Why Botter:

How Pro-Government Bots Fight Opposition in Russia

Online Appendix

Denis Stukal, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker

Appendix A Twitter Data Collection

We collected approximately 32 million tweets about Russian politics via the Twitter Streaming API using the list of 86 keywords and hashtags shown below. This list uses transliteration. A Cyrillic equivalent was used for all but italicized keywords.

Keywords: 37godvernulsya, 6maya, *6may*, biryulevo, bolotnaya, bolotnoyedelo, chestniyevybory, DMP, dukhovniyeskrepy, edro, *#Erdogan*, *#ExpelTurkeyFromNATO*, *#FreeSavchenko*, golodovka, gorozhaneprotiv, khvatitkormitkavkaz, khvatitvinitkavkaz, kirovles, komandanavalnogo, krovaviyrezhim, *#Latakia*, maidan, maidaner, maidanutiy, marshmillionov, medvedev, MinutaNeMolchaniya, #MinutaNeMolchaniya, narodniyskhod, navalniy, Nemtsov, *#Nemtsov*, #Nemtsov, Odessa, odinzavsekh, #olimpiada, #olimpiyskayazachistka, oppozicia, partiyazhulikovivorov, priamayaliniya, privet37god, PussyRiot, *PussyRiot*, putin, putinakh, putinvor, *#Putinkiller*, *#putinsgames*, #PutinUmer, pzhiv, rasserzhennye, rosuznik, *#RussianJet*, *#Russianplane*, russkiymarsh, #samolet, savchenko, #schitaemvmeste, sdnempobedi, sobyaninnashmer, Sochi2014, #sochi, *#sochi*, *#sochi2014*, *#sochi2014problems*, *#sochifail*, *#sochiproblems*, spasiboputinuzaeto, Strategiya31, Su-24, Su24, *Su24*, surkovskayapropaganda, suvkirove, svobodupolitzaklyuchennym, svoboduuznikam6maya, tolokno, tolokonnikova, triumfalnaya, udaltsov, #vitishko, vsezaodnogo, vysurkovskayapropaganda, zachestniyevibory, zanavalnogo, zhalkiy.

We composed this list of keywords to capture a large variety of tweets related to different aspects of Russian politics, including pro-government and anti-government online activity. Pro-government tweets are often produced by state-controlled mass media and can be captured using generic keywords including last names of Russian officials (Putin, Medvedev), opposition politicians (Navalny, Udaltsov), national holidays that loom large in the official narrative (sdnempobedi), pro-government campaigns (sobyaninnashmer), and generic terms like opposition, maidan, etc. Anti-government tweets exhibit a greater variety of sentiments, therefore we created a more diverse set of keywords, including the most popular references to state repressions (37godvernulsya, krovaviyrezhim) and propaganda (surkovskayapropaganda, vysurkovskayapropaganda); names of key opposition events and strategies (narodniyskhod, Strategiya31); and slogans that express negative sentiment towards the Russian government officials (putinvor, #Putinkiller).

Appendix B What can bots tweet about?

Although our data and methods do not allow us to identify the specific political actors that created each political bot, we are able to measure bots' political orientation (see details in the Data and Methods Section of the paper) and separate Twitter bots that post pro-Kremlin messages from other automated fake accounts. The pro-government Twitter bots that we are able to detect can engage in a number of different online activities. Tables 1 and 2 use bot-generated tweets from our collection to give examples of the types of posts made by Russian pro-government bots.

Table 1: Bots' tweets: Negative campaigning against opposition.

Date	Original tweet	English Translation
September 20, 2015	RT USER: Навальный все канает под бедного, а сам на джипе рассекает - это при двух-то судимостях! LINK	RT USER: Navalny keeps pretending to be poor, but drives an SUV – even though he has two previous convictions! LINK
September 20, 2015	RT USER: Для честности, Навальный должен был назвать митинг не "за сменяемость власти", а "за сменяемость власти на нас, русофобов."	RT USER: Were Navalny honest, instead of calling the rally "For the regular transfer of power", he would call it "For the transfer of power to us, Russophobs."
September 20, 2015	RT USER: Навальный опять рассказывает про жуликов воров и бандитов... Лично я уже устал от этой риторики, а вы? #Марьино LINK	RT USER: Navalny is talking crooks, thieves, and thugs again. Personally, I am sick and tired of this kind of rhetoric, aren't you? #Mar'yino LINK

Note: Retweeted personal accounts are masked with USER in compliance with Twitter's terms of service. URL links embedded in tweets are masked. On September 20, 2015, opposition rallies for for fair elections and regular transfer of power ("against no turnover in the government") took place in Moscow in the Maryino district. Alexey Navalny is a prominent Russian opposition leader who has been active in organizing mass political protests.

First, as Table 1 illustrates, pro-government bots can *broadcast anti-opposition and anti-*

protest sentiment in an effort to impact public opinion. The first tweet from Table 1 uses an unfavorable comparison (no turnover in opposition vs. no turnover in government) to discredit the protest cause. The second one capitalizes on the anti-American Russian nationalist sentiment that perceives the collapse of the Soviet Union as a national tragedy (Zevelev 2008) to portray opposition members as traitors. Finally, the third example expresses disappointment with the rhetoric of the Russian opposition in order to disengage people from the cause of a protest rally. This example also illustrates how bots can hijack protest-related hashtags. In particular, the protest-related hashtag #Maryino (the protest rally location in Moscow) is used here to increase the visibility of an anti-protest tweet.

Table 2: Bots' tweets: Cheerleading for the regime.

Date	Original tweet	English Translation
May 5, 2018	RT USER: 🇷🇺 Владимир Владимирович Путин Лучший Президент Великой России и Мира! Народ России сделал свой выбор и всё будет хорошо	RT USER: 🇷🇺 Vladimir Vladimirovich Putin is the Best President of our Great Country and the World! The people of Russia have made their choice, and everything will be great
September 20, 2015	RT USER: А вы поддерживаете Путина?! Да – #Ретвит Нет – избр. #опрос #Путин #политика LINK	RT USER: Does Putin have your support?! If yes, #Retweet. If no – add to Favorites. #poll #Putin #politics LINK
May 5, 2018	RT USER: Лайк и ретвит, если #Путин – наш президент! LINK	RT USER: Like and retweet if #Putin is our president! LINK

Note: Retweeted personal accounts are masked with USER in compliance with Twitter's terms of service. URL links embedded in tweets are masked. On September 20, 2015, opposition rallies for fair elections and regular transfer of power took place in Moscow. On May 5, 2018, a Russian opposition leader Alexey Navalny organized anti-Putin rallies across the country under the slogan "Putin is not our tsar."

Another strategy pro-government bots might pursue involves *cheerleading* and expressing pro-government sentiment, similar to what has been observed from Chinese government trolls

(King, Pan, and Roberts 2017). Table 2 shows examples of retweets that fall under this category. The first tweet in Table 2 is an example of a bot broadcasting excitement about Russian leaders. Although this tweet does not pretend to be an expression of a mass political sentiment, it does carry strong positive emotions that can pass to other Twitter users. The other two tweets pretend to be surveys that measure regime support as expressed via some online action (retweeting, liking, or adding to Favorites), and are intended to add pro-government content to the Twittersphere. The intention is especially clear, given that a "no" vote, as recorded by "adding to favourites," won't be visible to other users.

Appendix C How Do Bots Operate?

As people do not often follow bots on Twitter, there are limited mechanisms by which the latter can influence real human users. One of them is through *influencing search on Twitter*. A Twitter user who is interested in reading tweets about a particular topic can use the Twitter search panel to get a list of tweets that contain a search keyword and thereby be exposed to tweets posted by bots.

Another mechanism is via *using hashtags* that are employed to label the topic of a tweet and allow users to click on them to see other tweets with the same hashtag. As shown in Tables 1 and 2, bots also tweet and retweet using hashtags, which exposes Twitter users to their activity.

The third mechanism is via *coordinated campaigns targeting the list of trending topics* that Twitter shows its users. Although Twitter does not reveal the exact methodology behind the detection of trending topics, the platform states that this section is controlled by an algorithm identifying topics that are currently popular (FAQs 2019).

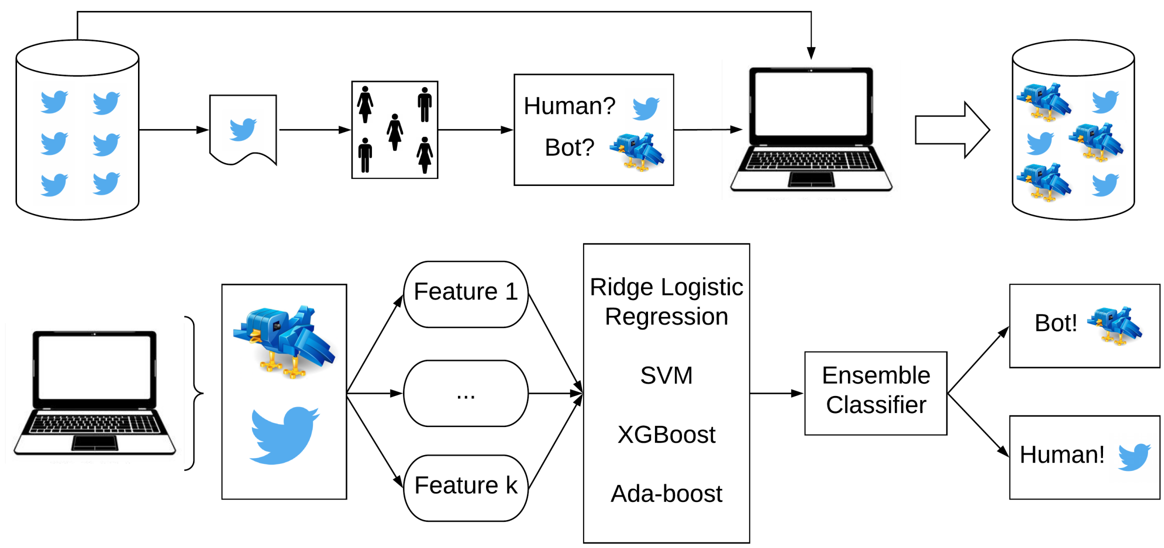
Finally, since the numbers of retweets and followers are often used by Twitter users as a heuristic to assess the importance of a particular message and the credibility or popularity of its author, bots could *strategically follow and retweet certain users* to affect how these are perceived by their human audiences (Varol and Uluturk 2020).

