Analyzing the Risk of Mass Shootings in the United States

by

Xue Lei

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Industrial Engineering

Program of Study Committee:
Cameron A. Mackenzie, Major Professor
Danial Davarnia
Guiping Hu
Qing Li
Jarad Niemi

Iowa State University

Ames, Iowa

2022

DEDICATION

This dissertation is dedicated to my beloved family. My parents and my husband support me with their patience, encouragement, care and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who made this work possible.

First of all, I would like to thank my major professor, Dr. Cameron A. Mackenzie, for his motivation, inspiration, support and patience throughout my Ph.D. study. He has helped me become a better person than I could ever be without him.

I would like to express gratitude to my committee members, Dr. Danial Davarnia, Dr. Guiping Hu, Dr. Qing Li, Dr. Jarad Niemi for their efforts, suggestions and contribution to this dissertation.

Many thanks to all faculty and staff in Department of IMSE, and all course instructors, who made my life at ISU an enjoyable experience.

ABSTRACT

Gun violence leads to tens of thousands of fatalities in the United States each year. Although a very small proportion of gun violence in the United States is due to mass shootings, the number of mass shootings in the United States represents a large proportion of global mass shootings. The number of mass shootings seems to be increasing in the United States. Since mass shootings take many innocent lives in the United States and generate significant media attention and public debate, it is necessary to understand the risks of mass shootings, analyze how the risk is changing over time, forecast the risk of mass shootings, and assess the risk for specific locations.

This research models the annual trend in mass shootings by using several different models, to include Poisson regression, change-point models, a time series model, and a hybrid of a time series and neural network. The analysis of the fitted results provides useful insight into whether the historical frequency of mass shootings has changed. Several different change-point models are analyzed to ascertain which change-point model provides the best fit to the frequency of mass shootings.

The forecasting performance of three forecasting models, a change-point model, a time series model, and a hybrid of a time series model with an artificial neural network model are compared. Each model is applied to forecast the frequency of mass shootings. Comparing among results from these models reveals advantages and disadvantages of each model when forecasting rare events such as mass shootings. The insights generated from the comparison are beneficial for selecting the best model and accurately estimating the risk of mass shootings in the United States.

Quantifying the risk of mass shootings in the United States provides a good picture of risk at the national level but further analysis is needed to understand the risk at a state and local level. Quantifying the risk of a mass shooting at specific locations provides a greater understanding of the chance of a mass shooting at a school, house of worship, or even a workplace. A scientifically rigorous and evidence-based method is proposed to calculate the probability that a mass shooting occurs in a state or a specific location. Quantifying the risk of mass shootings at different locations in different areas can provide insight into which areas should be protected against mass shootings.

CHAPTER 1. INTRODUCTION

1.1 Motivation

Mass shootings are a significant social problem in the United States in recent years. A substantial amount of media attention is frequently generated when mass shootings occur. The media attention generated by mass shootings emphasizes the tragic nature of these horrific events. On February 14, 2018, a 19-year-old killed 17 people and injured 17 at Marjory Stoneman Douglas High School in Parkland, Florida (Coughlan, 2018). On March 22, 2021, 10 people were killed at a supermarket in Boulder, Colorado (Macaya and Hayes, 2021). Although fatalities from mass shootings represent a very small part of all fatalities caused by gun violence (UC Davis Health, 2018), mass shootings create more public concern and anger than other types of shootings. Since mass shootings are a serious problem in the United States leading to scores of fatalities every year, mass shootings generate significant debates about the best ways to reduce the occurrence and fatalities from mass shootings.

Risk management is a discipline designed to assess and evaluate among strategies to mitigate risky events such as mass shootings. A principle of risk management is that it is first necessary to provide a quantitative assessment of the risk in order to properly determine how to best mitigate the risk (Hubbard, 2009; Kaplan and Garrick, 1981). A quantitative assessment of the risk is composed of the probability of the event and the consequences if the event occurs. In order to understand how to best mitigate the risk of mass shootings, it is necessary to first quantify the risk of mass shootings in the United States.

The risk assessment of mass shootings in the United States requires forecasting or estimating the frequency of mass shootings in the United States. Since the frequency in mass shootings seems to be increasing, modeling the historical trend can provide insight into how the risk of mass shootings may increase in the future. Forecasting the annual number of mass shootings in the future can help us understand the risk of a mass shooting in the United States and lead to better safety education and law enforcement. A mathematical model can help answer to what extent the rate of mass shootings is increasing and whether the rate has steadily increased or has suddenly changed. A mathematical model can forecast the number of mass shootings in the future, which is critical to a risk assessment of mass shootings.

Several mathematical models could be used to model and forecast the frequency of mass shootings in the United States, but the rare-event nature and annual variability in the number of mass shootings create obstacles to generating an accurate forecast and determining which model is most appropriate. A comparison of the fitting and forecasting performance of different mathematical models will help us learn the advantages and disadvantages of each mathematical model to forecast the number of mass shootings. Comparing among the results reveals more general insights into the usefulness of each mathematical model for forecasting rare events.

Since the majority of locations in the United States have not experienced a mass shooting, relying purely on the historical data of mass shootings may not provide an accurate picture of the risk of mass shootings. We cannot know whether a location will be one of the unlucky locations where an individual engages in a shooting event. Hence, the risk quantification of mass shootings at a specific location is useful to help understand the likelihood of such an event. Different organizations could use the quantified risk of mass shootings at different locations to help them make more informed trade-off decisions between mitigating the risk of mass shootings and other priorities.

This dissertation is aimed to assess the risk (likelihood) of mass shootings at a national level and at a local level. By fitting probabilistic models to the historical data on mass shootings, this dissertation forecasts the number of mass shootings, understands to what extent the frequency of mass shootings increased, and assesses if the rate of mass shootings has suddenly changed. Comparing different forecasting models of mass shootings generates beneficial insights on selecting the best model to accurately estimate the risk of mass shootings in the United States. This dissertation uses the probabilistic forecast on the risk of mass shootings to calculate the probability that a mass shooting occurs in a state or at a specific location.

1.2 Literature Review of Mass Shootings

A precise definition of what constitutes a mass shooting is important in order to be able to collect and analyze historical data. Two terms—an active shooter and a mass shooting—are commonly seen in the literature on mass shootings. The Department of Homeland Security defines an active shooter as an individual actively engaged in killing or attempting to kill people in a confined and populated area. In many cases, there is no pattern or method to active shooters' selection of victims (DHS, 2014). The definition of a mass shooting is usually defined based on the number of fatalities or injuries (Palermo and Ross, 1999; Lankford, 2013a; Levin and Madfis, 2009). The Federal Bureau of Investigation (FBI) definition states that a mass shooting is an incident in which four or more victims are murdered by a firearm—not including the perpetrator(s)—in a single event in one or more public locations, such as a work place, school, restaurant, house of worship, neighborhood, or other public setting (Krouse and Richardson, 2015; Smart, 2018). Mass shootings as defined in this dissertation are sometimes referred to public mass shootings to distinguish between other shootings that result in four or more fatalities but occur in private places (e.g., a home) or are more targeted shootings (e.g., gang violence).

The literature on mass shootings includes both qualitative and quantitative analyses. Qualitative analysis of mass shootings can be divided into several areas. Perhaps, the majority of this literature seeks to identify factors in or causes of mass shootings. Legislation, police response, violence, violent media, bullying, mental illness, and access to weapons may all be contributing factors in the presence of more mass shootings (Sherburne, 2003; Kelly, 2012a; Newman and Fox, 2009; Browne and Hamilton-Giachritsis, 2005; Lee, 2013; Metzl and MacLeish, 2015; Lemieux, 2014). Another area of research is a descriptive analysis of mass shootings. This research focuses on analyzing mass shootings by creating shooting profiles. The shooter's profile, the number of shooters, and police response time are described in a shooting profile. A shooting profile attempts to include as many details as possible for each mass shooting (Blair et al., 2014; Bagalman et al., 2013; Duwe, 2014). The last area of qualitative analysis is using typology to understand and classify mass shootings. Researchers continue to develop new typologies including different causes for suicide bombers, police killers, and school shootings to categorize mass shootings (Holmes and Holmes, 2001; Dietz, 1986; Boyd Sr, 2011).

Existing research on mass shootings contributes to a better understanding of the causes of individual mass shootings and the effects of improving laws and police response strategies to reduce the risk of mass shootings (Sherburne, 2003; Blair et al., 2014; Skeem and Mulvey, 2020). Particular strategies and policies are recommended to prevent and reduce the risk of mass shootings (Freilich et al., 2020; Nagin et al., 2020). Legislation, police response, violence, violent media, bullying, mental illness, and access to weapons may all be contributing factors in the presence of more or fewer mass shootings (Sherburne, 2003; Kelly, 2012a; Newman and Fox, 2009; Browne and Hamilton-Giachritsis, 2005; Lee, 2013; Metzl and MacLeish, 2015; Lemieux, 2014).

The existing quantitative literature on mass shootings usually tests for statistical significance and uses regression models. The tests for statistical significance compare different types of shooters to understand if the differences in genders or age of shooters is statistically significant for the analysis of mass shootings. Although the difference between genders may not be statistically significant, younger shooters are more likely to kill or injure people (Lankford, 2013b, 2016). Regression analysis shows that the number of weapons, the number of fatalities, and the location are statistically significant predictors of whether the perpetrator lives or dies (Lankford, 2015). The effect of gun culture and firearm laws on gun violence and mass shootings is examined using multiple regression models. The quantitative results show that all these factors have critical impacts on the number of fatalities (Lemieux, 2014). Serious illness rate, poverty percentage, and gun law permissiveness are linked with state-level mass shootings rate predicted by a Bayesian zero-inflated Poisson regression model (Lin et al., 2018). Simple linear regression models the trend of mass shootings and tests if the 1994–2004 federal assault weapons ban impacted this trend (DiMaggio et al., 2019). Some other machine learning methods are used to estimate the risk of rare violent crimes, but the methods are used in a highly provisional manner. It is not clear whether those methods can be used to forecast mass shootings (Berk and Sorenson, 2020; Reid and Beauregard, 2020).

The goal of contagion studies is related to the goals of this research. Contagion studies attempt to predict when future mass shootings will occur based on historical data and its probability. A significant increase in criminal activity occurs after several severe shooting incidents (Berkowitz and Macaulay, 1971). The earliest contagion studies applied a least-squares regression model, but these models do not sufficiently account for auto-correlation in the time series (Baron and Reiss, 1985; Phillips, 1983). The Cochrane-Orcutt procedure explicitly models the auto-correlation (Stack, 1989). The Poisson-based model of contagion assumes the frequency of mass shootings follows a Poisson distribution, and the probability of a future event is governed by the transition rate parameter λ (Hamilton and Hamilton, 1983; Coleman et al., 1964). In recent years, the series hazard model, based on a modified Cox proportional hazards model, has been introduced to temporal contagion studies (Dugan et al., 2005; Dugan, 2011; LaFree et al., 2009). This relatively new statistical method estimates the risk of future incidents based on past incidents and other relevant policies. An increase in past incidents triggers a higher risk of future incidents.

Much of the existing literature on mass shootings focuses on understanding and quantifying the factors that influence the prevalence of mass shootings, such as the demographic characteristics of the shooter, gun law legislation, poverty, gun ownership, and population (Sherburne, 2003; Blair et al., 2014; Boyd Sr, 2011; Duwe, 2020; Lin et al., 2018; DiMaggio et al., 2019; Webster et al., 2020; Fridel, 2020). More recently, some research has focused more on the trend in the frequency and severity in mass shootings in the United States. The lethality of mass shootings has increased in the 21st century (Densley and Peterson, 2019b; Lankford and Silver, 2020). Some research indicates that the number of mass shootings varies although the overall trend seems to be increasing (Blair and Schwieit, 2014; Duwe, 2020; Densley and Peterson, 2019a). If the annual number of mass shootings has increased, as the data seem to indicate, understanding if the increase has remained constant over time or if there are discontinuities or sudden changes in the rate of increase can provide insight into whether something has fundamentally changed in the frequency of mass shootings. Therefore, quantifying the risk of mass shootings, analyzing to what extent that risk has changed, forecasting how the risk might change, and decomposing the risk to focus on specific locations all represent important contributions to the current literature on mass shootings.

1.3 Database of Mass Shootings

Different sources collect data on mass shootings in the United States. The commonly used mass shootings data sources are the New York City Police Department (NYCPD) (Kelly, 2012a; O'Neill et al., 2016), the (FBI, 2016), Mother Jones (Follman et al., 2012), the Gun Violence Archive (2020aa), and the Violence Project (2020). Table 1.1 provides a comparison among these different databases.

Database	**Period**	**Total number of mass shooting incidents**	**Total number of mass shooting incidents in common period**
NYCPD	1966-2016	82	11 (2014-2016)
FBI	2001-2017	59	11 (2014-2016)
Mother Jones	1982-2018	107	17 (2014-2016)
Gun Archive	2014-2018	124	75 (2014-2016)
Violence Project	1966-2020	173	14 (2014-2016)

Table 1.1: Comparison between different mass shootings data sources

Since each database covers a different time period, comparing the number of mass shootings that each database records over a common set of years demonstrates how these databases differ. A common period recorded by four data source is from 2014 to 2016. During this time, the NYCPD and the FBI recorded the same number of mass shootings. Before 2013, Mother Jones used the same definition as the NYCPD and the FBI (i.e, 4 or more fatalities). After 2013, Mother Jones includes some shootings in which fewer than 4 victims died. If we cull the mass shootings data between 2014 and 2016 based on the FBI definition, the number of mass shootings in the Mother Jones database is equivalent to the NYCPD and FBI databases. The Gun Violence Archive has the broadest definition of a mass shooting and defines a mass shooting as four or move victims rather than limiting it to fatalities(Gun Violence Archive, 2020b). The Gun Violence Archive includes shootings that occur in private places such as an individual who shoots family members in a house. These types of shootings are not considered within the FBI definition of mass shootings. The Violence Project's definition of mass shootings is identical to that of the FBI, but as Table 1.1 indicates, the Violence Project identifies a few more mass shootings between 2014 and 2016 than the NYCPD and FBI although its definition is consistent with the FBI definition.

One of the model types used to estimate the number of mass shootings in the United States—change-point models—assumes that mass shootings is a non-homogeneous Poisson process (NHPP). These models require the time between each incident in a unit of time as small as possible. The FBI and NYCPD databases provide the month but not the day of the shooting. The Mother Jones and the Violence Project databases provide the day of each shooting. The Violence Project also provides the longest observation period. Given these reasons, this research uses the mass shootings data from the Violence Project for its analysis.

1.4 Outline and Contributions

This dissertation assesses and analyzes the risk (likelihood) of mass shootings at a national level and at a local level. The analysis is spilt into answering three main questions: (i) Has the rate at which mass shootings occur in the United States suddenly changed, and what can we learn about the risk of mass shootings based on that analysis? (ii) What is the best way to forecast the annual number of mass shootings in the United States? (iii) How can the risk of mass shootings at specific locations (e.g., a school, a house of worship) be quantified? Although some literature has addressed the first two questions, no research to my knowledge has addressed the third question. The manner in which this dissertation answers the first two questions represents a unique way to analyze mass shootings in the United States, and the results from that analysis are used to answer the the third question.

Statistical models can reflect the historical trend and help answer if the trend has been increasing steadily and forecast the number of mass shootings in the future. Chapter 2 fits the historical data on mass shootings in the United States to a Poisson regression model and to six different change-point detection models. The change point detects when the rate of mass shootings changes. The change-point models can be divided into two general categories. One category of models assumes the rate of mass shootings remains constant before the change point and remains constant after the change point. The other category of models assumes the rate of mass shootings is a function of time. Six different types of rate functions are explored.

Generating a probabilistic forecast for the annual number of mass shootings in the United States represents an important contribution to the field of risk analysis and our understanding of mass shootings. The Poisson regression model and the change-point detection models are compared according to different metrics. The use of change-point detection models determines to what extent the rate of mass shooting has changed. According to the change-point model that best fits the data, the model with two change points is indistinguishable from the model with no change points. The important insight is that the rate of mass shootings in the United States has steadily increased over time but no sudden change has occurred in this rate. The forecasts from the models exhibit substantial uncertainty. Although these models cannot tell us if any year will result in only a couple of mass shootings or in several mass shootings, we can use probabilities to quantify the likelihood of each number of mass shootings. This probabilistic forecast is an important tenet of quantitative risk assessments.

Chapter 3 compares three forecasting models, a change-point model, a time series model, and a hybrid of a time series model with an artificial neural network model. Each model is applied to forecast the frequency of mass shootings. Comparing among results from these models reveals advantages and disadvantages of each model when forecasting rare events such as mass shootings. The insights generated from the comparison are beneficial for selecting the best forecasting model of the number of mass shootings and accurately estimating the risk of mass shootings in the United States.

The contribution of this chapter leads to a deeper understanding of the different forecasting models when applied to a time series of rare events such as mass shooting. Comparing among these models separate the historical data on mass shootings into different training and testing sets while preserving the time series of the data. The hybrid model is relatively new, and the model tries to capture the annual variation in mass shootings. The variation does not follow a consistent pattern. The more the hybrid model is tuned to fit the variation in the training set, the worse the model performs on the testing set. The mean of the change-point model and the time series model exhibit much more less annual variation and are not influenced as much by the inclusion of a new observation.

Chapter 4 proposes a scientifically rigorous and evidence-based method to quantify the risk of mass shootings for a state or a specific location like a high school. Two different models—one model that assumes the likelihood is proportional to a state's population and a zero-inflated Poisson regression model that includes several independent variables—are used to calculate the risk of mass shootings for each state. The simulated probability of a mass shooting in a state and the historical percentage of mass shootings in a type of location are combined to calculate the probability a mass shooting occurs at a specific location in that state. Quantifying the risk at different locations can benefit local governments, education administrators, and business owners who need to consider the threat of a mass shooting and protect against one.

The hierarchical structure of this model represents an important advancement in conducting probabilistic risk analysis for mass shootings. A probability distribution over the number of mass shootings in the nation is decomposed via two ways to estimate the conditional probability that a state experiences a mass shooting given that a mass shooting occurs in the United States. One of these models is a zero-inflated Poisson regression model. The simulation that assigns a shooting at the national level to one of the 50 states requires a method to use both the zero-inflated or binomial part and the Poisson part of the model within the simulation. The state probabilities are used to calculate the risk of a mass shooting at specific locations. The database categorizes the locations of mass shootings into nine different types, and the research uses those percentages with the state probabilities of a mass shooting to generate the risk for a specific location. The research compares the probabilities based on different assumptions (assuming each location within a category has equal risk or assuming the risk at each location within a category is proportional to the number of people).

As discussed in Chapter 5, the risk quantification process at the national level, the state level, and at specific locations can help decision makers determine how much protecting against mass shootings should be prioritized versus other risks and needs. Although the nation can expect about a half-dozen mass shootings per year, the probability of a mass shooting in any state is relatively low and the the probability of a mass shooting in a specific location is very small. Decision makers should probably plan for these very severe and horrific events while keeping in mind the very low probabilities. This dissertation provides an important piece of information to help the decision makers do just that.

CHAPTER 2. MODELING AND FORECASTING MASS SHOOTINGS USING POISSON REGRESSION AND CHANGE-POINT MODELS

2.1 Introduction

Mass shootings are particularly horrific events and frequently grab the news cycle for multiple days when they occur. For example, on October 1, 2017, a gunman killed 58 people and injured 546 others during a country music festival in Las Vegas, Nevada (Dolliver and Kearns, 2019). On August 3, 2019, 22 people were killed and 24 people were injured at a Walmart store in El Paso, Texas (Statista, 2020). Although fatalities from mass shootings represent less than 0.5% of all firearm-related fatalities (UC Davis Health, 2018), mass shootings create more public concern and anger than other types of shootings. Mass shootings also drive a lot of debate over gun policy in the United States. Mass shootings frequently result in increased proposed legislation on firearms (Webster, 2017; Luca et al., 2020; Lankford and Silver, 2020).

Mass shootings appear to occur with greater frequency in the United States, which is worrisome. Some studies indicate the rate of mass shootings has tripled since 2011 (Cohen et al., 2014). The number of mass shootings in the United States from 1966-2012 is approximately one-third of the total public mass shootings in the world (Christensen, 2017). Mass shootings not only take the lives of innocent Americans but also leave lasting psychological impacts on the survivors and the families of victims.

Mass shootings are a serious problem in the United States leading to scores of fatalities every year and generate significant debate about the best ways to reduce the occurrence and fatalities from mass shootings. The definitions for mass shootings are usually defined based on the number of fatalities or injuries (Palermo and Ross, 1999; Lankford, 2013a; Levin and Madfis, 2009). This research follows the Federal Bureau of Investigation (FBI) definition that a mass shooting is an incident in which four or more victims are murdered by a firearm—not including the perpetrator(s)—in a single event in one or more public locations, such as a work place, school, restaurant, house of worship, neighborhood, or other public setting (Krouse and Richardson, 2015; Smart, 2018). This type of shooting may also be called a mass public shooting in the literature. The media attention generated by mass shootings emphasize the tragic nature of these horrific events.

Risk management is a discipline designed to assess and evaluate among strategies to mitigate risky events such as mass shootings. A principle of risk management is that it is first necessary to provide a quantitative assessment of the risk in order to properly determine how to best mitigate the risk (Hubbard, 2009; Kaplan and Garrick, 1981). A quantitative assessment of the risk is composed of the probability of the event and the consequences if the event occurs. In order to understand how to best mitigate the risk of mass shootings, it is necessary to first quantify the risk of mass shootings in the United States.

A risk assessment of mass shootings in the United States requires forecasting or estimating the frequency of mass shootings in the United States. Since the frequency in mass shootings seems to be increasing, modeling the historical trend can provide insight into how the risk of mass shootings may increase in the future. Forecasting the annual number of mass shootings in the future can help us understand the risk of a mass shooting in the United States and lead to better safety education and law enforcement.

Analyzing to what extent the annual rate of mass shootings is increasing over time is more difficult than simply studying the raw data on the number of mass shootings that occur every year. The annual number of mass shootings varies although the overall trend seems to be increasing (Blair and Schwieit, 2014; Duwe, 2020; Densley and Peterson, 2019a). If the annual number of mass shootings has increased—as the data seem to indicate—understanding if the increase has remained constant over time or if there are discontinuities or sudden changes in the rate of increase can provide insight into whether something has fundamentally changed in the frequency of mass shootings. A mathematical model can help answer to what extent the rate of mass shootings is increasing and whether the rate has steadily increased or has suddenly changed. A mathematical model can forecast the number of mass shootings in the future, which is critical to a risk assessment of mass shootings. If the frequency of mass shootings suddenly changed in recent years, then new policies might be necessary to reduce a suddenly more dangerous risk. More informed policy prescriptions, especially at the national level, can be enacted if a careful statistical analysis can present a clearer picture of the risk of mass shootings and what that risk might look like in future years.

The goal of this research is two-fold: (1) forecasting the annual number of mass shootings in the United States by fitting probabilistic models to the historical data; and (2) understanding to what extent the rate of mass shootings is increasing and whether the rate has suddenly changed. This research uses statistical models of the historical trend to answer these questions. Since the annual number of mass shootings represents count data, a Poisson regression model is fit to the historical data and provides a base-case mathematical model for comparison purposes. The research uses a Bayesian method to detect potential change points where the rate or frequency of mass shootings has changed. One type of change-point model assumes the rate of mass shootings remains constant before the change point and remains constant after the change point, and the change point identifies when the constant rate changes. The second type of change-point model assumes the rate of mass shootings is a function of time. Since the rate of mass shootings in these models is a function of time, the change point identifies when the rate suddenly changes because function parameters change. The Poisson regression model, the constant-rate change-point model, and the time-dependent change-point model are used to generate probabilistic forecasts of the number of mass shootings in the United States. The research compares and contrasts and among these models and derives insights about the risk of mass shootings based on these models.

As described in Section 2.2, a review of the literature reveals that number of mass shootings has increased, but little attempt has been made to forecast the number of mass shootings in the future. Section 2.3 introduces the Poisson regression and the different change-point models and explains why these might be appropriate models to forecast count data such as the number of mass shootings. The results of fitting the data on mass shootings to these models are presented in Section 2.4. The results are compared according to three different metrics, and the three best-performing models are used to forecast the number of mass shootings in 2020 and 2021. The forecast for 2020 is compared to the actual number of mass shootings in 2020. Concluding comments appear in Section 2.5, and we use the best-performing models to address the two goals of this research. We outline how this modeling and forecasting can help inform better policies about how to mitigate the risk of mass shootings in the United States.

2.2 Literature Review

2.2.1 Mass shootings

Existing research on mass shootings contributes to a better understanding of the causes of individual mass shootings and the effects of improving laws and police response strategies to reduce the risk of mass shootings (Sherburne, 2003; Blair et al., 2014; Skeem and Mulvey, 2020). Particular strategies and policies are recommended to prevent and reduce the risk of mass shootings (Freilich et al., 2020; Nagin et al., 2020). Legislation, police response, violence, violent media, bullying, mental illness, and access to weapons may all be contributing factors in the presence of more or fewer mass shootings (Sherburne, 2003; Kelly, 2012a; Newman and Fox, 2009; Browne and Hamilton-Giachritsis, 2005; Lee, 2013; Metzl and MacLeish, 2015; Lemieux, 2014).

Research in mass shootings has generally concluded that the frequency of mass shootings has increased since about the mid-2000s even when controlling for population growth. The number of active shooter incidents has increased from an average of 6.7 annual incidents in the early 2000s to 16.4 annual incidents in 2007-2013 (Blair and Schwieit, 2014). Duwe finds that mass shootings became more frequent in about 2008 based on a 5-year moving average of the number of mass shootings per 100 million U.S. residents (Duwe, 2020). The increase in frequency may be less revealing than the apparent increase in the deadliness of mass shootings. The annual number of mass shootings as well as the severity, as measured by the number of fatalities, has increased from 1966 to 2018 (Densley and Peterson, 2019a). The decade of the 2010s has seen a sharp increase in the number and percentage of mass shootings that result in high fatalities (e.g., 8 or more victims, 12 or more victims, and 16 or more victims) (Lankford and Silver, 2020). DiMaggio et al. (2019)

use linear regression to conclude that the federal assault weapons ban from 1999-2004 resulted in fewer fatalities from mass shootings during that time period.

The quantitative literature on mass shootings frequently uses tests for statistical significance and regression models. The tests for statistical significance compare different types of shooters to understand if differences in the genders or ages of shooters are statistically significant. Although the difference between genders may not be statistically significant, younger shooters are more likely to kill or injure people(Lankford, 2013b, 2016). Regression analysis shows that the number of weapons, the number of fatalities, and the location are statistically significant predictors for mass shooters who die or live (Lankford, 2015). The effect of gun culture and firearm laws on gun violence and mass shootings is examined using multiple regression models. The quantitative results show that all these factors have critical impacts on the number of fatalities (Lemieux, 2014). The rate of serious illness, poverty, and gun law permissiveness are correlated with the frequency of state-level mass shootings according to a Bayesian zero-inflated Poisson regression model (Lin et al., 2018). Berk and Sorenson, Reid and Beauregard use machine learning methods (e.g., gradient boosting, clustering, binomial and logistic regression) to predict the risk of violent rare events, but it is not clear if these methods can be used to forecast mass shootings (Berk and Sorenson, 2020; Reid and Beauregard, 2020).

The goal of contagion studies is perhaps the most similar to the goal of this research. Contagion studies attempt to predict when future mass shootings will occur based on historical data and its probability. A significant increase in criminal activity occurs after several severe shooting incidents (Berkowitz and Macaulay, 1971). The earliest contagion studies applied a least-squares regression model, but these models do not sufficiently account for auto-correlation in the time series (Baron and Reiss, 1985; Phillips, 1983). The Cochrane-Orcutt procedure explicity models the auto-correlation (Stack, 1989). The Poisson-based model of contagion assumes the frequency of mass shootings follows a Poisson distribution, and the probability of a future event is governed by the transition rate parameter λ (Hamilton and Hamilton, 1983; Coleman et al., 1964). In recent years, the series hazard model, based on a modified Cox proportional hazards model, has been introduced to temporal contagion studies (Dugan et al., 2005; Dugan, 2011; LaFree et al., 2009). This relatively new statistical method estimates the risk of future incidents based on past incidents and other relevant policies. More past incidents trigger a higher risk of future incidents.

This research contributes to the quantitative analysis of mass shootings by using a change-point model to capture the trend in the frequency of mass shootings. Change-point detection analysis usually considers two types of rate functions, which describe how the rate or average frequency changes over time: a constant rate (Achcar et al., 2007) and a time-dependent rate (Muss et al., 1987). A Poisson process with a non-constant rate is defined as a non-homogeneous Poisson process (NHPP). Within the existing literature, the power law process (PLP), the Musa-Okumoto process (MO), the Goel–Okumoto process (GO), the generalized Goel-Okumoto process (GGO), and the Weibull-geometric process (WG) are commonly used time-dependent rate functions (Mudholkar et al., 1995; Musa and Okumoto, 1984; Goel and Okumoto, 1978; Goel, 1985; Barreto-Souza et al., 2011).

2.2.2 Change-point models

The methods used to estimate the parameters in change-point models include both supervised methods and unsupervised methods(Aminikhanghahi and Cook, 2017). Supervised methods use a variety of classification methods, such as decision trees (Reddy et al., 2010), Bayes nets (Cleland et al., 2014), support vector machines (Zheng et al., 2008), nearest neighbor approaches (Wei and Keogh, 2006), hidden Markov models (Han et al., 2012), and conditional random field and Gaussian mixture models (Zheng et al., 2008). Supervised methods treat change-point detection as a state boundary detection for labeled data. Once classifiers classify different states, each state boundary (change point) can be found. Unsupervised methods are for unlabeled data. Unsupervised methods detect change points by dividing the time series into several consecutive time intervals. The time points connecting different segments are change points. Unsupervised methods include likelihood ratio (Aue et al., 2009; Li et al., 2020b), subspace models (Yamanishi and Takeuchi, 2002), probabilistic methods (Alippi et al., 2015; Li et al., 2021), kernel based methods (Feuz et al., 2014), graph based methods, and clustering (Rosenbaum, 2005; Zhang and Small, 2006). A Bayesian approach is an unsupervised method that detects when the change point occurs based on a probabilistic approach (Raftery and Akman, 1986). The Bayesian change-point methods have been applied to various change points detection problems, such as ozone measurements in Mexico City (Achcar et al., 2010; Cruz-Juárez et al., 2016) and the risk analysis of teenage drivers (Li, 2015; Li et al., 2018, 2017, 2020a).

