street designs that cater to how pedestrians may cross in a specific intersection. Although 2014 saw the least pedestrian fatalities since 1910, where there was an all-time low of 132 deaths, every preceding year after the implementation of the program saw an increase in death rates by 11% while injuries increased by 18%. Even though these collisions are defined as accident's, they are far from random given they exhibit trends and patterns which this paper will further analyze.

Hypothesis

Due to the normalization of legislation put forth by vision zero, pedestrian fatality rates have increased and will continue to increase.

Data Description

This data was extracted from NYCOpen Data, specifically, the Pedestrian Crash Data sector of the Motor Vehicle Collison dataset. This consisted of only cases where an MV104-AN form was filled out or there was $1000 worth of damage to the vehicle. For this paper in specific, daily data between the years 2014 through 2018 (a total of 1826 points) will be used to analyze the effects of Vision Zero. The data points are indexed in time order based on both death and injuries due to vehicle accidents in New York City using statistical software, R.

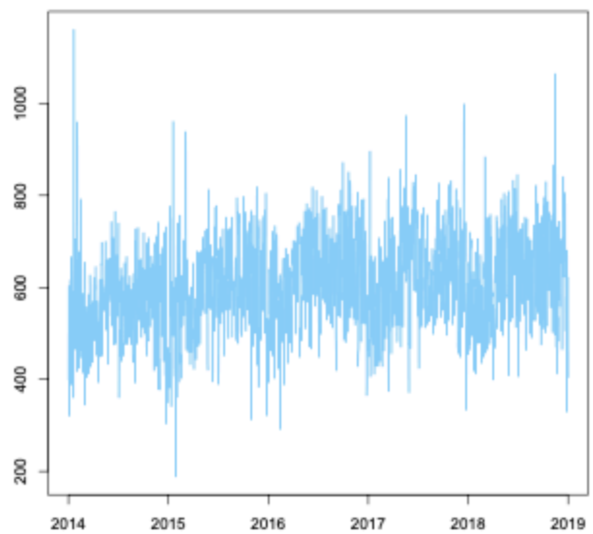

Figure 1: Pedestrian fatality over time

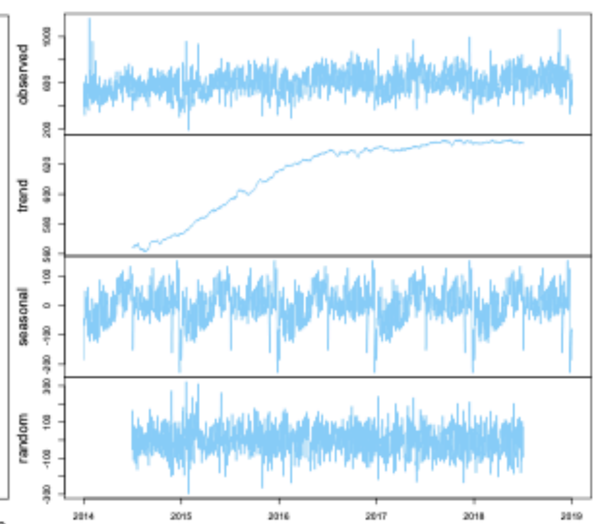

Figure 2: Pedestrian fatality decomposed into seasonal, trend, and random plots

As can be seen from *Figure 1*, there is stationarity which was confirmed using the Augmented Dicky Fuller Test having a p-value of .01, which is less than 5 %. Hence leading us to reject the null hypothesis that there is a unit root. Some volatility clustering's can be observed due to each crash event being independent of one another. For *Figure 2*, there is an upward trend as well as annual seasonality which is the first sign to use the modeling process, seasonal Arima.

