Twitter Data Analysis as a Lens for Understanding the Conflict in Ukraine: A Time Series Analysis

Rachel Brasier*

December 5, 2020

Abstract

This paper determines that there are structural differences between the patterns of daily Tweets since February 2014 containing the patriotic phrase "Glory to Ukraine" in Ukrainian versus in Russian. Ukrainian military fatalities and Ukraine's Independence Day have larger postive effects on Ukrainian-language than on Russian-language Tweets, while the shock of a controversy caused by a video released during the 2018 FIFA World Cup caused a larger positive response in daily volumes of Russian-language than in Ukrainian-language Tweets.

1 Introduction

This paper provides a preliminary examination of the patterns of daily Tweets since February 2014 containing the patriotic phrase "Glory to Ukraine" in Ukrainian and in Russian.[1] I consider three exogenous variables which might engender patriotic sentiment: fatalities, Independence Day holidays, and a global discussion of the phrase "Glory to Ukraine."

The larger motivation behind this paper is to determine if the volume of Tweets containing "Glory to Ukraine" can be interpreted as a measure of Ukrainian patriotic sentiment, and if that measure could help to explain a prior finding in my research that, controlling for the number of months into the war and total Russian imports and exports, monthly Ukrainian military fatalities have a significant negative effect on seasonally-adjusted trade between Ukraine and Russia. Such a measure of patriotic sentiment would provide a valuable step in determining how much of the effect of conflict intensity on trade is based on choice, and how much is logistical.

I find that there are structural differences between the patterns of daily Tweets since February 2014 containing the patriotic phrase "Glory to Ukraine" in Ukrainian versus in Russian. Ukrainian military fatalities and Ukraine's Independence Day have larger postive effects on Ukrainian-language than on Russian-language Tweets, while the shock of a controversy caused by a video released during the 2018 FIFA World Cup caused a larger positive response in daily volumes of Russian-language than in Ukrainian-language Tweets.

*M.S. Candidate, School of Economic, Political and Policy Sciences, The University of Texas at Dallas

[1] "Glory to Ukraine" is the most widespread patriotic call in Ukraine, comparable to "Vive la France" or "God bless America." See Atlantic Council UkraineAlert (2018).

2 Review of Literature

Applications of Twitter data to understand conflict is a nascent field of study. Since Twitter itself was founded in 2006 and reached global prominence only around 2008 to 2009, there has not been much time to observe conflicts in the Twitter age. However, there is a growing body of literature on the subjects of nationalism, political polarization, and conflict through the lens of Twitter.

Bodrunova, et al. (2019) studies the different shades of left and right political polarization in Germany, the U.S., and Russia. They find that rather than being conventionally left or right, Twitter users' views are constructed in shades of anti-institutionalism, nationalism, and pro- and anti-minority views. Political polarization is also the subject of Conover, et al. (2011), which employs network and cluster analysis to detect and measure political polarization.

With regard to conflict, Makhortykh and Lyebyedyev (2015) investigate the use of the `#SaveDonbassPeople` hashtag by both the pro-government and separatist sides of the conflict in eastern Ukraine. Pantti (2019) looks at the increasing polarization of conflict reporting, which can spur trends on Twitter. Rodon, et al. (2018) looks at Twitter support for the Catalan Diada, measured by volumes of Tweets using selected hashtags in Spanish and Catalan. Zeitzoff (2011) creates a set of hourly dyadic conflict intensity scores based on Twitter and other social media sources during the 2008-2009 Gaza Conflict to measure military responses directly following important junctures in the conflict.

3 Data and Methodology

All data sources used in this paper are complete and available covering the daily time span from February 1, 2014, through June 30, 2020.

I use Jefferson Henrique's `GetOldTweets` command-line method[2] to search Twitter for a selected text query and return a list of all Tweets published from February 1, 2014, through June 30, 2020. This paper employs a Tweet keyword-search method similar to that used by Siegel (2015) to analyze Sunni-Shia sectarian Twitter wars. This approach requires significant area and language knowledge, but allows for better control of Tweet content.

Daily Ukrainian military fatalities from the Ukrainian National Military History Museum's Memory Book will serve as a proxy for the intensity of the conflict on the Ukrainian side. It is not possible at this time to compare Ukrainian military fatalities with Russian or separatist military fatalities, as information on separatist and Russian military fatalities is limited. Furthermore, Ukrainian civilian fatalities and casualties data collected by the Human Rights Monitoring Mission in Ukraine are often coded monthly, not daily, so they cannot be applied to daily analysis.

A non-conflict factor worth mentioning is the 2018 FIFA World Cup in Sochi, Russia, where on July 7, Croatian defender Domagoj Vida released a video to his fans exclaiming "Слава Украине! Эта победа за Динамо и за Украину" ("Glory to Ukraine! This victory [over Russia] is for [Kyiv] Dynamo and for Ukraine.") The video generated extensive controversy when Russia complained over the use of nationalist rhetoric, while Vida claimed

[2] https://github.com/Jefferson-Henrique/GetOldTweets-python

that he only meant to show solidarity with his former teammates at Dynamo Kyiv. The controversy spurred a Twitter storm of Tweets including the phrases "Слава Україні" or the Russian variant "Слава Украине," which constituted the highest peak since 2014 in Tweets matching that query – a phenomenon that was not directly affected by contemporaneous military fatalities. This date will be marked as an indicator variable, and also used as a shock in the impulse response functions.

Similarly, there is a much higher volume of Tweets containing the patriotic expression "Glory to Ukraine" every year on August 24, Ukrainian National Independence Day. This date will be marked as an indicator variable d_2, and impulse response functions will verify that there is only a single-day effect on Tweets.

I analyze Russian-language and Ukrainian-language Tweets separately, both to allow for differences in trends and also to identify particular idiosyncracies which could be interesting for future study. I do not consider the effects of Russian-language Tweets on Ukrainian-language Tweets because aggregated at a daily level, these variables are highly serially correlated. At a sub-daily level as in the analysis performed by Zeitzoff (2011), or with a focus on network analysis as in Siegel (2015), it may be possible to draw meaningful conclusions about exchanges between Ukrainian and Russian Twitter users on this topic as in the `#Holodomor` analysis by Rutten et al. (2013).

Table 1: Yearly "Glory to Ukraine" Tweets by Language

		Downloaded Tweets						
Text Query	Language	2014	2015	2016	2017	2018	2019	2020*
`Слава Украине`	Russian	70,631	49,831	24,539	18,557	33,236	18,861	8,287
`Слава Україні`	Ukrainian	56,960	22,474	10,952	9,438	19,311	19,073	9,007

There are some basic trends in Tweets over time that are apparent from the aggregated Twitter data (see Figure 1). For the most part, Tweets containing any of the selected patriotic phrases were more frequent in 2014 and 2015, becoming less frequent in 2016, and maintaining visibly lower levels from 2016 onward.

This pattern roughly matches the trend in conflict intensity over the course of the war (see Figure 2). The heaviest fighting and highest numbers of Ukrainian military fatalities were recorded in 2014 and 2015, with more stable conflict zones in eastern Ukraine and the conflict line not experiencing any major shifts.

Consistent with other studies finding that conflict-related fatalities tend not to be autocorrelated, the Ukrainian military fatalities series exhibits very mild ARMA(1,1) behavior.[3] There is an outlier in the series due to a spike in fatalities on August 29, 2014, corresponding to the Ukrainian retreat from the battle of Ilovaisk (BBC 2019).

[3]Based on autocorrelation and partial autocorrelation function plot diagnostics and AIC statistics (See Figures 3 and 4 in Appendix.). Model R^2 is approximately 0.20.

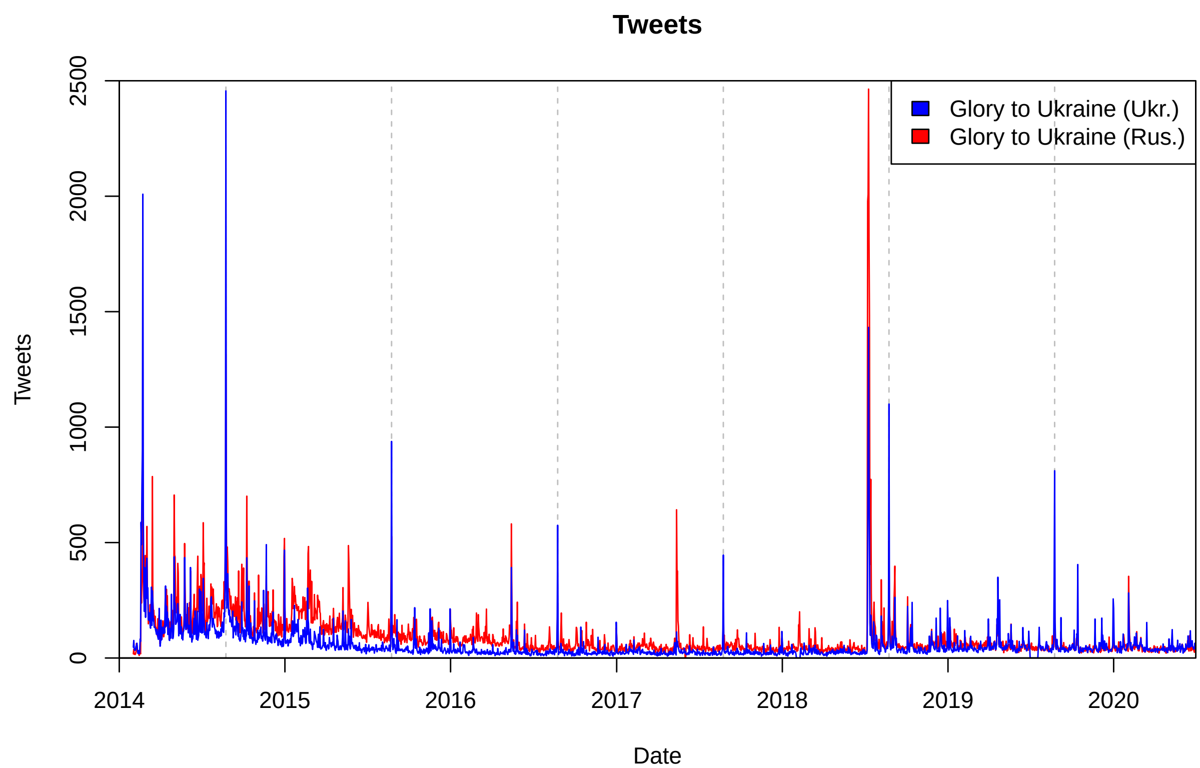


Figure 1: Daily "Glory to Ukraine" Tweets by Language. Dotted vertical lines represent Ukraine's annual Independence Day on August 24.

