Gasonline Prices Effects on Auto Fatalities in New York City, 2001-2008

Guoqing Wang

October 29, 2016

1 Introduction

Many works have found that the increasing gas price can lead to a decreasing trend of motor vehicle fatalities (Grabowski and Morrisey 2004, 2006; Leigh and Geraghty 2008; Leigh and Wilkinson 1991). In general, higher prices lead to fewer miles driven, and lead to fewer chance for crash involvement and fewer fatalites. The aim of this analysis is to examine the relationship between fatal auto crashes and gasoline price, controlling the seasonal effect, temperature, and some other socioeconomic effects. The data used in this research is obtained from the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration. We are focusing on the New York City data with the year range from 2001 to 2008. The Annual average regular grade unleaded gasoline prices were obtained from Energy Information Administration (EIA), U.S. Department of Energy, which again will be focused on the year from 2001 to 2008.

2 Method

2.1 Data description

In this study, we estimate the effect of gasoline prices on the incidence of traffic fatal crashes in New York City from 2001 to 2008.

2.1.1 Fatality data

The traffic fatality data used in this study were obtained from Fatality Analysis Reporting System (FARS) data, collected by the National Highway Traffic Safety Administration (NHTSA; US Department of Transportation (USDOT), 2016). The FARS is a census of all motor vehicle crashes involving a fatality. A crash included in the FARS has to involve a motor vehicle traveling on a roadway customarily open to the public and had to result in the death of person within 30 days of the crash. It contains detailed information on the time, location, vehicles, drivers, occupants in the crash.

2.1.2 Gasoline prices

Gasonline prices in New York City were obtained from Energy Information Administration (EIA; US Department of Energy (USDOE), 2016). The per-gallon prices are the average retail prices from all gasoline outlets in New York City. The gasoline price data are adjusted for inflation in January 2009 US dollars.

2.1.3 Control variables

In this study, the data commonly included in the previous analysis to explain the vehicle miles traveled (VMT), annual daily traffic (AADT) counts is unfortunately unavailable. We include the the monthly weather information (temperature, precipitation) and socioeconomic factors into our analysis. Climate is used as control variable because climate varies seasonally and is found to affect traffic crashes (e.g., Quddus 2008). The weather data were obtained from the National Ocean and Atmospheric Association which is collected at Central Park, New York (2016).

Unemployment rates are included in this analysis as a representative of socioeconomic factors. The unemployment rates are used as a control variable because economic conditions affect consumer's ability to afford gasoline, then affects the occurrences of fatal traffic crashes (Graham and Glaister, 2003). The unemployment rates from 2001 to 2008 were obtained from the New York State Department of Labor (2016).

2.2 Model

In order to estimate the relative risk of fatality in a traffic accident for each dollar increase of gasoline price, we employ the log-linear regression model to fit the number of traffic fatality in each month. We first perform exploratory analysis by plot fatality in each month against the number of months, and add smoothing curve to illustrate the marginal trend. Figure 1(a) shows that the number of traffic fatality has a clear seasonal effects pattern, which is higher in the summers and lower in the winters. We perform the similar exploratory analysis for fatality versus gasoline price, precipitation, and unemployment rate. Figure 1(b) shows a decreasing trend of fatality against gasoline price.

Fatality has a approximately linear relationship with temperature as shown in Figure 1(c). Higher temperature will generally lead to more Fatalities, which is reasonable because it will lead to more traffic in general. Figure 1(d) shows a nonlinear relationship between fatality and unemployment rate, but overall fatality is higher at lower unemployment rate. We also find that traffic fatality depends on precipitation in a nonlinear way, with increasing fatality at low and high precipitation. The relative effect size of gasoline price and precipitation on fatality is smaller than the effect of time and unemployment rate. Extracting the gasoline price signal among other effects is the key of time series modeling of gasoline price and traffic safety. Since the fatality data has a seasonal pattern, we first try to eliminate the effect of seasonality by regressing the fatality on a natural spline smoothing function with 4 degree of freedom each year based on the Poission regression model. To reiterate, we are interested in estimating the relationship between the traffic fatalities and gasoline prices while controlling for other factors. Based on the exploratory data analysis, we build the log-linear regression model as follows:

$$\begin{aligned}\log(E(Y_t)) = \beta_0 + \beta_1 Gas_t + \beta_2 Gas_{t-1} + \beta_3 Temperature_t + \sum_{i=1}^{3} \beta_4^{(i)} f_{\text{Precip},i}(t) \\ + \sum_{j=1}^{3} \beta_5^{(4)} f_{\text{UR},j}(t) + \sum_{k=1}^{32} \beta_6^{(k)} f_{\text{Months},k}(t),\end{aligned} \tag{1}$$

where α denotes the intercept. Y_t is the fatality in NYC in month t, Gas_t is the gasoline price in current month. Gas_{t-1} is in one month lagged. $f_{\texttt{Precip},i}(t)$, $f_{\texttt{UR},j}(t)$,$f_{\texttt{Months},k}(t)$ are the basis functions for natural splines of precipitations, unemployment rate, and months, respectively. Total number of basis functions of each factor equals to the number of degree of freedom for the fitted natural spline smoother.

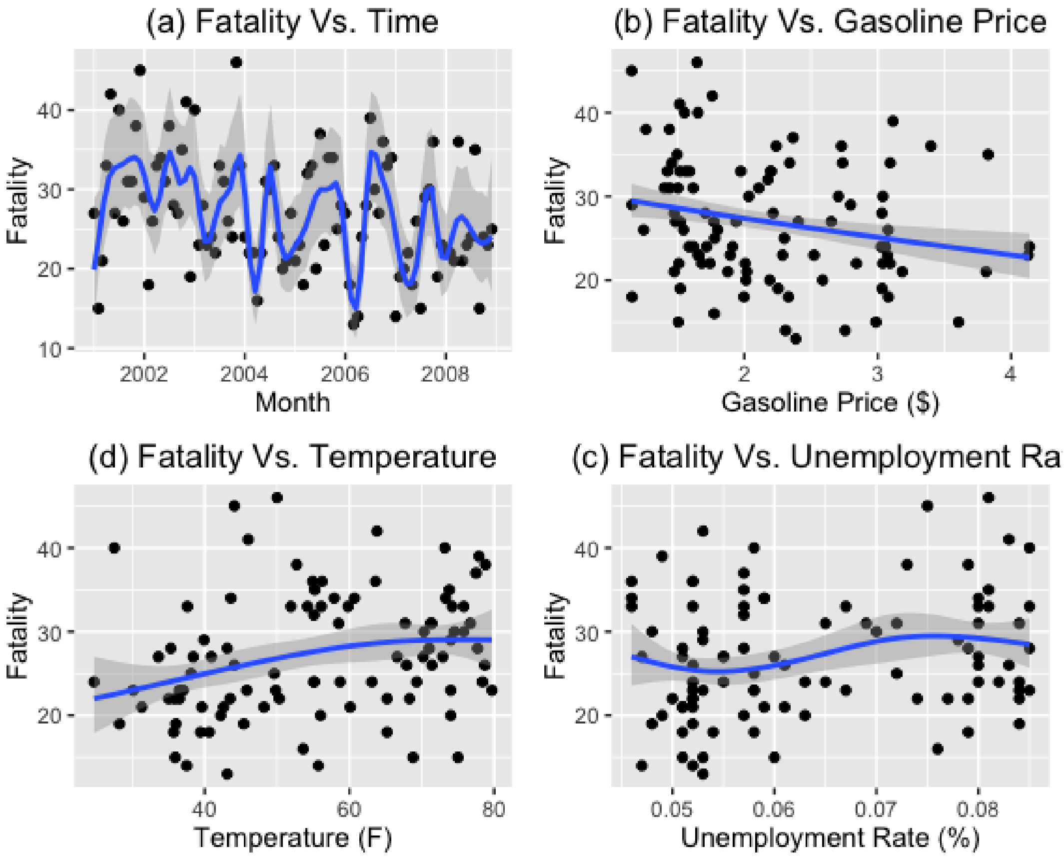

Figure 1: Exploratory data analysis: plotting the fatality against each potential predictor with a fitted smooth curve and 95% confidence interval with shaded grey area. a) Fitted with a natural spline smoother with 32 degree of freedom, fatalities shows a seasonality pattern; b) fitted with line, negative correlation between Fatality and Gasoline price; c)fitted with a natural spline smoother with 3 degree of freedom, fatality has a positive correlation with monthly averaged temperature; d)fitted with a natural spline smoother with 3 degree of freedom, fatality is higher at lower unemployment rates

