MGT 6203 - Data Analytics for Business

Impact on Gun Industry Company Stock Returns after Mass Shooting

Team 89

Javier Sanmiguel
Madhu Mudigonda
Shainu Prakash
Thanigaivel Shanmugam

1. Abstract

Recent decades have witnessed a series of high-profile mass shootings throughout the United States. While most homicides receive little attention from the public, mass shooting incidents are extremely salient. Since mass shootings are random occurrences, we want to estimate their impact on gun manufacturer's stock. Specifically, we want to see the trend of gun manufacturer stock performance after mass shootings.

It is expected that gun manufacturers will be affected by the response of these events. It is not clear what is the impact on the share price and return of those companies. Financial organizations and some shareholders are not supporting those companies, and some people are demanding changes in the legislation. However, it appears that some people respond to mass shootings by buying more guns, so gun sales are increasing.

2. Problem Statement

We analyze the impact of mass shooting events on the US gun industry stocks. General perception is that gun sales will increase after the mass shooting events. In this study, we want to analyze if there's a real correlation between mass shootings events on gun company stocks using the analytical tools. This could be very valuable information for Portfolio managers, Hedge funds, High frequency traders for risk management within their portfolios and also could be helpful information for companies who are into gun manufacturing or sales.

The question we want to answer is: Is there any impact on the gun industry stock returns after a mass shooting event?

And some additional questions

1. Is there correlation between the mass shooting and the stock performance?
2. Is there a long-term impact on the stock due to the high-profile mass shootings?

3. Initial Hypothesis

Our initial hypothesis and assumption is that the stock prices will get positively impacted after a mass shooting event. Though there might be a negative impact to the stock price, it will get settled. We think due to the fear factors, people will buy more guns, which will increase the revenue of the company which in turn increases the stock price.

4. Data and Feature Engineering

4.1 Data

We are using the mass shooting data from Kaggle us-mass-shootings-last-50-years, this dataset contains Mass Shootings in the United States of America (1966-2017) and it appears to meet the definition of mass shooting used by the FBI. This has 128 entries with 24 attributes, key fields we are interested in are location of the incident, date, fatalities, injured and total victims' info.

	location	date	fatalities	injured	total_victims
1	Uvalde, Texas	5/24/22	15	-	-
2	Buffalo, New York	5/14/22	10	3	13
3	Sacramento, California	2/28/22	4	0	4
4	Oxford, Michigan	11/30/21	4	7	11
5	San Jose, California	5/26/21	9	0	9
6	Indianapolis, Indiana	4/15/21	8	7	15

For the gun stocks dataset, we identified the following 10 companies in the gun industry as the universe for creating the gun industry stock portfolio as per the Yahoo News article on "*10 Best Gun Stocks to Invest In"* [8]:

1. AMMO, Inc. (NASDAQ: POWW) is a designer, developer, and seller of ammunition and ammunition component products for handguns and long guns in the US and internationally. It offers STREAK Visual Ammunition for shooters to track bullet paths, among other products
2. National Presto Industries, Inc. (NYSE: NPK) is a provider of houseware and defense and safety products in North America. The company's Defense segment manufactures 40mm ammunition, precision mechanical and electro-mechanical products, and medium caliber cartridge cases, alongside other products. This segment operates mainly under the US Department of Defense and DOD prime contractors.
3. American Outdoor Brands, Inc. (NASDAQ: AOUT) is a provider of outdoor products and accessories for outdoor enthusiasts in the US and internationally. The company offers hunting, shooting, personal security, fishing, camping, and defense products.
4. Big 5 Sporting Goods Corporation (NASDAQ: BGFV) is a sporting goods retailer in the western US. The company sells equipment for hunting, alongside other outdoor sports equipment and a selection of firearms.
5. Smith & Wesson Brands, Inc. (NASDAQ: SWBI) is a manufacturer and seller of firearms across the globe. The company offers handguns, including revolvers and pistols, long guns, including sporting rifles, bolt action rifles and muzzleloaders, and other products.
6. Sportsman's Warehouse Holdings, Inc. (NASDAQ: SPWH) is an outdoor sporting goods retailer in the US. The company offers hunting and shooting products like ammunition,

firearms, firearms safety and storage products, reloading equipment, and shooting gear products, among others.

7. Sturm, Ruger & Company, Inc. (NYSE: RGR) is a designer, manufacturer, and seller of firearms under the Ruger brand and trademark in the US. The company has two segments: Firearms and Castings.
8. Vista Outdoor Inc. (NYSE: VSTO) is a manufacturer of consumer products in the outdoor sports and recreation markets in the US and internationally. The company's Shooting Sports segment provides ammunition products like centerfire ammunition, rimfire ammunition, shotshell ammunition, and reloading components, alongside hunting and shooting accessories.
9. Axon Enterprise, Inc. (NASDAQ: AXON) is a manufacturer of technology and weapons products for the military, law enforcement, and civilians making up its consumer base.
10. Olin Corporation (NYSE: OLN) is a manufacturer and distributor of chemical products in the US, Europe, and internationally. The company's Winchester segment offers ammunition products for hunters, recreational shooters, and law enforcement, and small caliber military ammunition products for infantry and mounted weapons, among other related products

As we explored and analyzed the stock universe, we had to remove some companies from the list due to lack of history, due to their minimal exposure to gun manufacturing and also some of them. For example, some companies don't manufacture guns, some they're only retailers. We had to exclude those stocks from our analysis.

Below are the sample data points of the gun company stocks:

symbol <chr>	date <date>	open <dbl>	high <dbl>	low <dbl>	close <dbl>	volume <dbl>	adjusted
rgr	1/4/2010	10.01	10.54	10.01	10.24	788200	6.930286
rgr	1/5/2010	10.31	10.84	10.29	10.59	466700	7.167161
rgr	1/6/2010	10.68	11	10.49	10.51	455500	7.113017
rgr	1/7/2010	10.54	10.77	10.39	10.47	236800	7.085947
rgr	1/8/2010	10.42	10.61	10.28	10.41	203700	7.04534
rgr	1/11/2010	10.54	10.57	10.4	10.49	168000	7.099483

4.2 Key Variables

Key Variables	**Type**	**Notes**
Mass shooting event date	Independent	
Number of fatalities	Independent	
Number of injured	Independent	
Adjusted Close price	Dependent	The adjusted close price of the gun manufacturing stocks
Portfolio returns of the gun stock portfolio	Dependent	Derived from the adjusted close prices of the gun companies. We used market cap weights to create the portfolio
Simple Moving Average (SMA – Short 2 day) of the portfolio return	Dependent	Derived from the portfolio returns
Simple Moving Average (SMA – long 15 day) of the portfolio return	Dependent	Derived from the portfolio returns
Momentum of the portfolio return (2, 3, 5, 10, 15 and 20 day)	Dependent	Derived from the portfolio returns
Draw Down of the portfolio return	Dependent	Derived from the portfolio returns

Please note, we have more than one dependent variable listed above. The idea is, we will be studying the impact with different derived dependent or response variables to understand the impact of mass shooting events on the market returns of the top gun manufacturing companies in the industry.