4 Model Identification

This paper compares four basic models for understanding the behavior motivating users to post Tweets containing the phrase "Glory to Ukraine" in Russian and Ukrainian.

The first is a multivariate regression (LM) with no autoregressive component and no time dimension:

$$y = \alpha + \beta_1 x + \delta_1 d_1 + \delta_2 d_2 + \epsilon. \tag{1}$$

The second is an autoregressive model. Autocorrelation and partial autocorrelation function results[4] find that the AR(1) model is most appropriate for both series of Tweets, based on plot diagnostics and AIC statistics.

$$y_t = \alpha + \phi_1 y_{t-1} + \epsilon_t. \tag{2}$$

The third is a distributed lag model (DL) with no autoregressive component[5]:

$$\begin{aligned} y_t = \alpha &+ \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + \delta_{1,0} d_{1,t} + \delta_{1,1} d_{1,t-1} + ... + \delta_{1,q} d_{1,t-q} \\ &+ \delta_{2,0} d_{2,t} + \delta_{2,1} d_{2,t-1} + ... + \delta_{2,r} d_{2,t-r} + \epsilon_t. \end{aligned} \tag{3}$$

[4]See Figures 5, 6, 11, and 12 in Appendix.

[5]Lag lengths determined by comparison of AIC values (not shown).

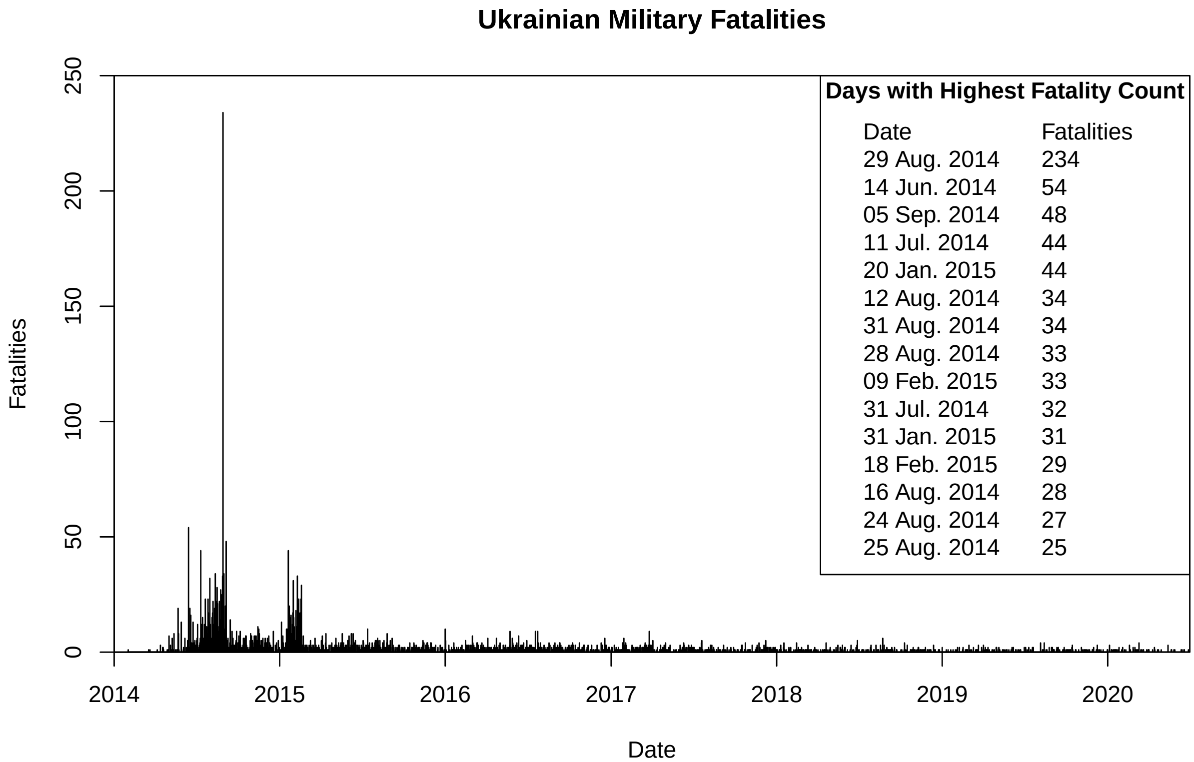


Figure 2: Daily reported Ukrainian military fatalities.

The fourth is an autoregressive-distributed lag model (ADL) with one autoregressive lag on y and lags for exogenous variables[6] (under the assumption of exogeneity):

$$y_t = \alpha + \phi_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + \delta_{1,0} d_{1,t} + \delta_{1,1} d_{1,t-1} + ... + \delta_{1,q} d_{1,t-q} \\ + \delta_{2,0} d_{2,t} + \delta_{2,1} d_{2,t-1} + ... + \delta_{2,r} d_{2,t-r} + \epsilon_t. \quad (4)$$

In all models, y represents daily volume of Tweets containing the selected phrase; x represents daily Ukrainian military fatalities, with maximum lag length p; d_1 is an indicator where $d_1 = 1$ if the date is August 24, Ukrainian Independence Day, and 0 otherwise, with maximum lag length q; and d_2 is an indicator where $d_2 = 1$ on July 7, 2018, the date the "Glory to Ukraine" FIFA World Cup controversy began, with maximum lag length r. Note that p, q, and r may differ between Russian-language and Ukrainian-language series.

5 Results

For both languages, the autoregressive distributed lag (ADL) model is best-performing, with the lowest model AIC and mean squared error (MSE), and all coefficients positive at all included lag lengths.

[6] Lag lengths determined by comparison of AIC values (not shown).

The multivariate linear regression model[7]) is worst-performing, with respective R^2 of 0.16 for Russian-language Tweets and 0.28 for Ukrainian-language Tweets. These poor results indicate that a time series approach is necessary to explain the variance in these data.

Table 2: Glory to Ukraine (Russian-Language Query)

	AR(1)		DL		ADL	
	Estimate	SE	Estimate	SE	Estimate	SE
$Intercept$	95.05	7.52	72.15	1.98	79.19	5.15
$Tweets_{t-1}$	0.79	0.01			0.7	0.02
$Fatalities_t$			1.58	0.34	0.53	0.27
$Fatalities_{t-1}$			1.38	0.34	0.82	0.3
$Fatalities_{t-2}$			1.22	0.34	0.98	0.3
$Fatalities_{t-3}$			1.47	0.32	1.42	0.29
$Fatalities_{t-4}$			1.3	0.32	1.28	0.29
$Fatalities_{t-5}$			0.8	0.32	0.76	0.29
$Fatalities_{t-6}$			1.04	0.32	0.74	0.28
$Fatalities_{t-7}$			1.05	0.32	0.53	0.25
$IndependenceDay_t$			403.7	34.83	332.45	27.45
$IndependenceDay_{t-1}$			94.09	34.84		
$WorldCup_t$			1931.83	85.15	595.03	75.24
$WorldCup_{t-1}$			2387.66	85.15	1448.16	94.62
$WorldCup_{t-2}$			1744.78	85.15	1084.28	100.78
$WorldCup_{t-3}$			1407.27	85.15	944.8	100.74
$WorldCup_{t-4}$			362.89	85.15		
$WorldCup_{t-5}$			329.49	85.15		
$WorldCup_{t-6}$			697.49	85.15		
$WorldCup_{t-7}$			207.45	85.15		
R^2	0.63		0.90		0.89	
MSE	5739.3		7247.3		4493.8	
AIC	26923.5		27432.8		26319.5	

Coefficients for AR(1), distributed lag (DL), and autoregressive distributed lag (ADL) models for explaining daily volume of Tweets containing "Glory to Ukraine" keyword in Ukrainian language.

The purely autoregressive AR(1) model is second-best performing for Russian Tweets, with an MSE of 5739.3, and third-best performing for Ukrainian Tweets, with an MSE of 7303.4. The model predicts Russian Tweets as:

$$\widehat{Tweets}_t = 95.05 + 0.79 * Tweets_{t-1},$$

[7]See Tables 4 and 5 in Appendix.

