Spatio-Temporal Trends in Mass Shooting Incidents in the United States

Ali Arab*

Abstract
The recent increase in mass shooting incidents in the United States has drawn attention to better understand this phenomenon. To this end, there have been recent efforts in collecting and analyzing data from mass shootings. Most of these studies focus on identifying the underlying factors related to mass shootings. However, there is also interest in understanding spatial and temporal patterns in mass shootings data. Our proposed method correlates mass shooting data (number of victims killed or injured) with State level predictor information on attributes such as population, access to mental health care, suicide rate, crime rate (e.g., gun death rate), as well as several gun policy indicators. We identified several factors to be statistically significant for these data (e.g., population, gun death rate, and access to mental health care) while none of the gun policy related predictors were significant (probably due to weak gun policies in most States). Critically, this method considers spatial dependence in the data and thus, it allows for detecting mass shooting hotspots. We also showed statistical evidence for increasing trend of frequency of mass shootings.

Key Words: Bayesian, Hierarchical Models, Spatial Statistics, Time-to-Event Data

1. Introduction

There is growing concerns regarding the increasing trend in gun violence in the United States to the extent that the American Medical Association considers it a public health crisis (See web link: http://www.ama-assn.org/ama/pub/news/news/2016/2016-06-14-gun-violence-lobby-congress.page). In particular, the recent surge in mass shootings in the U.S. has drawn attention of both experts and the general public, and generated many debates among researchers (e.g. see Metzl and MacLeish 2015).

While research on gun violence and mass shootings suffers from a general lack of reliable data (e.g., see Foran 2016), several academic and non-profit organizations have recently attempted to compile databases on data scraped from the web, and official sources (if and when available). The data bases developed by the Gun Violence Archive, and Stanford University's Mass Shootings in America project are examples of such efforts. In particular, evidence-based discussions on mass shootings often refer to these (or similar) databases as the main reliable sources (Follman, Aronsen, and Pan 2012; Fox and DeLateur 2013). However, there is a lack of rigorous statistical analyses for mass shootings data in these debates, and most studies rely on very basic statistical analyses which often simply summarize the data. To this end, we developed a modeling framework using data on U.S. mass shootings from a popular sources (Gun Violence Archive, and Stanford University's Mass Shootings in America project).

In this paper, our goal is to understand patterns and trends in recent mass shooting incidents. To this end, we look at spatial and temporal patterns, as well as, the patterns in occurrence of these incidents.

2. Data and Methods

We use data on mass shooting incidents from the Gun Violence Archive (see: Shooting-tracker.com) which is a comprehensive database of mass shootings since 2013. The data

*Georgetown University, 3700 Reservoir Rd. Washington, DC 20057

includes 1042 records of mass shootings during the period between January 1, 2013 and July 7, 2016. The Gun Violence Archive uses the definition described by the Federal Bureau of Investigation (FBI) for a mass shooting event such that the event is defined as a mass shooting incident if four or more persons are shot and/or killed in a single incident, at the same general time and location, not including the shooter. We obtained geographical coordinates for the data pints. The data for each state, in each year then were grouped together and the average of data coordinates was used as the centroid for the state.

We also consider several potential predictor variables for each state including population, gun death rates, suicide rates, mental health rates, and access to mental health care (Percent of mental health need met). These data are obtained from the Center for Disease Control and Prevention (CDC) and the Kaiser Family Foundation. Additionally, we consider gun policy indicators at the state level including child access prevention firearm law, safe storage or gun lock requirement, assault weapons ban, universal background checks requirement. These data were obtained from the websites smartgunlaws.org and gunlawscorecard.org, and whenever possible, the data were validated with the data from the National Rifle Association (NRA).

