Do Social Media matter? The impact of Twitter on US politics

Tiziano Rotesi
Bocconi University

August 30, 2017

Abstract

In this paper I study the impact that Twitter had on voting behavior during the 2008 and 2012 presidential elections. First I create a measure of the number of Twitter accounts over time and then study the correlation between this measure and political outcomes. In order to address endogeneity I implement an IV strategy where as instrument I exploit the popularity of MLB, NBA and NFL teams that sign new players with Twitter accounts at the time of the announcement, making therefore the social network more interested for their fans. By comparing OLS and IV estimates it emerges the presence of a bias, probably due to the fact that dissatisfied voters are overrepresented on the platform. IV estimates show that the impact on turnout is positive. Moreover I find a negative effect on the share of votes for Democratic Party. By using data from the Current Population Survey I also find that respondents tend to discuss less about politics in areas with a higher Twitter penetration. Taken together, these results suggest that the positive effect of Twitter on turnout is more likely to be a consequence of peer pressure at the time of the elections, rather than the result of a higher quality of information available to Twitter users.

1 Introduction

During the last decade Internet has changed the way we communicate and interact. Together with the strong increase in Internet penetration, the use of the Web has moved to a new model, in which users are the main source of content and do not just play a passive role. This introduces the possibility for the platforms that are driving this change to overcome the limits of pure entertainment, becoming a tool used to collect information, learn about peers and get involved in forms of collective action.

This paper studies the effect that Twitter had on political participation in the US. Twitter is one of the most popular microblogging platforms. It allows users to publish short public messages, called "tweets", that anyone can read, comment and share with others. This platform has gained a relevant spot in the public debate so that it is now common to read in the news the last tweets by major politicians or users' reactions to them. This media has also gained attention as one of the factors behind Obama's success in 2008 or the protest that took place during the Arab Spring.

Ex ante it is hard to predict the impact that Twitter had on participation and voting behavior. First, this platform could affect the amount of information available to users. Twitter could indeed enlarge the set of entertainment opportunities already available and thus crowd out more informative media as online newspapers. At the same time, through the network of contacts, users could be exposed to pieces of information that they would have ignored otherwise. The effect on participation would therefore be ambiguous, depending on which of these two elements prevails. A second dimension that seems relevant and novel with respect to traditional media is the social interaction between users. Social media and Twitter in particular are characterized by their ability to foster interaction, making users part of a public debate that would have been hardly accessible otherwise or, more simply, by allowing politically active participants to exert influence on their contacts. This could make users more engaged politically, in particular in areas and in moments characterized by a stronger political debate. Yet, there is a concern that this interaction takes place predominantly among like-minded users, potentially leading to ideological segregation and therefore viewpoints that are harder to change.

When trying to study this phenomenon the main issue that needs to be addressed is the one of endogeneity. It is indeed possible to imagine a number of unobservables that may be correlated with political preference and the level of activity on Twitter, for example the level of dissatisfaction with politics or how people are used to public debate. In order to identify the effect of Twitter I implement an IV strategy based on an instrument that exploits differences in popularity of teams from the three major sport leagues that receive new players with a Twitter account at the time of the transfer. The exact definition of the instrument will be given later in the paper, for the moment it should be noticed that the idea is relatively straightforward: Twitter, as a platform, offers content that is created by the users. The more interesting these users are, the more interesting the content. Clearly, sports players are a special kind of users because they appeal to a vast number of fans. Therefore, by being on Twitter an athlete should make his fans more willing to join the platform too, in order to receive messages posted by him[1]. Now, the decision to join Twitter could depend on a number of factors and it is not possible to exclude that players are influenced by how successful the social network is in the city they are living. A stronger take up among fans or friends would increase the probability a player decides to open an account and this would undermine the identification strategy as described above. For this reason the instrument will rely on a second element, the player movements from one team to the other. Transfers are indeed driven by the match between players' characteristics and teams' needs, factors that are unrelated to the political climate of the regions where fans are present.

OLS analysis shows no significant correlation between Twitter penetration and turnout, votes for Democratic Party or difference in vote share for the two candidates. Once I instrument Twitter penetration the picture changes. In particular I find that the effect on turnout is positive, while the effects on the vote share received by democrats is negative.

In order to gain a better understanding of the mechanisms behind this pattern I then move to survey data. By using answers available in the Current Population Survey (CPS) Civic Engagement Supplement (waves 2008 and 2013), I do not find any evidence that points towards a stronger interest in politics by the respondents. There is indeed a negative effect on how often respondents used to discuss politics with their family or friends and no effect on another important form of participation, that is contacting public officials.

Regarding the difference between OLS and IV estimates, I observe a negative bias for turnout and a positive bias for the vote share received by the Democratic party.

