THE EFFECT OF SOCIAL MEDIA ON ELECTIONS: EVIDENCE FROM THE UNITED STATES

Thomas Fujiwara

Princeton University and NBER, USA

Karsten Müller

National University of Singapore, NUS Business School, Singapore

Carlo Schwarz

Bocconi University, Department of Economics, Italy

Abstract

We study how social media affects election outcomes in the United States. We use variation in the number of Twitter users across counties induced by early adopters at the 2007 South by Southwest festival, a key event in Twitter's rise to popularity. We show that this variation is unrelated to observable county characteristics and electoral outcomes before the launch of Twitter. Our results indicate that Twitter lowered the Republican vote share in the 2016 and 2020 presidential elections, but had limited effects on Congressional elections and previous presidential elections. Evidence from survey data, primary elections, and text analysis of millions of tweets suggests that Twitter's relatively liberal content may have persuaded voters with moderate views to vote against Donald Trump. (JEL: L82, D72)

Teaching Slides

A set of Teaching Slides to accompany this article is available online as Supplementary Data.

The editor in charge of this paper was Romain Wacziarg.

Journal of the European Economic Association 2024 22(3):1495–1539 https://doi.org/10.1093/jeea/jvad058

© The Author(s) 2023. Published by Oxford University Press on behalf of European Economic Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Acknowledgments: We are grateful to Pablo Barberá, Levi Boxell, Matt Gentzkow, Ro'ee Levy, Alexey Makarin, Maria Petrova, Jesse Shapiro, James Snyder, Ekaterina Zhuravskaya, as well as seminar participants at Princeton University, Imperial College, Warwick University, Bocconi University, University of California San Diego, Insper, University of Sao Paulo, Inter-American Development Bank, IPEG/Pompeu Fabra, University of Zurich, OPESS, Università di Bergamo, the 2021 ASSA Conference, NBER Political Economy Spring 2021 Meeting, NBER IT and Digitization Summer Institute 2021, and the ENS Lyon for their helpful comments. We thank Pedro Cabra-Acela and Wenqing Yu for excellent research assistance. Müller acknowledges funding from the Singapore Ministry of Education (MOE) through a start-up grant and Presidential Young Professorship (grant numbers A-0003319-01-00 and A-0003319-02-00).

E-mail: fujiwara@princeton.edu (Fujiwara); kmueller@nus.edu.sg (Müller); carlo.schwarz@unibocconi.it (Schwarz)

1. Introduction

Does social media affect election outcomes? A popular narrative holds that Twitter played a decisive role in both recent American presidential elections and the United Kingdom's "Brexit" referendum. Many see this as part of social media's broader influence on political polarization and the re-emergence of populist politicians in many countries. The U.S. Federal Election Commissioner, for example, has argued that Facebook "has no idea how seriously it is hurting democracy" (NPR 2020a).¹

An alternative view suggests that social media platforms are biased against conservatives (e.g., NPR 2020b; Wall Street Journal 2020) and that its younger, relatively left-leaning user base is unlikely to tilt elections toward right-wing politicians (e.g., Boxell, Gentzkow, and Shapiro 2017, 2018). However, there is limited evidence that can be used to evaluate these contrasting (causal) claims.

This paper focuses on the effects of Twitter, a platform used by almost a quarter of American adults. We estimate how a county's number of Twitter users affects election results by exploiting a persistent network effect sparked by early Twitter adoption, building on Müller and Schwarz (2023).² Although it was launched in March 2006, Twitter's popularity increased rapidly after its advertising campaign at the South by Southwest festival (SXSW) in March 2007. The SXSW festival was also key for Twitter's geographical diffusion: counties with more SXSW followers who joined during the 2007 festival saw disproportionately higher growth of Twitter adoption compared to counties with SXSW followers who already joined before the festival. Consistent with path dependence in technology adoption, this difference in Twitter use across counties persists.

Our identification strategy leverages the 2007 SXSW festival as a shock to early Twitter adoption that is uncorrelated with pre-existing election results. Conditional on geographic controls and previous interest in the SXSW Twitter account, a county's number of SXSW followers who joined in March 2007 is essentially uncorrelated with a host of county characteristics. It is also unrelated to election outcomes before Twitter's launch (going back as far as 1924) and during the period it had fewer users (between 2006 and 2012). However, the number of SXSW followers who joined in March 2007 is correlated with Twitter usage in 2016, and has predictive power for the 2016 and 2020 presidential election results.

We estimate that a 10% increase in a county's number of Twitter users lowered the vote share of Republican presidential candidate Donald Trump by 0.2 percentage points (p.p.) in both the 2016 and 2020 presidential elections. The implied persuasion rates are 8.6% and 9.4%, respectively. These estimates are smaller than the estimated pro-Republican effect of Fox News (DellaVigna and Kaplan 2007; Martin and Yurukoglu 2017), the pro-Democrat effect of the Washington Post (Gerber, Karlan, and Bergan 2009), or the effect of get-out-the-vote canvassing on turnout (Gerber

^{1.} See, for example, The New Yorker (2016); New York Times (2017); Allcott and Gentzkow (2017); The Guardian (2018); UK Parliament (2019).

^{2.} Enikolopov, Makarin, and Petrova (2020) use a similar empirical strategy based on spatial variation in early adopters of the social media network VK in Russia to study its effects on protests.

and Green 2000), but larger than the effect of an independent anti-Putin Russian TV channel on vote shares (Enikolopov, Petrova, and Zhuravskaya 2011) or the effect of TV rollout on turnout (Gentzkow 2006).

For presidential elections before 2016, we find effects that are small and statistically indistinguishable from zero. The same holds true for House and Senate races, including the 2016 and 2020 elections. We thus detect a negative effect of Twitter adoption on Trump's vote share but do not do so for Republican candidates in congressional races in the same election. This pattern bolsters confidence that our estimates capture an effect of Twitter, which contains more content on presidential than congressional candidates.

To shed light on the mechanisms behind these results, we estimate Twitter's effect on vote choices reported in the Cooperative Congressional Election Survey (CCES), primary presidential candidates' approval in the Gallup Daily Tracker, and countylevel results in the 2016 and 2020 presidential primaries. Further, we explore data on the partisanship of political content on Twitter.

These exercises yield three findings. First, the CCES results indicate that Twitter's effect is driven by independents and moderates switching their votes toward the Democratic candidate (Hillary Clinton). This is consistent with Bayesian updating, since moderates presumably have weaker priors and are thus more likely to be persuaded.

Second, we find that Twitter also lowered Trump's vote share during the 2016 primaries, a finding we confirm using individual-level Gallup candidate approval ratings. We find that Twitter decreased Trump's approval ratings and increased Clinton's with only small effects on relatively more moderate Republican candidates.³

Third, we document that political content on Twitter has a pro-Democratic slant. We classify the slant of tweets based on two complementary approaches: one based on the network users follow, and one using the text of tweets in a machine-learning approach in the spirit of Gentzkow and Shapiro (2010). We apply these methods to the over 460 million tweets mentioning the presidential candidates in the 2012, 2016, and 2020 elections. We find that the number and attention (proxied by "likes") of tweets mentioning Trump was substantially larger than that of those mentioning Clinton and Joe Biden. Moreover, tweets about Trump in 2016 and 2020, 70% more likely to have Democratic rather than Republican slant.

Overall, our results indicate an effect that is specific to Trump and not other Republican candidates. One potential interpretation is that Trump's own behavior on Twitter created a "backlash" effect of moderate voters against him. Trump adopted Twitter as his preferred social media platform and differed from most high-profile candidates in its use, both in the amount of tweeting and also in its tone and content (Bursztyn, Egorov, and Fiorin 2020). Another interpretation is that Twitter has relatively pro-Democratic content persuading moderate voters about a high-profile candidate. These two interpretations are not only complementary but also likely to feed

^{3.} We also estimate effects for the 2016 and 2020 Democratic primaries, detecting a (positive) effect for Bernie Sanders in 2020.

on each other: Trump's tweets generate many other tweets that paint him in negative light (accusing of misinformation or inappropriate speech), which can in turn have persuasive effects.

Our work contributes to the literature on the impact of media on political outcomes. Expansions of traditional media such as newspapers, radio, broadcast television, and cable news have been associated with changes in voter turnout, polarization, and electoral outcomes.⁴ While a set of papers studies the effect of overall internet access, the effects of social media *per se* received less attention.⁵

A nascent literature studies the political effects of social media on protest participation (Howard et al. 2011; Enikolopov, Makarin, and Petrova 2020; Acemoglu, Hassan, and Tahoun 2017; Fergusson and Molina 2021), xenophobia (Müller and Schwarz 2023; Bursztyn et al. 2019; Müller and Schwarz 2020), and mental health (Braghieri, Levy, and Makarin 2022).⁶ Additionally, a burgeoning field of experimental research focuses on social media. Bond et al. (2012) and Jones et al. (2017) provide evidence that online messages on social networks affect voter turnout. Allcott et al. (2020) and Mosquera et al. (2020) find that individuals who deactivate Facebook react along many dimensions, including some measures of political polarization. Levy (2021) studies the effect of randomly assigning Facebook users subscriptions to conservative or liberal media outlets. Bail et al. (2018) estimate the effect of paying Twitter users to follow a bot with messages of the opposing political ideology. Bessone et al. (2022) studies how Facebook affected Brazilian politicians' behavior. Most related to our paper is recent unpublished work by Rotesi (2019), who finds social media negatively affected the Democratic vote share in the 2008–2016 presidential elections using variation in Twitter adoption resulting from transfers of NBA players with Twitter accounts.⁷

Existing research thus provides an incomplete picture. On one hand, social media has been painted as a key force behind political change, and experimental studies indeed suggest that social media affects individuals' self-reported political beliefs. On the other hand, it remains unclear whether social media can indeed persuade voters

^{4.} See, for example, Gentzkow (2006); Huber and Arceneaux (2007); DellaVigna and Kaplan (2007); Gerber, Karlan, and Bergan (2009); Gerber et al. (2011); Gentzkow, Shapiro, and Sinkinson (2011); Enikolopov, Petrova, and Zhuravskaya (2011); Campante and Hojman (2013); DellaVigna et al. (2014); Larcinese and Miner (2017); Martin and Yurukoglu (2017); Spenkuch and Toniatti (2018); Chen and Yang (2019).

^{5.} There is evidence that broadband internet (Falck, Gold, and Heblich 2014; Gavazza, Nardotto, and Valletti 2019; Campante, Durante, and Sobbrio 2017; Lelkes, Sood, and Iyengar 2017) and mobile internet (Manacorda and Tesei 2016; Guriev, Melnikov, and Zhuravskaya 2020) exert political effects.

^{6.} For reviews; see DellaVigna and Gentzkow (2010), Napoli (2014), Strömberg (2015), Enikolopov and Petrova (2015), and DellaVigna and Ferrara (2015) and in particular Zhuravskaya, Petrova, and Enikolopov (2020) for the case of social media.

^{7.} Our paper differs from Rotesi (2019) not only in its research design, but in that we estimate effects for a larger set of election years (including 2020) as well as Congressional elections. Rotesi (2019) only reports estimates pooling the 2008–2016 presidential elections, differently from our paper. This does not allow to see if his instrument is uncorrelated with election results before Twitter's launch and how its possible effects evolved over time.

and affect election results on a larger scale. Our paper sheds light on this question by focusing on how Twitter affects federal elections in the United States.

2. Background: Social Media and Politics

Most Americans use social media platforms or messaging applications. Data from the Pew Research Center suggest that the most popular services are YouTube (used by 73% of adults in the United States), followed by Facebook (69%), and Instagram (37%) (Pew Research Center 2019c). A total of 22% of adults in the United States use Twitter, a rate similar to that of Snapchat (24%) and Whatsapp (20%) users. On average, adult users spend more than an hour a day using social networks (eMarketer 2019).⁸

One popular perspective is that online networks, and social media in particular, may give rise to so-called "filter bubbles" (Pariser 2011) or "echo chambers" (Sunstein 2017). The idea is that social media—unlike traditional mass media outlets—may facilitate the provision and consumption of one-sided information, either through the use of algorithms or by allowing individuals to self-select into preferred content. While there is considerable empirical evidence supporting this idea (e.g., Conover et al. 2011; Weber, Garimella, and Batayneh 2013; Bessi et al. 2015; Del Vicario et al. 2016; Halberstam and Knight 2016; Schmidt et al. 2017; Levy 2021), other studies have found that individuals are exposed to a wide range of political opinions on social media (Barberá 2014; Bakshy, Messing, and Adamic 2015; Nelson and Webster 2017; Beam, Hutchens, and Hmielowski 2018), perhaps even more so than via traditional media outlets or personal interactions (Gentzkow and Shapiro 2011). Some work also challenges the notion that increased polarization due to online channels is quantitatively important (Flaxman, Goel, and Rao 2016; Guess 2018; Boxell, Gentzkow, and Shapiro 2019).

Much of the recent public discussion about the role of social media platforms has been shaped by controversies, including Cambridge Analytica's involvement in political campaigns (e.g., The Guardian 2018); the Russian Internet Research Agency's efforts to support Trump's campaign (e.g., New York Times 2017); and the role of misinformation ("fake news") (e.g., Allcott and Gentzkow 2017). Both Clinton and Trump have argued that these factors were instrumental in the 2016 election outcome, as has former president Barack Obama (The New Yorker 2016). As Brad Parscale, Trump's digital media director in 2016, put it: "Facebook and Twitter were the reason we won this thing. Twitter for Mr. Trump. And Facebook for fundraising" (Wired 2016). In Online Appendix Figure A.3, we document that discussions of social media have become increasingly frequent in major American news outlets.

Such controversies intensified in the aftermath of the 2020 election. Twitter permanently suspended Trump's account in the aftermath of the January 6, 2021,

^{8.} Pew bases its usage measures on the share of respondents who state they have ever used one of the online platforms. Twitter reported around 69 million monthly active users in 2019 (see Statista 2019), which yields a slightly higher share of around a third of the 210 million adults in the United States.

invasion of the Capitol. In October 2022, Elon Musk took ownership of the platform, reversed Trump's suspension, and released what he labeled the "Twitter Files." The journalists that first received the files argued they showed a pattern of Twitter taking politically motivated decisions to aid Democrats' electoral chances, such as moderation of news stories related to Biden's son's "laptop controversies" and providing relatively less visibility to conservative accounts (e.g., NPR 2022). The files also documented that Twitter regularly communicated with the FBI about addressing the spread of misinformation and dealing with foreign influences. Others have argued that the "Twitter Files" themselves were a political tool and, if anything, showed the platform's desire to remain politically neutral and limit the spread of misinformation (e.g., New York Magazine 2022).