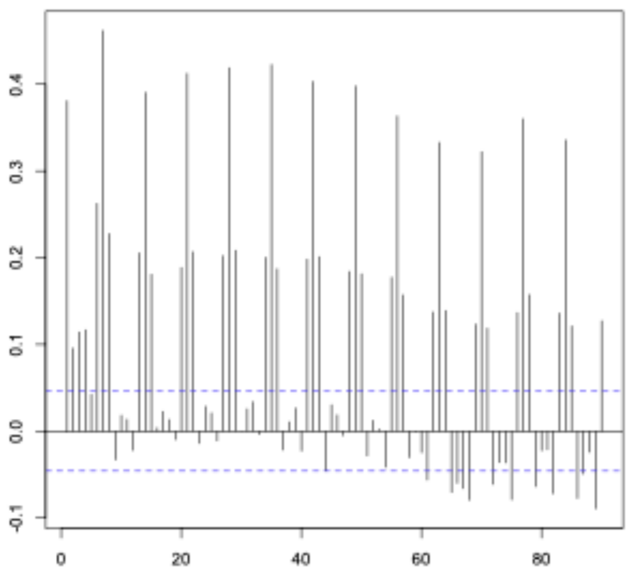

Figure 3: Autocorrelation Function

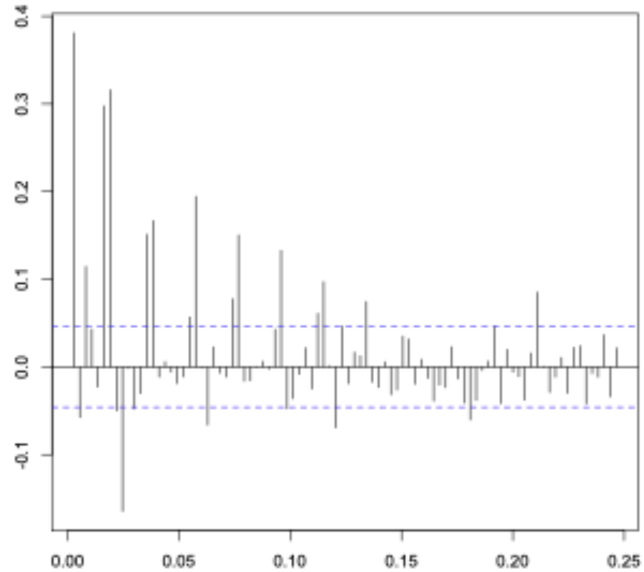

Figure 4: Partial Autocorrelation Function

Figure 3 depicts the original time series' autocorrelation function which exhibits the same seasonality shown in *figure 2.* The autocorrelation function is concentrated around positive lags fluctuating between 0.3 and 0.4 with three lag correlation. Due to the autocorrelation function being out of the confidence interval, differencing, square rooting as well as logging the original time series was done in effort to adjust seasonality, with no model improvement. While our partial autocorrelation function depicts a gradual decline into the confidence interval in which too could be improved with a seasonal ARIMA.

Methods

I. Model Specification

The modeling process chosen was a seasonal ARIMA in order to forecast the next 31 days of auto vehicle incidents. A seasonal ARIMA is formed by including additional seasonal terms to a ARIMA model, written: $ARIMA(p, d, q)(P, D, Q)_m$. The elements of the non-seasonal ARIMA are p: trend autoregression order, d: trend difference order, q: trend moving average order while the elements for the seasonal ARIMA are P: seasonal autoregressive order, D: seasonal difference order, Q: seasonal moving average order and lastly, m: the number of steps for a single seasonal period. The reason for this model choice is due to the annual seasonality in which is not just present in our original time series plot but also in its autocorrelation function plot.