This research models the frequency of mass shootings as a Poisson process. A Poisson regression model treats the annual number as count data and assumes the logarithm of the mean value is a linear function of time. A Bayesian approach is used to identify change points in a time series, and we compare the performance of the constant rate function, five time-dependent rate functions, and the Poisson regression model. Although this research does not explicitly model the contagion between two adjacent mass shootings, the time-dependent rate functions and the inclusion of change points should capture if one mass shooting leads to more mass shootings because the overall frequency of mass shootings would increase in the historical data. The posterior distributions of the parameters in the change-point models and the best-fit Poisson regression model are used to forecast the number of mass shootings in 2020 and 2021, and we compare the forecast with the actual number of mass shootings in 2020.

2.3 Forecasting Models

Mass shootings are rare events and represent less than one-half percent of gun homicides in the United States. Modeling the frequency of rare events can be challenging (Bier and Yi, 1995; Guns and Vanacker, 2012; Theofilatos et al., 2016). The Poisson distribution may be a good model for rare events because of its ability to model individual counts (0, 1, 2, ...) per a given time period (Moksony and Hegedűs, 2015; Thompson, 2001; Griffith and Haining, 2006). The change-point models estimate the time between events. Since the measurement of time could be measured in different units (e.g., minutes, days, years), modeling the time between events overcomes some of the challenges of modeling the frequency of rare events. Achcar et al. fit a change-point model to mine accidents in England. Their data set includes 109 mine accidents over 76 years, which occurs much less frequently than mass shootings in the United States (Achcar et al., 2007).

2.3.1 Poisson Regression

Since the annual number of mass shootings is relatively small (8 or fewer shootings in year), an ordinary least squares regression model would likely violate the assumption of constant variance. A Poisson regression treats the annual number of shootings as count data, where the count data Y follows a Poisson distribution with the mean μ. The probability that $Y = y$ is:

$$P(Y = y \mid \mu) = \frac{\mu^y}{y!} e^{-\mu} \tag{2.1}$$

The Poisson regression model assumes the logarithm of μ can be modeled by a linear function of an independent variable X. In this case, the independent variable X represents the year. For a year $X = j$ $(j = 1, \ldots, J)$, the corresponding mean count of year j is:

$$\mu_j = e^{\alpha + \beta j} \tag{2.2}$$

The maximum likelihood method is used to estimate the coefficients α and β.

2.3.2 Change-point Models of NHPP

Unlike the Poisson regression model which assumes the model parameters are constant over time, change-point models identify points in time when those parameters change. A change point could indicate that something historically has changed in the rate or frequency of mass shootings in the United States.

We assume all observed data points occur in the time interval $[0, T]$ where T is the end of the study period. Let $N(t)$ represent the number of events that occur in the time interval $[0, t), \quad t \geq 0$. We assume the time intervals between successive mass shootings are independently distributed. Each event occurs at a unique time t_i $(i = 0, \ldots, n)$ where there are a total of n events and

$0 = t_0 < t_1 \cdots < t_n < T$. If there are a total of Q change points in the entire interval $[0, T]$, then the change point τ_q represents the qth time when the rate of mass shootings changes, $q \in \{1, 2, 3, \dots, Q\}$. The parameter $\lambda(t)$ is the rate function, which shows that the rate of mass shootings is a function of time t. The parameter $\lambda_q(t)$ represents the rate function of mass shootings between the change point τ_{q-1} and the change point τ_q.

A Poisson process with a non-constant rate is defined as an NHPP. The rate function for an NHPP with Q change points is:

$$\lambda(t) = \begin{cases} \lambda_1(t), & 0 \le t < \tau_1 \\ \lambda_q(t), & \tau_{q-1} \le t < \tau_q, \quad q \in \{2, 3, , Q\} \\ \lambda_{Q+1}(t), & \tau_Q \le t \le T \end{cases} \tag{2.3}$$

In this research, we compare the performance of typically used rate functions (Mudholkar et al., 1995; Musa and Okumoto, 1984; Goel and Okumoto, 1978; Goel, 1985; Barreto-Souza et al., 2011). Six rate function forms including a constant rate, PLP, MO, GO, GGO and WG are:

$$\begin{aligned}
\lambda^{(\text{Constant})}(t) &= \lambda \\
\lambda^{(PLP)}(t) &= \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1} \\
\lambda^{(MO)}(t) &= \frac{\beta}{t+\alpha} \\
\lambda^{(GO)}(t) &= \alpha\beta e^{-\beta t} \\
\lambda^{(GGO)}(t) &= \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^{\gamma}} \\
\lambda^{(WG)}(t) &= \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{1 - pe^{-\left(\frac{t}{\beta}\right)^{\alpha}}}
\end{aligned} \tag{2.4}$$

where α, β, $\gamma > 0$ and $p \in (0, 1)$ are parameters of the rate functions. The mean value function $m(t)$ of each rate function is defined as

$$m(t) = \int_0^t \lambda(s)ds, \qquad t \ge 0. \tag{2.5}$$

The corresponding mean value functions for the different rate functions are:

$$
\begin{aligned}
m^{(\text{Constant})}(t) &= \lambda t \\
m^{(PLP)}(t) &= \left(\frac{t}{\beta}\right)^{\alpha} \\
m^{(MO)}(t) &= \beta \log\left(1 + \frac{t}{\alpha}\right) \\
m^{(GO)}(t) &= \alpha[1 - e^{-\beta t}] \\
m^{(GGO)}(t) &= \alpha[1 - e^{-\beta t^{\gamma}}] \\
m^{(WG)}(t) &= -\log\left(\frac{(1-p)e^{-\left(\frac{t}{\beta}\right)^{\alpha}}}{1 - pe^{-\left(\frac{t}{\beta}\right)^{\alpha}}}\right)
\end{aligned}
\tag{2.6}
$$

The mean value function can be used to calculate the probability of k events in a time interval $[t, t+s)$ (Cruz-Juárez et al., 2016):

$$
P(N(t+s) - N(t) = k) = \frac{[m(t+s) - m(t)]^k}{k!} \exp(-[m(t+s) - m(t)]) \tag{2.7}
$$

The parameters $\boldsymbol{\theta}$ needed to be estimated for the rate functions and the mean value functions are $\boldsymbol{\theta}_{Constant} = \lambda$, $\boldsymbol{\theta}_{PLP} = \boldsymbol{\theta}_{MO} = \boldsymbol{\theta}_{GO} = (\alpha, \beta)$, $\boldsymbol{\theta}_{GGO} = (\alpha, \beta, \gamma)$ and $\boldsymbol{\theta}_{WG} = (\alpha, \beta, p)$. If Q change points exist, we also need to estimate the change points $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_Q\}$. The parameters of the rate function before and after the change point τ_q are $\boldsymbol{\theta}_q$ and $\boldsymbol{\theta}_{q+1}$, respectively. The parameters that need to be estimated in the change-point models are $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{Q+1}$, and $\boldsymbol{\tau}$.

In this research, we use a Bayesian approach to estimate the parameters of the model. The likelihood function for data $\boldsymbol{D_T} = \{n; t_1, \ldots, t_n; T\}$ with at least one change point is based on a generalization of the exponential distribution when $\lambda(t)$ is a constant. The likelihood function of

$\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}\}$ is (Cruz-Juárez et al., 2016):

$$L(\boldsymbol{D_T}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}) \propto \prod_{i=1}^{N(\tau_1)} \lambda_1(t_i) e^{(-m_1(\tau_1))} \times \left[\prod_{j=2}^{Q} \left(\prod_{i=N(\tau_{j-1})+1}^{N(\tau_j)} \lambda_j(t_i) e^{-[m_j(\tau_j)-m_j(\tau_{j-1})]} \right) \right] \\ \times \prod_{i=N(\tau_Q)+1}^{N(T)} \lambda_{Q+1}(t_i) e^{(-m_{Q+1}(T)+m_{Q+1}(\tau_Q))} \tag{2.8}$$

where $m_i(t)$ is the mean value function corresponding to $\lambda_i(t)$ as provided by Equation (2.3).

The uniform distribution will be used as the prior distribution for each parameter to represent vague prior information and also for simplicity. Then the joint prior distribution over the parameters is $P(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau})$. The joint posterior distribution of parameters is:

$$P(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}|\boldsymbol{D_T}) \propto L(\boldsymbol{D_T}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}) P(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}). \tag{2.9}$$

We assume that the change-points are independent from the rate function parameters, and the rate function parameters are conditionally independent given the change points in a change-point model. Therefore, the joint posterior distribution is rewritten as:

$$P(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}|\boldsymbol{D_T}) \propto L(\boldsymbol{D_T}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{Q+1}, \boldsymbol{\tau}) P(\boldsymbol{\theta}|\boldsymbol{\tau}) \ldots P(\boldsymbol{\theta}_{Q+1}|\boldsymbol{\tau}) P(\boldsymbol{\tau}) \tag{2.10}$$

The MCMC method is used to sample from the posterior distributions for all of the parameters. A simplified method to apply the MCMC and sample from the posterior distribution is using the software package `Stan`. The library `rstan` enables us to run `Stan` in software `R` . We need to define the priors for the parameters and the likelihood function. From Equation (2.8), the likelihood function of the change-point model is not a known distribution function. We write user-defined distributions in `Stan` (Annis et al., 2017). After the priors and likelihood functions are defined, `Stan` will generate samples from the posterior distributions.

This research fits a model for each of the six rate functions across three different sets of change points: no change point or $Q = 0$, one change point $Q = 1$, and two change points $Q = 2$. If one change point exists, the prior distribution for τ_1 is uniformly distributed in the interval $(0, T)$. As depicted in Table 2.1, we assume uniform non-informative priors for the different parameters in rate functions. Since the prior distributions remain the same for a parameter regardless of whether the parameter is before or after a change point, Table 2.1 provides the priors of parameters for the six rate functions with no change point. For example, the prior for the constant rate function with no change points for λ is $U[0, 1]$. The priors for λ_1 and λ_2 are $U[0, 1]$ with one change point. The priors for λ_1, λ_2, and λ_3 are $U[0, 1]$ with two change points.

Rate function	**Priors**
Constant	$\lambda \sim U[0, 1]$
PLP	$\alpha \sim U[0, 100]$, $\beta \sim U[0, 1000000]$
MO	$\alpha \sim U[0, 1000000]$, $\beta \sim U[0, 100]$
GO	$\alpha \sim U[0, 1000000]$, $\beta \sim U[0, 100]$
GGO	$\alpha \sim U[0, 1000000]$, $\beta \sim U[0, 100]$, $\gamma \sim U[0, 100]$
WG	$\alpha \sim U[0, 100]$, $\beta \sim U[0, 1000000]$, $p \sim U[0, 1]$

Table 2.1: Prior distributions of parameters for the different rate functions with no change point

2.3.3 Model Performance Measurement Metrics

Three performance metrics, DIC, marginal likelihood, and RSS, are used to compare among the three different change-point models for each of the six different rate functions. The DIC is based on the difference between the posterior mean of the deviance and the deviance of the posterior means. The model with the smaller DIC value fits the data better. The equation to calculate DIC is:

$$\begin{aligned} DIC &= -2 \times (L - P) \\ L &= \log(L(\boldsymbol{D_T} | \hat{\boldsymbol{\theta}}_{\mathbf{1}}, \dots, \hat{\boldsymbol{\theta}}_{\boldsymbol{Q+1}}, \hat{\boldsymbol{\tau}})) \\ P &= 2 \times [L - \frac{1}{S} \sum_{i=1}^{S} \log(L(\boldsymbol{D_T} | \boldsymbol{\theta_{1,i}}, \dots, \boldsymbol{\theta_{Q+1,i}}, \boldsymbol{\tau_i}))] \end{aligned} \tag{2.11}$$

where $L(\boldsymbol{D_T}|\hat{\boldsymbol{\theta}}_{\boldsymbol{1}}, \ldots, \hat{\boldsymbol{\theta}}_{\boldsymbol{Q+1}}, \hat{\boldsymbol{\tau}})$ is the likelihood of the data given the posterior means of parameters, based on Equation (2.8). The parameter $\hat{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$, and S is the total number of samples drawn from the posteriors. The parameter $\boldsymbol{\theta_{q,i}}$ is the vector of parameters before the qth change point for sample i. The $L(\boldsymbol{D_T}|\boldsymbol{\theta_{1,i}}, \ldots, \boldsymbol{\theta_{Q+1,i}}, \boldsymbol{\tau_i})$ is the likelihood of the data given the ith sample of the posteriors (Gelman et al., 2013).

The model with the larger marginal likelihood is more strongly supported by the data. We use the harmonic mean of the likelihood of the MCMC iterations as an approximation of the marginal likelihood (Raftery et al., 2006). It is $\hat{P}(\boldsymbol{D_T}|\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}}, \boldsymbol{\tau}) = \left[\frac{1}{S}\sum_{i=1}^{S} \frac{1}{L(\boldsymbol{D_T}|\boldsymbol{\theta_{1,i}}, \ldots, \boldsymbol{\theta_{Q+1,i}}, \boldsymbol{\tau_i})}\right]^{-1}$.

RSS calculates the difference between the real data and the estimated mean value of the annual count (Archdeacon, 1994). The estimated mean value is calculated according to the mean value function $m(t)$ from Equations (2.6). The data set in the interval $(0, T)$ spans a total of J years, and the estimated mean annual count is $\hat{C}_j = m\left(t^{(j)}\right) - m\left(t^{(j-1)}\right)$ for $j = 1, \ldots, J$. Time $t^{(j)}$ represents the last time in year j (e.g., December 31) in year j. The parameter $N(t)$ represents the number of events that occur in the time interval $[0, t)$ in the data. The real number of events in year j is $C_j = N\left(t^{(j)}\right) - N\left(t^{(j-1)}\right)$ for $j = 1, \ldots, J$. Thus, RSS $= \sum_{j=1}^{J}(C_j - \hat{C}_j)^2$.

2.4 Analysis of Mass Shootings

2.4.1 Mass Shootings Data

This research focuses on mass shootings in a public place. The exact time of occurrence is needed in order to construct the change-point detection model. This research initially considered five main data sources: the New York City Police Department (NYCPD) (Kelly, 2012b; O'Neill et al., 2016), the FBI (FBI, 2016), Mother Jones (Follman et al., 2012), the Gun Violence Archive(Gun Violence Archive, 2020a), and the Violence Project(The Violence Project, 2020). These different data sources define a mass shooting differently and record a different number of mass shootings recorded for each year. As stated previously, this research uses the FBI's definition of a mass shooting as an incident in which four or more victims are murdered by a firearm in a public place. Since the change-point model considers mass shootings as an NHPP, a database that measures the time between each incident in a unit of time as small as possible is necessary. The Violence Project database defines a mass shooting identical to the FBI's definition, contains data observed over longest time period from 1966 to 2019 (Boerboom et al., 2020), and provides the date of each shooting. For this reason, this research uses the Violence Project data. The annual count of mass shootings as provided by Table A.1 in the Appendix. Since the first recorded time is on August 1, 1966, the first year (1966) represents the number of mass shootings from August 1, 1966 to December 31, 1966.

Table A.2 in the Appendix depicts the day t_i for each recorded mass shooting. The first mass shooting recorded by Violence Project took place on August 1, 1966, which corresponds to the starting time in the model $t_1 = 0$. All other times in the table are the number of days after the first shooting.

2.4.2 Model Results

Figure 2.1 depicts the real observations and estimated count values by the Poisson regression model. The Poisson regression model estimates an exponential growth in the number of annual mass shootings. The estimated coefficients are $\alpha = -67.45$ with a 95% confidence interval of [-88.70 -46.88] and $\beta = 0.03$ with a 95% confidence interval of [0.02, 0.04]. The variance also exhibits exponential growth. A visual look at Figure 2.1 indicates that the Poisson regression model appears to capture the data well and follow the trend of the data although outliers (e.g., year 1999 with 8 mass shootings) appear to exist.

The Poisson regression model indicates that the rate of mass shootings has increased over time and the mean annual number is increasing exponentially. This exponential increase is constant over time. Employing the change-point model with different rate functions for mass shootings may provide different insights into how the rate of mass shootings changes over time and indicate whether the rate has suddenly changed. We explore the results of fitting a model with no change points, one change point, and two change points.

When we use the MCMC method to sample from the posterior distributions for the mass shooting data, we consider a burn-in sample of size 15,000 to minimize the impact of initial values.

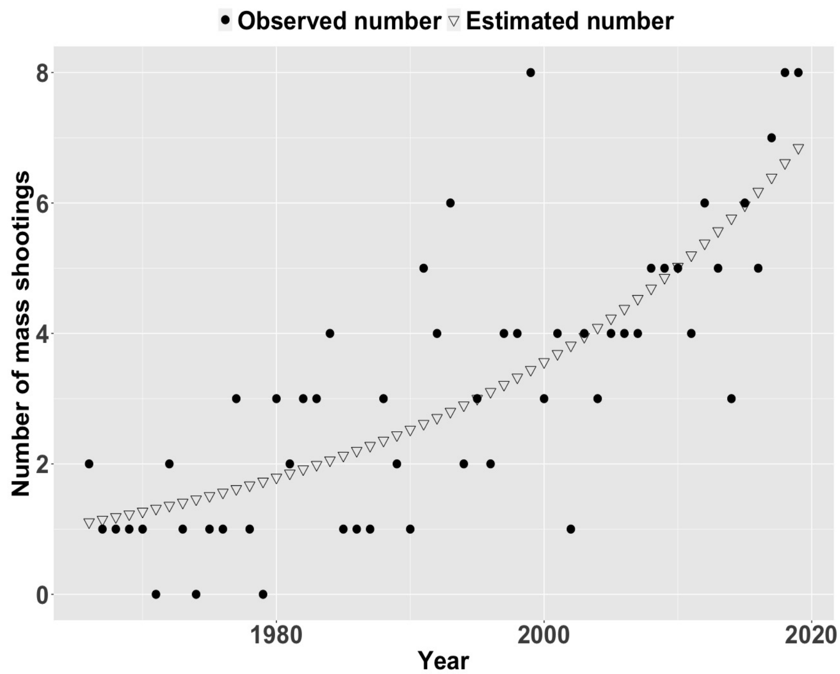
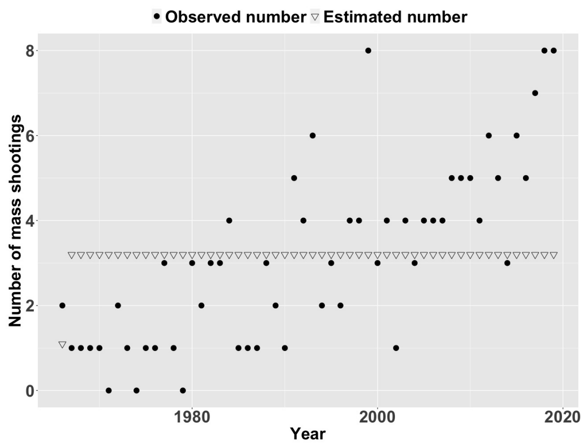


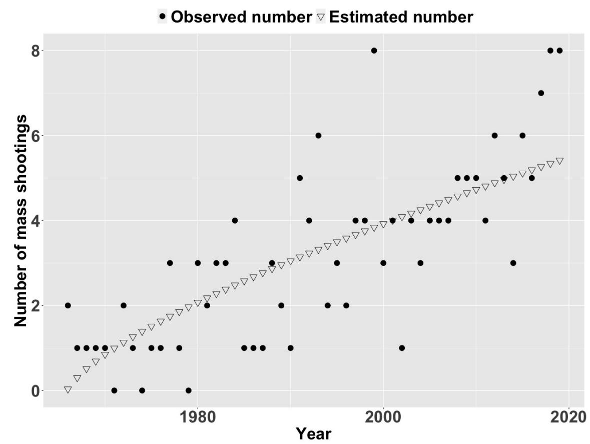
Figure 2.1: Observed value and estimated mean value from the Poisson regression model from 1966 to 2019

Tables A.3 – A.5 in the Appendix provide the posterior summaries of the parameters based on 6000 simulated samples after the burn-in phase.

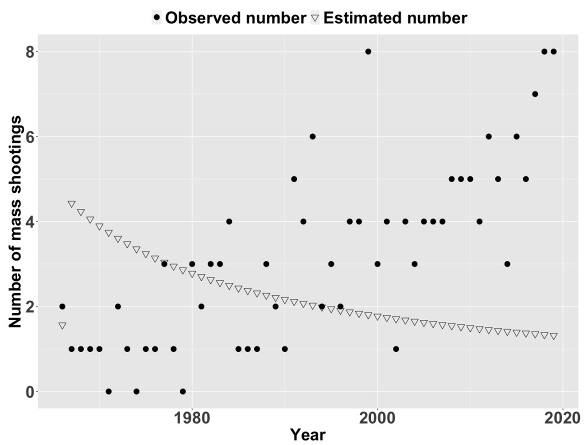
The fitted results are shown in the plot of the count data of mass shootings from 1966 to 2019. Figure 2.2 presents the observed values from the Violence Project data and the estimated mean values obtained by the models with no change point for each of the six different rate functions. Figure 2.3 shows the cumulative plots of Figure 2.2. The estimated mean values are calculated by averaging all of the samples generated from the MCMC simulations. The estimated mean rate is increasing in the PLP, GGO, and WG rate function models. The WG rate function generates the highest frequency of mass shootings. The estimated mean rate is decreasing in the MO and GO rate function models, which contradicts the observed trend of the count data. Since the PLP, GGO, and WG rate functions capture the increasing trend in annual count of mass shooting data, the cumulative estimated mean values of the PLP, GGO, and WG rate functions fit the cumulative observed values better than other rate functions in the Figure 2.3.



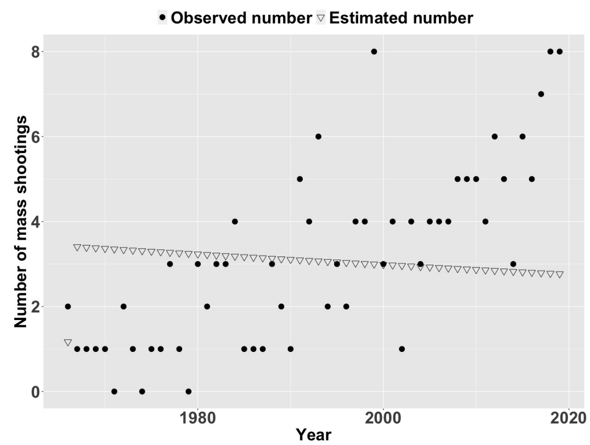
(a) Constant Rate



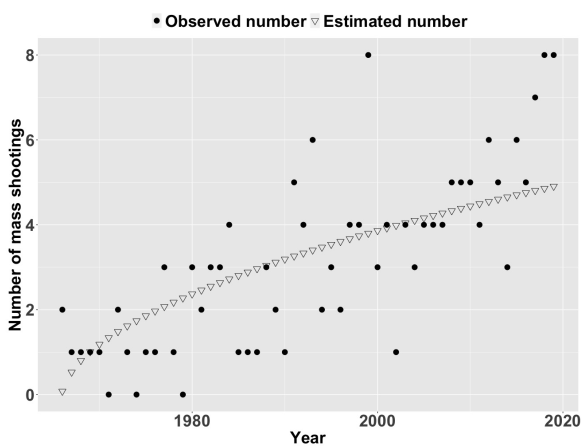
(b) PLP



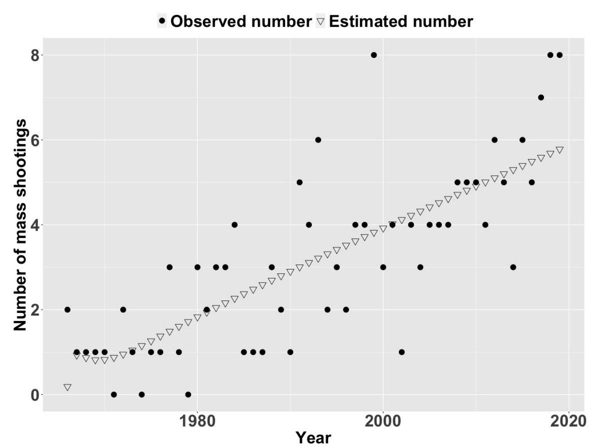
(c) MO



(d) GO

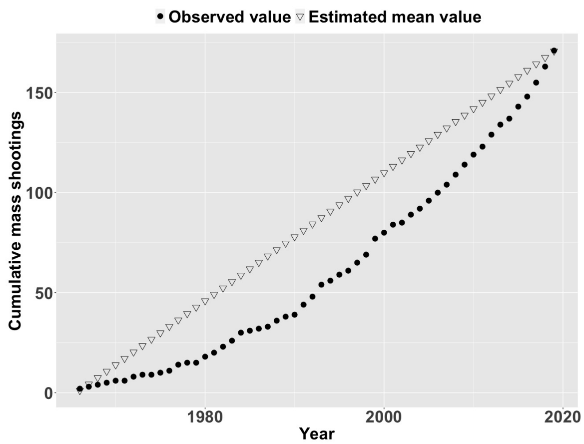


(e) GGO

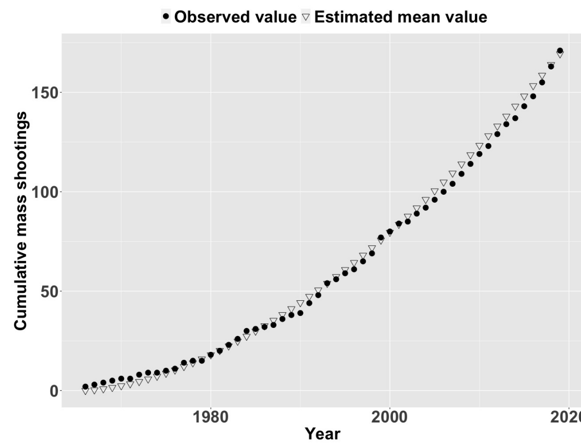


(f) WG

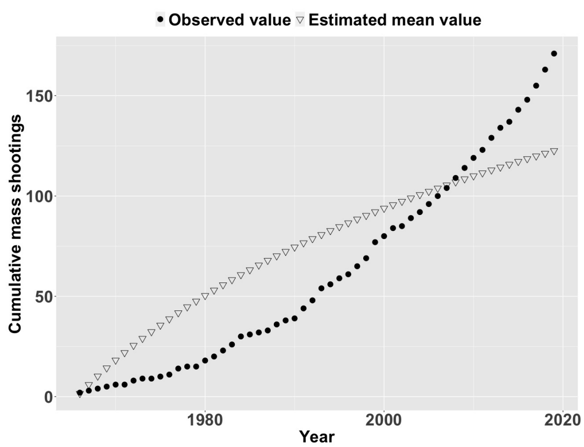
Figure 2.2: Observed value and estimated mean value from the model with no change point from 1966 to 2019



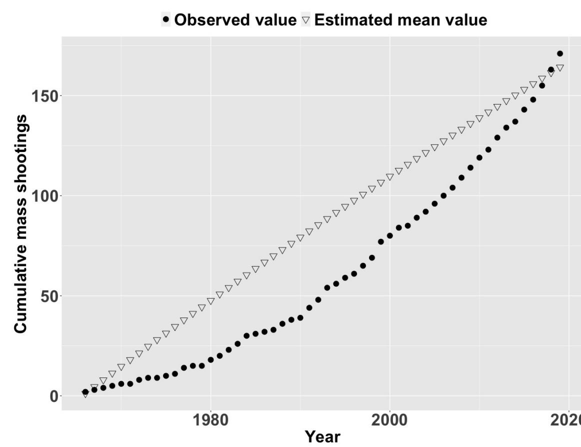
(a) Constant Rate



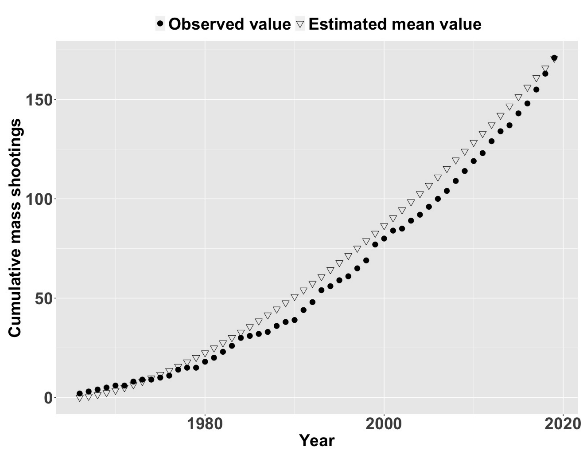
(b) PLP



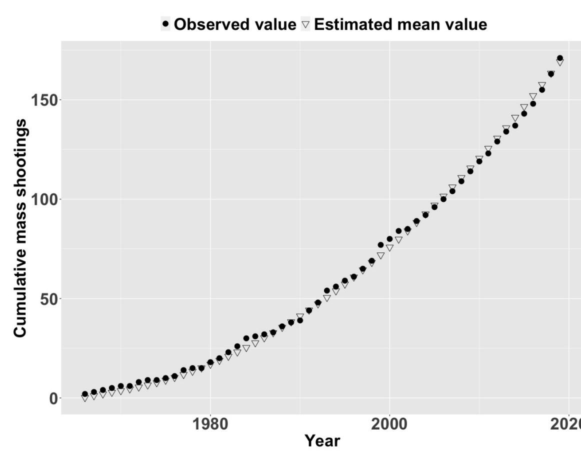
(c) MO



(d) GO



(e) GGO



(f) WG

Figure 2.3: Cumulative observed value and cumulative estimated mean value from the model with no change point from 1966 to 2019

Figures 2.4 and 2.6 present the estimated mean number of annual mass shootings obtained by the one change-point models and the two change-point models, respectively. Figure 2.7 show the cumulative plots of the two change-point models in Figure 2.6. (The cumulative plots with one change point are similar to those with two change points.) Tables 2.2 and 2.3 display the mean year at which the change point occurs and the corresponding 95% credible interval.

