

# Analysis of association between unemployment rate and traffic fatality across counties in Maryland

*Boyang Zhang*

## **Introduction**

Over 35,000 traffic fatalities occurred in 2015. In 2015, there was a 7.2% increase in the number of traffic-related fatalities compared to 2014 [1]. Apart from regulatory safety policies, such as DUI laws and seat belt enforcement, macroeconomic condition has been identified as another important factor related to temporal changes in fatality rates in previous studies. However, little research has been conducted on association between macroeconomic conditions (i.e. unemployment rate) and number of traffic-related fatalities across counties [2]. Therefore, we focus on quantifying the association between the unemployment rate and traffic-related fatalities across counties in Maryland.

In this report, we estimated association between unemployment rate and traffic fatalities across counties in Maryland. We used Fatality Analysis Reporting System (FARS) data provided by the National Highway Traffic Safety Administration (NHTSA) [1]. We performed a generalized linear mixed model with random county effects on monthly data for 24 counties in Maryland from 2005 to 2015. Through comparing association across counties, policy makers could take this factor into consideration when determining effective ways to allocate traffic safety resources.

## **Methods**

### **Data collection**

The number of traffic-related fatalities for each county in Maryland from 2005 to 2015 was obtained from the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration [1]. Population estimates (total number of people living) in each county in Maryland was obtained from U.S. Census Bureau [3]. County-level monthly unemployment rates were obtained from Local Area Unemployment Statistics [4]. All data were downloaded using the R programming language except post 2010 population estimates, which needed to be downloaded by hand [5].

## **Methods**

### **Data collection**

The number of traffic-related fatalities for each county in Maryland from 2005 to 2015 was obtained from the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration [1]. Population estimates (total number of people living) in each county in Maryland was obtained from U.S. Census Bureau [3]. County-level monthly unemployment rates were obtained from Local Area Unemployment Statistics [4]. All data were downloaded using the R programming language except post 2010 population estimates, which needed to be downloaded manually [5].

### **Key variable definitions**

This analysis estimates the statistical relationship between unemployment rate and traffic fatality via a cross-county analysis using all 24 counties in the State of Maryland. FARS data recorded traffic-related fatalities by date, but we decided to summarize the data on a monthly basis. The outcome in our study is number of traffic-related fatalities. We computed traffic-related fatalities for each month from January of 2005 through December of 2015. The key exposure variable is the unemployment rate and it was measured as monthly average for each county in Maryland during 2005 to 2015.

As for key confounding variables, we chose to focus on person-level information and climate conditions. Proportion of drunk drivers, drivers' age, number of people who are at risk, severe weather condition and

light condition could be other potential risk factors associated with total number of traffic fatalities in a specific county in a given month. For proportion of drunk drivers, we imputed missingness in the indicator of drunk drivers by setting threshold for blood-alcohol content (BAC) at 0.08, according to Maryland’s DUI laws. Then we calculated county-specific proportion of drunk drivers involved in fatal crashes in each month. Also, we included average drivers’ age in fatal crashes for each county and county-level population on a monthly basis. In addition, by defining 0 as clear weather and 1 as severe weather conditions, we obtained proportion of severe weather condition in a given month for each county. Similarly, we re-categorized light conditions into 3 levels, namely night without light, obscure light (night light, dawn and dusk) and daylight, and considered average light conditions as a measurement for light condition for each county in a given month. The range for light condition is from 0 to 2, where the larger number indicates clearer vision. To adjust for seasonality, we would take calendar time into consideration. Additionally, we assumed that number of people at risk in each county would not change significantly throughout a year. The detailed definitions and variable calculation will be discussed further in our supplementary document [10].

### Statistical Modeling

To quantify the relationship between traffic-related fatalities and the unemployment rate, we utilized a generalized linear mixed model where we assumed that number of traffic-related fatalities follows a Poisson distribution. A random intercept at the county level was included to account for the correlation in the number of traffic-related fatalities within a county over time.