1.1 Literature

This paper contributes to the economics literature that studies the link between media and political outcomes such as voter turnout. Stromberg (2004)

[1] There is another way, more mechanical, in which celebrities could affect Twitter's popularity. When searching for a name of a person that happens to be on Twitter, among the first Google search results, there's usually the link to the Twitter profile. Therefore, people that could be looking for a player's name on Google or Bing, would become aware of the existence of Twitter. About this, the support page of Twitter says: *"Your Twitter profile shows up in Google searches because Twitter has a high Google search rank. Keep in mind that the words you write in your Twitter profile or public Tweets may be indexed by Google and other search engines, and cause your profile or Tweets to come up in a search for those terms."* Source: https://support.twitter.com/articles/15349.

studies the effect of the introduction of radio in the United States between 1920 and 1940 and finds a positive impact on turnout and on the amount of relief funds coming from early New Deal programs. Gentzkow (2006) shows that the introduction of commercial TV led instead to a decrease in turnout, making also voters less informed. DellaVigna and Kaplan (2007) study again the effect of TV, but focus on Fox News Channel introduction across the United States to study how media bias affects voting. Gentzkow, Shapiro, and Sinkinson (2011) estimate the effect of newspapers entries and exits on political participation and find a positive effect that is stronger before the introduction of radio and television. Closer to the question addressed in this paper, the works by Campante, Durante, and Sobbrio (2013), Falck, Gold, and Heblich (2014) and Gavazza, Nardotto, and Valletti (2016) study the effect of broadband on voting behavior, respectively in Italy, Germany and England. In all cases, authors find a negative effect of Internet availability on turnout and no immediate impact on voting behavior.

These papers suggest as main mechanism the quality of information offered by the media. By providing new and relevant information, newspapers and radio had a positive impact on participation. On the other hand, at least at first, both TV and Internet where used as a source of entertainment, reducing also consumption of traditional media. To better study how important the quality of information is, Snyder and Stromberg (2010) and Hall and Snyder (2015) focus on variation in coverage of politics by newspapers. In areas where newspapers markets closely resemble U.S. congressional districts political coverage is higher, citizens are more informed and this produces a positive effect on politicians' behavior, on policies, but also on turnout.

With respect to this literature, this paper contributes by focusing on social media, that can be seen as characterizing the second step in the evolution of how Internet is commonly used, with more relevance given to user-generated content. The results presented here point towards the possibility that social media had a positive effect on political participation, at least for what concerns voter turnout. The mechanism may not be the one on which the papers mentioned above are focused though, as a lower propensity to discuss about politics may not be related to a better quality of information. One explanation that seems promising is close to the work by DellaVigna et al. (2016), that focus on social image concerns with people voting "because others will ask". Social media can indeed create a new kind of peer pressure that makes it harder for voters to ignore elections.

Closely related to this paper, the work by Petrova, Sen, and Yildirim (2017) studies how Twitter affected political competition by increasing campaign contributions for politicians active on the platform.

Regarding the way Internet and in particular social media are influencing the political debate, one hypothesis that has attracted attention in the public and academic debate is that new media are creating "echo chambers" where participants are only exposed to homogeneous opinions, increasing extremism and polarization, making the electorate more divided than ever. Gentzkow (2016) contains a good summary of this literature.

Other forms of collective action have been considered, in particular protests participation and boycotts. Acemoglu, Hassan, and Tahoun (2014) study how popular mobilization influenced stock market valuations of firms connected to Mubarak's regime. In their analysis, the authors show how the number of tweets referring to Tahrir square where a good predictor of subsequent protest participation, suggesting that Twitter and other social media where used as a vehicle for mobilizing citizens. More in line with the empirical strategy presented in this paper, the work by Enikolopov, Makarin, and Petrova (2016) studies the impact that VK, a Russian social network similar to Facebook, had on protests participation in Russia in 2011. The empirical strategy relies on instrumenting VK's popularity in an area using the city of origin of the students that were allowed to join the network before it opened to the public. Findings are that VK penetration increased the probability of a protest and the number of protesters. Hendel, Lach, and Spiegel (2016) study a consumer boycott on cottage cheese that was organized using Facebook in Israel, after an increase in the price that was considered by consumers to be unfair. The authors find that the boycott had a strong impact on sales, more so in areas with higher presence of social media, suggesting that these websites played a major role in coordinating the actions of customers.

Outside the field of Economics, the impact of Twitter and other social media on participation has attracted a lot of interest by researchers. A first strand of papers use data from surveys to study how the use of social networking websites like Facebook or Twitter correlates with acts of participation as voting. In general the correlation is positive, even though this literature suffers of a lack of identification. For a meta-analysis of this literature see Boulianne (2015). Another recent contribution is the work by Margetts et al (2016) in which the authors, after analyzing the behavior of online users regarding small acts of participation like signing a petition or sharing a message in a social network, suggest a concept of *Chaotic Pluralism* to describe collective action today. Another strand of literature relies on data downloaded from the platforms. Two examples are Barberá and Rivero (2015) and Barberá (2015). In the former, the authors use tweets to study ideological position of users that wrote about politics and find that individuals with extreme positions are overrepresented. In the latter, the author measures ideological position of millions of individuals and finds that users are usually embedded in ideologically diverse networks, suggesting that social media may mitigate political polarization.

1.2 Background - Twitter

Twitter is a microblogging platform that allows users to publish short messages, tweets (max 140 characters), that are received by their followers. Tweets can then be shared with others or commented, possibly creating complex discussions involving a high number of participants. The website was launched in July 2006 and quickly became a mass phenomenon. In 2015 Twitter still ranked in the top 10 most popular websites[2], with approximately 66 million active users in the US and 320 worldwide. A survey made by PewResearch in 2014 shows that 21% of respondents were using Twitter. With respect to Facebook, the first social network in term of users, there are some relevant differences. In particular, from the beginning Twitter has appeared to be focused on the public sphere while the other was marketed as a tool to stay in touch with friends. This difference is evident under two aspects. First, Twitter accounts are public, while on Facebook there is a higher attention to privacy. Second, links on Twitter are unidirectional ('followers'), while on Facebook they are reciprocal ('friendship'). These differences are also reflected in the way users exploit the network. In particular 41% of users on Twitter say that reading comments by politicians, celebrities or athletes is a reason they use the website[3], share that is significantly higher than for Facebook.

