Discussion Papers in Economics

No. 21/01

Britain has had enough of experts? Social networks and the Brexit referendum

Giacomo De Luca, Thilo R. Huning,
Paulo Santos Monteiro

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

This version: January 25, 2021

Britain has had enough of experts? Social networks and the Brexit referendum[*]

Giacomo De Luca[†] Thilo R. Huning[‡] Paulo Santos Monteiro[§]

Abstract

We investigate the impact of social media on the 2016 referendum on the United Kingdom membership of the European Union. We leverage 18 million geo-located Twitter messages originating from the UK in the weeks before the referendum. Using electoral wards as unit of observation, we explore how exogenous variation in Twitter exposure affected the vote share in favor of leaving the EU. Our estimates suggest that in electoral wards less exposed to Twitter the percentage who voted to leave the EU was greater. This is confirmed across several specifications and approaches, including two very different IV identification strategies to address the non-randomness of Twitter usage. To interpret our findings, we propose a model of how bounded rational voters learn in social media networks vulnerable to fake news, and we validate the theoretical framework by estimating how Remain and Leave tweets propagated differently on Twitter in the two months leading to the EU referendum.

JEL Codes: *D72 · D83 · L82 · L86*

Keywords: *Fake News · Social Networks · Social Media · Brexit*

[*] We would like to thank CAIDA for kindly sharing their data on internet outages with us.

[†] Faculty of Economics and Management, Free University of Bozen-Bolzano; e-mail: giacomo.deluca at unibz.it

[‡] Department of Economics and Related Studies, University of York; e-mail: thilo.huning at york.ac.uk

[§] Department of Economics and Related Studies, University of York; e-mail: paulo.santosmonteiro at york.ac.uk

1 Introduction

Over the last two decades social media transformed the way information is generated, consumed and shared (Gottfried and Shearer, 2016; Bialik and Matsa, 2017; Gavazza, Nardotto, and Valletti, 2019). About 50% of Europeans use social media daily, and 16% indicate social media as the major source of information regarding politics (Eurobarometer, 2016; Eurobarometer, 2020). Substituting real interactions with virtual ones through social media has been argued to affect social capital accumulation (Geraci et al., 2018; Antoci et al., 2019). Perhaps more importantly, social media have been blamed for increasing political polarization due to ideological segregation in news consumption (i.e. the repeated interaction within more homogeneous groups), as fostered by their selection algorithms which prioritize and propose information users may agree with (Glaeser and Sunstein, 2009; Gentzkow and Shapiro, 2011; Pariser, 2011).[1] The concern becomes even more relevant when considering the volume of "fake news", hosted on social media in the run-up to high stakes elections (Allcott and Gentzkow, 2017). Voters exposed to social media would then interact within "echo chambers", in which homophilic political information blending truthful and fake news represents a large share of the information considered (McPherson, Smith-Lovin, and Cook, 2001; Barberá et al., 2015; Levy and Razin, 2019).

The 2016 referendum on the United Kingdom membership of the European Union (the EU or "Brexit" referendum) has been widely reported to have been a watershed moment in the influence of fake news in electoral contests. Indeed, it led to the launching of a UK Parliamentary inquiry into Disinformation and fake news (Collins et al., 2019). In this paper, we investigate the impact of social media fake news on the outcome of the EU referendum. We do this by leveraging on 18 Million geo-located Twitter messages originating from the UK in the weeks before the EU referendum. Using electoral wards as our unit of observation, we explore how exogenous variation in the exposure to Twitter affected the share of votes in favor of leaving the EU.[2] Our estimates suggest that in electoral wards more exposed to Twitter the percentage who voted to leave the EU was smaller. This finding is confirmed across several specifications and approaches, including two alternative instrumental variable (IV) identification strategies to address the potential non randomness of Twitter usage (even after controlling for local authority fixed effects, and demographic and socio-economic characteristics of the electoral ward).

The first adopted identification strategy is based on an instrumental variable for internet diffusion similar to that proposed by Falck, Gold, and Heblich (2014), Campante, Durante, and Sobbrio (2018) and, in the

[1] The empirical evidence of information segregation among social media users is not consistent. For instance, Flaxman, Goel, and Rao (2016) report relatively mild levels of segregation for descriptive news articles on both Facebook and Twitter, although the level increases substantially for opinion pieces.

[2] Electoral wards are the smallest particles in the UK electoral geography with an average population of about 5,000 people (UK Office for National Statistics).

UK context, by Geraci et al. (2018). Specifically, in the UK the location of today's 5,564 asymmetric digital subscriber line (ADSL) local exchanges is predetermined by the historical location of the old telephone exchanges.[3] Broadband quality is determined by distance to the nearest local exchange (LE), and constitutes a powerful predictor of internet usage in the UK. The second identification strategy we propose is based on the frequency of unexpected internet outages in the weeks leading to the referendum. Internet outages are measured by the Center for Applied Internet Data Analysis, an internet telescope at the San Diego Supercomputer Center at UC San Diego who developed an operational prototype system that monitors the Internet in almost real-time, to identify macroscopic internet outages around the world geo-located via IP addresses (Benson et al., 2013).

Both identification strategies lead to the same conclusion, establishing a negative causal relationship between the use of Twitter and the percentage of votes in support of leaving the EU in the referendum. To support the validity of our findings, we conduct a battery of falsification tests to support both IV strategies. For one such test in particular, we look at local elections in the UK in 2002–2003 (before social media, but already in the internet age), and show that neither instrumental variable is correlated with pre-social media voting patterns. Thus, the instrumental variables we propose (which capture the availability, speed, and reliability of high-speed internet), are not correlated with historical political outcomes at the local level prior to the arrival of social media.

Our main finding, that electoral wards with greater Twitter exposure delivered a lower percentage of Leave votes, may seem at odds with the widespread perception that social media was a contributing factor to the referendum's outcome. The view that the Leave propaganda was more effective in social media and, in particular, on Twitter has been widely reported in the popular press (Siegel and Tucker, 2016; Field and Wright, 2018), and has also received support by data scientists employing machine learning methods to measure the stance of UK Twitter users (Grčar et al., 2017; Hänska and Bauchowitz, 2017). With this backdrop, we propose a simple theoretical framework which explains how the empirical finding that Twitter lowers the percentage of Leave votes is consistent with widespread propaganda supporting Leave on social media. The model features bounded rational voters connected in social networks, who use simple heuristics to filter fake news and, subsequently, behave as credulous Bayesians (a term coined by Glaeser and Sunstein, 2009), assigning excessive weight to messages supporting specific views of the world and to the "wisdom of crowds".[4]

The crucial assumption underlying our behavioral model is that voters apply a two-step behavioral heuristic

[3] The ADSL is the most common type of broadband internet connection. It is delivered through the wires that carry phone lines.

[4] A recent survey, reporting that only 20% of Europeans trust social media, is consistent with our underlying assumption (Eurobarometer, 2020).

when they engage with news on social media about the state of the world. They ignore entirely news feeds which they suspect to be fake news, and treat news which are sufficiently plausible as entirely unbiased. If fake news is perceived to be biased towards supporting a given state of the world, then the upshot is that news which proclaim that state are going to be discounted as fake by agents. Individuals connected in social networks subsequently share filtered information and update their beliefs using linear updating rules which have become standard in the literature on learning in networks (DeMarzo, Vayanos, and Zwiebel, 2003; Acemoglu, Ozdaglar, and ParandehGheibi, 2010; Golub and Jackson, 2010). The behavioral assumptions we make are supported by experimental evidence recently obtained by Kirill and Shum (2019), showing that individuals who interact in social networks both share news signals selectively and, at the same time, naïvely take signals at face value, ignoring the selection bias in the shared signals.

To generate a negative link between Twitter exposure and the support for Leave, our theoretical framework requires Leave supporting messages to be more likely perceived as fake news. Social media news supporting leaving the EU would then be less trusted by Twitter users, and often dismissed as fake. Relying on sentiment and language analysis on our rich data, we propose two complementary strategies substantiating this conjecture. First, we exploit the fact that bots are often associated with the spread of fake news. Building on the recent work by Gorodnichenko, Pham, and Talavera (2018), reporting substantially greater activity by bots supporting the Leave campaign, we document that Twitter users leaning more towards Leave were indeed more likely to display features generally not consistent with human behaviour, and therefore associated with bots, e.g., abnormally large volume of messages, activity in unusual (night) time, and the repetition of identical messages.

Second, we make use of the full database of 18 million tweets to measure how Remain and Leave supporting tweets were differently perceived by Twitter users. More specifically, we show that tweets in support of Remain are more likely to generate more followers for the user who originates the tweet. Instead, Leave supporting tweets do not generate new followers for the sender. We interpret this as evidence that tweets in support of Remain were regarded by users of Twitter to be genuine, or at the very least, from legitimate sources.

Finally, we show that wards with larger Twitter exposure were systematically reporting less change in support for leaving the EU over time. We measure the change in beliefs by the difference between the support for Leave at the EU referendum and the support for UKIP (a party created in the early 1990s with the explicit goal of leaving the EU) in the 2014 EU elections. We interpret this as evidence that social media lowers the informational gain from news about politics. This results seem consistent with results by Gavazza, Nardotto, and Valletti (2019), that greater internet penetration (and thus, a greater role for social media) seems to decrease the competitiveness of elections by favoring incumbents. We interpret this finding as evidence that social media delays learning, as users dismiss as fake and ignore news which are at odds with their prior beliefs about the state of the world. If, moreover, fake news is perceived by all to be biased in favor of a given (contrarian) worldview, this may generate a "wisdom of the crowds" curse: agents in social networks supporting the consensus world view are given undue weight, and legitimate expertise supporting the contrarian world view is dismissed as fake news.

The rest of the paper is organized as follows. In the next section we discuss the related literature in greater length. In Section 3 we outline a theoretical framework to help organise and interpret our empirical findings. The data and empirical strategy are presented in Sections 4 and 5, respectively. The empirical results are reported and discussed in Section 6. Finally, Section 7 offers some concluding remarks.

2 Related literature

This paper contributes to at least two separate literatures: on the quantitative impact of the internet on political outcomes; and the literature studying political information diffusion in social networks (such as Twitter and Facebook). First, and most directly, it provides additional insights on the impact of the internet (and social media) on political outcomes and behavior. Starting from the seminal contributions by Sunstein (2001) and Sunstein (2009), a growing theoretical literature has investigated the effects of the new forms of social interactions on electoral competition and political behavior. Gentzkow, Shapiro, and Stone (2015) review the theoretical literature on market determinants of media bias and propose an encompassing theoretical framework distinguishing between supply-side forces (biased media production), and demand-side forces (biased consumer preferences). Allcott and Gentzkow (2017) model fake news production and consumption, offering stylized empirical evidence from the 2016 US presidential election to understand how fake news emerges in the media market. Fake news emerges in equilibrium because consumers have a willingness to pay for partisan news, whilst fake news is cheap to produce and costly for consumers to detect. Grossman and Helpman (2019) study electoral competition with fake news and show that when parties can manipulate information, they face a trade-off between appealing to well informed voters and to gullible voters who can be easily misled by fake news. In their model, if there are limits to parties' ability to misreport their own platform, the possibility of fake news may lead to a polarized electorate.

A parallel empirical literature has developed methods to estimate the effect of internet and social media on political behavior. Falck, Gold, and Heblich (2014) study internet penetration in the German market using an identification strategy very similar to the one we propose in this paper. They find that greater broadband penetration decreases political participation, as measured by aggregate turnout, and argue that this follows from a drop in political information due to the substitution of traditional more informative media (e.g., TV and local newspapers which feature greater coverage of (local) political issues) for internet-based media. To address the endogeneity of broadband access, Falck, Gold, and Heblich (2014) adopt an IV strategy based on the distance of users to the telephony infrastructure, which was designed long before the internet era but affected the location of the internet infrastructure. While the length of the cable from the main wires is not important for telephony, it is crucial for the internet connectivity and speed. Our first IV strategy, based on the distance to LE, follows the same logic, and is also adopted by Campante, Durante, and Sobbrio (2018) and Geraci et al. (2018), the latter to study the impact of the internet on social capital formation in the UK.

Like Falck, Gold, and Heblich (2014), the study by Campante, Durante, and Sobbrio (2018) also finds a detrimental short-run impact of internet penetration on electoral turnout looking at Italian elections. However, they find evidence that the long-run effects might have the reverse sign, as disenchanted or demobilized voters are captured by political parties able to use the internet more effectively. Perhaps consistent with this interpretation, Miner (2015) uses a similar identification strategy and found that better internet access was detrimental to the incumbent ruling coalition in the Malaysian 2008 elections for state legislatures, and did not lower turnout. Crucially, in the context of Malaysia (which still has fragile democratic institutions according to the Economist Intelligence Unit) better internet access is not found by Miner (2015) to raise the exposure to fake news, and is instead associated with an increase in media trustworthiness, as it offers an escape from the censorship affecting other forms of state-controlled media.

Gavazza, Nardotto, and Valletti (2019), using detailed data on internet penetration in the UK, study how increased exposure to internet news content affects local election outcomes (at the electoral ward level, similar to our analysis) and local governments' policy choices. Their paper is particularly relevant to our study due to its focus on UK political outcomes, but also because our second identification strategy (based on internet outages) is related to that proposed by Gavazza, Nardotto, and Valletti (2019). They use the fact that internet network outages are more likely to occur during rain and thunderstorms and, hence, use weather (rainfall data) as an instrument for internet penetration. We instead use a direct measure of (unexpected) internet outages in the weeks leading to the referendum (thus, capturing a direct treatment effect instead of an intention to treat estimated by Gavazza, Nardotto, and Valletti, 2019). Similar to Falck, Gold, and Heblich, 2014, the study by Gavazza, Nardotto, and Valletti (2019) finds that an increase in internet penetration lowers voter turnout (with a -21% elasticity), and also favors incumbents. Rainfall is shown not to affect political outcomes before the diffusion of broadband, establishing support for the causal impact of the internet. We obtain similar falsification tests, by documenting the lack of predictive ability of our main instrumental variable (distance to the local exchange) for the UKIP vote share before the social media era and, thus, supporting the causal link between social media and electoral outcomes.