Under Musk's leadership, Twitter implemented a series of changes to the platform, most notably affecting which Tweets receive more visibility and what content creates ground for a suspension for the platform. Many argued this further politicized the platform and moved the content most users see toward a conservative slant (e.g., USA Today 2023; The Guardian 2023). At the time of writing, the landscape of American social media is rapidly changing, with new entrants such as Bluesky and Threads providing platforms with user interfaces that are similar to Twitter.

These events highlight both the prominence of social media's effects on political speech and the perceived need for regulation over social media platforms' moderation policies. However, whether social media actually affects electoral outcomes is largely unknown, and some have suggested that concerns about its effects may be overblown. As one example, in the 2016 presidential election, Trump received fewer votes from demographic groups with higher propensities to use social media or the internet more broadly (The Hill 2016; Boxell, Gentzkow, and Shapiro 2017, 2018). Indeed, Trump's broadest support came from older white voters without college education in rural areas, who are among the least likely people to use social media actively (Hargittai 2015; Pew Research Center 2015, 2018). These patterns seem difficult to square with the idea that online channels were an important driver of the 2016 presidential election result, although such observations also do not rule this out.

Further, most social media users—particularly on Twitter—appear to be disproportionately left-leaning. This is not surprising given that most Twitter users are relatively younger, educated, and from urban areas. While there appears to be a cluster of right-wing networks, Pew Research Center (2019d) estimates that, in 2018, 60% of Twitter users identified as Democrat and only 35% as Republican. Among Democrats, those on Twitter are considerably more liberal and focus less on finding common ground with Republicans (Pew Research Center 2020). In 2019, 26% of American Twitter users followed Obama, and 19% followed Trump (Pew Research Center 2019a). Survey evidence suggests that 80% of Twitter content is produced by people who strongly disapprove of Trump (Pew Research Center 2019b). "Liberals" are also more likely to get political news on Twitter or Facebook and follow more media and political accounts compared to "conservatives" (Pew Research Center 2014; Eady et al. 2019). Twitter and Reddit, which are often said to be pro-Trump factors, were considerably more popular among Clinton supporters before the 2016 election



FIGURE 1. Twitter reach by party. This figure plots data on the Twitter reach of Congress members. The sample includes all 901 senators and House representatives who were in office between 2007 and 2019 for whom we could identify a Twitter account. For each account, we plot the average number of tweets and followers, and the average number of "likes" and retweets of their tweets. Online Appendix Figure A.2 replicates the figure using medians instead of averages. The data were collected from Twitter in November 2019.

(Hargittai 2015). Although social media allows users to partially select which content they see, Twitter content disproportionately leans toward the Democratic party.

We provide additional evidence for the composition of political content on Twitter by analyzing the Twitter reach of Democratic and Republican politicians. We collected data on the Twitter accounts of all Senators and House Representatives from the 110th to the 115th Congresses (2007–2019). In Figure 1, we plot the average number of tweets and followers that members of each party have on Twitter, as well as the average number of retweets and "likes" their tweets receive. The patterns here again clearly indicate that Democratic politicians are more active on Twitter and have larger follower bases than their Republican counterparts. Tweets by Democrats also receive, on average, three times the number of "likes."⁹

^{9.} In Online Appendix Figure A.2, we confirm that these patterns are not driven by a small group of Congress members by showing that they also hold when we compare the median Twitter reach of Democrats and Republicans.

3. Data

The main analysis is based on a county-level dataset on election outcomes, political opinions, and Twitter use. It covers 3,065 counties in 48 states (we exclude Alaska and Hawaii) and the District of Columbia (except in congressional elections). County-level election results are from Dave Leip's Atlas of U.S. Presidential Elections and the MIT Election Lab. We complement our analysis with individual-level survey data on approval ratings from the Gallup Daily Tracker and voting data from the CCES. Our measure of Twitter usage is derived from an archive of 475 million geo-located tweets compiled by Kinder-Kurlanda et al. (2017). We combine this with newly collected data on Twitter's early adopters at the 2007 SXSW festival; data on the Twitter activity of U.S. Congress members; and a large corpus of tweets related to the 2012, 2016, and 2020 presidential elections. Additional county characteristics were obtained from the U.S. Census, the U.S. Religious Census, the American Community Survey (ACS), and the Bureau of Labor Statistics (BLS). We describe the individual data sources in more detail below. Online Appendix Table A.1 provides additional details and summary statistics.

Election Outcomes. We use county-level data on presidential election outcomes between 1924 and 2020 from Dave Leip's Atlas of U.S. Presidential Elections. From the same source, we also obtained county-level voting data for the Republican and Democratic primaries in 2016 and 2020. We complement this with county-level results on Senate and House elections from the MIT Election Lab for the 1996–2020 period. In all cases, we focus on two-party vote shares.¹⁰

Individual-Level Voting Decisions. The CCES is a nationwide survey that collects information on voter behavior in two waves (before and after the election). We focus on votes for Trump and Clinton in 2016 and 2020. The CCES contains a rich set of individual characteristics, including political affiliation, family income (in 12 bins), gender, race, education (in 6 bins), marital status, age, and interest in the news. Online Appendix Table A.2 provides summary statistics (weighted by sample weights). The CCES also uses administrative data on turnout records to verify its respondents have voted.

Presidential Candidate Approval. The Gallup Daily Tracker provides individuallevel survey data for a sample of 1,000 individuals per day since 2009.¹¹ During the 2016 presidential campaign, it fielded survey items regarding the approval of Republican and Democratic presidential candidates. This allows us to investigate

^{10.} While senatorial and presidential elections are decided at the state level and House elections at the congressional district level, counties are usually smaller geographic units and far more numerous. Additionally, unlike congressional districts, county boundaries are fixed over our sample period, allowing us to observe changes across years.

^{11.} The Gallup Daily Tracker for the 2020 election is not available at the time of writing.

Trump's pre-election approval relative to other candidates (e.g., Clinton or Ted Cruz). The data also include a rich set of individual characteristics, including political affiliation, county of residence, income (in 10 bins), gender, race, marital status, age, and education (in 6 bins). Online Appendix Table A.3 provides summary statistics.¹²

Twitter Usage. We construct a measure of county-level Twitter usage based on a sample of 475 million geo-coded tweets collected by Kinder-Kurlanda et al. (2017).¹³ The tweets were collected between 2014 and 2015 using the Twitter Streaming API by selecting a geographic bounding box around the mainland United States. The Streaming API continuously returns all geo-located tweets within the bounding box as long as the sample does not exceed 1% of all tweets. Information on a tweet's geo-location either come from the global positioning system (GPS) coordinates of a mobile phone or from a WiFi/IP address in case a computer is used. Both types of information allow for a precise assignment to a U.S. county. At the start of the data collection by Kinder-Kurlanda et al. (2017) in 2014, Twitter had not yet introduced the feature that allowed users to tag specific places in their tweets. This avoids collecting geo-located tweets based on arbitrary decisions by users (e.g., tagging their holiday location) and prevents users from tagging wrong locations (either intentionally or by accident).

The individual tweets from this dataset are already assigned to counties. Additionally, we collected the underlying user profiles for each tweet in the database. This allows us to construct a user-based measure by assigning users to the county from which they tweet most frequently. The resulting measure, which we use throughout the paper, is a proxy for the number of Twitter users per county, based on 3.7 million individual users (around 7% of the Twitter population in 2015). Figure 2(a) plots the number of Twitter users per capita across counties. Each user profile further provides us with a short biography and the date that each user joined Twitter. We use the join dates to construct a time-varying proxy of Twitter usage based on how many of the Twitter users had opened an account at each point in time.

The great advantage of this dataset is that it allows us to provide individual-level evidence for the adoption of Twitter in a county following the SXSW festival and further compare user profiles in different counties. The drawback of using geo-located Twitter data is that only a sub-sample of tweets is geo-located. To overcome concerns of measurement error in our Twitter measure, we validate the data in two ways. First, Online Appendix Figure A.1(a) shows that our Twitter usage measure's evolution closely tracks the number of daily Twitter users from Statista (2019), which were directly obtained from the platform. Secondly, our measure of county-level Twitter usage also strongly correlates with the number of Twitter users in a county based on the GfK Media Survey (see Online Appendix Figure A.1(b)). Lastly, note that measurement error is less of a concern in our setting because we largely rely on 2SLS

^{12.} For some estimations, we also collapse responses about approval of Trump to the county level using weighted averages based on the number of survey respondents in each county.

^{13.} These data are available in the Gesis Datorium at https://datorium.gesis.org/xmlui/handle/10.7802/1166.

(a) Twitter Usage per Capita



FIGURE 2. Twitter usage and identifying variation. These maps plot the proxy for social media usage based on data from Twitter and the identifying variation of our instrument. Panel (a) plots quintiles of the number of Twitter users per capita. Panel (b) plots the three types of counties relevant for our identification strategy: (1) the 47 counties with SXSW followers that joined Twitter both in March 2007 and the "pre-period" (light red); (2) the 108 counties with SXSW followers that joined in March 2007, but none in the "pre-period" (dark red); and (3) the 20 counties with SXSW followers that joined in the "pre-period," but none in March 2007 (blue).

estimation throughout the paper. We return to the discussion of measurement error in Section 5.

Twitter Data for the South by Southwest Festival. We collected data for our instrument for Twitter usage, based on early adoption during the SXSW festival, through the Twitter API. More specifically, we scraped the account data for 658,240 users who followed the Twitter account of SXSW Conference & Festivals (@SXSW) at the time of collection (January 2019). We assign these users to counties based on the location people report in their user profile.¹⁴

^{14.} Of the 44,625 SXSW followers who joined between 2006 and 2008, we are able to geo-code 25,830 (58%).

A user profile contains the month and year that they joined Twitter, which allows us to determine the number of SXSW followers in each county that joined Twitter in a particular month. The two key variables in our analysis are: (i) the number of SXSW followers that joined Twitter in the month of March 2007 and (ii) the number of SXSW followers that joined Twitter during 2006 (the year the platform was launched). We refer to (ii) as the number of SXSW followers who joined before the March 2007 festival. We also scraped the follower lists of SXSW followers who joined in March 2007, which allows us to investigate the connections of Twitter users to the SXSW festival. Further, we additionally collected tweets mentioning the festival, based on the term "SXSW," as well as a proxy for overall Twitter activity based on the 100 common English words. We use these measures to document the SXSW festival's impact on local Twitter adoption.¹⁵

Data on Political Twitter Activity. We scraped the tweets, user profiles, and followers of the 901 Senators and House Representatives from the 110th to 115th (2007-2019) Congress who have Twitter accounts. This includes 424 Democrats and 465 Republicans.¹⁶ In total, the data contain 4,300,579 tweets, which we use to analyze the Twitter reach of Democratic and Republican Congress members.

We complement this dataset with election-related tweets to shed light on the overall partisan slant of Twitter activity during the 2012, 2016, and 2020 elections. For each election, we obtained the universe of tweets mentioning the last name of a Democrat and Republican presidential candidates.¹⁷

To determine the likely political affiliation of Twitter users, we create two measures of political slant. The first measure is based on the political accounts a user is following. In particular, we check whether a user follows more Democrat or Republican Congress members on Twitter. If they follow more Republican than Democrats, all their tweets would be classified as Republican. In case a user either does not follow any Congress members or an equal number of Congress members from either party, their tweets are classified as neutral.¹⁸

The second measure of political slant is based on the similarity of the text of tweets to those sent by Republican or Democratic Congress members. We train a

^{15.} Data on SXSW 2007 attendants (e.g., their county of residence) is not available, despite our efforts to obtain it from the SXSW organizers on multiple occasions.

^{16.} The remaining twelve politicians are either Independents or switched their party affiliation.

^{17.} For the 2012 election, we use data collected by Diaz et al. (2016), comprising 24 million tweets containing either "Obama" or "Romney" for the period from July 1, 2012 through November 7, 2012. For 2016, we use the archive from Littman, Wrubel, and Kerchner (2016), which contains 280 million tweets, collected between July 13, 2016, and November 10, 2016. The 2020 election tweets are based on the archive from Chen, Deb, and Ferrara (2020), which covers the period from March 2020 to November 2020. To make these datasets comparable, we restrict the 2016 election sample to tweets mentioning either "Clinton" or "Trump" (112 million tweets). Similarly, we restrict the 2020 dataset to the time period from July 1, 2020 through November 3, 2020 and tweets mentioning either "Biden" or "Trump" (339 million tweets).

^{18.} The idea of using the Twitter network to determine a user's ideology is inspired by Barberá (2015).

L2 regularized logistic regression model separately for each election based on 901 Congress members' Twitter accounts to classify whether a tweet contains language frequently used by either Republican or Democratic politicians.¹⁹ We then use this classifier to predict a partisan score between 0 and 1 for each of our election-related tweets. These scores can be interpreted as the probability of a tweet with the same content being sent by a Republican. As such, our approach is similar to how Gentzkow and Shapiro (2010) measure newspaper slant. Both approaches lead to similar overall slant classifications for the election tweets in our data.

Additional County Characteristics. We collect county-level demographic control variables from the U.S. Census and the ACS. In particular, we use information on population, population share by age group and ethnicity, poverty rates, and education levels. We also obtained industry-level employment shares and unemployment rates from the BLS. Additional controls on county media usage patterns are from Simply Analytics. We also construct geographical controls such as the distance from Austin, TX, where SXSW takes place every year; population density; and county size (in square miles). For one set of results, we also use donation data from OpenSecrets.

4. The 2007 South by Southwest Festival and Early Twitter Adoption

The empirical strategy behind our main results exploits a shock to early-stage Twitter adoption connected to the 2007 SXSW festival, as in Müller and Schwarz (2023). This section discusses the key role of the festival in boosting the platform's popularity and documents how it created a persistent effect on its spatial diffusion.²⁰

Founded in March 2006, Twitter was largely unknown before SXSW 2007. Twitter's popularity increased dramatically after the festival, where Twitter strategically placed screens in the conference hallways and allowed users to sign-up by simply sending a text message to a predefined number. As a result, speakers and bloggers in attendance broadcasted the platform to the outside world, and Twitter went on to win the South by Southwest Interactive Web Award Prize.

The importance of SXSW 2007 has also been stressed by the platform's founders. As co-founder Evan Williams explained in a post on Quora (Quora 2011):

"We didn't actually launch Twitter at SXSW—SXSW just chose to blow it up. We launched it nine months before—to a whimper. By the time SXSW 2007 rolled around,

^{19.} We clean the text of the tweets by removing common words (stopwords) and by reducing the words in each tweets to their morphological roots (lemmatizing). The input is based on unigrams, bigrams, and trigrams from these tweets. We choose the optimal normalization strength using 10-fold cross-validation. The resulting classifier achieves high out-of-sample F1-scores, for example, 0.904 for the tweets during the 2020 presidential election. We provide additional details regarding the machine learning classifier in Online Appendix A.1, which also visualizes the most predictive terms identified by the classifiers.