This fact suggests the idea behind the instrument that we are using. The presence (or absence) of celebrities should affect users' interest in the platform and therefore Twitter penetration. Since players differ in their popularity both across regions and over time, we can exploit this variation to explain differences in Twitter popularity that are not related to political preferences.

2 Data

Before describing the data, it is necessary to specify that the analysis was carried at the Designated Market Area (DMA) level. DMAs are groups of counties defined by Nielsen on the basis of television market in such a way that all counties that belong to the same DMA have a similar TV offering[4]. These regions are not the same as metropolitan areas, even though in some cases the differences are small. In total there are 210 DMA regions. The reason why I use this level of aggregation is that Google Trends data, which are used to measure popularity, are available at DMA level but not at the county level. Electoral data, demographic controls and the measure of Twitter penetration were collected at the county level and then aggregated at DMA level. The

[2]Source: http://www.alexa.com/topsites

[3]For 11%, a major reason, 30% a minor reason. Source, Pew Research: http://www.pewinternet.org/2011/11/15/why-americans-use-social-media/

[4]From Nielsen website: "A DMA region is a group of counties that form an exclusive geographic area in which the home market television stations hold a dominance of total hours viewed."

sample of counties that I use is such that each county belongs entirely to one DMA[5].

The sample includes observations for DMAs that had data in both periods, 2008 and 2012. In total the sample contains 206 DMA regions.

Outcome variables and controls are standard to this literature and are described in subsection 2.1. Subsection 2.2 briefly illustrates the measure of Twitter penetration that I will then use as main regressor. Finally, subsections 2.3 and 2.4 show the data behind the instrument.

2.1 Electoral and Census Data

I collected data at the county level on turnout and voting behavior for Presidential Election in 2008 and 2012. The source is Dave Leip Atlas. Data include information on the number of valid votes and votes received by the candidates. Table 1 includes summary statistics for the outcome variables that I consider, once the data were aggregated at DMA level.

Controls were downloaded from Census and include age distribution across cohorts, income, race, gender and educational attainment[6]. Table 2 includes summary statistics for the variables that are included in the analysis.

Table 1: Outcome Variables - Summary Statistics

	2008		2012	
	Mean	SD	Mean	SD
Turnout	58.22	7.7	54.49	8.28
Democrat	47.61	10.33	45.26	11
Margin	21.4	10.77	23.55	11.67
N. Observations	206		206	

Note: *Turnout* is given by the ratio between the number of votes and the voting age population. *Democrat* represents that share of votes to the Democratic Party, while *Margin* is the difference in the two shares.

[5] In order to assign counties to DMA regions we used the file provided by Google that links cities to DMA regions: https://developers.google.com/adwords/api/docs/appendix/cities-DMAregions

[6] Other controls like percentage of rural area were not included because of the lack of variation in the 4 years considered.

Table 2: Control Variables - Summary Statistics

	2008		2012	
	Mean	SD	Mean	SD
Household Int Penetration	3.03	0.54	3.7	0.43
Population (log)	13.39	1.20	13.44	1.19
Density	2.22	0.21	2.31	0.23
Male	0.49	0.01	0.49	0.007
Under 18	0.24	0.03	0.23	0.02
Age 18-44	0.37	0.03	0.36	0.03
Age 45-64	0.25	0.02	0.26	0.02
Over 65	0.13	0.02	0.14	0.02
White (no hisp/lat)	0.73	0.17	0.71	0.18
Hispanic or Latino	0.06	0.1	0.09	0.12
High school or less	0.46	0.07	0.44	0.06
Some College	0.29	0.03	0.3	0.03
College or more	0.24	0.05	0.25	0.06
Poor	0.24	0.05	0.27	0.05
Income 75k or more	0.07	0.03	0.08	0.03
Language other than English	0.07	0.06	0.08	0.07
English "very well"	0.04	0.06	0.05	0.05
N. Observations	206		206	

Note: Controls are provided at the DMA level. *Household Int Penetration* refers to Residential Fixed High-Speed Connections per 1000 Households. Data were downloaded from Federal Communication Commission. Data are provided by county in a scale from 0 to 5 and were aggregated using population as weight.

2.2 Twitter Users

Twitter does not provide aggregated data on the number of active users and their geographic distribution. In order to build a measure of Twitter penetration across regions I relied on Twitter Search API[7] and downloaded information on accounts.

Figure 2 shows the screenshot of an account page. On the left, below the profile picture, it is possible to read username, description, location and the date the account was created. In the center of the page we can see the number of tweets (messages written by the user), the number of other accounts that

[7] API stands for Application Programming Interface. In this context it can be considered as a set of tools that Twitter makes available to interact with their database.

the user is following[8] (*Following*), the number of accounts that are following this user (*Followers*) and the number of messages that the user *liked*. In this case, we see that the username is *Robert M*[9], his location is Chicago and joined Twitter in November 2011. *Robert M* has written in total almost 400 tweets and is receiving direct updates from 254 other accounts. There are 440 users that are receiving every message written by *Robert M*.

It is important to underline that, while the creation date is always provided by Twitter and cannot be modified by the user, the location field contains information that is self reported. There are four cases that are typical, as showed in Table 3. The user could:

· Specify a location using GPS, as *Darcy*.

· Indicate a location that does not match with any place, as *fari*.

· Indicate a location that corresponds to a clearly identifiable place, as *Mindu*.

· Decide not to provide any information, as *deida*.