To interpret our empirical findings, we propose a simple model in which individuals filter information using simple heuristics and subsequently behave as credulous Bayesians. This follows a growing literature arguing that voters (and more generally news consumers) are consciously choosing the source of their political news, and weigh differently their credibility depending on their perceived ideological bias (Chiang and Knight, 2011; Durante and Knight, 2012). Thus, individuals select and filter their news sources. If the traditional media is perceived as less biased in comparison to social media, this filtering would lead voters assigning excessive weight to the traditional media. This is compatible with the survey in Allcott and Gentzkow (2017), who report how trust in information accessed through social media is substantially lower than trust in traditional outlets. On the other hand, there is evidence that traditional media (in particular, big broadcast conglomerates) exert "media power" and can sway voters' behavior to a greater extent than internet media sources (Prat, 2018). During the EU referendum, the Vote Leave campaign accused the big UK broadcasting groups (the BBC and ITV) of exerting media power (in the sense of Prat, 2018) by exaggerating the economic cost of leaving the EU – the so called "project fear" (Johnson, 2016; Moore and Ramsay, 2017). Arguably attempting to undermine media power, prominent Leave campaigner and UK conservative politician, Michael Gove, said the UK public "have had enough of experts" (*sic*).

Our second focus is on the role information diffusion in networks has in shaping electoral outcomes. There iss evidence that social network ties are important for shaping political outcomes through social media. A pioneering study by Bond et al. (2012) implements a randomized control trial of political mobilization messages, delivered to 61 million Facebook users during the 2010 US congressional elections. The messages shared were encouraging political participation. Receiving an encouragement to go to vote from a known friend on Facebook was shown to increase electoral participation. Tight social links thus establish trust in social networks. Another recent study examining the effects of Facebook on political contests is Liberini et al. (2020), who propose using variation in advertising prices for narrowly defined audiences as a measure of exposure intensity to political messages on social media. They show that during the 2016 US Presidential election, a larger exposure to online political ads made individuals less likely to change their initial voting intentions (with the effect particularly strong among Trump supporters). In another related paper, Fujiwara, Müller, and Schwarz (2020) exploit the staggered adoption of Twitter in the US, shaped by the fact that individuals exogenously exposed to an early Twitter advertisement campaign are more likely to have been early adopters of Twitter and, thus, more likely to be current users. Twitter is found to have had an impact on the 2016 US Presidential election, lowering the percentage of Republican voters. Finally, the political mobilization effect of social media may go beyond electoral contest. For example, by reducing the costs of coordination in collective action, social media can increase participation in mass protest, as shown by Enikolopov, Makarin, and Petrova (2020) in the case of the 2011 waves of protests in Russia.[5] Using internet outages as an instrumental variable (similar to our second empirical strategy), Müller and Schwarz (2020) establish a causal relationship between Facebook and violence against refugees in German municipalities.

A related literature investigates the impact of online social networks and social media on polarization and political fragmentation. Halberstam and Knight (2016) analyze nearly 500,000 communications during the 2012 US Presidential elections in a social network of about two million Twitter users and find that users affiliated with majority political groups, relative to the minority group, have more connections, are exposed to more information, and are exposed to information more quickly. This suggests that due to network homophily mainstream views spread more quickly in social networks such as Twitter, a finding which is consistent with our own theoretical interpretation of our empirical results. Closely related to our study, Gorodnichenko, Pham, and Talavera (2018) analyze the interaction between humans and bots on Twitter, using the EU referendum and the 2016 US presidential elections as a laboratory. Their results suggest that bots may have contributed to shape electoral outcomes, but that the degree of bots' influence depends on whether they provide information consistent with the priors of the human users. Measuring engagement through retweeting activity, bots also appear to be discounted by humans who engage more strongly with tweets generated by other humans than with tweets generated by bots. The latter result appears entirely consistent with our theoretical framework.

With respect to the literature discussed above we propose a simple mechanism through which social media may affect political behavior, which draws on some of the insights from the work on social learning with bounded rational voters (a literature not surveyed here, but which includes DeMarzo, Vayanos, and Zwiebel, 2003; Acemoglu, Ozdaglar, and ParandehGheibi, 2010; Golub and Jackson, 2010; Golub and Jackson, 2012; Azzimonti and Fernandes, 2018). Bounded rational voters will filter fake news spread on social media based on their own beliefs and the common knowledge about the ideological bias of fake news. When exposed to news from social media instead of more trustworthy and established sources, voters choose endogenously to update less based on the new information and assign undue weights to news which either

[5] A similar effect is highlighted by Manacorda and Tesei (2020), focusing on the role of mobile phones' diffusion on political mobilisation in Africa.

conform strongly with their priors or that appear less likely to be fake news because of its content. The next section presents our theoretical framework.

3 Social media and voting: theoretical framework

We propose a simple model of learning with fake news in a set-up with bounded rational voters who interact in social networks. We consider a referendum election race between two alternative platforms: $x = 0$, yielding the status quo; and $x = 1$, the policy reform (thus, in our context, the policy reform is for the UK to leave the EU). An electoral ward, which in the empirical section is our baseline unit of observation of electoral outcomes, is comprised of a finite (large) set of voters with names $i \in \mathcal{N} = \{1, 2, \ldots n\}$. Voters in an electoral ward are socially connected and, thus, share and aggregate information.

The aggregate benefits of the policy reform are unknown ex-ante and, in particular, there are two possible states of the world, $s \in \mathcal{S} = \{0, 1\}$, yielding the aggregate benefits of the policy reform. Voters have state dependent preferences over the discrete valued policy variable $x \in \{0, 1\}$, represented by the following utility function

$$u(x, s) = (1 - x) z^i + xs, \tag{1}$$

with $z^i \in [0, 1]$ a random preferences shock with cumulative distribution $G(z)$ across the population, and representing the private net benefits to individual i of preserving the status quo. For simplicity, we set $G(z)$ to be the uniform cumulative distribution.

Notwithstanding the heterogeneous preferences over the policy reform, everyone agrees over the preferred policy in the absence of uncertainty over the state of the word. If the state of the world is known to be $s = 1$, $x = 1$ is the preferred alternative for everyone, while if $s = 0$ is known, it is $x = 0$ the preferred alternative. Political disagreement stems from the combination of ex-ante uncertainty about the state of the world and heterogeneous preferences.

3.1 Prior beliefs and political preferences

Individual voters in a given electoral ward $w \in \mathcal{W}$ share common prior beliefs about s, such that the probability assigned to the state of the world $s = 1$ in ward w is denoted $\hat{\pi}_w \in [0, 1]$. Individuals vote for the policy which maximizes their subjective expected utility. Hence, the preferred policy for voter i in constituency $w \in \mathcal{W}$ given prior beliefs $\hat{\pi}_w$ is given by

$$\begin{aligned} x^{\star} &= \arg\max_{x\in\{0,1\}} E_{\hat{\pi}_w}\left[u\left(x,z,s\right)\right], \\ &= \arg\max_{x\in\{0,1\}} \hat{\pi}_w \left[\left(1-x\right)z^i + x\right] + \left(1-\hat{\pi}_w\right)\left[\left(1-x\right)z^i\right], \\ &= \arg\max_{x\in\{0,1\}} \left[z^i + x\left(\hat{\pi}_w - z^i\right)\right]. \end{aligned} \tag{2}$$

We establish the following Proposition:

Proposition 1. *Let $\hat{\pi}_w$ be the probability assigned by voters in electoral ward $w \in \mathcal{W}$ to state of the world $s = 1$. The probability that voter $i \in \mathcal{N}$ prefers the policy reform to the status quo is given by $G\left(\hat{\pi}_w\right)$. With $G\left(z\right)$ the uniform cumulative distribution, the vote share in favor of the policy reform is $\hat{\pi}_w$.*

The proof follows immediately from (2), which implies that the probability that a voter in electoral ward $w \in \mathcal{W}$ prefers the policy reform ($x = 1$) to the status quo ($x = 0$) is given by Prob $\left(z^i \leq \hat{\pi}_w\right) = G\left(\hat{\pi}_w\right)$. Since $G\left(z\right)$ is the uniform cumulative distribution function, the vote share in favor of the policy reform is given by $\hat{\pi}_w$.

3.2 Learning and fake news

Each day voters learn about the state of the world. Learning takes place over two stages. In Stage 1, voters receive private messages about the state of the world, $m \in \{0,1\}$, and update their beliefs accordingly (using a simple heuristic to filter fake news). In Stage 2, socially connected agents share their updated beliefs, and aggregate information using an averaging rule similar to that in DeMarzo, Vayanos, and Zwiebel (2003) and Golub and Jackson (2010) and Golub and Jackson (2012).

The private messages received by voters in Stage 1, come from three sources: traditional news media (for example, the online edition of Broadsheet UK newspapers); social media news feeds from legitimate sources; social media news feeds from fake sources. The proportion of individuals who receive news from social media feeds (either legitimate of fake) is denoted $\lambda_w \in [0,1]$. We refer to news from traditional media and from legitimate social media feeds as legit news, while fake social media feeds are fake news. Legit news are reports on experiments conducted by experts. These experiments yield message $m \in \{0,1\}$, with probability measure

$$\mathcal{P}: \Pr\Big(m\big|s=1,\text{legit}\Big) = 1 - \Pr\Big(m\big|s=0,\text{legit}\Big) = \alpha^m\left(1-\alpha\right)^{1-m}, \tag{3}$$

with $\alpha \in (1/2, 1)$ common knowledge, and where $\Pr\left(\bullet|s\right)$ denotes the probability of an event conditional on s. Since $\alpha > 1/2$, legit news are more likely to report the true state of the world.

Fake news, instead, are generated by bots and report on fake experiments. As in Azzimonti and Fernandes (2018), but adapting the terminology to capture the EU referendum, there are two types of bots: L-bots who only make claims in support of Leaving the EU, and R-bots who only make claims supporting Remain. Thus, messages by bots, $m \in \{0, 1\}$, are generated with probability measure

$$\widetilde{\mathcal{P}} : \Pr\Big(m|s = 1, \text{fake}\Big) = \Pr\Big(m|s = 0, \text{fake}\Big) = \Pr\Big(m|\text{fake}\Big) = \beta^{m} (1 - \beta)^{1-m}, \tag{4}$$

where $\beta \in (0, 1)$ is the proportion of L-bots, which is common knowledge. Of course fake news provide no information about the state of the world, since the messages are generated by a probability measure which is the same across the two states. The proportion of fake news in the social media is $\mu \in (0, 1)$.

Upon receiving messages, individuals update their beliefs using a simple computational heuristics which departs from the fully rational model in a way which echoes the Fryer Jr, Harms, and Jackson (2019) framework. In particular, we assume agents follow a two-step strategy. First, they compute the probability that the message is fake, given by

$$\Pr(\text{fake}|m) = \begin{cases} 0 & \text{if message is from traditional media,} \\ \mu\beta \left(\Pr(m = 1)\right)^{-1}, & \text{if message is from social media and } m = 1; \\ \mu (1 - \beta) \left(\Pr(m = 0)\right)^{-1}, & \text{if message is from social media and } m = 0; \end{cases} \tag{5}$$

Second, if the probability that the message is fake exceeds a given threshold level $\tau \in (0, \mu)$, agents ignore the message and hold on to their prior belief. If instead the probability that it is fake is less than τ, agents behave as if the message is legitimate with certainty and update their prior accordingly.

Next, we impose the following assumption

Assumption 1. *The proportion of bots supporting the state of the world under which the policy reform is preferred (in our context, leaving the EU) is large, in the following sense:*

$$\beta \geq \frac{\tau}{\mu}. \tag{6}$$

This assumption ensures that β the proportion of L-bots is not too small. Hence, in the context of the UK referendum on EU membership, we assume that there is a sufficiently large number of bots tweeting Leave supporting messages. Now, given Assumption 1, any social media message which supports the policy reform is perceived by agents to be fake news with a probability above the threshold level τ

$$\Pr\left(\text{fake}|m=1\right) \geq \tau, \tag{7}$$

even for agents who believe that $s = 1$ with certainty. As an upshot, all social media messages supporting the policy reform are discarded as fake news.

We establish the following Proposition:

Proposition 2. *Suppose agents compute the probability that a message is fake, and ignore messages which have a probability of being fake above a threshold level, τ. If the proportion of* L*-bots is large enough (satisfying Assumption 1), all social media messages supporting the policy reform will be discarded as fake news, regardless of the agent's prior beliefs, $\hat{\pi}^0$.*

The proof follows immediately from (5) and noticing that, given Assumption 1, $\mu\beta\left(\Pr\left(m=1\right)\right)^{-1} \geq \tau$ *almost surely*. Next, we turn our attention to how socially connected agents share their updated beliefs and aggregate information.[6]

3.3 Belief dynamics in social networks

We capture belief formation in social networks by adapting features of the DeMarzo, Vayanos, and Zwiebel (2003), and Golub and Jackson (2010) and Golub and Jackson (2012) models of naïve learning in networks. The social network is represented over the set of individuals in the ward, $\mathcal{N} = \{1, 2\ldots, n\}$, but not all individuals need to be connected. We assume that n is a large number, so that we may apply the law of large numbers. Interaction patterns are captured through an $n \times n$ nonnegative interaction matrix $\mathbf{T}$. The interaction matrix $\mathbf{T}$ is allowed to be asymmetric and, in particular, display one-sided interactions: $\mathbf{T}_{ij} > \mathbf{T}_{ji} = 0$. Specifically, the matrix $\mathbf{T}$ is obtained endogenously, as follows

$$\mathbf{T}_{ij} = \begin{cases} 0, & \text{if } i \text{ does not follow } j; \\ \\ 0, & \text{if } i \text{ follows } j\text{, but } j \text{ does not share a new posterior belief;} \\ \\ (1/n_i)\,, & \text{if } i \text{ follows } j\text{, and } j \text{ shares a new posterior belief;} \end{cases} \tag{8}$$

[6] Proposition 2 is purposely establishing a stark result (all leave supporting tweets are discarded as fake). Of course, Assumption 1 could be relaxed to achieve a less extreme result. However, we consider this stark example to illustrate clearly how our theoretical framework allows for biased learning in networks.

where $n_i \leq n$ is the number of contacts who individual i follows and who shares an updated belief (thus, each row of $\mathbf{T}$ is normalized to sum to unity). We also assume that the social network is dense, meaning that $n_i \to \infty$ as $n \to \infty$ (for all individuals i).