^{20.} SXSW is an annual conglomeration of parallel film, interactive media, and music festivals and conferences organized jointly that take place in March in Austin, TX.

we were starting to grow finally, and it seemed like all of our users (which were probably in the thousands) were going to Austin that year ... I don't know what was the most important factor, but networks are all about critical mass, so doubling down on the momentum seemed like a good idea. And something clicked."²¹

SXSW's immediate impact on Twitter's popularity in early 2007 can be seen in Figure 3(a), which plots our proxy for the daily number of tweets as well as the number of tweets explicitly mentioning SXSW. The figure shows that Twitter's growth rate accelerated during the festival, visible as the spike in SXSW-related tweets. The month-to-month growth rate of Twitter quadrupled with the start of the SXSW festival.²² After SXSW 2007, Twitter experienced further rapid growth (Venture Beat 2008). The platform went from an average of 5,000 tweets a day in 2007 to 300,000 in 2008, and 2.5 million in 2009 (Twitter 2010). In 2019, users sent roughly 500 million tweets a day.

We exploit that the SXSW festival had persistent effects on Twitter's spatial diffusion. This is likely the result of network effects that are key to the social media experience, as a larger number of users makes it more interesting for potential new users to join. Such a mechanism also applies at the local level. For example, a boost in the number of neighbors, personal connections, local businesses, and/or people who play a prominent role in an area should also boost the value of joining the platform for those living there. As Evan Williams' quote above notes, "networks are all about critical mass," and initial differences in adoption can lead to persistent differences in network adoption. Enikolopov, Makarin, and Petrova (2020) document a similar mechanism for spatial dispersion in the adoption of the Russian VK social network: The home towns of the first university students invited to be users in 2006 exhibited a higher number of users in 2011.

We provide further support for this hypothesis by investigating whether the inflow of early-stage adopters put these counties on a differential growth path of Twitter usage. Figure 3(b) plots the estimates of β_{τ} from the following panel event study regression at the county (c) and week (t) level:

$$tweets_{ct} = \sum_{\tau} \beta_{\tau} SXSW_{c}^{March2007} \times 1(t = \tau) + \sum_{\tau} \delta_{\tau} SXSW_{c}^{Pre} \times 1(t = \tau) + \theta_{c} + \gamma_{t} + \varepsilon_{ct}$$

where $tweets_{ct}$ is the log of (one plus) the number of tweets in county c on week t, $SXSW_c^{March2007}$ is the logarithm of (one plus) the number of SXSW followers in county c that joined Twitter on March 2007 and $SXSW_c^{Pre}$ is a similarly defined variable for followers that joined Twitter before March 2007. β_{τ} thus illustrates, conditional

^{21.} Online Appendix Figure B.1 provides Williams' full post describing the role of SXSW 2007.

^{22.} Our proxy for Twitter usage is created by scraping tweets that contain any of the 100 most common English words. Our data contain any tweet that contains at least one of these words. We should therefore obtain a large fraction of the English-speaking tweets at that point in time.



(a) Twitter Activity Around SXSW 2007

FIGURE 3. SXSW 2007 and the spread of Twitter. Panel (a) plots the total number of tweets and the number of tweets containing the term "SXSW" over time, smoothed using a 7-day moving average. Panel (b) plots the estimates of β_{τ} from the panel event study regression $tweets_{ct} = \sum_{\tau} \beta_{\tau} SXSW_c^{March2007} \times 1(t = \tau) + \sum_{\tau} \delta_{\tau} SXSW_c^{Pre} \times 1(t = \tau) + \theta_c + \gamma_t + \varepsilon_{ct}$ where $tweets_{ct}$ is the log of (one plus) the number of tweets in county *c* on week *t*, $SXSW_c^{March2007}$ is the logarithm of (one plus) the number of SXSW followers in county *c* that joined Twitter on March 2007 and $SXSW_c^{Pre}$ is a similarly defined variable for followers that joined Twitter before March 2007. We standardize the variables to have a mean of 0 and standard deviation of 1. The whiskers represent 95% confidence intervals based on standard errors clustered by state.

on county and week fixed effects, the difference in the number of tweets sent from counties with relatively larger numbers of SXSW followers that joined on March 2007. The variables are standardized to have a mean of 0 and standard deviation of 1. The whiskers represent 95% confidence intervals based on standard errors clustered at the state level. The sample includes the period between the third and fourteenth week of 2007.

Figure 3(b) illustrates that home counties of SXSW followers who joined during the festival in March 2007 saw a rapid, disproportionate increase in Twitter usage around the time of SXSW. Importantly, however, this increase came only after the SXSW festival, and we find no evidence for pre-existing trends. In Appendix Figure B.2 we show further that, in line with our expectations, March 2007 is also a clear outlier when it comes to the number of people who started following the SXSW festival. This is consistent with the idea that SXSW was a catalyst for the spread of Twitter in the United States.

Online Appendix Figure B.3(a) presents additional evidence on the long-term adoption effect of the 2007 SXSW festival. It plots estimates from a similar regression as the one in Figure 3(b) but in a county-quarter panel covering the period from Twitter's launch in 2006–2016. The dependent variable is substituted by the number of Twitter users per capita in a county based on our baseline measure. The resulting S-shaped pattern in the figure is consistent with models of technology adoption in the presence of network effects. More importantly, we find that the amount of early adopters in a county still matters for the amount of Twitter usage today.²³

5. Empirical Framework

Our identification strategy leverages the 2007 SXSW festival as a shock to early Twitter adoption. We show that, conditional on a set of controls (described in further detail below), a county's number of SXSW followers that joined Twitter in March 2007 is uncorrelated with levels and trends in election outcomes before Twitter's launch and during its early years. It is also uncorrelated with a host of observable county characteristics. This feature of the data can be interpreted as idiosyncratic factors (e.g., who attended the 2007 SXSW, who decided to join Twitter at the time), giving us a "natural experiment" or "exogenous shock" in Twitter adoption that allows to estimate its effect on election outcomes. This interpretation is, of course, not self-evident, and we provide several pieces of evidence to support it.

^{23.} Additionally, Online Appendix Figure B.3b shows just how dominant Twitter users connected to the SXSW festival were among early adopters. In 2007, we estimate that around 60% of Twitter users either followed the SXSW festival or followed someone who followed SXSW and joined in March 2007. As the number of Twitter users increased over time, the importance of SXSW followers in the platform declined. But as Online Appendix Figure B.3(a) shows, the festival created persistent differences at the county level. The next section outlines how we use the SXSW festival in our 2SLS estimates.

An important concern is that counties whose populations are more interested in the SXSW festival (and its Twitter account) may be systematically different from other counties. To address this issue, our empirical strategy exploits variation in the exact timing of when Twitter users interested in SXSW joined the platform across counties. In particular, our regressions control for the number of SXSW followers who joined in the months *before* the festival. Intuitively, our empirical strategy compares a "treatment" group of counties with SXSW followers that joined in March 2007 (during the festival) against a "control" group of counties with followers that joined before. While both groups of followers were interested in SXSW, we show that only the number of followers that joined on March 2007 are predictive of later Twitter adoption, consistent with the evidence that users that joined during the festival were key in the platform's diffusion. In contrast, counties with more users that joined *before* the festival during Twitter users in subsequent years.²⁴

The "treatment" and "control" counties are similar along several characteristics. Online Appendix Table A.4 compares the average characteristics of three types of counties relevant for our identification strategy: (1) the 47 counties with SXSW followers that joined Twitter both in March 2007 and the "pre-period;" (2) the 108 counties with SXSW followers that joined in March 2007 (but none in the "pre-period"); and (3) the 20 counties with SXSW that joined in the "pre-period" (but none in March 2007). Differences in vote shares in the 1996 presidential election, demographics (e.g., race, age, and education), and media consumption (e.g., share that watches Fox News) are quantitatively small or zero. This is particularly true for groups (2) and (3)—which are key to the identification—with *t*-tests indicating that differences between the two groups are not statistically different from zero.²⁵ The geographical variation in the three groups of counties is shown in Figure 2(b). As the results in Online Appendix Table A.4 suggest, the counties do not differ systematically in size and how distant they are from major American cities.

Moreover, observable individual characteristics of SXSW followers who joined Twitter in March 2007 and the "pre-period" are also similar. We validate this using data on Twitter user profiles we obtained from the platform. Online Appendix Table B.1 shows that followers who joined in March 2007 have similar first names and profile descriptions compared to those that joined before: users in both groups tend to have common names (e.g., "Michael" or "Chris") and use words such as "founder" or "tech" to describe themselves in their profiles. The correlations of the frequency of first names and terms used in their bios between the two groups are 0.63 and 0.89, respectively. We also investigate differences in the political leanings of the two groups using the network-based methods we outline in Section 3. In particular, we test whether the

^{24.} An alternative approach is to compare the counties of users who signed up for Twitter during SXSW 2007 with those of users who signed up during *other* festivals in the same year. We discuss the results from such an exercise in the robustness section below.

^{25.} Given the large number of county characteristics, we report Šidàk-corrected *t*-statistics, which are smaller than those generated by applying the Bonferroni correction.

users in March 2007 follow more Democrats or Republicans than the users in the "preperiod". We find that the political leanings of the two groups are nearly identical. A *t*-test rejects differences in the average political slant with a *p*-value of 0.93.

Specification. Motivated by the evidence above, our main results are based on estimating the following two equations:

$$Twitter \ users_c = \alpha + \beta \cdot SXSW_c^{March2007} + \gamma \cdot SXSW_c^{Pre} + \mathbf{X}_c \delta + \xi_c, \qquad (1)$$

$$y_c = \alpha' + \beta' \cdot SXSW_c^{March2007} + \gamma' \cdot SXSW_c^{Pre} + \mathbf{X}_c \delta' + \zeta_c,$$
(2)

where *c* indexes counties, $SXSW_c^{March2007}$ is the logarithm of (one plus) the number of SXSW followers in county *c* that joined Twitter on March 2007, and $SXSW_c^{Pre}$ is a similarly defined variable for followers who joined Twitter before March 2007. \mathbf{X}_c is a vector of control variables that hold constant geographical factors (e.g., population density, distance from Austin, TX), demographic factors (e.g., the share of the population in different age and ethnic groups), socioeconomic factors (e.g., the share of adults with a high school degree or employed in IT), factors related to the "China shock" (e.g., exposure to Chinese import competition), and previous election results. Note that the right-hand side of both equations is similar. *Twitter users_c* is the logarithm of the number of Twitter users in the county (during 2014–2015). y_c are election outcomes (e.g., vote shares), which we estimate in both levels and changes (e.g., y_c can be the vote share in 2016 or the change in vote shares between 2000 and 2016).

In a 2SLS framework, equations (1) and (2) are the first-stage and reduced form, while the second stage is

$$y_c = \varphi + \theta \cdot Twitter \, users_c + \pi \cdot SXSW_c^{Pre} + \mathbf{X_c}\rho + \varepsilon_c, \tag{3}$$

where $Twitter users_c$ is predicted from the first stage regression in equation (1). We weigh observations by turnout (total number of votes cast) in the 2000 presidential election.²⁶ We cluster standard errors at the state level.²⁷

Identification. Formally, the identification condition for the effect of Twitter users (θ) is that $E(SXSW_c^{March2007} \cdot \varepsilon_c) = 0$ holds. Intuitively, this states that, conditional on the $SXSW_c^{Pre}$ and other controls (\mathbf{X}_c), the number of SXSW followers who joined in March 2007 is uncorrelated with other determinants of political outcomes y_c , implying that it only affects political outcomes via Twitter usage (the "exclusion restriction").

^{26.} We weigh observations to make our sample representative of national election results. For example, due to many less populated counties tilting Republican, the unweighted average Republican vote share across counties in the 2016 elections is 64%. Weighing makes our sample mean match the national average of 46%. Moreover, we obtain similar results without using weights or using election-year turnout as weights (Online Appendix Table C.2), suggesting that effect heterogeneity along dimensions correlated with weights do not play an important role in our context (Solon, Haider, and Wooldridge 2015).

^{27.} We consider spatial standard errors using the methods described in Colella et al. (2019) for robustness.

We provide six pieces of evidence in support of this condition. First, as discussed above, including the $SXSW_c^{Pre}$ control implies that the identifying variation comes from comparing counties with similar observable characteristics.

Second, the coefficient of $SXSW_c^{Pre}$ is small and statistically insignificant in our first stage regressions. This provides us with a "placebo" test based on checking if it is also unrelated to political outcomes in the reduced form and two-stage least squares (2SLS) regressions. Intuitively, we have two variables that are correlated with interest in the SXSW festival among early Twitter adopters, but only one predicts Twitter users in later years, allowing us to disentangle interest in the festival from its effect via more Twitter users.

Third, we provide additional placebo tests for other festivals in 2007 and show that none of these other festivals is either correlated with higher Twitter adoption or changes in the Republican vote share today. This allows us to rule out that our effects are simply driven by the selection of users into attending festivals that are highly similar to SXSW.

Fourth, estimating equations (2) and (3) for different time periods shows that $SXSW_c^{March\ 2007}$ does not correlate with both levels and trends in election outcomes before Twitter's launch in 2006 and in its early years, when the platform had few users and was unlikely to affect election outcomes. Intuitively, outcomes in "treatment" and "control" counties behaved similarly before Twitter could plausibly affect elections.

Fifth, we detect a statistically significant effect of $SXSW_c^{March 2007}$ on Trump's vote share in 2016 and 2020 but not on House and Senate elections (neither in 2016 nor in 2020 or other periods between 2000 and 2018). This pattern is consistent with an effect of Twitter, since there is more content on presidential candidates than on congressional elections in the platform.

Sixth, results based on survey data suggest the effects are concentrated among moderate or independent voters, which is also the expected pattern from Twitter having a causal effect due to voter persuasion.

Stated differently, a violation of the identification condition would require an omitted variable that correlates with $SXSW_c^{March 2007}$, *Twitter users*_c, and y_c but is uncorrelated with: (i) $SXSW_c^{Pre}$ as well as followers of other festivals, (ii) levels and trends in election results before Twitter's launch and rise to popularity, (iii) the observable variables presented in Online Appendix Table A.4, and (iv) election results in congressional elections both during the Trump elections and before, while also (v) being correlated with vote choices of moderate voters. Our argument is that the existence of such an omitted variable is implausible to an extent that allows us to interpret θ as the effect of Twitter users on election outcomes.