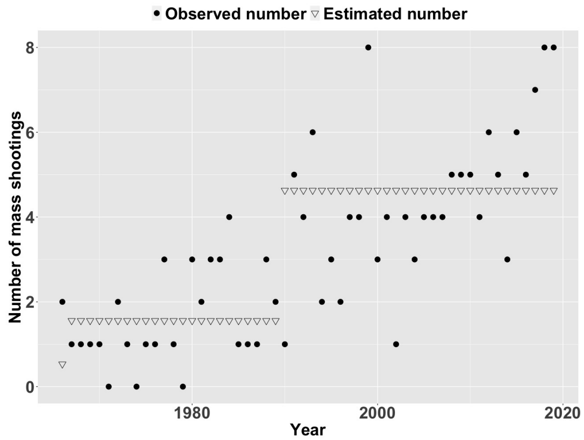
Rate function	**Change point(year)**	**95% credible interval (year)**
Constant	1990	[1988, 1991]
PLP	1967	[1966, 1968]
MO	1990	[1989, 1991]
GO	1990	[1990, 1991]
GGO	1967	[1966, 1968]
WG	1966	[1966, 1966]

Table 2.2: Change point estimates for the one-change-point model

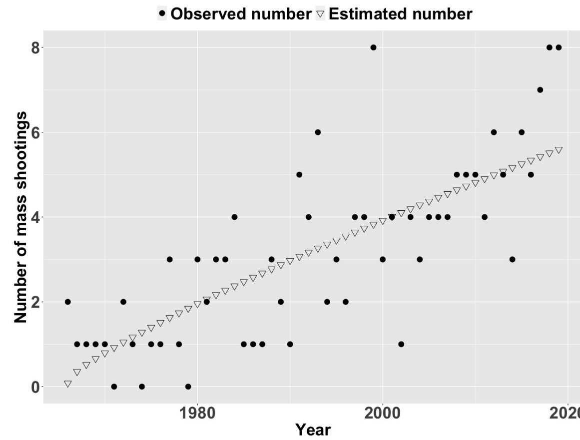
Rate function	**First change point (year)**	**95% credible interval (year)**	**Second change point (year)**	**95% credible interval (year)**
Constant	1991	[1989, 1993]	2015	[2012, 2018]
PLP	1966	[1966, 1967]	1967	[1967, 1968]
MO	1967	[1966, 1968]	1991	[1990, 1992]
GO	1966	[1966, 1967]	1967	[1967, 1968]
GGO	1966	[1966, 1967]	1966	[1966, 1967]
WG	1966	[1966, 1966]	1966	[1966, 1966]

Table 2.3: Change point estimates for the two-change-point model

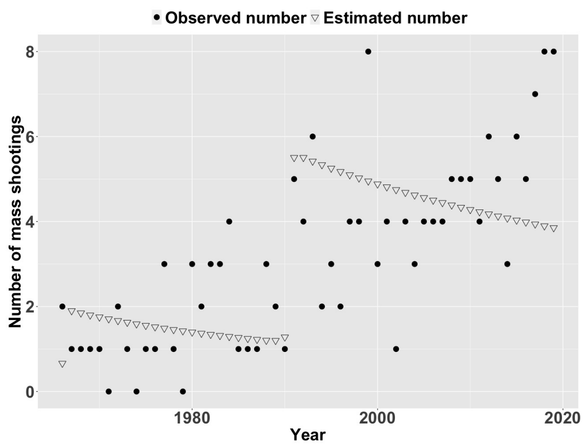
The PLP, GGO, and the WG rate functions for the one change-point model detect that a change point occurs in 1967, or approximately within the year of the first mass shooting. The two change-point models for these rate functions find that both change points occur on average in 1967 or earlier. The constant rate, MO, and GO rate functions detect a change point around 1990 with the one change-point model, which is approximately 24 years after the first mass shooting in the database. The two change-point model with the constant rate function detects change points in 1991 and 2015, and the MO rate function detects change points in 1967 and 1991. The GO rate function detects change points in 1966 and 1967.



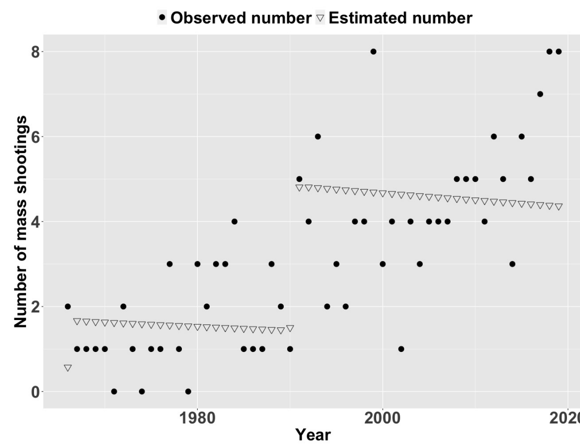
(a) Constant Rate



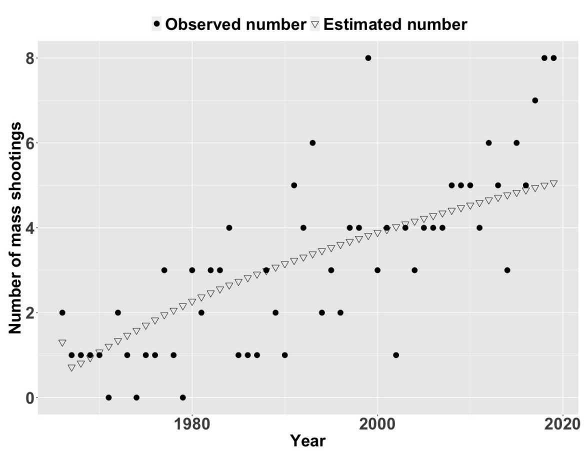
(b) PLP



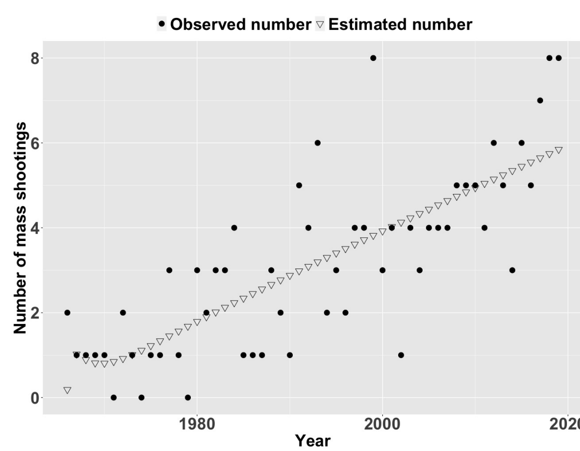
(c) MO



(d) GO

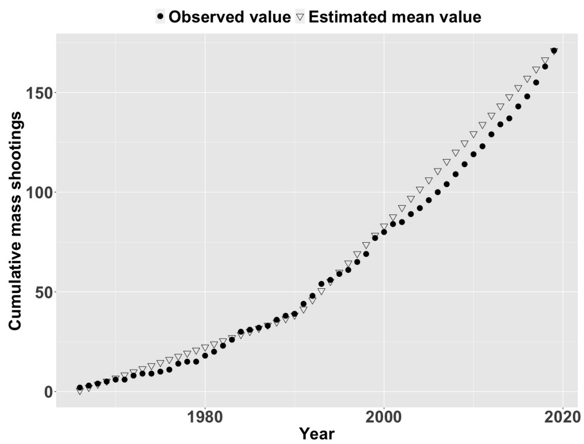


(e) GGO

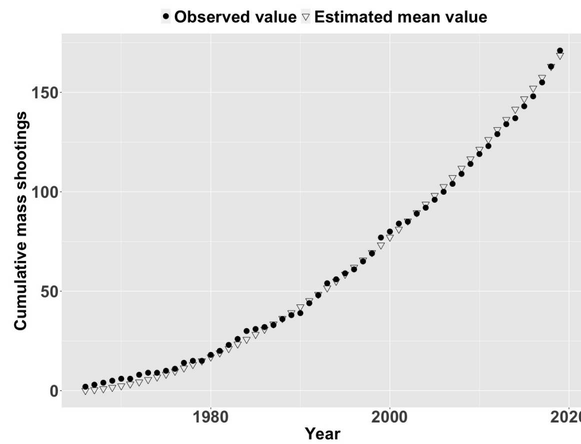


(f) WG

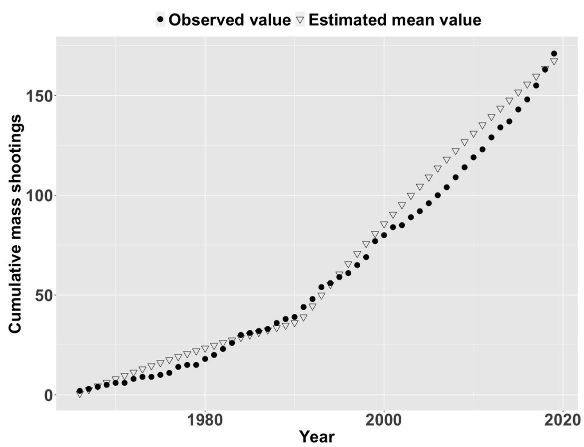
Figure 2.4: Observed value and estimated mean value from the model with one change point from 1966 to 2019



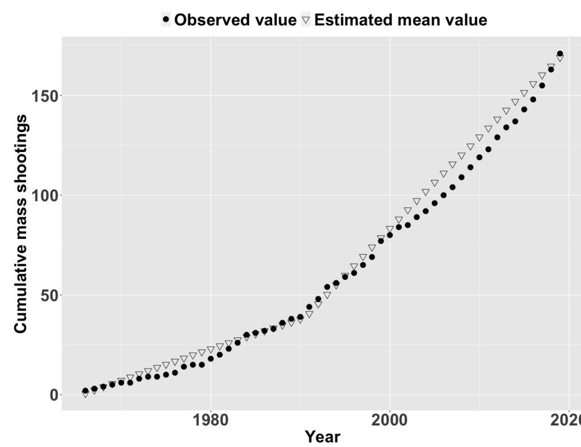
(a) Constant Rate



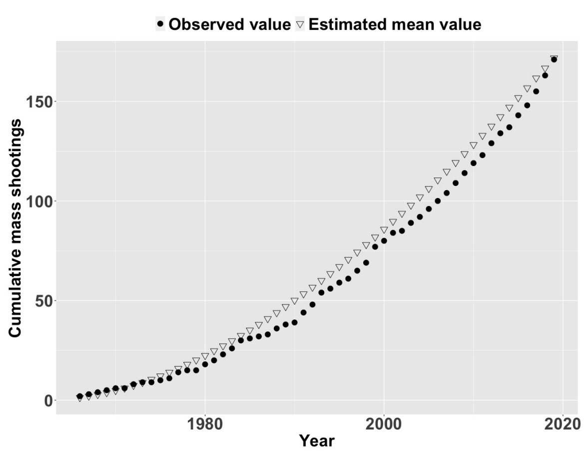
(b) PLP



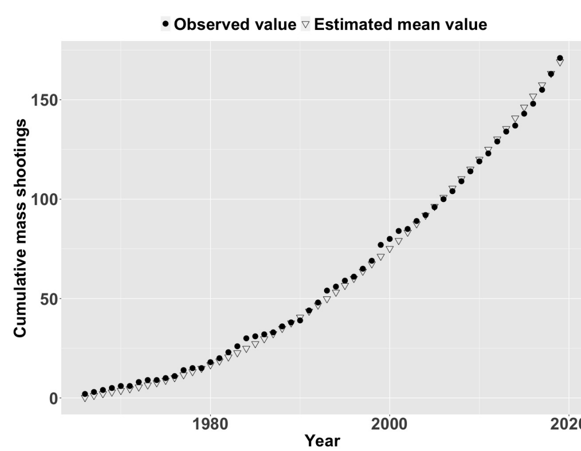
(c) MO



(d) GO

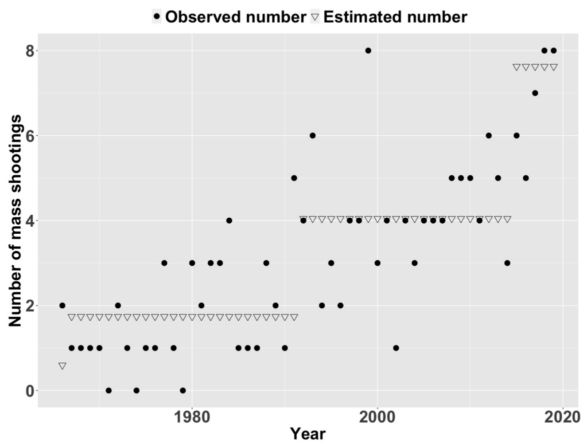


(e) GGO

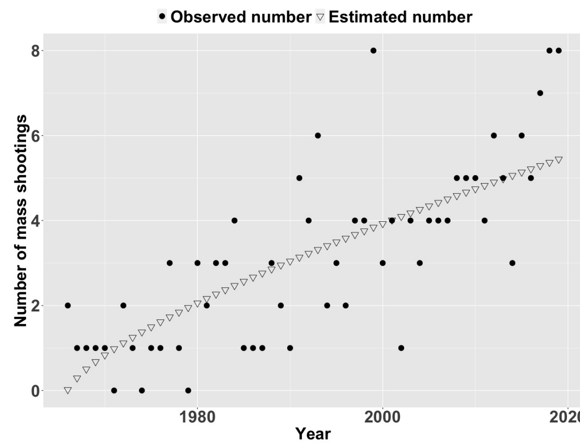


(f) WG

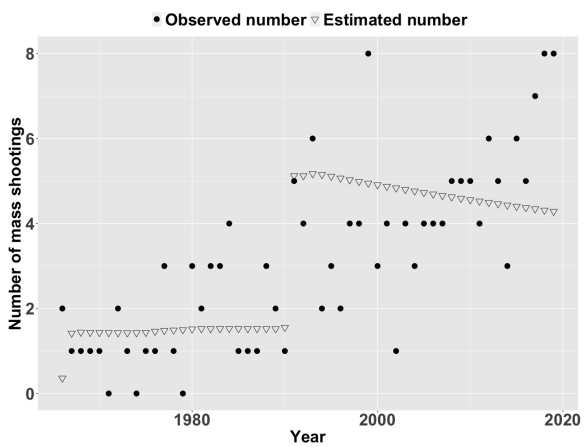
Figure 2.5: Cumulative observed value and cumulative estimated mean value from the model with one change point from 1966 to 2019



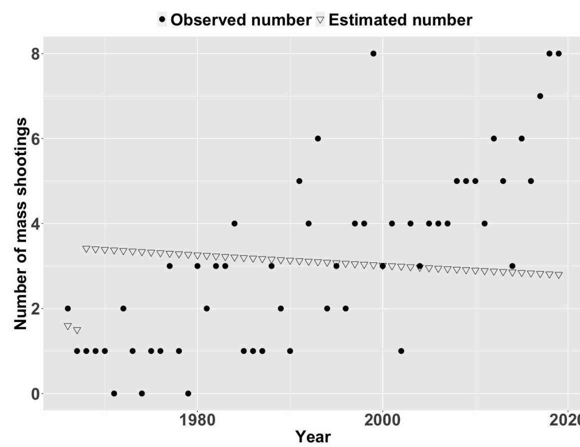
(a) Constant Rate



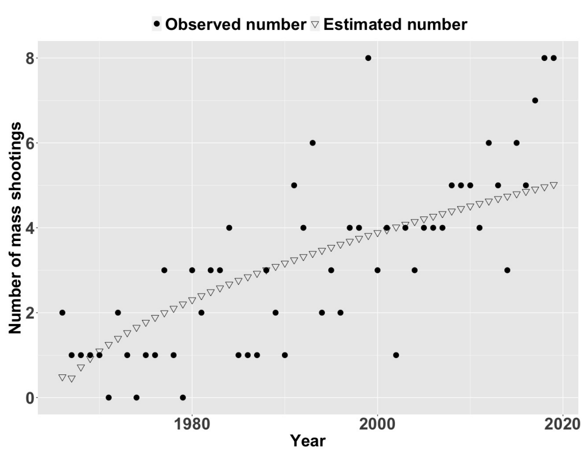
(b) PLP



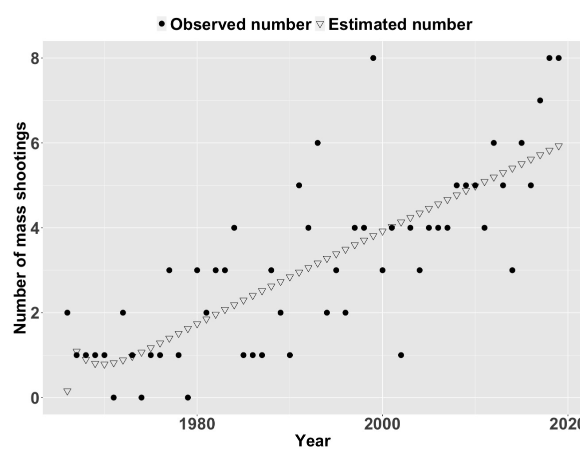
(c) MO



(d) GO

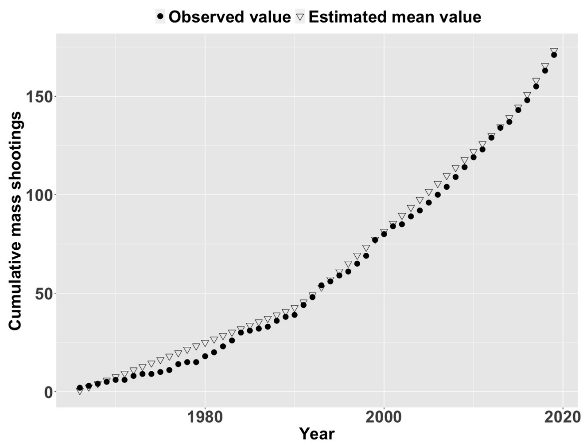


(e) GGO

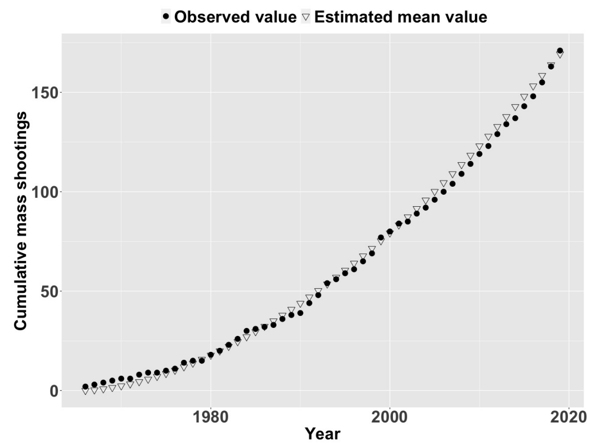


(f) WG

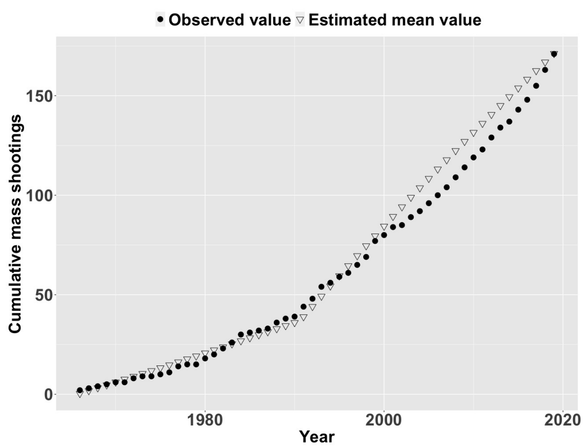
Figure 2.6: Observed value and estimated mean value from the model with two change points from 1966 to 2019



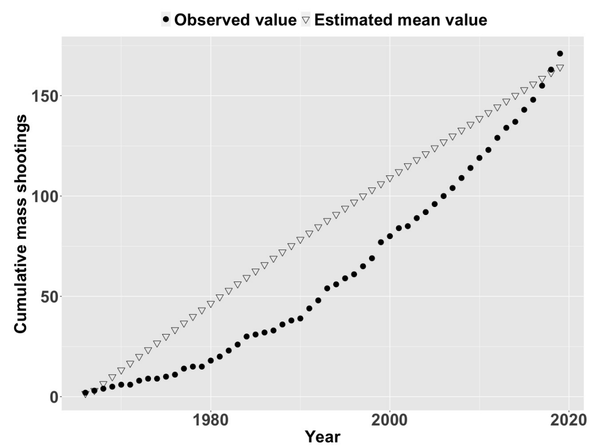
(a) Constant Rate



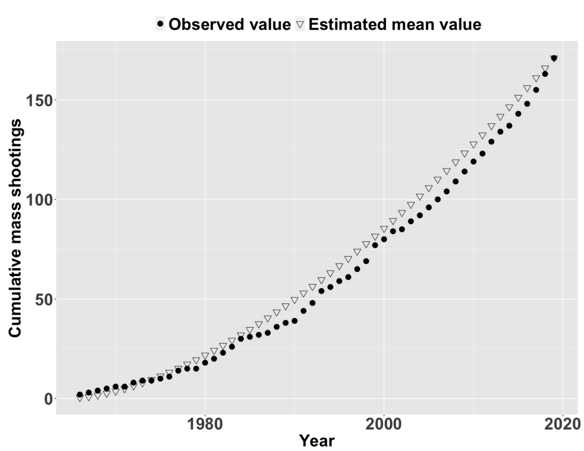
(b) PLP



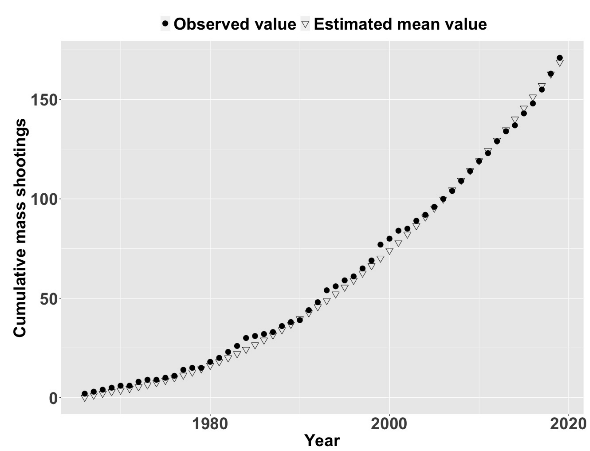
(c) MO



(d) GO



(e) GGO



(f) WG

Figure 2.7: Cumulative observed value and cumulative estimated mean value from the model with two change points from 1966 to 2019

The figures show an approximately linear relationship between the estimated mean value and time except for the points where the change points occur. The constant rate, MO, and GO rate functions have large discontinuities in the estimated mean values where the change points occur. These discontinuities occur with the MO and GO rate functions because the mean values for those rate functions decrease over time. The estimated mean values increase over time with the PLP, GGO, and WG rate functions.

The mean value plots and the cumulative mean value plots reveal that including more change points enables the constant rate function and the MO rate function to fit the data better. However, including more change points in the PLP, GGO, and WG rate function models does not improve their performance. The change points occur very early in time (before 1967), and the rate functions after the final change point are similar whether there are 0, 1, or 2 change points in these three models. Figures 2.2, 2.4, and 2.6 look almost identical for each of these three rate functions.

2.4.3 Model Performance Comparison

Although using graphs to compare the model's mean values with the observed annual number of mass shootings provides important insights into the performance of each model, statistical metrics can help us evaluate and compare the performance of each model. Each of the change-point model's performance is measured by DIC, marginal likelihood, and RSS. Table 2.4 depicts the performance of the different change-point models with different rate functions. Since the value of the marginal likelihood is small for each model, Table 2.4 presents the marginal likelihood on a logarithm scale. We can calculate RSS but not DIC or the marginal likelihood for the Poisson regression model. The RSS for the Poisson regression model is 92.

The PLP, GGO, and WG rate functions have similarly good performance based on all three metrics. The WG rate function performs the best for each number of change points (0, 1, and 2). and the WG rate function with two change points performs the best out of all change-point models according to the three metrics. The RSS for the Poisson regression model is better than the RSS of the WG rate function with 2 changes points, however. Since PLP, GGO, and WG rate functions

Rate function	No change points DIC	ML	RSS	One change point DIC	ML	RSS	Two change points DIC	ML	RSS
Constant	1954	-976	235	1919	-956	130	1911	-954	99
PLP	1916	-957	103	1915	-956	100	1915	-956	100
MO	2050	-1024	448	1921	-959	153	1918	-956	125
GO	2010	-1004	270	1920	-957	126	2008	-1002	263
GGO	1917	-958	114	1916	-958	106	1916	-957	105
WG	**1911**	**-955**	**97**	**1910**	**-954**	**97**	**1909**	**-954**	**96**

Table 2.4: Performance comparison of models with different number of change points (ML = marginal likelihood)

have similar performance for 0, 1, and 2 change points and the change points occur within the first two years of the data, these models suggest that the rate of mass shootings has steadily increased in the past five decades but without a sudden change in the rate. These metrics correspond to what we visually see in Figures 2.2, 2.4, and 2.6.

The constant rate function with two change points also performs well according to the three metrics, and it performs better than all of the models except for the WG rate function and the Poisson regression. This model detects change points in 1991 and 2015, which indicates that the rate of mass shootings in the United States has changed with a large discontinuity in the past five decades.

2.4.4 Forecasting the number of mass shootings in 2020 and 2021

Since the true goal of this research is to assess the risk of mass shootings in the United States, we want to use these models to forecast the number of mass shootings in the future in order to quantify the future risk. The Poisson regression model performs the best according to the RSS metric. The WG rate function with two change points out-performs the other change-point models, and the constant rate function with two change points performs better than the PLP, MO, GO, and GGO rate function models. Thus, this research compares the forecasts of the Poisson regression model, the constant rate function with two change points, and the WG rate function with two change points.

We use each of these three models to forecast the number of mass shootings in 2020 and 2021. Table 2.5 presents the average number of mass shootings in 2020 and 2021 and the corresponding 95% prediction interval for Poisson regression model and the 95% credible intervals for the two change-point models. The prediction interval of the Poisson regression model is based on the empirical simulation method (R-bloggers, 2015). The predicted value of the change-point models is calculated by simulating the Poisson process based on parameters sampled from the posterior distribution in the MCMC simulation.

Forecast of mass shootings in 2020	**Mean**	**95% credible (prediction) interval**
Poisson regression	7.1	[2, 13]
Constant (two change points)	7.6	[2, 14]
WG (two change points)	6.0	[2, 12]
Forecast of mass shootings in 2021	**Mean**	**95% credible (prediction) interval**
Poisson regression	7.3	[2, 13]
Constant (two change points)	7.6	[2, 14]
WG (two change points)	6.1	[2, 12]

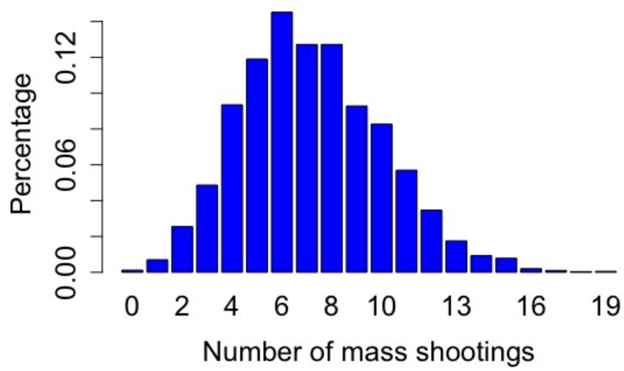
Table 2.5: Predicted average number of mass shootings of 2020 and 2021

The average number of mass shootings in 2020 and 2021 predicted by the three models and the 95% intervals are similar. The Poisson regression model forecasts an average of 7.1 and 7.3 mass shootings, the constant rate forecasts an average of 7.6 shootings, and the WG rate function forecasts an average of 6.0 and 6.1 mass shootings. The Poisson regression model forecasts approximately one more shooting than the WG rate function because the mean number of shootings from the Poisson model exhibits a convex shape with respect to time, and the rate is continuing to increase (see Figure 2.1). The mean number of shootings from the WG rate function exhibits an approximately linear trend with respect to time (see Figure 2.6), and the frequency of mass shootings is increasing at a slower rater than that of the Poisson model. The constant rate function forecasts the largest number of mass shootings. Since the second change point in the constant rate function occurs in 2015 (see Table 2.3), its forecast for 2020 and 2021 is a function of the number of mass shootings from 2015-2019. A large number of mass shootings (7 in 2017 and 8 in 2018 and 2019) occurred in those five years.

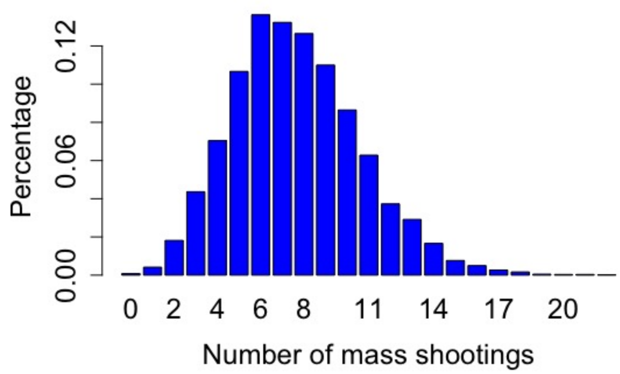
The 95% credible or prediction intervals range from 2-12 to 2-14, which indicate large variability and uncertainty in all three models. These 95% intervals are wider than an interval derived directly from the historical number of mass shootings, which range from 0 to 8 mass shootings a year. At least 1 mass shooting has occurred in ever year since 1980. Since the models' credible or prediction intervals are so much wider than the historical data, this might suggest that the models are forecast too many mass shootings. A probabilistic risk assessment based on the models can evaluate more precisely if these models forecast too many mass shootings.

Risk assessments frequently provide a probabilistic description of the risk or consequences of hazards in order to convey uncertainty (Hubbard, 2009; Kaplan and Garrick, 1981). Simulation is used to generate a probability distribution for the number of mass shootings for the Poisson regression model, the constant rate change-point model, and the WG rate change-point model. A Poisson distribution is simulated based on calculating $\mu = 7.1$ for 2020 from the Poisson regression model. The two change-point models are simulated by sampling from the posterior distribution and generating the number of mass shootings from each sample. These simulated forecasts are depicted as probability mass functions, or column charts, in Figure 2.8. All three figures demonstrate right-skewed distributions with the tail of distribution continuing to 12 or more shootings in 2020. The constant change-point model has the longest tail and the probability of 14 or mass shootings equals 0.035. The constant rate function has the longest tail since its forecast is a function of only five data points. All three forecasts have most of their distribution concentrated from 3-9 mass shootings, which align with historical trend.

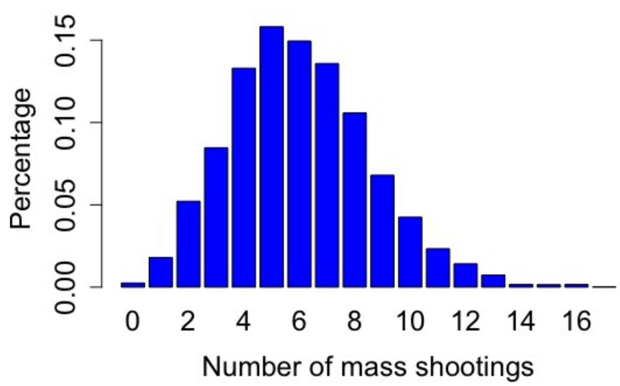
We evaluate the accuracy of these forecasts based on the year 2020. According to the Violence Project database version 3, 2 mass shootings occurred in 2020, which equals the minimum value in the 95% prediction or credible interval for all three of the models. According to the simulated distributions depicted in Figure 2.8, the probability of 2 or fewer mass shooting in 2020 is: (i) 0.034 in the Poisson regression model, (ii) 0.023 in the constant rate function model, and (iii) 0.068 in the WG rate function model. The year 2020 was unique, and the COVID pandemic likely contributed to the small number of mass shootings. All three models do forecast 2 mass shootings



(a) Poisson regression model



(b) Constant with two change points



(c) WG with two change points

Figure 2.8: Simulated distributions of the number of mass shootings in 2020

as possible, though not very likely. If there are 3 or fewer annual mass shootings in the United States in 2021 and 2022—which would be a positive development—this may suggest the models should be adjusted to forecast fewer mass shootings. If the number of mass shootings increases to 5-10 in 2021 and 2022, then these models would appear to generate valid forecasts.