Let $\hat{\pi}^0$ denote individual initial beliefs. Since individuals have common priors, $\hat{\pi}^0$ is the same for all members. Upon receiving a private message in Stage 1, member $i \in \mathcal{N}$ updates her beliefs using Bayes law, and thus we have that

$$\hat{\pi}_i^1 = \begin{cases} \hat{\pi}^0, & \text{if fake news is suspected;} \\ \\ \mathbf{Q} > \hat{\pi}^0, & \text{if } m = 1\text{, and the news is trusted as legit;} \\ \\ \mathbf{q} < \hat{\pi}^0, & \text{if } m = 0\text{, and the news is trusted as legit.} \end{cases} \tag{9}$$

Next, in Stage 2, individuals update beliefs by taking the weighted averages of their connection's average posterior beliefs and their prior belief, as follows

$$\hat{\Pi}^2 = \mathbf{T}\hat{\Pi}^1, \tag{10}$$

where $\hat{\Pi}^1$ and $\hat{\Pi}^2$ are column n-dimensional vectors with entries given by the individual beliefs, in turn, $\hat{\pi}_i^1$ and $\hat{\pi}^2$. The behavioral rule (10) for updating beliefs is motivated by the average-based updating rule developed in Golub and Jackson (2010). Although it is not fully rational, this rule will (under some regularity conditions) converge in the limit to a fully rational posterior belief, whilst only requiring simple computational heuristics.

We characterize the belief dynamics in large networks, as $n \to \infty$, and which are dense, so that all agents have a large number of contacts. Thus, we obtain the following difference equation for the individual beliefs

$$\begin{aligned} \hat{\pi}_w^2 &= \lim_{n_i \to \infty} \sum_{j=0}^{n_i} \mathbf{T}_{ij} \hat{\pi}_j^1, \\ &= \delta \mathbf{q} + (1-\delta) \left[\alpha^s (1-\alpha)^{1-s} \mathbf{Q} + \alpha^{1-s} (1-\alpha)^s \mathbf{q} \right], \end{aligned} \tag{11}$$

where

$$\delta = \left[\frac{\lambda_w \eta}{(1-\lambda_w) + \lambda_w \eta} \right], \tag{12}$$

is the probability that a contact in agent i's network shares a social media feed. This probability is increasing in λ_w the intensity of engagement with social media, and also the endogenous variable $\eta \in \{0, 1\}$. The latter represents the probability that an individual receives message $m = 0$ on social media and trusts the message as legit. It is a function of the prior $\hat{\pi}^0$ and, since all individuals on the network share a common prior, η is either 1 (all messages $m = 0$ on social media are trusted as legit and thus shared) or 0 (all messages $m = 0$ are suspected to be fake and thus not shared). Either way, we know from Proposition 2 that if social media messages are shared, these messages claim $m = 0$.

The upshot of this filtering of information, and consequent dismissal of all tweets claiming $m = 1$, is that beliefs are biased because agents that share messages in support of Remain have influence in excess of the accuracy of their information. Notice also that, if δ is sufficiently large, it is possible for learning to collapse altogether. This result echoes the one by Acemoglu, Ozdaglar, and ParandehGheibi (2010), who show that when there are forceful agents in social networks misinformation may be a persistent outcome. However, in our framework the lack of efficient learning results from the inefficient filtering of information flows.

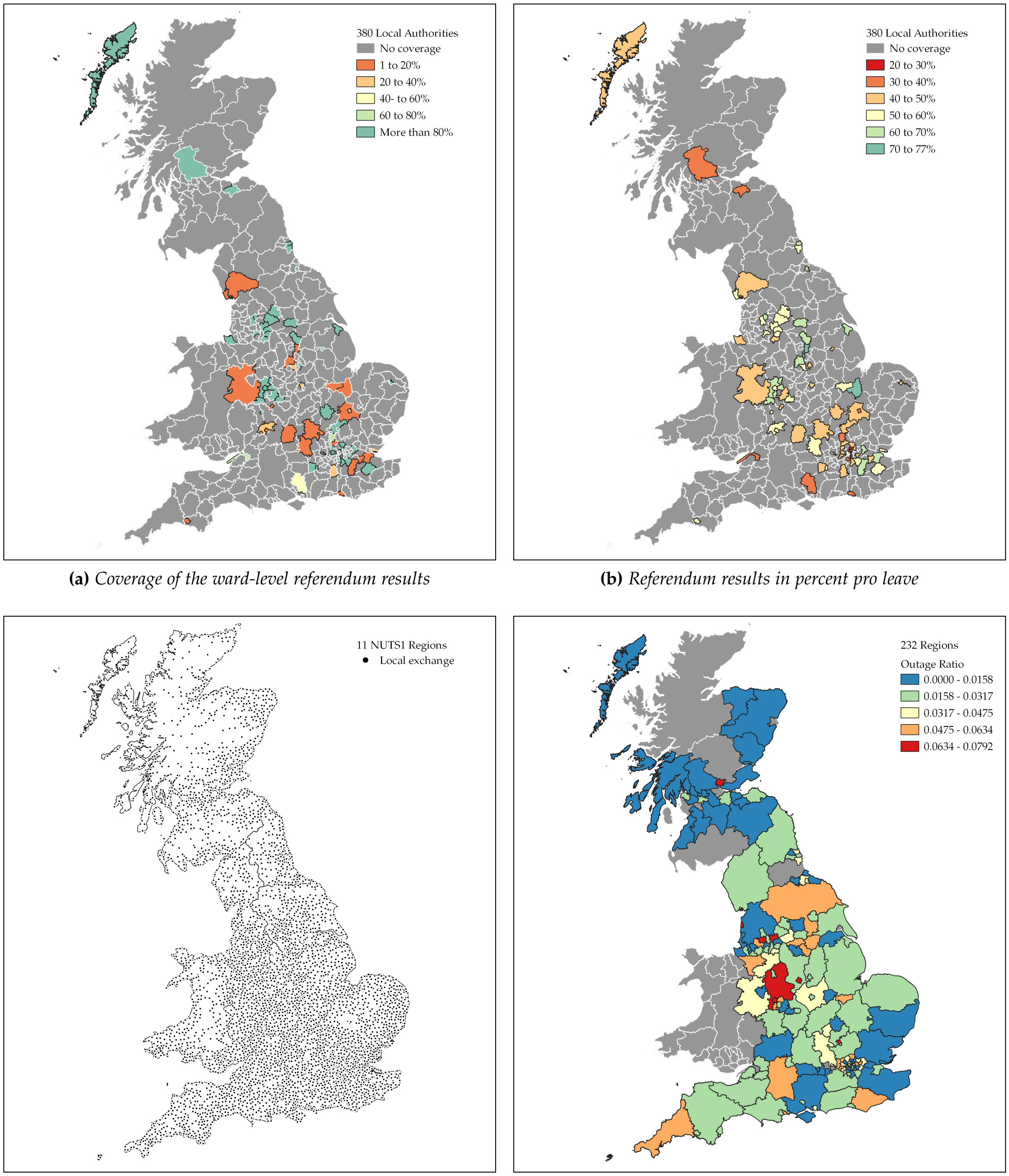
In the rest of the paper we study empirically the impact of Twitter usage on the 2016 Brexit referendum in light of the theoretical framework just laid out.

4 Data

4.1 EU referendum results

Official data on the EU membership referendum in the UK is released at the level of local authorities. To allow controlling for a variety of confounding factors and reduce unobserved heterogeneity, our analysis is at a more granular level, the ward. There are 9,456 wards in the UK. The BBC provides the absolute number of Leave and Remain votes and their respective shares for a sample of 1,283 wards.[7] A map displaying the geographic coverage of our main dependent variable, aggregated at the local authority level, can be found in Panel (a) of Figure 1. More specifically, it shows the share of wards in our sample for each local authority in the UK. Panel (b) reports our main dependent variable, the share of Leave votes, for the wards in our sample (again aggregated at the local authority level for visualization purposes). Our sample features a large variation in support for Leave and covers regions of England and Scotland.

[7] These data were obtained by the BBC under the Freedom of Information act, but were not possible to obtain when local authorities merged ballot papers from different wards before counting began. Given the lack of (geo-located) Twitter activity in five wards, and the log transformation of our Twitter Exposure variable, we work with 1,278 wards in our main analysis.



(a) *Coverage of the ward-level referendum results*

(b) *Referendum results in percent pro leave*

(c) *UK Internet Infrastructure*

(d) *Outage Ratio*

Figure 1: *Maps of the dependent variable and the instruments. Panel **(a)** shows the coverage of our ward-level referendum results within the UK. Panel **(b)** shows the outcome of the referendum as crude average over the ward results. Panel **(c)** shows the distribution of ADSL local exchanges. Panel **(d)** shows the frequency of Internet outages.*

4.2 Twitter data

A tweet is a short text message, which is by default visible to all other Twitter users.[8] Each message can also contain media, for example a link to a video, or a link to a website, and includes information on the date and time it was sent. To allow other users to search for tweets of their interest, users can also add one, or a list of hashtags, for example, by adding '#brexit' to a tweet they send, so it can be found by other users interested in that topic.

There are several ways of collecting data from Twitter. To understand their differences, let us briefly outline how Twitter, according to its documentation, processes a tweet.[9] A user can tweet using the App or the website. This information is then wrapped into a tweet object that carries the message (text, references to websites, pictures), auto-generated metadata, and a copy of information on the user at the time the tweet is sent (including the age of the account, the number of tweets, followers, and people this person follows). This object is then stored in at least two locations. First, in the network's database, so that users can now see this tweet and interact with it. Second, the tweet object is immediately archived. While the tweet in the network's database is dynamic and is altered upon interaction, the archived tweet object always remains in the state it was first stored in. For example, while a user can decide to delete a tweet on the network, the archive will still carry it. A lot of research has relied on accessing the dynamic data, which allows researchers for example to study directly how much a tweet was interacted with (for example Gorodnichenko, Pham, and Talavera, 2018). Twitter does not allow researchers to access dynamic data directly. Researchers interested in thousands or millions of messages would always have to rely on an application programming interface (API) to download these data, and Twitter's API comes with restrictions. The free API, used in the majority of related research, allows users to access Tweets only from the last seven days and only searches against a sample of tweets, which Twitter argues to be pre-filtered for "relevance not completeness".[10] The Premium API allows to download older tweets, but also here is not possible to select all tweets from a given time frame and location for download.

The alternative is to go back to (and rely on) the static archival data, which can be purchased from Twitter in its entirety using their Historical PowerTrack service. We follow this route and, thus, have purchased all archive entries that Twitter attributes to an UK origin using the revealed geographic coordinates of the tweets or based on Twitter's own algorithm, sent in the last two month preceding the EU Referendum (May and June 2016).

[8] The platform started with a 140 characters length limit.

[9] `https://developer.twitter.com/en/docs[Accessed11/01/2021]`

[10] `https://developer.twitter.com/en/docs/tweets/search/overview/standard` [Accessed 06/02/20].

Twitter delivered several hundred files in JSON format from their archive, containing 18,049,673 tweets in total.[11] From these raw data, we created three different data sets.

The first *tweet-level data set* includes all messages with geographic coordinates in the UK (which we refer to as geo-located), but also non-geo-located messages sent by an user who declared an address of residence in the UK. We report the summary statistics of this data set in Table 1 of the Appendix. From each tweet's metadata we extract whether a tweet is sent by a verified user and whether the tweet is a quote. From the timestamp of the tweet we obtain the week and hours of the day when the tweet was sent. We obtain the language of the tweet using Twitter's provided classification and create three dummies for the three most frequent languages.[12] To understand whether tweets are concerned with the European Union, and whether it supports Leave or Remain, we rely on a list of hashtags (Table A1 available in the Appendix). We generated this hashtag list by inspecting the 15,000 most used hashtags in our data set. We create a EU-related dummy that equals one to all tweets containing a hashtag in our list. All tweets that in addition allowed to infer a position from the hashtags were given either a Leave or a Remain dummy equal to one. Where questionable, we inspected individual tweets to understand their position. If hashtags from both lists were used, none of these two dummies was assigned a one.

The second *ward-level data set of geo-located tweets* is obtained from the subset of tweets which were exactly geo-located. We have 1,748,150 geo-located tweets for the UK. These tweets carry geographic longitude and latitude, and can be aggregated to the ward-level using GIS.[13] Since we want to measure the intensity of local Twitter usage by voters, and bots are clearly not voters, before aggregating tweets at the ward-level we leave out users suspected to be bots. In our benchmark analysis we simply remove the top 0.25% most active Twitter users from our sample.[14] We finally take the log of the ward-level count of tweets to obtain our *Twitter exposure* variable. Panel (a) of Figure 2, as a figurative example, shows these tweets sent from the Guildhall ward in the City of York. We then merge these data with the BBC data on ward level election results, and end up with 1,278 observations. The summary statistics of this data set is also shown in Table 1.

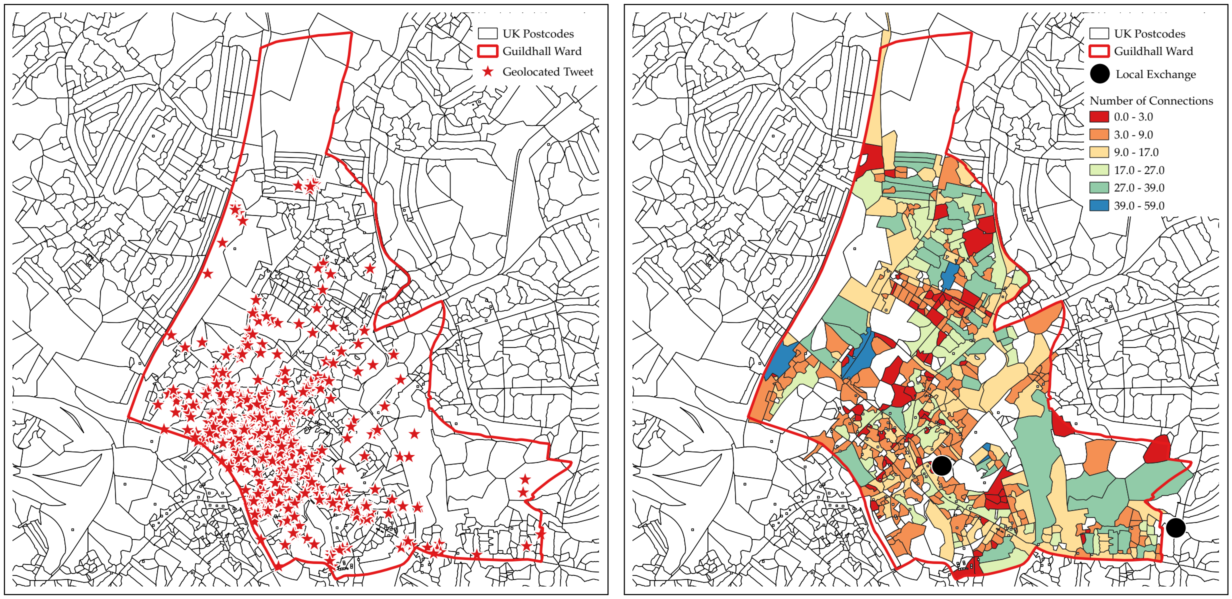
Finally, we generate a *user-level data set* counting 646,645 users. Using the raw time stamps provided by Twitter, we calculated the user's tweets per day. Following Gorodnichenko, Pham, and Talavera (2018) we also create several dummies which capture different features associated with bots (rather than human)

[11] JSON (JavaScript Object Notation) is an open standard file format to store and transmit data objects.