Table 3: Glory to Ukraine (Ukrainian-Language Query)

	AR(1)		DL		ADL	
	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	61.51	4.22	46.75	1.89	49.93	3.82
$Tweets_{t-1}$	0.58	0.02			0.61	0.02
$Fatalities_t$			1.26	0.33	0.45	0.26
$Fatalities_{t-1}$			0.71	0.32	0.38	0.28
$Fatalities_{t-2}$			0.74	0.33	0.64	0.28
$Fatalities_{t-3}$			1.08	0.3	1.18	0.26
$Fatalities_{t-4}$			0.55	0.3	0.66	0.26
$Fatalities_{t-5}$			0.33	0.31	0.44	0.26
$Fatalities_{t-6}$			0.41	0.3	0.33	0.26
$Fatalities_{t-7}$			0.64	0.31	0.41	0.24
$IndependenceDay_t$			986.39	33.24	837.24	27.03
$IndependenceDay_{t-1}$			79.6	33.25		
$WorldCup_t$			860.64	81.27	626.87	65.67
$WorldCup_{t-1}$			1382.54	81.27	1238.92	76.86
$WorldCup_{t-2}$			760.71	81.27	670.93	80.44
$WorldCup_{t-3}$			439.62	81.27	383.25	81.48
R^2	0.34		0.91		0.89	
MSE	7303.4		6601.4		4267.7	
AIC	27487.4		27214.9		26198.8	

Coefficients for AR(1), distributed lag (DL), and autoregressive distributed lag (ADL) models for explaining daily volume of Tweets containing "Glory to Ukraine" keyword in Ukrainian language.

and Ukrainian Tweets as:

$$\widehat{Tweets}_t = 61.51 + 0.58 * Tweets_{t-1}, \tag{5}$$

Comparison of AIC values (not shown) yields maximum lag lengths for the distributed lag (DL) model of 7 for *Fatalities*, 1 for *IndependenceDay*, and 7 for *WorldCup* for Russian-language Tweets, and 7, 1, and 3 respectively for Ukrainian-language Tweets. The DL model is third-best performing for Russian Tweets, with an MSE of 7247.3, and second-best performing for Ukrainian Tweets, with an MSE of 6601.4.

The autoregressive-distributed lag (ADL) model is best-performing, with the lowest MSE (4493.8 for Russian-language Tweets and 4267.7 for Ukrainian-language Tweets). In this model accounting for the autocorrelation of Tweet volumes on a given day with Tweet volumes on the previous day, AIC comparison (not shown) yields maximum lag lengths of 7 for *Fatalities*, 0 for *IndependenceDay*, and 3 for *WorldCup* for both languages. All coefficients are positive, indicating that Ukrainian military fatalities, Independence Day celebrations, and the World Cup video all motivate users to post more Tweets containing the patriotic phrase "Glory to Ukraine."

Further, days with Ukrainian military fatalities tend to inspire greater volumes of Tweets containing "Glory to Ukraine," with the effect remaining but weakening over the next seven days. The impact of Independence Day celebrations is highest the day of (August 24), carrying over weakly to the next day, August 25. The Independence Day effect appears to be higher for Ukrainian-language Tweets. The effect of the World Cup video appears to have elevated Tweet volumes for three days, with the effect larger for Russian-language than Ukrainian-language Tweets.

However, despite the promising statistics for the model, Ljung-Box tests[8] find that the DL and ADL models for the Russian and Ukrainian Tweet series all have an overall poor fit at all lag lengths, indicating that there is some element of the series behavior that they fail to capture. Summary analyses of the fatalities and Tweets data find that all three series are positively skewed, which might be corrected to a more normal distribution by logarithmic transformation; however, all series include zeroes which requires an alternative method for log transformation.

6 Conclusion

While none of the models explored, even the best-performing ADL model, is particularly effective at predicting daily Tweet volumes, they do consistently provide evidence to support some of my prior assumptions. There does appear to be a relationship between Ukrainian military fatalities and Tweets containing the patriotic Ukrainian phrase "Glory to Ukraine," particularly in the Ukrainian language. There are significantly higher volumes of Tweets on Ukrainian National Independence Day, August 24th. There may be other patriotic holidays (Day of the Defender of Ukraine, every October 14th starting in 2015; Victory Day on May 9th, also a major Russian holiday) that also generate stastically higher (but not as high) levels of Tweets. The very high peak in early 2014 visible in Figure 1 corresponds to the successful ouster of President Yanukovych in mid-February.

Fatalities can be modeled autoregressively, but more study will be necessary to determine the optimal model. Event count analysis such as the Bayesian Poisson VAR described by Brandt and Sandler (2012) may be more appropriate due to the generally low numbers and frequent days with zero fatalities (see Figure 2). In order to fully understand the dynamics in the fatalities time series, more research into individual events will be necessary. Furthermore, the Ukrainian military fatalities is just one measure of conflict intensity. Civilian attacks may not be closely correlated with military fatalities, but they are likely to generate similar news coverage and popular response. Here it may be useful to apply a method similar to Zeitzoff (2011) to create a complementary dataset of military actions based on news Tweets.

Models with a lagged fatalities term performed best, despite findings by Zeitzoff (2011), Rodon et al. (2018), and Rutten et al. (2013) which imply that Twitter activity happens quickly after an inciting event. It is possible that Ukrainian conflict reporting requires some time to disseminate, as with the crime reporting studied in Kounadi et al. (2015). Further research on communication methods may reveal that adding a categorical variable for day of the week may be appropriate.

[8]See Figures 8, 9, 14, and 15 in Appendix.

There may be a benefit to separating the conflict into sub-periods. In the case of fatalities, automatically generated breakpoints may be a good starting point to determine the general date range where I should expect to find news reporting on conflict developments (the Minsk agreement, ceasefires, the Normandy format negotiations, etc.). In the case of Tweets, there may be another major change in behavior in late 2018, possibly due to the Kerch Strait incident in the Sea of Azov. Here it will also be necessary to analyze the content of Tweets to verify that they are being correctly interpreted as expressions of patriotism and national pride, and are not in fact news-reporting Tweets.

The Twitter storm generated by the release of Vida's "Glory to Ukraine" video during the 2018 FIFA World Cup had a huge impact on Russian-language Tweets containing the phrase and a smaller effect on Ukrainian-language Tweets. However, as Rutten et al. (2013) find a generally large degree of intercommunication between Ukrainian- and Russian-language Twitter users, many of these conversations might be multilingual. Rutten et al. (2013) also use hashtag searches to validate assumptions about relevant keywords and sentiments, and consider overall Ukrainian Twitter user demographics, which is a necessary next step.

The Twitter data hint at a few potential trends which would be interesting to explore in future research including text and network analysis. First, the trends in Ukrainian and Russian Tweets might reflect the national emphasis on reasserting the dominance of the Ukrainian language and the growing stigma against communicating in the Russian language. User analysis could determine the growing relative prominence of Ukrainian-language over Russian-language Ukrainian patriotic Tweets is due to users switching from Russian to Ukrainian over time, or if Russian-speaking users stop posting Ukrainian patriotic phrases as the war progresses.

References

[1] Atlantic Council UkraineAlert. July 12, 2018. Why the Russian fuss over "Fascist" salute at World Cup backfired. The Atlantic Council. Accessed from https://www.atlanticcouncil.org/blogs/ukrainealert/why-the-russian-fuss-over-fascist-salute-at-world-cup-backfired/.

[2] BBC Ukrainian. Ukraine's deadliest day: The battle of Ilovaisk, August 2014. BBC Ukrainian. Accessed from https://www.bbc.com/news/world-europe-49426724.

[3] Bodrunova, S., Blekanov, I., Smoliarova, A., and Litvinenko, A. (2019). Beyond left and right: real-world political polarization in Twitter discussions on inter-ethnic conflicts. Media and Communication, 7(3), 119:132.

[4] Brandt, P., and Sandler, T. (2012). A Bayesian Poisson Vector Autoregression Model. Political Analysis, 20(3), 292:315.

[5] Conover, M. D., Ratkiewicz, J., Francisco, M., Goncalves, B., Flammini, A., and Menczer, F. (2011). Political polarization on Twitter.

[6] Kounadi, O., Lampoltshammer, T.J., Groff, E., Sitko, I., and Leitner, M. (2015). Exploring Twitter to analyze the public's reaction patterns to recently reported homicides in London. PLoS ONE, 10(3), 1:17.

[7] Makhortykh, M., and Lyebyedyev, Y. (2015). `#SaveDonbassPeople`: Twitter, propaganda, and conflict in Eastern Ukraine. The Communication Review, 18, 239:270.

[8] Memory Book. 2014-2020. ”Memory Book of Those Fallen for Ukraine.” Ukrainian National Military History Museum. http://memorybook.org.ua/ (accessed July 20, 2020).

[9] Pantti, M. (2019). The personalization of conflict reporting. (2019). Digital Journalism, 7(1), 124:145.

[10] Rodon, T., Martori, F., and Cuadros, J. (2018). Twitter activism in the face of nationalist mobilization: the case of the 2016 Catalan Diada. Universitat Oberta de Catalunya Revista D'Internet, Dret I Politica, 26, 16:29.

[11] Rutten, E., Fedor, J., and Zvereva, V., eds. (2013). Memory, Conflict and New Media: Web wars in post-socialist states, `#Holodomor`: Twitter and public discourse in Ukraine, 82-98.

[12] Siegel, A. (2015). Sectarian Twitter wars: Sunni-Shia conflict and cooperation in the digital age. Carnegie Endowment for International Peace.

[13] Zeitzoff, T. (2011). Using social media to measure conflict dynamics: an application to the 2008-2009 Gaza Conflict. Journal of Conflict Resolution, 55(6), 938:969.