Let  $\mu_i$  be the average traffic-related fatalities for  $i$  th county ( $i=1,2,\dots, 24$ ).

$$\log \mu_i = (\beta_0 + \gamma_i) + \beta_1 \text{unemprate} + \beta_2 \text{lgcond} + \beta_3 \text{population} + \beta_4 \text{ns(ym, 2)} + \beta_5 \text{age} + \beta_6 (\text{age} - 40)^+$$

Here,  $\beta_0$  is fixed intercept and  $\beta_1, \dots, \beta_6$  are fixed effects for unemployment rate, light condition, population, calendar time and age with its additional linear spline term respectively.  $\gamma_i$  represents additive random effect for county.

### Results

To explore the association between the number of traffic-related fatalities and economic conditions, we plotted the monthly traffic-related fatalities and unemployment rate during 2005-2015. In Figure 1, the four plots reveals trend for traffic fatalities against age, seasonal pattern for traffic fatalities, trend of unemployment rate and trend for traffic fatalities against proportion of severe weather in Baltimore City. For drivers’ age, we noticed that the fatalities displayed different patterns before 40 and after 40 years old. Hence, we applied a linear spline for age at 40 years old. As the traffic fatalities showed non-linear dependence on calendar date (i.e. year-month), we applied a natural spline for date with 2 degrees of freedom to account for seasonality. In Baltimore City, the unemployment rate reaches its peak around 2011 when the United States had entered a severe economic recession since 2008. The average unemployment rate in Baltimore City for year 2005 was 6.92% (SD:0.50), 10.72% (SD: 0.50) in 2011 and 7.7% (SD: 0.53) in 2015. We also performed the exploratory analyses on Anne Arundel County and Prince George’s County, which will be discussed in details in supplementary document [10].