Measurement error in county-level Twitter usage and $SXSW_c^{March 2007}$ is also unlikely to explain an effect in 2016 and 2020 presidential elections, but no effect in previous presidential elections or congressional elections. Moreover, $SXSW_c^{Pre}$ and the measures for other festivals are constructed similarly as $SXSW_c^{March 2007}$ and should thus have similar measurement error. However, $SXSW_c^{Pre}$ and the other festivals are uncorrelated with Twitter usage and election outcomes.

Lastly, another possible concern is that the SXSW adoption shock led to differences in the *composition* of Twitter users when compared to other U.S. counties. In

particular, one might be concerned that the SXSW festival lead to a more liberal Twitter population in the treated counties. While this would not influence the causal interpretation of our findings, it could make the local average treatment effect harder to interpret. Three pieces of evidence suggest that this appears to be an unlikely concern. First, as we show in Online Appendix Figure B.3(b), Twitter's user base became less connected to the SXSW festival over time and, in this process, likely reached people from more diverse backgrounds. Second, the findings of Müller and Schwarz (2023) indicate that the SXSW adoption shock was associated with an increase in hate crime with Trump's presidential run. This suggests that the shock eventually reached even the right-wing fringes of the political spectrum. Third, we can directly address this concern by comparing the profiles of Twitter users in SXSW home counties with those in the rest of the country. The results are presented in Online Appendix Table B.2. We find that the user profiles in SXSW counties are highly similar to the general Twitter population. If the Twitter population in SXSW counties was significantly more liberal, their Twitter and names and biographies should also be different, as Oliver, Wood, and Bass (2016) document that names predict political ideology. We find similar results when we look at which politicians users in the different counties follow. If anything, Twitter users in the "pre-period" counties appear to have a slightly more liberal Twitter network.

To be transparent, we want to stress what our findings do *not* imply. First, they cannot speak about social media platforms other than Twitter, such as Facebook. Our empirical strategy exploits a "shock" specific to early Twitter adoption, and we do not have a credible research design to estimate the effects of other platforms. While many other platforms share similarities with Twitter, such as being popular among younger and more educated people in urban areas (Pew Research Center 2019c), other platforms may have different effects on political outcomes. Second, our research design cannot separate the effect of particular types of social media content on Twitter (e.g., foreign governments or misinformation), but rather speaks to the overall effect of Twitter exposure. Third, like other papers in media economics, we estimate a "partial equilibrium" effect. In our case, we estimate the effect of adding Twitter users to a county while keeping other counties' Twitter use constant. We thus cannot address whether Twitter had a national-level effect on the election (e.g., Trump's tweets driving traditional media content).

6. Results

6.1. Main Results

First-Stage. Table 1 reports results from estimating equation (1) with different sets of control variables. The results indicate that counties with more SXSW followers who joined Twitter in March 2007 have higher numbers of Twitter users during 2014–2015. Since the variables are in logs, the coefficients can be interpreted as elasticities. A 10% increase in SXSW followers in March 2007 is associated with 5.2% more Twitter users. The results do not seem to be sensitive to the set of included covariates. For example,

	1	Dependent vo	ariable: Log(Twitter users	s)
	(1)	(2)	(3)	(4)	(5)
Log(SXSW followers, March 2007)	0.726***	0.683***	0.563***	0.524***	0.523***
	(0.087)	(0.079)	(0.055)	(0.048)	(0.048)
Log(SXSW followers, Pre)	0.104	0.110	0.059	0.059	0.058
	(0.101)	(0.076)	(0.098)	(0.082)	(0.082)
Population deciles	Yes	Yes	Yes	Yes	Yes
Census region FE	Yes	Yes	Yes	Yes	Yes
Geographical controls		Yes	Yes	Yes	Yes
Demographic controls			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes
China shock controls				Yes	Yes
1996 election control					Yes
Observations	3,065	3,065	3,064	3,064	3,064
R^2	0.92	0.93	0.95	0.95	0.95
Mean of DV	8.22	8.22	8.22	8.22	8.22
p-value: March 2007 = Pre	0.00	0.00	0.00	0.00	0.00

TABLE 1. South by Southwest 2007 and Twitter usage.

Notes: This table presents county-level regressions where the dependent variable is the number of Twitter users (in natural logarithm). *Log(SXSW followers, March 2007)* is the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow SXSW. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006, defined similarly. The bottom row reports *p*-values from *F*-tests for the equality of these coefficients. Regressions include the indicated control variables (see the Online Appendix for their descriptions). Observations are weighted by turnout in the 2000 presidential election. Standard errors in parentheses are clustered by state. *** p < 0.01, ** p < 0.05, * p < 0.1.

the distance from Austin, Texas (the location of SXSW) has no significant explanatory power. Importantly, the coefficients on SXSW followers before the 2007 festival are statistically insignificant and small in size: Twitter usage in 2014–2015 is not higher in areas with more SXSW followers who joined Twitter before March 2007.

Figure 4 presents the graphical representation of the estimates in column (5) of Table 1. Specifically, we show a binned scatter plot of *Twitter users*_c against $SXSW_c^{March 2007}$ after both variables are "residualized" by partialling out the control variables. The figure is constructed by dividing the X-axis variable into 40 equal-sized bins and plotting the average values of both variables in each bin.²⁸

Reduced Form and 2SLS Estimates. Table 2 shows the reduced form estimates from equation (2) and both ordinary least squares (OLS) and 2SLS estimates of equation (3), focusing on the Republican vote share in the 2016 and 2020 presidential elections. The specifications across columns match those in Table 1. Panel B indicates that the

^{28.} The fitted line is based on the unbinned data. Observations are weighted by turnout in the 2000 presidential election. This procedure guarantees the slope of the fitted line matches the estimate in column (5) of Table 1.



FIGURE 4. First Stage SXSW and Twitter usage. This figure presents a binned scatter plot of the relationship between Twitter users in 2014–2015 and the number of SXSW followers who joined Twitter in March 2007. Variables are residualized by partialling out SXSW followers who joined before March 2007, population deciles, Census region fixed effects, as well as geographical, demographic, socioeconomic, China shock, and 1996 election control variables (see Online Appendix for control variable definitions). The figure is constructed by dividing the X-axis variable into 40 equal-sized bins and plotting the average values of both variables in each bin. The fitted line is estimated using the unbinned data.

number of SXSW followers who joined Twitter in March 2007 is associated with a lower Republican vote share. Panel C presents the 2SLS effects of Twitter usage on vote shares. The 2SLS estimate in column (5) indicates that a 10% increase in the number of Twitter users in a county lowers Trump's vote share by 0.21 p.p. (e.g., a reduction from a 46.1% vote share to 45.8%).²⁹ We discuss potential differences between local average treatment effect (LATE) and average treatment effect (ATE) as well as heterogeneous treatment effects at the end of this section. The results for the 2020 presidential election shown in Table 2 columns (6)–(10) are nearly identical.

Figure 5 plots the reduced form and OLS estimates from Table 2 graphically, specifically the models in columns (5) and (10). These figures are constructed similarly to Figure 4 but show the Republican vote share on the Y-axis. The estimated slopes are negative.

Magnitudes and Persuasion Rates. Given the Electoral College and the closeness of individual state races, our effects are consistent with Twitter potentially affecting election results. For example, our results indicate that increasing the number of Twitter users by 30% lowered Trump's vote share by 0.63 p.p. This is larger than the margins

^{29.} The F-statistic of our estimated first-stage range from 70 to 120. This suggests that estimation and inference concerns related to weak instruments are unlikely to apply in our case.

		Dependent varic	ıble: Republican voi	te share in 2016			Dep. var.:	Republican vote sh	are in 2020	
	(1)	(2)	(3)	(4)	(5)	(9)	(1)	(8)	(6)	(10)
Panel A: OLS										
Log(Twitter users)	-0.065^{***} (0.009)	-0.067*** (0.008)	-0.013^{***} (0.004)	-0.011^{***} (0.003)	-0.007** (0.003)	-0.058^{***} (0.009)	-0.064^{***} (0.009)	-0.012^{***} (0.004)	-0.011^{***} (0.003)	-0.008^{***} (0.003)
Panel B: Reduced form										
Log(SXSW followers, March 2007) Log(SXSW followers, Pre)	$\begin{array}{c} -0.053^{***} \\ (0.011) \\ -0.021 \\ (0.016) \end{array}$	$\begin{array}{c} -0.058^{***} \\ (0.012) \\ -0.003 \\ (0.013) \end{array}$	-0.019*** (0.005) (0.006) (0.006)	$\begin{array}{c} -0.014^{***} \\ (0.004) \\ -0.002 \\ (0.006) \end{array}$	-0.011*** (0.004) 0.001 (0.004)	$\begin{array}{c} -0.046^{***} \\ (0.009) \\ -0.022 \\ (0.016) \end{array}$	$\begin{array}{c} -0.055^{***} \\ (0.010) \\ -0.005 \\ (0.013) \end{array}$	$\begin{array}{c} -0.017^{***} \\ (0.006) \\ -0.002 \\ (0.007) \end{array}$	$\begin{array}{c} -0.013^{***} \\ (0.005) \\ -0.004 \\ (0.007) \end{array}$	-0.011** (0.005) -0.001 (0.005)
Panel C: 2SLS										
Log(Twitter users)	-0.072***	-0.085***	-0.034***	-0.027***	-0.021***	-0.064***	-0.080***	-0.031**	-0.025***	-0.020**
Log(SXSW followers, Pre)	(0.020) -0.014 (0.020)	0.007 0.016)	0.002 0.002 (0.007)	(0.006) (0.006)	(0.005) (0.005)	-0.015 (0.020)	0.004 0.004 (0.015)	(1000) -0.000 (0.008)	(0.007) (0.007)	(0000) (0000)
Population deciles	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Census region FE Geographical controls	Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes	Yes	Yes Yes	Yes Yes	Yes Yes	Yes Yes
Demographic controls			Yes	Yes	Yes			Yes	Yes	Yes
Socioeconomic controls			Yes	Yes	Yes			Yes	Yes	Yes
1996 election control				102	Yes				102	Yes
Observations	3,065	3,065	3,064	3,064	3,064	3,065	3,065	3,064	3,064	3,064
Mean of DV Robust F-stat.	0.46 69.57	0.46 74.65	0.46 106.08	0.46 118.21	0.46 121.18	0.47 69.57	0.47 74.65	0.47 106.08	0.47 118.21	0.47 121.18
Notes: This table presents co March 2007) is the number of of SXSW followers who regis (see the Online Appendix for	unty-level regr Twitter users (tered at some p their descripti	essions where (in logs, with 1 oint in 2006, d¢ ons). The first-	the dependent added inside) v sfined similarly. stage regressio	variable is the who joined in N . Twitter users i ns for 2SLS re	Republican vc Aarch 2007 and are the number sults (Panel B)	the share in the follow South of users in 201-) are presented	2016 or 2020 by Southwest ((4–2015. Regres in Table 1, wit	presidential el SXSW). <i>SXSW</i> sisons include th th the F-stat fo	ection. Log(SXS followers, Pre i he indicated con r the excluded	<i>W followers,</i> is the number itrol variables instrument in

TABLE 2. Twitter and the 2016/2020 Republican vote share.

1516

Downloaded from https://academic.oup.com/jeea/article/22/3/1495/7308471 by guest on 07 June 2024

(10) to (6)); see text for further details. Observations are weighted by turnout in the 2000 presidential election. Standard errors in parentheses are clustered by state. *** p < 0.01, ** p < 0.05, * p < 0.05, * p < 0.1.



2016 Election

FIGURE 5. South by Southwest, Twitter, and the Republican vote share. Panel (a) presents a binned scatter plot of the relationship between the Republican vote share in the 2016 presidential election and the number of SXSW followers who joined Twitter in March 2007. Variables are residualized with respect to SXSW followers who joined before March 2007, population deciles, Census region fixed effects, as well as geographical, demographic, socioeconomic, China shock, and 1996 election control variables. The figure is constructed by dividing the X-axis variable into 40 equal-sized bins and plotting the average values of both variables in each bin. The fitted line is estimated using the unbinned data. Panels (b) and (d) replicate the exercise using Twitter users in 2014–2015 X-axis variable. Panel (c) and (d) show results for the 2020 election.

by which he won Michigan, Wisconsin, and Pennsylvania in 2016. Had Clinton won those states, she would have been elected president.

While the effects reported on Table 2 appear modest at first pass, they have to be interpreted taking into account that only a quarter of Americans are active Twitter users (Pew Research Center 2019c). To interpret the magnitudes of our estimates, we calculate a persuasion rate following DellaVigna and Gentzkow (2010). The persuasion rate can be approximated as $\theta \cdot t/e(1 - y)$, where θ is the 2SLS estimate, y is the average Republican vote share, e is the average exposure of American adults to Twitter, and t is the share of adults that turn out to vote. Using the estimate for θ from

columns (5) and (10) of Table 2, the persuasion rate is 8.6% and 9.4% for 2016 and 2020, respectively. It implies that, in 2016, one out of every twelve active Twitter users that voted for Clinton would not have done so if they had not been exposed to the platform.³⁰

Note also that we estimate county-level effects and persuasion rates, which may capture local spillovers and social interaction effects. In other words, they implicitly assume that one additional Twitter user can only affect one person's vote. However, if this additional user also exposes other non-users to content she sees on Twitter (e.g., via in-person conversations), then an "individual-level" persuasion rate would be smaller. Guriev, Melnikov, and Zhuravskaya (2020) show that persuasion rate formulas scale by a factor 1/N where N is the number of voters in a county "exposed" given each additional Twitter user. For example, if each new user has a 30% chance of exposing another voter in the county to the persuasive content she sees on Twitter, our county-level persuasion rates must be divided by 1.3 to be interpreted as individual-level persuasion rates.

The persuasion rates of Twitter are smaller than the persuasion rates of 3G internet on voting for establishment parties, which Guriev, Melnikov, and Zhuravskaya (2020) estimate to be 27%. It is also smaller than the estimated pro-Republican persuasion rate of Fox News, which Martin and Yurukoglu (2017) estimate to range between 27% and 58% (depending on the year).³¹ As a further comparison, Gentzkow, Shapiro, and Sinkinson (2011) estimate a persuasion rate of 12.9% for the effect of reading a local newspaper in the 1869–1928 period on voter turnout. All these persuasion rates are not on the "individual level," meaning they are based on effects estimated at a regional or county level. However, Twitter's persuasion rates are also somewhat smaller than others from studies that estimate effects with individual-level randomization: Gerber, Karlan, and Bergan (2009) report a 19.5% pro-Democrat persuasion rate for reading the Washington Post and DellaVigna and Gentzkow (2010) survey randomized getout-the-vote canvassing experiments and calculate persuasion rates in the 10% –15% range.