[12] More than 17 million tweets are classified as English, followed by Spanish with 102,555, and Portuguese with 40,311. Together they cover more than 99% of our tweets.

[13] It is important to note that for geographic operations on such a small-scale, the correct geographic referencing system is crucial. We relied on the British National grid (EPSG:27700) for all geographic operations.

[14] In Section 6.3 we show that adopting alternative criteria to identify bots does not affect our main results.



(a) *Geo-located tweets in this area*

(b) *Postcode-level connections and local exchanges*

Figure 2: *Map of the Guildhall ward in the City of York. The black lines show the boundaries of the postcodes.*

behavior, such as a dummy equal to one if a large number of tweets was sent between 0:00 and 06:00 in the morning (see Sections 6.3 and 6.4 for a more detailed discussion on bots definitions). Finally, to capture the relative position of the user on the EU referendum, we compute the share of pro Leave tweets as the number of pro Leave tweets over the sum of pro Leave and pro Remain tweets.

4.3 Instrumental variables

Our main identification strategy exploits the fact that the location of the current 5,564 UK asymmetric digital subscriber line (ADSL) local exchanges is predetermined by the historical location of the old telephone exchanges. The UK's ADSL architecture is organized via local exchanges (LEs), displayed in Panel (c) of Figure 1, which are themselves connected to the internet backbone, and connect households via several relays.

In order to gain a precise measure of the distance between the average internet user and its LE, we relied on connection data from the UK's regulatory body, the Office of Communications (Ofcom). These data allow us to link the 19,487,073 UK households with internet access to a given postcode. The UK has 1,69 million alphanumeric postcodes, on average only 1.6 km^2 in size. Panel (b) in Figure 2 displays the number of connections per postcode for the Guildhall ward of the city of York. The map also shows the LEs. For each postcode, we calculate the "as-the-crow-flies" distance to the closest LE, and then aggregated these measures to obtain a 'connection-weighted distance' per ward. Thus, for each electoral ward we obtain an appropriately weighted (by the number of connections) average distance to LE. This constitutes our first instrumental variable.

Our second identification strategy exploits abnormal internet outages in the weeks preceding the EU Referendum. The Internet Outage and Detection Analysis Project from the Center for Applied Internet Data Analysis of UC San Diego (CAIDA) provides high frequency data from a network telescope.[15] It relies on internet traffic unknown to most internet users to identify outages. Apart from signals exchanged between computers on the web, for example websites, streams, or professional data, the internet has a (relatively constant) background noise from viruses, scams, and wrongly addressed IP packages. The telescope exploits this noise by accepting incoming packages that are addressed to computers that should have technically never been contacted. The source of these falsely addressed packages (either wrongly configured or that have fallen prey of attackers giving a false return address) share their own IP address, which can be geo-located with some precision. CAIDA aggregates these data in two-hour periods for each region.

We use these data to construct an outage measure, by assuming that large drops in the amount of noise coming from one of these CAIDA regions proxies for network outages in that location. We proceed in two steps. First, we regress our traffic volume data on a set of all (two-hour) time and location fixed effects. We then use the residuals of this regression v_i, to detect abnormal traffic drops. More precisely, for each region i we compute the share of periods in which v_i drops by more than one local standard deviation. More formally,

$$\text{Outages}^i = \sum_{t=2}^{T} \frac{\mathtt{I}\left(v_t^i - v_{t-1}^i < -\sigma^i\right)}{T},$$

with σ^i the standard deviation of v_i in region i, T the number of periods for which the internet traffic is observed and $\mathtt{I}\left(\bullet\right)$ the indicator function. Panel (c) of Figure 1 shows the distribution of this ratio across the UK.

To combine these data with our ward-level data set, we overlay the CAIDA data with the wards in GIS and attribute to the wards the intersection-weighted average of the outage data. Even though CAIDA's regions are in many instances orthogonal to electoral divisions, the vast majority of wards falls within only one CAIDA region. For the rest, we rely on an average weighted by shared area.[16]

[15] `https://www.caida.org/projects/network_telescope/`.

[16] Consider that ward A shares ten percent of its geographic area with CAIDA area Y, and 90 percent of its area with CAIDA area Z. The outage ratio of A is then calculated by adding ten percent of Y's outage ratio to 90 percent of Z's outage ratio.

4.4 Additional controls

We gather data from a variety of sources to control for potential confounding factors. From British Ordnance Survey maps we calculate (log) area, a proxy for the distance to the equator, and the rainfall on the 23rd June 2016, the day of the referendum, using GIS.

Demographic and socio-economic control variables are obtained by aggregating the 2011 Census data from the Census output areas (the smallest geographical unit used by the Census) at the ward-level. These control variables include (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. It also includes information on the economic structure of the workforce, and in particular, the share of the population employed in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary.

The United Kingdom Independence Party (UKIP) vote share obtained in the 2014 EU elections (available at the local authority) and the UKIP vote share in the latest local elections preceding the EU referendum (available at the electoral ward-level) are obtained from the Democratic Dashboard established by the Democratic Audit, which is a research team based in the London School of Economics studying the electoral contests in the UK.

In Table 1 we also report the summary statistics for our instrumental variables, and these additional control variables.

5 Empirical Strategy

5.1 Social Media and the Geography of the Brexit Vote

According to the theoretical framework presented in Section 3, if fake news was predominantly tilted towards supporting Leave (as we assume in the model and show in more detail in Section 6.4), bounded rational voters more exposed to Twitter are less likely to engage with messages that support Leave, because they dismiss such messages as fake news. As an upshot, a greater exposure to Twitter is predicted to increase the support for Remain. A basic model to test this hypothesis is the following

$$Leave_{wl} = \alpha_l + \beta \; Twitter \; Exposure_{wl} + X^{'}_{wl}\gamma + \epsilon_{wl} \tag{13}$$

Table 1: *Descriptive statistics*

Variable	Observ.	Mean	St. Dev.	Min	Max
Tweet-level data set:					
EU-related tweet (dummy)	18,696,318	0.012	0.111	0	1
Pro Leave (dummy)	18,696,318	0.002	0.044	0	1
Pro Remain (dummy)	18,696,318	0.003	0.054	0	1
Change in number of followers	18,696,318	0.947	99.28	-139,044	156,656
User is verified (dummy)	18,696,318	0.008	0.087	0	1
Quoting another tweet (dummy)	18,696,318	0.082	0.275	0	1
Number of followers before tweet (in thousand)	18,696,318	2.469	30.35	0	11,748
Tweet was sent before 06:00	18,696,318	0.057	0.231	0	1
Ward-level data set:					
Share of Leave votes (in %)	1,278	52.28	14.29	12.16	82.51
Twitter Exposure	1,278	4.15	1.33	0	8.74
LE Distance (in km)[c]	1,278	1.57	0.82	0.26	5.90
Outages[d]	1,278	0.03	0.02	0.01	0.07
Log area (in square km)	1,278	1.43	1.05	-0.75	7.33
Distance from the Equator (in km)	1,278	5,811.93	134.55	5,579.98	6,477.04
Rain on June 23rd 2016[f]	1,278	4.12	5.59	0	36.53
Log population (in 1,000)	1,278	9.25	0.50	7.56	10.55
Share of females (in %)	1,278	50.64	1.55	35.72	55.30
Age Group 15–19 (in %)	1,278	5.93	1.56	2.30	29.34
Age Group 20–29 (in %)	1,278	13.87	6.07	5.27	67.92
Age Group 30–39 (in %)	1,278	13.74	3.73	6.24	27.17
Age Group 40–49 (in %)	1,278	13.85	1.61	3.71	20.85
Age Group 50–59 (in %)	1,278	12.70	2.25	3.54	19.31
Age Group 60–89 (in %)	1,278	20.72	6.55	5.30	43.14
Share of Managers	1,278	0.10	0.03	0.04	0.29
Share of Professionals	1,278	0.18	0.08	0.03	0.56
Share of Associate Professionals	1,278	0.13	0.04	0.05	0.33
Share of Administrative	1,278	0.12	0.03	0.06	0.29
Share of Trade	1,278	0.11	0.04	0.02	0.46
Share of Caring	1,278	0.10	0.03	0.04	0.28
Share of Sales	1,278	0.09	0.03	0.02	0.26
Share of Industry	1,278	0.08	0.04	0.01	0.23
Share of Elementary	1,278	0.12	0.05	0.03	0.33
Share of UKIP in the latest local election (in %)	1,278	9.45	9.50	0	52.21
Share of UKIP in the 2014 EU Parliament election (in %)	1,278	27.18	9.23	7.10	47.30

The data on ward-level referendum results can be accessed from bbc.co.uk/news/uk-politics-38762034, where a brief discussion of the methodology is also available [Last Accessed: 07.02.2020]. The coordinates of the geo-located tweets were transformed to the British National Grid coordinate system (EPSG:27700) and then intersected with the boundaries of UK wards from Boundary-Line™, which can be downloaded from ordnancesurvey.co.uk/business-government/products/boundaryline [Last Accessed: 12.05.20]. The local exchanges were gathered with data from and located with data from availability.samknows.com/broadband/exchange_search [State: 15.11.2018] and then aggregated to wards by connection with postcode-level data from ofcom.org.uk/research-and-data/multi-sector-research/infrastructure-research/connected-nations-2017/data-downloads [State: June 2016] A map is provided in Figure 1. Data from network telescope at the Center for Applied Internet Data Analysis of UC San Diego (caida.org). All calculations were run on data available in or transformed to the British National Grid coordinate system (EPSG:27700). Where regressions indicate distance to the equator, this is proxied by the coordinate system's abscissa. These numbers here are for reference only. Rainfall data were calculated in GIS using data from the Centre for Environmental Data Analsyis which can be downloaded from https://catalogue.ceda.ac.uk/uuid/87f43af9d02e42f483351d79b3d6162a Occupational shares are sourced from the UK ONS official labor market statistics (NOSIS). The data are obtained from the 2011 Census and are available at the very granular census output areas. Finally historical electoral data is from the Democratic Dashboard (Democratic Audit, Department of Government at the London School of Economics), and can be downloaded from https://democraticdashboard.com/data [Last Accessed: 10.12.2020].

where $Leave_{wl}$ is the share of votes cast in support of leaving the EU in ward w and local authority l, α_l is a set of local authority dummies capturing any unobserved heterogeneity at that geographic level, and $Twitter\ Exposure_{wl}$ is the (log) total number of tweets originated in ward w during the two months preceding the EU referendum, which proxies for the exposure to Twitter at the local level. Since we are interested in a measure of voters' exposure to Twitter, we exclude tweets originating from bots. A simple way to do so is to exclude the top 0.25% most active Twitter users before aggregating tweets at the ward level. In the section 6.3 we adopt several alternative criteria to exclude bots.

We start estimating this relatively parsimonious model, and then add gradually four sets of controls, included in the vector X'_{wl}. More specifically, we first add two basic geographic controls, (log) area and distance from the equator. Additionally, we also include local rainfall on the day of the referendum among the geographic controls, as it has been shown to influence turnout significantly (Gomez, Hansford, and Krause, 2007; Fujiwara, Meng, and Vogl, 2016; Lind, 2020). We then add demographic controls including: (log) population, the proportion of female population, and the share of population in the following age groups [15–19], [20–29], [30–39], [40–49], [50–59], [60–89]. The third set of variables, economic controls, include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Finally, political controls include the ward level UKIP vote share in the latest local elections (and the local authority level UKIP vote share in 2014 EU elections in the specifications without local authority fixed effects) following the logic that, at least before 2016, UKIP was the only major party in the UK with the explicit goal of leaving the European Union. This set of controls have been shown to explain the outcome of the EU Referendum well (Becker, Fetzer, and Novy, 2017).

The parameter of interest is β which estimates the impact of exposure to Twitter on the local support for Leave. Although equation 13 can be estimated by OLS, this is unlikely to lead to consistent estimates of the impact of Twitter on the referendum vote since, for instance, areas with greater Twitter use may systematically be more liberal or more conservative, despite our large set of control variables. In other words, we may still miss important factors driving both Twitter use intensity and the local support for Leave. This would lead to a classical omitted variable bias in the estimation of β. Similarly, although we have access to the universe of tweets generated in the two months previous to the referendum, we may still measure Twitter exposure with some error. We therefore turn to two alternative instrumental variable strategies, both proposed in the related literature (Falck, Gold, and Heblich, 2014; Geraci et al., 2018; Müller and Schwarz, 2020).

We first adopt as an instrumental variable the average distance of the ward level internet connections to the nearest local exchange (denoted $LE\ distance_{wl}$, hereafter). The identifying assumptions in this case is that the distance to local exchange affects the quality and speed of the broadband connections (and therefore the exposure to Twitter), but is not correlated with any other characteristic which may explain voting behaviour. The first assumption is formally tested in our first stage. The exclusion restriction is supported by the crucial argument proposed by Falck, Gold, and Heblich (2014) for Germany and further developed by Geraci et al. (2018) for the UK: the location of the local exchanges mainly follows cost minimisation criteria prevailing at the time of the development of the telephone grid. Crucially, the length of the wire between the user and the local exchange does not affect the quality of phone calls. Hence, we do not expect any link between the distance from local exchange and any socio-economic characteristic of the users. The length of the wire is instead one of the major determinants of internet connection's quality and speed.

The second instrumental variable strategy relies on local internet outages ($outages_c$). We recorded the number of outages, defined as a drop by more than a standard deviation of local traffic over any two hours window with respect to the previous two hours (after controlling for time and localities fixed effects). While the exogeneity of this second instrument is extremely plausible, and the exclusion restriction seems less demanding, it comes at the cost of having the variable defined at a less disaggregated level which does not allow a within local authority analysis. Formally, in our first stages we estimate the following equations

$$\begin{aligned} Twitter\ Exposure_{wl} &= \delta_l + \rho\ LE\ distance_{wl} + X'_{wl}\omega + \varepsilon_{wl} \qquad (14) \\ Twitter\ Exposure_{wa} &= \delta_a + \rho_a\ Outages_{ca} + X'_{wa}\omega + \varepsilon_{wa} \end{aligned}$$

where δ_a are now a set of fixed effects for NUTS1 areas in the UK. In Section 6.2 we propose a large number of falsification tests, each broadly supporting the validity of our exclusion restrictions.