Given these four ways bots can be used, the involvement of bots in cheerleading for the government can arguably suggest that pro-government bots might be targeting either regime supporters or the general audience with weak political preferences, as it is unlikely that regime opponents could change their preferences or leanings based on happy tweets about Vladimir Putin. The negative campaigning strategy on the other hand might target both anti-regime activists and the broader audiences, especially in times when there are higher risks of the popular engagement in the activities of the opposition (e.g. during mass political protests).

Appendix D Detecting Russian Twitter bots

This section provides an overview of the bot detection tool developed in Stukal et al. (2017) and employed in this paper for identifying Russian Twitter bots. As we explain in the Data and Methods Section of the paper, this is a supervised classifier that is trained on human-annotated data of Russian Twitter accounts. Figure 1 summarizes the idea behind the classifier in a flowchart that can be described as a sequence of steps.

Figure 1: Bot detection workflow



First, tweets are aggregated on account level, so that Twitter accounts become the unit of analysis. A Twitter account is represented by a set of tweets with associated meta-data that includes the user name, bio description, the number of followers and friends, the total number of tweets posted, date of joining Twitter, etc.

On steps 2 through 4, a relatively small subset of these collected accounts is randomly selected for annotation by human coders who assign labels "bot" or "human" to each account. Due to the difficulty of separating humans from bots, larger groups of human coders are desirable. The bot detection tool of Stukal et al. (2017) uses labeled accounts annotated by at least five trained human coders. Individual coders' labels are aggregated in a conservative way: an account receives the final label and is added to the labeled set if at least 75% of coders agreed on the assigned category. The remaining accounts that did not pass this inter-coder reliability requirement are not included in the labeled set.

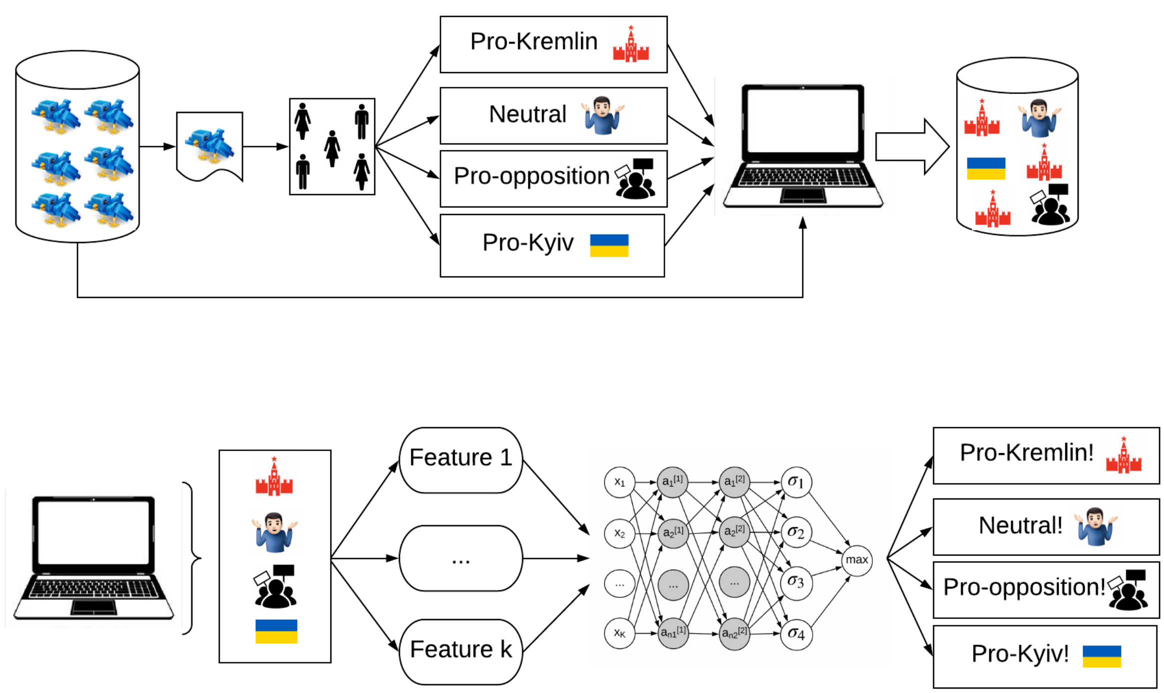
Step 5 involves supervised machine learning and is presented in more details on the lower panel of Figure 1. The machine learning algorithm's input includes the labels assigned to accounts and a set of features that describe the accounts and their tweeting activity. Then, the labeled set is split into training and test sets 10 times. Four supervised classifiers are trained on the training parts with the choice of hyper-parameters via cross-validation, and their predictions are aggregated using a voting ensemble classifier. For each of the training sets, the unanimous voting rule is applied to the predictions of the four classifiers (i.e. an account is predicted to be a bot if all four individual classifiers make this prediction). Then, these aggregated predictions are further aggregated across 10 training sets by applying the majority voting rule. Thereby, the final classifier is a majority-unanimous ensemble designed to maximise bot detection precision.

Finally, on step 6, the developed ensemble classifier is applied to the whole collection of accounts to scale up the human coding of bots.

Appendix E Measuring the political orientation of Russian bots

Since not every Russian-language Twitter bot is involved in spreading pro-Kremlin sentiment, Stukal et al. (2019) developed a tool to measure the expressed political orientation of bots. The overall architecture of the system (Figure 2) is similar to the one presented on Figure 1 in Appendix D, but the machine learning component relies on a feedforward neural network instead of an ensemble classifier.

Figure 2: Bot orientation measurement workflow



The labeled set includes accounts that were previously detected to be bots and were subject to human annotation for coding the orientation of the content they posted. Human coders could choose between four categories: Pro-Kremlin (accounts expressing support for the ruling party, Vladimir Putin, Dmitry Medvedev, their domestic and/or international policies), pro-opposition (criticizing a variety of domestic or international policies of the Russian government), pro-Kyiv (bots criticizing the Russian involvement in the crisis in Eastern Ukraine and the annexation of Crimea), and neutral (bots that either do not express any detectable political sentiment or expressing inconsistent political sentiment).

The machine learning part of this classifier takes the human-assigned labels at the predicted categorical variable and textual data (words and word pairs, mentions, hashtags, and domains of shared links) as features. These textual features are represented as binary ("dummy" or "one-hot") variables that are related to the outcome categorical variables through a sequence of non-linear transformations known as ReLU (rectified linear units that return the feature value for positive values and zero otherwise). The final layer of the neural network is the softmax function commonly referred to in the social sciences as the multinomial logistic regression. The output of the model is a set of predicted probabilities that an account belongs to each of the four possible categories.

Appendix F Robustness checks: Days before and after offline protests and online spikes

The offline demobilization framework for theorizing the strategic deployment of Twitter bots predicts that pro-government bots will be activated during offline protest rallies. However, in many cases, protest rallies can be anticipated, as they are often announced in advance[1] and promoted on social media, which informs the public ahead of time about when and where a rally will take place. Protest anticipation enables pro-government actors to deploy bots before the actual protest takes place. For this reason, under the offline demobilization hypothesis, we anticipate to see increases in bot activities not only on days of protests, but also on days preceding these protests (see the Theory and Hypotheses Section of the paper).

Anticipating spikes in the online opposition activity seems a harder task, as they may be driven by unanticipated events. However, in some cases, those spikes may originate from certain observable developments on the ground and can be foreseen.

We explore the potential anticipation of both the offline and online events in our main results in the paper that also include results for days following both types of events. The rationale for looking at days after spikes or protests is to better understand whether these events have long-lasting effects or, instead, those effects have high rates of decay.

Here, we extend this time frame and explore the dynamics of effect sizes for up to 7 days before and after the offline and online events.

As in the main text of the article, the reported lags and leads are cumulative. As one can see from Figures 3 – 6, the uncovered effects decrease mostly monotonically in the size of the lag, suggesting that the largest changes in the behavior of bots on all these dimensions happen on protest days or days with online spikes.

Figures 7 – 10 reveal a very similar pattern. In the case of relatively large effects (the Volume and Retweet diversity dimensions), they decline rapidly. The decay is less pronounced for the Cheerleading and Negative campaigning dimension, but is nevertheless clear from

[1] Russian law requires obtaining a permit from local authorities to hold a rally, which often involves lengthy negotiations about the timing, location, and the allowed number of participants.

