

Volume 42, Issue 2

Offline events and online engagement: user reaction following mass shootings

Veronica Espailat
Brown University

Abstract

This study seeks to analyze the impact of an offline event on online user sentiment. To conduct this analysis, this study implements a two-stage difference-in-difference regression model to determine whether an offline event has an impact on user sentiment. My data set primarily consists of a set of exogenous mass shootings, and relevant user data derived from the New York Times API. Moreover, this study implements sentiment analysis tools such as VADER and TextBlob to measure user sentiment. This study primarily finds a significant, negative effect on user sentiment following a mass shooting.

I am both thankful and grateful for the guidance of Dr. Brian Knight, Ph. D., and Dr. Neil Thakral, Ph. D. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Citation: Veronica Espailat, (2022) "Offline events and online engagement: user reaction following mass shootings", *Economics Bulletin*, Volume 42, Issue 2, pages 913-926

Contact: Veronica Espailat - veronica_espailat@alumni.brown.edu.

Submitted: June 28, 2021. **Published:** June 30, 2022.

1 Introduction

The integration of digital media and daily life allows for fluid engagement between online occurrences and offline events. Economic studies are beginning to exploit sentiment analysis and big data in relating online and offline events, such as considering the effect of sleep on tweeting behavior (Almond & Du, 2020), information dissemination on campaign sentiment and voter outcomes (Gorodnichenko, Pham, & Talavera, 2021), and media positioning on financial markets (Fedyk, 2022 (Forthcoming)). Online sentiment may proxy for personal activity that manifest to have significant real-world consequences (Card & Dahl, 2011; Eren & Mocan, 2018). As seen through the relevant literature, media maintains the ability to alter collective sentiment and can catalyze action. Robust literature exists around the effect of media on emotions. Card and Dahl (2011)'s piece on sporting events as negative unexpected emotional cues suggests a causal relationship between family violence and home team losses. Moreover, Philippe and Ouss (2018)'s piece on judicial sentencing following criminal events suggests that judges prescribe harsher punishment following highly publicized criminal events. When faced with these biases that skew individual behavior, measuring the potential impact of these unwarranted inclinations becomes critical.

Measuring changes in sentiment as a result of exogenous media events contributes to understanding the individual motivations for consequent action (for example, the analysis of media position in Fedyk (2022 (Forthcoming)) and the analysis of bot tweeting behavior in Gorodnichenko et al. (2021), contributing to the broader project of unearthing drivers in collective activity. As emphasized by Gorodnichenko et al. (2021), the use of bots on marginal, undecided voters in swing positions may disproportionately impact election outcomes. Fedyk (2022 (Forthcoming)) considers the disproportionate impact of media placement on trading outcomes. This paper differs from the existing literature in that it analyzes an organic sentiment driver which is both unrelated to the structure of the media platform (such as positioning) and to intentional drivers of emotional transformation (such as targeted ads and bots). By exploiting natural variation in mass shootings, this analysis considers a set of events isolated from the intention or mechanism of altering individual perception. This analysis questions how random phenomenon may have broad, unanticipated consequences.

This analysis seeks to contribute to existing research by relating offline, exogenous events, to online media engagement, and employing a novel data set compiled through the New York Times API. This study seeks to understand the degree to which offline events trigger changes in online user sentiment. I hypothesize that online user sentiment will change following an offline, exogenous event. I employ sentiment analysis tools to analyze the changes in user sentiment associated with significant, exogenous events that capture meaningful media attention. As my event sample primarily consists of mass shootings, I hypothesize that the sentiment score following this event sample will become more negative.

2 Data

This analysis integrates data from the Gun Violence Archive and the New York Times to compile a novel data set that relates offline events with online behavior. This data set is suitable to answering the hypothesis that sentiment score following exogenous events will shift (assuming the event is of significant magnitude) because it aggregates a data set of shocking, significant events with a database of user reactions to current events. Analyzing data from the New York Times, which has approximately 7.8 million subscribers, provides a base of users aware of current events

to measure exogenous event impact against.

Moreover, research relating mass media to tangible funding outcomes (Eisensee & Strömberg, 2007), family outcomes (Card & Dahl, 2011), and judicial action (Philippe & Ouss, 2018) supports the notion that offline events magnified through news sources garner reactions. In the case of Eisensee and Strömberg (2007), it is evident that natural calamities that do not necessarily have a direct impact on an individual's life affects their monetary decisions. Eisensee and Strömberg find a causal relationship between U.S. disaster relief, natural disaster timing, and unrelated noteworthy events such that disaster relief is higher when the natural disaster occurs at a period with limited unrelated noteworthy events. This suggests that because of a shift in attention due to media prevalence and crowd out, or lack thereof, unanticipated causal effects on funding occurred. Extrapolating further, mass shootings with significant media coverage are events likely to touch the lives of many, even if not personally affected, because of the significant media coverage and increased likelihood of media crowd out associated with this event.

2.1 Offline Events Sample

Using the publicly available New York Times API and Gun Violence Archive, I determine a sample of the top fourteen mass shootings from 2017-2020 in terms of injured victims, fatalities, and article references. Mass shootings provide the ideal event sample because of their perceived randomness and shortness in duration. Moreover, large scale exogenous, offline events are able to impact individual reaction to unrelated information within a time period because such events receive high media coverage and attention. Given the salience of mass shootings, the negative association with these events, and given exposure due to media positioning, this analysis is able to measure the impact of an event on individual reaction.