Appendix

Fatalities

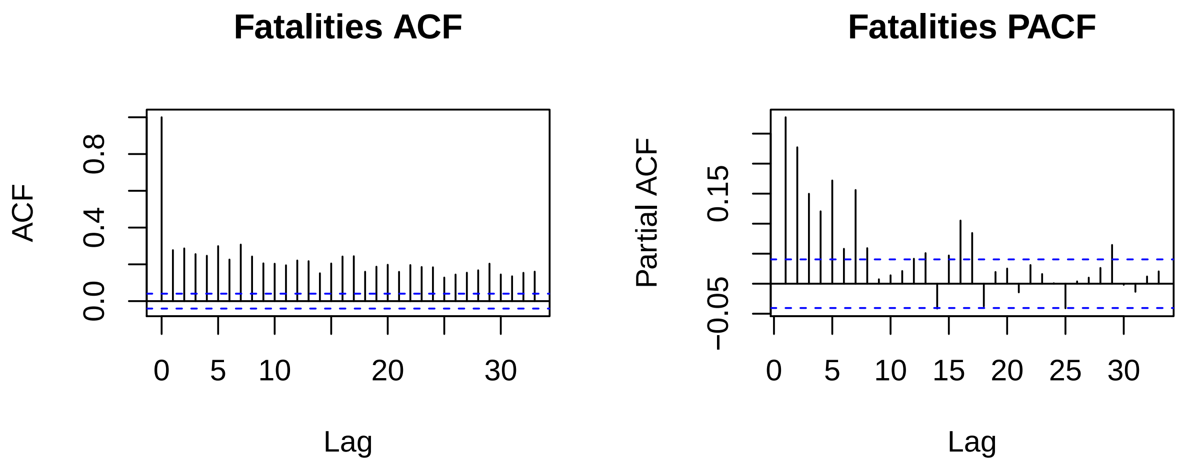


Figure 3: The fatalities ACF and PACF appear to demonstrate ARMA(1,1) behavior with no noticeable seasonality patterns. This is supported by the fact that the ARMA(1,1) model was the best-performing autoregressive model, with an AIC of 14643.93 and an overall R^2 of 0.2018.

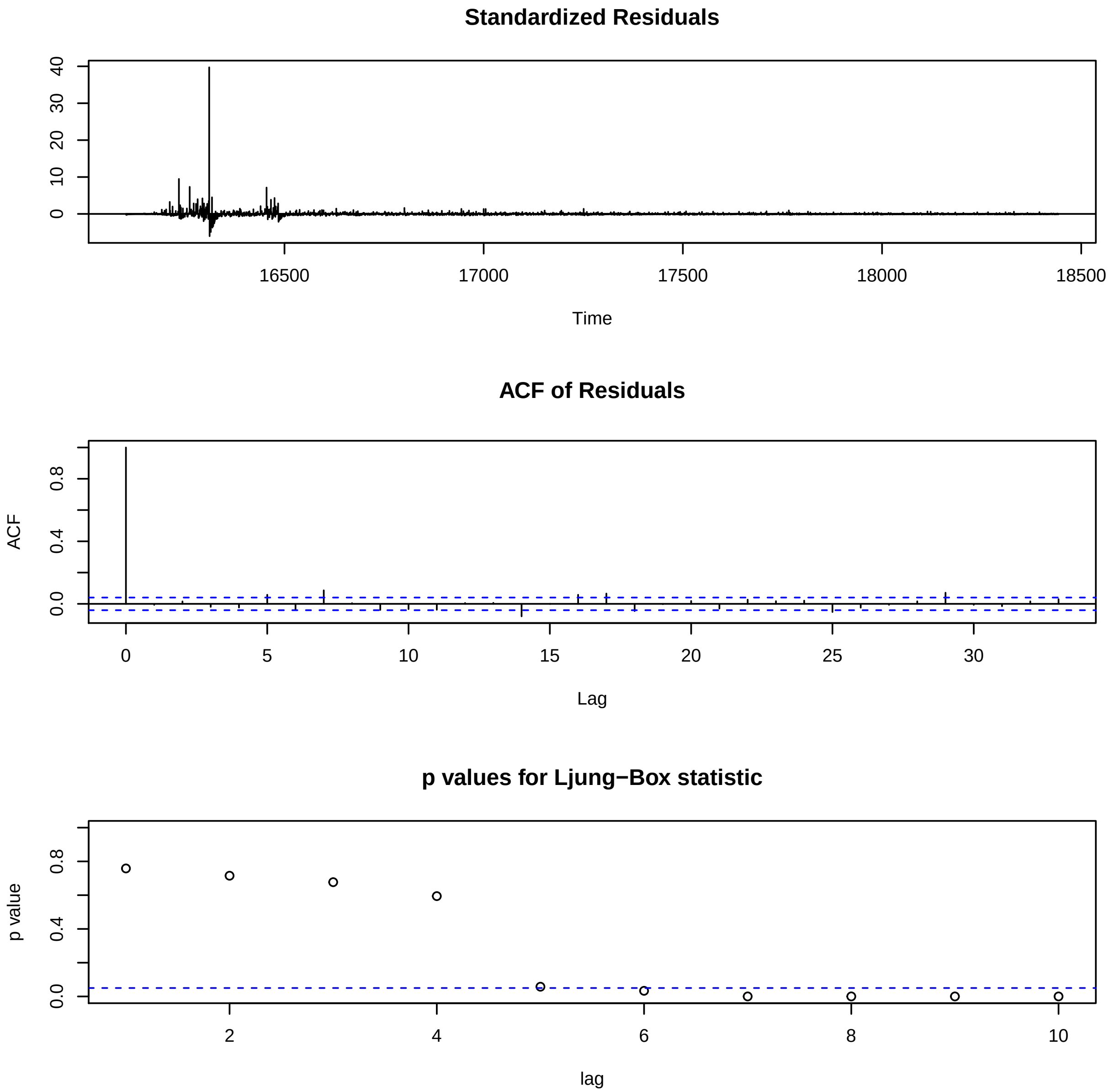


Figure 4: Despite a spike in residuals toward the beginning of the study period (due to the very large number of fatalities on August 29, 2014 (see Figure 2), the Ljung-Box test finds that overall the model fit is good.

Russian-Language Query "Glory to Ukraine"

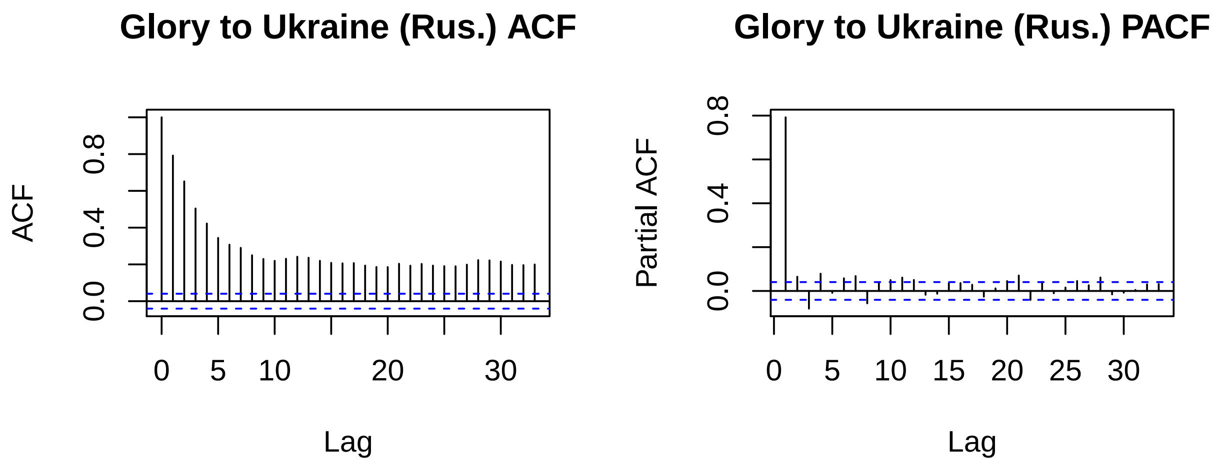


Figure 5: The Glory to Ukraine (Russian language Tweets) ACF and PACF show AR(1) behavior. The AR(1) model is the best-fitting autoregressive model, with an AIC of 26923.52 and an R^2 of 0.6276.

Table 4: Glory to Ukraine (Russian-Language Query) – No exogenous lags

	AR(1)		Linear Model		ADL	
	Estimate	SE	Estimate	SE	Estimate	SE
α	95.05	7.52	86.63	2.46	94.21	7.80
Φ_1	0.79	0.01			0.81	0.01
Fatalities			3.56	0.38	0.08	0.22
Independence Day			405.12	46.53	305.91	23.20
World Cup			1917.81	113.81	-164.90	57.20
R^2	0.63		0.16		0.89	
MSE	5739.3		12924.3		5327.4	

The ADL, the best-fitting model, finds that the Tweets can be modeled as:

$$Tweets_t = 94.1841 + 0.7684 * Tweets_{t-1} + 0.0670 * Fatalities_t \\ + 310.4646 * IndependenceDay_t - 96.2391 * WorldCup_t$$