II. Model Fitting

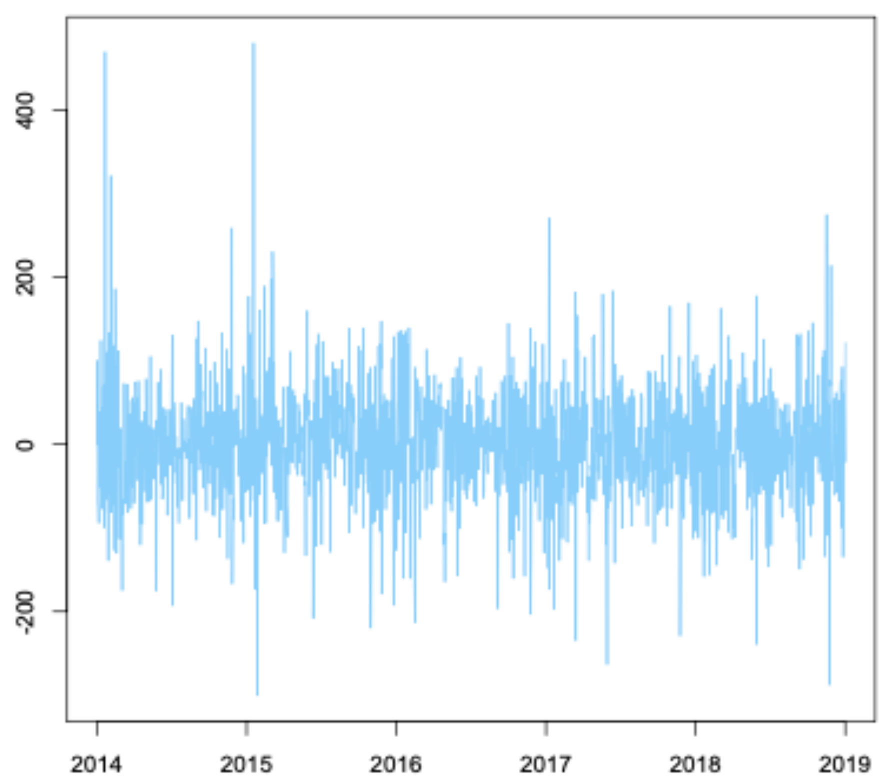

Figure 5: $SARIMA(5,1,2)(1,0,1)_7$

The parameters picked for this seasonal ARIMA was $SARIMA(5,1,2)(1,0,1)_7$ written as $$(1-\phi_1 L - ... - \phi_5 L^5)(1-\Phi_7 L^7)(1-L)y_t = c + (1+\theta_1 L + \theta_2 L^2)(1+\Theta_7 L^7)\varepsilon_t$$
While varying model parameters, this seasonal ARIMA had the lowest AIC value. Also, the sigma squared values as well as the residuals were considered when picking this model. Note, the value for frequency was picked to be 7 due to R's frequency limitation of 350 periods. The usual frequency used with daily data would be 7 since it attributes weekly seasonality but since this time series has annual seasonality, the ideal frequency would have been 365.