6 Results

6.1 Twitter usage and the Brexit vote

Table 2 reports the results of estimating equation 13 by simple OLS. The specification in the first column controls only for the Exposure to Twitter and local authority fixed effects. Twitter exposure is predicted to reduce the support to Leave in the EU referendum. The further columns add progressively geographic, demographic, economic, and political controls. The magnitude of the effect reduces in more loaded specifications, but the coefficient of interest remains negative and statistically significant. There are good reasons, as previously discussed, to believe that the results in Table 2 are biased.

Table 2: *The impact of Twitter exposure on the Brexit vote - OLS*

Dependent variable: Share of Leave votes	(1)	(2)	(3)	(4)	(5)
Twitter Exposure	-2.49***	-2.57***	-1.49***	-0.45***	-0.42**
	(0.37)	(0.38)	(0.25)	(0.17)	(0.17)
Local authority fixed effects	✓	✓	✓	✓	✓
Geographic controls		✓	✓	✓	✓
Demographic controls			✓	✓	✓
Economic controls				✓	✓
Political controls					✓
Observations	1,278	1,278	1,278	1,278	1,278
R-squared	0.75	0.75	0.79	0.90	0.90

Notes: Twitter Exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. We exclude the top 0.25% users to exclude bots' activity. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

In Table 3 we report the results of our first IV strategy, adopting the distance to the local exchange as the instrument for Twitter Exposure. The first stage generally passes the standard tests. We nevertheless report in square brackets the Anderson and Rubin weak instrument robust 95% confidence intervals. The results broadly confirm the negative and significant effect of Twitter Exposure on the support for Leave. The magnitude is significantly larger than in the OLS estimates, confirming the potential endogeneity concerns. According to the specification in column (5), a 1% increase in Twitter Exposure leads to a decrease by 5.3% in the Leave vote share. Alternatively, one standard deviation increase in Twitter Exposure leads to 0.5 standard deviation decrease in the local support for Leave.[17]

The results of our second IV strategy, relying on outages, are reported in Table 4.[18] It is important to notice that the variation we are exploiting here is not directly comparable to the previous results, as the outage information varies across CAIDA regions rather than across wards. We therefore replace local authority fixed effects with NUTS1 fixed effects in the model. The first stage weakens to some extent, which further justifies the use of Anderson and Rubin weak instrument robust 95% confidence intervals. Results are qualitatively identical, confirming that higher exposure to Twitter reduced the support for Leave. The

[17] First stage results are reported in the Appendix, in Table A2.

[18] Again, first stage results are reported in the Appendix, in Table A3.

Table 3: *The impact of Twitter exposure on the Brexit vote - IV: Distance from a LE*

Dependent variable: Share of Leave votes	(1)	(2)	(3)	(4)	(5)
Twitter Exposure	-9.44***	-7.67***	-9.03***	-5.85**	-5.28**
	(2.11)	(1.41)	(2.36)	(2.28)	(2.16)
	[-15.41, -6.26]	[-11.28,-5.27]	[-17.23,-5.07]	[-19.40,-1.94]	[-17.85,-1.55]
Local authority fixed effects	✓	✓	✓	✓	✓
Geographic controls		✓	✓	✓	✓
Demographic controls			✓	✓	✓
Economic controls				✓	✓
Political controls					✓
Observations	1,278	1,278	1,278	1,278	1,278
Kleibergen-Paap F-stat	9.737	16.30	15.36	7.707	7.538

Notes: IV estimates using the distance from LE as the instrument for Twitter Exposure. Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. We exclude the top 0.25% users to exclude bots' activity. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in five different age groups. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. Anderson Rubin weak instrument-robust 95% confidence intervals reported in squared brackets. *** p<0.01, ** p<0.05, * p<0.1.

coefficients for Twitter Exposure are larger than the corresponding coefficients in Table 3. It is worth noting, however, that the results in the most loaded specifications of both IV strategies considered – reported in column 5 in Tables 3 and 4 – are of similar magnitude.

Overall, the evidence strongly favours Twitter having a causal negative impact on the share of Leave vote. Through the lenses of the theoretical framework in Section 3, this result suggests that Leave supporting messages on Twitter are disproportionately perceived as being fake news. Consequently, Twitter users are more likely to discard Leave campaign material as illegitimate expertise and, thus, are more likely to vote for Remain. In the next sections, we first present a number of falsification tests to support our IV strategies. Next, we show the robustness of our main results when adopting different alternative approaches and assumptions. Finally, we explore more closely the empirical plausibility of the theoretical mechanism we have suggested. In particular, we use the 18 million tweets originating from the UK in the two months preceding the EU Referendum to provide evidence of a larger presence of fake news on the Leave field, and to study the reaction of Twitter users to specific tweeting activities related to the EU referendum.

Table 4: *The impact of Twitter exposure on the Brexit vote - IV: Internet Outages*

Dependent variable: Share of Leave votes	(1)	(2)	(3)	(4)	(5)
Twitter Exposure	-18.83*** (7.19) [-30.09,-13.59]	-19.47** (7.89) [-32.15,-13.88]	-26.82*** (9.66) [-60.82,-16.66]	-19.33** (8.03) [-53.55,-11.17]	-9.87** (4.95) [-27.80,-5.11]
NUTS1 fixed effects	✓	✓	✓	✓	✓
Geographic controls		✓	✓	✓	✓
Demographic controls			✓	✓	✓
Economic controls				✓	✓
Political controls					✓
Observations	1,278	1,278	1,278	1,278	1,278
Kleibergen-Paap F-stat	4.680	3.973	8.521	7.878	7.157

Notes: IV estimates using an aggregate measure of Internet outages in the two months preceding the referendum as the instrument for Twitter Exposure. Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. We exclude the top 0.25% users to exclude bots' activity. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections, and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. Anderson Rubin weak instrument-robust 95% confidence intervals reported in squared brackets. *** p<0.01, ** p<0.05, * p<0.1.

6.2 Falsifications

The results in Tables 3 and 4 rely on the assumption that the two instruments are exogenous. More specifically, our exclusion restriction is valid only if distance to the local exchange (and outages) do not affect voting behaviour other than through the frequency of Internet use and the exposure to social media. Intuitively, outages seem plausibly exogenous, whereas the distance to local exchange may generate some concerns: may better "connected" wards be systematically richer or feature a better educated or more liberal electorate? The arguments presented in section 5.1 in support of our exclusion restrictions are solid and convincing. Nevertheless we report in Table 5 a series of falsification tests, which further support the validity of the exclusion restrictions. For all tests we report the results of estimating the most complete specification of the model, in which all controls are included.

First, we collect information relative to 2002–3 UK local elections at the ward level, and show that both our instruments are not systematically correlated with voting patterns in these pre-social media elections.[19] Since a good deal of wards in our sample were reshaped over the last two decades, we run two alternative tests: in Panel A we reconstruct 2002–3 local election results for today's wards by weighting the relative results of past wards by their relative area in the current wards; in Panel B we instead run the falsification

[19] Twitter and Facebook were launched in 2006 and 2004, respectively.

test only on the subset of wards which remained unchanged over the period considered (about 2/3 of the total sample). Importantly, we do not include our UKIP support controls when implementing these tests. No regular pattern emerges, and crucially UKIP support – the single most influential political force behind the EU referendum, counting about 20,000 members (and half million voters) in 2004 (Abedi and Lundberg, 2009) – is not correlated with either of our instruments.

Next, we locate all Waitrose retail centers in the UK. These generally appeal to a richer, more liberal and more green-minded population (similar to Whole Food in the US). We then test whether the distance to local exchange (outage) has any explanatory power for the location of Waitrose. In Panel C, we report four alternative sets of regression results, in which the dependant variables are: the distance from the closest Waitrose, and dummies for wards located between *1 and 2*, *1 and 3*, and *1 and 4* km from Waitrose, in the logic of the doughnut approach, i.e. assuming that Waitrose may be located in a cluster of retail-oriented buildings, not suitable for residential use, but close to wards rich in potential customers. Results in Panel C do not suggest any correlation between the location of Waitrose and our instruments, which is reassuring.

Finally, we gathered ward level data on house prices in the UK for 2000, 2007, 2014, and for the year of the referendum 2016, and test whether there is any systematic relationship between them and the distance from the local exchange, or the number of internet outages in the months preceding the referendum. Once more, no clear pattern seems to emerge. Overall, the results in Table 5 support the validity of our IV strategies, and the two statistically significant coefficients (out of 32 regressions) are compatible with randomness.

6.3 Robustness tests

In the baseline analysis we exclude the 0.25% top most active Twitter users as a way of excluding bots from our measure of Twitter exposure. In Table 6 we replicate our main results, including our results from OLS and from both IV strategies, excluding the top 0.5% (Panel A) and the top 0.1% most active Twitter users (Panel B), and excluding bots identified following several alternative criteria proposed in Gorodnichenko, Pham, and Talavera (2018): users with more than 10 tweets in any day (Panel C); users with more than 25 tweets in any day (Panel D); users with more than 10 tweets in any day between 00:00 and 06:00 (Panel E); users with more than 25 tweets in any day between 00:00 and 06:00 (Panel F); users with more than 10 identical tweets in any day (Panel G); users with more than 25 identical tweets in any day (Panel H); and users whose account has been set up after the EU referendum announcement, hence potentially created with the explicit purpose of conditioning the related campaign (Panel I). All approaches produce

Table 5: *Falsification tests*

Instrument:	LE distance				Internet outage			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Pre-social media local elections (Reconstructed wards)								
Dependent variable:	UKIP 2002–3	Tories 2002–3	Labour 2002–3	Libdem 2002–3	UKIP 2002–3	Tories 2002–3	Labour 2002–3	Libdem 2002–3
Instrument	0.0001	0.015	0.0008	-0.005	0.01	0.68	1.35**	-0.31
	(0.0004)	(0.012)	(0.010)	(0.012)	(0.03)	(0.71)	(0.65)	(0.86)
Panel B: Pre-social media local elections (Only unchanged wards)								
Dependent variable:	UKIP 2002–3	Tories 2002–3	Labour 2002–3	Libdem 2002–3	UKIP 2002–3	Tories 2002–3	Labour 2002–3	Libdem 2002–3
Instrument	0.0001	0.0062	0.0080	-0.0051	-0.01	0.23	1.41	-0.67
	(0.0003)	(0.0143)	(0.0156)	(0.0159)	(0.03)	(0.73)	(0.86)	(1.05)
Panel C: Distance from Waitrose								
Dependent variable:	Dist. Waitrose	Waitrose 1-2	Waitrose 1-3	Waitrose 1-4	Dist. Waitrose	Waitrose 1-2	Waitrose 1-3	Waitrose 1-4
Instrument	-0.17	0.01	0.02	0.02	-45.53	0.50	-0.20	-0.20
	(0.19)	(0.02)	(0.02)	(0.02)	(56.26)	(0.54)	(0.95)	(0.95)
Panel D: Housing prices								
Dependent variable:	House pr. 2000	House pr. 2007	House pr. 2014	House pr. 2016	House pr. 2000	House pr. 2007	House pr. 2014	House pr. 2016
Instrument	0.01	0.02	0.03	0.05*	0.37	0.46	1.47	0.38
	(0.01)	(0.01)	(0.02)	(0.02)	(0.68)	(0.90)	(2.01)	(2.52)
Local authority fixed effects	✓	✓	✓	✓				
NUTS1 fixed effects					✓	✓	✓	✓
All controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Correlations between our instruments and: electoral results of the four main parties in the 2002–3 local elections, pre-dating the launching of social media (Panel A & B, with the latter using only wards which did not change over the period considered); distance from Waitrose (and three alternative doughnut dummies for wards located between 1 and 2 km, 1 and 3 km and 1 and 4 km - Panel C); and housing prices in 2000, 2007, 2014, and in the year of the referendum (Panel D). Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15–19], [20–29], [30–39], [40–49], [50–59], [60–89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls (not included in models in Panel A & B) include the UKIP vote share in 2014 EU elections (in columns 5-8 only), and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

qualitatively identical results and with remarkably similar magnitudes.

In the Appendix we also run our analysis on the full tweets dataset, without excluding bots. The results, reported in Panel A of Table A4, confirm that more intense Twitter exposure reduced the support for Leave, although the most demanding specification for the IV strategy relying on the distance from local exchange is no longer statistically significant when adopting the standard criteria. Secondly, by taking the log of our measure of Twitter Exposure, we drop five wards for which we have the EU referendum results but no tweets have been generated during the two months preceding the EU referendum. In Panel B of Table A4, we replicate the main analysis using the widespread transformation $log(1 + Tweets)$. Unsurprisingly, the results are almost identical.

Our analysis is based on the universe of tweets generated in the UK in the two months preceding the referendum. Retweets are, however, absent from our Twitter Exposure, as they are not geo-located. However, since all tweets in the archive contains user-related information including the user's total amount of tweets, we can calculate the difference between the total number of tweets between tweets. By attributing this change to retweeting activity, we can estimate a measure of retweets per ward, and create an alternative measure of Twitter Exposure including retweets. More specifically, we compute a per day retweeting activity for each user in our sample, and multiply the result by 62 days preceding the referendum, before aggregating retweets at the ward level. We then estimate the main models with this new measure. Panel C of Table A4 summarizes the results of this alternative strategy and broadly confirms our findings. A slight complication in the retweets-based analysis is that users can delete tweets from their account, which would potentially bias our estimated number of retweets, as it would produce negative retweets for users regularly deleting tweets. We address this concern by testing whether the presence of such users is in any way correlated with the support for Leave in the EU referendum. Results in Table A5 do not substantiate these concerns. In any case, the inevitable inaccuracies in estimating retweeting activity in the two months preceding the referendum by extrapolating from the users' activity between observed tweets and the relatively low Kleibergen Paap F-statistic for some of the specifications of the related models suggest caution in interpreting these results.

6.4 Mechanism

The robust negative relationship between Twitter Exposure and Leave support is consistent with the mechanism proposed by our model. In this section, we propose three alternative strategies to strengthen the case for our mechanisms. First, since our model suggests that the presence of fake news on social media constrains “learning” among voters, we directly test whether voters located in wards more exposed to Twitter were featuring lower learning. We create a dummy variable capturing the 15% wards featuring lowest learning among voters, defined as the absolute value of the difference between the support for Leave at the EU referendum and the support for UKIP (a party created in the early 1990s with the explicit goal of leaving the EU) in the latest 2014 EU elections. The underlying assumption is that without any new information and absent any learning, voters would have remained of the same opinion they had in 2014. We then replicate our analysis testing whether Twitter Exposure had an impact on the probability of low learning.