4.3 Feature Engineering

At a high level, feature engineering is the process of selecting, manipulating, and transforming raw data into features that can be used in supervised learning models like linear regression. In this section, we will describe our process to create the model data features, preparing the training and test datasets, and model parameters used for creating the models.

We mainly used two datasets for our models. One is, US mass shooting event data and the other one is stock market data on gun industry stocks.

The US mass shooting events dataset is fairly straightforward and it didn't require much transformation. It contained the event date, number of fatalities, number of injured, total victims, and the location information. For our model, we considered number fatalities, number injured, total victims and event date as the key variables from this dataset.

In the case of the gun stock price dataset, we used trading date, and adjusted close price as key variables from the financial stock dataset. We have selected adjusted close price instead of daily closing price because adjusted close price is normalized price after taking into account corporate actions like stock splits and dividend distributions. Our thought process behind selecting adjusted close price is that it will give an accurate measure of the impact without sudden swings due to stock-splits and dividend distribution.

As we are working with multiple gun industry company stock prices, we had to create an index to measure the impact on all the stock prices versus individual stock prices. For that reason, we created a portfolio with weights based on market capitalization [9]. It is a fairly commonly used approach in the financial industry to create the financial models for a group of companies.

Following are the steps we followed for transforming the stock price data for modeling:

1. Created a portfolio using the gun industry company stock prices based on market capitalization weights.
2. Computed the normalized daily returns on the portfolio using the Tidyquant package.
3. As our objective was to test if there's an impact on the market returns of the gun companies based on the US mass shooting event, we thought technical indicators like Simple Moving Average (SMA) and Momentum of the portfolio return would help us. So, we added four new features to the portfolio dataset: SMA_Short-term, SMA_Long-term, MOM_Short-term, MOM_Long-term.

Simple moving average (SMA) calculates the average of a selected range of prices, usually closing prices, by the number of periods in that range, aids in determining if an asset price will continue or if it will reverse a bull or bear trend. Momentum(MOM) measures the rate of the rise or fall of stock prices.

4.3.1 Data cleansing

US Mass Shootings data from Kaggle had 24 attributes, this was reduced to 5 columns, location, date, fatalities, injured, total victims. This location attribute contained the locality and state, so created a new attribute State with info from location attribute, also created a new column Year for analysis.

Next to help in joining the data to stock market info, we had to adjust the date field value to be the next working day, the day the stocks will see an impact of the event.

Saturday = Event Date + 2, which is Monday
Sunday = Event Date + 1, which is Monday

Final data set is set to have location, state, date, year, fatality, injured and total victims.

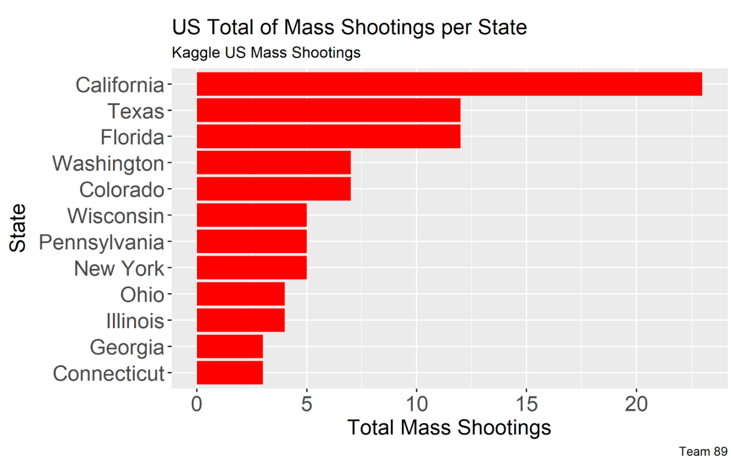
location <chr>	state <chr>	date <date>	year <int>	fatalities <dbl>	injured <dbl>	total_victims <dbl>
Uvalde, Texas	Texas	5/24/2022	2022	22	17	39
Buffalo, New York	New York	5/16/2022	2022	10	3	13
Sacramento, California	California	2/28/2022	2022	4	0	4
Oxford, Michigan	Michigan	11/30/2021	2021	4	7	11
San Jose, California	California	5/26/2021	2021	9	0	9
Indianapolis, Indiana	Indiana	4/15/2021	2021	8	7	15

We have used the Tidyquant R package to download the daily stock prices for the 2 gun manufacturing companies with Yahoo Finance as source.

Merged the mass shooting data with the stock performance data based on the date field.

4.3.2 Exploratory Data Analysis

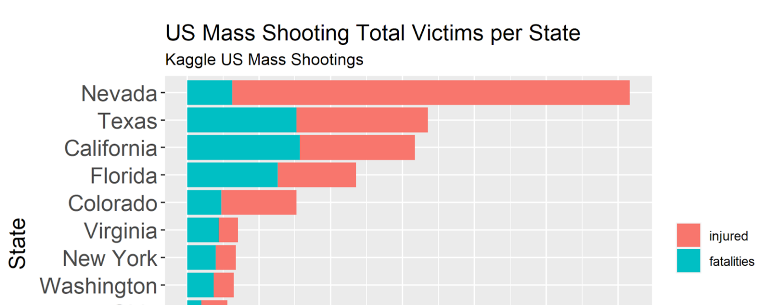
Analyzed the mass shooting data individually to see the pattern of the data with respect to state vs no of events, total victim's vs state, years vs no of events etc. This helped to understand the data and to identify data issues if any.



Total Mass Shootings vs. State

Highest: California: 23

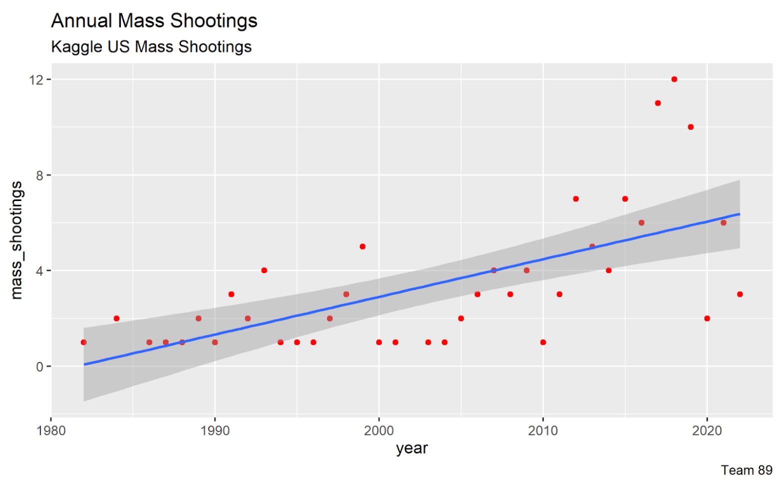
The data looks distributed across 12 states with California having the largest number of mass shooting events.