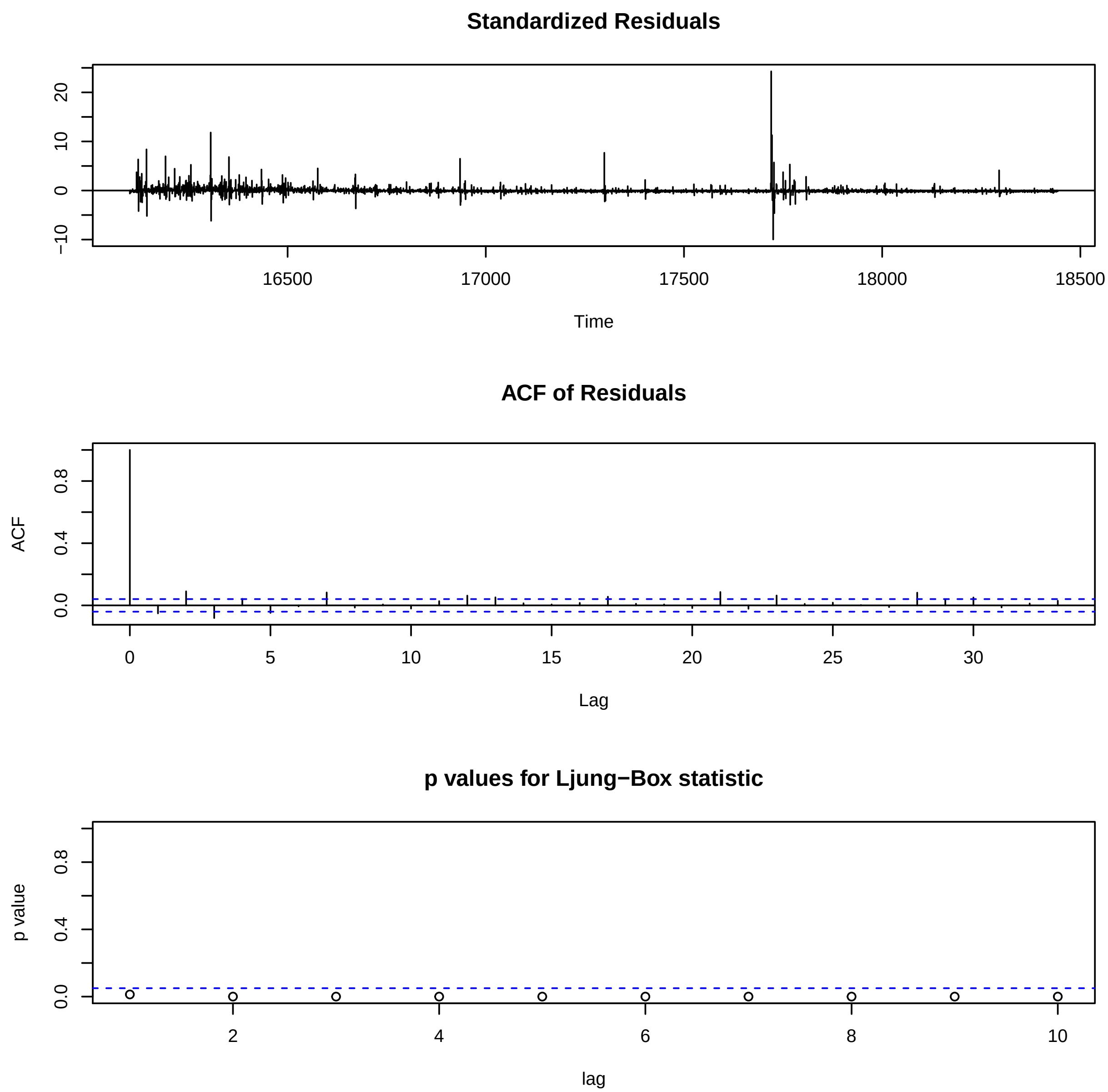


Figure 6: Although the overall explanatory power of the autoregressive model for the Glory to Ukraine (Russian) Tweets was high, the Ljung-Box test finds that the model has a significantly poor fit at all tested lag lengths. It exhibits a very high residual spike in July 2018, when the Russian Tweets peaked.

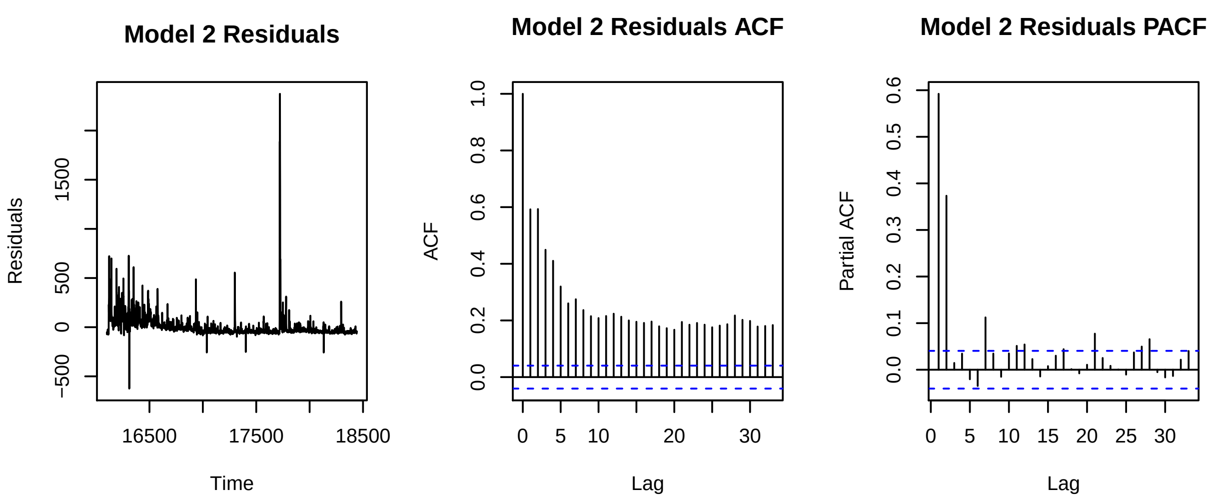


Figure 7: The linear model residuals appear to exhibit unmodeled AR(2) behavior.

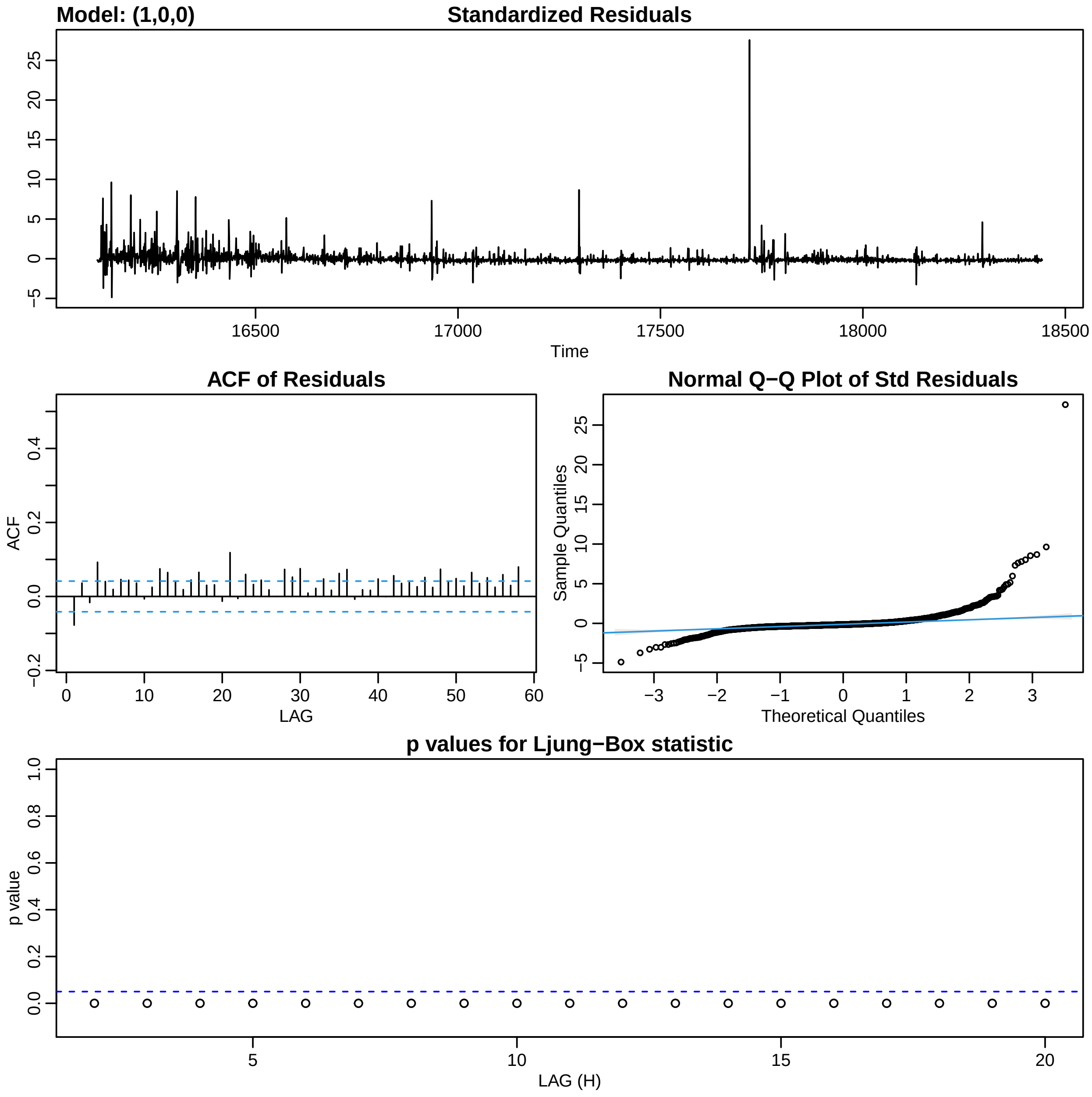


Figure 8: Despite the high R^2, the Ljung-Box test finds that the ADL model has an overall poor fit for all lags.