Robustness to Controls and Coefficient Stability. Although the estimated effects are always negative and significant at the 1% level, comparing columns shows that the effect sizes vary somewhat with the inclusion of controls, especially demographic and socioeconomic ones. This sensitivity to controls is perhaps expected given their explanatory power over vote shares. The results in other papers that explore effects on

^{30.} The persuasion rate re-scales effect sizes by how many individuals are exposed to the platform and how many are not already persuaded. For marginal changes in exposure, the formula is $f = \frac{dy}{de} \cdot \frac{t}{1-y}$ (DellaVigna and Gentzkow 2010). Since our estimate θ is the semi-elasticity $\frac{dy}{de} \cdot e$, we obtain $f = \theta \cdot \frac{t}{e(1-y)}$. In 2016, y = 0.46, t = 0.55, while y = 0.47 and t = 0.62 in 2020. We assume e = 0.25 for both periods. This implicitly assumes that Twitter usage among voters is the same as the overall population. If voters are over-represented among Twitter users, the persuasion rate would be smaller. On one hand, Twitter users are younger (which is associated with lower turnout) but more educated (which is associated with higher turnout) than the general population.

^{31.} DellaVigna and Kaplan (2007) estimate a smaller persuasion rate of 11.6% for Fox News.

vote shares in similar frameworks, such as DellaVigna and Kaplan (2007), Martin and Yurukoglu (2017), and Autor et al. (2020), show a similar sensitivity to controls.

To further probe the sensitivity of our results, we provide three separate pieces of evidence. First, we apply two approaches, from Altonji, Elder, and Taber (2005) and Oster (2019), designed to gauge the potential importance of unobservable variables in driving the results. We compare our reduced form (Panel B) specifications with all controls (columns (5) and (10)) to those with the fewest controls (columns (1) and (6)). We obtain an "Oster- δ " of approximately 7 for both 2016 and 2020. This suggests, in order for the true effect to be zero, unobservable variables would have to be seven times as "important," in terms of driving selection into treatment and explaining the outcome, as the (numerous) controls added in columns (5) and (10). We also obtain an "Altonji et al- δ " of 1.7 for 2016 and 1.6 for 2020, which have a similar interpretation as the "Oster- δ ;" note that these values are smaller because they do not take into account how much of the variance in the treatment (in our case, Twitter usage) is explained by the controls.³²

Both approaches indicate that our results are robust when compared to usual cutoffs and to other papers published in economics journals. Both Altonji, Elder, and Taber (2005) and Oster (2019) argue that δ s above one are an appropriate "robustness standard." Moreover, Oster (2019) calculates the "Altonji et al- δ " for 45 results published in the *American Economic Review*, *Quarterly Journal of Economics*, and *Journal of Political Economy*. Only 20% of these results have a "Altonji et al- δ " above the ones we find in our context. Moreover, Oster (2019) reports only 9.1% of these results survive the cutoff of an "Oster- δ " above one.

Second, Online Appendix Table C.1 provides further evidence for the robustness of our findings to the inclusion of different controls. It allows for more flexible interactions between observable characteristics by using partialling-out LASSO variable selection (Chernozhukov, Hansen, and Spindler 2015a, 2015b).³³ More specifically, we estimate regressions in which we allow for state fixed effects and all possible interactions of the control variables (i.e., up to 654 controls). The LASSO then selects controls that either predict the vote shares, SXSW participation, or Twitter usage. All estimates are statistically significant and of similar magnitude as our baseline estimates.

Third, Online Appendix Figure D.3(a) plots the estimated θ of our 2SLS equation (3) while flexibly allowing the included control variables to vary. The

^{32.} The R^2 of regressions with the fewest controls (columns (1) and (6)) is approximately 0.6, while it is 0.94 for the regressions with the most controls (columns (5) and (10)). Thus, observable controls explain a large part of the variation in vote shares but only generate a modest change in coefficient sizes. Intuitively, this is what generates large δ s. The difference between the calculated "Altonji et al- δ " and "Oster- δ " is that the former only considers the coefficients and R^2 of the two specifications being compared, while the latter takes into account how much of the variance in treatment can be explained by observed controls. Note that the "Oster- δ " is derived under more general conditions than the "Altonji et al- δ ." In all calculations, we set R^{max} (the hypothetical R^2 that one would obtain if all relevant unobserved controls were included) to 1. This is the most conservative assumption possible.

^{33.} Note that, unlike most other regressions reported in the paper, these regressions are unweighted.

resulting "specification curves" suggest that our results (for both the 2016 and 2020 elections) are robust to how our regressions are specified. The estimated coefficients are always negative, almost always statistically significant at the 5% level, and in the overwhelming number of specifications considerably more negative than our "baseline estimates," which is marked by the vertical line.

Placebo Test: Interest in SXSW Before March 2007. The coefficients on $SXSW_C^{Pre}$ in Table 2 are statistically insignificant and substantially smaller than those on $SXSW_c^{March 2007}$. As discussed in Section 5, this provides support for our identification condition (exclusion restriction). Suppose that our instrument merely captured that counties with an interest in SXSW's Twitter account during the platform's early years also differ in (unobservable) characteristics that predict the 2016 election outcome. If this was the case, the coefficients on $SXSW_C^{Pre}$ should be similar in size to those on $SXSW_c^{March 2007}$. Intuitively, we have two variables that are correlated with interest in the SXSW festival, but only one predicts Twitter users in later years, allowing us to disentangle interest in the festival (and its correlates) from its effect via more Twitter users.

Placebo Test: Other Festivals in 2007. We also provide an additional placebo check by investigating five other festivals that are similar in nature to SXSW (Burning Man, Coachella, Pitchfork, Austin City Limits, and the Electric Daisy Carnival).³⁴ One of these placebo festivals, Austin City Limits, takes place in the same city as SXSW. For this exercise, we construct analogous measures to our instrument for other festivals using the same exact procedure: the Twitter followers of the respective festival that joined in the month the festival took place in 2007. In Table 3, we then show that none of these festival variables, including that for Austin City Limits, has predictive power for future Twitter usage. More importantly, the variables are also essentially uncorrelated with 2016 and 2020 presidential vote shares. These results provide further support for our identifying assumption and suggest that a spurious relationship between counties' Twitter users interest in festivals during 2007 and future election outcomes cannot explain our findings.

Further Robustness and Additional Tests. The Online Appendix presents a number of additional sensitivity checks. In Online Appendix Table C.2, we consider changes to the baseline regression specification. In particular, we allow for unweighted regressions; weighting by the relevant election-year turnout (2016 or 2020); alternative functional forms of the pre-SXSW user variable; restrict the sample to the sub-sample of counties where we observe either SXSW followers who joined in March 2007 or the pre-period; allow for flexible spatial correlation in the standard errors; and per capita measures of Twitter usage. Further, Online Appendix Table C.4 replaces our baseline

^{34.} These festivals take place in the Black Rock Desert (NV), Indio (CA), Chicago (IL), Austin (TX), and Las Vegas (NV), respectively.

	Twi	tter Followers of	^e Festival join	ning in Festi	ival Monti	'n
	SXSW (1)	Burning Man (2)	Coachella (3)	Pitchfork (4)	EDC (5)	ACL (6)
Panel A: First Stage (Dep	endent varia	ıble: Twitter Us	sage)			
Followers Festival Month	0.167*** (0.015)	-0.005 (0.012)	0.011 (0.012)	-0.006 (0.006)	0.010 (0.008)	-0.009 (0.018)
Panel B: Reduced form (I	Dependent va	ariable: Republ	lican Vote S	hare 2016)		
Followers Festival Month	-0.004*** (0.001)	0.001 (0.001)	-0.000 (0.001)	-0.000 (0.000)	0.000 (0.001)	-0.000 (0.001)
Panel C: Reduced form (Dependent v	ariable: Repub	lican Vote S	hare 2020)		
Followers Festival Month	-0.003** (0.001)	0.000 (0.001)	-0.001 (0.001)	-0.000 (0.000)	0.000 (0.001)	-0.000 (0.001)
Census region FE Population controls Geographical controls Demographic controls Socioeconomic controls China shock controls 1996 election control Observations	Yes Yes Yes Yes Yes Yes 3 064	Yes Yes Yes Yes Yes Yes 3 064	Yes Yes Yes Yes Yes Yes 3 064	Yes Yes Yes Yes Yes Yes 3 064	Yes Yes Yes Yes Yes Yes Yes 3 064	Yes Yes Yes Yes Yes Yes Yes 3 064

TABLE 3. Placebo tests for other festivals.

Notes: This table presents the first stage and reduced from estimates using equation (3). The dependent variable is either Twitter usage (Panel A) or the vote share of the Republican party in the 2016 and 2020 presidential elections (Panels B and C). The independent variables are number of followers of different festivals that joined in the respective festival month (in logs with 1 added inside). To make the coefficients comparable, we standardized these variables to have mean 0 and standard deviation 1; this means the results for SXSW are not directly comparable in magnitude to those in the main text. Similarly to our baseline specification, all regressions control for the number of festival followers that joined Twitter before the festival. As the festivals take place after the SXSW festival, we additionally control for the number of SXSW followers from March 2007. All regressions are weighted by turnout in the 2000 presidential election. All regressions include the controls from columns (5) and (10) in Table 2. EDC is the Electric Daisy Carnival and ACL is Austin City Limits. Standard errors in parentheses are clustered by state. *** p < 0.01, ** p < 0.05, * p < 0.1.

time-invariant measure of Twitter usage (making use of all available user data) with a time-varying measure based on how many users were using the platform in a given year. None of these adjustments make a substantial difference in the magnitudes or statistical significance of the estimates.

Online Appendix Table D.1 reports results for additional outcome variables, all of which support the idea that Twitter exerts a pro-Democrat effect. In column (1), we use a probit IV model to investigate the likelihood of a county switching from Obama in 2008 to Trump in 2016. The coefficient suggests that a one standard deviation increase in Twitter usage is associated with a -24% lower probability of a county

		Dep. v	ar.: Reput	blican vot	e share in	
	2000 (1)	2004 (2)	2008 (3)	2012 (4)	2016 (5)	2020 (6)
Panel A: Reduced form						
Log(SXSW followers, March 2007)	-0.003	-0.005	-0.006	-0.003	-0.011***	-0.011**
	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)
Log(SXSW followers, Pre)	0.001	0.001	-0.000	-0.002	0.001	-0.001
	(0.004)	(0.003)	(0.006)	(0.005)	(0.004)	(0.005)
Panel B: 2SLS						
Log(Twitter users)	-0.005	-0.009	-0.011	-0.007	-0.021***	-0.020**
	(0.004)	(0.006)	(0.009)	(0.008)	(0.008)	(0.009)
Log(SXSW followers, Pre)	0.001	0.001	0.000	-0.001	0.002	0.000
	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)
Observations	3,064	3,064	3,064	3,064	3,064	3,064
Mean of DV	0.48	0.51	0.46	0.47	0.46	0.47
Robust F-stat.	121.18	121.18	121.18	121.18	121.18	121.18

TABLE 4. Twitter and the Republican vote share, 2000–2020.

Notes: This table presents county-level regressions where the dependent variable is the Republican vote share in presidential elections. *Log(SXSW followers, March 2007)* is the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow South by Southwest (SXSW). *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006, defined similarly. *Twitter users* are the number of users in 2014-2015. All regressions control for population deciles, Census region fixed effects, and the full set of controls (as in column (5) and (10) of Table 2). The first-stage regression for 2SLS results (Panel B) are presented in column (5) of Table 1, with the F-stat for the excluded instrument in the bottom row. On Panel (A), the coefficients on *Log(SXSW followers, March 2007)* for 2000, 2004, 2008, and 2012 are jointly statistically insignificant (*p*-value=0.329). Further, the average effect in 2016 and 2020 is statistically distinct from the average effect in 2000, 2004, 2008, and 2012 (*p*-value=0.002). Observations are weighted by turnout in the 2000 presidential election. Standard errors in parentheses are clustered by state. *** *p* < 0.01, ** *p* < 0.05, * *p* < 0.1.

electing Obama in 2008 and Trump in 2016. Columns (2) and (3) look at changes in campaign donations to Democrats and Republicans between 2000 and 2016. We find a positive and statistically significant effect for donations to Democrats, and no effect for Republicans. Lastly, columns (4) and (5) look at approval ratings for President Trump in 2017 based on data from the Gallup Daily Poll. We find that exposure to social media is associated with a decrease in Trump's popularity, and more so among Republicans.

The 2007 SXSW Shock and Previous Election Outcomes. Table 4 repeats the analysis from column (5) of Table 2 using Republican vote share in the previous presidential elections of 2000, 2004, 2008, and 2012 as the dependent variable.³⁵ All estimates for

^{35.} Note that the right-hand side of the regressions remains the same as in the previously reported regressions. That is, the instrument is the same, and the "endogenous variable" in the 2SLS is Twitter users measured in 2014–2015. We explore the role of time-varying Twitter user measures on Online Appendix Table C.4.

those years are substantially smaller than the ones for 2016 and 2020 and statistically insignificant. For 2000, 2004, and 2008, this can be interpreted as a placebo or pretrends test: conditional on the covariates, our instrument is uncorrelated with outcomes before Twitter's launch (2000 and 2004) and when the platform had few users (2008). In 2012, Twitter already had a substantial user base, so our estimates can be interpreted as a genuine "zero effect"; we return to a comparison of the 2020, 2016, and 2012 results in Section 6.3. The estimates for 2000, 2004, 2008, and 2012 are also jointly statistically insignificant (p-value=0.329). The average effect in 2016 and 2020 is statistically different from the average effect in 2000, 2004, 2008, and 2012 (p-value=0.002).

In Online Appendix Figure C.1, we further show the reduced form estimates for presidential election going back as far as 1924. We find that our instrument is also uncorrelated with any of the earlier presidential election results. As discussed in Section 5, this result lends additional support for our exclusion restriction. If our instrument merely captured uncontrolled differences across counties, these should also correlate with vote shares in previous elections.