Total Victims vs. State

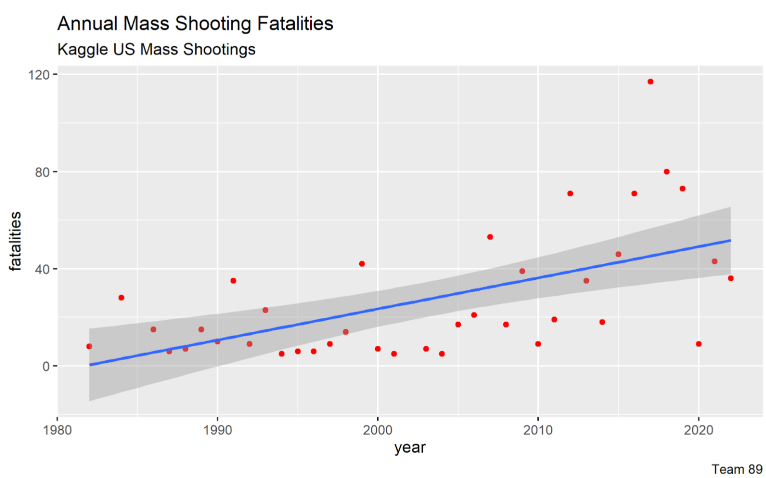
Highest: Nevada
Concentration: 5 States

Total number of victims is high in Nevada, with the highest injured, Texas, California, and Florida show high numbers of fatalities.



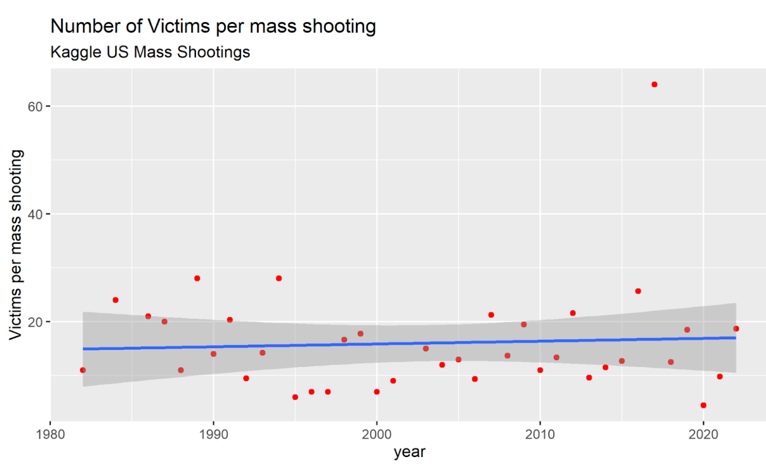
Year vs. # of Mass Shootings

Mass shootings are increasing every year



Year vs. Fatalities

Number of fatalities is also increasing every year.