The accuracy of these models may also be evaluated by considering the extreme number of mass shootings as forecast by the models. The models may overestimate the risk of mass shootings in the United States. According to the Violence Project database, the maximum number of mass shootings that have occurred in a year is 8. The Poisson regression estimates a 31% chance, the constant rate function estimates a 36% chance, and the WG rate function estimates a 16% chance of 9 or more mass shootings in 2020. Since all of these models assume the number of mass shootings follows a Poisson distribution, the large probabilities of 9 or mass shootings are due to the right-skewness of the Poisson distribution. Although it would be wrong to forecast a 0% chance of 9 or more mass shootings because the largest number of mass shootings in a year has been 8, a 30-36% chance of 9 or more mass shootings perhaps overestimates the likelihood of such a large number. Sixteen percent may be reasonable although the WG rate function also considers 11, 12, or 13 mass shootings as realistic scenarios, which is much greater than the historical maximum.

2.5 Conclusions

This research studies and analyzes the historical mass shootings data from 1966-2019. We fit a Poisson regression model and several change-point models with different types of rate functions and numbers of change points to model the historical trend in mass shootings. We compare the goodness-of-fit for the models to select the most accurate model. We use the Poisson regression model, the constant rate function with two change points, and the WG rate function with two change points to forecast the number of mass shootings in 2020 and 2021. We compare the forecast results of these three models with each other and with the actual number of mass shootings in 2020.

One of the goals of this research is to understand how the rate of mass shootings is changing. According to the Poisson regression model, the average number of mass shootings is increasing exponentially, and this exponential increase has been constant over time. Since the Poisson regression model results in the smallest RSS of all the models, the average number of mass shootings may be increasing at a cosntant exponential rate. The WG rate function model performs very well and generates very similar results regardless if the model contains 0, 1, or 2 change points. If the change points are included in the model, the WG rate function identifies these change points in the first year or two of the time series. After the change point, if one exists, the average number of mass shootings increases approximately linearly according to the WG rate function. Neither the Poisson model nor WG rate function model reveals a sudden change or discontinuity in the frequency of mass shootings, but the rate has steadily increased over time. The WG rate function model exhibits a slower increase in the rate than the Poisson regression model.

The constant rate function with two change points does identify a sudden change or discontinuity in the rate of mass shootings with the second change point occurring in 2015. The constant rate function suggests that 4 mass shootings occurred on average before 2015, and that number increased to 7.6 after 2015. If the constant rate function model is believed, identifying what factors or events may have led to such a sharp increase in the years immediately prior to 2015 could provide insight into how to mitigate the risk of mass shootings. However, the constant rate function performs worse than the Poisson regression model and the WG rate function with two change points, and the constant rate function assumes a constant rate before and after the change points. We caution against using the constant rate function to forecast mass shootings and prefer the Poisson regression model or the WG rate function change-point model.

The WG rate function and the Poisson regression models both indicate that the frequency of mass shootings has consistently increased since 1966 without a sudden discontinuity. If these models are correct, the media attention and the public concern over mass shootings accurately reflect that mass shootings are becoming a more important problem in the United States. This persistent increase in mass shootings may be due to factors that have also consistently increased since 1966, such as the U.S. population or the number of firearms in the United States. If a model based on these factors is going to be used to forecast the number of mass shootings in the future, then it would be necessary to first forecast the underlying factor(s) as well as mathematically describe the relationship between those factors and the frequency of mass shootings. Contagion theory or the copycat effect could also explain this consistent increase in which more frequent mass shootings lead to even more mass shootings.

In addition to analyzing how the rate of mass shootings has changed, the probabilistic forecasts generated by these models provide important insights into the risk of mass shootings in the United States. All of the models suggest that there is substantial uncertainty in number of mass shootings. According to the models, we expect that 6-8 annual mass shootings will occur, but we should not be surprised if 2 or 3 shootings occur or even if 11 or 12 shootings occur. Although a policy maker interested in understanding and reducing the risk of mass shootings may desire a more precise estimate, these models caution against providing too narrow of range. Only 2 mass shootings occurred in 2020, which represents an extreme, low number of mass shootings according to the models. Although 2020 was perhaps an anomaly, this demonstrates the usefulness of a probabilistic forecast that can assess the likelihood of such a small number. The forecasts of the models, even that of the WG rate function (which has the narrowest distribution of the three models), may contain more uncertainty or variability than what the historical data indicates.

Modeling and forecasting the number of mass shootings should be used to make good risk management decisions to reduce the risk mass shootings. A proper understanding of the risk of mass shootings can help us determine if we as a society should allocate resources to reduce the risk and the trade-offs that we may need to make between our freedoms and the benefit to society. Different actions and strategies will have different effects on mitigating the risk of mass shootings. We can model how different strategies may impact the probabilistic forecasts and the models' rates of mass shootings in order to identify the most cost-effective strategies. We hope that with the help of models such as those presented in this research, reasonable actions and decisions can be reached.

CHAPTER 3. COMPARING DIFFERENT MODELS TO FORECAST THE NUMBER OF MASS SHOOTINGS IN THE UNITED STATES: AN APPLICATION OF FORECASTING RARE EVENT TIME SERIES DATA

3.1 Introduction

Public mass shootings, in which 4 or more individuals are killed from a shooting in a public setting, are a major social problem in the United States and generate a significant amount of media attention and debate over the best strategies to reduce their risk. The United States accounts for approximately one-third of all public mass shootings in the world (Christensen, 2017). Research in mass shootings finds that the rate of mass shootings in the United States has increased in the 21st century (Cohen et al., 2014; Blair and Schwieit, 2014; Duwe, 2020) although these conclusions depend in part on the definition of mass shootings (King and Jacobson, 2017). Equally as disturbing, the lethality of mass shootings has increased in the 21st century (Densley and Peterson, 2019b; Lankford and Silver, 2020). Much of the existing literature on modeling and forecasting the trend in mass shootings focuses on associating different factors such as poverty, gun laws, gun ownership, and population with the prevalence of mass shootings (Duwe, 2020; Lin et al., 2018; DiMaggio et al., 2019; Webster et al., 2020; Fridel, 2020). Little research has analyzed or tested different types of mathematical models that could be used to forecast the frequency of mass shootings in the United States.

Being able to accurately model and forecast the number of mass shootings in the United States should help us understand and analyze the risk of these events and should lead to more informed discussions of how best to mitigate the risk. Mass shootings are rare events, and accurately forecasting rare events is problematic and statistically challenging (Goodwin and Wright, 2010). According to the Violence Project (The Violence Project, 2020), the maximum number of mass shootings that has occurred in a single year is 8 with most years since 2000 seeing 2-6 mass shootings. Models that explicitly incorporate uncertainty may be the best approach to forecasting rare events (Balesdent et al., 2016; Chabridon et al., 2018). Bayesian models incorporate uncertainty in both model parameters and future forecasting results (El-Gheriani et al., 2017; Martin et al., 2015). Poisson models can also be appropriate for modeling rare events because the rare events can be considered a recurrent process (Cook and Lawless, 2007), and the Poisson model does not require normally distributed errors (Winahju and Irhamah, 2016). Attempting to model rare events can also lead to overfitting due to a limited set of data for training the model. Potential solutions to overfitting are penalized regression (e.g., Ridge regression, Lasso regression) (Pavlou et al., 2015; Ying, 2019) and bootstrapping (Choe et al., 2000; Muchlinski et al., 2016).

Several models could be used to forecast the frequency of mass shootings in the United States, but the rare-event nature and annual variability in the number of mass shootings create obstacles to generating an accurate forecast and determining which model is most appropriate. This research compares three models to forecast the annual number of mass shootings, a Bayesian change-point model, the autoregressive integrated moving average (ARIMA) model, and a hybrid of an ARIMA and neural network model. We compare the fitting and forecasting performance of these models. The comparison helps us learn the advantages and disadvantages of each model to forecast the number of mass shootings. Comparing among the results reveals more general insights into the usefulness of each model for forecasting rare events.

A change-point model detects times when the stochastic process or time series changes. The change-point model often models recurrent events in which the rate of occurrence changes with time (Achcar et al., 2010; Gyarmati-Szabó et al., 2011; Guarnaccia et al., 2016; Achcar et al., 2008; Cruz-Juárez et al., 2016). Probabilistic methods of change-point models typically follow a Bayesian approach (Alippi et al., 2015; Raftery and Akman, 1986) and have been used to measure ozone levels in Mexico City (Cruz-Juárez et al., 2016), tuberculosis in New York City (Achcar et al., 2010), the risk of teenage drivers (Li, 2015; Li et al., 2018, 2017, 2020a), and the trend in mass shootings (Xue et al., ndd).

ARIMA is one of the most widely used forecasting models for time series (Van Der Voort et al., 1996; Chen and Tiao, 1990; Contreras et al., 2003; Stergiou, 1989). The ARIMA model can express different time series through its flexible parameters (Box et al., 2011) and can tackle non-stationary time series (Liu et al., 2016). ARIMA models have been applied to predict crime in many countries, including the Philippines (Orong et al., 2018), Australia (Payne and Morgan, 2020), China (Chen et al., 2008), and the United Kingdom (Islam and Raza, 2020). A bivariate ARIMA model is used to investigate the relationship between crime and arrests in Oklahoma City (Chamlin, 1988), and an ARIMA model studies the impact of COVID-19 stay-at-home orders on the gun violence in Buffalo, New York (Kim and Phillips, 2021).

ARIMA models may not be ideal for forecasting rare events in part because the ARIMA equation is a linear equation, but some examples exist in the literature of using ARIMA to forecast rare events. An empirically based smoothing technique combined with ARIMA is used to forecast the occurrence of rare events (strong earthquakes in Parkfield, California) (Ho and Bhaduri, 2015). ARIMA is applied to forecast drought in the Jordan River basin where 0-2 severe droughts occur and 4 moderate droughts occur (Shatanawi et al., 2013). A resampling strategy is proposed to forecast rare events with an ARIMA mdoel when the training data is imbalanced, which can be a feature of rare events (Moniz et al., 2016). An autoregressive model combined with a change-point detection model is used to detect outliers in a time series (Yamanishi and Takeuchi, 2002).

The third type of model used in this paper to forecast mass shootings is a hybrid of ARIMA and an artificial neural network (ANN). ANN is a popular machine learning tool because of its ability to model nonlinearity (Patra et al., 1994; Chattopadhyay and Rangarajan, 2014) and learn from data (Grossberg and Merrill, 1992; El-Sharkawi et al., 1991). Neural networks have been applied to forecast time series of rare events (Fong et al., 2012; Pisa et al., 2019; Fong and Deb, 2015). The hybrid ARIMA-ANN model is proposed for time series forecasting (Zhang, 2003). The hybrid ARIMA-ANN model frequently has a better prediction accuracy than either the pure ARIMA model or ANN model (Wang et al., 2013; Faruk, 2010; Zhang et al., 2018; Tseng et al., 2002). Some of the literature finds that the hybrid model performs better than the ARIMA model for time series forecasting based on limited historical data (Zhang, 2003; Lu et al., 2004; Wang et al., 2013).

This research fits the time series of mass shootings in the United States as recorded by the Violence Project (The Violence Project, 2020) from 1966-2020 to each of the three models: a change-point model with a time-dependent rate function, the ARIMA model, and the ARIMA-ANN hybrid model. Such a comparison requires several unique approaches. Since comparing among statistical models often separates data into training and testing sets, the comparison among these models separates the historical data on mass shootings into different training and testing sets while preserving the time series of the data. The hybrid model is relatively new, and we compare its ability to fit historical data and forecast the future with these other models for rare events. The results of this comparison lead to a discussion of the advantages and disadvantages of using each type of model to forecast the annual number of mass shootings. This discussion may be broadly applicable to other types of applications. Comparing these models contributes significantly to our understanding of the risk of mass shootings and forecasting rare events.

Section 3.2 introduces each of the forecasting models with an explicit focus on the hybrid ARIMA-ANN model because it is less well known. Section 3.3 compares among the different models, and we examine how choosing a different number of nodes in the hybrid model substantially impacts the performance of this model. We also study the effect of including the number of mass shootings in the most recent year 2020 on the forecast of each model. We conclude in Section 3.4 with some insights from this study.

3.2 Forecasting models

This section introduces the three models that are used to forecast mass shootings: the Bayesian change-point model, the ARIMA model, and the hybrid ARIMA-ANN model.

3.2.1 Change-point Model

A non-homogenous Poisson process (NHPP) refers to a Poisson process where the arrival rate changes over time. A change-point model can use a time-dependent rate function to model a NHPP. Commonly used time-dependent rate functions are the power law process, the Musa-Okumoto process, the Goel–Okumoto process, the generalized Goel-Okumoto process, and the Weibull-geometric process (WG) (Barreto-Souza et al., 2011; Goel and Okumoto, 1978; Goel, 1985; Mudholkar et al., 1995; Musa and Okumoto, 1984). The change-point model identifies one or more points in time when the parameters of the rate function changes. Bayesian methods can be used to detect change points or more accurately the posterior distribution for these change points (Raftery and Akman, 1986). After the change-point model is fit to the historical data, we can generate a probabilistic forecast of future events by simulating the NHPP by sampling parameters from the posterior distribution.

Since the Violence Project data for mass shootings contain the date of each mass shooting, the change-point model with the time-dependent rate function can be fit to the historical data on mass shootings by modeling the time between each mass shooting. Since mass shootings have become more frequent over time, a NHPP is a reasonable model for this event. Lei et al. (Xue et al., ndd) fit the different time-dependent rate functions to the mass shootings for zero, one, and two change points. They find that the WG rate function performs the best according to three performance metrics: deviance information criterion, marginal likelihood, and residual sum of squares. Thus, we use the change-point model with the WG rate function in this research to model and forecast the annual number of mass shootings. The WG rate function is:

$$\lambda(t) = \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{1-\rho e^{-\left(\frac{t}{\beta}\right)^{\alpha}}} \tag{3.1}$$

where $\lambda(t)$ is the rate at time t, and $\alpha > 0$, $\beta > 0$, and $\rho \in (0,1)$ are parameters of the rate functions.

As explained in (Xue et al., ndd), this rate function is used to derive the likelihood function for the observed mass shootings data. We assume uniform prior distributions for the parameters in the rate function and the change points. The software package `Stan` which is run via the `R` library `rstan` applies a Markov Chain Monte Carlo sampling technique to generate a posterior distribution for the rate function parameters and the change points.

3.2.2 ARIMA model

The ARIMA model assumes that future observations are linearly dependent on past observations and random errors. The parameters of the non-seasonal ARIMA model are p, d, and q. The parameter p is the order of autoregression. The parameter d is the differencing number. The parameter q is the order of the moving average model (Brockwell et al., 2002; Box et al., 2011). The ARIMA model can be expressed as:

$$\phi(B)\nabla^d(y_t - \mu) = \theta(B)\varepsilon_t \tag{3.2}$$

where y_t is the observation at time t, ε_t is the random error at time t, μ is the mean value, and B is backward shift operator. The backward shift operator causes the observation that it multiplies to be shifted backwards in time by one period. In our case, $By_t = y_{t-1}$. The functions $\phi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$, $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$, and $\nabla^d = (1-B)^d$. The parameters $\varphi_1 \ldots \varphi_p$ are the autoregressive parameters to be estimated. The parameters $\theta_1 \ldots \theta_q$ are the moving average parameters to be estimated. The random errors ε_t are independently and identically distributed with zero mean and a constant variance.

The first step of fitting the ARIMA model to data is choosing the values for p, d, and q. We use the Akaike Information Criterion (AIC) to select the best order of the ARIMA model. An approximate calculation of the AIC is based on the sum of squared residuals (RSS) (Faruk, 2010):

$$\text{AIC} = k\left[\log\left(\frac{2\pi\text{RSS}}{n}\right) + 1\right] + 2(p+q) \tag{3.3}$$

where k is the number of observations in the ARIMA model. Given the order of the ARIMA model, the parameters of model can be estimated by the maximum likelihood estimation (Dent, 1977; Azrak and Melard, 1993).

Python software packages `pmdarima.arima` use `auto.arima` function to estimate parameters in the ARIMA model (Smith, 2020). After setting the maximum values for p and q, the `auto_arima` function will test all different value combinations of p and q and select the best one with the smallest AIC. The `auto_arima` package uses the Augmented Dickey-Fuller test to determine if the time series is stationary (Mushtaq, 2011). If the time series is not stationary, `auto_arima` will provide a suitable value of d.

3.2.3 Hybrid ARIMA-ANN model

The ARIMA model considers the linear combinations of inputs for modeling a time series. However, the nonlinear combinations of inputs may also be needed for the time series data. The ANN model is a widely used model to capture nonlinearities in data (Sane, 1994). The unique advantage of using the ANN is there are no prior assumptions about the form of the model. The form of the ANN model depends on the data. The hybrid AIMRA-ANN models the time series data via a linear part and a nonlinear part. The model can be expressed as:

$$y_t = L_t + N_t \tag{3.4}$$

where L_t is the linear model and N_t is the nonlinear model at time t. The linear model L_t is estimated by the ARIMA model and denoted as $\hat{L}_t$.

The residual at time t, e_t, is obtained by:

$$e_t = y_t - \hat{L}_t \tag{3.5}$$

The analysis of residuals indicates whether the ARIMA model fully captures the time series. The nonlinear component of the residuals can be modeled by using the ANN model. The function h is generated by the ANN model as a function of the preceding n residuals before time t:

$$e_t = h(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t \tag{3.6}$$

where e_t is the current residual, $e_{t-1}, e_{t-2}, \ldots, e_{t-n}$ are the n most recent residuals before time t. The model $\hat{N}_t = h(e_{t-1}, e_{t-2}, \ldots, e_{t-n})$ is the estimate of e_t. Residuals should be normalized and mapped to the range $[0, 1]$ before being input in the ANN model.

The architecture of the ANN model is very flexible (Wang, 2003). Three types of layers exist in the ANN model. The input layer consists of different inputs. The output layer exports the outputs of the model. The hidden layer connects the input layer and output layer. Unlike the input and the output layer, the hidden layer can have more than one layer. The most commonly applied ANN structure is the single hidden layer and back propagation ANN (Haykin, 1999). In this research, the ANN model estimates the current residual e_t based on the previous $t-1$ residuals. The input layer has multiple input nodes. The output layer only has one output node. Multiple hidden nodes exist in the hidden layer. The general ANN architecture considered in this paper is shown in Figure 3.1.

The activation functions embedded in the ANN model allow the model to capture nonlinearity. The activation functions used for each node define the output of that node through some inputs. Many different activation functions can be used in the ANN model, such as the sigmoid (Sig) function, the hyperbolic tangent (Tanh) function, the SoftPlus function, and the binary step function (Mourgias-Alexandris et al., 2019; Lin and Wang, 2008; Glorot et al., 2011; Do et al., 2016). The Sig and Tanh functions are used as the activation functions for the hidden layer and the output layer, respectively, in this research. The form of these two activation functions are:

$$\mathrm{Sig}(x) = \frac{1}{1 + \exp(-x)} \tag{3.7}$$

$$\mathrm{Tanh}(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)} \tag{3.8}$$

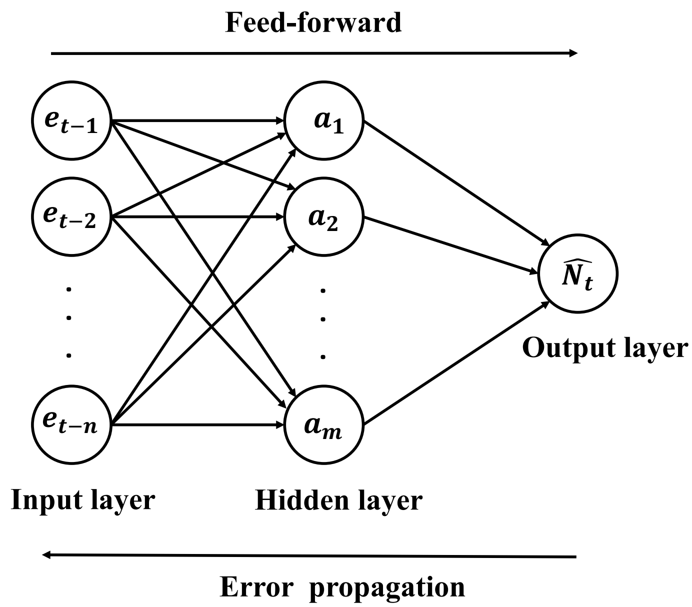


Figure 3.1: Single hidden layer neural network

The mathematical relations between the three layers in the Figure 3.1 can be described by the activation functions. There are I data points to train the ANN model for the nonlinear part of the annual count of mass shootings. For data point i ($i \in I$), $\mathbf{x^{(i)}} = (e_{t-1}^{(i)}, e_{t-2}^{(i)}, \ldots, e_{t-n}^{(i)})$ is the output of the input layer where n is the number of nodes in the the input layer, or more simply, the number of inputs. The corresponding output of the hidden layer is $\mathbf{a^{(i)}} = (a_1^{(i)}, a_2^{(i)}, \ldots, a_m^{(i)})$, where m is the number of nodes in the hidden layer. The relationship between the input layer and the hidden layer is:

$$\mathbf{a^{(i)}} = \mathrm{Tanh}(\mathbf{W^{[1]}}\mathbf{x^{(i)}} + b^{[1]}) \tag{3.9}$$

where $\mathbf{W^{[1]}}$ and $b^{[1]}$ are the parameters for the hidden layer. Similarly, the relationship between the hidden layer and the output layer is:

$$\hat{N_t}^{(i)} = \mathrm{Sig}(\mathbf{W^{[2]}}\mathbf{a^{(i)}} + b^{[2]}) \tag{3.10}$$

where $\mathbf{W}^{[\mathbf{2}]}$ and $b^{[2]}$ are the parameters for the output layer. The cost function used for back propagation to update all parameters should be a measurement of accuracy, such as the mean squared error J (Grossi and Buscema, 2007):

$$J = \frac{1}{I} \sum_{i=1}^{I} \left(e_t^{(i)} - \hat{N_t}^{(i)} \right)^2 \tag{3.11}$$

where $e_t^{(i)}$ is true residual at time time t for data point i as obtained from the ARIMA model.

A potential problem raised with the ANN model is overfitting. Overfitting often happens when the model has a complex structure and many parameters. Regulation methods can reduce the effect of the problem. The regulation term can be added to the cost function to prevent forming a large neural network. The regulation term penalizes large weights and results in fitting a less complex model. Another way to avoid overfitting is to reduce some nodes of the hidden layer (Phaisangittisagul, 2016). This dropout method frequently performs better than adding a regulation term for complex neural networks, but adding a regulation term is easier to apply. Since the ANN model in this research only has a single hidden layer, it is not too complex. An $L2$ regulation term is added to the cost function:

$$\begin{aligned} J_{regularized} &= J + L2 \\ &= \frac{1}{I} \sum_{i=1}^{I} \left(e_t^{(i)} - \hat{N_t}^{(i)} \right)^2 + \frac{\lambda}{2} \left(\mathbf{W}^{[\mathbf{1}]\intercal} \mathbf{W}^{[\mathbf{1}]} + \mathbf{W}^{[\mathbf{2}]\intercal} \mathbf{W}^{[\mathbf{2}]} \right) \end{aligned} \tag{3.12}$$

Another problem that needs to be solved is selecting the number of input nodes n and the number of hidden nodes m shown in Figure 3.1. It is time consuming to try every different combination of n and m. Different methods have been proposed to find the optimal architecture of the ANN model (Benardos and Vosniakos, 2002; Arifovic and Gencay, 2001; Benardos and Vosniakos, 2007). One architecture selection strategy suggests a sequential network construction (SNC) (Moody and Utans, 1994). The SNC for the ANN model is depicted in Figure 3.2. This process can be summarized in two steps. The first step is to select the number of hidden nodes, and the second step is to select choosing the number of input nodes given the hidden nodes.

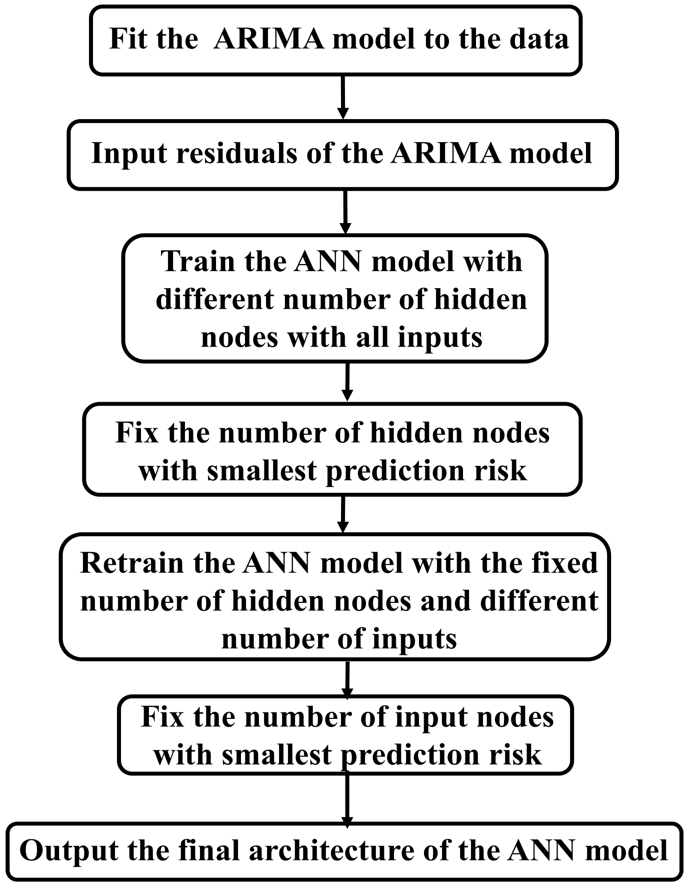


Figure 3.2: The architecture selection of the ANN model

The prediction risk represents the expected prediction performance of the model. By comparing the prediction risk of different models, we can select the model with the best generalization ability. The general definition of prediction risk is the expected mean squared error for the test data set. In many cases, calculating the expected value of the mean squared error is challenging because of a limited test set. Hence, we need to estimate the prediction risk. Other methods to estimate predication risk include cross validation and algebraic estimation (Craven and Wahba, 1978; Akaike, 1970; Barron, 1984; Moody, 1991). We let the ANN model train over all of the data and calculate the prediction risk by the algebraic estimation. The estimation based on all available data is:

$$\hat{P} = J * \left(1 + \frac{2Q}{I}\right) \tag{3.13}$$

where $\hat{P}$ is the estimated prediction risk, J is the mean squared error of the ANN model trained over all available data, and Q is the number of weights used in the ANN model. Based on the single hidden layer neural network shown in Figure 3.1, $Q = n \times m + m$.

3.3 Comparing Among Different Forecasting Models

The data of mass shootings are available from different sources. The commonly used mass shootings data sources are New York City Police Department (NYCPD) (Kelly, 2012a; O'Neill et al., 2016), FBI (FBI, 2016), Mother Jones (Follman et al., 2012), Gun Violence Archive (Gun Violence Archive, 2020a), and Violence Project (The Violence Project, 2020). One of the model types used to estimate the number of mass shootings in the United States—change-point models—assumes that mass shootings is a non-homogeneous Poisson process (NHPP). These models require the time between each incident in a unit of time as small as possible. The Violence Project databases provide the day of each shooting. The Violence Project also provides a long observation period, from 1966-2019. Given these reasons, this research uses the mass shootings data from the Violence Project (The Violence Project, 2020).

Table A.1 in the Appendix shows the annual count of mass shootings recorded by the Violence Project from 1966 to 2019. The first mass shooting recorded by the Violence Project took place on August 1, 1966, which corresponds to the starting time in the change-point model $t_1 = 0$. The ARIMA and hybrid ARIMA-ANN models use the annual number of shootings rather than the number of days between each shooting.

The Violence Project data on mass shootings covers the years 1966-2019. In order to compare the forecast accuracy among the three models, it is necessary to divide the data into a training set and a testing set. Since the data is a time series, randomly dividing the data into a training and testing set is incorrect. Instead, the training set is established as the annual number of mass shootings from 1966 to year T and the testing set is the annual number of mass shootings from year $T + 1$ to 2019. The final year T of the training set varies during this analysis, and the proportion of years in the testing set ranges from 10% to 30% of the total number of years. Our comparison among the three models analyzes the root mean squared error (RMSE) on the training set and on the testing set data and also explores how the models perform when forecasting the annual number of mass shootings in the future.

3.3.1 Comparison of Model Performance with Different Size Training Sets

The last year of the training set T changes from 2003 to 2014. For each training set and its corresponding test set, we fit three different types of forecasting models, the change-point model with the WG rate function, the ARIMA model, and the hybrid ARIMA-ANN model. The Python package `pmdarima.arima` is used to select p, q, and d for the ARIMA model for each training set. We limit the domain of p and q to be between 0 and 5 and the domain of d to be between 1 and 3. The `auto_arima`, which is imported into the Python package, selects $p = 0$, $q = 1$, and $d = 1$ for all of the training sets. Given the ranges of these parameters, the ARIMA(0-1-1) model results in the best fit for the data.

For the hybrid model, ARIMA(0-1-1) is used to model the linear part of the hybrid ARIMA-ANN model. The inputs for the ANN model are the residuals from the ARIMA(0-1-1) model. Each training set may provide a different architecture for the ANN model. As shown in Figure 3.2, the number of hidden nodes is selected before the number of input nodes. The maximum number of nodes in the hidden layer is set to 10 and the ANN model is trained with a different constraint on the maximum number of input nodes, 3, 5, 7, or 10. For each training set, the best architecture of the ANN model is based on the prediction risk calculated by Equation 3.13. The first step trains the fully connected ANN model with all the available input nodes ($n = 3, 5, 7,$ or 10) and varies the number of hidden nodes m from 0 to 10. The number of hidden nodes m is selected with the smallest prediction risk when the number of input nodes n is fixed at 3, 5, 7 or 10. Then we fix the number of hidden nodes m at selected value. The ANN model is then retrained with the number of inputs ranging between 0 and the maximum number of input nodes (3, 5, 7 or 10). The number of input nodes n is chosen for the ANN model with the smallest prediction risk. This architecture selection process is repeated for each training set with the years of the training ranging from 1966-

2003 to 1966-2014. The architecture selection results for different training sets when we consider the different maximum numbers of input nodes presented in the Appendix, Table B.1, Table B.2, Table B.3 and Table B.4.