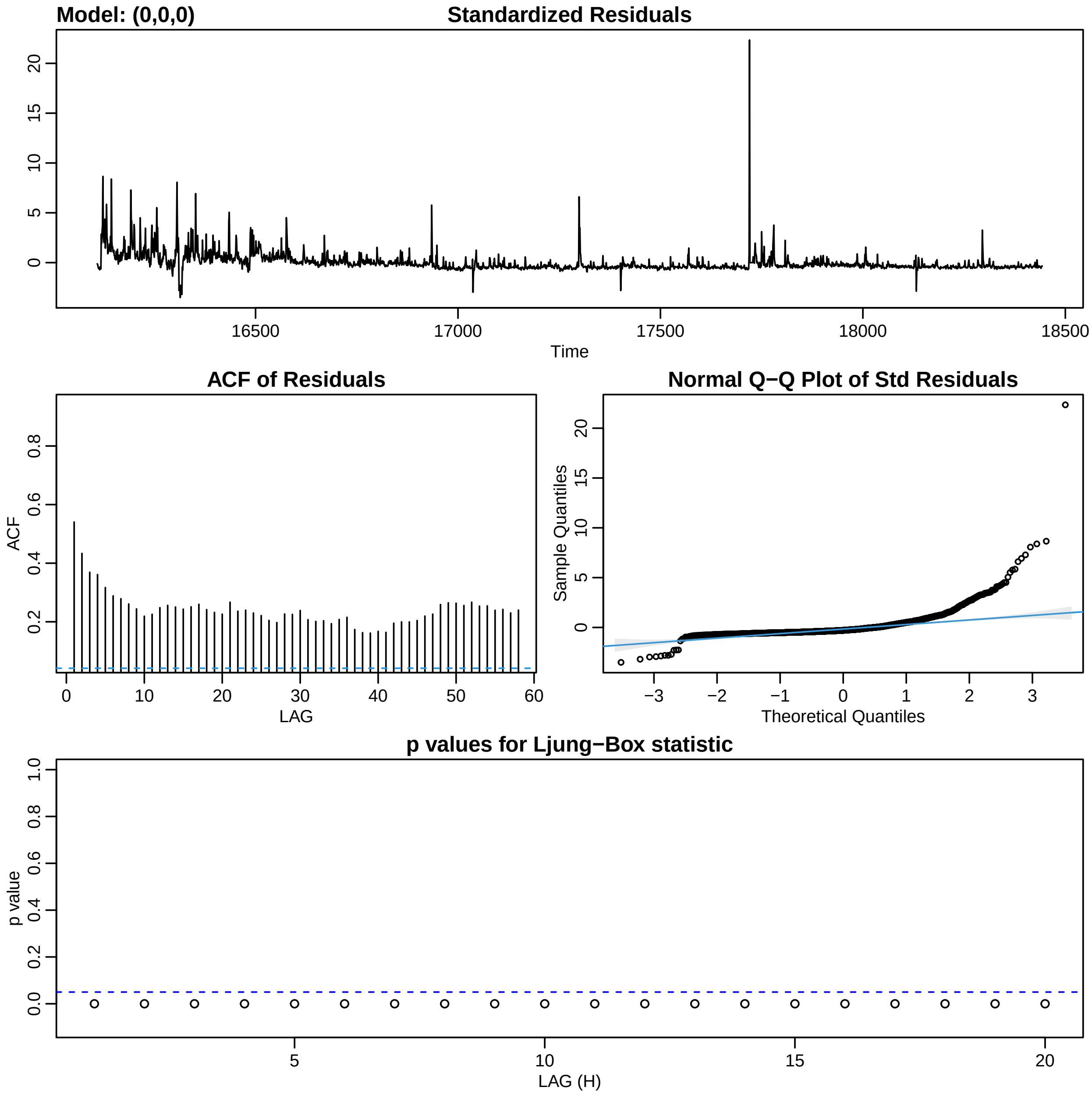


Figure 9: Despite the high R^2, the Ljung-Box test finds that the ADL model has an overall poor fit for all lags.

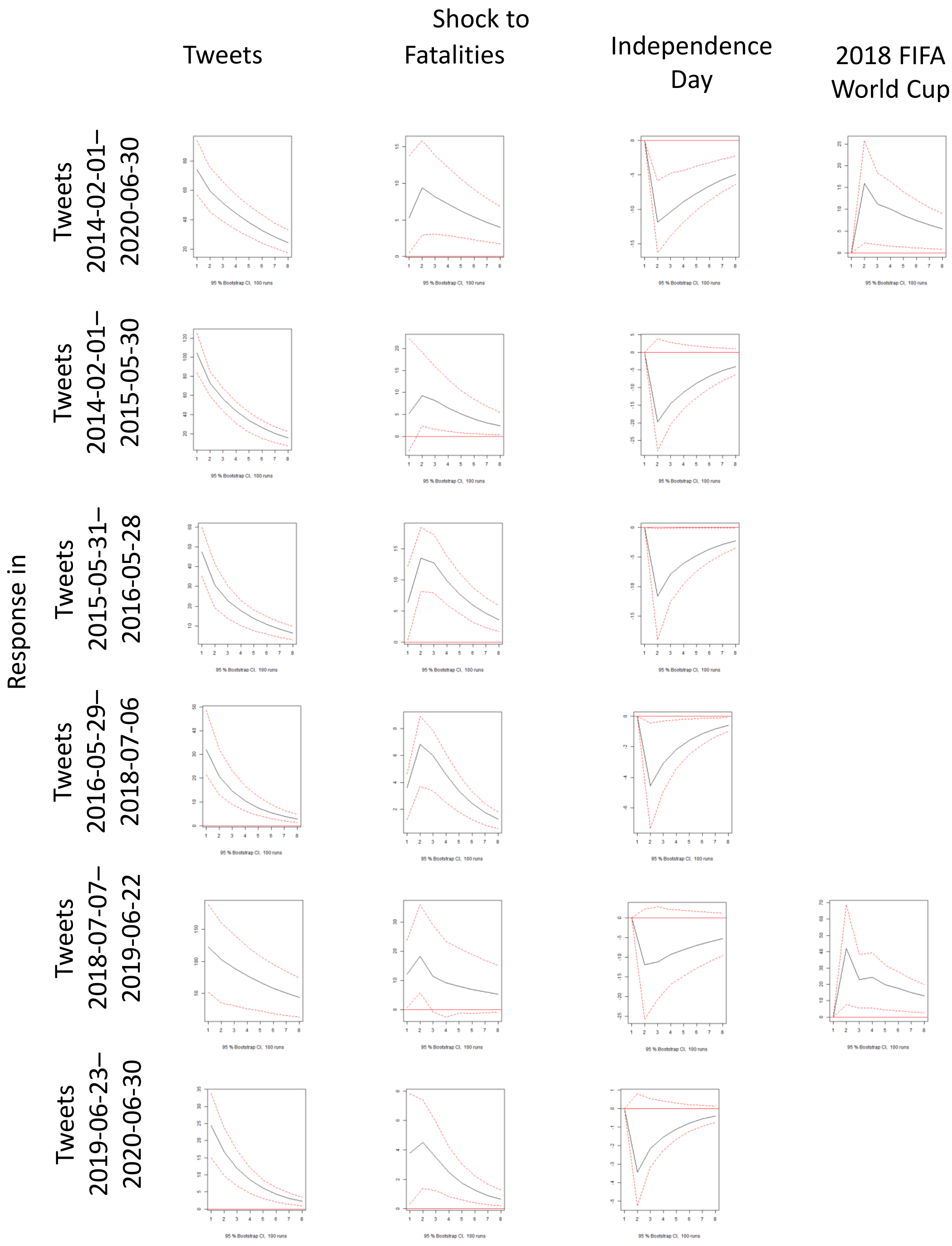
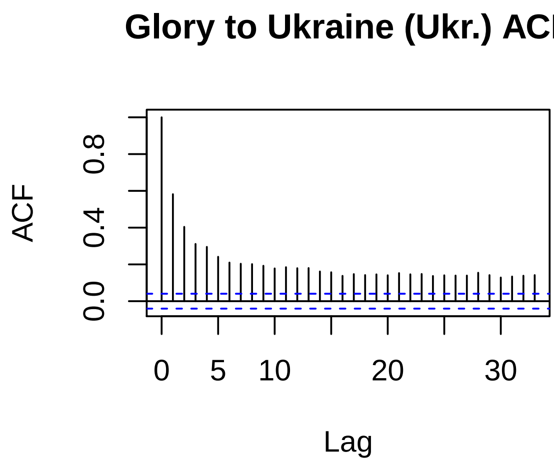


Figure 10: Impulse response function plots showing responses in Tweet volumes to shocks in Tweets, fatalities, Independence Day, and World Cup video during overall time period and windows (selected via breakpoints method).

Ukrainian-Language Query "Glory to Ukraine"



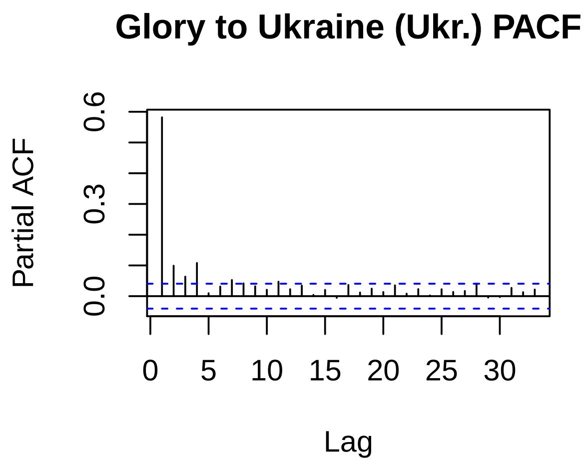


Figure 11: The Glory to Ukraine (Ukrainian language Tweets) ACF and PACF show AR(1) behavior. The best-fitting autoregressive model, with an AIC of 27487.38 and an R^2 of 0.3386, is indeed the AR(1).