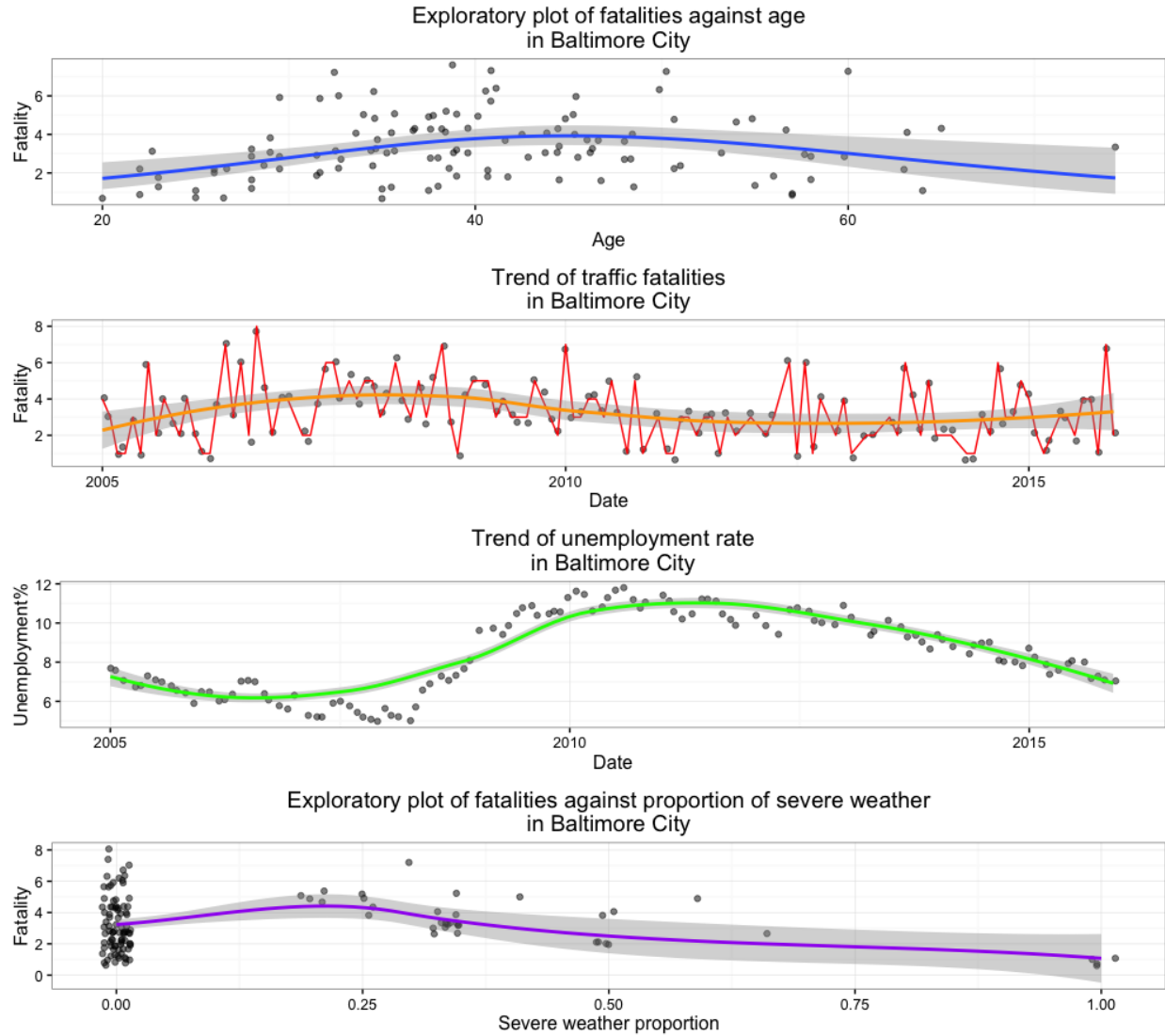


Figure 1: Exploratory analysis of Baltimore City and Montgomery County. They are exploratory plots for fatalities against age, trend of traffic fatalities, trend of unemployment rate and traffic fatalities against proportion of severe weather in Baltimore City from 2005 to 2015. Note that blue line represents a smoothing line for traffic fatalities against age using natural spline with 2 degrees of freedom while the trend for other variables (marked in orange, green and purple) are displayed in a loess smoothing line. The shaded area for all plots are point-wise confidence interval.

Based on our prior knowledge and exploratory analyses, we identified potential main effects, such as person-level information (i.e. age, proportion of drunk drivers and population), climate information (i.e. weather and light conditions), calendar time and economic conditions (i.e. unemployment rate). Moreover, interaction effects between age, weather, light conditions and unemployment rate were included in the final model if statistically significant. In order to obtain unbiased estimates of coefficients and standard errors, we divided dataset into training set (50%) and test set (50%). Using likelihood ratio test as a criterion, we included statistical significant main effects, such as unemployment rate, light conditions ( $P=4.23 \times 10^{-15}$ ), age ( $P=4.52 \times 10^{-14}$ ) and county-level population ( $P=8.44 \times 10^{-9}$ ). Since we noticed seasonal pattern in exploratory analysis, we need to include a natural spline of calendar time with two degrees of freedom to account for seasonality based on our prior knowledge.

In addition, we performed likelihood ratio test to test whether each of the proposed interaction effects contributes significantly to model fitness but none of them was statistically significant. As for random effects, we only included random intercept at county level for the reason that random county slope for unemployment rate did not show statistical significance. All our variable inclusion was performed on training set.

Table 1: Summary of exp(coef) and its 95% confidence intervals for fixed effects

	Estimates of exp(coef)	Lower CI of exp(coef)	Upper CI of exp(coef)
Intercept	2.7408	2.2987	3.2680
unemployment rate	0.9620	0.9341	0.9908
light condition1	1.4219	1.2562	1.6094
light condition2	1.0913	0.9608	1.2394
population	1.4423	1.3292	1.5650
ns(year-month,2)1	0.8533	0.6253	1.1645
ns(year-month,2)2	0.8313	0.7339	0.9418
age	1.4388	1.2952	1.5982
(age - 40) <sup>+</sup>	0.6574	0.5898	0.7328