RMSE is used to compare the different models' performances over the different sizes of the training set (Hyndman and Koehler, 2006). Figure 3.3 displays the training RMSE for each model with the various size of training set data. The RMSE for the change-point model with the

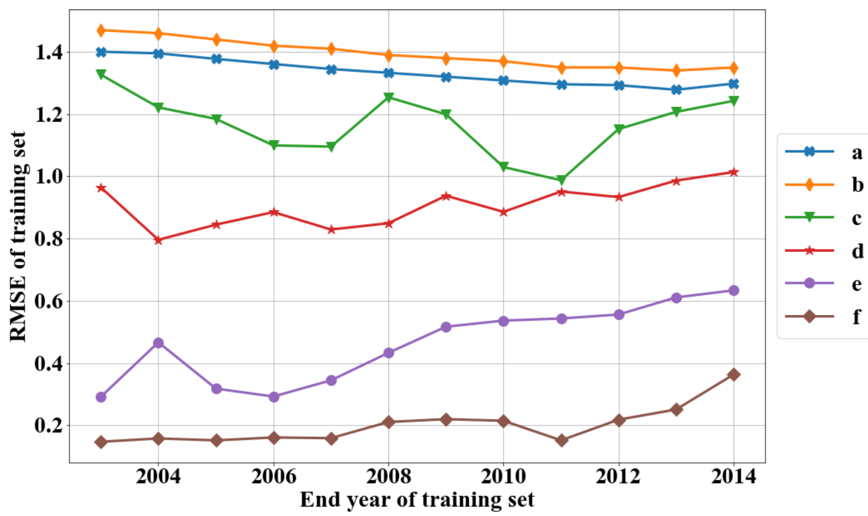


Figure 3.3: The training RMSE of different models over different training sets (a: change-point model with WG rate function, b: ARIMA model, c: hybrid ARIMA-ANN with maximum 3 input nodes, d: hybrid with maximum 5 input nodes, e: hybrid with maximum 7 input nodes, f: hybrid with maximum 10 input nodes)

WG rate function is based on the mean annual counts of the model. The hybrid ARIMA-ANN model always has the smallest RMSE over the different training sets. The change-point model with the WG rate function and the ARIMA model have very similar performance for the training set data, and the RMSE decreases for both models as the training set gets larger except for the largest training set (years 1966-2014). The maximum number of input nodes affects the training RMSE for the hybrid ARIMA-ANN model. The ANN model with a largest maximum number of input nodes (10) has the smallest training RMSE.

Figure 3.4 depicts the test RMSE for each model with the different training sets. The test RMSEs for the change-point model, the ARIMA model, and the ARIMA-ANN model with a

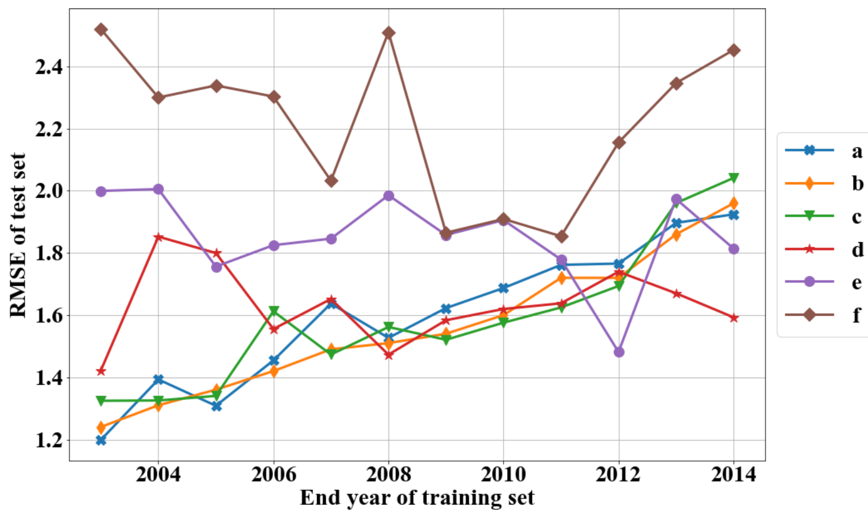
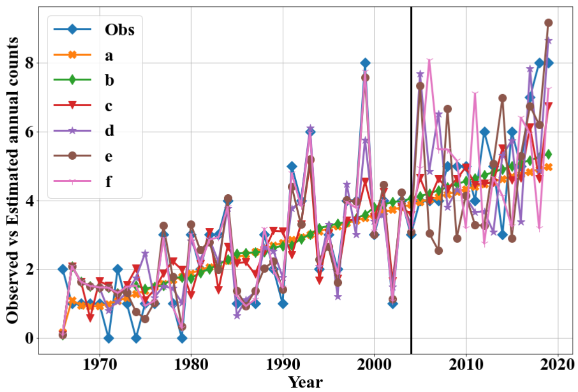


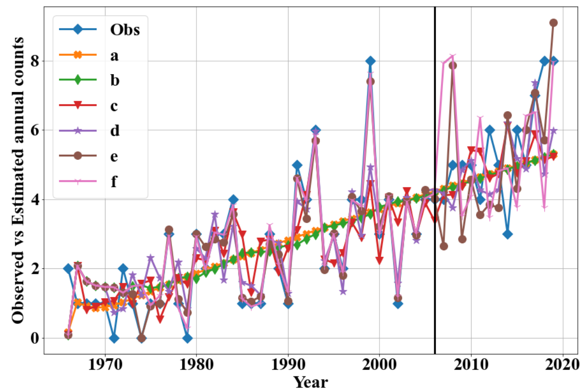
Figure 3.4: The test RMSE of different models over different training sets (a: change-point model with WG rate function, b: ARIMA model, c: hybrid ARIMA-ANN with maximum 3 input nodes, d: hybrid with maximum 5 input nodes, e: hybrid with maximum 7 input nodes, f: hybrid with maximum 10 input nodes)

maximum of three input nodes generally increase as the size of the testing set decreases. The other hybrid ARIMA-ANN models (maximum 5, 7, and 10 input nodes) may have overfitting issues. Although these models have the smallest training RMSE, they frequently have the largest test RMSEs. The hybrid model with a maximum of 5 input nodes looks to perform the best out of all the models when the testing begins with years 2014 or 2015, and the test RMSE remains relatively constant for the different testing sets.

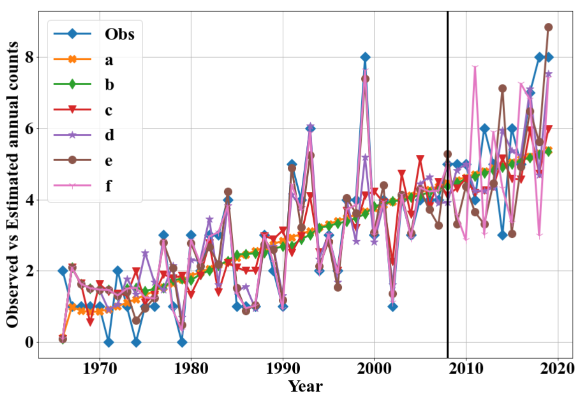
The training RMSE and test RMSE provide exact errors on how the different models fit the mass shootings data from the Violence Project database. Comparing the models' outputs with the annual number of mass shootings enables us to understand the results more intuitively. Figure 3.5 depicts some plots showing these comparisons. The remaining comparison plots are presented in Appendix Figure B.1. The plot for the change-point model depicts the mean annual counts from the model. The change-point model and the ARIMA model provide very similar estimates and capture the increasing trend in the number of mass shootings. While the ARIMA model generally suggests almost a linear trend over time with little variation, the hybrid ARIMA-ANN model follows the variation of the annual counts of mass shootings quite well for the training set



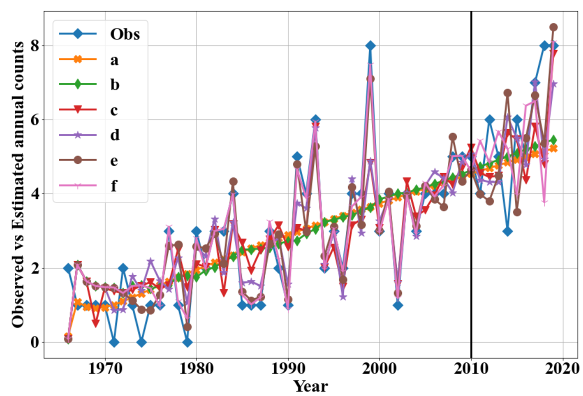
(a) Training set 1966-2004



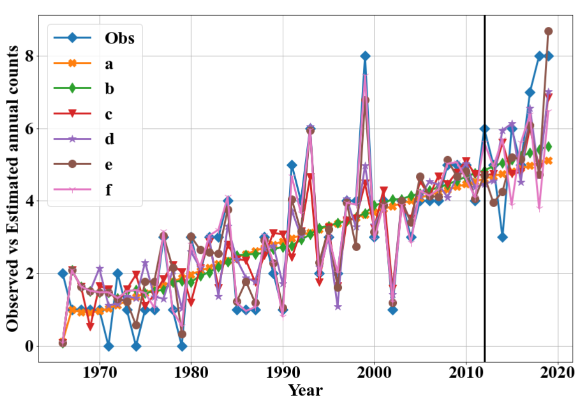
(b) Training set 1966-2006



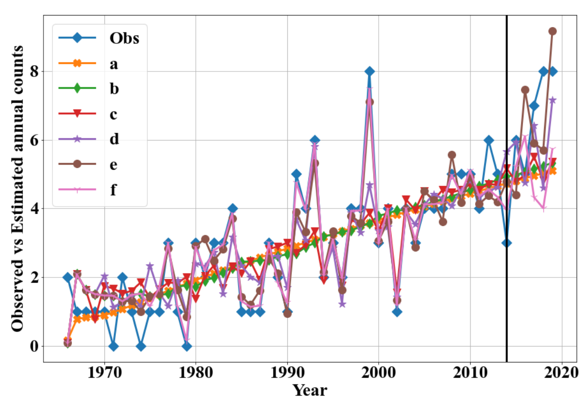
(c) Training set 1966-2008



(d) Training set 1966-2010



(e) Training set 1966-2012



(f) Training set 1966-2014

Figure 3.5: Observed and estimated annual counts from different models with using different training sets (obs: real observations from the Violence Project database, a: change-point model with WG rate function, b: ARIMA model, c: hybrid ARIMA-ANN with maximum 3 input nodes, d: hybrid with maximum 5 input nodes, e: hybrid with maximum 7 input nodes, f: hybrid with maximum 10 input nodes)

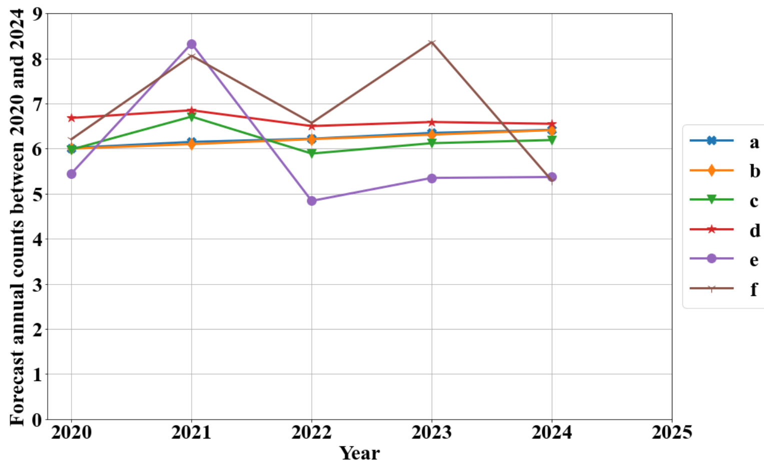
data. The hybrid model is trying to capture a pattern in the variation from year to year. Although the testing sets also depict substantial annual variation, there is not really a pattern. The hybrid models, especially those models with a greater maximum number of input nodes, correctly forecast substantial variation in the annual number of mass shooting, but they generally fail to forecast accurately if a year will have fewer (i.e., 3 or 4) mass shootings or more (i.e., 7 or 8) mass shootings.

According to the above comparison, the hybrid ARIMA-ANN models with a maximum of 7 and 10 input nodes may suffer from overfitting. The large test RMSEs for these two hybrid models indicate that the fluctuation pattern of annual shootings does not continue in the same way. The number of mass shootings in a year exhibits a lot of randomness, which is difficult if not impossible to forecast accurately. The hybrid ARIMA-ANN model with a maximum of 5 input nodes generates a good RMSE for both the training and testing sets, and perhaps this model appropriately balances between reflecting the trend in mass shootings and capturing some of the variation. Another way to forecast the variation in mass shootings is with a prediction interval for the ARIMA model or a credible interval of the change-point model.

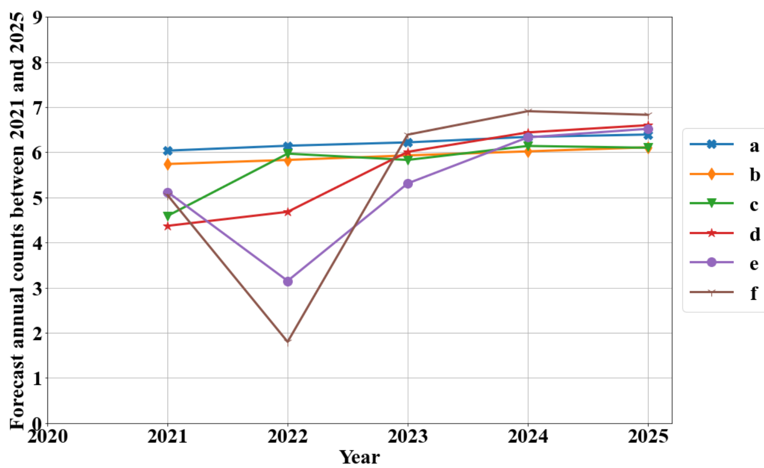
3.3.2 Forecasting Results for the Future

In addition to using testing sets comprised of historical data to compare the models results, we also analyze how the models use the entire set of data to forecast the number of mass shootings 5 years into the future. Each model is trained on the data from 1966 to 2019 in order to forecast mass shootings from 2020 to 2024. The Violence Project recently completed its data for mass shooting in 2020, a year in which only one mass shooting occurred. Each model is also trained on the data from 1966 to 2020 in order to forecast mass shootings from 2021 to 2025. Comparing the forecast of 2020-2024 and the forecast of 2021-2025 can provide insight into the sensitivity of the models to a recent change (2 mass shooting in 2020). Figure 3.6(a) shows the forecasted number of mass shootings in each year from 2020 to 2024 based on the historical data from 1966 to 2019.

The change-point model, the ARIMA model, and the hybrid models with a maximum of 3 or 5 input nodes predict a relatively constant number of mass shootings (between 6 and 7 shootings).



(a) between 2020 and 2024



(b) between 2021 and 2025

Figure 3.6: Forecasting of the annual counts of mass shootings (a: change-point model with WG rate function, b: ARIMA model, c: hybrid ARIMA-ANN with maximum 3 input nodes, d: hybrid with maximum 5 input nodes, e: hybrid with maximum 7 input nodes, f: hybrid with maximum 10 input nodes)

The hybrid models with a maximum of 7 or 10 input nodes forecast much more variation with approximately 8 mass shootings in 2021 but only 5 in 2024. The two models' forecasts diverge in 2023 as their forecasts differ by approximately 3 shootings.

As depicted in Figure 3.6(b), the hybrid models with a maximum of 7 and 10 input nodes are very sensitive to the additional data point of one mass shooting in 2020. These two models have similar forecasts to the other four models in 2021, but the two models forecast a relatively small number of mass shootings (approximately 3 shootings for the 7-input-node model and 2 shootings for the 10-input-node model) in 2022. The other four models predict between 4.5 and 6.5 mass shootings in 2022. All six models forecast a relatively similar number of mass shootings (approximately 6 ± 1 shootings) for the years 2023-2025. Each of the six models that included the data point from 2020 forecasts fewer shootings than the same model if the data point from 2020 is not included. A sudden and recent decrease in the number of mass shootings impacts all of the models' forecasts although it impacts the hybrid models with a large number of inputs the most.

The prediction interval of a forecasting model provides a range in which the future observation will fall with a certain probability. The wider prediction interval means more uncertainty exists in the forecast. We compare the prediction intervals of the forecasted number of mass shootings in 2020 given the data from 1966-2019. We also compare the prediction intervals for 2021 when the data of 2020 is included in training set. Table 3.1 depicts the 95% prediction intervals estimated by different models in 2020 and 2021.

Model	**2020**	**2021**
Change-point model(WG)	[2, 12]	[2, 12]
ARIMA	[3.33, 8.66]	[2.80, 8.68]
ARIMA-ANN (max 3 input nodes)	[3.74, 8.22]	[1.98, 7.20]
ARIMA-ANN (max 5 input nodes)	[4.21, 9.15]	[1.66, 7.08]
ARIMA-ANN (max 7 input nodes)	[1.71, 9.19]	[1.96, 8.28]
ARIMA-ANN (max 10 input nodes)	[2.96, 9.46]	[1.73, 8.37]

Table 3.1: Prediction intervals in 2020 and 2021

The ARIMA-ANN model with 3 input nodes provides the narrowest prediction interval for the forecasts. The width of prediction interval for the change-point model is the widest, which is likely due to the highly skewed posterior distribution in the change-point model. Including the two mass shooting in 2020 changes the models' prediction intervals except for that of the change-point model. The change in 2020 brings more uncertainty with the forecasts of the ARIMA model and the ARIMA-ANN models with 3 or 5 input nodes. The change in 2020 decreases the widths of the prediction intervals for the ARIMA-ANN models with 7 or 10 input nodes. Including another data point in these relatively wide prediction intervals decreases the uncertainty in these two models' forecasts.

3.4 Conclusions

This paper compares the performance of different models to forecast the annual number of mass shootings. Three types of models are compared, the change-point model with a WG rate function, the time series ARIMA model, and the hyrbid ARIMA-ANN model. The hybrid model has four different variants, depending on the maximum number of input nodes. The last year of the training set is varied in order to analyze the performance of the models on slightly different testing sets while keeping the time series elements of the data intact. The models' forecasts for the first half of the decade of the 2020s are compared especially as it relates to whether or not the number of mass shooting in 2020 is included.

Since this paper only examines the performance of these models on one data set, making sweeping conclusions about when each type of model should be used may not be wise. However, the performance and forecasting results can provide more general insights into the advantages and disadvantages of these models and specific insights into the annual number of mass shootings. The hybrid ARIMA-ANN model, especially if the ANN model has a large number of input nodes, fits the training set time series the best. The hybrid model reflects the substantial variation in the historical data of annual mass shootings. Conversely, the ARIMA model depicts a relatively stable trend over time and its RMSE for the training set is the largest of all of the models. The mean of the change-point model depicts a very consistent trend over time. As a probabilistic model, the change-point model's distribution also reflects the large variation in each year.

Although the hybrid models with a maximum of 7 and 10 input nodes have the smallest RMSE for the training set, these two models frequently have the largest RMSE for the testing set. This likely suggests that the hybrid model, especially with a large number of input nodes, can suffer from overfitting. These models try to capture the variation and seem to look for a pattern in the variation, but any pattern that may exist in the variation of the training set does not necessarily hold true in the testing set. The hybrid model often forecasts a large number of mass shootings (e.g., 7 or 8) in one year followed by a small number (e.g., 3 or 4) in the following year. The experiments reveal that the RMSEs for the testing set for the change-point model, the ARIMA model, and the hybrid model with a maximum of 3 input nodes increase as fewer data points are included in the training set, or equivalently as more data points are included in the testing set. The RMSEs for the testing set for the other hybrid models do not show a trend but vary a lot. The hybrid models with a maximum of 5 and 7 input nodes have the smallest test RMSE of all the models when the training set has the largest number of data points. This result may not be generalizable, however, especially because the hybrid model with a maximum of 10 input nodes has the largest test RMSE for that same training set.

This research is unique in that it compares different forecasting models to predict the number of mass shootings in the future. Comparing different forecasting models sheds insight into the advantages and disadvantages of each model. The hybrid ARIMA-ANN model can be tuned to follow variation in the data, but the pattern of the variation may not continue into the future. The mean of the change-point model and the ARIMA model exhibit much more less annual variation and are not influenced as much by the inclusion of a single data point.

CHAPTER 4. QUANTIFYING THE RISK OF MASS SHOOTINGS AT SPECIFIC LOCATIONS

4.1 Introduction

Public mass shootings are horrific events. Mass shootings randomly occur at different types of locations, such as schools, houses of worship, or work places. On December 14, 2012, 28 students and teachers were murdered in a mass shooting at Sandy Hook Elementary School in Newtown, Connecticut (Ray, 2022). A shooter killed 26 people, including an unborn child, and wounded 22 other people at the Sutherland Springs Church on November 5, 2017 in Sutherland Springs, Texas (Gutierrez et al., 2017). At the Molson Coors Beverage Co. in Milwaukee, Wisconsin on February 26, 2020, a 51-year-old employee killed five coworkers (Bosman et al., 2020). Each year, mass shootings take scores of innocent live and have significant societal costs (Soni and Tekin, 2020).

Academic, religious, and public-sector organizations and private businesses consider the best ways to protect people in their buildings from active shooter situations and how best to mitigate the risk of mass shootings. For example, some schools hire armed security officers and restrict access to a school building. College campuses and private businesses may conduct training to instruct people on what to do during an active shooter event (Jonson, 2017). Some states enact legislation that restricts firearm access to reduce the risk of a mass shooting (Wintemute et al., 2019). In California, the gun violence restraining order (GVRO) can be served against an individual to prohibit that person from having a gun, ammunition, or magazines (California Courts, 2020).

The extent of risk mitigation activities and the amount of money spent to reduce the risk should depend on the risk of a mass shooting. Mass shootings remain a rare event, and the vast majority of buildings or organizations will not experience a mass shooting. However, mass shootings are tragic and often gruesome events that take the lives of innocent people who are just going to school or work, shopping, or worshipping. An organization does not know if one of its locations will be one of the unlucky locations where an gunman opens fire. Estimating the likelihood of a mass shooting at a location can help decision makers understand the risk of a mass shooting. They can use this knowledge to make better informed decisions and make trade-offs between mitigating the risk of a mass shooting at their location and other priorities, including other low-probability high-consequence events.

Several difficulties and challenges occur when attempting to quantify the risk of mass shootings at different locations. A mass shooting, sometimes referred to as a public mass shooting, occurs when or more individuals kill at least 4 people (not including the shooter) via firearms in a public location. Most locations and even some states have never experienced a mass shooting, but the chance of a mass shooting at that location or in a state is not zero. Mass shootings seem to be largely random events with rare occurrence, and quantifying both the randomness and rareness requires some important modeling assumptions and mathematical tools. Mass shootings are one type of gun violence and its relation to other types of gun violence is unknown. The risk of mass shootings for a state might be purely dependent on the number of people in the state or might also depend on other factors such as the economic conditions and gun violence in the state.

The goal of this chapter is to present a scientifically rigorous and evidence-based method to calculate the probability that a mass shooting occurs in a state or a specific location like a university campus. Probabilistic methods can be used to capture both the randomness and rareness of mass shootings in order to assess the risk of mass shootings at a national level, at a state level, and at specific locations. Probabilistic methods can forecast mass shootings, which are extreme events with low probability but severe consequences (Duwe et al., 2021). Prior research has assessed the risk of mass shootings in the United States and has studied whether the risk of mass shootings varies from state to state, but no research has attempted to connect a probabilistic risk analysis of mass shootings at the national level to a risk assessment for each state and then to a probability for a specific location.

The unique contribution of this chapter is the development of a method to calculate the probability of a mass shooting at a specific location that is based on the available historical data.

Specifically, we propose a hierarchical probability estimation method by beginning with a national risk assessment, decomposing the national risk picture to a state-level risk assessment, and then decomposing the state risk assessment to specific locations within that state. Different modeling assumptions are potentially legitimate for these tasks, specifically whether the likelihood of a mass shooting is solely dependent on the number of people or is correlated with other factors. The national-level risk model is the change-point model explained in Chapter 2. We use two different models—one based purely on each state's population and the other is a zero-inflated Poisson regression model based on other factors—to assess the probability of mass shootings in a state. We use the state's probability of a mass shooting and the percentage of mass shootings in a type of locations to calculate the probability of a mass shooting at a given type of location (e.g., a school, a house of worship). The final step calculates the probability of a mass shooting at a specific location, such as a specific school or a specific shopping mall. We calculate this final probability assuming each location type has the same risk and assuming the risk depends on the number of people.

Section 4.2 reviews the literature on different models that have been proposed to quantify and compare the risks of mass shootings Section 4.3 outlines the modeling approach and provides details of how the risk is decomposed from the national level to the state level to specific locations. Section 4.4 provides the results of the models and compares among the different modeling assumptions, especially between the two state-level risk models. We calculate probabilities at specific locations in California, the state with the most mass shootings, and Iowa, a state that has never had a mass shooting. Section 4.5 concludes what we have found based on our previous analysis. Quantifying the risk of a mass shooting at different locations can enable local governments, business owners, and school administrators to better understand the chances of such an event and make more informed decisions about mitigation alternatives.

4.2 Literature Review

The existing literature has focused on the increasing rate of mass shootings in the United States. Some risk factors might be associated with the increasing trend of mass shootings (Lin et al., 2018), including anger and resentment (Kwon and Cabrera, 2019a,b) and income inequality and household income (Cabrera and Kwon, 2018). The number of public mass shootings increased in the United States from 1982 to 2012 but did not match the trend of other types of homicides, such as general homicide and stranger homicide (Dillon, 2014). Towers et al. (2019) use negative binomial linear regression to model the per-capital rate of public mass shootings from 1995 to 2018. Since 1995, the per-capita rate of public mass shootings with at least four people killed has not significantly increased, but the rate of shootings with at least six people killed increased between 1995 to 2018 (Towers et al., 2019).

Regional variability in mass shootings exists in the United States (Lemieux, 2014; Gius, 2015; McNeal et al., 2020). Blum and Jaworski (2017) use social disorganization theories and strain theory to identify a non-random spatial pattern of mass shootings that reveals that some areas of the United States have a greater risk of mass shootings than other areas. States that have more permissive gun laws and gun ownership may have higher rates of mass shootings than other states (Reeping et al., 2019; Fridel, 2021) although those findings use a broader definition of a mass shooting than the definition in this dissertation.

Spatial risk analysis of rare events have used a modified logistic regression for the occurrence of landslides (Guns and Vanacker, 2012; Bai et al., 2011), a generalized linear model for nation state failure (Calabrese and Elkink, 2016), and a hidden Markov model for the spread of a rare disease (Forbes et al., 2013). A simple bi-linear regression finds that the latitude and longitude of mass shootings linearly dependent on the year (Andersen et al., 2018). A kernel density estimation and K-means clustering algorithm are used to identify different levels of mass shootings in the United States (D'anna, 2020). In 2019, almost 10% of mass shootings occurred at schools or places of interests (Nance et al., 2020).

Research that identifies the occurrence of mass shootings in different states and if differences among states are due to gun law legislation frequently use a Poisson regression model (Gius, 2015; Duwe et al., 2002) or a negative binomial regression model (Fridel, 2021; Webster et al., 2020). The Poisson regression model is a special case of the negative binomial regression model, the latter of which does not assume that the mean and variance are equal (Hilbe, 2011; Gardner et al., 1995; Berk and MacDonald, 2008). A zero-inflated Poisson regression model is used to investigate the impact of state-level gun ownership rate, serious mental illness rate, poverty percentage, and gun law permissiveness on state-level mass shootings (Lin et al., 2018).

Although a significant amount of research has studied the national trend of mass shootings in the United States, and other research has analyzed differences among states for mass shootings in order to identify reasons for those differences, research has not attempted to connect the national trend to state differences. The goal of much of this research has been to statistically analyze the influence of different factors rather than to provide a probabilistic risk assessment of a mass shooting in a state much less the risk in a specific location. The existing analysis of spatial risk of mass shootings does not provide a rigorous and systematic method to calculate the risk of mass shootings at specific locations. Quantifying the risk for specific locations could benefit decision makers, such as mayors, police commissioners, and school administrators. In order to fill the gap, we propose a framework to quantify the risk of mass shootings for different locations based on the historical data from 1966 to 2020.

4.3 Modeling Framework

The modeling framework estimates the risk at three different levels, as illustrated in Figure 4.1. First, we generate a probability distribution over the annual number of mass shootings in the Untied States through the use a change-point model as presented earlier in this dissertation. Second, we create two different models to calculate the probability that a mass shooting occurs in a specific state given that a mass shooting occurs in the United States. One model assumes this conditional probability equals the proportion of U.S. citizens living in a state. The other model is a zero-inflated Poisson regression model that considers population, a state's gross domestic product, and a more general definition of mass shooting to determine the condition probability of a public mass shooting in a state given one occurs nationally. We use simulation to generate the marginal probability and the average number of mass shootings for each state based on the probability distribution of mass shootings in the entire United States. The third and final step uses the risk assessment at the state level to estimate the likelihood of a mass shooting a a specific location. We quantify the risk for a type of location in a state, such as the probability of a mass shooting at a schools or a retail store. Then, we estimate the probability of a mass shooting at a specific location such as a specific school.

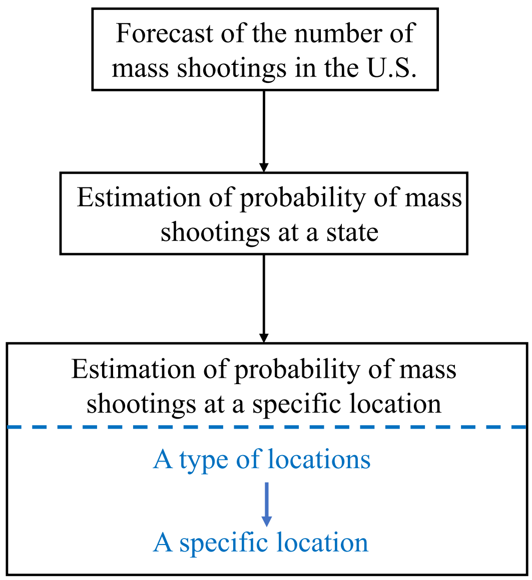


Figure 4.1: Probability estimation of mass shootings at a specific location

4.3.1 Quantifying the Risk of Mass Shootings in the United States

The probability distribution on the number of mass shootings in the United States is derived from the change-point detection model in the previous research. Our prior research compared change-point detection models with constant or time-dependent rate functions and different number of change points. The change-point model that seemed to provide the best fit to the historical number of mass shootings data is the change-point model with the Weibull-geometric (WG) rate function. This chapter uses the change-point model with WG rate function and two change points to forecast the number of mass shootings in the United States.