Figure 3: Volume of tweets: days before events

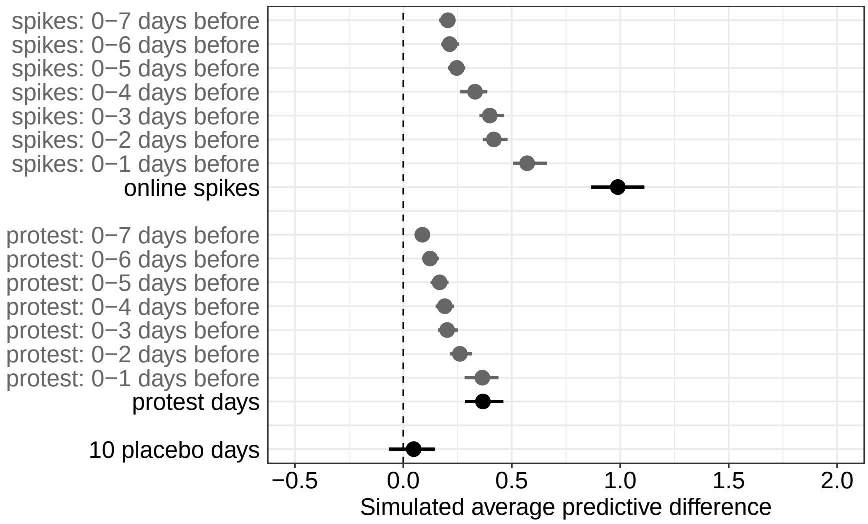
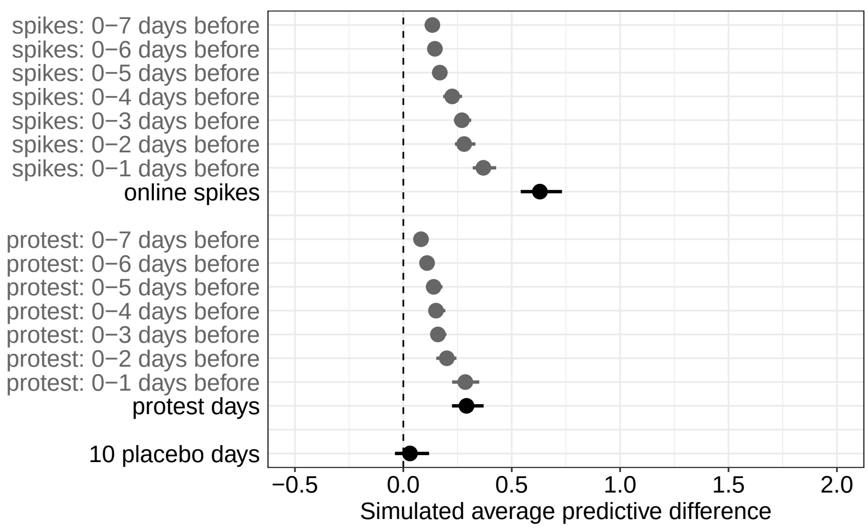


Figure 4: Retweet diversity: days before events



Figures 9 and 10.

Figure 5: Cheerleading: days before events

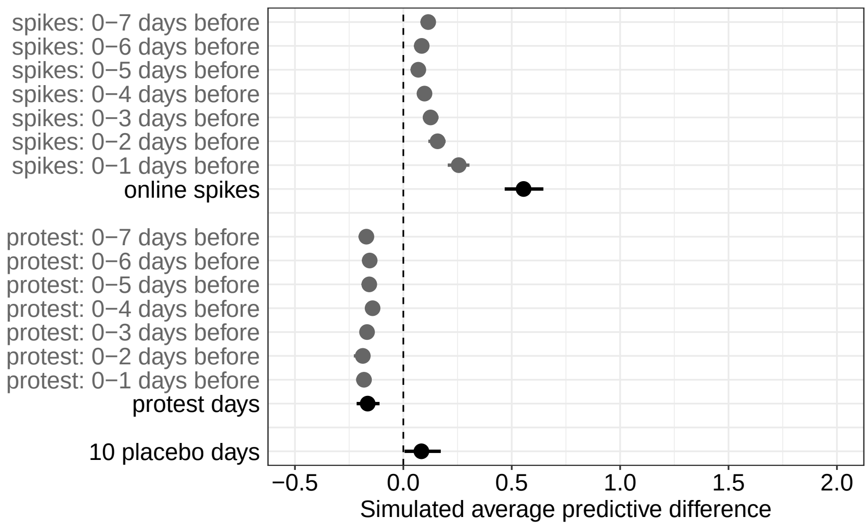


Figure 6: Negative campaigning: days before events

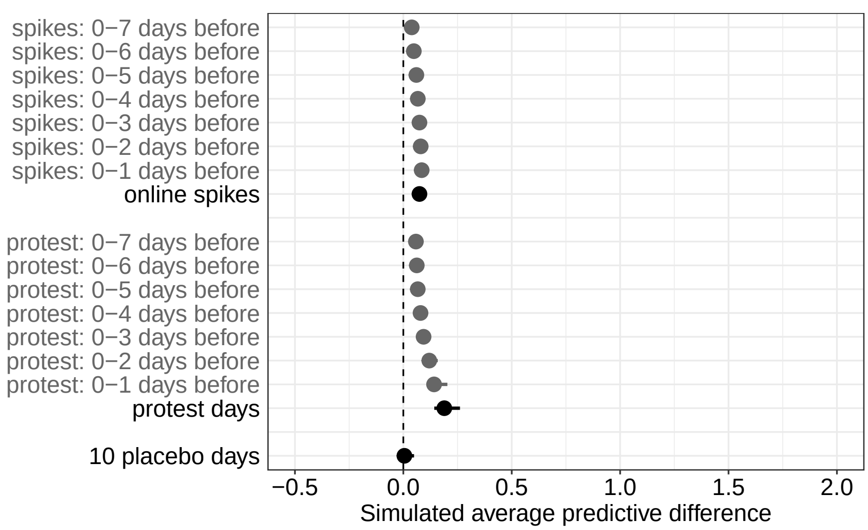


Figure 7: Volume of tweets: days after events

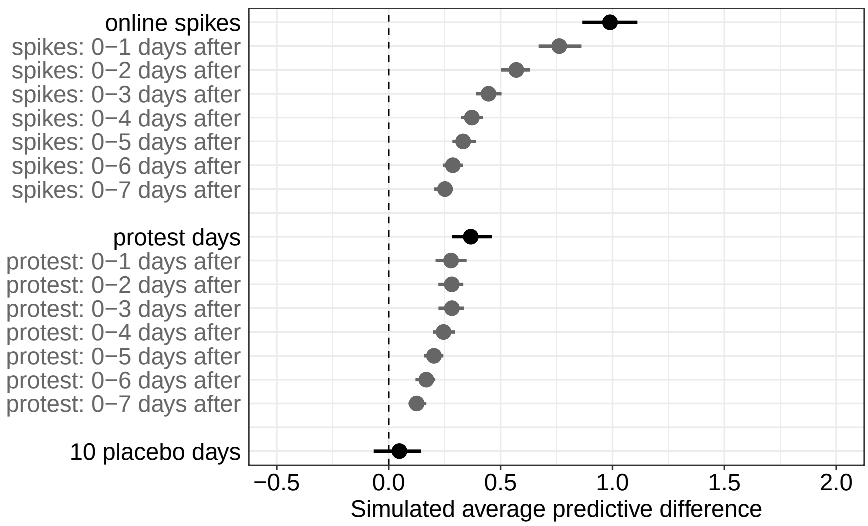


Figure 8: Retweet diversity: days after events

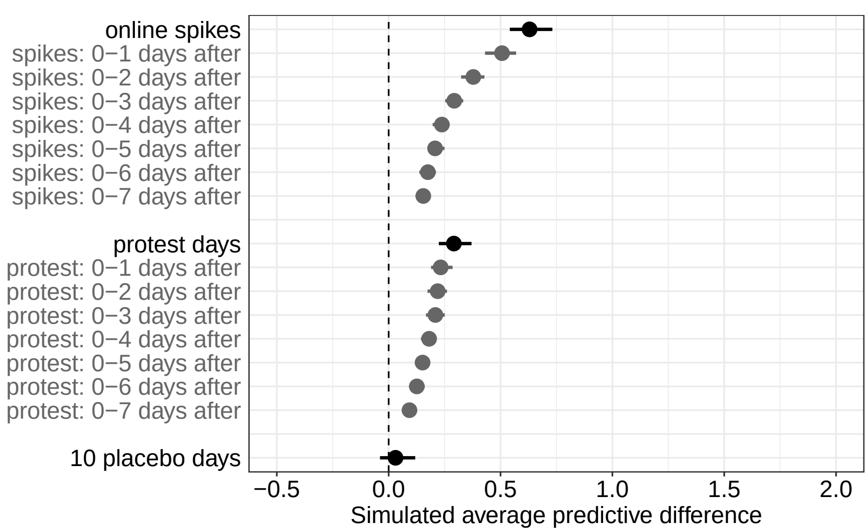


Figure 9: Cheerleading: days after events

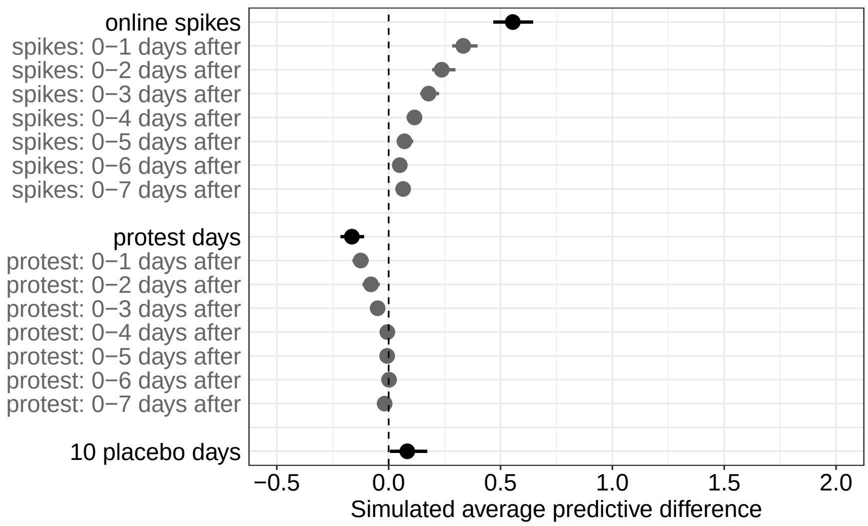
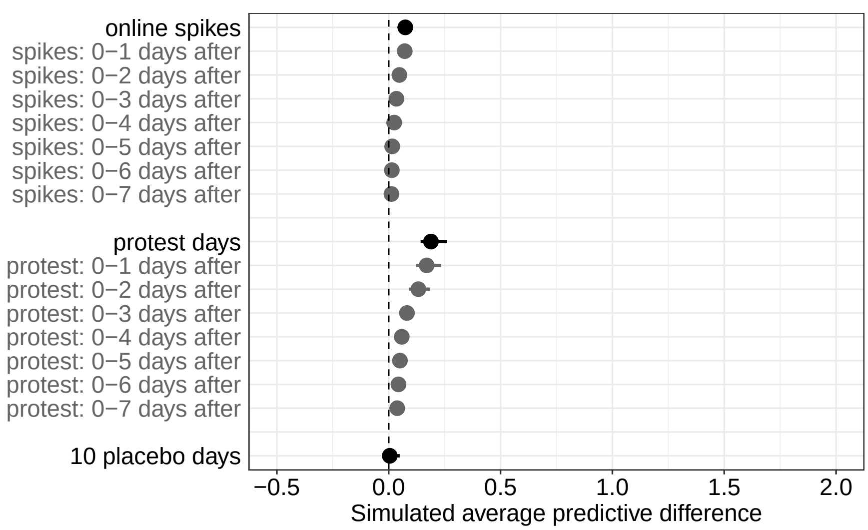