2.2 Online Engagement Sample

Employing the New York Times Archive and Community APIs, I extract user comment data on unrelated articles published up to seventy-two hours before the exogenous event occurrence.¹ Furthermore, to remove the potential for bias associated with an article's content, only articles with comments before and after the event are considered. Thus, the sample seeks to describe changes within a single article before and after the event. To control for temporal confounding variables, I employ a parallel control time trend based on data from the preceding month.² Each event has, on average, 355 associated articles, and 42,536 associated comments. User and comment data includes a user identification number, publication time, comment content, and comment identification number.

¹The New York Times Archive API extracts relevant article URLs, and the New York Times Community API extracts relevant comment data from a specific article URL.

²For the four weeks before the event, I extract all articles up to seventy-two hours before the same time and day of the original event. This analysis applies this temporal delineation irrespective of treated event's temporal proximity to each other.

Table I: Offline Events Summary Statistics

Event	Victims	Treated URLs	Treated Comments	Avg Vader Score	Avg TextBlob Score
(1) Christchurch, New Zealand, Attack (2019)	91	480	71,349	0.051	0.078
(2) Dayton, Ohio, Shooting (2019)	27	220	21,649	0.059	0.087
(3) El Paso, Tex, Shooting (2019)	46	200	18,798	0.083	0.095
(4) Las Vegas, Nev, Shooting (2017)	500	215	33,729	0.035	0.061
(5) Little Rock, Ark, Shooting (2017)	25	399	54,988	0.030	0.062
(6) Orlando, Fla, Shooting (2016)	103	362	40,036	0.047	0.079
(7) Parkland, Fla, Shooting (2018)	34	406	53,257	0.044	0.070
(8) Pittsburgh, Pa, Shooting (2018)	18	380	34,190	0.050	0.074
(9) San Bernardino, Calif, Shooting (2015)	40	507	40,302	0.021	0.069
(10) Santa Fe, Tex, Shooting (2019)	23	381	50,249	0.049	0.084
(11) Sutherland Springs, Tex, Shooting (2017)	47	340	43,660	0.029	0.076
(12) Thousand Oaks, Calif, Shooting (2018)	15	353	36,673	0.032	0.061
(13) Virginia Beach, Va, Shooting (2019)	17	502	50,458	0.054	0.080
(14) Washington, D.C., Shooting (2020)	22	234	46,176	0.053	0.079

2.3 Engagement Proxy

This study relies on comment sentiment as a user engagement proxy to measure user reaction. The sentiment score serves as an engagement proxy because it allows for the measurement of individual tone and response to a text. To determine a comment’s sentiment score, I employ both [Hutto and Gilbert \(2015\)](#)’s VADER and [Loria \(2018\)](#)’s Textblob sentiment analysis tools. Both methods employ lexical dictionaries to assess the intensity and tone of the text fragment and provide sentiment scores ranging from -1 and 1. [Table I](#) displays figures regarding the sentiment group by treatment status and before or after event status.

3 Empirical Strategy

I apply [Gardner \(2021\)](#)’s two-step regression model to test the hypothesis that offline events impact online behavior.³ This analysis exploits treatment versus the control group variation, and before and after event group variation. The treatment group in my analysis consists of the articles published within the three days before the exogenous event that have comments both before and after the event. The control group for my analysis consists of the aggregated parallel time-trend for the four weeks before the true event occurred.

$$Y_{t,c} = \alpha + H_{t,c} + D_{t,c} + H_{t,c} * D_{t,c} + \epsilon_{t,c} \quad (1)$$

$$Y_{t,c} = \beta_0 + \widehat{D}_{t,c} + \widehat{H}_{t,c} + \widehat{H_{t,c} * D_{t,c}} + \beta_1 * E_{t,c} + u_{t,c} \quad (2)$$

In this regression, the dependent variable for comment c is the comment sentiment score before or after event designation t . H refers to the hour-of-day, and D refers to the day-of-week. E represents an event dummy. $\epsilon_{t,c}$ and $u_{t,c}$ refer to error coefficients. β_1 refers to the change in sentiment score. A key facet of this two-step model is that Equation (1) includes data only from the control group. Equation (1) estimates the time-trend fixed effects for the hour-of-day, the day-of-week, and the interaction of the hour-of-day and day-of-week. Equation (2) then includes the predicted fixed effects ($\widehat{H}_{t,c}$, $\widehat{D}_{t,c}$, and $\widehat{H_{t,c} * D_{t,c}}$) from Equation (1) when analyzing the effect of the event dummy on the comment sentiment score exclusively for the treatment time period.

³[Thakral and Tô \(2020\)](#) apply this two-stage methodology to analyze consumption habits.