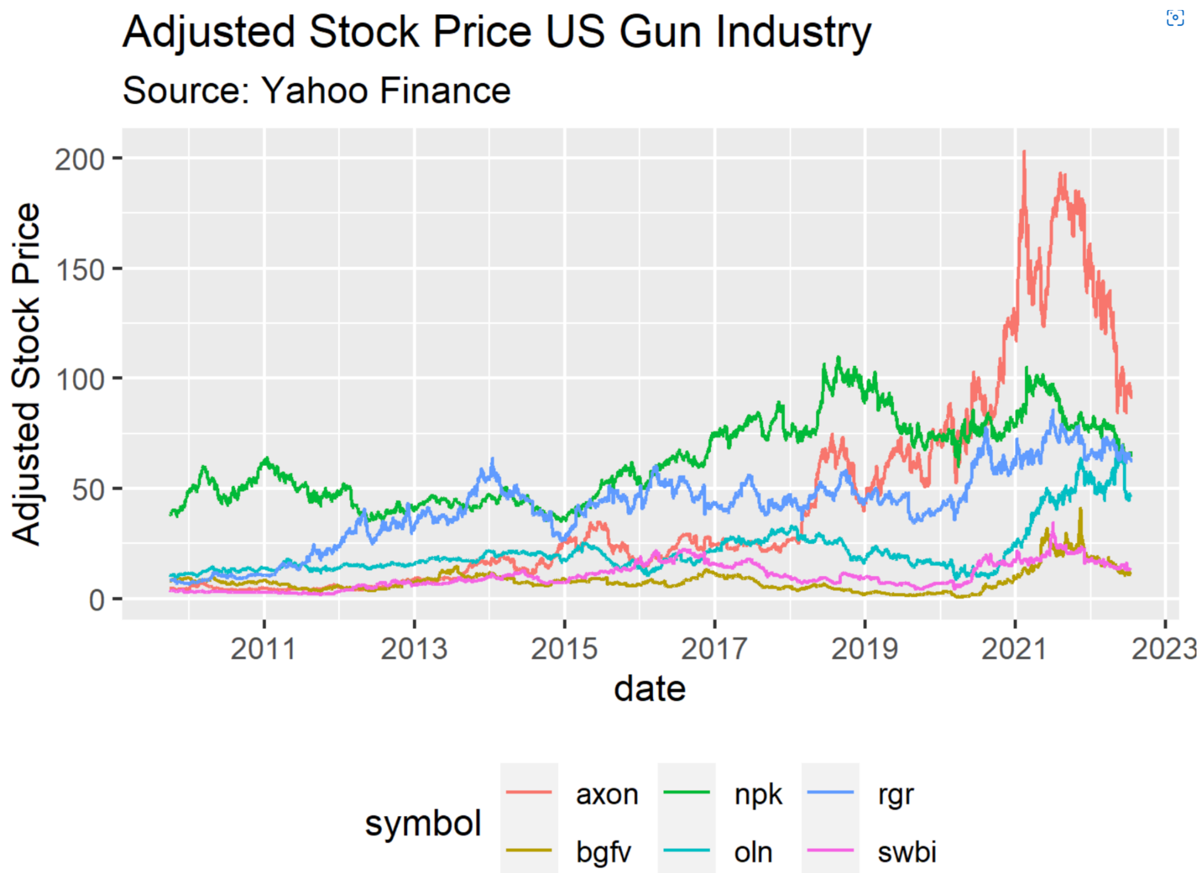


Year vs. Victims per Mass Shootings

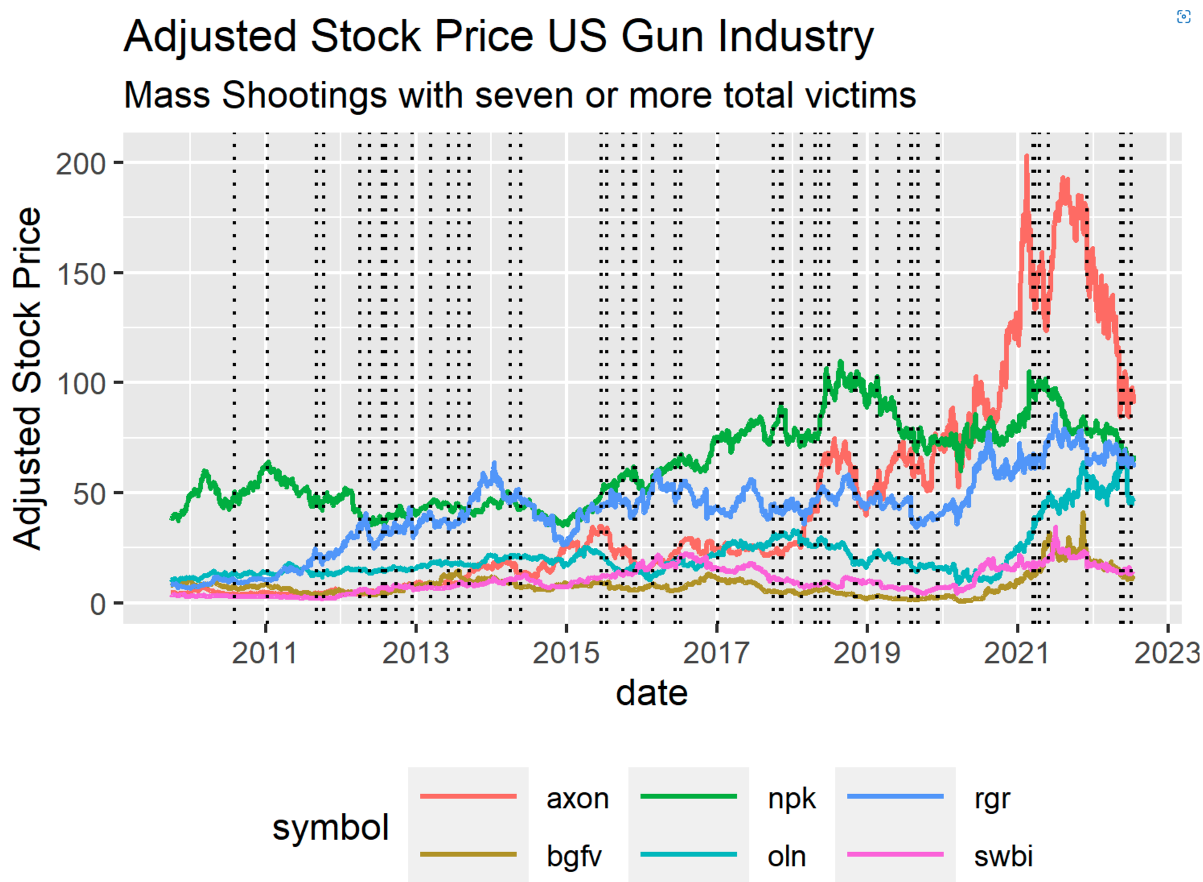
The number of victims per shooting is about the same since 1982; around 17

The mass shooting data has shown the trends that are in line with the normal conception that the events are increasing year over year and the fatalities because of that.

Analysis of the stock data and the combined data and its impact. Impact of the mass events to the stock prices just after the event occurred.

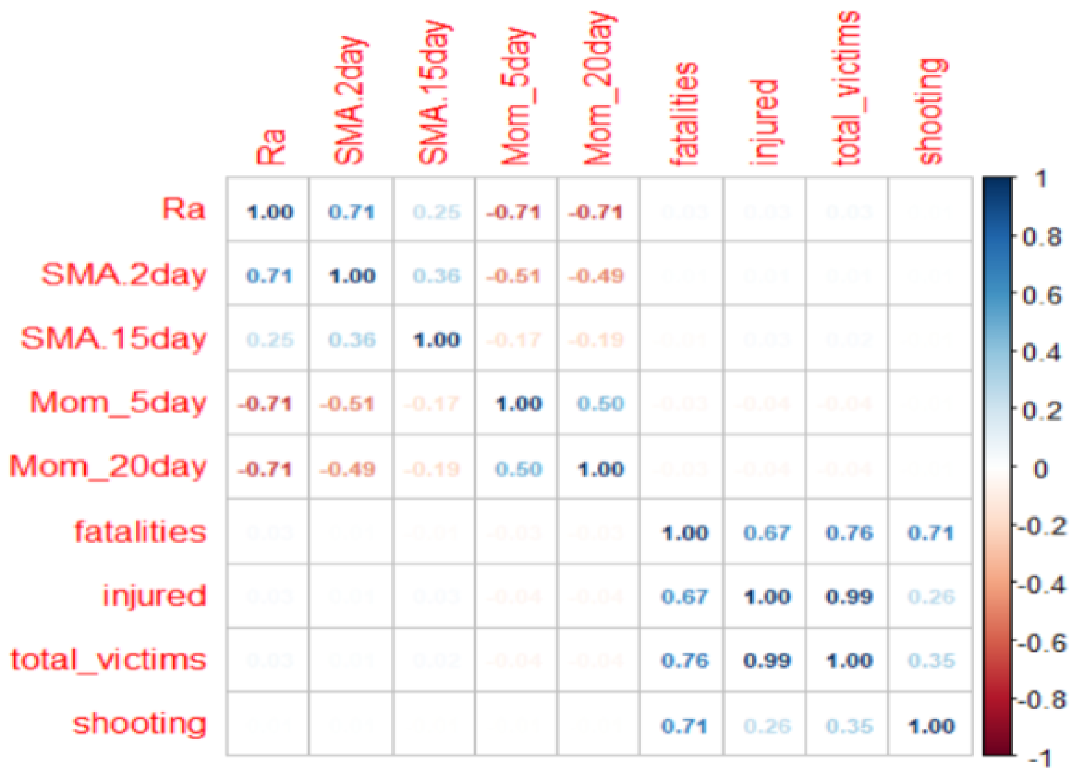


This graph shows the adjusted stock prices of the selected stocks with a marker on the dates mass shooting occurred. The chart shows some companies' stock prices showing a downward trend whenever a mass shooting event occurred.