I collected a random sample of user ids and matched locations to Counties in the US. Table 4 summarizes the results of this operation when considering a subsample of approximately 33 million accounts. Over the 33 million accounts in the subsample, 69% of them did not include any location. I could match in total 6 million accounts, of which 1.5 million at the county level. The remaining 4.5 million are either foreign users or accounts that I could only associate to a country or a state. Table 4 reports also the average number of tweets, the average number of followers and the average number of likes for these subsamples. We can notice that on average, the accounts that leave the location field empty appear to be less active that the others. Moreover, if we select only accounts with at least 100 tweets, two out of three of them are providing some location. The analysis that is shown below relies on a larger sample of accounts. In particular the accounts that were matched at the county level were in total 3.2 million.

We then aggregate accounts at the DMA level. Figure 3 shows the kernel density of the number of accounts per 1000 inhabitants for the 206 DMA regions that we include in our sample. Figure 4 shows instead the number of accounts that were created per week in six DMA regions, for the period 2008-2010. From our data we can notice how Twitter popularity grew faster starting from early 2009. This fact is confirmed by Figure 5, that shows the

[8] On Twitter links are unidirectional. We say that user A is *following* user B when A is receiving all messages written by B. User B, instead, will not receive any update about A, unless he follows A back.

[9] The usernames provided in this and in the other examples have been changed in order not give precise reference to any real account.

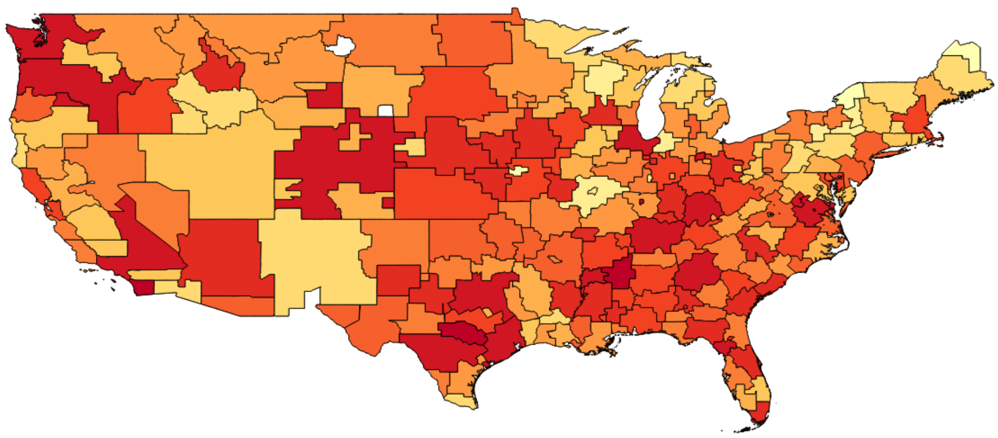
Table 3: Four examples of Twitter accounts

- Username *Darcy*:
  - Location: 43.64, -79.36 (Toronto)
  - Created: 2006-12-29 05:28:30
  - Followers: 506
  - Tweets: 6849
  - Description: Analytics, insights and all the things....
- Username *fari*:
  - Location: On the road
  - Created: 2006-12-29 20:02:50
  - Followers: 29092
  - Tweets: 59970
  - Description: I'm Fari: I'm looking for the awesome...
- Username *Mindu*:
  - Location: Chicago, Illinois
  - Created: 2008-12-06 14:01:52
  - Followers: 409
  - Tweets: 567
  - Description: Hi There!
- Username *deida*:
  - Location: -
  - Created: 2007-17-01 23:10:17
  - Followers: 51
  - Tweets: 1
  - Description: -

Table 4: Collected accounts

	Accounts	%	Avg n. Tweets	Avg n. Followers	Avg n. Likes
Total	33,129,071	100	1,009	201	147
Empty location	22,956,140	69	392	65	63
Some location	10,172,931	31	2,400	506	338
Matched	6,056,009	18	1,693	516	266
Matched to county	1,523,558	4.6	1,737	658	364
Empty, 100+ tweets	2,182,070		4,084	622	640
Some, 100+ tweets	4,418,421		5,504	1,129	768

Figure 1: Distribution of accounts by DMA in 2012



number of tweets over time (data from Twitter). Finally, Figure 1 shows the distribution of accounts in 2012.

2.3 Players

I collected a dataset of names of players from the three major leagues in the US that were active in the period 2007-2013. I then identified those who have a Twitter account and the day they joined. Table 5 report statistics on the number of players and accounts that were created over time. In the analysis that I present here I focus on accounts that have more than 100,000 followers,therefore I report statistics also for this subsample. Figure 6 plots the evolution of the number of accounts over time, for the three leagues.

Table 5: Players

League	Accounts	Avg n. Tweets	Avg n. Followers
NFL	2,341	7,878	51,755
NBA	591	7,250	340,371
MLB	905	3,305	50,176

As the instrument relies on tracking the transfers of players with a Twitter account I also collected data on transfers for the three leagues. For each transfer there are information on the date it was announced, the players and teams involved. Drafted players were included as well.

2.4 Popularity

In order to measure how popular teams were over time and across regions I used Google Trends[10]. Data were downloaded at the DMA level, with monthly observations. I used Google Trends in two different ways. First, I used it to get a measure of the popularity of each team across DMAs, over time. Then I compared the average popularity of different teams in order to assign more weight to those who were receiving more queries on Google search. In order to exclude the possibility that these scores were directly affected by Twitter I downloaded them for the period 2004-2008.