Table II: Summary Statistics by Population

Phase 1: Event Treatment						
	Observations	Mean	Median	Standard Deviation	Min	Max
VADER						
Control	455,713	0.044	0.028	0.289	-0.994	0.995
Treatment	139,801	0.046	0.030	0.289	-0.993	0.993
TextBlob						
Control	455,713	0.073	0.056	0.174	-1.000	1.000
Treatment	139,801	0.074	0.057	0.174	-1.000	1.000
Phase 2: Treated Events						
	Observations	Mean	Median	Standard Deviation	Min	Max
VADER						
Before	98,050	0.045	0.027	0.290	-0.993	0.993
After	41,751	0.046	0.036	0.287	-0.993	0.982
TextBlob						
Before	98,050	0.073	0.054	0.177	-1.000	1.000
After	41,751	0.077	0.065	0.168	-1.000	1.000

All regressions are clustered at the URL level to mitigate URL level differences that may bias the regression.⁴ I also include two robustness checks: whether (1) results differed from the top 50% most salient events and whether (2) the results were driven by a compositional change in readership.⁵ This analysis suggests a causal 0.009 and 0.002 point decrease in VADER and TextBlob sentiment scores respectively.

4 Results

4.1 Impact of Shootings on Online Sentiment

The estimates from Equation (2) show a significant effect of exogenous shootings on online user engagement.

⁴While the analysis does not measure the sentiment of the articles, there is no reason to expect systematic differences between the treatment and control articles given that all articles were published before the mass shootings and that mass shootings are arguably random events. Moreover, all identification is derived from within-article variation, comparing the sentiment of comments for treatment articles, relative to control articles, before and after the mass shooting.

⁵To address the ambiguity concerning whether the changes in sentiment are driven by the exogenous event or driven by changes in the composition of users, I created a sub-sample of repeat engagers and analyzed the effect of the exogenous event on the sentiment score, as well as the effect on the frequency of user commenting.

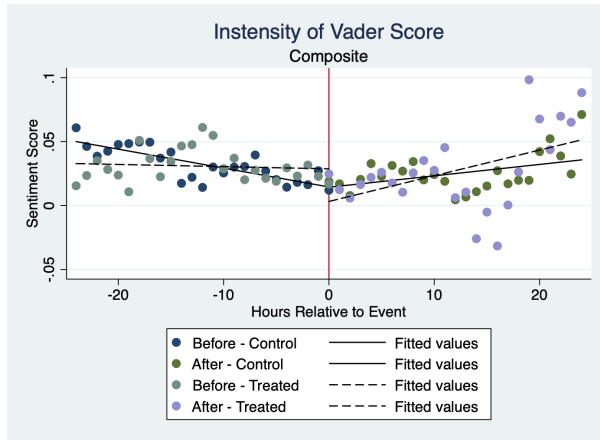


Figure 1: This graph displays the VADER sentiment score by the hours relative to the cut-off. The red line symbolizes the before and after event period. Moreover, the grey dashed lines display the treatment group, while the solid line displays the control group. In this case, the control group refers to the time-trend, and the treatment group refers to the real event time period. This graph suggests that there is no change following the event for the control group. However, there is a significant decrease following the event for the treatment group.

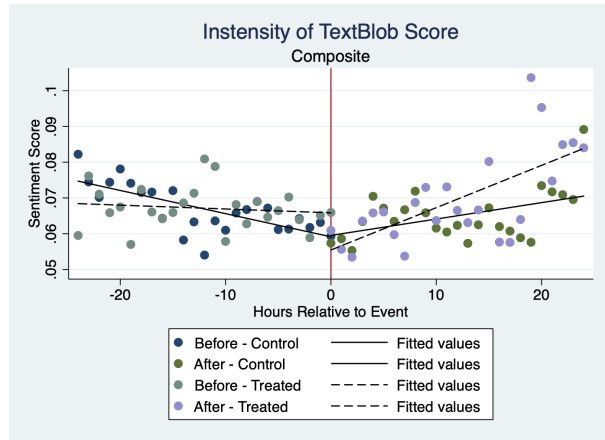


Figure 2: This graph displays the TextBlob sentiment score by the hours relative to the cut-off. This graph follows the same structure as Figure 1. Like Figure 1, this graph displays a significant decrease following the event for the treatment group. The control group exhibits no effect following the event marker.

Figures 1 and 2 display a graphical construction of the relevant effect of an exogenous, offline event on the engagement of online users. Figures 1 and 2 focus on displaying the effect of an exogenous event on the VADER and TextBlob sentiment scores respectively.⁶ This graph suggests an increase in negativity of the sentiment score following the event for the treatment group. Although these two graphs capture a similar change in sentiment score following the exogenous event, the two metrics differ in their scale. Figures 3 and 4 show how the sentiment score changes in response to each shooting event separately. These graphs decomposing the effect of exogenous shootings on online user engagement are relevant because they show the differential impact of nuances such as number of victims and degree of coverage. They also describe which events had the most dramatic change in sentiment before and after the event.

⁶These two metrics, although attempting to measure the same text sentiment, differ in their classification methodology. These tools employed to measure the sentiment score of relevant comments are far from perfect and suffer from significant biases that prevent consistent sentiment scoring, as seen through Han, Zhang, Zhang, Yang, and Zou (2018).

Table III: Two-Step Regression of Mass Shootings on Sentiment Score

	(1) Full Sample - VD	(2) Full Sample - TB	(3) Most Victims - VD	(4) Most Victims - TB
After Event	-0.009*** (0.000)	-0.002*** (0.001)	-0.017** (0.003)	-0.003** (0.002)
Constant	-0.009*** (0.001)	-0.004*** (0.000)	0.001** (0.002)	0.000*** (0.001)
Sample Size	595,514	595,514	301,131	301,131
Cluster by URL	Yes	Yes	Yes	Yes

* Significant figures at 10 percent.