For the observation period $[0, T]$, the parameters need to be estimated are α, β, and p. The WG rate function is:

$$\lambda^{(WG)}(t) = \frac{\frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}}{1 - pe^{-\left(\frac{t}{\beta}\right)^{\alpha}}} \tag{4.1}$$

Since our model has two change points ($t = \tau_1$, $t = \tau_2$), each parameter in equation 4.1 has three values, one set of values before the first change point, one set of values after the first but before the second change point, and one set of values after the second change point. The mean value function of the WG rate function is:

$$m^{(WG)}(t) = \int_0^t \lambda^{(WG)}(s)ds = -\log\left(\frac{(1-p)e^{-\left(\frac{t}{\beta}\right)^{\alpha}}}{1 - pe^{-\left(\frac{t}{\beta}\right)^{\alpha}}}\right), \qquad t \geq 0. \tag{4.2}$$

The posterior distributions for all parameters in the WG rate function before and after the change points are estimated by using a Bayesian approach. The occurrence of mass shootings is considered as a non-homogeneous Poisson process. The number of mass shootings in each year is simulated based on the posterior distributions of all parameters.

4.3.2 Quantifying the Risk of Mass Shootings in a State

4.3.2.1 Model 1: Population Assumption

Model 1 calculates the conditional probability that a mass shooting occurs in a state given a mass shooting in the United States in a simple way. The conditional probability is equal to the proportion of U.S. population residing in a state:

$$\begin{aligned} &P(\text{mass shooting in state } s|\text{mass shooting in the U.S.}) \\ &\qquad = \frac{\text{population of state } s}{\text{total population of the U.S.}} \end{aligned} \tag{4.3}$$

The predicted number of mass shootings in the United States for the future is provided by the change-points detection model with the WG-rate function. Based on the posterior distribution of the number of mass shootings we generate N forecast samples of the annual number of mass shootings in the United States. For each mass shooting in a forecast sample, we use to simulate in which state that mass shooting occurs. Repeating this process for each of the N forecast samples, we generate N forecast samples of the annual number of mass shootings in each state.

4.3.2.2 Model 2: Zero-Inflated Poisson Regression

A commonly used regression model for the count data is the Poisson regression model. The probability mass function of m events in the i^{th} sample for the standard Poisson regression model is:

$$P(y_i = m) = \frac{\exp(-\lambda_i)\lambda_i^m}{m!} \tag{4.4}$$

where the mean rate $\lambda_i = \exp(\mathbf{x}_i'\theta)$. The coefficients for the input variables $\mathbf{x}_i'$ is the vector θ.

The Poisson regression model may not be appropriate for very rare events in which a data set contains a lot of zeros. The Poisson regression model assumes that the conditional variance is equal to the conditional mean, which may be violated when the data set contains a lot of zeros. Although the United States has experienced at least one mass shooting in every year since 1966, several states have never experienced a mass shooting according to the Violence Project database. We want to estimate the annual risk for each state, and the vast majority of states never have a mass shooting in a single year. Fitting a standard distribution such as the Poisson distribution to data with an excess number of zeros frequently provides a poor fit (Heilbron, 1994). The zero-inflated Poisson regression model has been proposed to model count data with an excess number of zeros (Lambert, 1992; Mullahy, 1986; Heilbron, 1989; Cameron and Trivedi, 2013). The zero-inflation Poisson is a mixing distribution (Böhning, 1998; Lambert, 1992; Fong and Yip, 1993; Yip, 1991). The zero-inflated Poisson regression model splits the data in two parts, zeros and counts. The zero-inflated Poisson model can identify the factors that may cause extra zeros. The zero-inflated

Poisson regression model has been applied to analyze mass shootings in Australia and New Zealand (McPhedran and Baker, 2011).

The zero-inflated Poisson regression splits the count data y into two parts. The first zero-inflated part models whether the count data is zero or non-zero, and the second part uses a Poisson regression model to determine the number of events. The probability that the ith sample is 0 from the first part is p_i. The probability that $y_i = 0$ is $P(y_i = 0) = p_i + (1 - p_i)\exp(-\lambda_i)$, and the probability that $y_i = m$ is $P(y_i = m) = (1 - p_i)\exp(-\lambda_i)\lambda_i^m/m!$ for $m = 1, 2, \ldots$.

The binomial distribution with a logit link function is used to model the zero-inflated part. The probability of a 0 p_i is parameterized by a logistic function of the observable vector of independent variables $\mathbf{z}_i$ (Lambert, 1992). The proportion of zeros is $p_i = \frac{\exp(\mathbf{z}_i'\gamma)}{1+\exp(\mathbf{z}_i'\gamma)}$, where γ is the vector of coefficients for $\mathbf{z}_i'$. The link function is $\log(\frac{p_i}{1-p_i}) = \mathbf{z}_i'\gamma$. For the Poisson count part, the Poisson distribution follows the Poisson regression model where the conditional mean is $\lambda_i = \exp(\mathbf{x}_i'\theta)$, where conditional mean is a function of parameters θ and covariates $\mathbf{x}_i$.

The maximum likelihood method can be used to estimate the coefficients of zero-inflated Poisson regression (Beckett et al., 2014). The joint log-likelihood is maximized by maximizing $\log L(\theta)$ and $\log L(\gamma)$. By setting the partial derivatives of the log-likelihood function with respect to θ and γ equal to 0, we can estimate the parameters used in the zero-inflated Poisson regression model.

After we use historical data and contributing factors to determine the parameters of the zero-inflated Poisson regression model, we need a method to link the model from 4.3.1, the probability distribution on the annual number of mass shootings in the nation, to this state-level model based on the zero-inflated Poisson regression model. For example, if the national model produces 5 mass shootings in the United States, how should the zero-inflated Poisson regression model probabilistically determine in which states those 5 mass shootings occur? Since the zero-inflated Poisson regression model contains both a probability that a state does not experience a mass shooting and the average number of mass shootings in a state, we want to use both pieces of information to create a model for the state-level risk that depends on the national model.

Algorithm 1 provides the algorithm to simulate in which state a mass shooting occurs. The algorithm begins with the N forecast samples from the national model. For each sample j, $j = 1, 2, \ldots, N$, there are n_j mass shootings in the United States. Simulation is used to assign each of the n_j shootings to 1 of 50 states. For a mass shooting, step 1 in the algorithm provides the procedure to simulate a set of candidate states C in which a mass shooting might occur. These candidate states are determined by generating a uniform random number for each state U_s and comparing it to the probability of no mass shooting p_s as calculated from the zero-inflated Poisson regression model. Step 2 uses the mean rate of the Poisson process λ_s for just the set of candidate states and simulates in which state the mass shooting occurs where the probability of the mass shooting occurring in state s' is the ratio of $\lambda_{s'}$ to the sum of all λ_s for s in C.

Algorithm 1 Simulation of mass shootings in each state

Inputs N, p_s, λ_s for each state $s = 1, 2, \ldots, 50$
Output The simulated number of mass shootings in each state corresponding to each of the N samples

```
while j ≤ N do
    if n_j > 0 then
        while k ≤ n_j do
            Step 1: Select candidate states that might have a mass shooting
                    (1) For state s, generate a uniform random number U_s
                    (2) For state s, get the probability of no mass shooting p_s
                        estimated by zero-inflated Poisson regression model
                    (3) If U_s > p_s, then treat state s as a candidate state
                    (4) Return the set of candidate states C
            Step 2: Simulate in which state the mass shooting occurs
                    (1) For states in candidate states set C, get the mean λ_s,
                        s ∈ C from zero-inflated Poisson regression model
                    (2) The probability the mass shooting occurs in state s' is λ_{s'} / Σ_{s∈C} λ_s
                    (3) Generate a uniform random number U to determine in which state
                        the mass shooting occurs based on the probabilities in (2)
            k ← k + 1
        end while
    end if
    j ← j + 1
    Record the simulation results for each forecast sample
end while
```

Models 1 and 2 separately simulate the annual number of mass shootings in each state. The simulated results are used to calculate the expected number of mass shootings and the probability that at least one mass shooting occurs in each state in a year.

4.3.3 Quantifying the Risk of Mass Shootings at A Specific Location

The third step uses the probability of a mass shooting in a state to assess the risk of a mass shooting at a specific location. The historical data on mass shootings provide the proportion of mass shootings that occur in different types of locations. We use that information to estimate the risk of a mass shooting in a specific location, such as the probability of a mass shooting at K-12 school in a given state.

Two methods are used to calculate the probability of a mass shooting at specific location. The first method assumes that the risk is a function of the proportion of people at a specific location, such as the proportion of all K-12 students in a state enrolled in a specific high school. According to this method, the calculation of the probability of mass shootings at a specific location in state s is:

$$\begin{aligned} P(\text{mass shooting at a specific location in state } s) = {} & P(\text{mass shooting in state } s) \\ & \times \text{percentage of mass shootings at that location type} \\ & \times \frac{\text{number of people at a specific location}}{\text{total number of people at that type of location}} \end{aligned} \tag{4.5}$$

The second method assumes that within a type of location, each location in the state carries the same risk of a mass shooting. For example, all K-12 schools in a state have the same probability of a mass shooting. For this method, the calculation of the probability of mass shootings at a specific location in state s is:

$$\begin{aligned} P(\text{mass shooting at a specific location in state } s) = P&(\text{mass shooting in state } s) \\ &\times \text{percentage of mass shootings at that location type} \\ &\div \text{total number of locations of that type in state } s \end{aligned} \tag{4.6}$$

4.4 Results

4.4.1 Data Sources

Several databases on mass shootings sources are available, such as New York City Police Department (NYCPD) (Kelly, 2012a; O'Neill et al., 2016), FBI (FBI, 2016), Mother Jones (Follman et al., 2012), Gun Violence Archive (Gun Violence Archive, 2020a), and Violence Project (The Violence Project, 2020). These different data sources use different definitions of mass shootings and record different numbers of mass shootings in each year. The FBI and the Violence Project define a mass shooting as an incident with four or more victims killed by a firearm in a public place, which is the definition used in this dissertation. The national model (change-point detection model) in this research forecasts the number of mass shootings in the United States. The national model requires the time between incidents in a unit of time as small as possible. Since the Violence Project database provides the day of each shooting and carries the longest observation period (1966-2020), we use the mass shootings data from the Violence Project database.

The two different models that assess the risk of mass shootings for each state both require the population of each state, which is collected from the U.S. Census Bureau official website (U.S. Census Bureau, 2021). Model 2, the zero-inflation Poisson regression model, incorporates additional independent variables that may correlate with the risk for each state. Relying only the Violence Project database of mass shootings may not provide an accurate picture of the risk at a location. For example, Delaware has not experienced a single mass shooting from 1966 (the first mass shooting in the Violence Project database (The Violence Project, 2020)), but using that information to conclude that the likelihood of a mass shooting in Delaware is zero is incorrect. The Gun Violence Archive (Gun Violence Archive, 2020a) includes shootings that occur in both public and private locations as well as targeted shootings (e.g., a gang shooting). The Gun Violence Archive thus records more shootings than the Violence Project. We hypothesize that a state's risk of a mass shooting is correlated with the number of shootings in the Gun Violence Archive. The final independent variable is each state's unemployment rate (Keenen, 2019), which is collected from the U.S. Bureau of Labor official website (BLS, 2022)

4.4.2 Forecasting Mass Shootings in the United States

Prior research forecast the number of mass shootings in the United States using the change-point detection model with the WG rate function as described in Subsection 4.3.1. Fitting the model to the data from the Violence Project reveals that the two change points occur within the first two years of the data set (before 1968) and forecast the same number of mass shootings as the WG rate function with no change points. Figure 4.2 show the fitting results compared with the real annual counts provided by the Violence Project. According to the WG rate function,

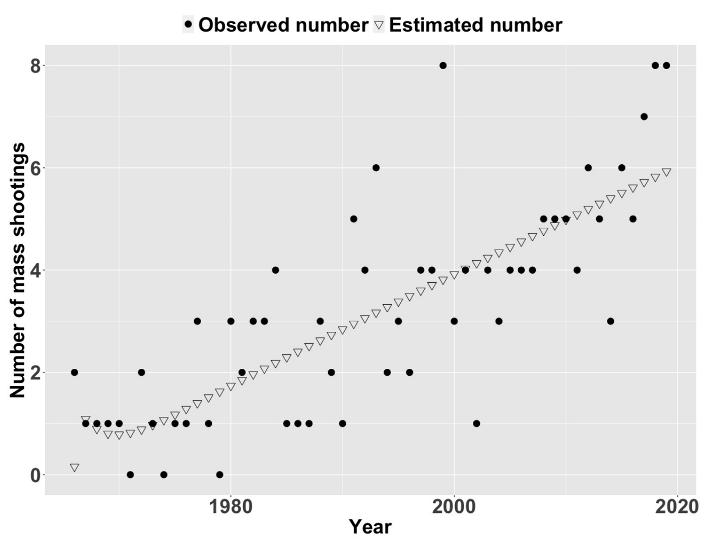


Figure 4.2: Observed value and estimated mean value from the model with WG rate function and two change points

the expected number of mass shootings in 2022 is 6 with a 95% chance there are between 2 and 12 shootings in the United States.

4.4.3 Risk of Mass Shootings in a State

4.4.3.1 Parameter Estimation for Model 2

As discussed in Subsection 4.4.1, we hypothesize four independent variables that can be used to assess the state-level risk using the zero-inflated Poisson regression model. These four variables are the number of mass shootings recorded by the Gun Violence Archive, the state unemployment rate (Iowa Community Indicators Program, 2021; BLS, 2022), the population of state (U.S. Census Bureau, 2021), and the year. The Gun Violence Archive contains shootings for the years 2014-2020, and we use data for the those seven years to fit the model and estimate the parameters. The number of mass shootings in each state in each of the seven years as recorded by the Violence Project is the dependent variable. We try different combinations of the independent variables for fitting the zero-inflated part and the Poisson regression part. Table 4.1 shows the best-fit models.

The log-likelihood value of a regression model is used to represent the goodness of fit for a model. The higher the value of the log-likelihood, the better a model fits the data (Rossi, 2018). We use the Wald test to check if reducing the independent variables significantly changes the fit of the regression model (Zeileis et al., 2008). The model using all four independent variables is the full model. Models using fewer than all four independent variables are reduced models. If the Wald test's p-value is small, we can conclude the fit of the reduced model significantly changes compared to the full model. The log-likelihood and p-value of Wald test are shown in Table 4.1.

Zero-inflated part	Poisson count part	Log-likelihood	Wald test's p-value
Population	GVA+Unemployment rate+Year	-97.74	full model
Population	GVA+Year	-111.6	0.0045
Population	Unemployment rate+Year	-101.4	0.0003
Population	GVA+Unemployment rate	-107.3	< 0.0003

Table 4.1: Different fits for the zero-inflated Poisson regression model (GVA = Gun Violence Archive)

Table 4.1 indicates that the model using the Gun Violence Archive, unemployment rate, and year as predictors for the Poisson count rate has the largest log-likelihood. The Wald test's p-values of the reduced models are all less than 0.01, which suggests that these reduced models do not perform as well as including all four independent variables in the model. Coefficients corresponding to each variable for the full regression model are depicted in Table 4.2. According to these results, as the number of Gun Violence Archive shootings in a state increases, the mean number of shootings in a state given that a mass shooting occurs in that state (i.e., λ_i) also increases.

Zero-inflated part			
Variable	Estimate	Standard error	P-value
Intercept	3.97	1.03	1.08×10^{-4}
Population	-7.06×10^{-7}	2.11×10^{-7}	8.41×10^{-4}
Poisson count part			
Variable	Estimate	Standard error	P-value
Intercept	1.71×10^{2}	8.43×10^{-1}	$<2.00\times10^{-16}$
GA	4.93×10^{-2}	1.38×10^{-2}	3.48×10^{-4}
Unemployment rate	-3.82×10^{-1}	1.35×10^{-1}	4.55×10^{-3}
Year	-8.49×10^{-2}	2.56×10^{-4}	$<2.00\times10^{-16}$

Table 4.2: Inputs considered of zero-inflated Poisson regression model

4.4.3.2 Simulated Results for Each State for Both Models

We use the simulated results for the national model in order to simulate the annual number of shootings for each of the two state-level models. Model 1, which assumes the risk in each state is proportional to a state's population, determines in which state a mass shooting occurs via each state's probability given in equation 4.3.2.1. The simulation of Model 2 relies on algorithm 1 in order include both the count part and the Poisson part of the zero-inflated Poisson regression model. We first compare the two models by simulating how well these models produce results that aligns with the actual number of mass shootings in each state from 2014-2020. We choose these 7 years because the Gun Violence Archive data compasses those years. We randomly choose 7 data points from the $N = 15000$ samples in the national model and simulate in which state each mass shooting occurs. The simulation is used to calculate a 7-year average for each state. We repeat this process 1000 times, and the results are depicted in Table 4.3. The real value is the 7-year average according to the Violence Project. The mean and standard deviation are the average and the standard deviation of 1000 simulated 7-year averages. A 90% confidence interval is assessed using the student-t distribution with 6 degrees of freedom. If the real value is less (greater) than the mean, the p-value is assessed as the proportion of simulations in which the 7-year average is less (greater) than or equal to the real value.

The two models produce similar results, and the states with the greatest risk of a mass shooting are the most populous states. The sum of the absolute differences between the real 7-year averages and the simulated means of Model 1 is 4.55, and the sum of the absolute differences between the real 7-year averages and simulated means of Model 2 is 4.54. The mean values of Model 2 are slightly more dispersed than the mean values of Model 1. For example, the most populous state California has a simulated mean of 1.23 annual mass shootings according to Model 2 but only a simulated mean of 0.73 mass shootings in Model 1. Several states in Model 2 have a mean of 0.00 mass shootings (rounded to the nearest hundredth) whereas Wyoming has a mean of 0.01 mass shootings according to Model 1. The standard deviations for Model 1 are frequently greater than the standard deviations for Model 2, which leads to slightly wider 90% confidence intervals for Model 1 than that of Model 2. The real 7-year average for each state is contained within the 90% confidence intervals for all 50 states in Model 1. The real 7-year average for the states of Illinois, Mississippi, Nevada, Oregon, and Wisconsin are not contained within the 90% confidence intervals from Model 2. The p-values for many states that have experienced a mass shooting are greater than 0.2 for both models. Model 2 has three p-values less than 0.05.

We assess the risk of a mass shooting for each state for a single year by using the 15,000 samples from the national model to simulate the two state-level models. Figure 4.3 shows the expected annual number of mass shootings in each state for Model 1, and Figure 4.4 shows the expected number for Model 2. For states with a very low risk of mass shootings (less than 0.02 annual mass shootings), these states carry slightly more risk according to Model 1 than according to Model 2. The most populous states that exhibit the greatest risk of a mass shooting (e.g.,

		Model 1				Model 2			
State	Real	Mean	Std	90% CI	P-value	Mean	Std	90% CI	P-value
AL	0.00	0.09	0.12	[0, 0.32]	0.54	0.04	0.08	[0, 0.19]	0.73
AK	0.00	0.02	0.05	[0, 0.11]	0.89	0.00	0.00	[0, 0]	1.00
AZ	0.00	0.13	0.14	[0, 0.4]	0.39	0.07	0.10	[0, 0.26]	0.64
AR	0.00	0.06	0.09	[0, 0.24]	0.66	0.00	0.02	[0, 0.04]	0.99
CA	1.00	0.73	0.32	[0.12, 1.35]	0.26	1.23	0.42	[0.41, 2.06]	0.35
CO	0.00	0.11	0.13	[0, 0.36]	0.45	0.07	0.10	[0, 0.27]	0.62
CT	0.00	0.06	0.10	[0, 0.25]	0.64	0.00	0.02	[0, 0.05]	0.97
DE	0.00	0.02	0.05	[0, 0.12]	0.87	0.00	0.00	[0, 0]	1.00
FL	0.71	0.41	0.25	[0, 0.89]	0.18	0.35	0.23	[0, 0.78]	0.10
GA	0.00	0.20	0.18	[0, 0.55]	0.25	0.36	0.22	[0, 0.79]	0.09
HI	0.00	0.02	0.06	[0, 0.14]	0.84	0.00	0.01	[0, 0.03]	1.00
ID	0.00	0.04	0.07	[0, 0.17]	0.78	0.00	0.00	[0, 0.01]	0.99
IL	0.14	0.23	0.19	[0, 0.59]	0.53	0.89	0.35	[0.2, 1.58]	0.02
IN	0.00	0.12	0.14	[0, 0.39]	0.43	0.12	0.13	[0, 0.37]	0.43
IA	0.00	0.06	0.09	[0, 0.24]	0.65	0.00	0.02	[0, 0.05]	0.97
KS	0.00	0.06	0.09	[0, 0.23]	0.67	0.00	0.02	[0, 0.05]	0.97
KY	0.00	0.07	0.11	[0, 0.28]	0.61	0.02	0.05	[0, 0.11]	0.92
LA	0.00	0.09	0.12	[0, 0.32]	0.54	0.03	0.07	[0, 0.17]	0.80
ME	0.00	0.03	0.06	[0, 0.14]	0.83	0.00	0.01	[0, 0.01]	1.00
MD	0.14	0.12	0.13	[0, 0.37]	0.55	0.14	0.14	[0, 0.41]	0.61
MA	0.00	0.13	0.13	[0, 0.38]	0.41	0.11	0.13	[0, 0.36]	0.47
MI	0.29	0.17	0.16	[0, 0.48]	0.33	0.16	0.16	[0, 0.46]	0.30
MN	0.00	0.10	0.12	[0, 0.34]	0.50	0.04	0.08	[0, 0.19]	0.76
MS	0.14	0.06	0.09	[0, 0.23]	0.35	0.00	0.02	[0, 0.03]	0.01
MO	0.14	0.12	0.13	[0, 0.36]	0.56	0.12	0.13	[0, 0.38]	0.57
MT	0.00	0.02	0.05	[0, 0.11]	0.89	0.00	0.01	[0, 0.01]	1.00
NE	0.00	0.04	0.07	[0, 0.17]	0.77	0.00	0.01	[0, 0.03]	0.99
NV	0.14	0.06	0.09	[0, 0.24]	0.36	0.00	0.02	[0, 0.04]	0.02
NH	0.00	0.03	0.07	[0, 0.16]	0.81	0.00	0.01	[0, 0.03]	1.00
NJ	0.14	0.16	0.15	[0, 0.46]	0.71	0.20	0.17	[0, 0.53]	0.58
NM	0.00	0.04	0.07	[0, 0.17]	0.78	0.00	0.01	[0, 0.01]	1.00
NY	0.00	0.35	0.23	[0, 0.79]	0.08	0.20	0.17	[0, 0.52]	0.25
NC	0.00	0.20	0.17	[0, 0.52]	0.24	0.21	0.17	[0, 0.55]	0.21
ND	0.00	0.02	0.05	[0, 0.11]	0.90	0.00	0.00	[0, 0.01]	1.00
OH	0.14	0.21	0.17	[0, 0.55]	0.54	0.21	0.17	[0, 0.55]	0.53
OK	0.00	0.07	0.11	[0, 0.28]	0.60	0.01	0.04	[0, 0.09]	0.92
OR	0.14	0.08	0.11	[0, 0.29]	0.41	0.01	0.05	[0, 0.1]	0.06
PA	0.43	0.23	0.18	[0, 0.58]	0.23	0.25	0.19	[0, 0.62]	0.26
RI	0.00	0.02	0.05	[0, 0.12]	0.88	0.00	0.01	[0, 0.01]	1.00
SC	0.14	0.10	0.12	[0, 0.32]	0.50	0.05	0.09	[0, 0.22]	0.29
SD	0.00	0.02	0.05	[0, 0.12]	0.88	0.00	0.01	[0, 0.02]	1.00
TN	0.29	0.12	0.13	[0, 0.37]	0.21	0.11	0.12	[0, 0.35]	0.18
TX	0.86	0.53	0.28	[0, 1.08]	0.17	0.64	0.30	[0.05, 1.23]	0.28
UT	0.00	0.06	0.09	[0, 0.24]	0.63	0.01	0.03	[0, 0.06]	0.97
VT	0.00	0.01	0.04	[0, 0.09]	0.93	0.00	0.00	[0, 0.01]	1.00
VA	0.14	0.16	0.15	[0, 0.45]	0.69	0.26	0.20	[0, 0.64]	0.47
WA	0.29	0.14	0.14	[0, 0.41]	0.27	0.09	0.12	[0, 0.32]	0.13
WV	0.00	0.03	0.07	[0, 0.16]	0.79	0.00	0.01	[0, 0.01]	1.00
WI	0.29	0.11	0.13	[0, 0.36]	0.19	0.06	0.09	[0, 0.23]	0.06
WY	0.00	0.01	0.04	[0, 0.08]	0.94	0.00	0.00	[0, 0]	1.00

Table 4.3: The real 7-year average and simulated 7-year averages for each state

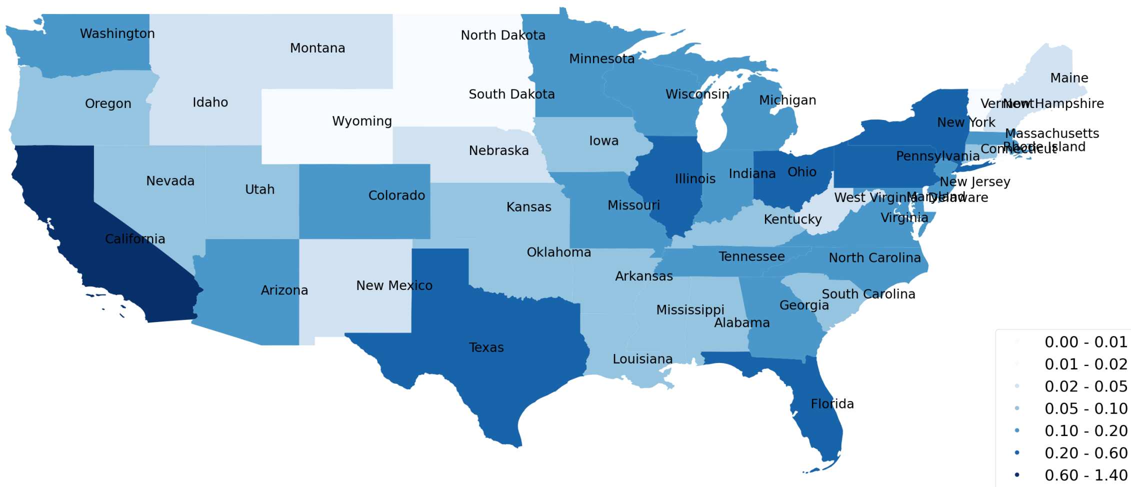


Figure 4.3: Expected annual number of mass shootings for each state from Model 1 based on population

California, Texas) have a greater number of mass shootings according to Model 2 than according to Model 1.

We compare the two models with the actual average number of mass shootings each state has experienced according to the Violence Project database. We use two error metrics to represent the error in each model. We consider the mean absolute error (MAE) and the mean arctangent absolute percentage error (MAAPE). MAE depicts the average magnitude of error generated by different models, and MAAPE captures how far the simulated averages are from the corresponding actual average number of mass shootings. MAAPE is used instead of the mean absolute percentage error because some states have never experienced a mass shooting. The mean absolute percentage error will thus have division by zero. The arctangent function used by MAAPE can transfer infinity to finite radians (Kim and Kim, 2016). We make this comparison for three groups: (i) all 50 states, (ii) the 5 most populous states (California, Texas, Florida, New York, and Pennsylvania) which account for almost 50% of all mass shootings, and (iii) the other 45 states. The comparison results are in the Table 4.4

Comparing between the two models using MAE and MAAPE reveals the same conclusions. The error metrics for both models are very similar. Model 2 has a slightly smaller error for the

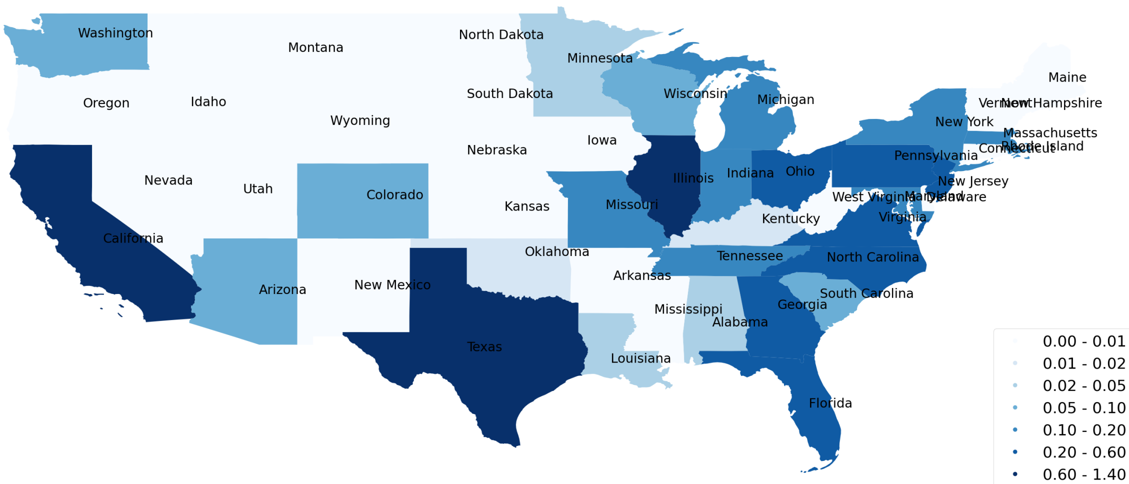


Figure 4.4: Expected annual number of mass shootings for each state from Model 2 based on the zero-inflated Poisson regression

	Number of states	Percetage of mass shootings	Model 1 - MAE	Model 2 - MAE	Model 1 - MAAPE	Model 2 - MAAPE
Group 1	50	100%	0.0914	0.0926	1.1430	1.1990
Group 2	5	44%	0.2925	0.2492	0.6090	0.5941
Group 3	45	56%	0.0691	0.0752	1.2023	1.2663

Table 4.4: MAE and MAAPE in different groups of States

most populous states (Group 2). Model 1 has a slightly smaller error for all the other 45 states or for all 50 states.

4.4.4 Risk of Mass Shootings at Specific Locations

We extend both models of the state-level risk assessment to assess the risk at specific locations in two different states, California and Iowa. California has experienced the most mass shootings. Iowa has never experienced a mass shooting. The annual probability of a mass shooting in Iowa is 0.058 form Model 1 and 0.005 from Model 2. The Violence Project provides the percentage of mass shootings in different types of locations. We use those percentages and each model's state-level probability to calculate the probability of a mass shooting in 9 different locations in each state. Table 4.5 depicts these probabilities.