2.1 Hierarchical Spatial Model

We consider a hierarchical spatial modeling framework to model the spatially-referenced mass shooting fatalities data. The data are assumed to follow a Poisson distribution:

$$Y_i|\lambda_i \sim \text{Poisson}(\lambda_i). \tag{1}$$

We then consider a log-linear regression model for the Poisson intensities:

$$\log(\lambda_i) = \beta_0 + \sum_{j=1}^{k} \beta_k X_{ik} + \eta_i, \tag{2}$$

where X_k's denote the predictor variables, and β_j's ($j = 0, 1, \ldots, k$) denote the intercept and the regression coefficients for the predictor variables. We considered several potential predictors and our model selection lead us to consider four predictor variables for each state: population, proportion of mental health need met (denotes by "Menta Health Care"), suicide rate, and gun deaths rate.

We considered spatially correlated error terms η_i's in Equation (2). The spatial dependence is described based on an exponential function of Euclidean distances between coordinates of the data (the average of the data coordinates within each state was used as the centroid coordinates for that state). The spatial model is described below

$$\begin{aligned} \Sigma &= \sigma_\eta^2 \mathbf{R}(\theta), \\ R(\theta) &= \exp(-\theta d). \end{aligned}$$

2.2 Modeling Time-to-Event Data

In order to model the temporal trend in the frequency of mass shooting events, we adapt the piecewise exponential (PEXP) model described in Arab et al. (2012). Let $X_1, X_2, \ldots, X_n$ denote the time between mass shooting events. The PEXP model assumes

$$E\left(X_j\right) = \frac{\delta}{\mu} j^{\delta-1}, \quad \delta > 0 \text{ and } \mu > 0, \tag{3}$$

where parameter δ describes the “reliability status” of the system such that if $\delta < 1$ the system is deteriorating (i.e., more frequent “failures” or “events”), while $\delta > 1$ denotes “reliability improvement” (i.e., less frequent “failures” or “events”). Note that if $\delta = 1$ the non-homogenous Poisson process considered in PEXP, reduces to a homogenous Poisson process (i.e., constant rate of “failures” or “events”).

3. Results and Discussion

Inference was conducted based on Bayesian estimation and Markov Chain Monte Carlo (MCMC) method. We assigned relatively non-informative priors for the unknown parameters in both models. We used the package spBayes (Finely et al. 2007) in R to fit a spatial Poisson regression with 25,000 MCMC iterations and 5000 burn-in. Similarly, we used 25,000 MCMC iterations with 5000 burn-in and the freeware OpenBUGS (Lunn et al. 2009) to fit the PEXP model for the time between mass shooting events. It should be noted that the two analyses are conducted independently.

Results for the spatial model are shown in Table 1. Both population and gun deaths rates are positively associated with number of victims killed in mass shootings. These results are reasonable and intuitive given that most mass shootings occur in metropolitan and highly populated areas, where gun crime is usually high (and thus gun deaths rates are typically high). Mental Health Care and suicide rates are negatively correlated with number of victims killed in mass shootings. The variable Mental Health Care describes the proportion of mental health need met in a state, and thus, the result is intuitive that the states with lower mental health care needs met tend to have higher mass shooting fatalities. Also, lower suicide rates are associated with higher mass shooting fatalities. We are not sure how to interpret this result in this context as it is not necessarily a direct measure of mental health and may be tied to different factors including socio-economic factors. It should be noted that the associations described in this model are merely used as proxies to potentially capture the variability in the mass shooting fatalities data and the author has no intention in attempting to describe the underlying factors that may “trigger” mass shootings.

The spatial surface for the observed and fitted values are shown in Figures 6 and 7, respectively. Based on the comparison of these two surfaces, it is apparent that the model is underestimating some of the values and particularly, it is unable to provide a good fit for high values. To investigate this issue further, we compare the fitted and observed values directly. Figure 8 shows the plot of fitted versus observed values. As shown in this figure, the fitted values are underestimating the observed values for California, Florida, and Texas. This is due to the relatively high numbers of total fatalities in these three states during the period of the study.