** Significant figures at 5 percent.

*** Significant figures at 1 percent.

Robust standard errors in parenthesis.

This regression table employs the model described in Equations 1 and 2. This regression seeks to determine the effect of mass shootings on the VADER and TextBlob sentiment scores. Based on these results, a causal relation can be extrapolated between sentiment score and event timing.

Table IV: Two-Step Regression Analyzing Compositional Change

	(1) Frequency	(2) VD	(3) TB	(4) Most Victims - VD	(5) Most Victims - TB
After Event	-4.271 (0.597)	-0.009** (0.003)	-0.002*** (0.001)	-0.016** (0.004)	-0.001** (0.002)
Constant	2.743 (0.426)	-0.008** (0.002)	-0.004*** (0.001)	-0.001** (0.003)	0.000** (0.001)
Sample Size	473,218	473,218	473,218	236,443	236,443
Cluster by URL	Yes	Yes	Yes	Yes	Yes

* Significant figures at 10 percent.

** Significant figures at 5 percent.

*** Significant figures at 1 percent.

Robust standard errors in parenthesis.

This regression table employs the model described in Equations 1 and 2. This regression table seeks to determine the effect of a mass shooting on both frequency of commenting and sentiment score for the repeat commenter population. Like in Table III, "VD" refers to the effect on the VADER Score analyzing the entire sample of repeat commenters, "TB" refers to the effect of the TextBlob score analyzing the sample of repeat commenters, and "Most Victims" refers to the top 50% of events.

4.2 Robustness

These estimates remain statistically significant and become more salient when the sample is restricted to the top 50% of events, defined by number of victims (in terms of injuries and deaths).

Moreover, the results from Table III are statistically significant. The VADER estimates are more precise than the TextBlob estimates (99% confidence level relative to 99% and 95% in Columns 2 and 4 respectively).

Table IV seeks to unpack the compositional analysis by restricting the sample to repeat commenters. The estimates ascertained in Table IV are consistent with those determined in Table IV. Table IV suggests that a repeat user comments 4.27 fewer comments following a negative, exogenous event. These results remain statistically significant as sample size in both the repeat commenter population and event sample set constricts. Column 2 suggests a statistically significant 0.009 decrease in VADER comment sentiment. Columns 3 and 5 also exhibit significant decreases in sentiment score, although smaller in magnitude than columns 2 and 4. This table suggests that there is quite a large, significant decrease in commenting frequency following a mass shooting for the population of repeat commenters. Moreover, Columns 2 and 4 suggest that there is a causal, negative relationship between a mass shooting and online user sentiment which becomes greater in magnitude as the scale of the event increases (evident through Column 4).

4.3 Biases

This analysis displays inconsistency among text analysis classification, and noise created by these sentiment analysis tools; however, this inaccuracy supports rather than undermines the validity of these results. A key driver of this misclassification is the use of sarcasm in text and the algorithm's inability to account for the undertones of human speech evident in colloquial text. While individuals may have sarcastic comments that read both positively and negatively, according to (Riloff et al., 2013) and (Kovaz, Kreuz, & Riordan, 2013) sarcasm tends to follow a more positive literal interpretation. Given this assumption, the sarcastic bias would bias the regression upwards, reducing the effect of an exogenous mass shooting on user sentiment score. This logic still holds when considering the limitations of bias associated with sarcasm categorization in the case of a negative exogenous event, and we can expect an exacerbation of this positive bias in the case that individuals are reacting more saliently and including greater degrees of sarcasm within their texts. Considering the results exhibited in Tables III and IV, removing this bias would increase, rather than decrease, negativity. The remaining significance despite this potential positive bias speaks to the credibility of these results. However, we cannot extrapolate further regarding the potential bias attributed to sarcasm, and whether there is a differential effect on treatment or control groups. Sarcasm proves to be one of the primary confounding variables in sentiment analysis and text classification of written colloquial language. This analysis attempts to minimize sarcasm's confounding effect by playing against the positive bias. Due to the use of negative events, we reduce the likelihood of an overestimated negative effect caused by the sarcastic bias as this bias leans in the positive direction.

5 Discussion and Conclusion

The implications of these findings support the hypothesis that offline events have a significant impact on online activity. This analysis helps to inform questions regarding user behavior and user perspectives. The empirical findings suggest that following significant offline events, users exhibit atypical textual responses and frequency of engagement. A primary strength of this type of analysis is that it estimates individual perspectives following major events. Rather than employing a proxy to approximate user reaction, sentiment analysis reaches the user directly. For example, this study

shows that individuals in this sample react negatively following the news of a mass shooting. This direct individual insight can be used to inform subsequent decisions related to gun policy.

Moreover, existing research is only beginning to employ sentiment analysis as a tool to further relate the effects of major events on individual reaction. This limited economics literature can be partially attributed to the existing biases in sentiment analysis and noise present in this data. This analysis contributes to the relevant economics literature by providing a convincing scenario of decreasing the limitations of sentiment analysis and text classification. In analyzing a negative set of exogenous events, rather than a positive set, this analysis minimized the limitation of the bias associated with the categorization of sarcasm.