While these findings make it unlikely that our instrument is correlated with pro-Democratic attitudes at the county level, a possible concern could be that we are picking up "anti-populist" attitudes, which could have harmed Trump's electoral results. To address this concern, we turn to the historical case study of Ross Perot's political campaign in 1992 and 1996. Perot, a billionaire businessman, also ran as a "third-party candidate" on a populist platform. However, when we replace the dependent variable with the third-party vote share in the 1992 and 1996 presidential election (see Online Appendix Table C.5), we find no evidence that our instrument is associated with lower vote shares for Ross Perot. This makes it unlikely, we are capturing differences in "demand for populism" across counties.³⁶

Effects on Vote Share Changes. We also consider specifications of equations (2) and (3) using vote share *changes* instead of *levels* as the dependent variable. All our estimates based on changes take differences relative to the base year 2000 (akin to the approach in Autor et al. (2020)) and use the full set of controls (as in columns (5) and (10) of Table 2). Figure 6(a) plots the reduced form estimates for changes in the Republican vote share in presidential elections.

The results corroborate the previously presented evidence based on specifications in levels. Our instrument is essentially uncorrelated not only with *levels* but also with *changes* (or *trends*) in election outcomes during the 2000–2012 period. Given our arguments above, this also lends support for our identification strategy. Again the estimates for 2000, 2004, 2008, and 2012 are also jointly indistinguishable from 0 (*p*-value=0.398). The reduced form effects for 2016 and 2020 are similar to the one estimated using levels. For example, the estimated effect (θ) for 2016 using *changes* is

^{36.} In Appendix Table C.6, we also investigate the vote shares in the 2020 democratic primaries. Here, we find a positive association between Twitter exposure and the vote share of Bernie Sanders, often described as a left-wing populist. This further speaks against the hypothesis of "anti-populist" sentiment.



(a) Changes in Republican Vote Share

(b) Change in Voter Turnout

FIGURE 6. Twitter and presidential elections—Reduced form. These figures plot reduced form estimates $\hat{\beta}'$ from county-level regressions as in equation (2). They measure the effect of Log(1 + SXSW followers, March 2007), while controlling for Log(1 + SXSW followers, Pre), on changes in the Republican vote share in presidential elections relative to the year 2000 in Panel (a), and changes in the ratio of voter turnout to voting-age population relative to 2000 in Panel (b). All regressions control for population deciles and Census region fixed effects, and the full set of controls (as in columns (5) and (10) of Table 2). On Panel (a), the estimates for 2004, 2008, and 2012 are jointly statistically insignificant (*p*-value=0.398). Further, the average effect in 2016 and 2020 is statistically distinct from the average effect for 2004, 2008, and 2012 (*p*-value=0.001). Regressions are weighted by turnout in the 2000 presidential election. Whiskers represent 95% confidence intervals based on standard errors clustered by state.

-0.017, similar to the one estimated in levels (-0.021).³⁷ Unsurprisingly, the average effect in 2016 and 2020 is statistically different from the average effect in 2000, 2004, 2008, and 2012 (*p*-value=0.001).

Effects on Turnout and Congressional Elections. Figure 6(b), Figure 7(a), and Figure 7(b) replicate the exercise in Figure 6(a) using voter turnout and vote shares in House and Senate elections as the outcomes. We do not find a statistically significant association between our instrument and election turnout except for 2020. Before 2020, the estimated point effects are usually small. For example, the upper bound on the 95% confidence interval of the 2SLS estimate for the effect of turnout in the 2016 election implies that a 10% increase in Twitter users raises turnout by 0.036 p.p. (Online Appendix Table D.2).

In the 2020 election, which saw the highest turnout rate in more than a century (NPR 2020c), we find that Twitter is associated with a larger fraction of the voting-age population casting their ballot. Why did Twitter have an effect on 2020 turnout but not in the previous election? One possible explanation could be that calls to turn out, and vote were widespread on the platform, partially because of an initiative by Twitter itself that was not present in 2016 (Twitter 2020). Further, the 2020 election was unique in its prevalence of mail and early voting because of the COVID-19 pandemic. Another

^{37.} Online Appendix Table C.3 present the OLS and 2SLS estimates.



FIGURE 7. Twitter and congressional election results—Reduced form. These figures plot reduced form estimates $\hat{\beta}'$ from county-level regressions as in equation (2). They measure the reduced form effect of Log(1 + SXSW followers, March 2007), while controlling for Log(1 + SXSW followers, Pre), on the Republican vote share in House and Senate elections since 2000. For House elections in Panel (a), the dependent variable is the change in the Republican vote share since 2000. For Senate elections in Panel (b), the dependent variable is the change in the Republican vote share from 6, 12, or 18 years ago (to accommodate senators' 6-year terms). All regressions control for population deciles and Census region fixed effects and the full set of controls (as in columns (5) and (10) of Table 2). Regressions are weighted by turnout in the 2000 presidential election. The whiskers represent 95% confidence intervals based on standard errors clustered by state.

possibility is that Twitter served as a platform to convey information on how to vote by mail or before election day.

The coefficients for House and Senate races are more noisily estimated, particularly for the smaller sample of Senate races (where only a third of seats is renewed every 2 years). Overall, there is little evidence suggesting an effect of Twitter on congressional elections, including in 2016, the 2018 midterm election, and 2020. Finding an effect on presidential vote shares but not in these "down-ballot" races is perhaps expected since content on presidential candidates (and in particular on Trump in 2016 and 2020) is more common on Twitter than content on congressional races.³⁸

^{38.} Online Appendix Table D.3 presents the reduced form estimates for House and Senate races. To accommodate the Senate's 6-year terms, we take changes relative to 2000, 1998, and 1996, instead of always using 2000 (as we do for other outcomes).

Discussion of Identification Condition. As discussed in more detail in Section 5, there are five pieces of evidence supporting our identification condition: (i) our empirical strategy compares relatively similar counties (Section 5); (ii) the placebo test based on the coefficient on $SXSW_c^{Pre}$ and other festivals; (iii) the instrument being uncorrelated with election outcomes in the 1924–2012 period; and (iv) absence of a statistically significant effect of Twitter on House and Senate races; while at the same time (v) the instrument being correlated with vote choices of moderate voters in particular.

Given this, a violation of the identification condition would require an omitted variable that correlates with the instrument, Twitter usage, and Trump's vote share in 2016 and 2020 but is uncorrelated with: (i) $SXSW_c^{Pre}$ and other festivals, (ii) levels and trends in election results before Twitter's launch and rise to popularity, (iii) the observable variables presented in Online Appendix Table A.4, and (iv) election results in congressional elections. At the same time, such omitted variable would also (v) be more strongly correlated with the vote choices of moderates and independents than "partisans." Our argument is that the existence of such omitted variable is unlikely.

Toward an Average Treatment Effect. As with any instrument, our 2SLS results identify a LATE. In our setting, the "compliers" are counties with higher Twitter usage as a result of the inflow of SXSW attendees. While the negative treatment effect of Twitter usage for these counties is in itself an interesting finding, the ATE for the United States overall—and therefore the overall impact of Twitter on elections—may differ. However, two pieces of evidence suggest that our estimates, despite this concern, allow us to infer information about the ATE.

A first indication comes from comparing the OLS and 2SLS results in Table 2. Both estimates are always negative and relatively similar in magnitude. Second, we build on the approach suggested by Andrews and Oster (2019) and more formally investigate the external validity bias of our estimates. For experimental settings, Andrews and Oster (2019) suggest using the observable heterogeneity in estimated treatment effects within the experimental sample to learn about the ATE in the overall population. Similarly, we can use the heterogeneity of treatment effects within counties that provide the variation that identify our results to approximate the ATE for the United States as a whole. Using all included control variables from our main specification for the prediction of heterogeneity of the treatment effect, the Andrews and Oster (2019) approach suggests that the ATE should, if anything, be *larger* than the LATE, we estimate in our baseline results. This seems plausible as the more urban counties for which we have variation in our instrument tend to be Democratic strongholds, and thus likely have fewer independents, for which we find the largest persuasion effects (in survey data). We provide additional details on our approach in Online Appendix E³⁹

^{39.} As our setting differs from the one discussed in Andrews and Oster (2019), some adjustments to our baseline estimation were required. We estimate the treatment effect exclusively in the subset of counties for which either $SXSW_c^{March 2007}$ or $SXSW_c^{Pre}$ are not equal to 0. Then, we define a treatment indicator variable equal to 1 for counties with SXSW followers who joined in March 2007.

6.2. Effects Are Concentrated Among Moderate Voters

If social media indeed matters for election outcomes, we would expect there to be heterogeneous effects across groups of voters. In particular, Bayesian updating suggests that voters who do not hold strong priors about a particular party should be more likely to be persuaded. We test this prediction using individuals' voting decisions from the 2016 and 2020 CCES. In particular, we estimate the following instrumental variable Probit regression:

$$y_{ic} = \varphi + \theta \cdot Twitter \, users_c + \pi \cdot SXSW_c^{Pre} + \mathbf{X}_{ic}\rho + \varepsilon_{ic}, \tag{4}$$

where y_{ic} is an indicator variable equal to 1 if an individual *i* living in county *c* voted for Trump in the 2016 or 2020 election and 0 for Clinton.⁴⁰ The definition of the countylevel variables *Twitter usage*_c and *SXSW*_c^{Pre} remains unchanged. **X**_{ic} is now a vector of individual-level control variables including age, gender, race, family income, and education. We again instrument for county-level Twitter usage based on the SXSW followers who joined in March 2007.

Table 5 presents results from estimating equation (4). In Panel A and B, Column (1) suggests county-level Twitter usage has a statistically significant negative effect on the likelihood to vote for Trump. The marginal effect implies that a 10% increase in the number of Twitter users in a county would lower Trump's vote share by 0.49 p.p. in the 2016 and 0.46 p.p. in the 2020 presidential election.⁴¹

Columns (2)–(6) report results estimated separately by voters' reported party affiliation. The effect is strongest for voters who identify as independents, and thus likely to not hold strong priors. The results suggest weaker or zero effects for those with stronger political views, whether Republican or Democrat.

Indeed, we find larger marginal effects for Log(Twitter users) among younger voters. For independents, the estimate for young voters is 20% larger than for older voters (-0.071 compared to -0.059). Among moderate Republicans and Democrats, the estimated coefficients are close to 0 for voters aged 50+ but sizeable for younger voters (although they are not statistically significant at conventional levels). Because young voters are less likely to vote for Trump, this implies larger elasticities of vote outcomes with respect to Twitter for those below 50 relative to the baseline probabilities.

A potential concern with this exercise is that party affiliation may itself be affected by Twitter usage. We thus present further support for Twitter having persuasion effects on moderates using county-level data that is not subject to such concerns. In particular, we estimate our county-level specification (equation 3) for the 2016 elections, but

^{40.} Note that we use data on all CCES respondents, not only those in the 2016 wave (i.e., respondents from the 2018 and 2020 waves were asked how they voted in 2016). The results are similar if we only use the 2016 wave. In unreported regressions, we do not find an effect on votes for Jill Stein.

^{41.} This effect size is within the range of the county-level estimates presented in Table 2. Appendix Table C.7 shows this baseline result is robust to using only individuals with validated turnout and/or who stated that they originally intended to vote for Trump in the pre-election wave of the CCES.

		Dependent	variable: V	oted for Trum	p in	
	Full Sample (1)	Strong Dem. (2)	Mod. Dem. (3)	Indep. (4)	Mod. Rep. (5)	Strong Rep. (6)
Panel A: 2016 Election	n					
Log(Twitter users)	-0.135***	0.018	-0.079	-0.209***	-0.045	-0.022
	(0.045)	(0.054)	(0.067)	(0.058)	(0.043)	(0.037)
Observations	146,579	44,241	30,745	13,625	27,865	28,559
Mean of DV	0.492	0.026	0.111	0.611	0.919	0.980
Marginal effect	[-0.049]	[0.001]	[-0.014]	[-0.073]	[-0.006]	[-0.001]
Panel B: 2020 Electio	n					
Log(Twitter users)	-0.130***	-0.069	-0.002	-0.186***	0.000	-0.080
	(0.042)	(0.075)	(0.120)	(0.063)	(0.041)	(0.079)
Observations	43,617	13,108	9,670	4,430	7,388	8,395
Mean of DV	0.475	0.012	0.074	0.516	0.908	0.978
Marginal effect	[-0.046]	[-0.002]	[0.000]	[-0.069]	[0.000]	[-0.004]

TABLE 5. Twitter and individuals' vote decisions in 2016/2020.

Notes: This table presents results estimated using IV probit models, as in equation (4). The dependent variable is a dummy for individuals in the CCES who voted for Trump in 2016 or 2020. *Log(Twitter users)* is instrumented using the (log) number of SXSW followers that joined Twitter in March 2007. All regressions control for the (log) number of SXSW followers that joined Twitter at some point in 2006, family income, gender, education levels, marital status, news interest, and age, as well as county-level population deciles and Census region fixed effects. Regressions are weighted by survey weights. Standard errors in parentheses are clustered by state. *** p < 0.01, ** p < 0.05, *p < 0.10.

split counties based on how consistently they voted for either the Republican or the Democratic party. Specifically, we define "swing counties" as counties that were not consistently won by one party in the presidential elections from 2000 to 2012. Because we find no effect of Twitter on vote shares before 2016 (see Table 4), splitting the sample by swing counties is not subject to potential concerns that this variable might itself be affected by social media.

Online Appendix Table D.4 shows the results. For the 2016 presidential elections, we find that Twitter usage only negatively impacts the Republican vote share in "swing" counties. We find no evidence for Republican or Democratic strongholds, where people likely have the strongest priors. The patterns are similar in 2020. But here, we also find a small effect on counties that usually vote Republican, which could suggest that the effect on moderate Republicans we find in the CCES also apply to the county level.

6.3. Potential Mechanisms

The findings above suggest that Twitter had an effect in 2016 and 2020, but not during previous presidential elections. We address three potential explanations for this pattern:

lack of familiarity with social media (a *learning channel*), changes in social media's "slant" (a *content channel*), and Trump's role as an outsider candidate (a *political shock channel*).

The first factor could be the reach of and familiarity with social media content. In 2008, social media was a relatively new type of technology. Only a quarter of American adults used *any* social media platform and only 10% of internet users posted political commentary on social media (Pew Research Center 2009, 2011). Online Appendix Figure A.1(a) shows that Twitter, which was founded in 2006, only had around one million users during the 2008 elections, compared to 40 million in 2012, 67 million in 2016, and 69 million in 2020 (Statista 2019, 2020). Twitter's limited reach and novelty might have initially restricted its impact on voters.

The second possible explanation is that social media's content changed between 2008 and 2016. It is conceivable that, similar to changes in the slant of cable news (e.g., Martin and Yurukoglu 2017), the content to which Twitter users are exposed has become more left-leaning over time.