Given the high number of transfers, in order to select only players that would receive a high level of attention by the users I selected a subset based on charts of top 30 players that are released at the end of each season. For NBA I used the players that were invited to play at the All-Star game.

3 Empirical Specification

In order to investigate the effect that a stronger presence of Twitter had on turnout, we need to relate the geographic variation in Twitter penetration to geographic variation in turnout. The basic framework for our analysis is given by the following fixed effect model:

$$Y_{dt} = \beta_0 + \beta_1 Twitter_{dt} + X'_{dt}\beta_2 + \delta_t + \delta_d + \epsilon_{dt} \tag{1}$$

[10] From A Hands-on Guide to Google Data by Davidowitz and Varian (2015): *Google Trends reports an index of search activity. The index measures the fraction of queries that include the term in question in the chosen geography at a particular time relative to the total number of queries at that time.*

where t indexes years of election (2008 and 2012) and d indexes DMA regions. Outcome variables are presented in Table 1. The variable $Twitter$ measures the (log) number of accounts per 1000 inhabitants in the DMA region. I control for the set of census variables described above, including also the same variables in 2008 interacted with a time trend. Finally, I include year fixed effects and DMA fixed effects.

A critical issue in estimating the previous model is that of omitted-variable bias. The success of a new social network like Twitter is likely to be determined by characteristics of the population that we can only partially capture using demographics like age or education. These characteristics may be in turn correlated with turnout or voting behavior. To address this issue I implement an instrumental variable approach to exploit the fact that celebrities influenced Twitter's success by making the platform more interesting with their presence. As instrument I use differences in popularity across regions of teams who received players that have joined Twitter, where popularity is measured using Google Trends.

The specification I consider is the following one:

$$Twitter_{dt} = \alpha_0 + \alpha_1 Z_{dt} + X'_{dt}\alpha_2 + \delta_t + \delta_d + \nu_{dt} \tag{2}$$

The instrument is constructed according to the following formula:

$$Z_{dt} = \sum_c Incoming\_Twitter_{ct} \cdot Popularity_{cd} \tag{3}$$

Where c indexes teams. $Twitter_{dt}$ indicates the log of the number of accounts per 1000 inhabitants in region d at time t. $Incoming\_Twitter_{ct}$ measures the number of players that joined team c between period t and $t-1$ and that had a Twitter account when the transfer was announced. Only players that appear in the top charts described above are considered here. $Popularity_{cd}$ refers to the measure of popularity of team c in region d, calculated using Google Trends for the period 2004-2008.

In words, the instrument captures the shock that is generated when a transfer is announced. The player will receive a new wave of interest coming in particular from the fans of the team he will be playing for in the next season. In case the player has a Twitter account we expect part of this wave of interest to be transformed in new accounts on the social network, as some supporters will be interested in following the new member of their team. [11]

The identifying assumption is that, conditional on observables, the popularity in region d of the team that has received a new player with a Twitter

[11] An alternative way to describe this instrument is to make a parallel with Bartik instruments, named after Bartik (1991). The distribution of teams' popularity across regions can be seen as playing the same role of the distribution of local industry shares (each team being a different industry). Moreover, the number of incoming players with a Twitter account for each team works as national growth rates for the sectors.

account is orthogonal to unobserved determinants of voting behavior in region d. Moreover, in order for the exclusion restriction to be valid we need to assume that players, using Twitter, do not have a direct effect on political attitudes, for example by inviting their followers to vote.

4 Results

I study the effect of Twitter on political participation and voting behavior by focusing on three different outcomes. First, I look at the effect on turnout, then I consider the vote share received by the Democratic Party and the difference in the share of votes obtained by the two parties.

Baseline results are presented in Table 6. Table 6 has the following structure. Columns (1) and (2) show the OLS results for the panel regression described above. Columns (3) and (4) refer instead to the IV regression. All specifications consider as explanatory variable Twitter penetration. The first rows refer to the three different outcomes that I consider: turnout, vote share for Democratic Party and margin of victory. The fourth and fifth rows instead report results regarding the first stage regression. I show results with and without the set of controls described in section 2. All regressions include year and county fixed effects, together with the interaction between controls in 2008 and a time trend.

The F-stat of excluded instrument refers to the Kleinbergen-Paap F-statistic and is equal to 13.37 when I include controls while it is 58.42 without controls. The instrument appears therefore to be relevant. By looking at the first stage regression we can notice that the sign is as expected, with Z having a positive effect on Twitter penetration.

The OLS regressions presented in table 6 highlight a weak correlation between Twitter penetration and the three outcomes. By looking at columns (3) and (4) we can see how the IV estimates of the effect of Twitter penetration on turnout are positive. A similar picture emerges for the vote share received by Democratic Party and for the margin of victory. OLS estimates point towards a very weak correlation between the outcome variable and Twitter penetration. Results change in the IV regressions and show a negative effect on the vote share received by democratic party and no in the absolute difference between the shares of votes obtained by the two candidates, once we include controls.

By comparing OLS and IV regressions in table 6 we notice the presence of a bias in the OLS estimates. This bias is negative for turnout and positive for the vote share for Democratic Party. In other words, I find that individuals that are less likely to vote and that lean towards the Democratic Party are overrepresented on Twitter. One possible interpretation is that Twitter strongly attracts people who are dissatisfied either with politics or with society in general, by offering them a tool to express their opinion. People from these

Table 6: Estimates of the impact on Twitter on Turnout, share for Democratic Party and Margin of victory.