Table 1 shows the summary of exponential transformation of coefficients and its 95% confidence interval for fixed effects. By interpreting the results, we know that at same level of light condition, population size, age and date (i.e. year-month), the expected traffic-related fatalities with 1 percent increase in unemployment rate is 0.962 times (95%CI: (0.9341, 0.9908)) as that with original unemployment rate. That is to say, after adjusting for drivers' age, county-level population size, calendar time (i.e. year-month) and average light condition, we estimate a 3.8 % (95%CI: (0.92 %,6.59%)) decrease in the average number of traffic-related fatalities per 1 percent increase in the unemployment rate. For drivers who are younger than 40 years old, the expected counts of traffic fatalities increases 43.88% (95%CI: (29.52%, 59.82%)) with 1 year older. Similarly, for drivers that are older than 40 years old, 1 unit increases in age is associated with additional 43.88 % decreases (95%CI: (59.82 %,29.52%)) in expected traffic-related fatalities.

## Discussion and Limitation

A major finding of this study is a statistically significant ( $P= 0.014$ ) inverse relationship between county level unemployment rate and traffic-related fatality after controlling for several demographic and geographic factors also linked to traffic fatalities. To be precise, after adjusting for drivers' age, county-level population size, calendar time (i.e. year-month) and average light condition, we estimate a 3.8 % (95%CI: (0.92 %,6.59%)) decrease in the average number of traffic-related fatalities per 1 percent increase in the unemployment rate. Also, through plotting observed values and fitted values for traffic fatalities in figure 2, our model captures overall seasonal pattern but not extreme cases in both Baltimore City and Montgomery County.

