Summary on the Analysis of Russian Facebook Propaganda Efficacy

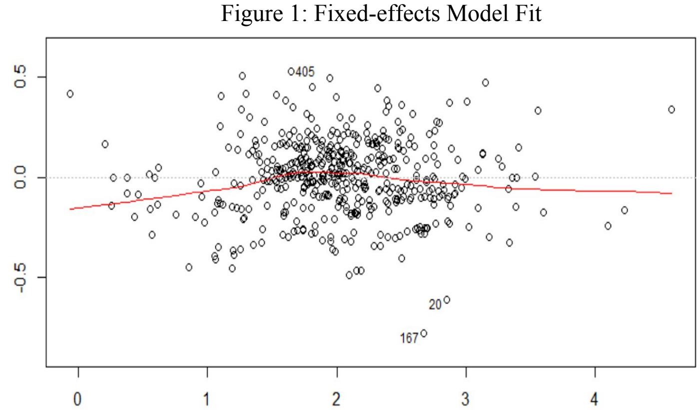
Written By: Donald Cooper, Jan Danel, Paul Franklin, and Tiger Hu

On February 16, 2018 Special Counsel Robert S. Mueller III indicted 13 Russian individuals and three Russian organizations for engaging in operations to interfere with U.S. political and electoral processes, including the 2016 presidential election. This was a significant step forward in exposing a surreptitious social media campaign and holding the accountable those responsible for this attack. The indictment spells out in exhaustive detail the breadth and systematic nature of this conspiracy, dating back to 2014, as well as the multiple ways in which Russian actors misused online platforms to carry out their clandestine operations. In conjunction with this investigation, authorities ordered Facebook to locate and disclose data relating to the Russian Internet Research Agency sponsored Ads placed on their site.[1] This data is the basis of our analysis.

The data we selected to use was the same data collected and used during the The House Intelligence Committee Minority Investigation by Special Counsel Robert S. Mueller who indicted 13 Russian individuals and three Russian organizations, including the IRA, for engaging in operations to interfere with U.S. political and electoral processes such as the 2016 presidential election. Facebook released "a total 3,519 total advertisements [that] were identified to have been purchased" with over "11.4 million American users exposed to those advertisements". The data has 25 different variables which include AdIDs, Adtext, Clicks, Impressions, Locations, CreationDate, and EndDate, all of which we use in our analysis. In order to interpret and use most of the data we convert the variables in character or string format to categorical data. While the data cleaning provided multiple variables as potential predictors, our Exploratory Data Analysis and intuition behind the origin of the data lead us to use only four variables for our first model in predicting the Click-Through-Rate (CTR) within each state: AdWordCount, AdDuration, Ad_Freq_state, ElectYear. When deciding between these variables and the others listed within the Method of Analysis section, much of our decisions were based on the fact that these four variables would directly affect both impressions and/or clicks by either a factor of higher count value (i.e variables AdWordCount, AdDuration, and Ad_Freq_state) or by a binary factor of whether the ad appeared during a certain year or not (ElectYear).

Using a fixed-effects multiple regression model[2], the analysis examines the effectiveness of russian political propaganda by accounting for the many differences between the states while simultaneously measuring the effectiveness of ads. This is used to account and control for the differences across states in order to make them as similar as possible. So, when testing for the efficacy of the ads, the result reveals how individual ads affected states differently or similarly, and if they had a greater or lesser impact in specific states. The model indicated that Russian ads had a significant effect on particular states likely due to the combination of higher CTR's[3], frequency in which ads appeared in these states, higher content, and both the release and duration of the content. In other words, the model showed that the average effect of our predictors on CTR percentage after controlling for state fixed-effects was greater than the baseline, being the state of Maryland simply for comparative purposes. The regression model has applications in showing how information through advertising is spread. More importantly, this algorithm provides detail on which ads are targeting and spreading within specific locations based on the content of the ad.

Figure 1: Fixed-effects Model Fit



[1] "Social Media Advertisements." U.S. House of Representatives Permanent Select Committee on Intelligence. Accessed April 28, 2020. https://intelligence.house.gov/social-media-content/social-media-advertisements.htm.