III. Model Fitting

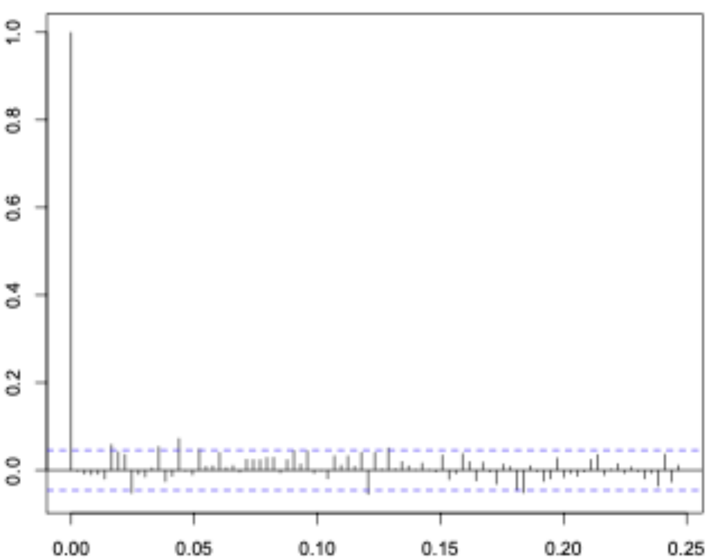

Figure 6: Seasonal Arima Autocorrelation Function

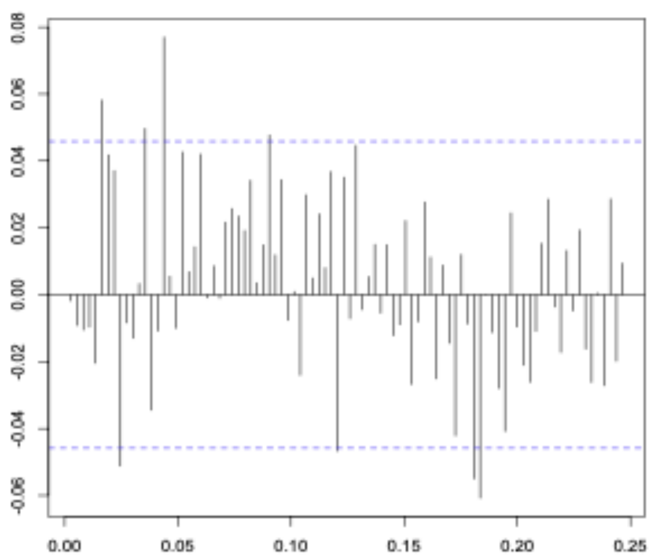

Figure 7: Seasonal Arima Partial Autocorrelation Function

In our chosen model's ACF, it can be observed that all of the lags are within the confidence interval except for a significant lag at 0, which is an improvement from the original autocorrelation function. Another important thing to note is that seasonality has been removed from our ACF which was the original goal put forth when picking a seasonal ARIMA. As for our PACF, we can see there are no longer many significant lags. It can also be noted that this PACF is almost white noise where it rarely gets all lags within the 95th confidence interval bands which shows far more improvement than our original PACF.