We are interested in examining the effect of gas price in current month and lagged one month. So we include the Gas_t and Gas_{t-1} into the model, indicating those two effects, respectively. In order to check the effect of precipitation and unemployment rate on fatality, we perform a likelihood ratio test to compare the model without those two effects. The result shows fatality in NYC is not significant affected by unemployment rate and precipitation (p-value = 0.5031). Therefore, we exclude the unemployment rate and precipitation effect. Our final model is the following:

$$\log(E(Y_t)) = \beta_0 + \beta_1 Gas_t + \beta_2 Gas_{t-1} + \beta_3 Temperature_t + \sum_{k=1}^{32} \beta_4^{(k)} f_{\text{Months},k}(t), \tag{2}$$

There is some over-dispersion in this model with dispersion parameter equal to 1.488. Therefore, we instead use the quasi-possion family to inflate the result. To check the model assumption, independent assumption of Poisson regression is violated in the raw data: with observations close in time likely to be more similar than those distant in time. However, this 'autocorrelation' is usually not intrinsic to the outcome series, but rather due to autocorrelation in the explanatory variables that are predictors of the outcome. After controlling for seasonality, long-term patterns, the exposure of interest and other explanatory variables, residual autocorrelation will tend to be much smaller than in the raw outcome data, and is usually not

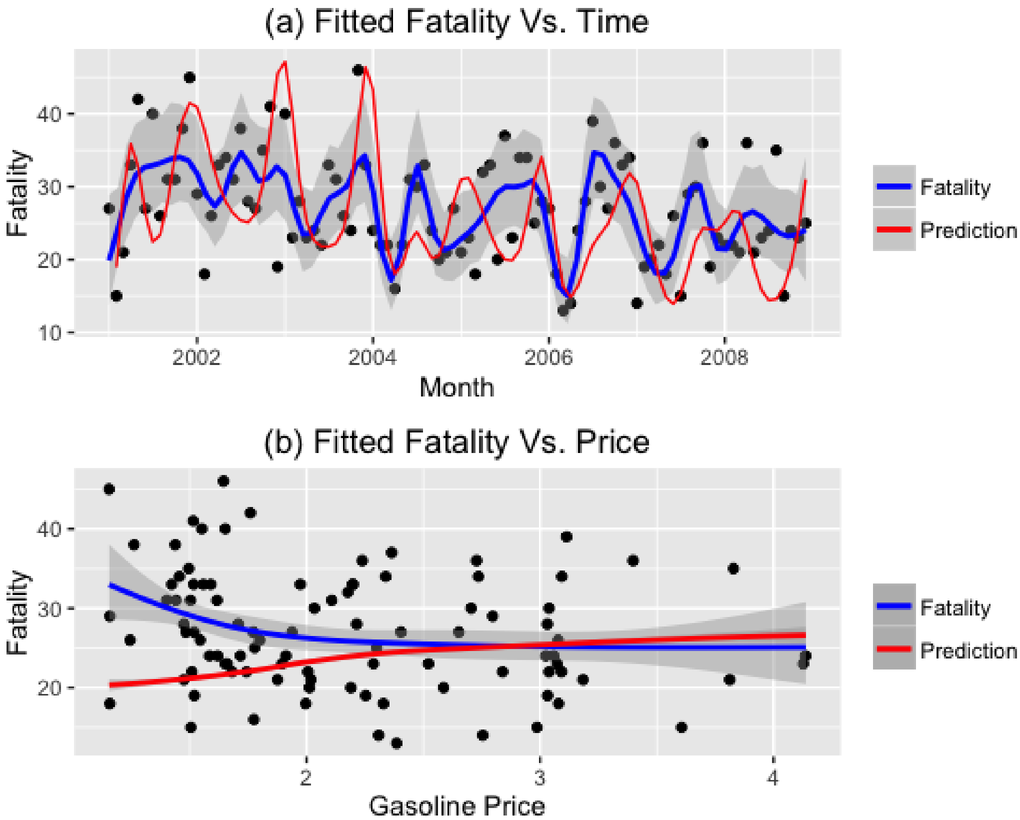

Figure 2: Observed and fitted values. a) Fitted values are at gasoline price equal to \$2.5, and temperature at $50F$. b) Fitted values are at time equal to April, 2014, and temperature equals to $67F$.

a major concern. To check the high influential points, there are 11 out of 96 observations highly influence on the result. After refit the model with highly influential point removed, the effect of gasoline price on fatality doesn't change substantially. Detailed model checking is included in supplemental materials.