[2] Having individual specific intercepts $\boldsymbol{\alpha}_i$, $i = 1,...,n,$ where each of these can be understood as the fixed effect of entity i, the model will have fixed effects on the response variable. For the formula above, $\boldsymbol{\alpha}_i$ are entity-specific intercepts that capture heterogeneities across entities and are denoted as $\boldsymbol{\gamma}_n Dn_i$ (Econometrics with R, 10.3).

[3] Note that CTR remains a change in terms of percentage. This is calculated within all model regressions as CTR*100 to account for the response already being in percentage terms when it comes to interpreting the model,

We also created a classification model to predict whether an ad is propaganda in comparison to legitimate political ads. The classification model is especially practical in that anyone can pull up a Facebook ad, enter their location and race, copy and paste the ad's text and it will tell you whether it thinks the ad is propaganda. This may be advised because it typically correctly identifies propaganda 88.99% of the time historically, but we hope others will build upon this progress and build a model that can even more reliably identify disingenuous Facebook advertisements. Facebook and government officials (Such as the U.S. Cyber Command or the NSA[4]) could potentially employ this algorithm with a greater degree of accuracy, since they have access to ad creation hour time stamps, and can verify sponsors. In the past, they have easily identified propaganda by looking at receipts, since the Russians bought ads under their actual agency name, and paid in rubles. After being exposed in 2017, however, they probably shifted their tactics and have begun funneling money through shell companies pretending to be legitimate U.S. political interest groups. In addition, the number of advertisers on Facebook has increased dramatically in recent years (from 3 million in 2016 to 7 million in late 2019[5]), so it is increasingly difficult to verify the legitimacy of each firm advertising on Facebook. This means Facebook and others will have to rely more heavily on propaganda characteristics and traits (thematic content, syntax, target demographics, etc.) if they are to successfully root out propaganda. The IRA will also likely expand its campaigns to include other social media platforms (They actually recently upgraded their office building), so unless these platforms gather detailed metrics on each advertisement, they will have to rely on an advertisement's facade to combat disinformation.

Our analysis concludes that this was no doubt a strategic endeavor conducted by the Russian Internet Research Agency (IRA) to create upheaval in the United States political arena and to also create chaos in the diverse social dynamics scattered across different populations across the United States. Our regression analysis using transformation and jackknifed methods gathers that the IRA purposely targeted states with higher populations, often including major cities, as focal points for the dissemination of propaganda generating controversial topics or falsified facts pertaining to current affairs within the United States during a range of three years which included a presidential election year. Additionally, states were targeted with specific ads focused on specific content, whether false or true, that were relatable to the political and/or sociological ideologies of those individuals within that state. While the first model does well in explaining the way IRA Propaganda spreads and where, we found that classification models can fairly accurately predict whether or not an ad is propaganda. This study also concluded that predictors like target location, number of exclamation points, and thematic predictors (racial topics) are very important for separating observations into pure groups. Although the best classification model is only 37.99% better than random guessing, it appears that propaganda detection rate can be improved upon the addition of informative predictors, and given augmented classification models. Finally, true positive rates could improve with the application of a boosted model or linear support vector machine. Either way, a new model should employ a cross-validation style technique to study whether prediction rates change given different subsamples of the Propublica data.

In terms of our original political inclination going into this study, we were not able to address this given the limited number of ads that each state had and the complexity of politics. Since politics varies widely across states and even within states, we were not able to have sufficient and appropriate data to analyze the effects of fake Russian ads within individual states. The closest we were able to get was finding the most common words within states but there were not sufficient ads for almost all states to establish a theme for each individual state. By having the specific ad and its information in more local, or state, levels the data will be better suited to try and tackle the relationship between the type of ads that the Russian created to interfere in our election and the relationship that they have with politics. This could answer questions such as if the Russians ads focus mostly on creating division within or between states, if they take on both sides of politics (Republicans and Democrats) or if they mostly back one side, etc.

****Please access these GitHub accounts for interactive visuals of our analysis results:***

https://jandanel.github.io/Fall-2020-Capstone/

https://pbf2tp.github.io/

[4] Nakashima, Ellen. "U.S. Cyber Command Operation Disrupted Internet Access of Russian Troll Factory on Day of 2018 Midterms." The Washington Post. WP Company, February 27, 2019. https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7ff322e9_story.html.

[5] Clement, J. "Facebook Active Advertisers 2019." Statista, January 30, 2020. https://www.statista.com/statistics/778191/active-facebook-advertisers/.