Source: Yahoo finance and Kaggle's Mass Shooting in United States (2018-2022)

Below is the correlation plot for the key independent and dependent variables from US mass shooting dataset and the derived variables for the gun industry stock portfolio: SMA_2day, SMA_15day, Mom_5day, Mom_20day, number of fatalities from the event, injured, total_victims, and shooting as dummy variable to indicate if there’s mass shooting event on a given date.



5. Modeling

5.1 Introduction to Modeling Approach

The primary goal of this project is to analyze the impact of US mass shooting events on the gun industry company stock prices. As discussed in the previous section, we have used US Mass shooting event data from the Kaggle datasets[5], Active Shooter resources from FBI[1] in conjunction with stock price data from Yahoo Finance and R's TidyQuant package[4]. To analyze and understand the impact of mass shooting events on the gun stocks, we employed widely used machine learning and statistical technique Linear regression for building models.

Linear regression is useful for studying the linear relationship between a response variable and one or more independent variables.

At a high level, there are mainly two types of linear regression: 1. Simple linear regression and 2. Multiple linear regression. Below are the brief descriptions of both types of linear regression: .

1. Simple linear regression: Simple linear regression is a statistical technique that is useful for finding relationships between two continuous variables. One being the response and other independent variables.

2. Multiple linear regression: In multiple linear regression, there will be two or more independent variables to predict or to explain the relationship to response or target variable.

The main idea behind the linear regression is to find the line equation that best fits and explains the relationship between independent and dependent (response) variables with an objective to minimize the error in predictions. Error is basically the distance between the response variable and the regression line equation.

In our case, we used linear regression to study the relationship between the gun industry company stock portfolio returns and the US mass shooting event factors like fatalities, number of injured and total victims of the event. We have employed both simple and multiple linear regression in our modeling. In the upcoming sections, we will describe the model data preparation process, our models from iterative process, evaluation process and key metrics used, and finally results and discussion.

5.2 Modeling

As described in the previous section, we will be using linear regression for creating the models for this project. Modeling is an iterative process. In this section, we describe different models we created and the process used for selecting the final best performing model.

5.2.1 Training, Validate and Test Splitting

Before we get into modeling, one of the important steps is to create training and test data sets. This is very important for evaluating the performance of the model. Every dataset contains patterns called real and random effects. Any model building exercise is to be able to predict the values for the response variable with new real-world data using the model created. When we build a model with a dataset, it gets exposed to both real and random patterns. Ideal model is one which perfectly identifies the real effects / patterns in the data. But, an ideal model is not possible in the real world and also it can lead to overfitting. That means, the model may perform well with training data, but may perform poorly with test data points or on any new data point that we use for predicting. That's why it is very important to understand the accuracy of the model by exposing it to test data.

Here are the commonly acceptable approaches for preparing the training and test data:

1. Random splitting: For evaluating the model, we need new data. In this approach, the dataset is randomly split into training and test data using some ratio. This approach may not work for time series data as we may end up using the latest data points to predict past data points.

2. Temporal splitting: Using the temporal variable is the most reliable way of splitting the data into training and test data whenever there is a date or time component available in the dataset and if we are trying to predict something in the future. In this approach, we can use past data points as training data and the recent data points as a test set.

In our case, we used temporal splitting for creating the training and test data sets as we are dealing with time series data. We used the date variable for splitting the data using a merged dataset of US mass shooting events and gun company stock portfolio.

Based on the generally accepted approach for temporal splitting, we used the recent for test dataset and the past data for training dataset. Our dataset consists of data from 2010-01-01 to 2022-06-15.

For the training dataset, we have used the data range from 2010-01-01 to 2019-12-31 and for the test dataset, we have used 2020-01-01 to 2022-06-15.

In our project, we didn't use validation dataset as we didn't employ any hyperparameters tuning for linear regression. Hyperparameters are the tuning parameters for mathematical algorithms to tune the learning of the data. We used standard OLS linear regression without any hyperparameters. We can include the hyperparameter tuning as our future work.

5.2.2 Model parameters

In this section, we will describe the model parameters and the parameters we used in our models for tuning the learning of the models. Basically, model parameters are the parameters in the model that are determined from the training data set. These are fitted parameters. [10]

In our case, model parameters are date, portfolio returns, SMA_2 day, SMA_15day, MOM_5day, MOM_20day, number of injured, fatalities, and total victims. While building our models, we tuned the duration of the SMA, MOM, and number fatalities in the training data to fit our models. The criteria we used to tune these parameters is by checking the significance of the feature in the model with a p-value significance of 0.05 level.

5.2.3 Models

In this section, we present the models we have created to answer our research question, that is to check the impact of US mass shooting events on the daily portfolio returns of the gun stocks.

Using our training dataset, we built different models using SMA_2 day, SMA_15day, MOM_5day and MOM_20day as response variables. First, we will discuss the models with SMA as response variable, then we will discuss MOM based models.

In the SMA based models, we created 2 types of models. One with SMA_2day as response variable and fatalities, injured and total_victims as independent feature variables. Another kind is, using SMA_15 day as response or dependent variable and fatalities, injured and total_victims as independent feature variables. Total_victims is an interaction with injured and fatalities combined.

Below is the table containing different SMA based models and their characteristics like coefficients, R^2, Adjusted R^2 and significance of the feature variables.

	Formula	R^2	Adj R^2	p-value
Model 1	SMA_2day ~ fatalities + injured Coeffs: intercept:8.629e^-04, fatalities: 3.240e^-05, injured:1.754e^-05	0.0002287	-0.0005667	fatalities = 0.88505, injured = 0.66125
Model 2	SMA_15day ~ fatalities + injured Coeffs: intercept:8.629e^-04, fatalities: -1.885e^-04, injured:4.039e^-05	0.003216	0.002423	fatalities = 0.01993*, injured = 0.00525**
Model 5	SMA_2day ~ total_victims Coeffs: intercept:8.645e^-04, total_victims: 1.934e^-05	0.0002273	-0.0001702	total_victims = 0.44962
Model 6	SMA_15day ~ total_victims Coeffs: intercept:8.645e^-04, total_victims: 1.259e^-05	0.0007356	0.0003383	total_victims = 0.174

Table - 5.2.3.a

We will be describing the results and evaluate the models in depth in the “Results and Discussion” section. But, at high-level we can observe that SMA_15day ~ fatalities + injured is the only model with slightly Adj R^2 and also the p-value for the features number of fatalities and injured variables are significant at 0.05 significance level.