The results, reported in Table 7, are consistent with the mechanisms described in the model: wards featuring larger exposure to Twitter were more likely to display low learning. In other words, voters more exposed to Twitter were less likely to change their opinion over time. In Table A6 in the Appendix, we show that replicating the analysis with two alternative measures of learning (a dummy for the bottom 30% wards in terms of learning in Panel A, and a continuous measure of learning in Panel B), produce patterns similar to the ones presented in Table 7.

We then exploit our rich database to test more directly whether messages supporting Leave are perceived as less trustworthy by Twitter users. We propose two complementary strategies. First, relying on the fact that bots are often associated with the spread of fake news, we test whether users tweeting more pro Leave are more likely to be bots, thereby lending support to the assumption proposed in our model. This part of the analysis is therefore run at the user level. We proceed in several steps. We first create a dummy, denoted by *EU-related Tweet*, which identifies all tweets including a list of hashtags related to the EU. We then create two dummy variables, *Pro Remain Tweet* and *Pro Leave Tweet*, capturing tweets featuring hashtags in support of Remain and Leave, respectively (Table A1 in the Appendix lists all hashtags used). For each Twitter user we measure the pro Leave orientation on the EU referendum by the share of pro Leave tweets, computed as the number of pro Leave tweets over the sum of pro Leave and pro Remain tweets. The dependent variable is a dummy identifying users who display features characterizing bots (*Bot dummy*).

Table 8 reports the results of this exercise, where we also include controls for verified users (dummy), and the number of their followers. We adopt several alternative criteria to identify bots. In column (1) the dummy for bots identifies the top 0.25% most active Twitter users, the simple criterion adopted in the benchmark model of our ward-level analysis. In the resting columns we follow again the bots-identifying criteria proposed by Gorodnichenko, Pham, and Talavera (2018): users with more than 10 tweets in any day (column 2); users with more than 25 tweets in any day (column 3); users with more than 10 tweets in any day between 00:00 and 06:00 (column 4); users with more than 25 tweets in any day between 00:00 and

Table 6: *Robustness tests on the impact of Twitter exposure on the Brexit vote*

Dependent variable: Share of Leave votes									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: excluding top 0.5% most active users									
Twitter Exposure	-2.55***	-1.53***	-0.40**	-8.75***	-8.13***	-4.63**	-18.53***	-25.93***	-10.02**
	(0.36)	(0.24)	(0.17)	(1.82)	(1.98)	(1.86)	(6.95)	(9.06)	(4.77)
Panel B: excluding top 0.1% most active users									
Twitter Exposure	-2.27***	-1.29***	-0.29*	-9.94***	-9.99***	-6.73*	-19.25**	-28.41***	-10.24*
	(0.37)	(0.25)	(0.16)	(2.34)	(3.11)	(3.90)	(7.78)	(10.63)	(5.58)
Panel C: Excluding users with more than 10 tweets in any day									
Twitter Exposure	-2.61***	-1.61***	-0.43***	-8.80***	-8.08***	-4.37**	-19.27***	-28.02***	-12.56*
	(0.36)	(0.25)	(0.16)	(1.83)	(1.98)	(1.90)	(7.16)	(10.13)	(7.08)
Panel D: Excluding users with more than 25 tweets in any day									
Twitter Exposure	-2.46***	-1.57***	-0.48***	-9.31***	-8.74***	-5.17**	-19.95**	-29.49***	-11.59*
	(0.36)	(0.26)	(0.16)	(1.90)	(2.09)	(2.38)	(8.13)	(11.29)	(6.53)
Panel E: Excluding users with more than 10 tweets in unusual time (00:00-06:00) in any day									
Twitter Exposure	-2.36***	-1.38***	-0.37**	-9.34***	-9.15***	-5.54**	-17.62***	-24.65***	-8.79**
	(0.35)	(0.26)	(0.18)	(2.11)	(2.63)	(2.66)	(6.12)	(9.14)	(4.35)
Panel F: Excluding users with more than 25 tweets in unusual time (00:00-06:00) in any day									
Twitter Exposure	-2.31***	-1.29***	-0.32*	-9.76***	-9.83***	-6.53*	-19.17**	-28.74**	-10.15*
	(0.37)	(0.27)	(0.18)	(2.25)	(2.82)	(3.41)	(7.56)	(11.37)	(5.40)
Panel G: Excluding users with more than 10 identical tweets in any day									
Twitter Exposure	-2.64***	-1.77***	-0.36**	-9.56***	-8.80***	-5.99**	-19.57***	-28.60***	-12.02*
	(0.35)	(0.27)	(0.16)	(2.06)	(2.14)	(2.99)	(7.06)	(9.93)	(6.38)
Panel H: Excluding users with more than 25 identical tweets in any day									
Twitter Exposure	-2.69***	-1.78***	-0.42**	-9.01***	-8.20***	-5.09**	-20.28**	-31.98***	-13.70*
	(0.37)	(0.26)	(0.17)	(1.68)	(1.84)	(2.22)	(7.96)	(11.98)	(7.38)
Panel I: Excluding users whose account was created after EU referendum announcement									
Twitter Exposure	-2.29***	-1.33***	-0.37**	-10.21***	-10.52***	-7.20	-18.93***	-23.00***	-7.50*
	(0.36)	(0.26)	(0.17)	(2.56)	(3.74)	(5.09)	(7.00)	(8.31)	(3.99)
Local authority fixed effects	✓	✓	✓	✓	✓	✓			
NUTS1 fixed effects							✓	✓	✓
Geographic controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Economic controls			✓			✓			✓
Political controls			✓			✓			✓
Estimation method	OLS			IV: Distance from LE			IV: Internet Outage		
Observations	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278

Notes: Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day when excluding bots, defined as: top 0.5% most active users (Panel A); top 0.1% most active users (Panel B); users with more than 10 tweets in any day (Panel C); users with more than 25 tweets in any day (Panel D); users with more than 10 tweets in the period 00:00-06:00 in any day (Panel E); users with more than 25 tweets in the period 00:00-06:00 in any day (Panel F); users with more than 10 identical tweets in any day (Panel G); users with more than 25 identical tweets in any day (Panel H); users whose account has been open after the EU referendum announcement (Panel I). Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15–19], [20–29], [30–39], [40–49], [50–59], [60–89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 7-9 only), and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: *The impact of Twitter exposure on "learning" about the Brexit vote*

Dependent variable: Low learning dummy (bottom 15%)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Twitter Exposure	0.06***	0.03**	0.001	0.33***	0.30***	0.22*	0.32***	0.35**	0.31*
	(0.01)	(0.01)	(0.01)	(0.08)	(0.10)	(0.12)	(0.11)	(0.14)	(0.16)
				[0.20, 0.57]	[0.14, 0.62]	[0.01, 0.84]	[0.20, 0.53]	[0.15, 0.86]	[0.10, 0.94]
Local authority fixed effects	✓	✓	✓	✓	✓	✓			
NUTS1 fixed effects							✓	✓	✓
Geographic controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Economic controls			✓			✓			✓
Political controls			✓			✓			✓
Estimation method		OLS			IV: Distance from LE			IV: Internet Outage	
Observations	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278
Kleibergen-Paap F-statistic				9.737	15.36	7.538	4.680	8.521	7.157

Notes: The dependent variable is a dummy equal one for the bottom 15% wards in terms of learning, defined as the absolute change between the Leave and the UKIP vote shares at the 2014 EU elections. Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15–19], [20–29], [30–39], [40–49], [5-0-59], [60–89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 7-9 only), and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. For the IV estimates in column 4-9, Anderson Rubin weak instrument-robust 95% confidence intervals are reported in squared brackets. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8: *Are Pro Leave users more likely to be bots?*

Dependent variable: Bot dummy	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Using the full sample of Twitter users (Obs. = 646,645)								
Share of Pro	0.128***	0.323***	0.135***	0.125***	0.0561***	0.524***	0.497***	-0.00955***
Leave Tweets	(0.00349)	(0.00518)	(0.00363)	(0.00355)	(0.00239)	(0.00427)	(0.00527)	(0.00106)
Panel B: Using the sample of Twitter users tweeting on EU related topics (Obs. = 62,136)								
Share of Pro	0.0557***	0.118***	0.0622***	0.0509***	0.0261***	0.0794***	0.110***	0.00218*
Leave Tweets	(0.00365)	(0.00550)	(0.00379)	(0.00374)	(0.00249)	(0.00460)	(0.00565)	(0.00113)
Bot definition:	top 0.25% user	>10 tweets in any day	>25 tweets in any day	>10 tweets in any night	>25 tweets in any night	>10 ident. tweets/day	>25 ident. tweets/day	post EU ref. ann.
User controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The dependent variable is a dummy for suspect bots according to several criteria: users in the top 0.25% of the user distribution (column 1); users with more than 10 tweets in any day (column 2); users with more than 25 tweets in any day (column 3); users with more than 10 tweets in the period 00:00-06:00 in any day (column 4); users with more than 25 tweets in the period 00:00-06:00 in any day (column 5); users with more than 10 identical messages in any day (column 6); users with more than 25 identical messages in any day (column 7); users whose account has been open after the EU referendum announcement (column 8). Share of Pro Leave Tweets is defined as Pro Leave tweets/(Pro Leave tweets+Pro Remain tweets). Panel A runs the test on the full sample of Twitter users. Panel B restricts the analysis on the sample of Twitter of users twitting on EU related subjects. User controls include: dummy for verified user and user's number of followers. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

06:00 (column 5); users with more than 10 identical tweets in any day (column 6); users with more than 25 identical tweets in any day (column 7); and users whose account has been set up after the EU referendum announcement (column 8).

In Panel A we run our regression on the full sample of UK Twitter users. All coefficients for the share of pro Leave tweets are positive and significant, with the exception of column (8). This implies that users tweeting more pro Leave are more likely to be bots, and therefore spread fake news, as compared to the average UK Twitter user. Given the bot definition adopted in column (8), however, the latter test is somewhat unsound. We are asking whether Twitter users leaning more pro Leave are more likely to have created their account after February 2016, as compared to any other Twitter user in the UK (a large share of users never tweeted on Brexit matter).

In order to improve the accuracy of the test, we restrict our analysis to the sample of users tweeting on EU related topics in the two months before the referendum in Panel B. In this context, the question becomes: among users actively tweeting on EU related matter, are relatively Leave oriented users more likely to be bots? Results in Panel B consistently suggest that this was indeed the case, irrespective of the criterion adopted to identify bots.

In our third strategy we exploit the full sample of 18 million tweets to further inspect the mechanism. Our theoretical model suggests that users react differently to messages they perceive to be fake. The full dataset allows us to trace users over time, since any of their tweets includes the number of followers they have at the moment they are shared. We can therefore estimate the change in users' popularity, measured as the change in the number of followers, which certain types of tweets generate. Intuitively, tweets which are perceived as trustworthy may spread further across users, and more importantly may grant more followers to the original tweeting user.

We first test whether tweeting about the EU was on average granting more followers. The results, reported in Table 9, suggest that tweeting about EU matter "in general" did not increase a user's number of followers. Results in column (1), including only *EU-related Tweet* on the right-hand side of the equation, suggest that EU related tweets were less popular then other tweets. The negative coefficient turns positive and non-significant from in column (2), where we include basic user characteristics: the number of followers before the tweet considered, and whether it was a verified user. The coefficient of interest remains non significant in the rest of the table, when we control for the tweet containing a direct quotation on another tweet (column 3), and a set of language dummies (column 4), and day of the week and hour of the day fixed effects (column 5). Hence, EU related tweets are not considered abnormally trustworthy as compared to tweets on other topics.

Table 9: *The effect of EU-related tweets on popularity on Twitter*

Dependent variable: Change in Number of Followers	(1)	(2)	(3)	(4)	(5)
EU-related Tweet	-0.274***	0.118	0.122	0.111	0.180
	(0.101)	(0.156)	(0.157)	(0.153)	(0.156)
User account controls		✓	✓	✓	✓
Quoting another Tweet dummy			✓	✓	✓
Main language dummies				✓	✓
Hour & day of week fixed eff.					✓
Observations	18,049,673	18,049,673	18,049,673	18,049,673	18,049,673

Notes: EU-related Tweet is a dummy for tweets featuring one of the EU-related hashtags (see Appendix Table A1). User account controls include: the user's number of followers before the tweet considered, and a dummy for verified users. The dummy for quoting another tweet is equal to one for tweets directly quoting another tweet. From column (4) we include dummies for the three main languages (English, Spanish and Portuguese). Column (5) includes fixed effects for the days of the week and the hours of the day. Standard errors are clustered at the date level. *** p<0.01, ** p<0.05, * p<0.1.

Next, we test if tweets supporting Remain (Leave) were eliciting different reactions on Twitter. More specifically, according to our model we expect that Leave supporting tweets are more likely to be perceived as fake news, and thus should be disregarded more often and generate fewer additional followers. Thus, we estimate an alternative model in which we include on the right-hand side the two dummy variables, *Pro Remain Tweet* and *Pro Leave Tweet*, instead of *EU-related Tweets*. Results are reported in Table 10.

While the coefficient for the *Pro Remain Tweet* dummy in consistently positive and statistically significant, suggesting that tweeting pro Remain in the last two months preceding the referendum was increasing popularity on Twitter, the coefficient for the *Pro Leave Tweet* dummy is consistently smaller in magnitude and not statistically significant. In other words, tweets featuring hashtags associated with a Remain position are relatively more popular, which is consistent with our interpretation according to which pro Leave tweets were regarded as less trustworthy.

Finally, we employ once more our measure of Twitter user orientation on the EU referendum, the number of pro Leave tweets over the sum of pro Leave and pro Remain tweets, to question whether tweets sent by relatively more Leave leaning users generate a systematically different number of followers. Table 11 reports the results of this exercise, applied on the sub-sample of tweets by users with at least one tweet disclosing their orientation on the EU referendum. Indeed, tweets from relatively more pro Leave users obtain systematically less followers. Compared to a fully committed pro Remain user, a fully committed pro Leave user is predicted to obtain 0.1 less additional followers in response to a new tweet, which represents about 10% of the average change in followers.