Figure 10: Negative campaigning: days after events

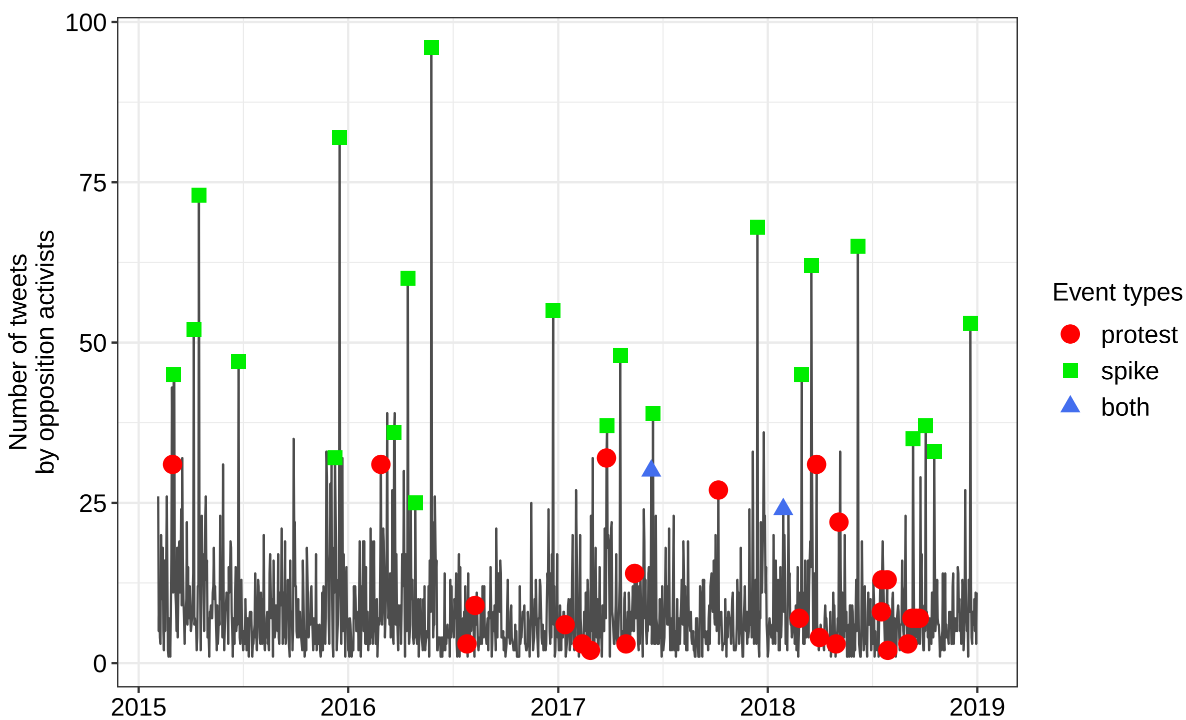


Appendix G Identifiability assumption

The identifiability assumption requires that the effect of protest rallies be separable from both the effects of online spikes and other events on the ground. This assumption may be violated if protests systematically coincide with other events that could require pro-government bots to be activated.

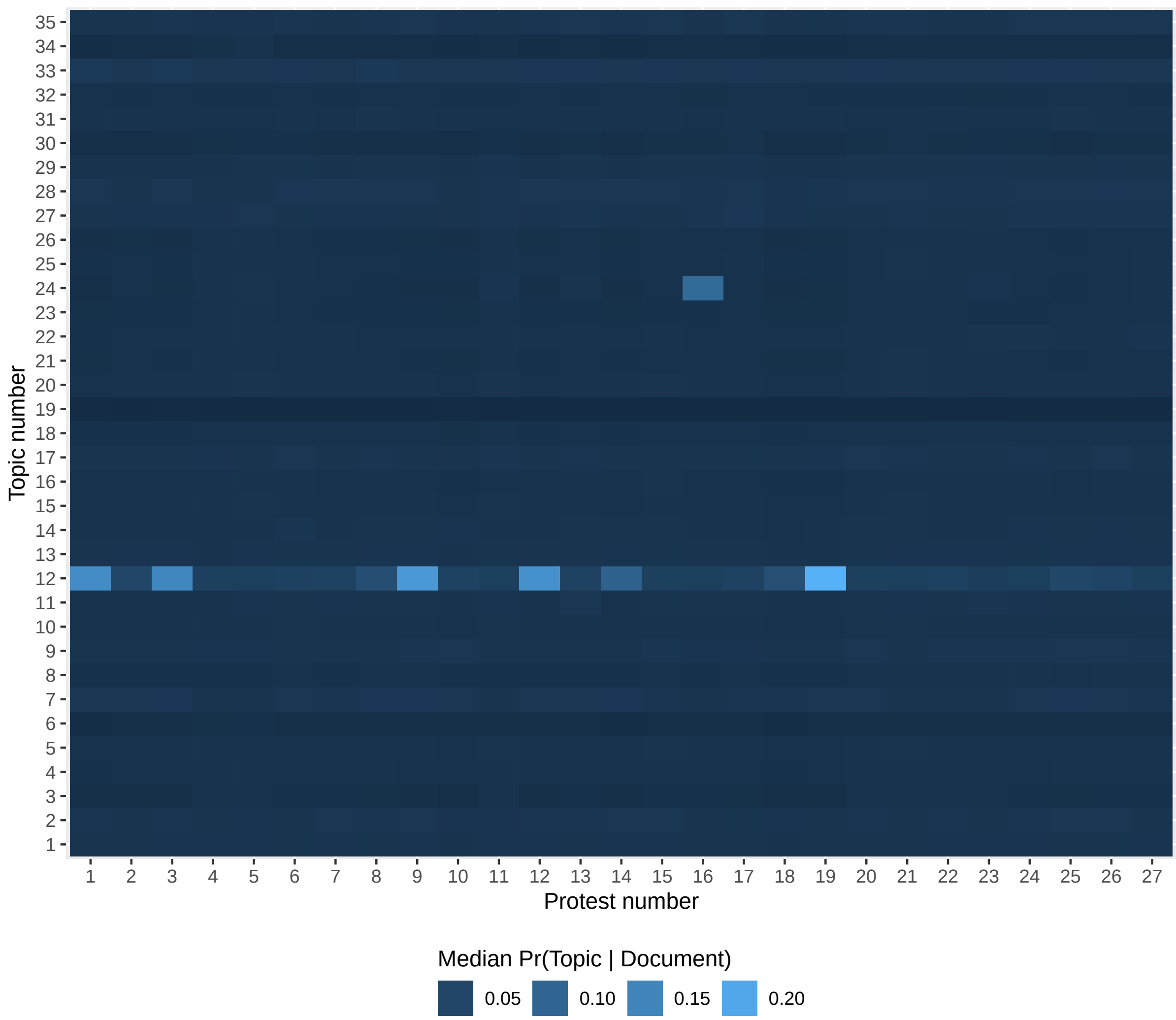
Here, we provide empirical evidence in support of our assumption. First, Figure 11 presents the dynamics of the total number of tweets posted by the Russian opposition with squares marking the days of online spikes as measured in this article and dots marking protest days. As one can see, these two types of events do not systematically coincide, which implies that their effects are separable.

Figure 11: Separability of protests and spikes in online opposition activity



In order to evaluate the separability of protest events from other events on the ground, we apply a Structural Topic Model (Roberts et al. 2014), a popular method for summarizing large collections of texts via identifying their topical structure. We fit a model with 35 topics to the tweets posted by pro-government bots and extract tweet-level topic distributions. Then, we measure topic prevalence on protest days by computing the median of topic probabilities across all tweets posted by pro-government bots on a given day for each topic separately.

Figure 12: Topic distributions across protest rallies



If there are any confounders, i.e. any events that systematically co-occur with protest rallies, we expect to see one of the two patterns. One possibility is that most protest days will score high on topics that are unrelated to protest rallies. The other possibility is that protest-related topics will not stand out as highly probable on protest days.