	OLS		IV	
	(1)	(2)	(3)	(4)
	Dependent Variable: Turnout			
Twitter	-0.955	-0.366	2.536	9.621**
	(1.114)	(-0.36)	(1.773)	(4.724)
	Dependent Variable: Democratic			
Twitter	1.815	-0.0796	-4.044***	-11.01***
	(1.286)	(-0.07)	(1.407)	(4.050)
	Dependent Variable: Margin of victory			
Twitter	2.315	0.312	5.242**	-0.381
	(2.393)	(0.13)	(2.637)	(7.324)
	First Stage			
Z	-	-	0.01***	.0085***
	-	-	(.0016)	(.0015)
F-stat Z	-	-	58.42	13.37
Controls	No	Yes	No	Yes
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
DMA regions	206	206	206	206
Observations	412	412	412	412

Note: Controls are described in Section 2. Clustering is done at the DMA level. F-stat refers to Kleibergen-Paap rk Wald F statistic.
Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

minorities would either turnout less because they had lost trust in politics or feel closer to the Democratic Party.

4.1 Survey Data

Before I comment these results and compare them with the literature, it is necessary to further investigate the relationship that emerges from the data. In particular it seems relevant to consider data from the Civic Engagement Supplement to the Current Population Survey, waves 2008 and 2013. This survey is particularly interesting because it contains questions regarding interest towards politics and forms of participation that are complementary to voting during the elections. As the data do not have a panel structure as in the case above, I estimate the following model:

$$Y_{i(dt)} = \beta_0 + \beta_1 Twitter_{dt} + X'_{idt}\beta_2 + \delta_t + \delta_d + \epsilon_i \quad (4)$$

Where i indexes the respondent, t indexes time and d indexes DMA regions. As controls I include the same set of controls that were used previously and add dummies for income level (three categories), age (five categories), gender, educational attainment (four categories), race (four categories) and employment status (three categories). $Twitter_{dt}$ is instrumented using the same strategy described above. I use dummy variables also as outcome, therefore the specification assumes a linear probability model. In particular I consider two outcomes. First, whether the respondent answered either *"Less than once a month"* or *"Not at all"* to the question *"How often did you discuss politics with family or friends?"*. Second, whether the respondent answered *"Yes"* to the question *"Please tell me whether or not you have done any of the following in the last 12 months, that is since November 2012: Contacted or visited a public official at any level of government to express your opinion?"*.

Results are presented in table 7. IV regressions show a positive effect of Twitter on the probability of not getting involved in discussions about politics and a negative effect on the likelihood of contacting a public official.

4.2 Discussion

From the analysis that relies on electoral data we see that Twitter had a significant effect on how people vote, increasing turnout and reducing the vote share for the Democratic party. It seems therefore that Social Media are having an effect that is significantly different from the one that is described in Falck, Gold and Heblich (2014) about the introduction of Internet as they observed a negative effect on turnout, but no effect on vote shares received by political parties. This is partially in line with the results by Campante, Durante and Sobbrio (2013) that suggest that the effect of Internet on politics has evolved over time. In this sense we can see Social Media as characterizing a second phase of the use of Internet. Regarding the effect on the margin of victory, this

Table 7: Estimates of the impact on Twitter, CPS survey data

	OLS		IV	
	(1)	(2)	(3)	(4)
	Dependent Variable: Never Discuss			
Twitter	-0.086**	-0.0098	0.042	0.225**
	(0.036)	(0.041)	(0.1)	(0.111)
First Stage				
Z	-	-	0.01***	.0095***
	-	-	(.0023)	(.0015)
F-stat Z	-	-	12.30	17.68
Observations	70034	70034	70034	70034
	Dependent Variable: Contact Public Off.			
Twitter	-0.013	-0.011	-0.073	-0.045
	(0.013)	(0.016)	(0.052)	(0.047)
First Stage				
Z	-	-	0.01***	.0095***
	-	-	(.0023)	(.0015)
F-stat Z	-	-	12.41	17.57
Observations	70843	70843	70843	70843
Controls	No	Yes	No	Yes
DMA FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Note: Controls are described in Section 2. Clustering is done at the DMA level (158 clusters). F-stat refers to Kleibergen-Paap rk Wald F statistic.
Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

result is not consistent with the hypothesis that sees social networks as echo chambers that increase polarization. Instead the result presented here seems closer to what is suggested by Barberá (2015) in his analysis of conversation patterns among Twitter users where he shows that most users are exposed to ideologically diverse network of contacts that lead to more moderate positions. Finally, an effect on turnout, although weaker, is what was also found in works cited by Boulianne (2015), especially those that were relying on panel datasets.

The analysis of CPS data adds interesting insights to the analysis, as it points towards a reduction in the level of interest towards politics that would be compatible with the idea of Twitter working as an entertainment platform, similarly as what was observed for TV or Internet.

Taken together, these results do not seem to be consistent, especially if we look at this question focusing on the dichotomy between information and entertainment that is usually considered in the literature. Indeed it is hard to explain these results by only considering the effect on information available to voters and a substitution mechanism among competing media that are more or less informative. It would be hard to justify an increase in turnout that occurs together with a reduction in the willingness to discuss about politics.

This relates to the nature of Twitter: it can be used both as a pure entertainment tool or as a source of information[12], but it is above all a platform on which users interact. This element reminds in particular the work of DellaVigna et al. (2016) who estimate a model in which people vote because of the pressure exerted by their peers that may ask about their choice. The increase in turnout could therefore be explained as being determined by the desire of each user to conform. This effect appears strong during the elections, that are the moment in which the interest towards politics reaches a peak, and then vanishes as the political debate goes back to normal.