Overall, the results discussed in this section suggest that sourcing political information through Twitter

Table 10: *The effect of Leave/Remain Tweets on popularity on Twitter*

Dependent variable: Change in Number of Followers	(1)	(2)	(3)	(4)	(5)
Pro Remain Tweet	-0.063	0.501***	0.508***	0.495***	0.534***
	(0.120)	(0.185)	(0.185)	(0.182)	(0.197)
Pro Leave Tweet	-0.098	0.335	0.342	0.317	0.356
	(0.217)	(0.264)	(0.265)	(0.261)	(0.248)
User account controls		✓	✓	✓	✓
Quoting another Tweet dummy			✓	✓	✓
Main language dummies				✓	✓
Hour & day of week fixed eff.					✓
Observations	18,049,673	18,049,673	18,049,673	18,049,673	18,049,673

Notes: Pro Remain and Pro Leave Tweet are dummies for tweets featuring one of the EU-related hashtags listed in Appendix Table A1. User account controls include: the user's number of followers before the tweet considered, and a dummy for verified users. The dummy for quoting another tweet is equal to one for tweets directly quoting another tweet. From column (4) we include dummies for the three main languages (English, Spanish and Portuguese). Column (5) includes fixed effects for the days of the week and the hours of the day. Standard errors are clustered at the date level. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: *The effect of user orientation in the EU referendum on popularity on Twitter*

Dependent variable: Change in Number of Followers	(1)	(2)	(3)	(4)	(5)
Share of Pro	-0.058	-0.101***	-0.101***	-0.100***	-0.101***
Leave Tweets	(0.031)	(0.034)	(0.034)	(0.034)	(0.034)
User account controls		✓	✓	✓	✓
Quoting another Tweet dummy			✓	✓	✓
Main language dummies				✓	✓
Hour & day of week fixed eff.					✓
Observations	3,696,776	3,696,776	3,696,776	3,696,776	3,696,776

Notes: Share of Pro Leave Tweets is defined as Pro Leave tweets/(Pro Leave tweets+Pro Remain tweets). User account controls include: the user's number of followers before the tweet considered, and a dummy for verified users. The dummy for quoting another tweet is equal to one for tweets directly quoting another tweet. From column (4) we include dummies for the three main languages (English, Spanish and Portuguese). Column (5) includes fixed effects for the days of the week and the hours of the day. Standard errors are clustered at the date level. *** p<0.01, ** p<0.05, * p<0.1.

constrained learning over the EU referendum debate. The more substantial presence of fake news on the Leave campaign, some of which generated and spread by bots, ended up affecting the relative credibility of Leave supporting tweets, thereby preventing voters from using them when reconsidering their position on the EU referendum.[20]

[20] An alternative possible mechanism could be related to turnout. As discussed in Section 2 there exists some evidence from previous studies that internet penetration lowers turnout and benefits the incumbency (Falck, Gold, and Heblich, 2014; Campante, Durante, and Sobbrio, 2018; Gavazza, Nardotto, and Valletti, 2019). In the context of the EU referendum, turnout was particularly high and it has been argued that it may have been one of the factors favoring vote Leave (Zhang, 2018). So could it be that the negative link found between Twitter use and the vote leave share is driven by turnout? To investigate this alternative mechanism, we run the same causal regression as in the baseline analysis, but this time using turnout as the dependent variable (computed as total voters

7 Conclusion

Social media and the diffusion of political messages in social networks has emerged as one of the most disruptive innovations in electoral competition in the twenty first century, with many scholars and political experts concerned that the big social media platforms such as Facebook and Twitter are harmful to democratic institutions. In this light, the 2016 UK referendum on the EU membership was a watershed moment, in which the intensity of social media campaigning and political messaging on social networks became even more salient.

In this paper we use data on 18 million tweets (messages on the social network Twitter) produced by users geo-located in the UK and in the two months preceding the EU referendum to investigate the impact of Twitter on the referendum outcome. We establish a causal negative relationship between exposure to Twitter and the Leave vote share. Next, we explore the mechanism behind this causal relationship. We argue that fake news on Twitter about the EU referendum was biased towards supporting vote Leave. Consequently, bounded rational voters receiving political messages on Twitter and applying simple behavioral rules to learn about the state of the world discard news supporting vote Leave as fake news and warrant undue weight to Twitter messages supporting Remain.

To support this mechanism, we follow three steps. First, exploiting the performance of UKIP in earlier political contests in the UK, we show that in regions more exposed to twitter there is greater persistence of political outcomes, suggesting that Twitter exposure constrained learning. Second, using text analysis and other metadata associated with the 18 million tweets, we identify Leave supporting tweets and Remain supporting tweets. We show that users who predominantly send Leave supporting tweets are more likely to be Bots (where Bots are defined as users displaying non-human patterns of behavior). This constitutes direct evidence that fake news was more likely to be supporting vote Leave. Finally, we show that users tweeting Leave supporting messages were less likely to generate new followers. This constitutes direct evidence that Leave supporting messages were less likely to be perceived to be produced by legitimate experts.

These results have important policy implications concerning the impact of social media on political contests. When fake news is pervasive in social networks, voters will be less willing to engage with information acquired through social media. If, moreover, fake news is perceived to be predominantly biased towards a given (contrarian) worldview, expertise in social media supporting the opposite view will exert an influence in excess of the accuracy of its information. This leads to a "wisdom of the crowds" phenomena, whereby

over above 20 population). The results are reported in Table A7 in the Appendix, and allow us to dismiss this alternative mechanism. No significant relationship is found between turnout and the intensity of Twitter use.

forceful agents emerge in social networks (supporting the consensus worldview), and legitimate evidence supporting the contrarian worldview is dismissed as fake news. Such outcome is inefficient because it delays and may even stop learning in social networks.

On the other hand, we believe our findings yield insights regarding the coexistence of traditional media and social media, and trustworthiness in the market for news (Gentzkow and Shapiro, 2008). If traditional media is perceived as trustworthy (for example, if all information reported by the BBC in the UK is trusted), then the big corporations do indeed exhort media power and are able to sway voters behavior to a greater extent than internet media sources (Prat, 2018). At any rate, the mediation between represented and representatives via the internet cannot explain the success of the Leave campaign, which brings other mediators (other forms of media, the political system and parties) back into focus. It is an open question whether through competition or regulatory intervention, in the long-run social media will graduate to standards of scientific trustworthiness which are similar to those enjoyed by traditional media.

References

Abedi, Amir and Thomas Carl Lundberg (2009). "Doomed to failure? UKIP and the organisational challenges facing right-wing populist anti-political establishment parties". *Parliamentary Affairs* 62.1, pp. 72–87.

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi (2010). "Spread of (mis)information in social networks". *Games and Economic Behavior* 70.2, pp. 194–227.

Allcott, Hunt and Matthew Gentzkow (2017). "Social media and fake news in the 2016 election". *Journal of Economic Perspectives* 31.2, pp. 211–36.

Antoci, Angelo, Laura Bonelli, Fabio Paglieri, Tommaso Reggiani, and Fabio Sabatini (2019). "Civility and trust in social media". *Journal of Economic Behavior & Organization* 160, pp. 83–99.

Azzimonti, Marina and Marcos Fernandes (2018). "Social media networks, fake news, and polarization". *NBER Working Paper No w24462*.

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau (2015). "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological Science* 26.10, pp. 1531–1542.

Becker, Sascha O, Thiemo Fetzer, and Dennis Novy (2017). "Who voted for Brexit? A comprehensive district-level analysis". *Economic Policy* 32.92, pp. 601–650.

Benson, Karyn, Alberto Dainotti, Kimberly C Claffy, and Emile Aben (2013). "Gaining insight into as-level outages through analysis of internet background radiation". In: *2013 IEEE Conference on Computer Communications Workshops*. IEEE, pp. 447–452.

Bialik, Kristen and Katerina Eva Matsa (2017). “Key trends in social and digital news media”. *Pew Research Center*.

Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler (2012). “A 61-million-person experiment in social influence and political mobilization”. *Nature* 489.7415, pp. 295–298.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio (2018). “Politics 2.0: The multifaceted effect of broadband internet on political participation”. *Journal of the European Economic Association* 16.4, pp. 1094–1136.

Chiang, Chun-Fang and Brian Knight (2011). “Media bias and influence: Evidence from newspaper endorsements”. *The Review of Economic Studies* 78.3, pp. 795–820.

Collins, Damien, Clive Efford, J Elliot, Paul Farrelly, Simon Hart, Julian Knight, and G Watling (2019). *Disinformation and "fake news": Final Report*. London: The House of Commons.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel (2003). “Persuasion bias, social influence, and unidimensional opinions”. *The Quarterly Journal of Economics* 118.3, pp. 909–968.

Durante, Ruben and Brian Knight (2012). “Partisan control, media bias, and viewer responses: Evidence from Berlusconi’s Italy”. *Journal of the European Economic Association* 10.3, pp. 451–481.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020). “Social media and protest participation: Evidence from Russia”. *Econometrica* 88.4, pp. 1479–1514.

Eurobarometer (2016). *Media use in the European Union*. Directorate-General for Communication (European Commission).

– (2020). *Media use in the European Union*. Directorate-General for Communication (European Commission).

Falck, Oliver, Robert Gold, and Stephan Heblich (2014). “E-lections: Voting Behavior and the Internet”. *American Economic Review* 104.7, pp. 2238–65.

Field, Matthew and Mike Wright (Oct. 17, 2018). “Russian Trolls Sent Thousands of Pro-Leave Messages on Day of Brexit Referendum, Twitter Data Reveals”. *Telegraph*.

Flaxman, Seth, Sharad Goel, and Justin M Rao (2016). “Filter bubbles, echo chambers, and online news consumption”. *Public Opinion Quarterly* 80.S1, pp. 298–320.

Fryer Jr, Roland G, Philipp Harms, and Matthew O Jackson (2019). “Updating beliefs when evidence is open to interpretation: Implications for bias and polarization”. *Journal of the European Economic Association* 17.5, pp. 1470–1501.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl (2016). “Habit formation in voting: Evidence from rainy elections”. *American Economic Journal: Applied Economics* 8.4, pp. 160–88.

Fujiwara, Thomas, Karsten Müller, and Carlo Schwarz (2020). "The effect of social media on elections: Evidence from the United States". *Available at SSRN No. 3719998*.

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* 86.5, pp. 2092–2135.

Gentzkow, Matthew and Jesse M Shapiro (2008). "Competition and truth in the market for news". *Journal of Economic Perspectives* 22.2, pp. 133–154.

– (2011). "Ideological segregation online and offline". *The Quarterly Journal of Economics* 126.4, pp. 1799–1839.

Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2015). "Media bias in the marketplace: Theory". In: *Handbook of Media Economics*. Vol. 1. Oxford: North-Holland, pp. 623–645.

Geraci, Andrea, Mattia Nardotto, Tommaso Reggiani, and Fabio Sabatini (2018). "Broadband internet and social capital". *IZA Discussion Paper No. 11855*.

Glaeser, Edward L and Cass R Sunstein (2009). "Extremism and social learning". *Journal of Legal Analysis* 1.1, pp. 263–324.

Golub, Benjamin and Matthew O Jackson (2010). "Naive learning in social networks and the wisdom of crowds". *American Economic Journal: Microeconomics* 2.1, pp. 112–49.

– (2012). "How homophily affects the speed of learning and best-response dynamics". *The Quarterly Journal of Economics* 127.3, pp. 1287–1338.

Gomez, Brad T, Thomas G Hansford, and George A Krause (2007). "The Republicans should pray for rain: Weather, turnout, and voting in US presidential elections". *The Journal of Politics* 69.3, pp. 649–663.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2018). "Social media, sentiment and public opinions: Evidence from Brexit and US Election". *NBER Working Paper No. w24631*.

Gottfried, Jeffrey and Elisa Shearer (2016). "News use across social medial platforms 2016". *Pew Research Center*.

Grčar, Miha, Darko Cherepnalkoski, Igor Mozetič, and Petra Kralj Novak (2017). "Stance and influence of Twitter users regarding the Brexit referendum". *Computational Social Networks* 4.1, p. 6.

Grossman, Gene M and Elhanan Helpman (2019). "Electoral competition with fake news". *NBER Working Paper No. w26409*.

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, pp. 73–88.

Hänska, Max and Stefan Bauchowitz (2017). "Mapping Twitter's information sphere in the lead-up to the Brexit referendum: How eurosceptic views outpaced their rivals". *AoIR Selected Papers of Internet Research*.

Johnson, Boris (Feb. 28, 2016). "Don't be taken in by project fear–Staying in the EU is the risky choice". *The Daily Telegraph*.

Kirill, Pogorelskiy and Matthew Shum (2019). "News we like to share: How news sharing on social networks influences voting outcomes". *CAGE Online Working Paper Series No. 427*.

Levy, Gilat and Ronny Razin (2019). "Echo chambers and their effects on economic and political outcomes". *Annual Review of Economics* 11, pp. 303–328.

Liberini, Federica, Michela Redoano, Antonio Russo, Ángel Cuevas, and Ruben Cuevas (2020). "Politics in the Facebook Era—Evidence from the 2016 US Presidential Elections".

Lind, Jo Thori (2020). "Rainy day politics. An instrumental variables approach to the effect of parties on political outcomes". *European Journal of Political Economy* 61, p. 101821.

Manacorda, Marco and Andrea Tesei (2020). "Liberation Technology: Mobile Phones and Political Mobilization in Africa". *Econometrica* 88.2, pp. 533–567.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks". *Annual Review of Sociology* 27.1, pp. 415–444.

Miner, Luke (2015). "The unintended consequences of internet diffusion: Evidence from Malaysia". *Journal of Public Economics* 132, pp. 66–78.

Moore, Martin and Gordon Ramsay (2017). *UK media coverage of the 2016 EU Referendum campaign*. King's College London.

Müller, Karsten and Carlo Schwarz (2020). "Fanning the flames of hate: Social media and hate crime". *Journal of the European Economic Association*. forthcoming.

Pariser, Eli (2011). *The filter bubble: What the Internet is hiding from you*. London: Penguin.

Prat, Andrea (2018). "Media power". *Journal of Political Economy* 126.4, pp. 1747–1783.

Siegel, Alex and Joshua A Tucker (July 20, 2016). "Here's what 29 million tweets can teach us about Brexit". *Washington Post*.

Sunstein, Cass R (2001). *Republic. com*. Princeton, NJ: Princeton University Press.

– (2009). *Going to extremes: How like minds unite and divide*. Oxford: Oxford University Press.

Zhang, Aihua (2018). "New findings on key factors influencing the UK's referendum on leaving the EU". *World Development* 102, pp. 304–314.