6 References

- Almond, D., & Du, X. (2020). Later bedtimes predict president trump's performance. *Economics Letters*, 197, 109590. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0165176520303554> doi: <https://doi.org/10.1016/j.econlet.2020.109590>
- Card, D., & Dahl, G. B. (2011, 02). Family violence and football: The effect of unexpected emotional cues on violent behavior*. *The Quarterly Journal of Economics*, 126(1), 103-143. Retrieved from <https://doi.org/10.1093/qje/qjr001> doi: 10.1093/qje/qjr001
- Eisenstein, T., & Strömberg, D. (2007, 05). News Droughts, News Floods, and U. S. Disaster Relief*. *The Quarterly Journal of Economics*, 122(2), 693-728. Retrieved from <https://doi.org/10.1162/qjec.122.2.693> doi: 10.1162/qjec.122.2.693
- Eren, O., & Mocan, N. (2018, July). Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics*, 10(3), 171-205. Retrieved from <https://www.aeaweb.org/articles?id=10.1257/app.20160390> doi: 10.1257/app.20160390
- Fedyk, A. (2022 (Forthcoming)). Front page news: The effect of news positioning on financial markets. *Journal of Finance*. Retrieved from <https://drive.google.com/file/d/1GufF1q56YA19a0b86j2EAIYekeKB8QIB/view>
- Gardner, J. (2021). Two-stage differences in differences.
- Gorodnichenko, Y., Pham, T., & Talavera, O. (2021). Social media, sentiment and public opinions: Evidence from #brexit and #uselection. *European Economic Review*, 136, 103772. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0014292121001252> doi: <https://doi.org/10.1016/j.euroecorev.2021.103772>
- Han, H., Zhang, Y., Zhang, J., Yang, J., & Zou, X. (2018, 08). Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *PLOS ONE*, 13(8), 1-11. Retrieved from <https://doi.org/10.1371/journal.pone.0202523> doi: 10.1371/journal.pone.0202523
- Hutto, C., & Gilbert, E. (2015, 01). Vader: A parsimonious rule-based model for sentiment analysis of social media text..
- Kovaz, D., Kreuz, R., & Riordan, M. (2013, 11). Distinguishing sarcasm from literal language: Evidence from books and blogging. *Discourse Processes*, 598-615.
- Lee, E. (2021, May). The new york times tops 7.8 million subscribers as growth slows. *New York Times*. Retrieved from <https://www.nytimes.com/2021/05/05/business/media/nyt-new-york-times-earnings-q1-2021.html>

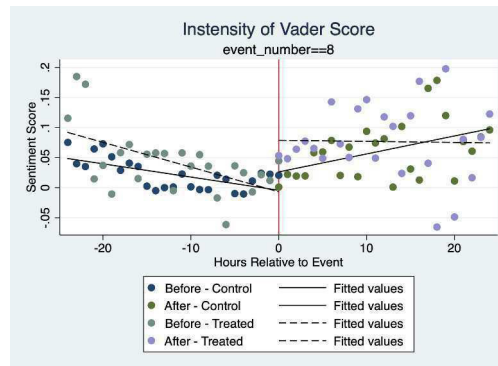
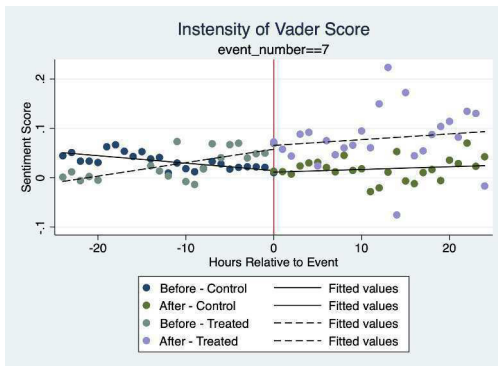
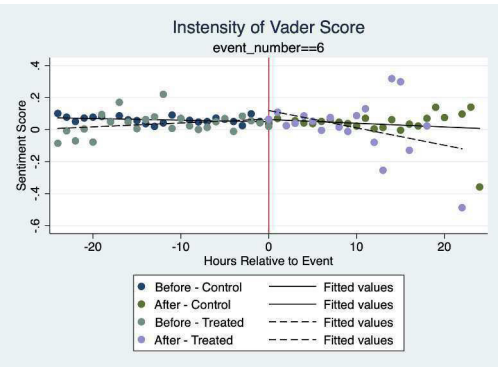
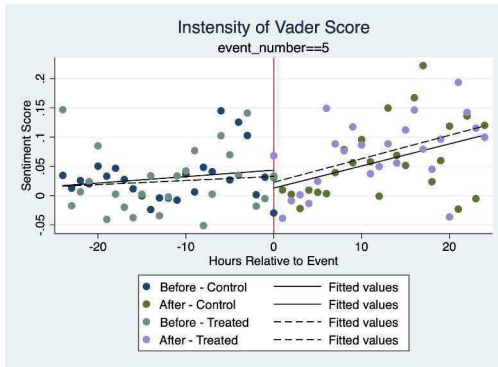
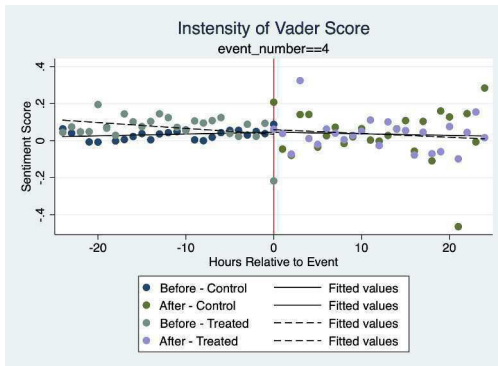
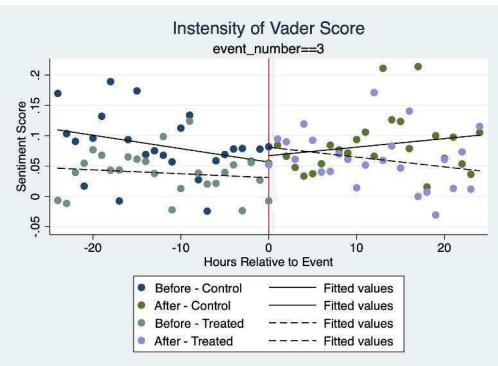
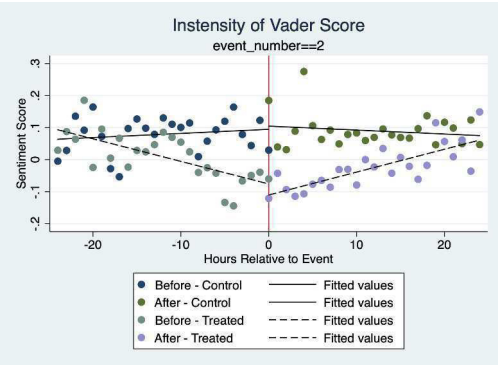
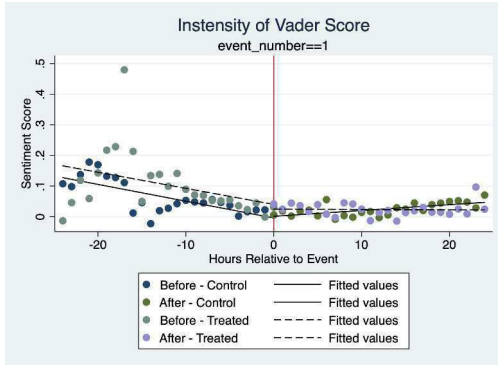
Loria, S. (2018). textblob documentation. *Release 0.15, 2*.