Results for the time-to-event model, the PEXP model, are shown in Table 2. The mean parameter μ is estimated as 0.32 with standard deviation of 0.0712 (i.e., mass shootings occur on average every 2 to 5 days). The “reliability status” parameter δ is estimated as 0.8626 (with standard deviation of 0.031). The confidence interval for δ is $(9.8001, 0.922)$ and thus, we conclude that the “system reliability” is deteriorating. In other words, during the period January 1, 2013 to July 7, 2016, the mass shootings have statistically significantly increased in frequency. Our results confirm the results of previous studies (Follman, Aronsen, and Pan 2012; Fox and DeLateur 2013) and the general perception of increased frequency of mass shootings over the recent years based on statistical evidence.

Future work should extend the proposed method to provide a better modeling approach to account for extreme cases and rare events (such as those in Califronia, Florida, and Texas). This may be done by considering a double hurdle model as described in Balderama et al. (2016). Also, a more rigorous search for potential predictor variables at the local

Table 1: Posterior results for the PEXP model for time between events.

Paramater	Posterior Mean	Posterior St. Dev.	95 CI%
Intercept	3.56	0.57	(2.44, 4.59)
Population	0.76	0.09	(0.58, 0.95)
Mental Health Care	-1.65	0.68	(-2.97, -0.31)
Suicide Rate	-0.14	0.03	(-0.20, -0.08)
Gun Deaths	0.17	0.03	(0.10, 0.22)

Table 2: Posterior results for the PEXP model for time between events.

Paramater	Posterior Mean	Posterior St. Dev.	95 CI%
μ	0.32	0.0712	(0.1994, 0.477)
δ	0.8626	0.031	(0.8001, 0.922)

levels should be considered. Mass shooting events are rare events in nature, and this would complicate finding statistical relationships between mas shooting data and predictor data, however, by collecting data on the shooter, as well as the circumstances that may have lead to the shooting event, we may be able to better understand this complex phenomenon.

Time between mass shootings during January 2013 and July 2016 have a significantly decreasing trend since the 95% CI for δ is smaller than 1.

REFERENCES

Arab A., Rigdon S. E. and A. P. Basu (2012). Bayesian Inference for Piecewise Exponential Model for the Reliability of Multiple Repairable Systems. *Journal of Quality Technology*. Volume 44, Issue 1, pages 28-38.

Arab, A., Jackson, M. and C. Kongoli (2014). Spatio-Temporal Analysis of Malaria Trends and Weather Effects in Western Africa. *Malaria Journal*.

Arab, A. (2015). Spatial and Spatio-Temporal Models for Modeling Epidemiological Data with Excess Zeros. *Int. J. Environ. Res. Public Health*, 12, pages 10536-10548.

Arab A, Courter J. R., and J. Zelt (2016). A Spatio-Temporal Comparison of Avian Migration Phenology Using Citizen Science Data. *Spatial Statistics*. In print.

Balderama, E., Gardner, B., and Reich, B. J. (2016). A spatial?temporal double-hurdle model for extremely over-dispersed avian count data. *Spatial Statistics*. In print.

Finley, A. O., Banerjee, S. and B. P. Carlin (2007). spBayes: An R Package for Univariate and Multivariate Hierarchical Point-Referenced Spatial Models. *Journal of Statistical Software*, 19(4), 1-24.

Follman, M., Aronsen, G., and Pan, D. (2012). A guide to mass shootings in America. *Mother Jones*, 15, 119.

Foran, C. (2016).The Missing Data on Gun Violence. *The Atlantic*, published on January 21, 2016; Accessed on September 23, 2016.

Fox, J. A., and DeLateur, M. J. (2013). Mass shootings in America: moving beyond Newtown. *Homicide Studies*, 1088767913510297.

Lunn D., Spiegelhalter D., Thomas A., and Best, N. (2009). The BUGS project: evolution, critique and future directions. *Stat Med*, 28(25):3049?3067.