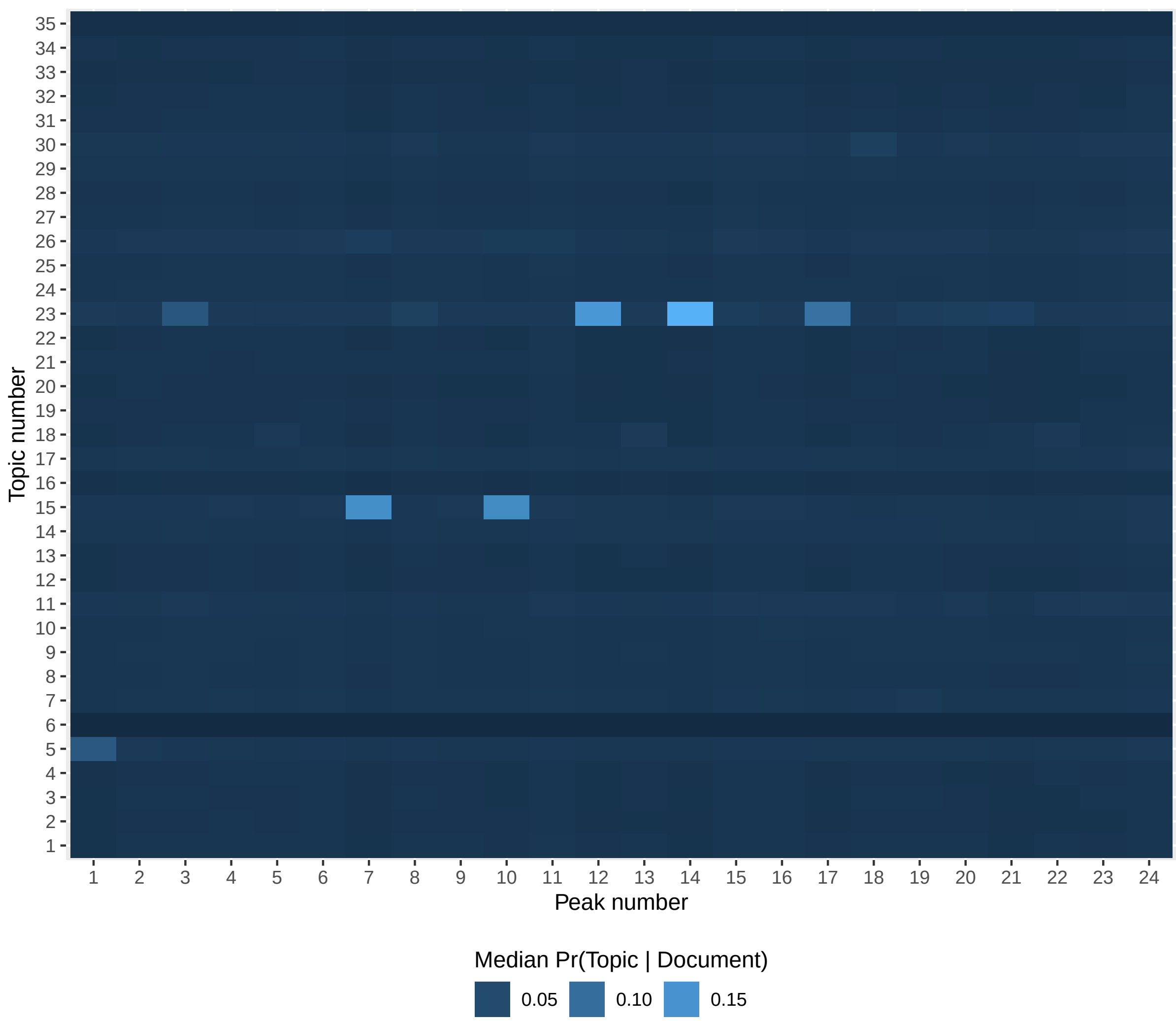
Figure 12 shows the median topic probabilities across all protest days (for the exact dates of each protest and additional protest-related information see Table 3 in Appendix I). Each colored rectangle on the graph corresponds to a topic (vertical axis) on a given protest day (horizontal axis). The darker the color, the smaller the median probability of a topic over all tweets pro-government bots posted on that day. As one can see from Figure 12, only one topic (Topic 12) stands out systematically as highly probable on most protest days.

To understand what this topic is about, we employ two methods. First, we use the FREX statistic (Bischof and Airoldi 2012) to extract the most relevant words that are both frequent and exclusive for each topic. Second, we extract the most representative tweets for a each topic. Both of these approaches reveal that Topic 12 is associated with protests. Indeed, its seven key words are "rally," "Alexey," "thousand," "Boris," "attention," "iremeslo," and "Nemtsov." The most representative tweet reads: "RT Navalny and his supporters managed to turn into a political rally even a memorial event. Such disgusting bastards @Current$_{p}olicy.''$

In addition to analyzing topic distributions across tweets posted on protest days, we perform a similar analysis for the days of increased opposition activity. We estimate another structural topic model with 35 topics using tweets posted on the days with online spikes. The results (shown in Figure 13) reveal that two topics (Topic 23 and Topic 15) stand out as prominent on days with online spikes. In order to interpret them, we followed the same procedure and employed the FREX statistic and the most representative tweet. The seven key words for Topic 23 include "direct," "Alexey," "rally," "line," "iremeslo," "Navalny," and "imerkouri," whereas key words for Topic 15 are "Nadezhda," "rada," "lawyer," "supreme," "deputy," "Savchenko," "position." The most representative tweets read: "RT Alexey Navalny has gathered another multi-thousand rally in Novosibirsk. They talk about the upcoming Kasyanov's rise to power" and "RT Savchenko called Poroshenko a slacker. The scandalously famous Ukrainian pilot and member of parliament Nadezhda Savchenko."

As no other topic stands out as dominant, this can serve as additional evidence in support of our claim that pro-government bots respond to anti-government activities and not to some specific political topic or policy. Overall, the textual analysis of the tweets does not provide evidence against the identifiability assumption.

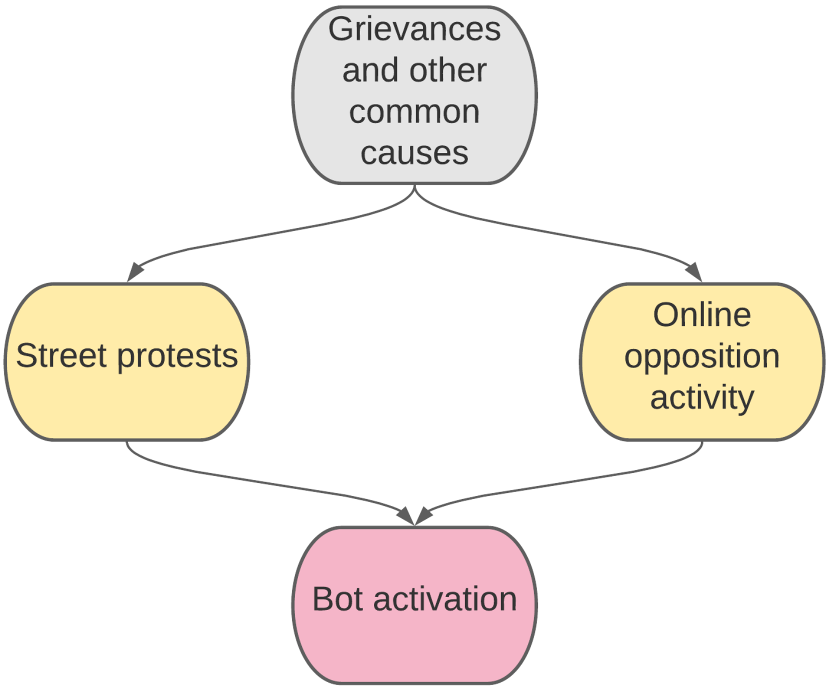
Figure 13: Topic distributions across online spikes



Appendix H Causal graph

This section presents the causal graph for bot activation that presents our reasoning in a clear graphical format.[2]

Figure 14: Causal graph for bot activation



This study focuses on two major ways for popular grievances to manifest in a non-democratic setting. Street protests are one way and online opposition activity is the other one.

Pro-government bots can get activated in response to street protests or spikes in the online opposition activity, or as a reaction of pro-government actors to the anticipation of the upcoming online or offline mobilization.

As street protests and online opposition activity could potentially be driven by common underlying causes, our regression models described in the paper include both protest and online opposition activity variables.

[2] We thank an anonymous reviewer for the idea of this graph.

Appendix I Protest rallies in Russia

Here we report the exact days of mass protest rallies in Russia in 2015–2018 that we used to test the offline hypothesis. We focused on large protest rallies that gathered at least 1000 people. Our choice was driven by two key considerations. First, these protests were popular enough to provoke a response from pro-government actors. Second, as our primary source of data on protest rallies in Russia is the ICEWS project that automatically extracts data from mass media, we could only collect data on rallies that were large enough to attract media attention.

Table 3: Offline protest rallies

No.	Date	Location
1	2015-03-01	Moscow
2	2015-09-20	Moscow
3	2016-02-27	Moscow
4	2016-07-26	national
5	2016-08-09	Moscow
6	2017-01-13	St. Petersburg
7	2017-02-12	St. Petersburg
8	2017-02-26	Moscow
9	2017-03-26	national
10	2017-04-29	national
11	2017-05-14	Moscow
12	2017-06-12	national
13	2017-10-07	national
14	2018-01-28	national
15	2018-02-25	Moscow
16	2018-03-27	Kemerovo
17	2018-04-01	Moscow region
18	2018-04-30	Moscow
19	2018-05-05	national
20	2018-07-18	Moscow
21	2018-07-19	Moscow
22	2018-07-28	national
23	2018-07-29	national
24	2018-09-02	Moscow
25	2018-09-09	Far East
26	2018-09-16	St. Petersburg
27	2018-09-22	national

Appendix J Augmented Protest Dataset

One of the potential explanations for the moderate evidence that we find in support of the offline demobilization hypothesis is that we missed some of the protests pro-government elite groups were especially afraid of. In order to suppress these protests, the elite groups could have simultaneously deployed social media bots and used a variety of unofficial channels and mechanisms that exist in Russia to prevent mass media from covering political events organized by the opposition (Vardanyan 2017). In this case, we could miss these protest events and underestimate the effects of street protests on bot activity.

In order to address this concern, we augment our data with extra protest events that were carefully selected into the Mass Mobilization in Democracies dataset (Weidmann and Rød 2019) and also a new dataset on protests in Russia by Lankina and Tertytchnaya (2019). The augmented set of protests is presented in Table 4.

We use this augmented dataset to re-run the analysis presented in the main text of the article and present the findings in Figures 15 – 18 by juxtaposing the original simulated average predictive differences and the new ones. As these graphs show, the results remain substantively unchanged on all dimensions.