Results

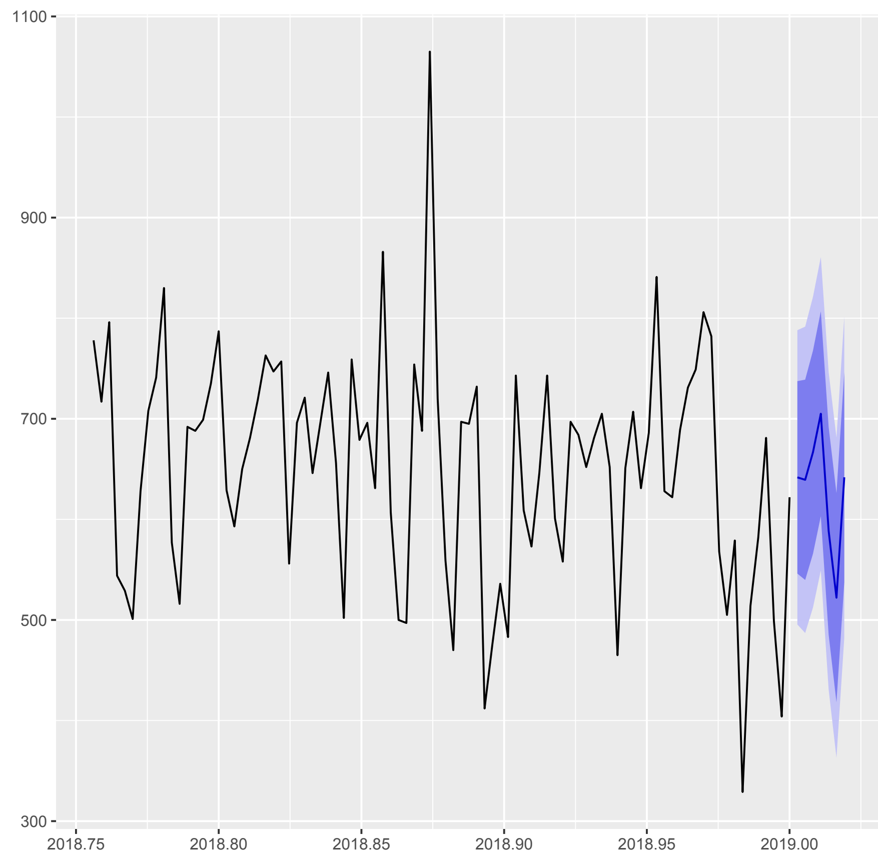

Figure 8: Forecasting pedestrian fatality rates over the month of January with 80th and 95th confidence intervals

As depicted above, pedestrian crash rates were forecasted for the month of January. The forecast mimics the behavior of the historical data and depicts a positive followed by a negative spike. These spikes could be best explained by January being a holiday season. The trend seen in the first couple of days could be attributed to the higher presence of drunk drivers and overall intoxicated people in the streets between dusk until dawn. What is important to note here is that on February 2016, the Dusk and Dawn campaign ran by NYPD, DOT and TLC was implemented in effort to not just inform drivers and pedestrians of the higher fatality rates during these hours but also, increase staffing on sight of more accident prone locations and increase deployment of inspectors to monitor speeding for TLC-licensed drivers. Despite this sector of vision zero bringing down fatalities by 25 percent within the first year of implementation, the years proceeding had opposite results. This could again be due to normalization of these programs over time. The lower spike depicted above could be best explained by it being the end of the first week of January and that less people are gathering between considerably dangerous hours. Also, people are still on vacation so auto vehicles on the street will still be low.

Table 1: Accuracy of 7-day prediction for pedestrian fatality rates with confidence intervals

Days	Lo 95	Predicted	Hi 95	Actual
1	495.4741	641.7716	788.0691	429
2	487.0700	639.3294	791.5889	502
3	512.4485	666.8782	821.3079	504
4	549.2444	705.0186	860.7929	598
5	430.8879	588.6356	746.3834	456
6	363.0883	522.1125	681.1367	376
7	482.2135	641.8135	801.4136	583

As could be seen in *table 1*, most of the predicted values were within the confidence intervals. Something interesting to note is that the actual values strayed towards the lower bound, which means the actual values skew negatively. This could be due to the seasonal ARIMA model overestimating. Nonetheless, the values have still increased from previous years, confirming the hypothesis that vision zero may have been successful at first, but as years passed, became not as effective. In the future, I hope to narrow down the data to not just time but also to contributing motor vehicle accident type. Given that Vision Zero has documentation on when all of its laws were passed within each of their programs, it would be far more insightful to narrow my data selection so I could better analyze why such spikes could have occurred within a specific law implementation period. It would also be interesting to see a comparison study between New York's vision zero and that of Europe's because Europe's vison zero has non diminishing results over time. All in all, these seemingly random events are not so random and can be analyzed in effort to reach vision's zero goal is having zero pedestrian fatalities.

‘

Work Cited

"Stats of the State of New York." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 10 Apr. 2018, www.cdc.gov/nchs/pressroom/states/newyork/newyork.htm.

Furfaro, Danielle, and Tina Moore. "NYC Traffic Injuries Are up despite Drop in Fatalities." *New York Post*, New York Post, 3 Jan. 2019, nypost.com/2019/01/02/nyc-traffic-injuries-are-up-despite-drop-in-fatalities/.

"Vision Zero In New York City." *Vision Zero*, www1.nyc.gov/content/visionzero/pages/.

Hyndman, Rob. "Forecasting with Daily Data." *Portrait of the Author*, 17 Sept. 2013, robjhyndman.com/hyndsight/dailydata/.

"Vision Zero." *Vision Zero: Year 5 Report*, Mar. 2019, www1.nyc.gov/assets/visionzero/downloads/pdf/vision-zero-year-5-report.pdf.