Figure 2 shows fitted results over the observed result by fixing the gasoline price equal \$2.5 and temperature and $50F$ in 2(a) and time equal to April, 2014, and temperature equals to $67F$ in 2(b). The fitting performance is not consistent across all gas price due to lack of seasonality information which has relatively higher effective size.

3 Results

The coefficients in model 2 explain the effects of each factor. Table 1 shows the result. We first notice that the effects of gasoline price and temperature are not statistically significant in this model (p-value < 0). Even though the model include some nonsignificant effects, it still explains some effects. Controlling the temperature effects, seasonality and gasoline prices in prior one months, with 1 dollar increase in gasoline price, we estimates the traffic fatality will increase 6.22% (C.I. $(45.64\%, 107.01\%)$). In addition, for months with similar temperature, current month gasoline price and time, we would estimate the traffic fatality increase 4.20% (C.I (-42.98%, 89.78%)) with one dollar increase in gasoline price one month prior.

Effects	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	1.41	0.91	1.55	0.13
Gas_t	0.06	0.34	0.18	0.86
Gas_{t-1}	0.04	0.31	0.13	0.89
Temperature	0.01	0.01	1.30	0.20

Table 1: Results of model 2: coefficients of interception, two effects of gasoline price on traffic fatalities, temperature. None of these factors is statistically significant.

4 Discussion

In this study, we intend to examine and estimate the effect of gasoline price on the traffic fatality. Here we choose to use data from New York City to build the regression model. The result shows non significant effect of gasoline with controlling the effect of other potential confounding effects. Comparing with the results previous literature, where they found some significant effects of gasoline price, the main difference is due to the quality of data collection. The data of traffic fatality and weather condition we used is based on the level of New York City, and it has been highly aggregated in the month level. Also, the data on some other important effects, for example vehicle miles traveled (VMT) and annual daily traffic (AADT) counts, are not publicly available for New York City. These effects are potentially important effects in the model.

In the future study, we could first try to collect some more measured confounding effects, to make the results more comparable with other works. The other approach we will examine is to build the model in a higher level, for instance nation level or state level, to estimate the effect of gasoline price. However, this approach may encounter the the adjustment of variation between state and within state. Generalized mixed effect model will be more appropriate to be built than generalized linear model that we used in this study.

References

Grabowski DC, Morrisey MA, 2004. Gasoline prices and motor vehicle fatalities *J Policy Anal Manage.* 3(1), 23:575–593.

Grabowski DC, Morrisey MA, 2006. Do higher gasoline taxes save lives? *Econ Lett.* 90:51–55.

Graham, D.J., Glaister, S., 2003. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. *JUrban Studies.* 40 (8), 1591–1607.

Leigh JP, Geraghty EM, 2008. High gasoline prices and mortality from motor vehicle crashes and air pollution. *J Occup Environ Med.* 50:249–254.

Leigh JP, Wilkinson JT, 1991. The effect of gasoline taxes on highway fatalities. *J Policy Anal Manage.* 10:474–481.

National Highway Traffic Safety Administration, 2016. Fatality analysis reporting system encyclopedia: National statistics. `<http://www-fars.nhtsa.dot.gov/Main/index.aspx>` (retrieved 22.10.2016)

National Weather Service, 2016. Central Park, NY Historical Data < `http://www.weather.gov/okx/CentralParkHistorical>` (retrieved 22.10.2016)

New York State Department of Labor, 2016. Labor Statistics for the New York City Region. `<https://www.labor.ny.gov/stats/nyc/>` (retrieved 22.10.2016)

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. *JAccident Analysis and Prevention.* 40 (4), 1486–1497.

U.S. Energy Information of Administration, 2016. New York City Total Gasoline through Company Outlets Price by All Sellers. < `https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_y35ny_m.htm>` (retrieved 22.10.2016)