Table 4: Augmented data on offline protest rallies

No.	Date	Location	Source
1	2015-02-01	Tomsk	MMAD
2	2015-03-01	Moscow	ICEWS
3	2015-03-03	Moscow	MMAD
4	2015-03-29	Novosibirsk	MMAD
5	2015-04-05	Novosibirsk	MMAD
6	2015-04-14	Irkutsk	LT
7	2015-04-19	Cherepovets	LT
8	2015-05-01	Saint Petersburg	MMAD
9	2015-06-06	Moscow	MMAD
10	2015-07-29	Aver'yanovka	MMAD
11	2015-08-30	Saint Petersburg	MMAD
12	2015-09-20	Moscow	ICEWS
13	2015-09-21	Moscow	LT
14	2015-09-29	Yekaterinburg	LT
15	2015-11-07	Khabarovsk	MMAD
16	2015-11-20	Kayakent	MMAD
17	2016-02-27	Moscow	ICEWS
18	2016-03-27	Yekaterinburg	LT
19	2016-07-26	national	ICEWS
20	2016-08-09	Moscow	ICEWS
21	2017-01-13	St. Petersburg	ICEWS
22	2017-02-12	St. Petersburg	ICEWS
23	2017-02-26	Moscow	ICEWS
24	2017-03-26	national	ICEWS
25	2017-04-29	national	ICEWS
26	2017-05-14	Moscow	ICEWS
27	2017-06-12	national	ICEWS
28	2017-10-07	national	ICEWS
29	2018-01-28	national	ICEWS
30	2018-02-25	Moscow	ICEWS
31	2018-03-27	Kemerovo	ICEWS
32	2018-04-01	Moscow region	ICEWS
33	2018-04-30	Moscow	ICEWS
34	2018-05-05	national	ICEWS
35	2018-07-18	Moscow	ICEWS
36	2018-07-19	Moscow	ICEWS
37	2018-07-28	national	ICEWS
38	2018-07-29	national	ICEWS
39	2018-09-02	Moscow	ICEWS
40	2018-09-09	Far East	ICEWS
41	2018-09-16	St. Petersburg	ICEWS
42	2018-09-22	national	ICEWS

Note: LT stands for protests missing from ICEWS, but included in Lankina and Tertytchnaya (2019) dataset. ICEWS stands for protest rallies either mentioned in both datasets or missing from Lankina and Tertytchnaya (2019).

Figure 15: Volume of tweets: Augmented dataset

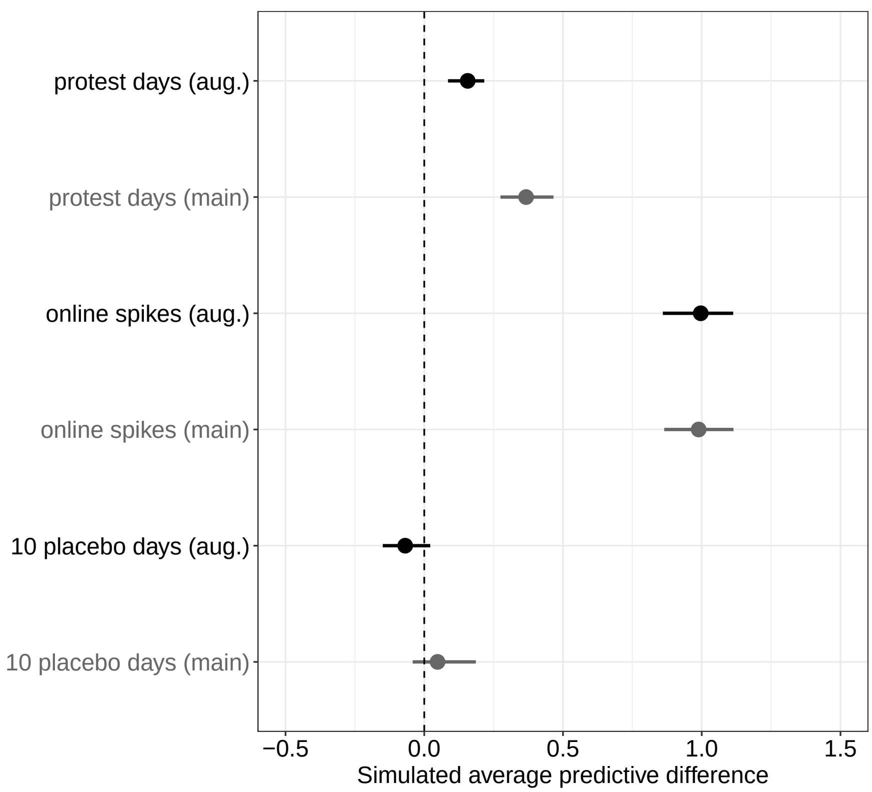


Figure 16: Retweet diversity: Augmented dataset

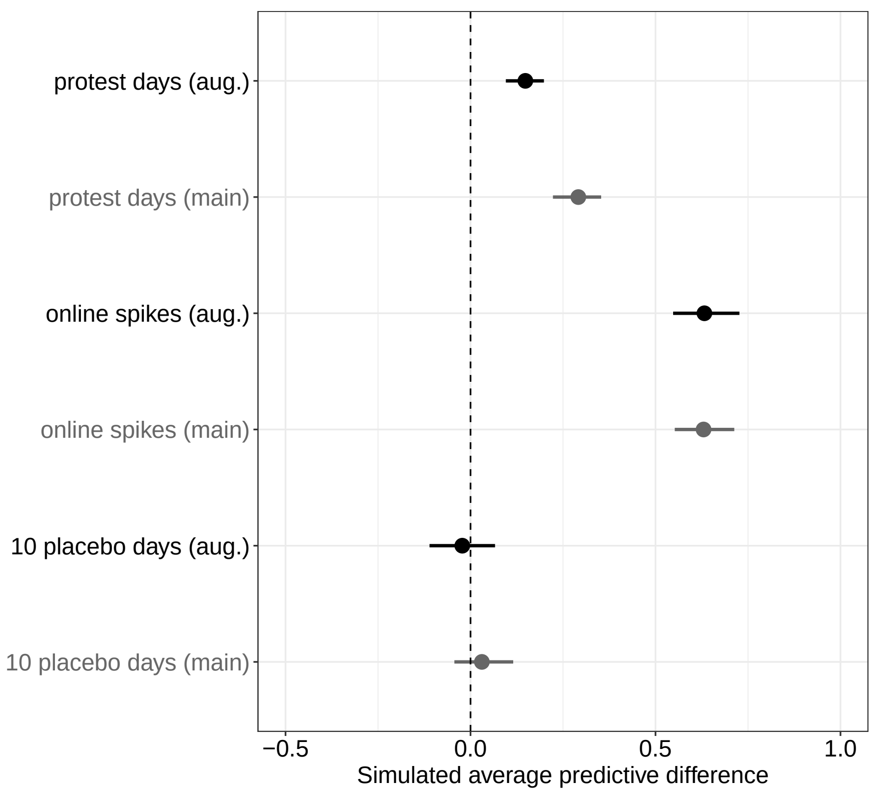


Figure 17: Cheerleading: Augmented dataset

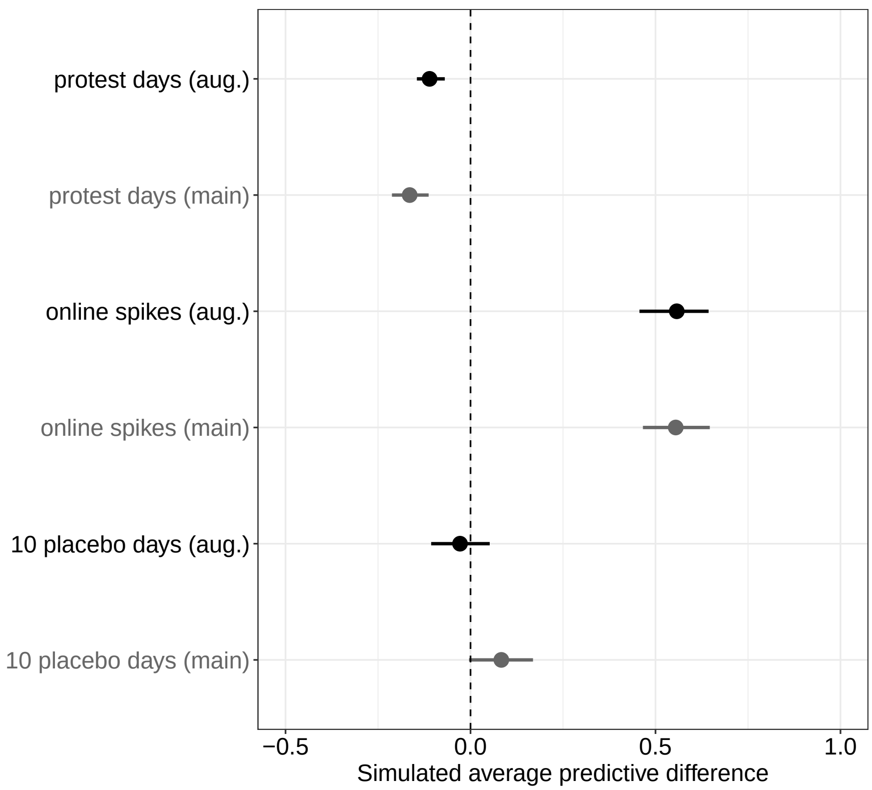
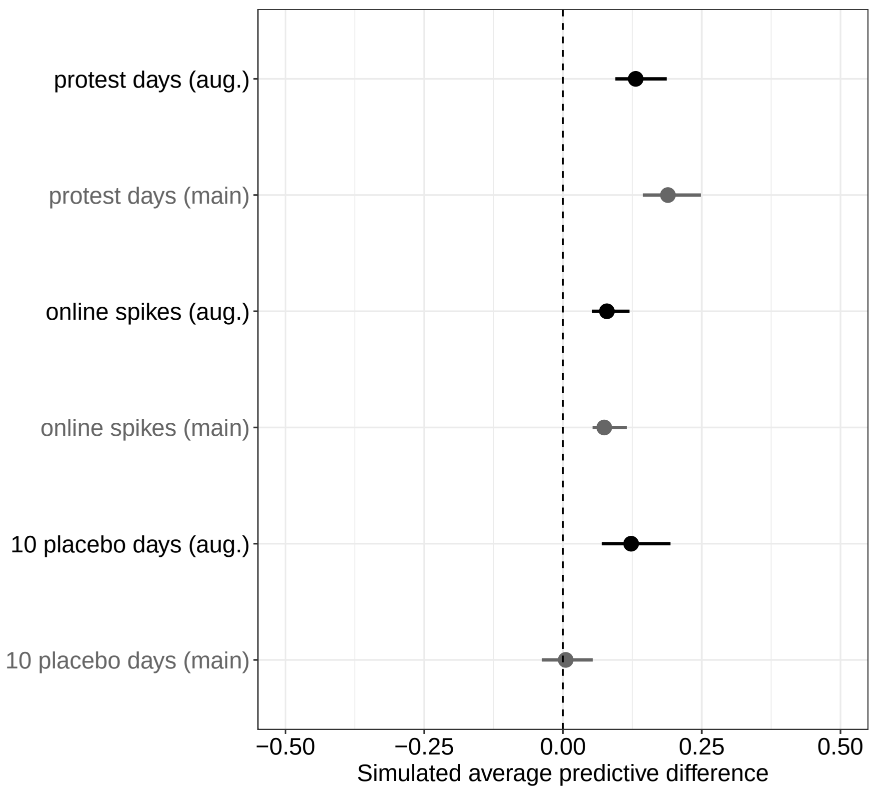