For what concerns the bias that seems to emerge from the comparison between OLS and IV estimates, a possible interpretation is that voters that are particularly dissatisfied with politics are overrepresented on Twitter. This hypothesis is in line with what Barberá and Rivero (2015) find when they analyze a dataset of tweets written during the Spanish legislative elections in 2011 and the 2012 US presidential elections.

5 Conclusions

To summarize, in the analysis presented above I study the impact that Twitter had on voting behavior during the 2008 and 2012 US presidential elections. I find that Twitter penetration had a positive effect on turnout and a negative effect on the share of votes received by Democratic party. The preliminary

[12] How Twitter is used as a source of information is studied in this report by PewResearchCenter: http://www.journalism.org/2015/07/14/the-evolving-role-of-news-on-twitter-and-facebook/

results that emerge are only partially in line with the literature that studied the impact of Internet or traditional media on politics. This difference could be determined by the different nature of these media. In the context of social media, users are the main source of content, they are not just passively receiving entertainment opportunities or information and this element opens to the possibility of having users directly influencing each other.

References

Acemoglu, Daron, Tarek A. Hassan and Ahmed Tahoun. 2014. "The Power of the Street: Evidence from Egypt's Arab Spring," National Bureau of Economic Research Working Paper 20665

Barberá, Pablo. 2015. "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the United States." Working Paper.

Barberá, Pablo and Gonzalo Rivero. 2015. "Understanding the political representativeness of Twitter users." *Social Science Computer Review* 2015, 33 (6), 712-729.

Bartik, Timothy. 1991. "Who Benefits from State and Local Economic Development Policies?" *W.E. Upjohn Institute.*

Boulianne, Shelley. 2015. "Social media use and participation: a meta-analysis of current research." *Information, Communication and Society* vol 18, Iss 5.2015.

Campante, Filipe R., Ruben Durante, and Francesco Sobbrio. 2013. "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation." National Bureau of Economic Research Working Paper 19029.

DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *Quarterly Journal of Economics* 122 (3): 11871234.

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2016. "Voting to Tell Others." National Bureau of Economic Research Working Paper 19832.

Enikolopov, Ruben, Alexey Makarin and Maria Petrova. 2016. "Social Media and Protest Participation: Evidence from Russia." CEPR Discussion Paper No. DP11254. Hendel Lach Spiegel 2016

Falck, Oliver, Robert Gold, and Stephan Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review* 104 (7):2238-2265.

Gavazza, Alessandro, Mattia Nardotto and Tommaso Valletti. 2016. "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies." Working Paper.

Gentzkow, Matthew. 2006. "Television and Voter Turnout." *Quarterly Journal of Economics* 121 (3): 931-72. **Gentzkow, Matthew.** 2016. "Polarization in 2016." *Toulouse Network for Information Technology Whitepaper.*

Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review* 101 (7): 2980-3018.

Hall AB, James M. Snyder J. 2015. "Information and Wasted Votes: A Study of U.S. Primary Elections." *Quarterly Journal of Political Science* 10 (4):433-459.

Hendel, Igal, Saul Lach, and Yossi Spiegel. 2016. "Consumers' Activism: The Cottage Cheese Boycott." Working paper.

Jurgens, David, Tyler Finnethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. "Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice." Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM).

Leip, David. *Dave Leip's Atlas of U.S. Presidential Elections.*

Margetts, Helen, Peter John, Scott Hale and Taha Yasseri. 2016. "Political Turbulence: How Social Media Shape Collective Action." *Princeton University Press.*

Petrova, Maria, Ananya Sen, and Pinar Yildirim. 2017. "Social Media and Political Donations: New Technology and Incumbency Advantage in the United States." Working Paper.

Snyder, James M., Jr., and David Strömberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy* 118 (2): 355-408.

Strömberg, David. 2004. "Radio's Impact on Public Spending." *Quarterly Journal of Economics* 119 (1): 189-221.

Figures

Figure 2: Screenshot of a profile on Twitter.