Now, below we will present the MOM based models and their characteristics:

	Formula	R^2	Adj R^2	p-value
Model 3	MOM_5day ~ fatalities + injured Coeffs: intercept:5.290e^-05, fatalities: -4.487e^-05, injured:-1.294e^-04	0.00211	0.001316	fatalities = 0.921, injured = 0.110
Model 4	MOM_20day ~ fatalities + injured Coeffs: intercept:1.077e^-04, fatalities: -4.569e^-04, injured: -6.590e^-05	0.002109	0.001315	fatalities = 0.317, injured = 0.419
Model 7	MOM_5day ~ total_victims	0.0021	0.001703	total_victims = 0.0215*

	Coeffs: intercept:6.203e^-05, total_victims: -1.191e^-04			
Model 8	MOM_20day ~ total_victims Coeffs: intercept:6.550e^-05, total_victims: -1.134e^-04	0.00188	0.001484	total_victims = 0.0296*

Table - 5.2.3.b

IN the MOM based models, we can observe that the MOM_5day ~ total_victims is the better model based on R^2 and also the total_victims variable is significant at 0.05 level.

We created models using log transformed data to see if we can improve the fit of the models to the training data. We briefly present here in the below table with log transformed models.

	Formula	R^2	Adj R^2	p-value
Model 8.a	MOM_20day ~ log (total_victims + 1) Coeffs: intercept:0.0001114, log(total_victims + 1): -0.0022919	0.0007297	0.0003324	log (total_victims + 1) = 0.175
Model 8.b	log (MOM_20day + c +1) ~ total_victims c: abs(min(mom_20day) Coeffs: intercept: 1.663e^-01, total_victims: -9.859e^-05	0.001974	0.001577	total_victims = 0.0258*
Model 8.c	log (MOM_20day + c +1) ~ log(total_victims + 1) Coeffs: intercept: 0.1663123, log (total_victims + 1): -0.0020589	0.0008177	0.0004204	log (total_victims + 1) = 0.152

Table - 5.2.3.c

For the log transformed models, we only implemented it for the MOM_20-day model due to time and resource constraints. As we observe from the above table, there's only one model with slightly better Adj. R^2 and total_victims feature at 0.05 significance. But it is no better than a regular linear-linear model. Linear-linear mode has better Adj R^2 and p-value for features.
In the upcoming "Results and Discussion" section we will dive deep into analyzing the models using key performance metrics and plots.

5.3 Key metrics for model evaluation

In this section, we describe the key metrics we used for evaluating the models. We mainly used R2 and adjusted R^2 to check the explanatory power of the model. For model performance and to check accuracy, we used Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Mean Absolute Percentage Error (MAPE).

R^2: This metric represents the part of the variance of the dependent variable explained by the independent variables of the model. It measures the strength of the relationship between your model and the dependent variable.

Adjusted R^2: The main difference between adjusted R-squared and R-square is that R-squared describes the amount of variance of the dependent variable represented by every single independent variable, while adjusted R-squared measures variation explained by only the independent variables that actually affect the dependent variable.

R^2 tends to increase with an increase in the number of independent variables. This could be misleading. Thus, the adjusted R-squared penalizes the model for adding furthermore independent variables (k in the equation) that do not fit the model.

RMSE: RMSE is also called the Root Mean Square Deviation. It measures the average magnitude of the errors and is concerned with the deviations from the actual value. RMSE value with zero indicates that the model has a perfect fit. The lower the RMSE, the better the model and its predictions. Below are the formulas for RMSE and RMSE%:

$$RMSE = \sqrt{\frac{1}{n}\sum e_t^2} \qquad RMSE\% = \frac{\sqrt{\frac{1}{n}\sum e_t^2}}{\frac{\sum d}{n}}$$

RMSE gives more importance to most significant errors. It penalizes the large errors.

MAE: This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent.

$$MAE = \frac{1}{n}\sum |e_t|$$

MAPE: The Mean Absolute Percentage Error (MAPE) is one of the most commonly used KPIs to measure forecast accuracy. MAPE is the sum of the individual absolute errors divided by the demand (each period separately). It is the average of the percentage errors.

$$MAPE = \frac{1}{n}\sum \frac{|e_t|}{d_t}$$

In the upcoming section, we will be using the above defined metrics to discuss the results of our models.

5.4 Results and Discussion

Our primary objective of the project is to study the impact of US Mass shootings on the gun industry company stock returns. Our initial hypothesis is that stock prices will get affected negatively due to mass shooting events in the short term. But, in the long term it may bring in more revenue for gun companies as supporters of the gun industry will buy more guns out of fear of congress passing new gun laws. Due to time and resource constraints, we could not completely analyze the long term impact on stock returns, revenues, or sales. In this section, we will attempt to explain the short term impact on the daily returns of the gun industry company stock daily returns using financial technical indicators like Simple Moving Average (SMA) and Momentum (MOM) against the key variables from the US mass shooting data.

We have iteratively built several models with different model parameters using SMA and MOM as response variables. Here's a brief description of model parameters we used to tune to get a better model with explainability:

SMA duration: We tried the SMA with a shorter duration of 2 days and slightly longer duration of 15 day to understand the impact on returns.

Companies in the portfolio: During exploratory data analysis, what we observed was not all companies have direct exposure to firearms manufacturing. In order to understand the impact and get a better model with explainability, we investigated each of the company profiles from our initial list of 10 companies we selected. Based on our analysis, only Sturm, Ruger & Company, Inc. (NYSE: RGR) and Smith & Wesson Brands, Inc. (NASDAQ: SWBI) have direct exposure to firearms manufacturing. Portfolio returns with those two companies in the portfolio demonstrated some negative short term impact.

Filter on the number of fatalities: Number of fatalities in the US mass shooting dataset is also a key parameter that we tuned. Unfortunately, we didn't see much significance in the model with lower fatalities. So, based on our tuning we settled with fatalities >=7 to build a model with slightly explainability.

We have selected the following two models based on their feature significance at 95% confidence interval, R^2 and adjusted R^2 values.

First, we will discuss the SMA_15day based model. In this model, we tried to analyze the relationship between SMA_15day of the portfolio return against the fatalities and injured variables from the US mass shooting model.