Locations	Percentage	Model 1		Model 2	
		CA	IA	CA	IA
Overall	1	0.5162	0.0579	0.7077	0.0049
K-12 school	0.0751	0.0533	0.0044	0.0884	0.0004
College, university	0.0520	0.0372	0.0030	0.0621	0.0002
Government building, place of civic importance	0.0347	0.0250	0.0020	0.0418	0.0002
House of worship	0.0636	0.0453	0.0037	0.0753	0.0003
Retail	0.1676	0.1149	0.0097	0.1866	0.0008
Restaurant, bar, nightclub	0.1329	0.0923	0.0077	0.1511	0.0006
Workplace	0.3064	0.1998	0.0177	0.3142	0.0014
Place of residence	0.0867	0.0612	0.0051	0.1013	0.0004
Outdoors	0.0809	0.0573	0.0047	0.0949	0.0004

Table 4.5: The probability of mass shootings in different locations in two states

We use the probabilities in Table 4.5 to assess the risk of a mass shooting at a specific location such as a specific school in California and a specific school in Iowa. As explained in Subsection 4.3.3, two methods are used. Equation 4.5 is a person-weighted method where the risk is proportional to the number of students in the school. Equation 4.6 assumes that each K-12 school a state has the same probability of a mass shooting.

The total number of registered students K-12 in Iowa is 540,496 (Iowa Department of Education, 2020). The total number of registered students in California is 6,620,140 (California Department of Education, 2020). With 5,423 students, River Springs Charter is the largest K-12 school in California. The largest K-12 school in Iowa is Lincoln High School with 2,423. California has 10,303 K-12 schools and Iowa has 1765 K-12 schools. The probabilities of a mass shooting at River Springs Charter and Lincoln High School using the two different state-level models and the two methods for assessing the risk at specific locations (person-weighted and equal probability) are provided in the Table 4.6.

	Model 1		Model 2	
	Person-weighted	Equal probability	Person-weighted	Equal probability
River Springs Charter (CA)	3.1×10^{-5}	3.7×10^{-6}	4.3×10^{-5}	5.1×10^{-6}
Lincoln High School (IA)	1.9×10^{-5}	2.4×10^{-6}	1.7×10^{-6}	2.2×10^{-7}

Table 4.6: The probability of mass shootings at two schools in different states

The annual likelihood of a mass shooting at River Spring Charters in California ranges between 3.7×10^{-6} and 4.3×10^{-5}. The annual chance of a mass shooting at Lincoln High School in Iowa ranges between 2.2×10^{-7} and 1.9×10^{-5}. The choice of the model and assumptions can increase or decrease the probability by a factor of 10 or even 100. The risk of a mass shooting at River Springs Charter is about 10 times greater than the risk at Lincoln High School.

4.5 Conclusions

This research proposes a hierarchical model to assess the risk of mass shootings at the national level, within each state, and at specific locations. The national risk assessment is based on a change-point detection model. Two models translate the forecast on the number of mass shootings in the United States to a state-level risk assessment. Model 1 assumes the risk is proportion to the population in each state. Model 2 uses a zero-inflated Poisson regression where population, year, unemployment rate, and a database with a broader definition of mass shootings are the independent variables. The state-level risk assessments are used to assess the risk at a specific location such as a named high school. Two methods are used in this assessment, one based on the proportion of people at that location and the other assuming each location has the same risk of a mass shooting.

The change-point detection model forecasts an average of 6 mass shootings in the United States in 2022 although the model contains substantial variation. Both state-level models generate similar risk assessments for each state based on this national forecast. Given the similarity in state-level risk assessments, it seems that the risk that each state faces depends to a large extent on the state's population, which tracks well with the Violence Project data. Whether including more factors such as the unemployment rate and the number of shootings from the Gun Violence Archive contributes to a better assessment and understanding of the risk at the state level is unclear. The range in the expected number of mass shootings in Model 2 is a little greater than the range in Model 1. The expected number of mass shootings in 3 states exceeds 0.6 in Model 2 but only 1 state in Model 1. The expected number of mass shootings in 21 states is less than 0.01 in Model 2 but only 5 states in Model 1.

The probability calculations in this chapter provide a base-level assessment of the risk for a state and a specific location. In a constrained resource environment where state and local governments, school administrators, law enforcement agencies, and private businesses need to consider carefully what issues deserve additional resources and money, understanding the probability of different risks can help guide and prioritize those investments. Even if a mass shooting has a very small probability of occurrence, on the order of 10^{-6} per year, the decision maker for the location might still want to mitigate the risk because the consequences of a mass shooting are so tragic. A mass shooting is a specific type of shooting that requires at least four fatalities. Mitigating the risk of a mass shooting will also reduce the risk of a shooting in which fewer than four individuals die, which is likely greater than the chance of a mass shooting. Protecting against a mass shooting can help prevent an active shooter incidence from resulting in a mass shooting.

An individual responsible for a location would likely be interested in knowing how the base-level probability of a mass shooting changes if warning signs appear that indicate an individual might engage in an active shooter event. For example, how does the probability of a mass shooting at River Springs Charter or Lincoln High School change if one of its students is diagnosed with severe mental health and potential violent tendencies or if a student threatens another student or teacher? Future work can model how a probability can be updated based on warning signals. Bayesian methods where the prior probability is the base-level risk assessment may be well suited for such an analysis.

The unique contribution of this research is providing a novel risk quantification process for a specific location. The number of mass shootings for each state are simulated based on a forecast of the number of mass shootings in the United States and two different models that provide the relative risk for each state. The risk of a mass shooting at specific location are calculated based on the probability of a mass shooting in a state and the observed percentage of mass shootings in different types of locations. The quantification process can be applied to various locations and generate insights into which areas or locations exhibit the greatest risk of a mass shooting.

CHAPTER 5. CONCLUSIONS

5.1 Summary

This research models the frequency of mass shootings in order to forecast the annual number of mass shootings in the United States. The forecast is used to assess each state's risk of a mass shooting and calculate the risk at a specific location. The Poisson regression model and the change-point detection model are used to quantify the frequency of mass shootings in the United States. Six different types of rate functions with zero, one, or two change points are fit to the historical data. They are constant rate, PLP rate, MO rate, GO rate, GGO rate, and WG rate functions. The different models are compared using several performance metrics.

The Poisson regression model outperforms the other models based on the RSS metric. A limitation of only considering the RSS is that RSS is sensitive to the outliers. Some of the change-point detection models have similar RSS to that of the Poisson regression model. Other performance metrics (marginal likelihood and DIC) compare the change-point models based on a probabilistic analysis of the fitted models. The WG rate function performs the best of the change-point models according to the DIC, marginal likelihood, and RSS metrics.

The Poisson regression model indicates that the average number of mass shootings may be increasing at a constant exponential rate. The WG rate function suggests that the rate of mass shootings has increased at an approximately linear rate in the past 50 years without any sudden discontinuities. The constant rate function with two change points suggests that the rate of mass shootings changed around 2015. Based on their fit to the historical data, the Poisson regression model and the WG rate function change-point model seem to be the two best models to forecast the annual number of mass shootings in the United States.

Several models can be used to forecast the number of mass shootings. Each model uses different assumptions, and Chapter 3 explores different forecasting models in order to generate more insights into how to forecast rare events, such as mass shootings, with limited historical data and substantial annual variability. By comparing the forecasting results among three types of models, the change-point detection model, ARIMA model, and hybrid ARIMA-ANN model, a better understanding of benefits and drawbacks of each modeling method can be achieved. Both the change-point model and the ARIMA model tend to depict the overall trend of the mass shootings data. In contrast, the hybrid ARIMA-ANN model tries to capture the variation of data by modeling the residuals from the ARIMA model through the ANN. The randomness and rareness of mass shootings makes it challenging to identify a pattern of variation by using ANN. Hence, the historical variation captured by the ANN might not be useful to forecast the future variation.

How can we use the forecast of the annual number of mass shootings to obtain a clearer picture of the risk of mass shootings at a local level? The third task of this research is to use the results of the previous analysis to quantify the risk of mass shootings at specific locations. Two models are used to the risk of mass shootings in each state: one model in which the risk is proportional to a state's population and the zero-inflated Poisson regression model that includes several factors such as population, the unemployment rate, and a database of shootings. The risk of a mass shooting in a state is used to calculate the risk at specific locations using two different assumptions: (i) each location within a type has equivalent risk or (ii) the risk is proportional to the number of people at a location.

Although this dissertation is not able to provide exact information about when and where mass shootings will occur, the analysis presents the future trend of mass shootings in the United States. The probabilistic forecasting results indicate the states and locations that exhibit the greatest risk of mass shootings. Policymakers, public safety officers, school administrators, and even private sector businesses can use this analysis to determine the level of resources that should be used to prevent and protect against mass shootings.

5.2 Future Work

This research focuses on the number and the frequency of mass shootings. Another important element in the risk of mass shootings is the number of people injured and killed, which could also be modeled and forecast. Some research finds that the increased risk of mass shootings is due more to a tendency for recent mass shootings to be more deadly than they were 10-20 years ago (Lankford and Silver, 2020; Duwe, 2020). Duwe et al. (2021) calculate that there is a 35% probability that a mass shooting as deadly as the 2017 Las Vegas shooting will occur again in the United States before 2040.. Future work could combine the probabilistic analysis from this dissertation on the frequency of mass shootings with a forecast of the fatalities and injuries from mass shootings.

This research compares among different forecasting models to a single data set, the historical data on mass shootings. Applying these types of forecasting models to multiple time series, especially time series on other rare events, would enable us to make stronger conclusions about the benefits and drawbacks of each modeling approach. Other time series data with similar rates of frequencies could be severe natural disasters in the United States, armed military conflicts, and fatal aviation accidents.

The research provides a base-level probability of a mass shooting at specific locations. Future work could understand how this prior probability changes if a location receives new information such as threats from an individual and individuals who exhibit mental health problems. Bayesian belief networks in which a prior probability is updated as new information arrives could be an appropriate modeling technique.

Another area of future research will be to design a risk mitigation model in order to determine the best allocation of resources to mitigate the risk of mass shootings in the United States. An optimization model could consider several potential mitigation strategies. Each mitigation strategy will have an associated cost and effectiveness. The goal will be to determine which set of strategies most effectively reduces the risk of mass shootings subject to a budget constraint.

5.3 Applications

The findings in this dissertation can be used to improve the risk assessment of mass shootings. It also can be used as a baseline for the national policymakers. They could design mitigation strategies to control the risk of mass shootings and perhaps lower the probability or the annual number of mass shootings. The risk quantification of mass shootings at specific locations enables local governments to learn which areas should be protected more again mass shootings.

Mass shootings are a serious problem in the United States, but as these results demonstrate, the number of mass shootings in the United States only averages about 6-7 shootings per year. Except for the most populous states (e.g., California, Texas), the chance of a mass shooting is less than 50% for the vast majority of states. Several states have about a 1-in-50 or 1-in-100 chance of a mass shooting in any year. The probability that a specific location experiences a mass shooting is 1-in-10,000 or less.

To what extent should the country, states, or specific locations work to reduce the risk of mass shootings? The probabilities suggest that mass shootings are rare enough that perhaps spending a significant amount of money and effort to decrease the probability may not be wise. The consequences of a mass shooting, if one does occur, include fatalities, injuries, psychological trauma, and frequently substantial coverage in the news media. Questions are raised after a mass shooting about what an organization could have done differently in order to have prevented a mass shooting. These consequences are significant enough that reducing the risk of a mass shooting may be reasonable. Implementing measures to reduce the risk of a mass shooting can also reduce the risk of other events such as shootings that do not rise to the level of a mass shooting. Without planning for and training to respond quickly to active shooter incidents, these active shooter incidents could result in more mass shooting events.

BIBLIOGRAPHY

Achcar, J., Martinez, E., Ruffino-Netto, A., Paulino, C., and Soares, P. (2008). A statistical model investigating the prevalence of tuberculosis in new york city using counting processes with two change-points. *Epidemiology & Infection*, 136(12):1599–1605.

Achcar, J. A., Loibel, S., and Andrade, M. G. (2007). Interfailure data with constant hazard function in the presence of change-points. *REVSTAT*, 5:209–226.

Achcar, J. A., Rodrigues, E. R., Paulino, C. D., and Soares, P. (2010). Non-homogeneous poisson models with a change-point: an application to ozone peaks in mexico city. *Environmental and Ecological Statistics*, 17(4):521–541.

Akaike, H. (1970). Statistical predictor identification. *Annals of the institute of Statistical Mathematics*, 22(1):203–217.

Alippi, C., Boracchi, G., Carrera, D., and Roveri, M. (2015). Change detection in multivariate datastreams: Likelihood and detectability loss. *arXiv preprint arXiv:1510.04850*.

Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367.

Andersen, M. L., Valone, S. H., Vakhitova, V. K., Chachra, V., and Sommers, P. M. (2018). The mathematics and geography of mass shootings in the us, 1982-2017. *Journal of Student Research*, 7(1):45–49.

Annis, J., Miller, B. J., and Palmeri, T. J. (2017). Bayesian inference with stan: A tutorial on adding custom distributions. *Behavior Research Methods*, 49(3):863–886.

Archdeacon, T. J. (1994). *Correlation and regression analysis: a historian's guide*. Univ of Wisconsin Press.

Arifovic, J. and Gencay, R. (2001). Using genetic algorithms to select architecture of a feedforward artificial neural network. *Physica A: Statistical Mechanics and Its Applications*, 289(3-4):574–594.

Aue, A., Hörmann, S., Horváth, L., Reimherr, M., et al. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087.

Azrak, R. and Melard, G. (1993). Exact maximum likelihood estimation for extended arima models. In *Developments in Time Series Analysis*, pages 110–123. Springer US.

Bagalman, E., Caldwell, S. W., Finklea, K. M., and McCallion, G. (2013). Public mass shootings in the United States: Selected implications for federal public health and safety policy. *Congressional Research Service*.

Bai, S., Lü, G., Wang, J., Zhou, P., and Ding, L. (2011). Gis-based rare events logistic regression for landslide-susceptibility mapping of lianyungang, china. *Environmental Earth Sciences*, 62(1):139–149.

Balesdent, M., Morio, J., and Brevault, L. (2016). Rare event probability estimation in the presence of epistemic uncertainty on input probability distribution parameters. *Methodology and Computing in Applied Probability*, 18(1):197–216.

Baron, J. N. and Reiss, P. C. (1985). Same time, next year: Aggregate analyses of the mass media and violent behavior. *American Sociological Review*, pages 347–363.

Barreto-Souza, W., de Morais, A. L., and Cordeiro, G. M. (2011). The weibull-geometric distribution. *Journal of Statistical Computation and Simulation*, 81(5):645–657.

Barron, A. (1984). Predicted squared error: a criterion for automatic model selection. self-organizing methods in modeling: Gmdh-type algorithms. forlow sj ed. *SJ Marcel-Dekker, New York*.

Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., and Pal, N. (2014). Zero-inflated poisson (zip) distribution: parameter estimation and applications to model data from natural calamities. *Involve, a Journal of Mathematics*, 7(6):751–767.

Benardos, P. and Vosniakos, G. C. (2002). Prediction of surface roughness in cnc face milling using neural networks and taguchi's design of experiments. *Robotics and Computer-Integrated Manufacturing*, 18(5-6):343–354.

Benardos, P. and Vosniakos, G. C. (2007). Optimizing feedforward artificial neural network architecture. *Engineering Applications of Artificial Intelligence*, 20(3):365–382.

Berk, R. and MacDonald, J. M. (2008). Overdispersion and poisson regression. *Journal of Quantitative Criminology*, 24(3):269–284.

Berk, R. A. and Sorenson, S. B. (2020). Algorithmic approach to forecasting rare violent events: An illustration based in intimate partner violence perpetration. *Criminology & Public Policy*, 19(1):213–233.

Berkowitz, L. and Macaulay, J. (1971). The contagion of criminal violence. *Sociometry*, pages 238–260.

Bier, V. M. and Yi, W. (1995). The performance of precursor-based estimators for rare event frequencies. *Reliability Engineering & System Safety*, 50(3):241–251.

Blair, J. P., Martaindale, M. H., and Nichols, T. (2014). Active shooter events from 2000 to 2012. *FBI Law Enforcement Bulletin*, 7.

Blair, J. P. and Schwieit, K. W. (2014). A study of active shooter incidents in the United States between 2000 and 2013. *US Department of Justice*.

BLS (2022). Unemployment rates for states. https://www.bls.gov/web/laus/laumstrk.htm.

Blum, D. and Jaworski, C. G. (2017). Spatial patterns of mass shootings in the united states, 2013–2014. In *Environmental Criminology*. Emerald Publishing Limited.

Boerboom, S., Pekel, K., and Dodds, M. (2020). Mass shooting data amp; research. https://www.theviolenceproject.org/mass-shooter-database.

Böhning, D. (1998). Zero-inflated poisson models and ca man: A tutorial collection of evidence. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 40(7):833–843.

Bosman, J., Smith, M., and Vigdor, N. (2020). 5 killed in shooting at molson coors in milwaukee. https://www.nytimes.com/2020/02/26/us/milwaukee-shooting-miller-coors.html.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2011). *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons.

Boyd Sr, B. R. (2011). *Mass murder in the new millennium a thesis restructuring mass murder typologies*. Diploma thesis.

Brockwell, P. J., Davis, R. A., and Calder, M. V. (2002). *Introduction to time series and forecasting*, volume 2. Springer.

Browne, K. D. and Hamilton-Giachritsis, C. (2005). The influence of violent media on children and adolescents: a public-health approach. *The Lancet*, 365(9460):702–710.

Cabrera, J. F. and Kwon, R. (2018). Income inequality, household income, and mass shooting in the united states. *Frontiers in public health*, 6:294.

Calabrese, R. and Elkink, J. A. (2016). Estimating binary spatial autoregressive models for rare events. In *Spatial Econometrics: Qualitative and Limited Dependent Variables*. Emerald Group Publishing Limited.

California Courts (2020). Gun violence restraining orders. https://www.courts.ca.gov/33961.htm?rdeLocaleAttr=en.

California Department of Education (2020). Enrollment: Other choices level. https://dq.cde.ca.gov/dataquest/DQ/enrotherstart.aspx.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Chabridon, V., Balesdent, M., Bourinet, J.-M., Morio, J., and Gayton, N. (2018). Reliability-based sensitivity estimators of rare event probability in the presence of distribution parameter uncertainty. *Reliability Engineering & System Safety*, 178:164–178.

Chamlin, M. B. (1988). Crime and arrests: An autoregressive integrated moving average (arima) approach. *Journal of Quantitative Criminology*, 4(3):247–258.

Chattopadhyay, P. B. and Rangarajan, R. (2014). Application of ann in sketching spatial nonlinearity of unconfined aquifer in agricultural basin. *Agricultural Water Management*, 133:81–91.

Chen, C. and Tiao, G. C. (1990). Random level-shift time series models, arima approximations, and level-shift detection. *Journal of Business & Economic Statistics*, 8(1):83–97.

Chen, P., Yuan, H., and Shu, X. (2008). Forecasting crime using the arima model. In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 5, pages 627–630. IEEE.

Choe, W., Ersoy, O. K., and Bina, M. (2000). Neural network schemes for detecting rare events in human genomic dna. *Bioinformatics*, 16(12):1062–1072.

Christensen, J. (2017). Why the US has the most mass shootings. https://www.cnn.com/2015/08/27/health/u-s-most-mass-shootings/index.html.

Cleland, I., Han, M., Nugent, C., Lee, H., McClean, S., Zhang, S., and Lee, S. (2014). Evaluation of prompted annotation of activity data recorded from a smart phone. *Sensors*, 14(9):15861–15879.

Cohen, A. P., Azrael, D., and Miller, M. (2014). Rate of mass shootings has tripled since 2011, new research from Harvard shows. https://www.motherjones.com/politics/2014/10/mass-shootings-increasing-harvard-research/.

Coleman, J. S. et al. (1964). Introduction to mathematical sociology. *Introduction to mathematical sociology*.

Contreras, J., Espinola, R., Nogales, F. J., and Conejo, A. J. (2003). Arima models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18(3):1014–1020.

Cook, R. J. and Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

Coughlan, S. (2018). 2018 'worst year for us school shootings'. https://www.bbc.com/news/business-46507514.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403.

Cruz-Juárez, J. A., Reyes-Cervantes, H., and Rodrigues, E. R. (2016). Analysis of ozone behaviour in the city of Puebla–Mexico using non-homogeneous poisson models with multiple change-points. *Journal of Environmental Protection*, 7(12):1886.

D'anna, M. (2020). *Black Swan Shootings: A Model for Predicting the Worst of the Worst Mass Shootings*. PhD thesis, George Mason University.

Densley, J. and Peterson, J. (2019a). Opinion: We analyzed 53 years of mass shooting data. attacks aren't just increasing, they're getting deadlier. https://www.latimes.com/opinion/story/2019-09-01/mass-shooting-data-odessa-midland-increase.

Densley, J. and Peterson, J. (2019b). Opinion: We analyzed 53 years of mass shooting data. attacks aren't just increasing, they're getting deadlier.

Dent, W. (1977). Computation of the exact likelihood function of an arima process. *Journal of Statistical Computation and Simulation*, 5(3):193–206.

DHS (2014). US Department of Homeland Security active shooter response. https://www.dhs.gov/xlibrary/assets/active_shooter_booklet.pdf.

Dietz, P. E. (1986). Mass, serial and sensational homicides. *Bulletin of the New York Academy of Medicine*, 62(5):477.

Dillon, L. (2014). *Mass shootings in the United States: An exploratory study of the trends from 1982-2012*. PhD thesis.

DiMaggio, C., Avraham, J., Berry, C., Bukur, M., Feldman, J., Klein, M., Shah, N., Tandon, M., and Frangos, S. (2019). Changes in US mass shooting deaths associated with the 1994–2004 federal assault weapons ban: Analysis of open-source data. *Journal of Trauma and Acute Care Surgery*, 86(1):11–19.

Do, T.-T., Doan, A.-D., and Cheung, N.-M. (2016). Learning to hash with binary deep neural network. In *European Conference on Computer Vision*, pages 219–234. Springer.

Dolliver, M. J. and Kearns, E. M. (2019). Is it terrorism?: Public perceptions, media, and labeling the las vegas shooting. *Studies in Conflict & Terrorism*, pages 1–19.

Dugan, L. (2011). The series hazard model: An alternative to time series for event data. *Journal of Quantitative Criminology*, 27(3):379–402.

Dugan, L., LaFree, G., and Piquero, A. R. (2005). Testing a rational choice model of airline hijackings. In *International Conference on Intelligence and Security Informatics*, pages 340–361. Springer.

Duwe, G. (2014). *Mass murder in the United States: A history.* McFarland.

Duwe, G. (2020). Patterns and prevalence of lethal mass violence. *Criminology & Public Policy*, 19(1):17–35.

Duwe, G., Kovandzic, T., and Moody, C. E. (2002). The impact of right-to-carry concealed firearm laws on mass public shootings. *Homicide Studies*, 6(4):271–296.

Duwe, G., Sanders, N. E., Rocque, M., and Fox, J. A. (2021). Forecasting the severity of mass public shootings in the united states. *Journal of Quantitative Criminology*, pages 1–39.

El-Gheriani, M., Khan, F., and Zuo, M. J. (2017). Rare event analysis considering data and model uncertainty. *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, 3(2).

El-Sharkawi, M., Oh, S., Marks, R., Damborg, M., and Brace, C. (1991). Short term electric load forecasting using an adaptively trained layered perceptron. In *Proc. of First International Forum on ANNPS*, pages 3–6.

Faruk, D. Ö. (2010). A hybrid neural network and arima model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4):586–594.

FBI (2016). Active shooter resources. https://www.fbi.gov/about/partnerships/office-of-partner-engagement/active-shooter-resources.

Feuz, K. D., Cook, D. J., Rosasco, C., Robertson, K., and Schmitter-Edgecombe, M. (2014). Automated detection of activity transitions for prompting. *IEEE Transactions on Human-Machine Systems*, 45(5):575–585.

Follman, M., Aronsen, G., and Pan, D. (2012). US mass shootings, 1982-2020: Data from Mother Jones' investigation. https://www.motherjones.com/politics/2012/12/mass-shootings-mother-jones-full-data/.

Fong, D. Y. and Yip, P. (1993). An em algorithm for a mixture model of count data. *Statistics & probability letters*, 17(1):53–60.

Fong, S. and Deb, S. (2015). Prediction of major earthquakes as rare events using rf-typed polynomial neural networks. In *Encyclopedia of Information Science and Technology, Third Edition*, pages 227–238. IGI Global.

Fong, S., Nannan, Z., Wong, R. K., and Yang, X.-S. (2012). Rare events forecasting using a residual-feedback gmdh neural network. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 464–473. IEEE.

Forbes, F., Charras-Garrido, M., Azizi, L., Doyle, S., and Abrial, D. (2013). Spatial risk mapping for rare disease with hidden markov fields and variational em. *The Annals of Applied Statistics*, pages 1192–1216.

Freilich, J. D., Chermak, S. M., and Klein, B. R. (2020). Investigating the applicability of situational crime prevention to the public mass violence context. *Criminology & Public Policy*, 19(1):271–293.

Fridel, E. E. (2020). Comparing the impact of household gun ownership and concealed carry legislation on the frequency of mass shootings and firearms homicide. *Justice Quarterly*, pages 1–24.

Fridel, E. E. (2021). Comparing the impact of household gun ownership and concealed carry legislation on the frequency of mass shootings and firearms homicide. *Justice quarterly*, 38(5):892–915.

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological bulletin*, 118(3):392.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gius, M. (2015). The impact of state and federal assault weapons bans on public mass shootings. *Applied Economics Letters*, 22(4):281–284.

Glorot, X., Bordes, A., and Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323.

Goel, A. and Okumoto, K. (1978). An analysis of recurrent software failures on a real-time control system. In *Proceedings of ACM Conference*, pages 496–500.

Goel, A. L. (1985). Software reliability models: Assumptions, limitations, and applicability. *IEEE Transactions on Software Engineering*, (12):1411–1423.

Goodwin, P. and Wright, G. (2010). The limits of forecasting methods in anticipating rare events. *Technological forecasting and social change*, 77(3):355–368.

Griffith, D. A. and Haining, R. (2006). Beyond mule kicks: The poisson distribution in geographical analysis. *Geographical Analysis*, 38(2):123–139.

Grossberg, S. and Merrill, J. W. (1992). A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*, 1(1):3–38.

Grossi, E. and Buscema, M. (2007). Introduction to artificial neural networks. *European Journal of Gastroenterology & Hepatology*, 19(12):1046–1054.

Guarnaccia, C., Quartieri, J., Tepedino, C., and Rodrigues, E. R. (2016). A time series analysis and a non-homogeneous poisson model with multiple change-points applied to acoustic data. *Applied acoustics*, 114:203–212.

Gun Violence Archive (2020a). Gun violence archive. https://www.gunviolencearchive.org/.

Gun Violence Archive (2020b). Gun violence archive explainer. https://www.gunviolencearchive.org/explainer.

Guns, M. and Vanacker, V. (2012). Logistic regression applied to natural hazards: rare event logistic regression with replications. *Natural Hazards and Earth System Sciences*, 12(6):1937–1947.

Gutierrez, G., Atencio, M., Chuck, E., Abdelkader, R., and Walters, S. (2017). About half of the texas church shooting victims were children. https://www.nbcnews.com/storyline/texas-church-shooting/about-half-victims-texas-church-shooting-were-children-n817896.

Gyarmati-Szabó, J., Bogachev, L. V., and Chen, H. (2011). Modelling threshold exceedances of air pollution concentrations via non-homogeneous poisson process with multiple change-points. *Atmospheric environment*, 45(31):5493–5503.

Hamilton, L. C. and Hamilton, J. D. (1983). Dynamics of terrorism. *International Studies Quarterly*, 27(1):39–54.

Han, M., Lee, Y.-K., Lee, S., et al. (2012). Comprehensive context recognizer based on multimodal sensors in a smartphone. *Sensors*, 12(9):12588–12605.

Haykin, S. (1999). Neural networks, a comprehensive foundation, prentice-hall inc. *Upper Saddle River, New Jersey*, 7458:161–175.

Heilbron, D. (1989). Generalized linear models for altered zero probabilities and overdispersion in count data. *Unpublished Technical report, University of California, San Francisco, Department of Epidemiology and Biostatistics.*

Heilbron, D. C. (1994). Zero-altered and other regression models for count data with added zeros. *Biometrical Journal*, 36(5):531–547.

Hilbe, J. M. (2011). *Negative binomial regression.* Cambridge University Press.

Ho, C.-H. and Bhaduri, M. (2015). On a novel approach to forecast sparse rare events: applications to parkfield earthquake prediction. *Natural Hazards*, 78(1):669–679.

Holmes, R. M. and Holmes, S. T. (2001). *Murder in America*. Sage.

Hubbard, D. (2009). *The Future of Risk Management: Why It's Broken and How to Fix it*. John Wiley & Sons.

Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688.

Iowa Community Indicators Program (2021). Annual unemployment rates by state. https://www.icip.iastate.edu/tables/employment/unemployment-states.

Iowa Department of Education (2020). Public school building. https://educateiowa.gov/document-type/public-school-building-prek-12-enrollment-grade-race-and-gender.

Islam, K. and Raza, A. (2020). Forecasting crime using arima model. *arXiv preprint arXiv:2003.08006*.

Jonson, C. L. (2017). Preventing school shootings: The effectiveness of safety measures. *Victims & Offenders*, 12(6):956–973.

Kaplan, S. and Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, 1(1):11–27.

Keenen, A. (2019). Patterns in the locations of US mass shootings. Available at https://shareok.org/handle/11244/323404.

Kelly, R. W. (2012a). *Active Shooter: Recommendations and analysis for risk mitigation*. New York City Police Dept and United States of America. https://www.calhospitalprepare.org/post/active-shooter-recommendations-and-analysis-risk-mitigation.

Kelly, R. W. (2012b). *Active Shooter: Recommendations and analysis for risk mitigation*. New York City Police Dept and United States of America.

Kim, D.-Y. and Phillips, S. W. (2021). When covid-19 and guns meet: A rise in shootings. *Journal of Criminal Justice*, 73:101783.

Kim, S. and Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679.

King, D. M. and Jacobson, S. H. (2017). Random acts of violence? examining probabilistic independence of the temporal distribution of mass killing events in the united states. *Violence and victims*, 32(6):1014–1023.

Krouse, W. J. and Richardson, D. J. (2015). Mass murder with firearms: Incidents and victims, 1999-2013. Congressional Research Service, Library of Congress. https://www.hsdl.org/?abstract&did=787620.