A third reason is that Trump's political rise constituted a considerable shock to the U.S. political system. In this view, Twitter may not have partian effects *per se*. Instead, the platform may have served as a conduit for spreading sentiments or information about Trump, either because of his own prominent behavior on the platform or because of other users' content about him.

Two pieces of evidence presented above are consistent with the political shock channel. First, if Twitter had little effect before 2016 because it was not widely used, its effect on vote shares should systematically increase over time. We do not find evidence supporting this idea in effects for presidential, House, and Senate elections (Figures 6 and 7). Instead, we find a discontinuous negative effect in the 2016 presidential election that persists in 2020. Second, we do not find significant effects for the 2016, 2018, and 2020 House and Senate elections. This implies that Twitter usage lowered Trump's vote share without significantly affecting other Republican candidates *on the same election day*.

Results from the 2016 Republican Primaries. We provide additional evidence for a Trump-specific effect of Twitter exposure by investigating the 2016 county-level Republican primaries results. The primaries allow us to focus on the favorability of different candidates among Republican voters. The results from this exercise are presented in Table 6. We find that Twitter usage is associated with a lower vote share for Trump. We also find a positive effect on the vote share of John Kasich, the most moderate of the major Republican candidates.⁴²

Results from Gallup Approval Ratings. A similar pattern emerges when we use data from the Gallup Daily Tracker, which contains approval ratings for three other

^{42.} In Appendix Table C.6, we also investigate the voting behavior in the 2016 and 2020 Democratic primaries. Twitter usage appears to be associated with a higher support for Bernie Sanders in 2020.

	Dependent	variable: Vo	ote share in	Republican I	Primary of
	Trump (1)	Cruz (2)	Rubio (3)	Bush (4)	Kasich (5)
Panel A: Reduced form					
Log(SXSW followers, March 2007)	-0.030** (0.012)	0.004 (0.009)	0.016 (0.010)	-0.002 (0.001)	0.017** (0.008)
Panel B: 2SLS					
Log(Twitter users)	-0.044** (0.017)	0.005 (0.013)	0.024 (0.015)	-0.003 (0.002)	0.025** (0.011)
Observations Mean of DV Robust F-stat.	2,831 0.48 69.54	2,831 0.23 69.54	2,831 0.09 69.54	2,831 0.01 69.54	2,831 0.15 69.54

TABLE 6. Twitter and vote shares in the 2016 Republican primaries.

Notes: This table presents county-level regressions where the dependent variable is the vote share of the indicated candidate in the Republican party primaries in 2016. Log(SXSW followers, March 2007) is the number of Twitter users (in logs, with 1 added inside) who joined in March 2007 and follow SXSW. *SXSW followers, Pre* is the number of SXSW followers who registered at some point in 2006, defined similarly. *Twitter users* are the number of users in 2014–2015. All regressions control for population deciles, Census region fixed effects, and the full set of controls (as in columns (5) and (10) of Table 2). The first-stage regressions for 2SLS results (Panel B) are analogous to the one presented in Table 1, except for the different sample of counties for which primary results are available. The F-stat for the excluded instrument is provided in the bottom row. Observations are weighted by turnout in the 2000 presidential election. Standard errors in parentheses are clustered by state. *** p < 0.01, ** p < 0.05, * p < 0.1.

Republican presidential candidates who ran alongside Trump during the primaries (Ted Cruz, Marco Rubio, and Kasich). Table 7 shows the results of running individual-level IV probit regressions as in equation (4), where the dependent variable is an indicator variable equal to 1 if the respondent approved of a specific candidate. As in Table 5, we differentiate between respondents' political affiliation.⁴³

Table 7 confirms our main county-level result from general elections and primaries: Twitter usage is associated with the lower approval of Trump, especially among independents (who are presumably more likely to be persuaded by social media content). We also find lower approval of Cruz, who is substantially more right-wing than other presidential primary candidates in recent years (FiveThirtyEight 2015). We find no link between Twitter use and approval of the more moderate Republican candidates, Rubio and Kasich. For the Democrat candidates, we find an effect for Clinton's but not Sanders' approval.⁴⁴

^{43.} We pool people who identify as leaning Republicans or Democrats with independents, because—in contrast to the CCES data—only a few individuals in the survey are classified as "leaners."

^{44.} Note that Online Appendix Table C.6 reports a significant positive effect for Sanders county-level primary vote share in 2020, but not in 2016. As mentioned earlier, Gallup Daily Tracker data are not

	Depe	ndent variable	e: Approved	of candidate	during prime	ıries
	Trump (1)	Cruz (2)	Rubio (3)	Kasich (4)	Sanders (5)	Clinton (6)
Panel A: Republica	ans					
Log(Twitter users)	-0.108^{***}	-0.086^{**}	-0.051	0.018	0.031	0.148***
Observations Mean of DV	19,974 0.647	11,959	8,344 0.779	8,995 0.665	16,099 0.238	20,983
Marginal effect	[-0.038]	[-0.029]	[-0.014]	[0.006]	[0.009]	[0.022]
Panel B: Independ	ents and Lean	ers				
Log(Twitter users)	-0.065^{**} (0.028)	-0.006 (0.035)	-0.015 (0.043)	0.050	0.059 (0.043)	0.154***
Observations Mean of DV	22,852 0.329	12,135 0.392	8,080 0.516	8,280 0.581	17,356 0.595	23,813 0.380
Marginal effect	[-0.021]	[-0.002]	[-0.006]	[0.019]	[0.021]	[0.054]
Panel C: Democrat	ts					
Log(Twitter users)	-0.052 (0.051)	-0.116^{**} (0.054)	-0.036 (0.056)	0.076 (0.051)	0.004 (0.050)	0.081** (0.038)
Observations Mean of DV Marginal effect	20,866 0.107 [-0.009]	11,098 0.195 [-0.030]	7,460 0.271 [-0.012]	7,547 0.502 [0.029]	16,059 0.808 [0.001]	21,454 0.807 [0.021]

TABLE 7. Twitter and candidate approval during the 2016 primaries.

Note: This table presents results estimated using IV probit models, as in equation (4). The dependent variable is a dummy for individuals who approved the respective presidential candidate during the presidential primaries in 2015 and 2016. We restrict the sample to the period before Trump became the presumptive nominee in June 2016. *Log(Twitter users)* is instrumented using the number of SXSW followers that joined Twitter in March 2007. All regressions control for the (log) number of SXSW followers that joined Twitter at some point in 2006, income, gender, education, and marital status, as well as county-level population deciles and Census region fixed effects. Regressions are weighted by survey weights. Standard errors in parentheses are clustered by state. ***p < 0.01, **p < 0.05, *p < 0.10.

Taken together, these results are consistent with Twitter turning voters against voting for Trump in particular and not against the Republican party more generally. Our results may also explain the absence of an effect in the 2008 and 2012 elections: Obama's opponents John McCain and Mitt Romney were widely considered to be moderate Republicans (e.g., more similar to Kasich than Trump or Cruz).⁴⁵

available for 2020 at the time of writing. Note also that Gallup collects data on candidate approval (as opposed to vote choice) for the entire population (as opposed to Democratic primaries votes), which may explain some differences between these results and the ones based on county-level primaries results.

^{45.} As previously discussed, the number of Twitter users in 2008 was relatively small, but by 2012, it was relatively close to its 2016 level (Online Appendix Figure A.1a).

Slant of Election-Related Tweets. We provide further support for the hypothesis that Trump's 2016 campaign and presidency triggered opposition on Twitter by analyzing the content of more than 460 million tweets mentioning the last name of presidential candidates during the 2012, 2016, and 2020 presidential campaigns.

First, we classify the tweets' slant as Republican, Democrat, or neutral using two approaches described in Section 3. In the first case, we classify the political affiliation of Twitter users by counting the number of Democrat and Republican Congress members they follow. If a user follows more Democrats than Republicans, they are classified as Democrat, and vice-versa. Tweets sent by a user classified as Democrat are classified as Democrat, and so forth.⁴⁶ In the second case, we classify individual tweets (not users) following an approach in the spirit of Gentzkow and Shapiro (2010) and train a L2 regularized logistic regression classifier to predict whether a tweet is more likely to have a Democrat or Republican slant, depending on its content's similarity to tweets sent by Congress members. If a tweet's content has higher similarity with those of Democratic Congress members, it is classified as Democratic, and as Republican otherwise.⁴⁷

Figure 8 plots the amount of Twitter attention directed at the Republican and Democratic presidential candidates in the 2012, 2016, and 2020 elections, as well as the tweets' estimated slant based on users' following of Congress members. To account for the attention and popularity of tweets, we base the graphs on the number of "likes" the tweets mentioning the last name of candidates received. In Online Appendix Figure D.1, we confirm that the results are similar using the number of tweets and when we base the slant classification on the text of the tweets (Online Appendix Figure D.2).⁴⁸

Panel (a) shows the number of "likes" for tweets mentioning the Republican presidential candidate (Romney and Trump), while Panel (b) provides similar evidence for the Democrats (Obama, Clinton, and Biden). There are three noteworthy facts presented in the figure. First, there was a sizable growth in the overall volume of Twitter content mentioning presidential candidates. Second, the number of "likes" for

^{46.} Users who follow an equal number of Democrats and Republican or no Congress members are classified as neutral. This approach is similar in spirit to Barberá (2015), who uses the network of Twitter users to create a measure of ideology. Because we are only interested in a binary measure of partisan slant and not the ideological distance of users, we do not estimate the full Bayesian ideal point mode. The advantage of our simplified approach is that it is faster to compute and the resulting measure is easier to interpret.

^{47.} See Section 3 and Online Appendix A.1 in the Appendix for more details. In unreported robustness checks, we confirm that these findings are robust using different slant cutoffs for classifying Republican and Democratic tweets, such as only classifying tweets for which the class probabilities are above 75% or 90%.

^{48.} Twitter users can choose to "like" each tweet they see in their "timeline." A user can only "like" a particular tweet once. "Likes" thus provide an useful metric since they capture the popularity and attention that the content received. For example, if an account sent millions of Republican-slanted tweets about Clinton, but such account had few followers and thus few users who can "like" the message, it would not meaningfully affect measures based on "likes," but could potentially do so for measures based on number of tweets.



(a) Tweets about Republican Presidential Candidates

(b) Tweets about Democratic Presidential Candidates



FIGURE 8. Twitter's partisan slant across presidential elections. These figures present the number of "likes" received by tweets that contain the last name of the candidates in the 2012, 2016, and 2020 presidential elections, depending on whether the tweet was classified as having a Republican, Democratic, or neutral slant. We classify the slant of a tweet based on the Twitter network of the user who sent the tweet. If the user follows more Democratic than Republican Congress members, they will be classified as a Democrat, and vice versa. Users who follow an equal number of Democrats and Republican or no Congress members are classified as neutral.

tweets mentioning Trump is larger than those mentioning his opponents (the difference is fourfold in 2016 and almost threefold in 2020). Note that a "like" for a tweet mentioning a candidate can occur for tweets that are positive or negative about the candidate, so the overall size of the bars are not informative about slant or sentiments of Twitter content. Third, Figure 8 also breaks down the share of tweets mentioning the candidates by slant. The content sent by users classified as Democrats are more sizable than that from those classified as Republicans. In particular, the amount of attention (proxied via "likes") on Twitter content mentioning Trump posted by Democrats is almost twice as large as the amount posted from Republicans (for both 2016 and 2020). On the other hand, content on Biden was more likely to have a Democrat slant and content on Clinton was almost equally likely to have a Democrat or Republican slant.

This pattern of results suggests that, in 2016 and 2020, Twitter became a vehicle for spreading opinions, particularly from Democratic-slanted users, on Trump. This may, in turn, have persuaded voters with weaker priors—independents and perhaps more moderate Republicans—to vote against Trump in the presidential election. This is likely due to a combination of (i) Twitter (and other social media) users being more likely to be young, well-educated, live in dense urban areas, and support the Democratic party (see Section 2), and (ii) Trump's prominent behavior on the platform himself, which could also have had its own direct effect as moderate voters' "backlash" against the content of his tweets.

It is difficult to separate whether our results are better represented as "Presidentialrace-specific" or "Trump-specific." The 2012 election differed from 2016 and 2020 both in terms of having less Twitter content referring to it (as Online Appendix Figure D.2 indicates) and also for having a more moderate Republican candidate (Romney) that behaved differently on social media and may not have attracted as much Democrat-slanted content.

7. Conclusion

Election officials around the globe are concerned about social media's increasing influence on voting decisions (e.g. NPR 2020a). At the time of writing, there is a heated debate about whether platform providers should "moderate" election-related content in the United States (e.g. Politico 2020). Exploiting variation based on a shock to Twitter's initial rise to popularity, our paper provides some of the first empirical evidence that social media can affect election outcomes.

We find that Twitter lowered the Republican party's vote share in the 2016 and 2020 presidential elections. While this finding runs counter to a popular narrative that places social media at the heart of Trump's election win, it is consistent with a growing body of evidence showing that social media users were less, not more likely to vote for Trump in 2016 or hold polarized views (Boxell, Gentzkow, and Shapiro 2017, 2018).

We also provide support for the idea that the demographics of Twitter users may account for the platform's partisan effects. People who use Twitter are 25 p.p. more likely to identify as Democrats rather than Republicans, and Democratic politicians are more popular on Twitter than Republican ones. Our work suggests that this environment not only reflects selection of like-minded individuals, but also affects voting decisions, particularly for people with more moderate views.