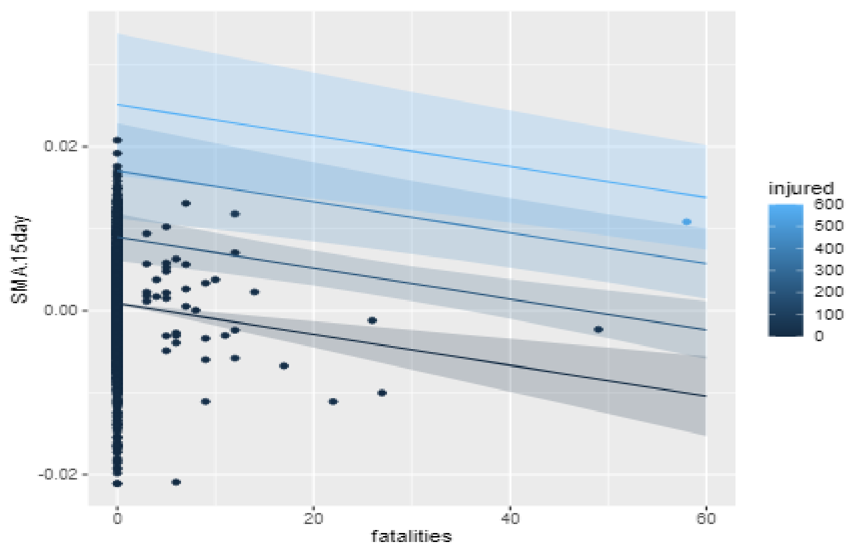
Technical indicator SMA indicates the trends in the portfolio returns [7]. If SMA is moving up, it means there's an upward trend. If SMA is moving down, then the trend is down. SMA duration in our model is one of the model parameters. While building models, we tried with both short term duration 2 days and slightly longer term of 15 days.

Here's the model coefficients, R^2 and adj. R^2 of the SMA_15day model:

Formula	R^2	Adj R^2	p-value
SMA_15day ~ fatalities + injured Coeffs: intercept:8.629e^-04, fatalities: -1.885e^-04, injured:4.039e^-05	0.003216	0.002423	fatalities = 0.01993*, injured = 0.00525**

Table - 5.4.a

Below is the regression plot for the SMA_15day ~ fatalities + injured model:

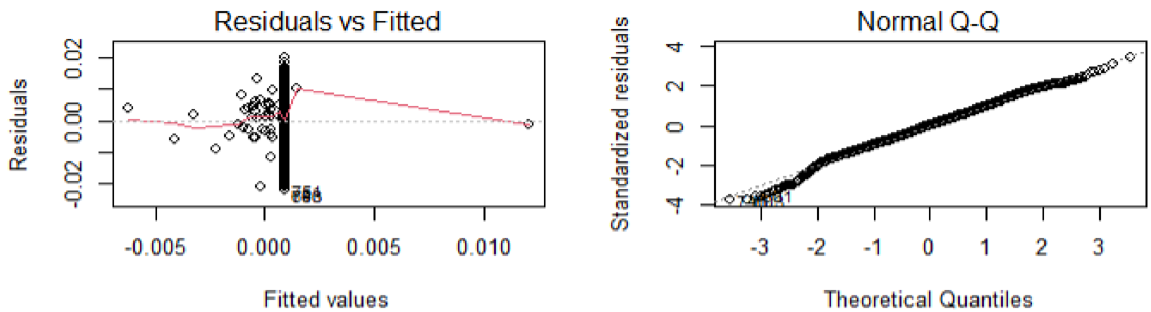


In the above SMA_15day vs fatalities + injured regression plot, we can observe the regression lines for fatalities vs SMA_15day with different regression bands for injured. Feature variable Injury is indicated as different blue bands with regression lines. We can clearly see that there's some slight linear relationship with SMA_15day. As fatalities increase, there's a slight negative impact on the SMA_15day portfolio return.

Now, we will look at the goodness of fit plots for the SMA_15day model to evaluate if the model has explainability power.

We will be looking at Residual vs fitted plot. This scatter plot shows the distribution of the residual errors vs fitted values. If there exists some pattern, then that means it is a sign of non-linearity. In the below plot, we don't see any major pattern. Points are somewhat randomly distributed.

Other plot is, Q-Q plot. It is a quantile-quantile plot used to validate the normality assumption of the dataset. If the data comes under normal distribution, then the points should fall on the diagonal line. In the below Q-Q plot, although it is not completely normal, most of the points are on the straight line. Vertical line near zero indicates 0 fatalities or no events.



Now, let's briefly discuss the model based on the MOM technical indicator as a response variable. In this model, we tried to investigate if the momentum exhibits any impact against the total victims of the US mass shooting event.

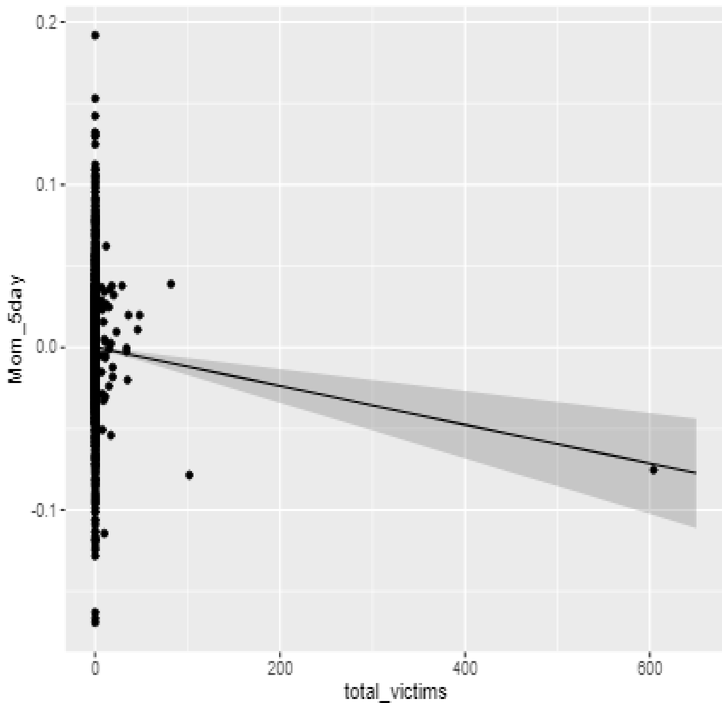
Technical indicator momentum is useful for determining the strength or weakness of the price. It measures the rate of rise or fall in the stock prices [6]. In our project, we used momentum of the gun company stock portfolio.