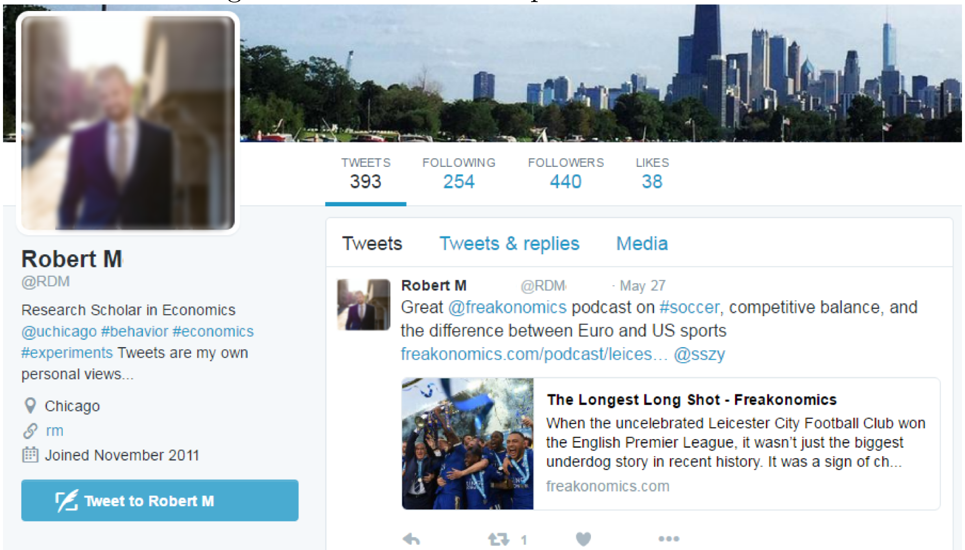


Figure 3: Distribution Twitter penetration

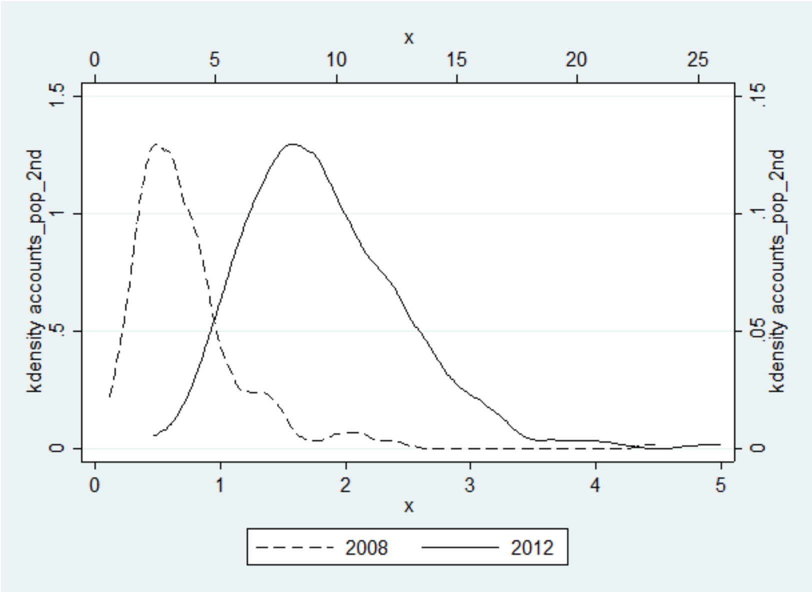


Figure 4: New accounts per week

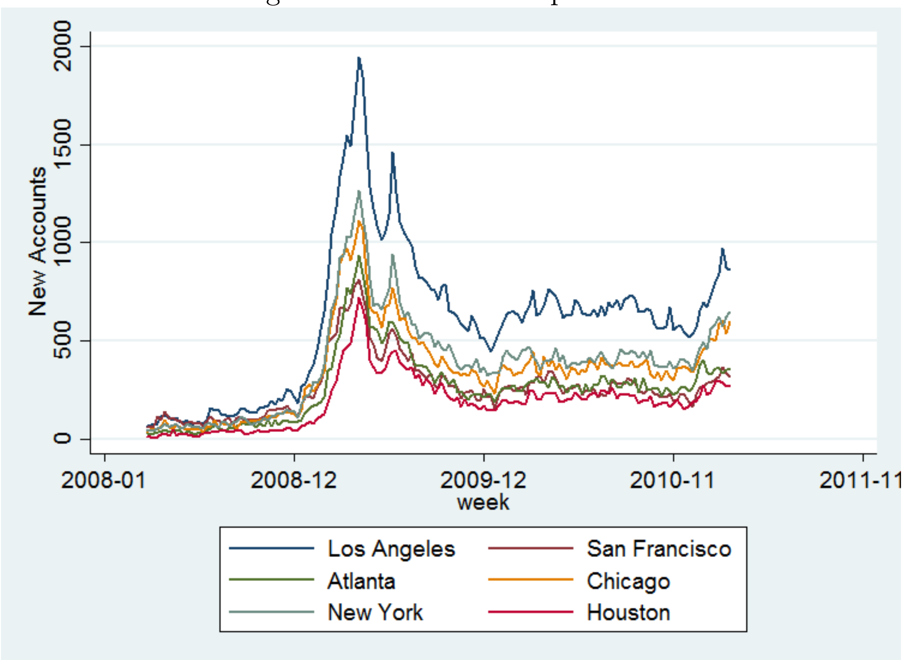


Figure 5: Tweets per day - Source: Twitter

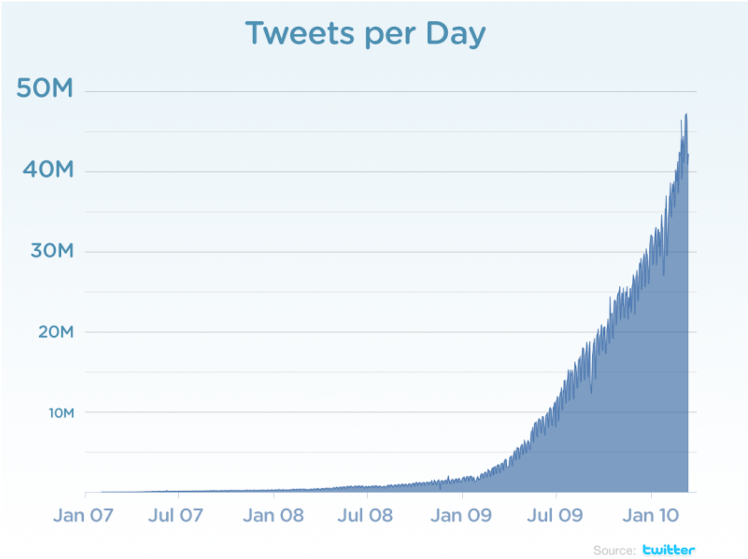


Figure 6: Accounts over time - Players