References

- Acemoglu, Daron, Tarek A. Hassan, and Ahmed Tahoun (2017). "The Power of the Street: Evidence from Egypt's Arab Spring." *The Review of Financial Studies*, 31, 1–42.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow (2020). "The Welfare Effects of Social Media." *American Economic Review*, 110(3), 629–76.
- Allcott, Hunt and Matthew Gentzkow (2017). "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, 31(2), 211–36.
- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of Political Economy*, 113, 151–184.
- Andrews, Isaiah and Emily Oster (2019). "A Simple Approximation for Evaluating External Validity Bias." *Economics Letters*, 178, 58–62.
- Autor, David, David Dorn, Gordon Hanson Kaveh Majlesi et al. (2020). "Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure." *American Economic Review*, 110(10), 3139–83.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky (2018).
 "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences*, 115, 9216–9221.
- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015). "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science*, 348, 1130–1132.
- Barberá, Pablo (2014). "How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US." New York University Working paper.
- Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis*, 23, 76–91.
- Beam, Michael A., Myiah J. Hutchens, and Jay D. Hmielowski (2018). "Facebook News and (De)polarization: Reinforcing Spirals in the 2016 US Election." *Information, Communication & Society*, 21, 940–958.
- Bessi, Alessandro, Fabiana Zollo, Michela Del Vicario, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi (2015). "Trend of Narratives in the Age of Misinformation." PLOS ONE, 10, 1–16.
- Bessone, Pedro, Filipe R Campante, Claudio Ferraz, and Pedro Souza (2022). "Social Media and the Behavior of Politicians: Evidence from Facebook in Brazil." Working paper, NBER.
- Bond, Robert M., Fariss. Christopher J, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler (2012). "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature*, 489, 295.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro (2017). "Greater Internet Use Is Not Associated with Faster Growth in Political Polarization Among US Demographic Groups." *Proceedings of the National Academy of Sciences of the United States of America*, 114, 10612–10617.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro (2018). "A Note on Internet Use and the 2016 U.S. Presidential Election Outcome." *PLOS ONE*, 13, 1–7.
- Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro (2019). "Cross-Country Trends in Affective Polarization." Working Paper No. 26669.
- Braghieri, Luca, Ro'ee Levy, and Alexey Makarin (2022). "Social Media and Mental Health." American Economic Review, 112(11), 3660–93.
- Bursztyn, Leonardo, Georgy Egorov, Ruben Enikolopov, and Maria Petrova (2019). "Social Media and Xenophobia: Evidence from Russia." Working Paper No. 26567, NBER, http://www.nber.org/papers/w26567.
- Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin (2020). "From Extreme to Mainstream: How Social Norms Unravel." *American Economic Review*, 110(11), 3522–3548.

- Campante, Filipe, Ruben Durante, and Francesco Sobbrio (2017). "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation." *Journal of the European Economic Association*, 16, 1094–1136.
- Campante, Filipe and Daniel Hojman (2013). "Media and Polarization: Evidence from the Introduction of Broadcast TV in the US." *Journal of Public Economics*, 100, 79–92.
- Chen, Emily, Ashok Deb, and Emilio Ferrara (2020). "# Election2020: The First Public Twitter Dataset on the 2020 US Presidential Election." arXiv preprint arXiv:2010.00600.
- Chen, Yuyu and David Y. Yang (2019). "The Impact of Media Censorship: 1984 or Brave New World?" *American Economic Review*, 109(6), 2294–2332.
- Chernozhukov, Victor, Christian Hansen, and Martin Spindler (2015a). "Post-Selection and Post-Regularization Inference in Linear Models with Many Controls and Instruments." *American Economic Review*, 105(5), 486–90.
- Chernozhukov, Victor, Christian Hansen, and Martin Spindler (2015b). "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach." *Annual Review of Economics*, 7, 649–688.
- Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig (2019). "Inference with Arbitrary Clustering." IZA Discussion Papers 12584, Institute of Labor Economics (IZA), https://EconPapers.repec.org/RePEc:iza:izadps:dp12584.
- Conover, Michael D., Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini (2011). "Political Polarization on Twitter." In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*.
- Del, Vicario Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi (2016). "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences*, 113, 554–559.
- DellaVigna, Stefano, Ruben Enikolopov, Vera Mironova, Maria Petrova, and Ekaterina Zhuravskaya (2014). "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal: Applied Economics*, 6, 103–32.
- DellaVigna, Stefano and Eliana La Ferrara (2015). "Chapter 19 Economic and Social Impacts of the Media." In *Handbook of Media Economics, Handbook of Media Economics*. Vol. 1, edited by Simon P. Anderson, Joel Waldfogel, and David Strömberg . North-Holland, pp. 723–768. http://www.sciencedirect.com/science/article/pii/B978044463685000019X.
- DellaVigna, Stefano and Matthew Gentzkow (2010). "Persuasion: Empirical Evidence." Annual Review of Economics, 2(1), 643–669.
- DellaVigna, Stefano and Ethan Kaplan (2007). "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122, 1187–1234.
- Diaz, Fernando, Michael Gamon, Jake M Hofman, Emre Kıcıman, and David Rothschild (2016). "Online and Social Media Data as an Imperfect Continuous Panel Survey." *PloS one*, 11, e0145406.
- Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker (2019). "How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data." *SAGE Open*, 9, 2158244019832705.
- eMarketer (2019). "US Time Spent with Social Media 2019, By Debra Aho Williamson."
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020). "Social Media and Protest Participation: Evidence from Russia." *Econometrica*, 88, 1479–1514.
- Enikolopov, Ruben and Maria Petrova (2015). "Chapter 17 Media Capture: Empirical Evidence." In *Handbook of Media Economics, Handbook of Media Economics*, vol. 1, edited by Simon P. Anderson, Joel Waldfogel, and David Strömberg, pp. 687–700. North-Holland, http://www. sciencedirect.com/science/article/pii/B9780444636850000176.
- Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya (2011). "Media and Political Persuasion: Evidence from Russia." *The American Economic Review*, 101(7), 3253–3285.
- Falck, Oliver, Robert Gold, and Stephan Heblich (2014). "E-lections: Voting Behavior and the Internet." *American Economic Review*, 104(7), 2238–65.
- Fergusson, Leopoldo and Carlos Molina (2021). "Facebook Causes Protests." Documentos CEDE 018002, Universidad de los Andes—CEDE, https://ideas.repec.org/p/col/000089/018002.html.

- FiveThirtyEight (2015). "Let's Be Serious About Ted Cruz From The Start: He's Too Extreme And Too Disliked To Win."
- Flaxman, Seth, Sharad Goel, and Justin M. Rao (2016). "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly*, 80(S1), 298–320.
- Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti (2019). "Internet and Politics: Evidence From UK Local Elections and Local Government Policies." *The Review of Economic Studies*, 86, 2092–2135.
- Gentzkow, Matthew (2006). "Television and Voter Turnout." *The Quarterly Journal of Economics*, 121, 931–972.
- Gentzkow, Matthew and Jesse M. Shapiro (2010). "What Drives Media Slant? Evidence from US Daily Newspapers." *Econometrica*, 78, 35–71.
- Gentzkow, Matthew and Jesse M. Shapiro (2011). "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics*, 126, 1799–1839.
- 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.
- Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw (2011). "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *The American Political Science Review*, 105, 135–150.
- Gerber, Alan S. and Donald P. Green (2000). "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *The American Political Science Review*, 94, 653– 663.
- Gerber, Alan S., Dean Karlan, and Daniel Bergan (2009). "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics*, 1, 35–52.
- Guess, A (2018). "(Almost) Everything in Moderation: New Evidence in Americans' Online Media Diets."
- Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya (2020). "3G Internet and Confidence in Government." *The Quarterly Journal of Economics*, 136, 2533–2613.
- Halberstam, Yosh and Brian Knight (2016). "Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter." *Journal of Public Economics*, 143, 73–88.
- Hargittai, Eszter (2015). "Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites." The Annals of the American Academy of Political and Social Science, 659, 63–76.
- Howard, Philip N., Aiden Duffy, Deen Freelon, M.M. Hussain, Will Mari, and Marwa Maziad (2011)."Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?" Project on Information Technology and Political Islam Working paper.
- Huber, Gregory A. and Kevin Arceneaux (2007). "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science*, 51, 957–977.
- Jones, Jason J., Robert M. Bond, Eytan Bakshy, Dean Eckles, and James H. Fowler (2017). "Social Influence and Political Mobilization: Further Evidence From a Randomized Experiment in the 2012 U.S. Presidential Election." *PLOS ONE*, 12, 1–9.
- Kinder-Kurlanda, Katharina, Katrin Weller, Wolfgang Zenk-Möltgen, Jürgen Pfeffer, and Fred Morstatter (2017). "Archiving Information from Geotagged Tweets to Promote Reproducibility and Comparability in Social Media Research." *Big Data & Society*, 4, 2053951717736336.
- Larcinese, Valentino and Luke Miner (2017). "The Political Impact of the Internet on US Presidential Elections." STICERD Economic Organisation and Public Policy Discussion Papers Series 63, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE, https://ideas.repec.org/p/cep/stieop/63.html.
- Lelkes, Yphtach, Gaurav Sood, and Shanto Iyengar (2017). "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Affect." *American Journal of Political Science*, 61, 5–20.
- Levy, Ro (2021). "Social Media, News Consumption, and Polarization: Evidence from a Field Experiment." *American Economic Review*, 111(3), 831–70.

- Littman, Justin, Laura Wrubel, and Daniel Kerchner (2016). "2016 United States Presidential Election Tweet IDs."
- Manacorda, Marco and Andrea Tesei (2016). "Liberation Technology: Mobile Phones and Political Mobilization in Africa." CEPR Working Paper No. 11278 https://ideas.repec.org/p/cpr/ceprdp/ 11278.html.
- Martin, Gregory J. and Ali Yurukoglu (2017). "Bias in Cable News: Persuasion and Polarization." American Economic Review, 107(9), 2565–2599.
- Mosquera, Roberto, Mofioluwasademi Odunowo, Trent McNamara, Xiongfei Guo, and Ragan Petrie (2020). "The Economic Effects of Facebook." *Experimental Economics*, 23, 575–602.
- Müller, Karsten and Carlo Schwarz (2020). "Fanning the Flames of Hate: Social Media and Hate Crime." *Journal of the European Economic Association*, 19, 2131–2167.
- Müller, Karsten and Carlo Schwarz (2023). "From Hashtag to Hate Crime: Twitter and Antiminority Sentiment." *American Economic Journal: Applied Economics*, 15(3), 270–312.
- Napoli, Philip M. (2014). "Measuring Media Impact." The Lear Center. http://www.learcenter.org/ pdf/measuringmedia.pdf.
- Nelson, Jacob L. and James G. Webster (2017). "The Myth of Partisan Selective Exposure: A Portrait of the Online Political News Audience." *Social Media + Society*, 3, 2056305117729314.
- New York Magazine (2022). "The 'Twitter Files' Is What It Claims to Expose ."
- New York Times (2017). "The Fake Americans Russia Created to Influence the Election, By Scott Shane."
- NPR (2020a). "Facebook Clamps Down On Posts, Ads That Could Undermine U.S. Presidential Election."
- NPR (2020b). "Facebook Keeps Data Secret, Letting Conservative Bias Claims Persist."
- NPR (2020c). "President-Elect Joe Biden Hits 80 Million Votes In Year Of Record Turnout."
- NPR (2022). "Elon Musk Is Using the Twitter Files to Discredit Foes and Push Conspiracy Theories." Oliver, J. Eric, Thomas Wood, and Alexandra Bass (2016). "Liberellas versus Konservatives: Social
- Status, Ideology, and Birth Names in the United States." *Political Behavior*, 38, 55–81.
- Oster, Emily (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37, 187–204.
- Pariser, E. (2011). The Filter Bubble: What The Internet Is Hiding From You. Penguin Books Limited, https://books.google.com/books?id=-FWO0puw3nYC.
- Pew Research Center (2009). "The Internet's Role in Campaign 2008." Pew Research Center Working paper.
- Pew Research Center (2011). "Social Networking Sites and Our Lives." Pew Research Center Working paper.
- Pew Research Center (2014). "Political Polarization & Media Habits." Pew Research Center Working paper.
- Pew Research Center (2015). "Americans' Internet Access: 2000-2015." Pew Research Center Working paper.
- Pew Research Center (2018). "For Most Trump Voters, 'Very Warm' Feelings for Him Endured." Pew Research Center Working paper.
- Pew Research Center (2019a). "About One-In-Five Adult Twitter Users in the U.S. Follow Trump." Pew Research Center Working paper.
- Pew Research Center (2019b). "National Politics on Twitter: Small Share of U.S. Adults Produce Majority of Tweets." Pew Research Center Working paper.
- Pew Research Center (2019c). "Share of U.S. Adults Using Social Media, Including Facebook, Is Mostly Unchanged Since 2018." Pew Research Center Working paper.
- Pew Research Center (2019d). "Sizing Up Twitter Users." Pew Research Center Working paper.
- Pew Research Center (2020). "Democrats On Twitter More Liberal, Less Focused On Compromise Than Those Not On The Platform." Pew Research Center Working paper.
- Politico (2020). "Biden Campaign Lashes Out at New York Post."
- Quora (2011). "What is the process involved in launching a start-up at SXSW?"
- Rotesi, Tiziano (2019). "The Impact of Twitter on Political Participation." University of Lausanne Working Paper.

- Schmidt, Ana Lucía, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi (2017). "Anatomy of News Consumption on Facebook." *Proceedings of the National Academy of Sciences*, 114, 3035–3039.
- Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge (2015). "What Are We Weighting For?" *Journal of Human resources*, 50, 301–316.
- Spenkuch, Jörg L. and David Toniatti (2018). "Political Advertising and Election Results." *The Quarterly Journal of Economics*, 133, 1981–2036.
- Statista (2019). "Twitter: Number of Monthly Active U.S. Users 2010-2019."
- Statista (2020). "Leading Countries Based on Number of Twitter users as of October 2020."
- Strömberg, David (2015). "Media and Politics." Annual Review of Economics, 7, 173–205.
- Sunstein, Cass R. (2017). # Republic: Divided Democracy in the Age of Social Media. Princeton University Press.
- The Guardian (2018). "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach, by Carole Cadwalladr and Emma Graham-Harrison."
- The Guardian (2023). "Twitter's Rightwing Takeover Is Complete. Why Are Liberals Still on It?"
- The Hill (2016). "Stop Blaming Facebook for Trump's Election Win, By Keith N. Hampton and Eszter Hargittai."
- The New Yorker (2016). "Obama Reckons with a Trump Presidency, By David Remnick."
- Twitter (2010). "Measuring Tweets, by Kevin Weil."
- Twitter (2020). "Empowering US Voters on National Voter Registration Day, by Bridget Coyne."
- UK Parliament (2019). "Disinformation and 'fake news': Final Report, By Digital, Culture, Media and Sport Committee."
- USA Today (2023). "Is Elon Musk's Twitter More Conservative? Marjorie Taylor Greene, Other GOP Stars Win Big."

Venture Beat (2008). "Hitwise: Twitter Traffic Is, in Fact, Going up but Still Not Big, by Eric Eldon." Wall Street Journal (2020). "Twitter's Partisan Censors."

Weber, Ingmar, Venkata R. Kiran Garimella, and Alaa Batayneh (2013). "Secular vs. Islamist Polarization in Egypt on Twitter." In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13*, p. 290–297, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/2492517. 2492557.

Wired (2016). "Here's How Facebook Actually Won Trump the Presidency, By Issie Lapowsky."

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov (2020). "Political Effects of the Internet and Social Media." *Political Effects of the Internet and Social Media*, 12, 415–438.

Supplementary Data

Supplementary data are available at *JEEA* online.

© The Author(s) 2023. Published by Oxford University Press on behalf of European Economic Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com