Table 5: Glory to Ukraine (Ukrainian-Language Query) – No exogenous lags

	AR(1)		Linear Model		ADL	
	Estimate	SE	Estimate	SE	Estimate	SE
α	61.51	4.22	54.24	1.93	58.65	4.22
Φ_1	0.58	0.02			0.66	0.02
Fatalities			2.38	0.30	0.38	0.23
Independence Day			987.81	36.58	836.65	23.63
World Cup			853.38	89.47	48.38	58.26
R^2	0.34		0.28		0.89	
MSE	7303.4		7987.5		4788.8	

The ADL, the best-fitting model, finds that the Tweets can be modeled as:

$$Tweets_t = 58.6497 + 0.6595 * Tweets_{t-1} + 0.3806 * Fatalities_t + 836.6485 * IndependenceDay_t,$$

with $WorldCup_t$ being statistically insignificant.

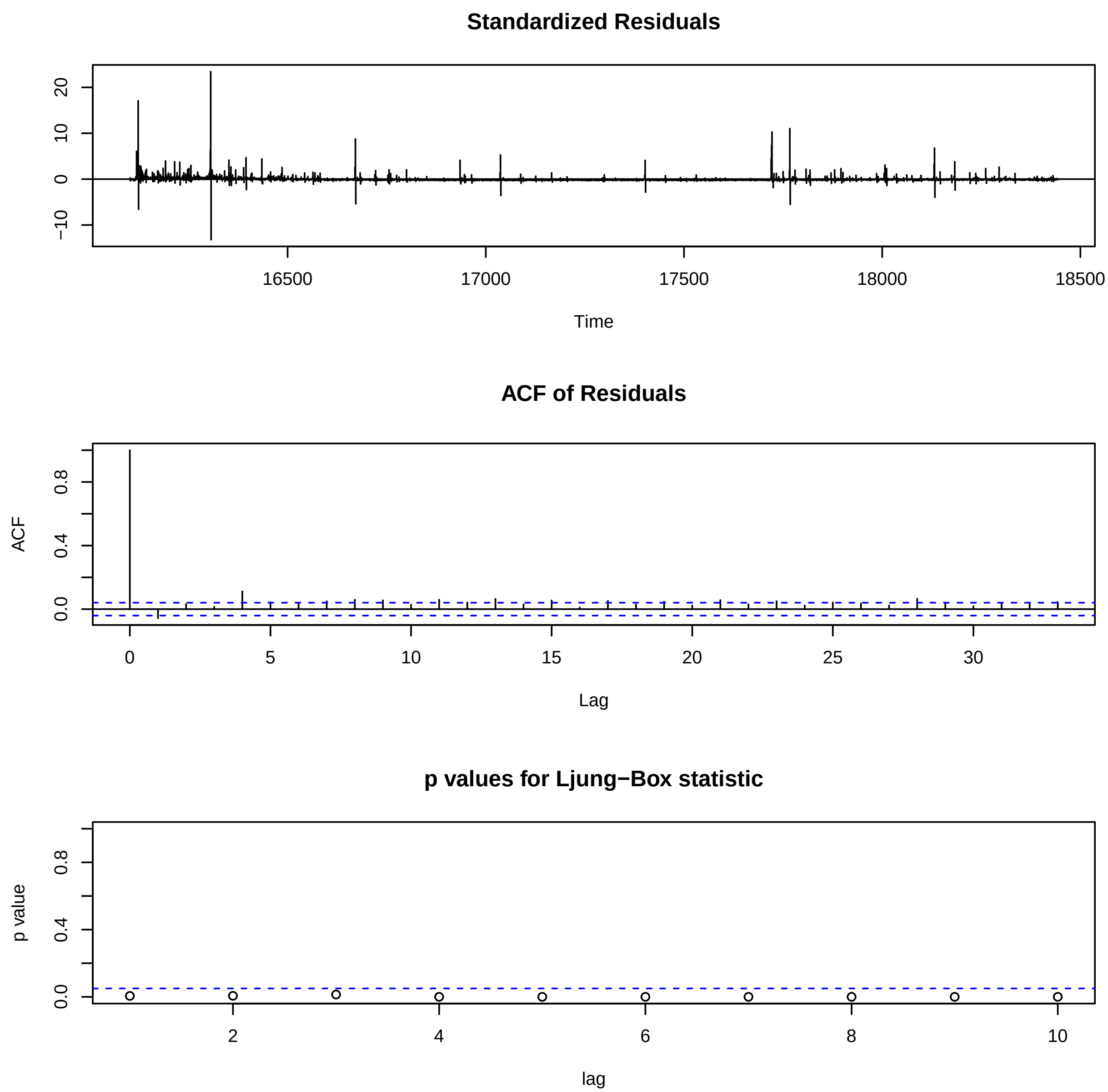


Figure 12: While it does not have the same residuals spike in July 2018, the Glory to Ukraine (Ukrainian) Tweets autoregressive model is also found to be ill-fitting at all tested lag lengths.

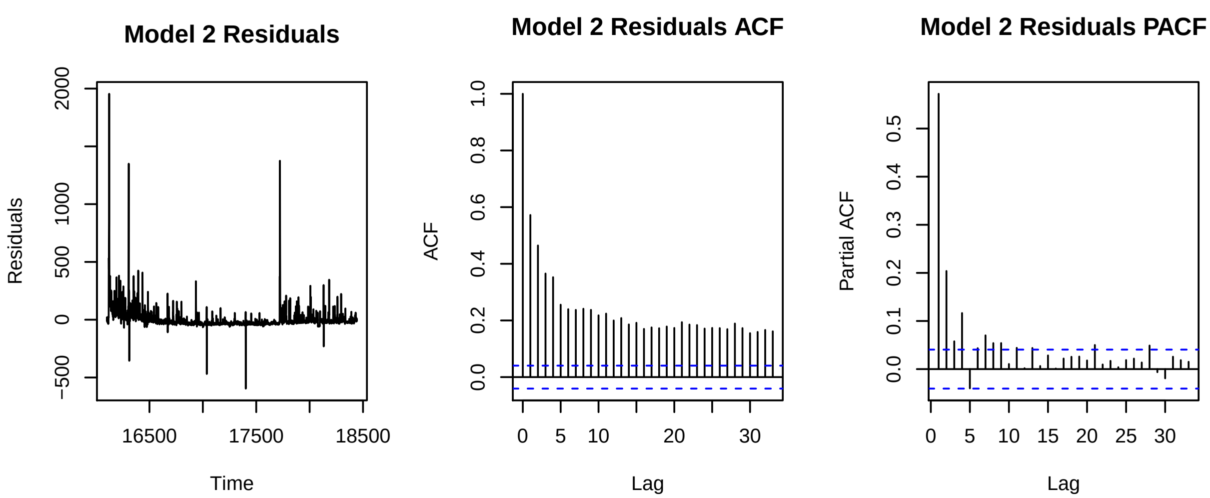


Figure 13: The linear model residuals appear to exhibit unmodeled AR(1) behavior.