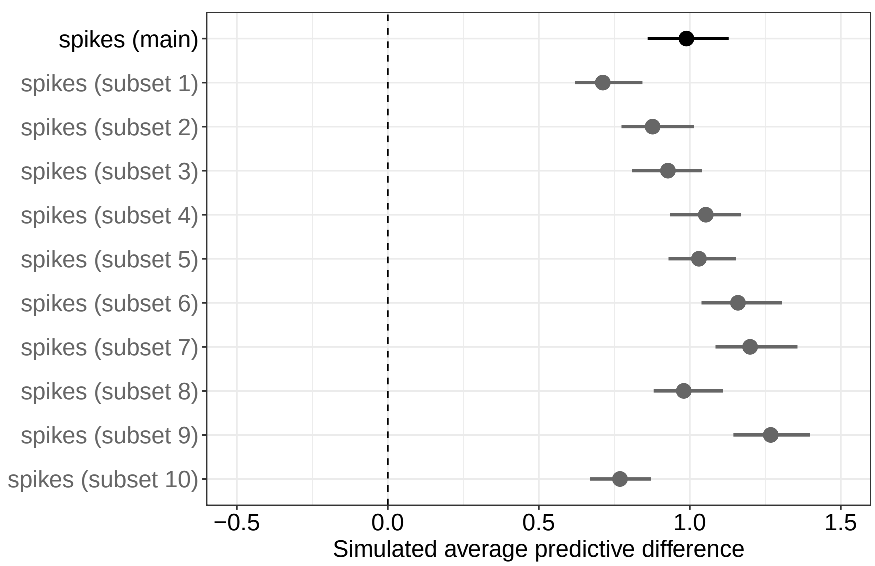
Figure 18: Negative campaigning: Augmented dataset



Appendix K Robustness check: subsets of activists

We test the online agenda control framework using data on spikes in the tweeting activity of 15 opposition leaders, activists, and media accounts that we selected based on our area expertise. In order to address potential concerns about the robustness of our results to changes in the composition of our opposition activists, we also replicate our results using 10 subsets of 10 activists randomly drawn from our pool of 15 pro-opposition Twitter accounts.[3] Figures 19 – 22 show that our main results remain robust to changes in the group of opposition activists analyzed.

Figure 19: Volume: Subsets of activists



[3]We are grateful to an anonymous reviewer for suggesting this robustness check.

Figure 20: Retweet diversity: Subsets of activists

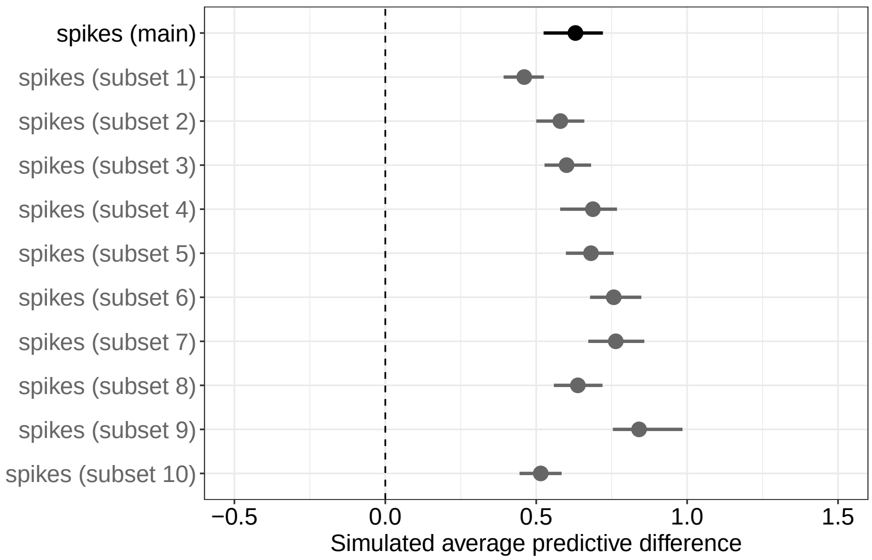


Figure 21: Cheerleading: Subsets of activists

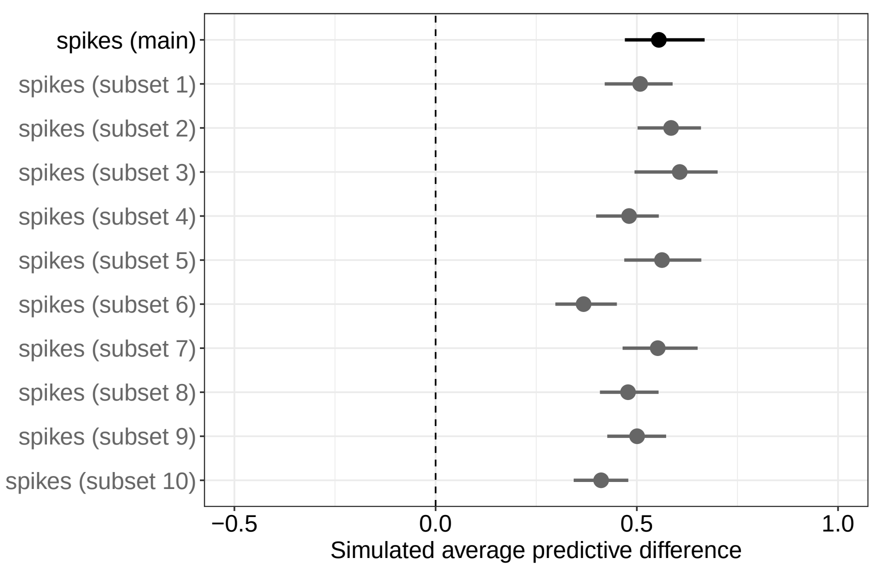
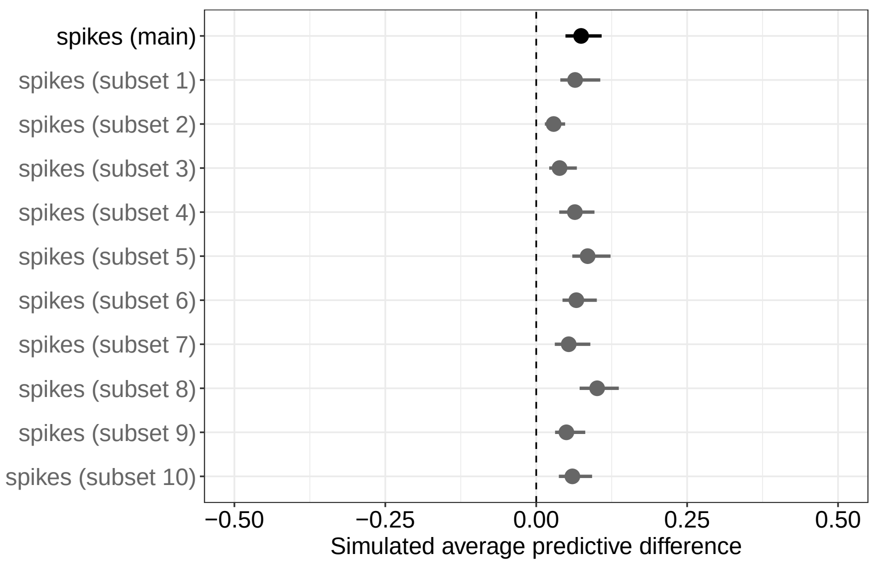


Figure 22: Negative campaigning: Subsets of activists



Appendix L Sentiment Analysis

This article measures the sentiment of bots at the account level (see Appendix E for details). An alternative approach would be to measure the sentiment of tweets. For example, we can use the popular dictionary-based approach that infers tweet-level sentiment based on the total number of positive and negative words in the tweet. If a tweet has more positive words than negative, the tweet is regarded as positive; if negative words outnumber positive words, we regard it as negative. We employ the state-of-the-art Russian sentiment dictionary RuSentiLex (Loukachevitch and Rusnachenko 2018) for counting positive and negative words and then count the number of positive and negative tweets mentioning Vladimir Putin or Alexey Navalny as our dependent variables. The results are shown in Figures 23 – 24.

In addition to the expected sentiment (positive for tweets mentioning Vladimir Putin and negative for tweets about Alexey Navalny), we also report the results for the opposite sentiment as an additional robustness check.

As one can see from Figures 23 – 24, the observed effects tend to be significantly attenuated for either type of sentiment, although, in the case of cheerleading, adding sentiment detection makes it possible to recover the expected effect of protest days.