Kwon, R. and Cabrera, J. F. (2019a). Income inequality and mass shootings in the united states. *BMC public health*, 19(1):1–8.

Kwon, R. and Cabrera, J. F. (2019b). Socioeconomic factors and mass shootings in the united states. *Critical Public Health*, 29(2):138–145.

LaFree, G., Dugan, L., and Korte, R. (2009). The impact of british counterterrorist strategies on political violence in northern ireland: Comparing deterrence and backlash models. *Criminology*, 47(1):17–45.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Lankford, A. (2013a). A comparative analysis of suicide terrorists and rampage, workplace, and school shooters in the United States from 1990 to 2010. *Homicide Studies*, 17(3):255–274.

Lankford, A. (2013b). A comparative analysis of suicide terrorists and rampage, workplace, and school shooters in the United States from 1990 to 2010. *Homicide Studies*, 17(3):255–274.

Lankford, A. (2015). Mass shooters in the USA, 1966–2010: Differences between attackers who live and die. *Justice Quarterly*, 32(2):360–379.

Lankford, A. (2016). Fame-seeking rampage shooters: Initial findings and empirical predictions. *Aggression and Violent Behavior*, 27:122–129.

Lankford, A. and Silver, J. (2020). Why have public mass shootings become more deadly? assessing how perpetrators' motives and methods have changed over time. *Criminology & Public Policy*, 19(1):37–60.

Lee, J. H. (2013). School shootings in the US public schools: Analysis through the eyes of an educator. *Review of Higher Education & Self-Learning*, 6(22).

Lemieux, F. (2014). Effect of gun culture and firearm laws on gun violence and mass shootings in the United States: A multi-level quantitative analysis. *International Journal of Criminal Justice Sciences*, 9(1).

Levin, J. and Madfis, E. (2009). Mass murder at school and cumulative strain: A sequential model. *American Behavioral Scientist*, 52(9):1227–1245.

Li, Q. (2015). *Recurrent-Event Models for Change-Points Detection*. PhD thesis, Virginia Tech. https://vtechworks.lib.vt.edu/bitstream/handle/10919/78207/Li_Q_D_2015.pdf?isAllowed=y&sequence=1.

Li, Q., Guo, F., and Kim, I. (2020a). A non-parametric bayesian change-point detection method in the recurrent-event context. *Journal of Statistical Computation and Simulation*.

Li, Q., Guo, F., Kim, I., Klauer, S. G., and Simons-Morton, B. G. (2018). A bayesian finite mixture change-point model for assessing the risk of novice teenage drivers. *Journal of Applied Statistics*, 45(4):604–625.

Li, Q., Guo, F., Klauer, S. G., and Simons-Morton, B. G. (2017). Evaluation of risk change-point for novice teenage drivers. *Accident Analysis & Prevention*, 108:139–146.

Li, Q., Liu, L., Li, T., and Yao, K. (2021). Bayesian change-points detection assuming a power law process in the recurrent-event context. *Communications in Statistics-Simulation and Computation*, pages 1–23.

Li, Q., Yao, K., and Zhang, X. (2020b). A change-point detection and clustering method in the recurrent-event context. *Journal of Statistical Computation and Simulation*, 90(6):1131–1149.

Lin, C.-W. and Wang, J.-S. (2008). A digital circuit design of hyperbolic tangent sigmoid function for neural networks. In *2008 IEEE International Symposium on Circuits and Systems*, pages 856–859. IEEE.

Lin, P.-I., Fei, L., Barzman, D., and Hossain, M. (2018). What have we learned from the time trend of mass shootings in the US? *PloS one*, 13(10):e0204722.

Liu, C., Hoi, S. C., Zhao, P., and Sun, J. (2016). Online arima algorithms for time series prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Lu, J.-C., Niu, D.-X., and Jia, Z.-Y. (2004). A study of short-term load forecasting based on arima-ann. In *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826)*, volume 5, pages 3183–3187. IEEE.

Luca, M., Malhotra, D., and Poliquin, C. (2020). The impact of mass shootings on gun policy. *Journal of Public Economics*, 181:104083.

Macaya, M. and Hayes, M. (2021). March 23, 2021 colorado mass shooting news. https://www.cnn.com/us/live-news/boulder-colorado-shooting-3-23-21/index.html.

Martin, S. L., Stohs, S. M., and Moore, J. E. (2015). Bayesian inference and assessment for rare-event bycatch in marine fisheries: a drift gillnet fishery case study. *Ecological Applications*, 25(2):416–429.

McNeal, R. S., Schmeida, M., Bryan, L. D., and Kunkle, S. M. (2020). State-level policy response to mass shootings: A timeseries analysis. In *Handbook of Research on Mass Shootings and Multiple Victim Violence*, pages 397–417. IGI Global.

McPhedran, S. and Baker, J. (2011). Mass shootings in australia and new zealand: A descriptive study of incidence. *Justice Policy Journal*, 8(1).

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.

Moksony, F. and Hegedűs, R. (2015). The use of poisson regression in the sociological study of suicide. *Corvinus Journal of Sociology and Social Policy*, 5(2).

Moniz, N., Branco, P., and Torgo, L. (2016). Resampling strategies for imbalanced time series. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 282–291. IEEE.

Moody, J. and Utans, J. (1994). Architecture selection strategies for neural networks: Application to corporate bond rating prediction. In *Neural Networks in the Capital Markets*, pages 277–300. Citeseer.

Moody, J. E. (1991). Note on generalization, regularization and architecture selection in nonlinear learning systems. In *Neural Networks for Signal Processing Proceedings of the 1991 IEEE Workshop*, pages 1–10. IEEE.

Mourgias-Alexandris, G., Tsakyridis, A., Passalis, N., Tefas, A., Vyrsokinos, K., and Pleros, N. (2019). An all-optical neuron with sigmoid activation function. *Optics Express*, 27(7):9620–9630.

Muchlinski, D., Siroky, D., He, J., and Kocher, M. (2016). Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Political Analysis*, pages 87–103.

Mudholkar, G. S., Srivastava, D. K., and Freimer, M. (1995). The exponentiated weibull family: A reanalysis of the bus-motor-failure data. *Technometrics*, 37(4):436–445.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of econometrics*, 33(3):341–365.

Musa, J. D. and Okumoto, K. (1984). A logarithmic poisson execution time model for software reliability measurement. In *Proceedings of the 7th international conference on Software engineering*, pages 230–238. Citeseer.

Mushtaq, R. (2011). Augmented dickey fuller test.

Muss, J., Iannino, A., and Okumoto, K. (1987). Software reliability: Measurement, prediction, application.

Nagin, D. S., Koper, C. S., and Lum, C. (2020). Policy recommendations for countering mass shootings in the United States. *Criminology & Public Policy*, 19(1):9–15.

Nance, M. L., DeSimone, J. D., Lorch, S. A., Passarella, M., Cronin, K. M., Kreinces, J., and Myers, S. R. (2020). Locations of mass shootings relative to schools and places frequented by children. *JAMA pediatrics*, 174(11):1109–1110.

Newman, K. and Fox, C. (2009). Repeat tragedy: Rampage shootings in american high school and college settings, 2002-2008. *American Behavioral Scientist*, 52(9):1286–1308.

O'Neill, J., Miller, J., and Waters, J. (2016). Active shooter: recommendations and analysis for risk mitigation. https://www.calhospitalprepare.org/post/active-shooter-recommendations-and-analysis-risk-mitigation.

Orong, M. Y., Sison, A. M., and Hernandez, A. A. (2018). Mitigating vulnerabilities through forecasting and crime trend analysis. In *2018 5th International Conference on Business and Industrial Research (ICBIR)*, pages 57–62. IEEE.

Palermo, G. B. and Ross, L. E. (1999). Mass murder, suicide, and moral development: Can we separate the adults from the juveniles? *International Journal of Offender Therapy and Comparative Criminology*, 43(1):8–20.

Patra, J. C., Panda, G., and Baliarsingh, R. (1994). Artificial neural network-based nonlinearity estimation of pressure sensors. *IEEE Transactions on Instrumentation and Measurement*, 43(6):874–881.

Pavlou, M., Ambler, G., Seaman, S. R., Guttmann, O., Elliott, P., King, M., and Omar, R. Z. (2015). How to develop a more accurate risk prediction model when there are few events. *Bmj*, 351.

Payne, J. and Morgan, A. (2020). Covid-19 and violent crime: A comparison of recorded offence rates and dynamic forecasts (arima) for march 2020 in queensland, australia.

Phaisangittisagul, E. (2016). An analysis of the regularization between l2 and dropout in single hidden layer neural network. In *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, pages 174–179. IEEE.

Phillips, D. P. (1983). The impact of mass media violence on US homicides. *American Sociological Review*, pages 560–568.

Pisa, I., Santín, I., Vicario, J. L., Morell, A., and Vilanova, R. (2019). Data preprocessing for ann-based industrial time-series forecasting with imbalanced data. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pages 1–5. IEEE.

R-bloggers (2015). Prediction intervals for poisson regression. https://www.r-bloggers.com/2015/12/prediction-intervals-for-poisson-regression/.

Raftery, A. E. and Akman, V. (1986). Bayesian analysis of a poisson process with a change-point. *Biometrika*, 73(1):85–89.

Raftery, A. E., Newton, M. A., Satagopan, J. M., and Krivitsky, P. N. (2006). Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.

Ray, M. (2022). Sandy hook elementary school shooting. https://www.britannica.com/event/Sandy-Hook-Elementary-School-shooting.

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13.

Reeping, P. M., Cerdá, M., Kalesan, B., Wiebe, D. J., Galea, S., and Branas, C. C. (2019). State gun laws, gun ownership, and mass shootings in the us: cross sectional time series. *bmj*, 364.

Reid, J. A. and Beauregard, E. (2020). Exploring a machine learning approach: Predicting death in sexual assault. *Journal of Criminal Justice*, 71:101741.

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4):515–530.

Rossi, R. J. (2018). *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons.

Sane, W. (1994). Neural networks and statistical models. In *Proc. of the Nineteenth Annual SASUsers Group Inter-national Conference*.

Shatanawi, K., Rahbeh, M., and Shatanawi, M. (2013). Characterizing, monitoring and forecasting of drought in jordan river basin. *Journal of Water Resource and Protection*, 2013.

Sherburne, E. G. (2003). Bureau of alcohol, tobacco, firearms and explosives. *Advanced Emergency Nursing Journal*, 25(4):294.

Skeem, J. and Mulvey, E. (2020). What role does serious mental illness play in mass shootings, and how should we address it? *Criminology & Public Policy*, 19(1):85–108.

Smart, R. (2018). Mass shootings: definitions and trends. rand corporation. https://www.rand.org/research/gun-policy/analysis/essays/mass-shootings.html.

Smith, T. G. (2017-2020). pmdarima. https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html.

Soni, A. and Tekin, E. (2020). How do mass shootings affect community wellbeing? Technical report, National Bureau of Economic Research.

Stack, S. (1989). The effect of publicized mass murders and murder-suicides on lethal violence, 1968–1980: A research note. *Social Psychiatry and Psychiatric Epidemiology*.

Statista (2020). Worst mass shootings in the u.s. 2019. https://www.statista.com/statistics/476101/worst-mass-shootings-in-the-us/.

Stergiou, K. (1989). Modelling and forecasting the fishery for pilchard (sardina pilchardus) in greek waters using arima time-series models. *ICES Journal of Marine Science*, 46(1):16–23.

The Violence Project (2020). Mass shooting database locations. https://www.theviolenceproject.org/mass-shooter-database/.

Theofilatos, A., Yannis, G., Kopelias, P., and Papadimitriou, F. (2016). Predicting road accidents: a rare-events modeling approach. *Transportation research procedia*, 14:3399–3405.

Thompson, W. J. (2001). Poisson distributions. *Computing in Science & Engineering*, 3(3):78–82.

Towers, S., Wallace, D., and Hemenway, D. (2019). Temporal trends in public mass shootings: high-capacity magazines significantly increase fatality counts, and are becoming more prevalent. *medRxiv*.

Tseng, F.-M., Yu, H.-C., and Tzeng, G.-H. (2002). Combining neural network model with seasonal time series arima model. *Technological Forecasting and Social Change*, 69(1):71–87.

UC Davis Health (2018). Deaths from firearms in the U.S., 2018. https://health.ucdavis.edu/what-you-can-do/facts.html.

U.S. Census Bureau (2021). State population totals: 2010-2020. https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-state-total.html.

Van Der Voort, M., Dougherty, M., and Watson, S. (1996). Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5):307–318.

Wang, L., Zou, H., Su, J., Li, L., and Chaudhry, S. (2013). An arima-ann hybrid model for time series forecasting. *Systems Research and Behavioral Science*, 30(3):244–259.

Wang, S.-C. (2003). Artificial neural network. In *Interdisciplinary Computing in Java Programming*, pages 81–100. Springer.

Webster, D. W. (2017). The true effect of mass shootings on americans. *Annals of Internal Medicine*, 166(10):749–750.

Webster, D. W., McCourt, A. D., Crifasi, C. K., Booty, M. D., and Stuart, E. A. (2020). Evidence concerning the regulation of firearms design, sale, and carrying on fatal mass shootings in the united states. *Criminology & Public Policy*, 19(1):171–212.

Wei, L. and Keogh, E. (2006). Semi-supervised time series classification. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 748–753. ACM.

Winahju, W. S. and Irhamah, I. (2016). Modeling the rare event using bivariate poisson integer autocorrelation. In *The 1st International Conference on Mathematics: Education, Theory & Application*.

Wintemute, G. J., Pear, V. A., Schleimer, J. P., Pallin, R., Sohl, S., Kravitz-Wirtz, N., and Tomsich, E. A. (2019). Extreme risk protection orders intended to prevent mass shootings: a case series. *Annals of internal medicine*, 171(9):655–658.

Xue, L., Cameron, M., and Qing, L. (). Analysis and forecasting of mass shootings using change point detection. Unpublished manuscript.

Yamanishi, K. and Takeuchi, J.-i. (2002). A unifying framework for detecting outliers and change points from non-stationary time series data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 676–681.

Ying, X. (2019). An overview of overfitting and its solutions. In *Journal of Physics: Conference Series*, volume 1168, page 022022. IOP Publishing.

Yip, P. (1991). Conditional inference on a mixture model for the analysis of count data. *Communications in Statistics-Theory and Methods*, 20(7):2045–2057.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in r. *Journal of statistical software*, 27(8):1–25.

Zhang, G. P. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159–175.

Zhang, J. and Small, M. (2006). Complex network from pseudoperiodic time series: Topology versus dynamics. *Physical Review Letters*, 96(23):238701.

Zhang, L., Zhang, G., and Li, R. (2018). Water quality analysis and prediction using hybrid time series and neural network models. http://ir.jkuat.ac.ke/bitstream/handle/123456789/3925/JAST_Volume%2018_Issue%204_Pages%20975-983.pdf?sequence=1&isAllowed=y.

Zheng, Y., Liu, L., Wang, L., and Xie, X. (2008). Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web*, pages 247–256. ACM.

APPENDIX A. Modeling and Forecasting Mass Shootings Using Poisson Regression and Change-Point Models

Table A.1 and A.2 show the annual count data and the time data of mass shootings from the Violence Project. Table A.3, A.4, and A.5 present the posteriors summaries of parameters in the different rate functions for the change-point models with different numbers of change points.

Year	Number of mass shootings	Year	Number of mass shootings
1966	2	1993	6
1967	1	1994	2
1968	1	1995	3
1969	1	1996	2
1970	1	1997	4
1971	0	1998	4
1972	2	1999	8
1973	1	2000	3
1974	0	2001	4
1975	1	2002	1
1976	1	2003	4
1977	3	2004	3
1978	1	2005	4
1979	0	2006	4
1980	3	2007	4
1981	2	2008	5
1982	3	2009	5
1983	3	2010	5
1984	4	2011	4
1985	1	2012	6
1986	1	2013	5
1987	1	2014	3
1988	3	2015	6
1989	2	2016	5
1990	1	2017	7
1991	5	2018	8
1992	4	2019	8

Table A.1: Year count data of mass shootings from the Violence Project

ith event	t_i(days)	ith event	t_i(days)	ith event	t_i(days)	ith event	t_i(days)	ith event	t_i(days)
1	0	36	8081	71	11944	106	15163	141	17988
2	101	37	8201	72	11944	107	15197	142	18006
3	447	38	8438	73	11987	108	15294	143	18006
4	595	39	8717	74	12043	109	15361	144	18089
5	979	40	9194	75	12089	110	15573	145	18108
6	1512	41	9200	76	12136	111	15577	146	18201
7	2128	42	9215	77	12194	112	15785	147	18226
8	2150	43	9223	78	12279	113	15789	148	18302
9	2351	44	9228	79	12317	114	15813	149	18410
10	3136	45	9354	80	12555	115	15942	150	18440
11	3631	46	9400	81	12573	116	16005	151	18486
12	3848	47	9564	82	12599	117	16062	152	18559
13	4007	48	9587	83	12747	118	16073	153	18675
14	4040	49	9825	84	12812	119	16100	154	18709
15	4336	50	9860	85	13011	120	16222	155	18718
16	4932	51	9928	86	13349	121	16431	156	18797
17	5071	52	9976	87	13482	122	16460	157	18813
18	5100	53	9981	88	13531	123	16496	158	18825
19	5391	54	9988	89	13588	124	16671	159	18881
20	5550	55	10179	90	13841	125	16729	160	18907
21	5752	56	10370	91	13980	126	16779	161	18947
22	5848	57	10467	92	13997	127	16794	162	19021
23	5859	58	10573	93	14095	128	16846	163	19066
24	6027	59	10723	94	14096	129	16923	164	19076
25	6055	60	10778	95	14105	130	17017	165	19157
26	6275	61	10853	96	14262	131	17055	166	19179
27	6496	62	11333	97	14419	132	17101	167	19285
28	6538	63	11359	98	14474	133	17150	168	19347
29	6557	64	11437	99	14530	134	17200	169	19348
30	6563	65	11452	100	14661	135	17359	170	19375
31	6800	66	11535	101	14796	136	17452	171	19474
32	7319	67	11553	102	14860	137	17603	172	19474
33	7567	68	11553	103	15089	138	17841	173	19555
34	7865	69	11609	104	15093	139	17870		
35	8016	70	11904	105	15156	140	17945		

Table A.2: Occurring time data of mass shootings from the Violence Project

Rate function	Parameters-posterior means(standard deviation)
Constant	$\lambda - 0.0088(0.0007)$
PLP	$\alpha - 1.72(0.13)$, $\beta - 993.54(225.34)$
MO	$\alpha - 7906.57(1055.53)$, $\beta - 98.08(1.85)$
GO	$\alpha - 876.98(98.63)$, $\beta - 0.0011(0.0002)$
GGO	$\alpha - 45440.12(52173.83)$, $\beta - 8.40 * 10^{-09}(2.36 * 10^{-08})$, $\gamma - 1.56(0.12)$
WG	$\alpha - 1.86(0.16)$, $\beta - 1260.91(300.36)$, $p - 0.79(0.21)$

Table A.3: Posteriors summaries of parameters in different rate functions for the change-point model with no change point

Rate function	Parameters-posterior means(standard deviation)
Constant	$\lambda_1 - 0.0043(0.0007)$, $\lambda_2 - 0.0127(0.0011)$, $\tau_1 - 9026.53(248.69)$
PLP	$\alpha_1 - 41.43(32.78)$, $\beta_1 - 79585.08(61862.00)$, $\alpha_2 - 1.79(0.17)$, $\beta_2 - 1128.97(318.33)$, $\tau_1 - 460.51(1237.53)$
MO	$\alpha_1 - 13747.71(3272.28)$, $\beta_1 - 71.70(14.41)$, $\alpha_2 - 14178.00(3652.14)$, $\beta_2 - 352.89(41.93)$, $\tau_1 - 9093.93(130.56)$
GO	$\alpha_1 - 438.75(263.64)$, $\beta_1 - 0.0017(0.0016)$, $\alpha_2 - 1818.67(906.03)$, $\beta_2 - 0.0010(0.0005)$, $\tau_1 - 9091.01(114.13)$
GGO	$\alpha_1 - 3550.63(18185.56)$, $\beta_1 - 35.46(32.68)$, $\gamma_1 - 35.07(32.90)$, $\alpha_2 - 49674.64(51936.29)$, $\beta_2 - 6.58 * 10^{-09}(1.94 * 10^{-08})$, $\gamma_2 - 1.62(0.14)$, $\tau_1 - 506.19(156.36)$
WG	$\alpha_1 - 51.20(30.16)$, $\beta_1 - 98880.81(57616.13)$, $p_1 - 0.47(0.27)$, $\alpha_2 - 1.89(0.15)$, $\beta_2 - 1308.44(294.81)$, $p_2 - 0.82(0.20)$, $\tau_1 - 96.38(485.82)$

Table A.4: Posteriors summaries of parameters in different rate functions for the change-point model with one change point

Rate function	Parameters-posterior means(standard deviation)
Constant	$\lambda_1 - 0.0048(0.0009)$, $\lambda_2 - 0.0111(0.0006)$, $\lambda_3 - 0.0209(0.003)$, $\tau_1 - 9157.97(376.81)$, $\tau_2 - 17937.44(549.12)$
PLP	$\alpha_1 - 50.57(27.05)$, $\beta_1 - 974532.73(562225.95)$, $\alpha_2 - 51.72(27.53)$, $\beta_2 - 1051286.15(493964.04)$, $\alpha_3 - 1.73(0.13)$, $\beta_3 - 1011.23(231.34)$, $\tau_1 - 32.49(23.19)$, $\tau_2 - 67.28(22.40)$
MO	$\alpha_1 - 57170.27(22612.39)$, $\beta_1 - 201.99(96.84)$, $\alpha_2 - 61697.46(20971.30)$, $\beta_2 - 358.94(137.58)$, $\alpha_3 - 43301.30(6822.99)$, $\beta_3 - 726.56(79.63)$, $\tau_1 - 3917.12(1662.16)$, $\tau_2 - 9080.51(86.18)$
GO	$\alpha_1 - 11.78(18.85)$, $\beta_1 - 1.16(1.40)$, $\alpha_2 - 499.97(278.46)$, $\beta_2 - 49518.30(30111.85)$, $\alpha_3 - 900.18(85.01)$, $\beta_3 - 0.0011(0.0002)$, $\tau_1 - 224.22(82.82)$, $\tau_2 - 370.67(66.66)$
GGO	$\alpha_1 - 1.00(1.01)$, $\beta_1 - 57.28(25.66)$, $\gamma_1 - 56.66(26.93)$, $\alpha_2 - 110540.7(55982.53)$, $\beta_2 - 55.84(26.53)$, $\gamma_2 - 57.82(26.57)$, $\alpha_3 - 37165.87(44008.65)$, $\beta_3 - 8.03 * 10^{-09}(1.95 * 10^{-08})$, $\gamma_3 - 1.55(0.11)$, $\tau_1 - 23.56(21.71)$, $\tau_2 - 62.41(26.15)$
WG	$\alpha_1 - 4914.21(2895.44)$, $\beta_1 - 965193.41(577814.04)$, $p_1 - 0.524(0.28)$, $\alpha_2 - 5065.02(2806.05)$, $\beta_2 - 1043789.78(529375.10)$, $p_2 - 0.47(0.27)$, $\alpha_3 - 1.92(0.17)$, $\beta_3 - 1372.41(323.28)$, $p_3 - 0.82(0.25)$, $\tau_1 - 38.11(22.78)$, $\tau_2 - 69.29(21.74)$

Table A.5: Posteriors summaries of parameters in different rate functions for the change-point model with two change points

APPENDIX B. Comparing Different Models to Forecast the Number of Mass Shootings in the United States: An Application of Forecasting Rare Event Time Series Data

Table B.1, Table B.2, Table B.3 and Table B.4 present the architecture selection results of the ANN model with different maximum number of input nodes for the different training sets.

Figure B.1 depicts the estimation and forecasting by models and real observations when we use different sizes of training set.

Training set observation period	The nubmer of inputs	The number of hidden nodes
1966-2003	1	3
1966-2004	3	3
1966-2005	2	3
1966-2006	4	2
1966-2007	3	3
1966-2008	2	3
1966-2009	2	3
1966-2010	4	3
1966-2011	3	3
1966-2012	3	3
1966-2013	1	3
1966-2014	1	3

Table B.1: Architecture selection results of the ANN model for different training sets with maximum three input nodes

Training set observation period	The nubmer of inputs	The number of hidden nodes
1966-2003	4	4
1966-2004	4	5
1966-2005	4	5
1966-2006	4	5
1966-2007	4	5
1966-2008	4	5
1966-2009	3	4
1966-2010	4	5
1966-2011	4	5
1966-2012	4	4
1966-2013	3	3
1966-2014	4	4

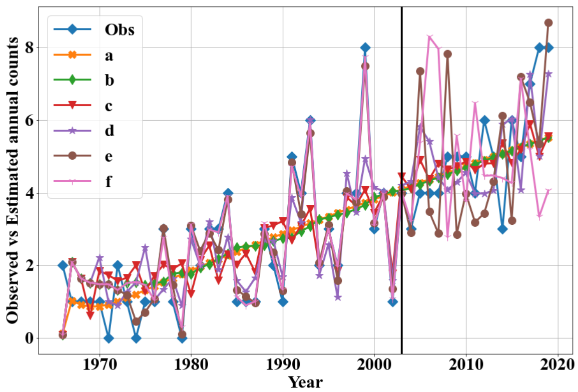
Table B.2: Architecture selection results of the ANN model for different training sets with maximum five input nodes

Training set observation period	The number of inputs	The number of hidden nodes
1966-2003	6	7
1966-2004	4	7
1966-2005	6	7
1966-2006	9	7
1966-2007	8	7
1966-2008	6	7
1966-2009	8	7
1966-2010	5	7
1966-2011	5	7
1966-2012	5	6
1966-2013	4	7
1966-2014	9	7

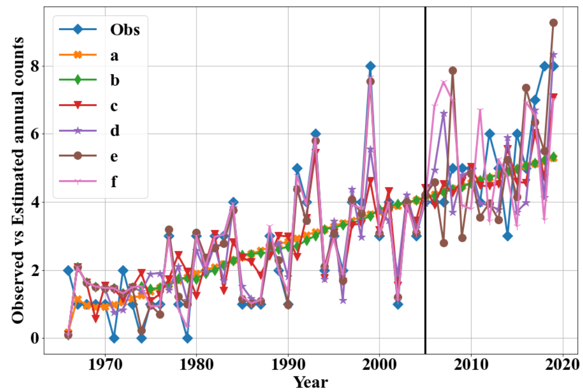
Table B.3: Architecture selection results of the ANN model for different training sets with maximum seven input nodes

Training set observation period	The number of inputs	The number of hidden nodes
1966-2003	7	10
1966-2004	7	9
1966-2005	6	9
1966-2006	9	9
1966-2007	7	9
1966-2008	4	9
1966-2009	4	10
1966-2010	5	10
1966-2011	7	10
1966-2012	8	10
1966-2013	5	10
1966-2014	8	9

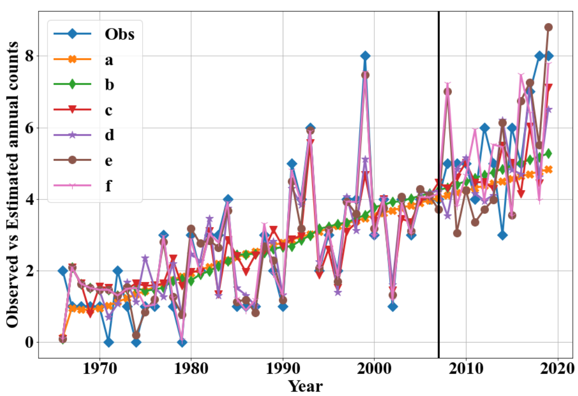
Table B.4: Architecture selection results of the ANN model for different training sets with maximum ten input nodes



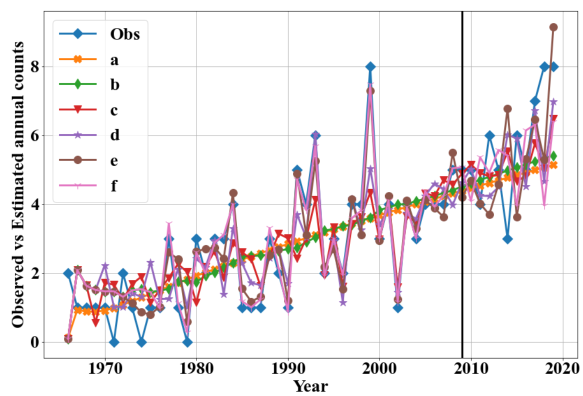
(a) Training set 1966-2003



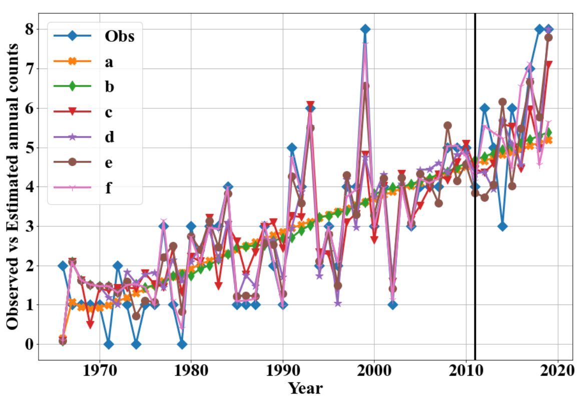
(b) Training set 1966-2005



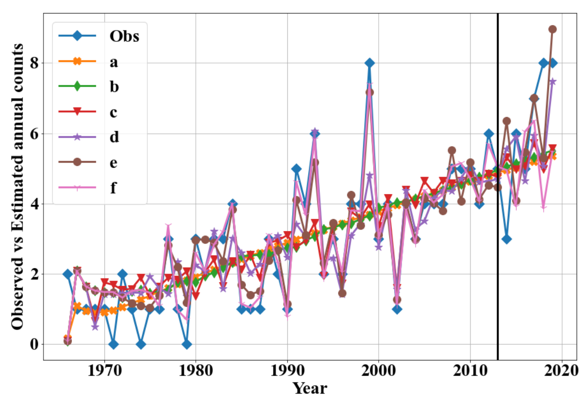
(c) Training set 1966-2007



(d) Training set 1966-2009



(e) Training set 1966-2011



(f) Training set 1966-2013

Figure B.1: Observed and estimated annual counts from different models with using different training sets (obs: real observations from the Violence Project database, a: change-point model with WG rate function, b: ARIMA model, c: hybrid ARIMA-ANN model with maximum 3 input nodes, d: hybrid ARIMA-ANN model with maximum 5 input nodes, e: hybrid ARIMA-ANN model with maximum 7 input nodes, f: hybrid ARIMA-ANN model with maximum 10 input nodes)