Here's the model coefficients, R^2 and adj. R^2 of the MOM_5day model:

Formula	R^2	Adj R^2	p-value
MOM_5day ~ total_victims Coeffs: intercept:6.203e^-05, total_victims: -1.191e^-04	0.0021	0.001703	total_victims = 0.0215*

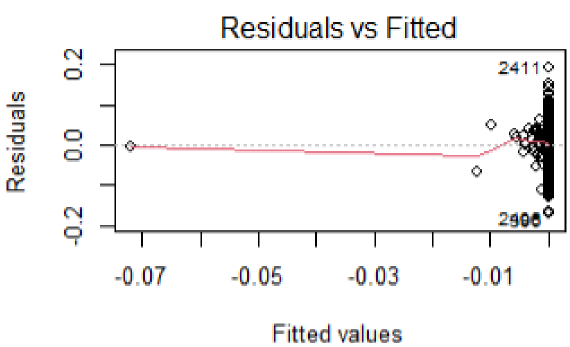
Table - 5.4.b

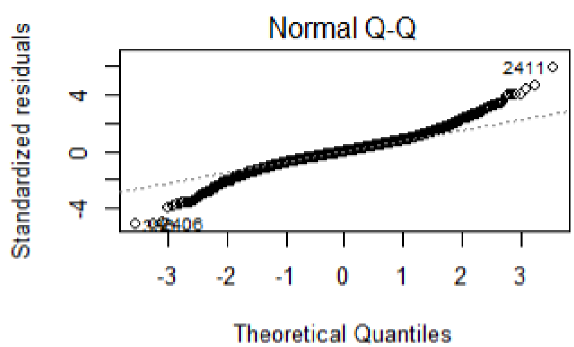
Below is the regression plot for the MOM_5day ~ total_victims model:



From the above MOM_5day vs total_victims regression plot, we can observe the regression line for total_victims vs SMA_15da. We can observe that there's some slight linear relationship with MOM_5day. Vertical line near zero indicates 0 fatalities or no events.

Below are the residual vs fitted and Normal Q-Q plot. Residual vs fitted plot is not much different from the SMA_15dya. But, in the case of Q-Q plot there are fat tails on the left and right-hand side. It indicates that the distribution may be skewed may not be completely normal distribution.





Here's the metrics from the prediction using the test dataset:

Formula	R^2 & Adj R^2	RMSE	MAE	MAPE
SMA_15day ~ fatalities + injured Coeffs: intercept:8.629e^-04, fatalities: -1.885e^-04, injured:4.039e^-05	R^2: 0.003216 Adj R^2: 0.002423	0.005951607	0.004613019	132.9048
MOM_5day ~ total_victims Coeffs: intercept:6.203e^-05, total_victims: -1.191e^-04	R^2: 0.0021 Adj R^2: 0.001703	0.03653719	0.02632611	99.66629

RMSE and MAE are slightly lower for SMA_15 day model and MOM_15day has lower MAPE. As lower numbers are better for eros and higher R^2 and adj. R^2 numbers are good for selecting the models, we think SMA_15day is probably a slightly better model with explanatory power. R^2 and adj R^2 indicate that models don't have huge explanatory power.

From the selected SMA_15day model, we can infer that:

- For unit increase in fatalitities, the 15-day average daily return of the portfolio drops by 1.885e^-04.
- That means, it is roughly 21% in daily return with respect to intercept while holding all else constant.

6. Summary

In the previous sections, we explained about our data acquisition process for US mass shooting and gun company stock price datasets, key variables from the dataset, feature engineering steps like data cleansing and exploratory data analysis, modeling approach, models we created, the key metrics definition, analysis for selecting the models, regression plots and goodness of fit of the models. In this section, we will summarize our findings, discuss the challenges and future work to conclude.

We have set out an initial project hypothesis to find the impact of US mass shooting events on the gun industry company stock returns. From our analysis, we can conclude that there's a very small impact on the daily stock returns of the gun companies portfolio between 5 to 15 day period. To study the long term impact, we need to explore other datasets like revenue, sales, company financial information data from these companies.

We briefly here mention the challenges we faced in connecting the two different datasets and also regarding the gun companies. As some of the Mass shootings may fall on the weekends or holidays, we had to make calculated judgments to roll forward the event to the following trading date to study the impact. Also, most of the companies from our initial list didn’t have direct exposure to firearms manufacturing as they’re into retail, distributions, and some are government contractors. So, we excluded those companies from our analysis as our focus was only on firearm manufacturing companies. We ended up creating a small portfolio of only two companies for our analysis.

In the future, for short term study, we can also explore other data points like options chain transactions of these gun company stocks, twitter sentiment analysis using StockTwits, volume or transactions of the stocks. Overall, this project gave us an opportunity to study about the gun industry, US mass shooting event trends, their impact on stock market returns and finally we were able to apply the key analytical concepts and tools we studied in this course to the real world problem . We hope that in the future studies, we can perform more thorough analysis and motivate the companies and other involved parties to stop these nonsensical mass shootings by showing a negative impact on their business and society.

7. References

1. FBI. (2016, June 9). *Active shooter resources*. FBI. Retrieved July 16, 2022, from https://www.fbi.gov/about/partnerships/office-of-partner-engagement/active-shooter-resources

2. Daniel, W. (2022, May 25). *Gunmakers' stocks rise after Texas School shooting leaves 19 children and two teachers dead*. Fortune. Retrieved July 16, 2022, from https://fortune.com/2022/05/25/uvalde-school-shooting-gun-stocks-rise/

3. Gopal, A., & Greenwood, B. N. (2017, May 18). *Traders, guns, and money: The effects of mass shootings on stock prices of firearm manufacturers in the U.S.* PloS one. Retrieved July 16, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5436715/

4. Dancho, M. (2022, May 20). Core functions in tidyquant. Retrieved July 16, 2022, from https://cran.r-project.org/web/packages/tidyquant/vignettes/TQ01-core-functions-in-tidyquant.html

5. Usmani, Z.-ul-hassan. (2022, May 25). *US mass shootings*. Kaggle. Retrieved July 17, 2022, from https://www.kaggle.com/datasets/zusmani/us-mass-shootings-last-50-years?datasetId=2778&sortBy=dateRun&tab=collaboration&select=US%2BMass%2BShootings%2BMay%2B24%2B2022.csv

6. Team, T. I. (2022, May 11). *Momentum indicates stock price strength*. Investopedia. Retrieved July 17, 2022, from https://www.investopedia.com/articles/technical/081501.asp

7. Hayes, A. (2022, June 13). *Simple moving average (SMA)*. Investopedia. Retrieved July 17, 2022, from https://www.investopedia.com/terms/s/sma.asp

8. Farooq, F. (2021, July 12). 10 Best Gun Stocks to Invest In, from https://www.yahoo.com/video/10-best-gun-stocks-invest-191658480.html

9. Hayes, A. (2022, January 21). Capitalization weighted Index.Investopedia. Retrieved July 21, 2022, from https://www.investopedia.com/terms/c/capitalizationweightedindex.asp

10. Hayes, A. (2022, January 21) Model Parameters and Hyperparameters in Machine Learning. https://towardsdatascience.com/model-parameters-and-hyperparameters-in-machine-learning-what-is-the-difference-702d30970f6