Philippe, A., & Ouss, A. (2018). “no hatred or malice, fear or affection”: Media and sentencing. *Journal of Political Economy*, 126(5), 2134-2178. Retrieved from <https://doi.org/10.1086/699210> doi: 10.1086/699210

Riloff, E., Qadir, A., Surve, P., Silva, L., Gilbert, N., & Huang, R. (2013, 01). Sarcasm as contrast between a positive sentiment and negative situation. *Proceedings of EMNLP*, 704-714.

Thakral, N., & Tô, L. T. (2020). Anticipation and consumption. *Mimeo*.

7 Appendix



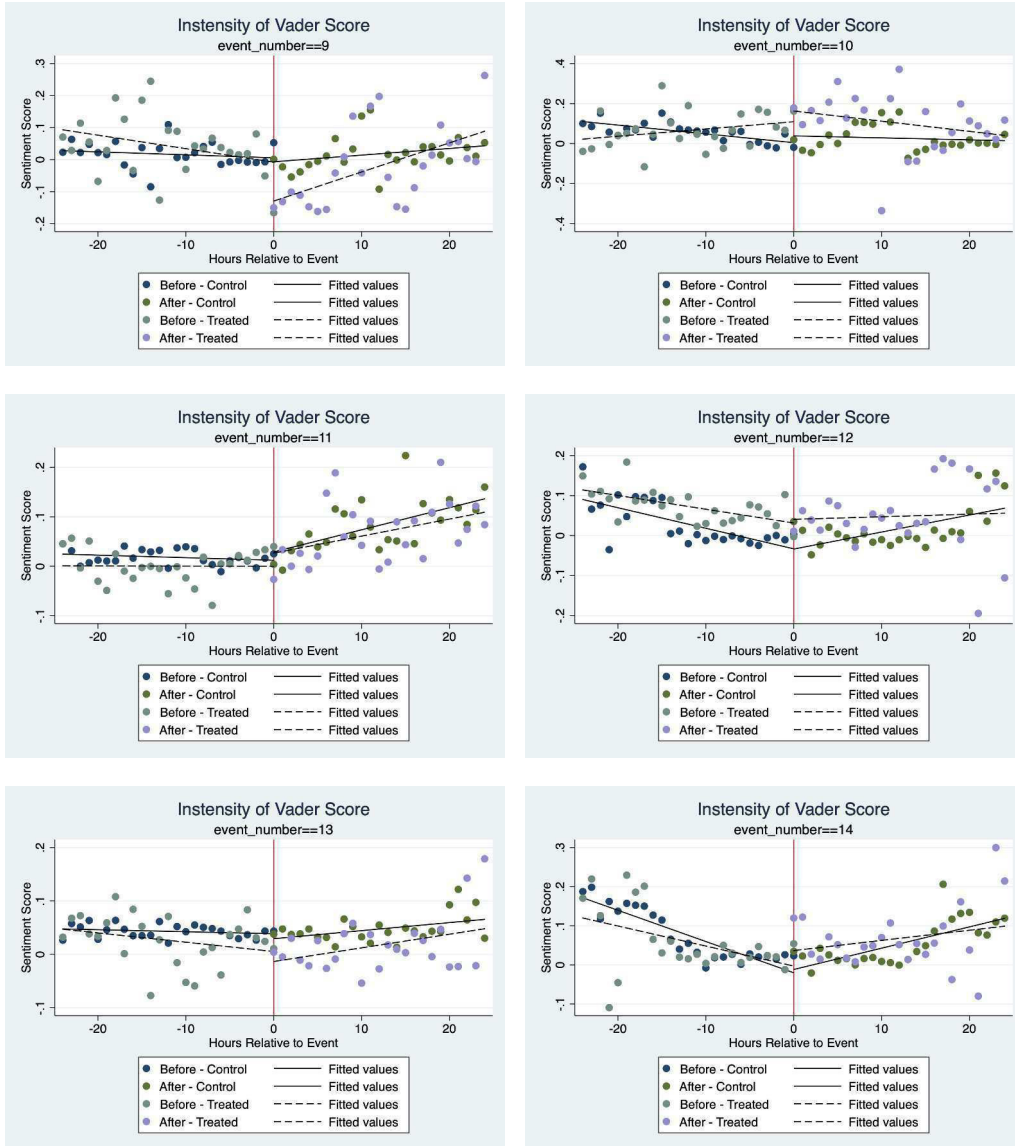
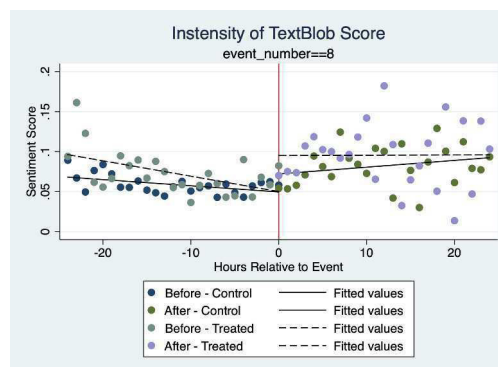
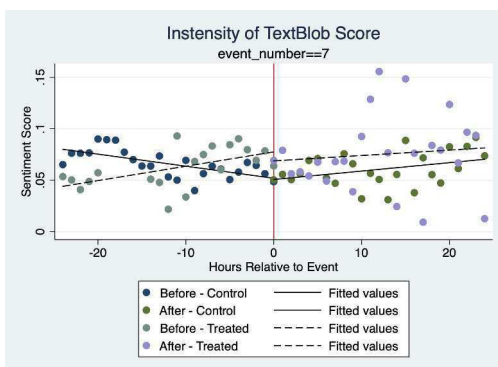
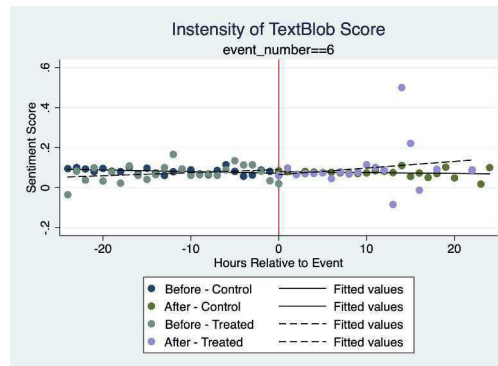
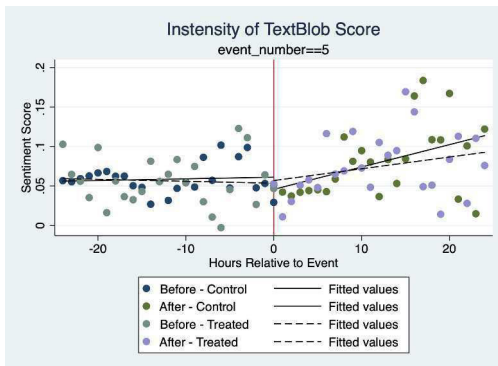
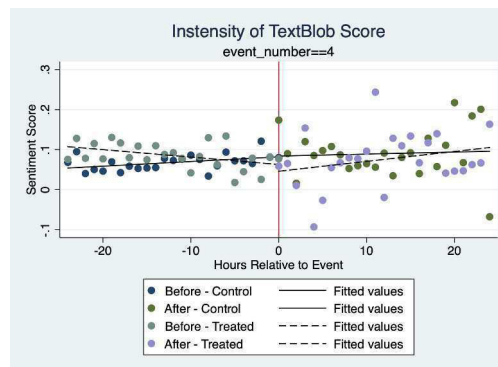
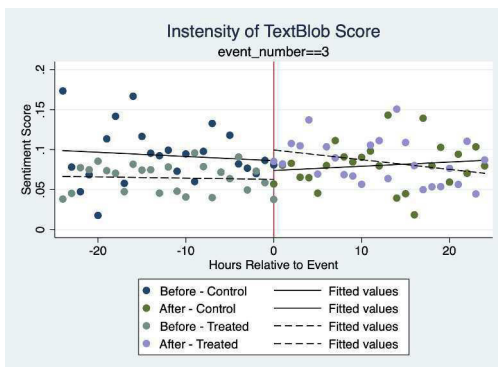
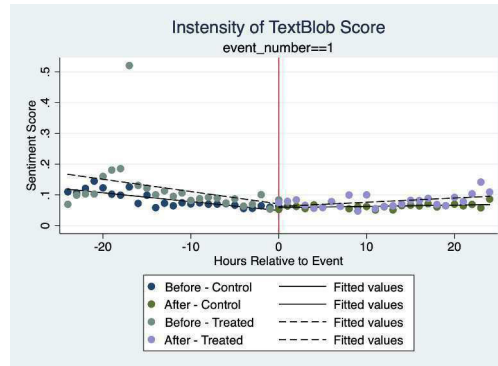
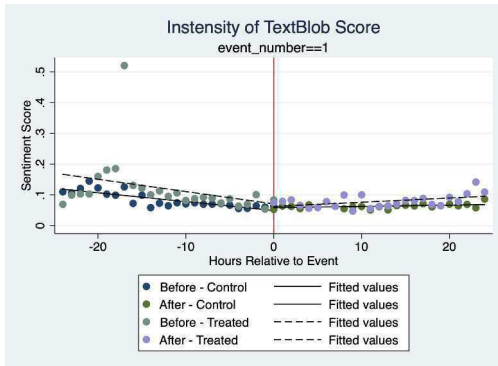