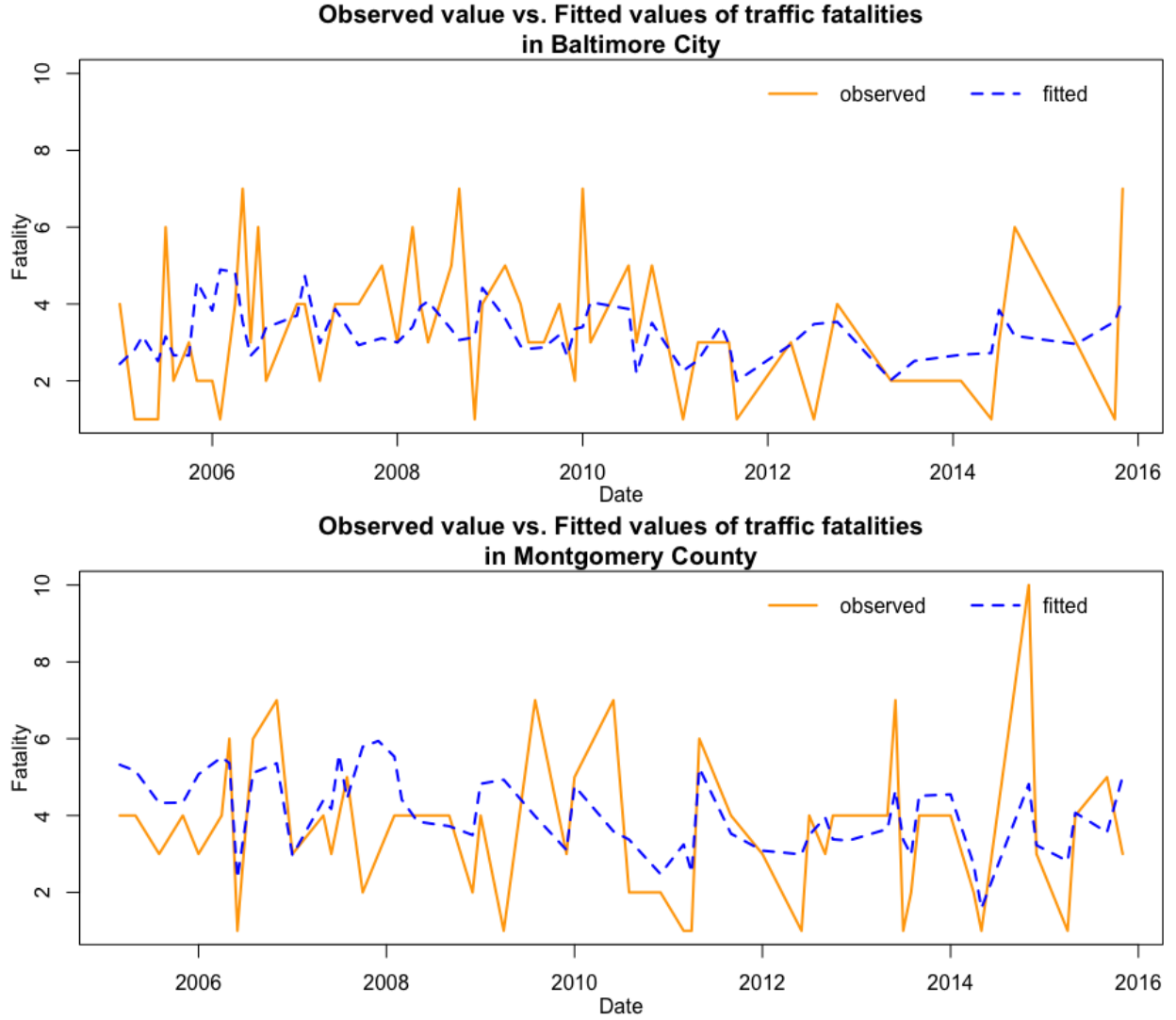


Figure 2: Observed traffic fatalities against fitted values in Baltimore City and Montgomery County. Note that the orange solid line represents observed traffic fatalities in test set and the blue dashed line represents fitted values for traffic fatalities in either city using our model.

This association could have multiple explanations. One possible explanation is that the unemployed may experience less psychosocial stress compared to the stress level at work. The low level of psychosocial stress may allow drivers to drive safely and therefore lead to a decrease in traffic fatalities. Or perhaps, due to higher unemployment rate, less vehicle density will be expected on road during rush hours and thus lower regional traffic-related fatalities [6]. Further analysis could be conducted to account for psychosocial stress

level for individuals and to see whether association is causation.

There are some limitations associated with this study. Firstly, this study only measures fatalities associated with traffic crashes on a monthly basis and therefore summarizing data may have some information loss. Secondly, the model may only capture overall seasonal pattern but not extreme cases as shown in figure 2. Further analysis needs to be conducted to search for other potential factors related to traffic fatalities to build a better prediction model. Thirdly, this study only measures the fatalities associated with traffic crashes. As it neglects nonfatal crashes, it may not capture the total impact of economic conditions (i.e. unemployment rate) on overall traffic safety. If we could obtain information for nonfatal crashes, we could probably capture the total impact of economic conditions (i.e. unemployment rate) on overall traffic safety and improve prediction accuracy. In this way, policy makers could rely on our results and allocated their resources more efficiently.

## Reproducibility

All data analyses performed in this report can be reproduced in the R markdown file: `trafficfatality.Rmd`. To reproduce exact results, data obtained online should be the same version of data we downloaded on 2016-10-01.

## Reference

- [1] National Highway Traffic Safety Administration (2015).
- [2] The Relationship Between Regional Economic Conditions and the Severity of Traffic Crashes; Thomas L. Traynor; Traffic Injury Prevention.Vol. 10 , Iss. 4,2009
- [3] U.S. Census Bureau: Annual Estimates of the Resident Population for Counties. <http://www.census.gov/popest/data/counties/totals/2015/CO-EST2015-01.html>
- [4] Local Area Unemployment Statistics (LAUS) - Workforce Information & Performance <https://www.dllr.state.md.us/lmi/laus/>
- [5] American FactFinder <http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk>
- [6] Alexander C. Wagenaar (1983). Unemployment and Motor Vehicle Accidents in Michigan. Report Number: UMTRI-83-45
- [7] Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3),7-10. URL <http://CRAN.R-project.org/doc/Rnews/>
- [8] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [9] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- [10] Boyang Zhang. Supplementary Document For Traffic Fatality Project.