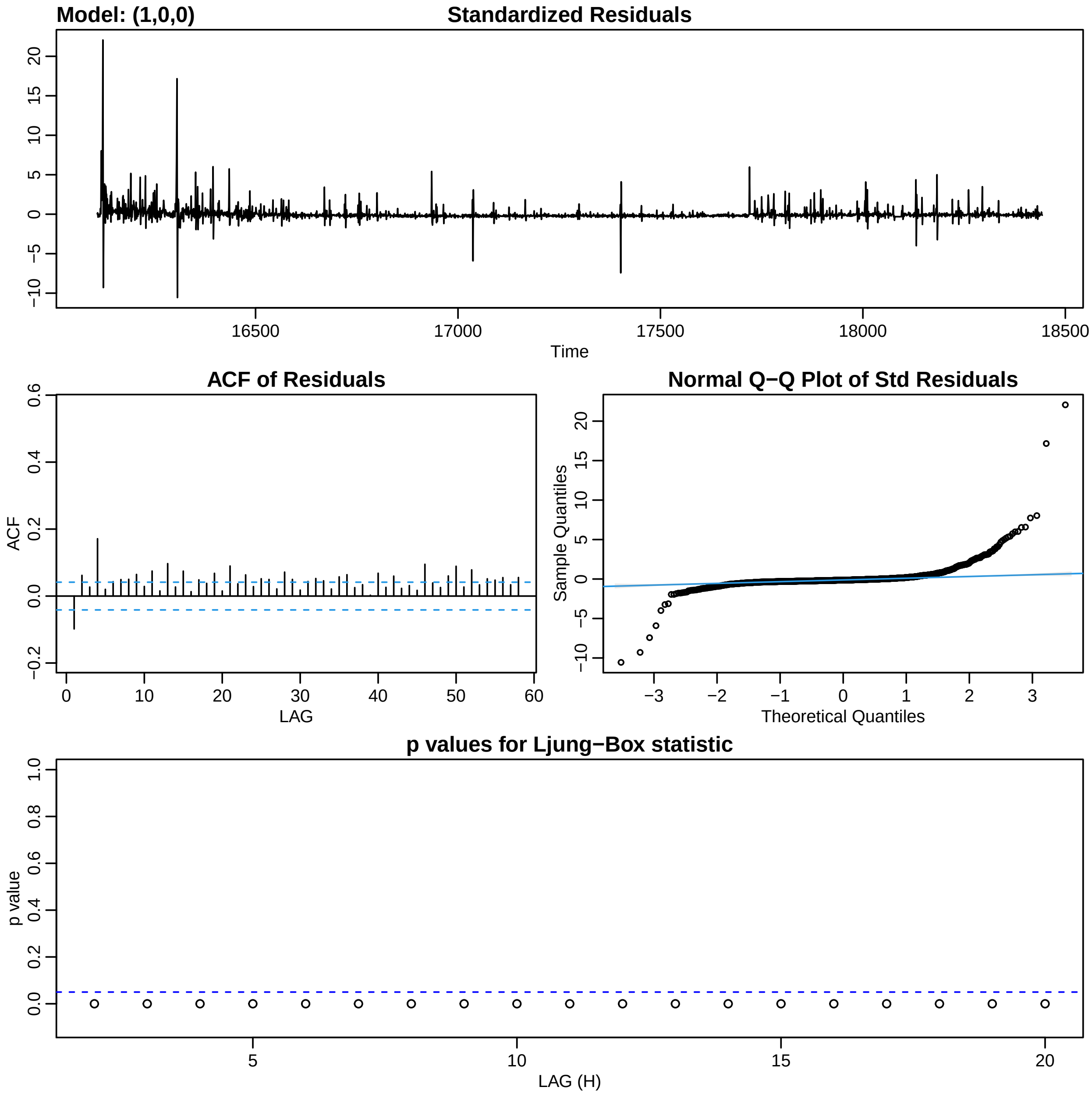


Figure 14: Despite the high R^2, the Ljung-Box test finds that the ADL model has an overall poor fit for all lags.

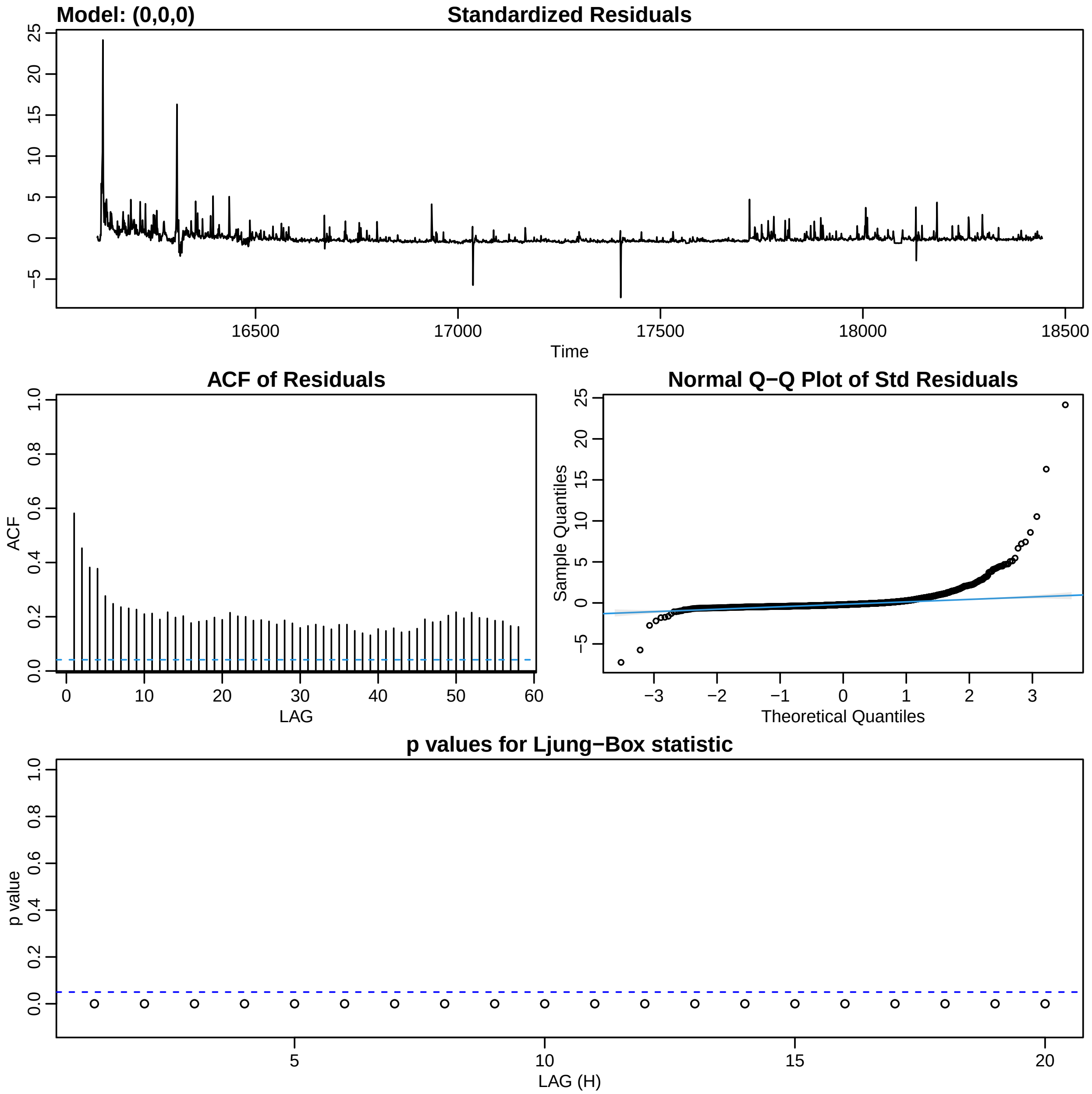


Figure 15: Despite the high R^2, the Ljung-Box test finds that the ADL model has an overall poor fit for all lags.

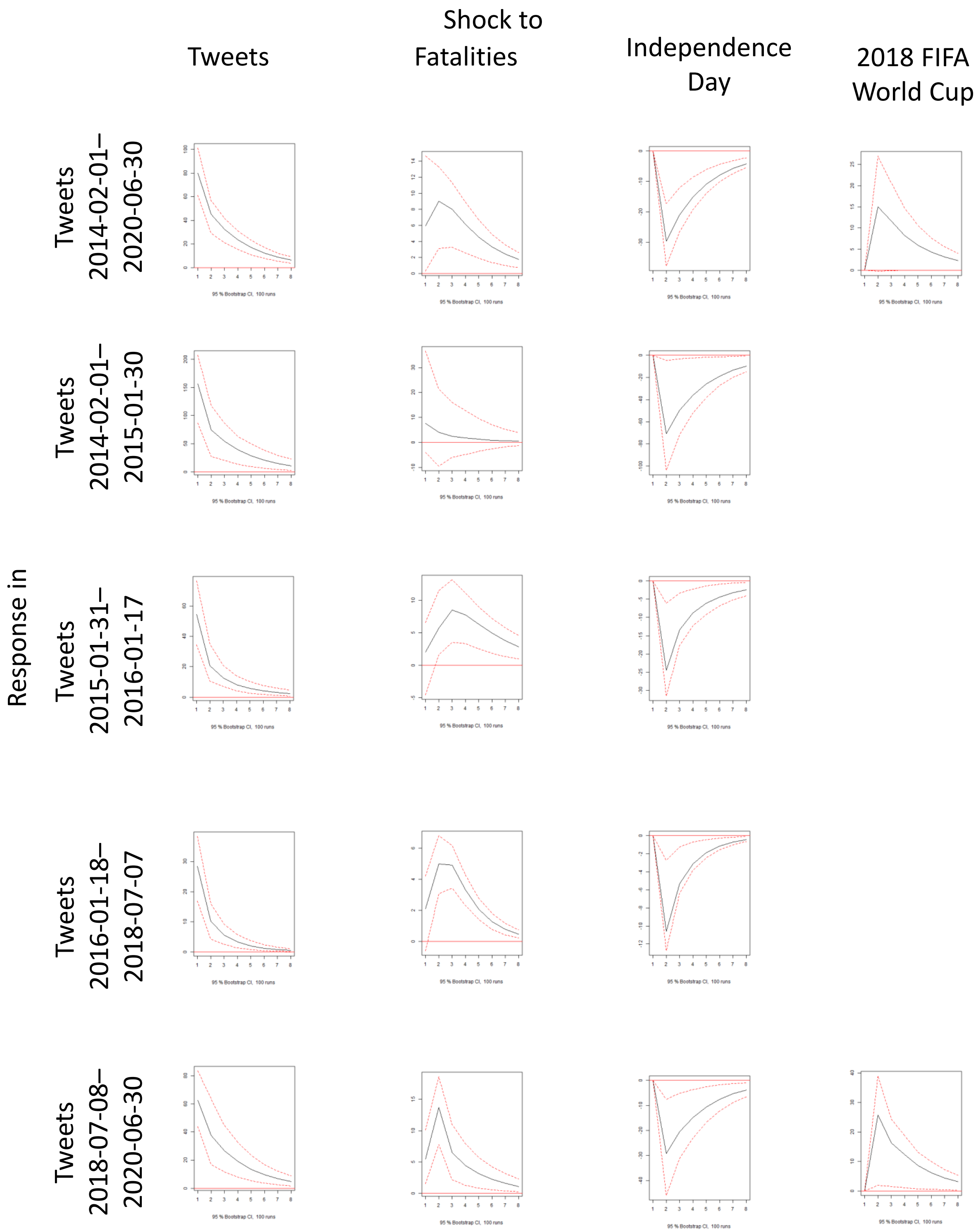


Figure 16: Impulse response function plots showing responses in Tweet volumes to shocks in Tweets, fatalities, Independence Day, and World Cup video during overall time period and windows (selected via breakpoints method). Note that these breakpoints differ from those implied by the Russian-language Tweets, as the World Cup had a smaller effect on Ukrainian-language Tweets.