The key limitation of this approach to testing hypotheses is related to the pitfalls of sentiment analysis. Indeed, the presence of negative words in a tweet does not necessarily mean that the expressed or implied sentiment towards the President is also negative (similarly, for the opposition leader). That is why, for the purposes of this study, we are more confident that the highly conservative and validated approach we describe in Appendix E yielded as pro-government only those accounts that mention the autocrat and the opposition leader with the expected sentiment. Automatic sentiment analysis of tweets serves thereby as an extra source of random noise that attenuates the estimated effects. Further research is required for improving existing sentiment analysis techniques for the Russian language to make them more reliable.

Figure 23: Cheerleading: Sentiment Analysis

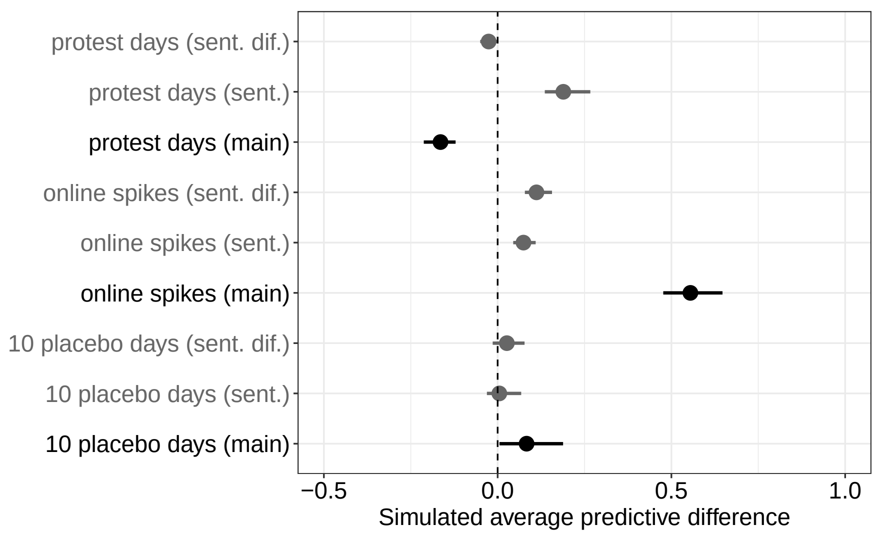
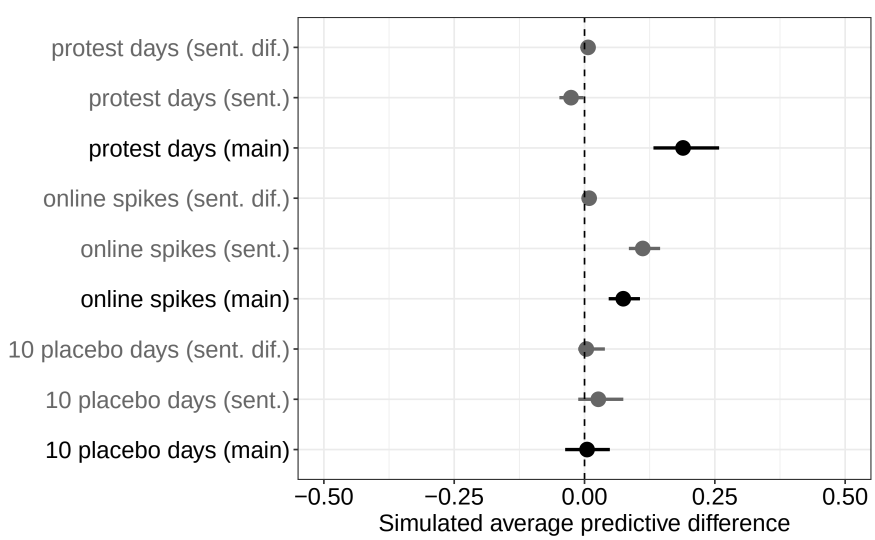


Figure 24: Negative campaigning: Sentiment Analysis



Appendix M Retweet Diversity

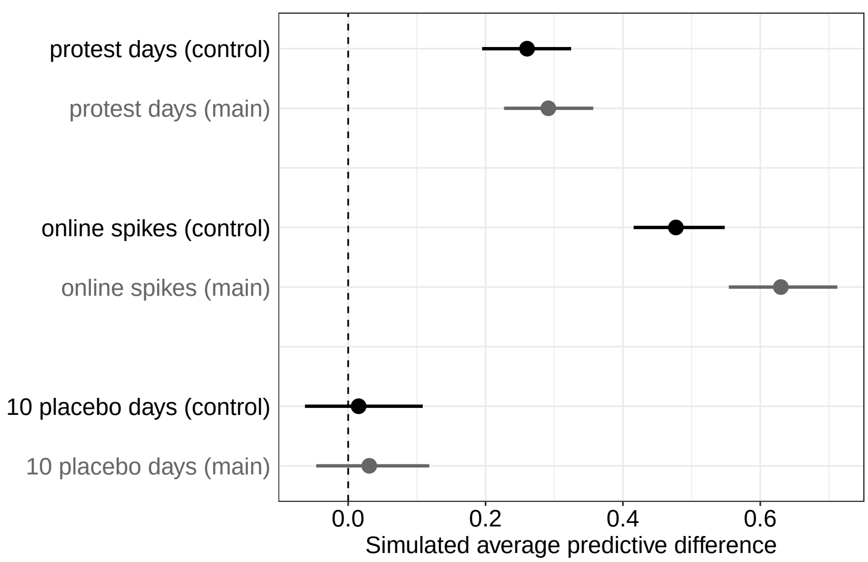
A potential explanation for the relatively large observed effects on the retweet diversity dimension is high correlation with the total number of bot-generated tweets.[4] In order to explore, to what extent this is true, we estimate the following mixed-effects model:

$$\mathbb{E}(Y_{it}|D_t, \gamma_{m[t]}) = \exp(\beta_{0i} + \beta_{1i} \times \text{Protest}_t + \beta_{2i} \times \text{Online}_t + \beta_{3i} \times \text{Placebo}_t$$
$$+ \beta_{4i} \times \text{TotalTweets}_{it} + \gamma_{m[t]}),$$

where Y_{it} is the retweet diversity variable for pro-government bot i on day t; Protest$_t$, Online$_t$, and Placebo$_t$ are binary independent variables that equal 1 if there is a street protest, spike in the online opposition activity on day t, or day t is a randomly selected placebo day, respectively; TotalTweets$_{it}$ is the total number of tweets pro-government bot i posted on day t; $\gamma_{m[t]}$ are month-year random effects.

The results of estimating this model are shown in Figure 25. As one can see, the retweet diversity effects remain substantively unchanged after adding the total number of tweets as a control variable.

Figure 25: Retweet Diversity with Total Volume



[4]We are grateful to an anonymous reviewer for bringing this possibility to our attention.

Appendix N Alternative Estimation Approaches

To date, there is an ongoing debate about the advantages and disadvantages of multilevel modeling and some more traditional approaches, including utilizing fixed effects. A recent study (Hazlett and Wainstein 2020) highlights the benefits of a unifying approach, a.k.a. correlated random effects (Wooldridge 2010), in a linear regression setting. There is, however, controversy about the appropriate model and estimation routine in the case of nonlinear models, especially in the case of unbalanced panels (Cameron and Trivedi 2005; Wooldridge 2018).

Even though we consider this debate unresolved, and, therefore, the jury is still out on whether we should use correlated random effects when analyzing our (non-linear) models, in order to check the robustness of the main results reported in the paper, we performed additional robustness checks using the pooled Poisson model with cluster-robust standard errors in a correlated random effects setting in Stata.[5] We found nothing in these preliminary analyses suggesting that our findings would not be robust to such a change in estimation or inference routines; results available from the authors upon request.

References

Bischof, Jonathan M., and Edoardo M. Airoldi. 2012. "Summarizing Topical Content with Word Frequency and Exclusivity." In *Proceedings of the International Conference on Machine Learning (Edinburgh, Scotland, UK).* https://icml.cc/2012/papers/113.pdf.

Cameron, Colin A., and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* Cambridge University Press.

FAQs, Twitter Trends. 2019. *How are trends determined?* https://help.twitter.com/en/using-twitter/twitter-trending-faqs.

[5]We are not aware of an R package for implementing this approach for non-linear models.

Hazlett, Chad, and Leonard Wainstein. 2020. "Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches." *Political Analysis,* 1–20. https://doi.org/10.1017/pan.2020.41.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument." *American Political Science Review* 111 (3): 484–501.

Lankina, Tomila V., and Katerina Tertytchnaya. 2019. "Protest in Electoral Autocracies: A New Dataset." *Post-Soviet Affairs* 36 (1): 20–36.

Loukachevitch, Natalia, and Nicolay Rusnachenko. 2018. "Extracting Sentiment Attitudes from Analytical Texts." In *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2018).*

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, and Joshua A. Tucker. 2017. "Detecting Bots on Russian Political Twitter." *Big Data* 5 (4): 310–324. http://online.liebertpub.com/doi/abs/10.1089/big.2017.0038.

———. 2019. "For Whom the Bot Tolls: A Neural Networks Approach to Measuring Political Orientation of Twitter Bots in Russia." *SAGE Open* 9 (2): 1–16.

Vardanyan, Denis. 2017. *Glavniy Organ Vlasti [Main governing authority].* https://newtimes.ru/articles/detail/123871.

Varol, Onur, and Ismail Uluturk. 2020. "Journalists on Twitter: Self-branding, Audiences, and Involvement of Bots." *Journal of Computational Social Science* 3 (1): 83–101. https://doi.org/10.1007/s42001-019-00056-6.

Weidmann, Nils B., and Espen Geelmuyden Rød. 2019. *The Internet and Political Protest in Autocracies.* Oxford University Press.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press.

———. 2018. "Correlated Random Effects Models with Unbalanced Panels." *Journal of Econometrics,* 137–150. https://doi.org/10.1017/pan.2020.41.

Zevelev, Igor. 2008. "Russia's Policy toward Compatriots in the Former Soviet Union." *Russia in Global Affairs* 6 (1): 33–45.