Figure 3: Intensity of Vader Score by Event.



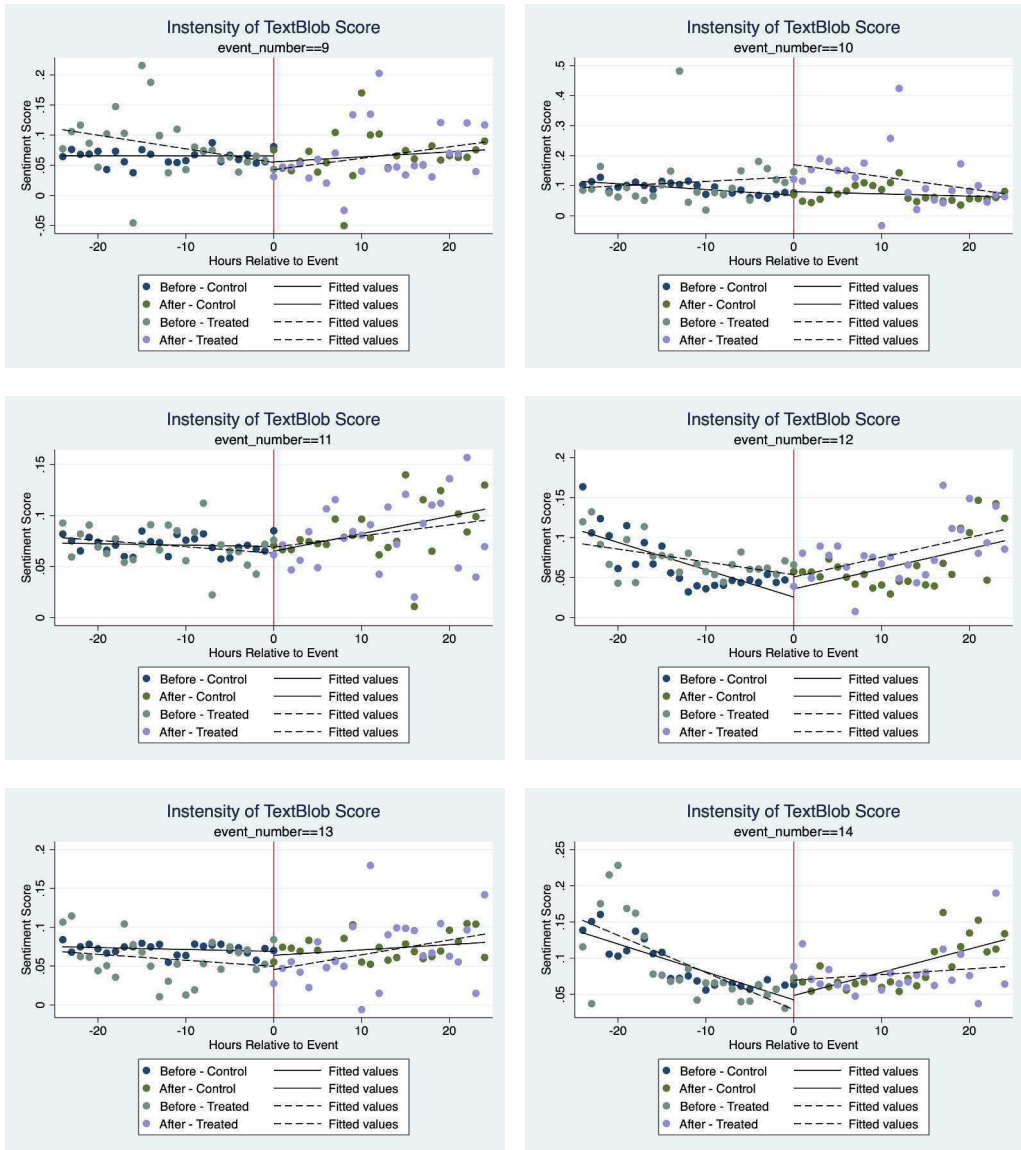


Figure 4: Intensity of TextBlob Score by Event.