Metzl, J. M., and MacLeish, K. T. (2015). Mental Illness, Mass Shootings, and the Politics of American Firearms. *American Journal of Public Health*, 105(2), 240?249. http://doi.org/10.2105/AJPH.2014.302242

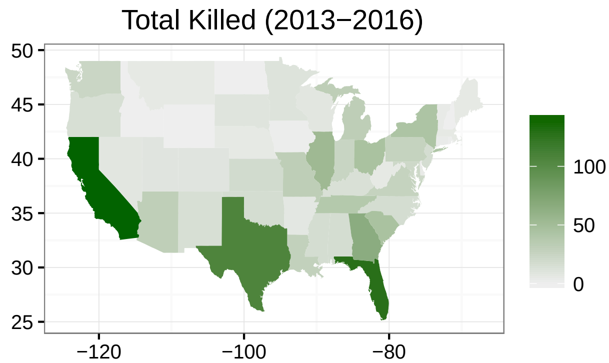

Figure 1: Total killed in mass shootings in the United States between January 1, 2013 and July 7, 2016.

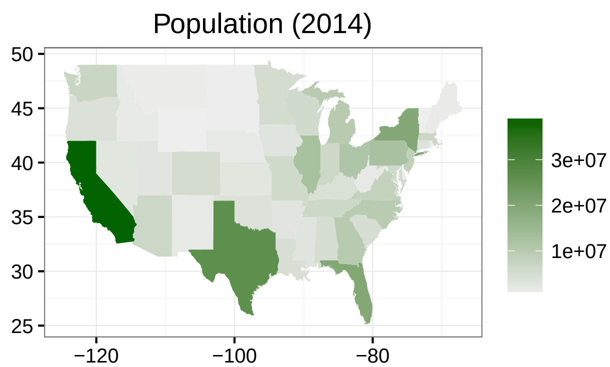

Figure 2: Population for each State according to 2014 data.

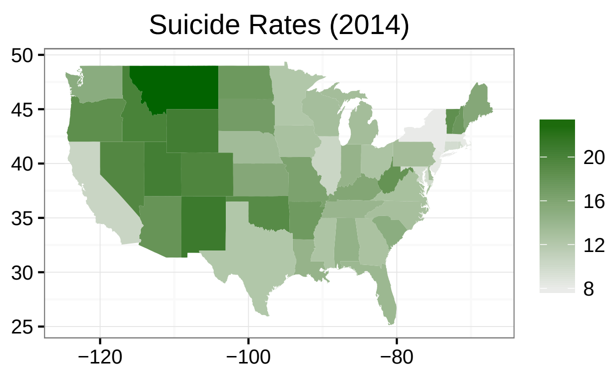

Figure 3: Suicide rates for each State according to 2014 data.

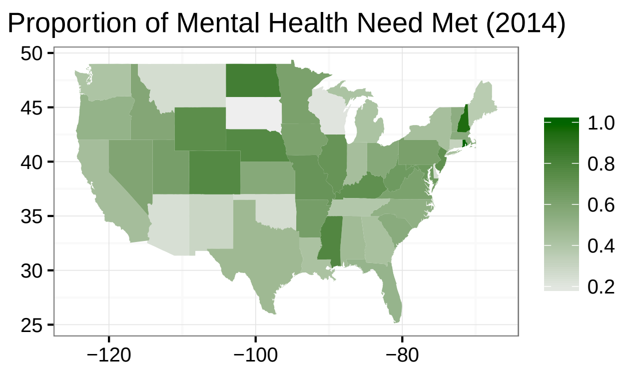

Figure 4: Proportion of mental Health care need met for each State according to 2014 data.

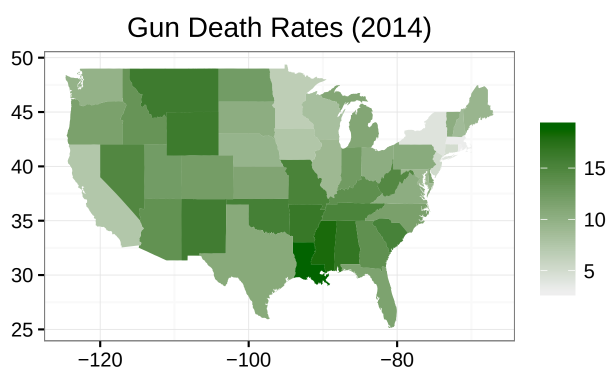

Figure 5: Gun related deaths rates for each State according to 2014 data.

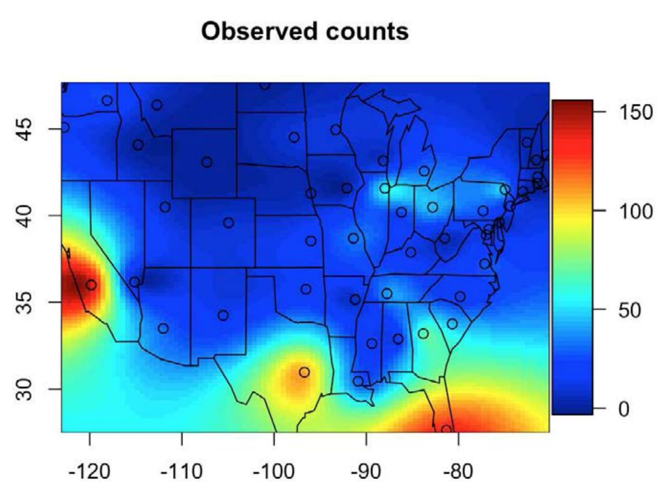

Figure 6: Spatial surface of observed values.

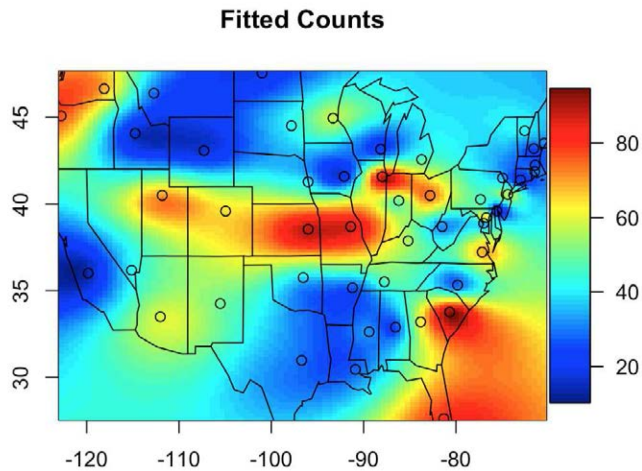

Figure 7: Spatial surface of fitted values.

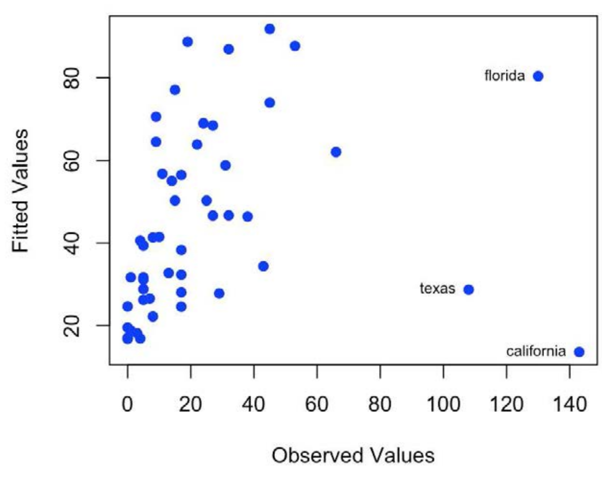

Figure 8: Scatterplot of fitted values versus observed values. The three most extreme cases of lack of fit are highlighted (Florida, Texas, and California).