A Appendix (For online publication only)

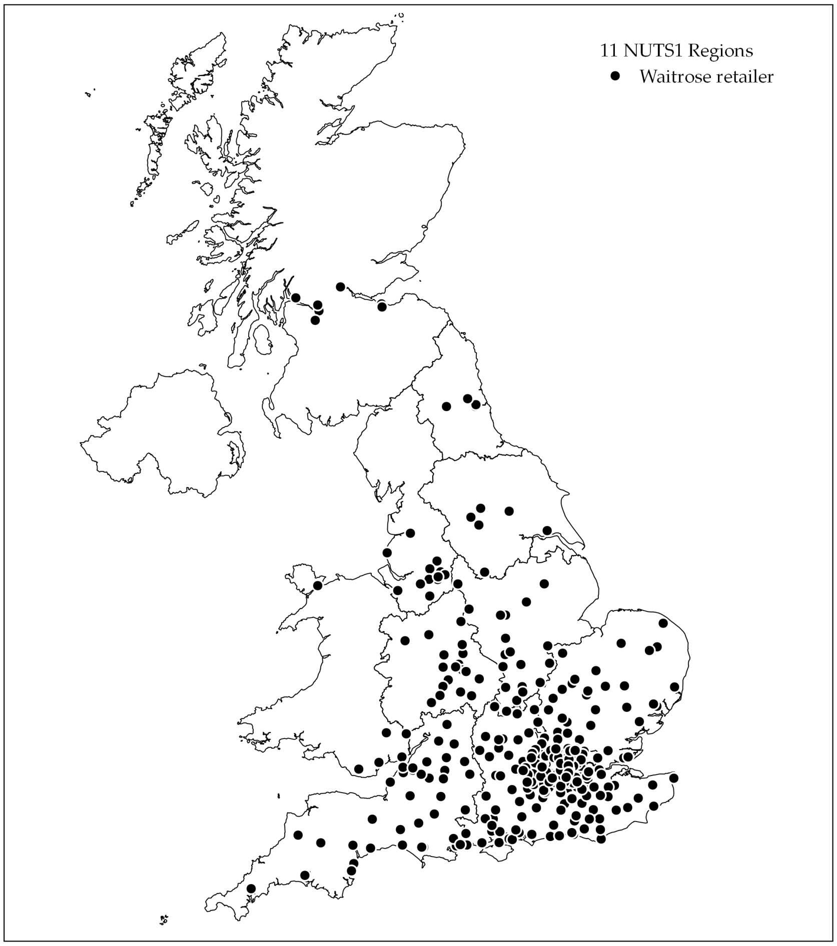


Figure A1: *Map of UK Waitrose markets*

Table A1: *List of EU related hashtags*

Hashtags Leave:
Article50
BeLeave
BetterOffOut
BorisJohnson
borisjohnson
BREXIT
BRexit
BrExit
Brexit2
BrexitAFilm
BrexitBudget
BrexitClub
BrexitDilemma
Brexiteers
Brexiters
brexiters
BrexitIn5Words
BrexitInFiveWords
brexitparty
brexitthemovie
BrexitVote
BritainFirst
Britexit
euleave
Farage
farage
fishingforleave
Frexit
independence
Independence
independent
IVotedLeave
ivotedleave
LabourLeave
Leave
leave
LEAVE
LeaveCampaign
LeaveEU
leaveeu
leaveEU
leavers
leaves
LeaveWins
leaving
Lexit
lexit
NigelFarage
proudtobebritish
ProudToBeBritish
RespectTheMandate
TakeBackControl
takebackcontrol
TheresaMay
UKIP
ukip
Ukip
unionjack
Vote_Leave
VotedLeave
VoteLeave
voteleave
Voteleave
VOTELEAVE
VoteLeaveTakeControl
VoteOut
voteout

Hashtags Remain:
2ndReferendum
2ndVoteWE
betterin
BetterIn
BetterTogether
bettertogether
Borexit
Bregret
Bremain
bremain
brexiteffect
brexitfail
CatsAgainstBrexit
CurseBorisJohnson
DogsAgainstBrexit
Engexit
eustay
INtogether
IVotedRemain
JoCox
jocox
JoCoxMP
JoCoxRIP
LabourIn
LabourInForBritain
leadnotleave
Leadnotleave
LoveOurEUStaff
MoreInCommon
moreincommon
NoMoreLies
NotInMyName
notinmyname
NotInOurName
notmyvote
NotMyVote
NoToGove
PostRefRacism
postrefracism
REGREXIT
Remain
remain
REMAIN
RemaIN
remaIN
Remainers
RemainIn
RemainINEU
RIPJo
RIPJoCox
ripjocox
SadiqKhan
secondreferendum
SNPin
standupforeurope
StrongerIn
strongerin
StrongerIN
Strongerin
strongerineurope
StrongerTogether
strongertogether
the48percent
TheIndecentMinority
UKinEU
VoteBeaver
VotedRemain
VoteIn
votein
VoteIN
VoteRemain
voteremain
VoteREMAIN
VoteStay
votestay
WhatHaveWeDone
whathavewedone

Other EU hashtags:
brexit
bbceureferendum
bbcreferendum
BBCReferendum
Brexit
BrexitOrNot
BritainVotes
EU
eu
Eu
EUDebate
EUParliament
EUref
EURef
euref
EuRef
Euref
EUREF
EUreferendum
EUReferendum
eureferendum
eureferendum2016
EUrefmids
EURefReady
eurefresult
EURefResult
EURefResults
EUrefresults
eurefresults
EuRefResults
Europe
europe
European
european
EuropeanUnion
EUvote
HowShouldIVote
Immigrants
Immigration
Immigration
InOrOut
Inorout
InorOut
InOut
ITVEURef
Itveuref
ITVEUref
Itvreferendum
Ivevoted
iVoted
Ivoted
IVoted
Ivoted
LSEBrexitVote
MirrorLiveEU
politicalcartoon
politicians
poll
PollingDay
pollingday
PollingStation
referendum
Referendum
ReferendumDay
refugee
RefugeeCrisis
refugees
RegisterToVote
sovereignty
UKDecides
UKref
UKreferendum
useyourvote
vote
Vote
VOTE
voted
voterregistration
voting
whyvote
YourVoteMatters

Notes: The following list was created by scanning the top 15,000 hashtags (by number of tweets). All hashtags listed were considered as EU-related, whereas the first two subsets were considered as pro Leave, and pro Remain, respectively.

Table A2: *First stage - IV: Distance from LE*

Dependent variable: Twitter Exposure					
	(1)	(2)	(3)	(4)	(5)
Distance from LE	-0.21***	-0.25***	-0.15***	-0.10***	-0.10***
	(0.07)	(0.06)	(0.04)	(0.04)	(0.04)
Local authority fixed effects	✓	✓	✓	✓	✓
Geographic controls		✓	✓	✓	✓
Demographic controls			✓	✓	✓
Economic controls				✓	✓
Political controls					✓
Observations	1,278	1,278	1,278	1,278	1,278
R-squared	0.38	0.41	0.59	0.60	0.60

Notes: Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: *First stage - IV: Internet Outages*

Dependent variable: Twitter Exposure					
	(1)	(2)	(3)	(4)	(5)
Internet Outages	-10.19**	-9.61**	-5.48***	-4.89***	-5.03***
	(4.71)	(4.82)	(1.88)	(1.74)	(1.88)
NUTS1 fixed effects	✓	✓	✓	✓	✓
Geographic controls		✓	✓	✓	✓
Demographic controls			✓	✓	✓
Economic controls				✓	✓
Political controls					✓
Observations	1,278	1,278	1,278	1,278	1,278
R-squared	0.10	0.12	0.52	0.55	0.55

Notes: Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections, and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: *The impact of Twitter exposure on the Brexit vote - further robustness tests*

Dependent variable: Share of Leave votes									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A:									
Twitter Exposure (with bots)	-2.28***	-1.31***	-0.35**	-10.13***	-10.44***	-7.02	-18.90***	-23.39***	-7.52*
	(0.36)	(0.26)	(0.17)	(2.53)	(3.72)	(5.00)	(6.93)	(8.43)	(3.90)
				[-17.7,-6.6]	[-24.0,-5.6]	[-206.5,-1.9]	[-30.2,-13.5]	[-48.6,-14.8]	[-16.3,-4.1]
Kleibergen-Paap F-statistic				7.548	5.961	2.226	5.170	8.077	8.843
Panel B:									
Twitter Exposure	-2.55***	-1.52***	-0.44**	-9.74***	-9.53***	-5.85**	-20.39**	-29.82***	-12.38*
(Log(1+Tweets))	(0.38)	(0.26)	(0.18)	(2.17)	(2.45)	(2.35)	(8.36)	(10.97)	(6.76)
				[-15.9,-6.5]	[-18.4,-5.4]	[-21.9,-1.8]	[-34.0, -14.5]	[-76.8,-18.0]	[-60.8,-6.0]
Kleibergen-Paap F-statistic				9.715	15.77	7.451	3.960	7.320	4.782
Panel C:									
Twitter Exposure	-1.48***	-0.92***	-0.25**	-8.84***	-8.52***	-5.97	-16.27**	-24.39*	-7.76*
(including Retweets)	(0.24)	(0.16)	(0.11)	(2.39)	(2.83)	(3.94)	(7.61)	(12.62)	(4.43)
				[-19.4,-5.4]	[-26.8,-4.4]	[-∞, +∞]	[-31.1,-11.0]	[-123.2,-13.3]	[-37.1,-3.7]
Kleibergen-Paap F-statistic				6.717	7.041	2.340	3.446	4.354	4.956
Local authority fixed effects	✓	✓	✓	✓	✓	✓			
NUTS1 fixed effects							✓	✓	✓
Geographic controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Economic controls			✓			✓			✓
Political controls			✓			✓			✓
Estimation method		OLS			IV: Distance from LE			IV: Internet Outage	
Observations	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278

Notes: Twitter exposure is measured as: the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day (Panel A); the log(aggregate number of tweets at the ward level in the two months preceding the referendum day +1) in Panel B; the (log) aggregate number of tweets and retweets at the ward level in the two months preceding the referendum day, where retweets are obtained by multiplying by 62 days a per day retweeting activity computed on the basis of users activity (Panel C). Except in Panel A we exclude the top 0.25% users to exclude bots' activity. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15–19], [20–29], [30–39], [40–49], [50–59], [60–89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 7-9 only), and the UKIP vote share in the latest local elections. Anderson Rubin weak instrument-robust 95% confidence intervals are reported in squared brackets. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: *Falsification exercise: Negative Retweets*

Dependent variable: Share of Leave votes	(1)	(2)	(3)	(4)	(5)	(6)
Negative Retweets Dummy	-0.147	-0.625	-0.376	-0.150	-1.098*	-0.490
	(0.546)	(0.479)	(0.408)	(0.706)	(0.655)	(0.416)
Local authority fixed effects	✓	✓	✓			
NUTS1 fixed effects				✓	✓	✓
Geographic controls		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓
Economic controls			✓			✓
Political controls			✓			✓
Observations	1,278	1,278	1,278	1,278	1,278	1,278

Notes: Negative Retweets Dummy equals 1 for wards with aggregate negative retweets. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 4-6 only), and the UKIP vote share in the latest local elections. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: *The impact of Twitter exposure on "learning" - alternative measures of learning*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A:									
Dependent variable: Low learning dummy (bottom 30%)									
Twitter Exposure	0.09***	0.05***	0.01	0.44***	0.47***	0.31*	0.45**	0.61***	0.57**
	(0.02)	(0.01)	(0.01)	(0.13)	(0.16)	(0.19)	(0.19)	(0.23)	(0.25)
				[0.28, 0.75]	[0.25, 0.93]	[0.06, 1.12]	[0.29, 0.74]	[0.32, 1.43]	[0.27, 1.65]
Panel B:									
Dependent variable: Learning (continuous variable)									
Twitter Exposure	-2.46***	-1.49***	-0.45**	-9.27***	-8.90***	-5.14**	-7.48*	-8.66**	-10.00**
	(0.35)	(0.25)	(0.18)	(2.05)	(2.30)	(2.10)	(4.10)	(4.39)	(4.96)
				[-15.1,-6.1]	[-17.0,-5.0]	[-17.4,-1.4]	[-12.2, -4.8]	[-20.2,-4.0]	[-28.1,-5.2]
Local authority fixed effects	✓	✓	✓	✓	✓	✓			
NUTS1 fixed effects							✓	✓	✓
Geographic controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Economic controls			✓			✓			✓
Political controls			✓			✓			✓
Estimation method		OLS			IV: Distance from LE			IV: Internet Outage	
Observations	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278
Kleibergen-Paap F-statistic				9.737	15.36	7.538	4.680	8.521	7.157

Notes: The dependant variable is a dummy equal one for the bottom 30% wards in terms of learning, defined as the absolute value of the change between the Leave and the UKIP vote shares at the 2014 EU elections (Panel A), and the absolute value of the change between the Leave and the UKIP vote shares at the 2014 EU elections (Panel B). Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 7-9 only), and the UKIP vote share in the latest local elections. For the IV estimates in column 4-9, Anderson Rubin weak instrument-robust 95% confidence intervals are reported in squared brackets. Standard errors are clustered at the local authority level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A7: *The impact of Twitter exposure on the EU referendum turnout*

Dependent variable: Referendum Turnout	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Twitter Exposure	-0.03***	-0.00	-0.00	0.02	0.11	-0.03	-0.08*	0.01	0.03
	(0.01)	(0.00)	(0.00)	(0.07)	(0.08)	(0.07)	(0.04)	(0.04)	(0.06)
				[-0.04, 0.12]	[0.01, 0.35]	[-0.07, 0.71]	[-0.17,-0.02]	[-0.06, 0.11]	[-0.10, 0.24]
Local authority fixed effects	✓	✓	✓				✓	✓	✓
NUTS1 fixed effects				✓	✓	✓			
Geographic controls		✓	✓		✓	✓		✓	✓
Demographic controls		✓	✓		✓	✓		✓	✓
Economic controls			✓			✓			✓
Political controls			✓			✓			✓
OLS	✓	✓	✓						
IV: Internet Outage				✓	✓	✓			
IV: Distance from LE							✓	✓	✓
Observations	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278	1,278
R-squared	0.40	0.61	0.62	0.01	0.14	0.42	-0.08	0.36	0.37
Kleibergen-Paap F-statistic				4.680	8.521	7.157	9.737	15.36	7.538

Notes: Referendum Turnout is (imperfectly) measured as the total votes expressed over the total ward population aged 20-89. Twitter exposure is measured as the (log) aggregate number of tweets at the ward level in the two months preceding the referendum day. We exclude the top 0.25% users to exclude bots' activity. Geographic controls include: (log) area, distance to the equator, and local rainfall on the day of the referendum. Demographic controls include: (log) population, the proportion of female population, and the share of population in the following age groups [15-19], [20-29], [30-39], [40-49], [50-59], [60-89]. Economic controls include the share of population occupied in the nine standard UK occupation categories: Managers, Professionals, Associate Professionals, Administrative, Trade, Caring, Sales, Industry, and Elementary. Political controls include the UKIP vote share in 2014 EU elections (in columns 7-9 only), and the UKIP vote share in the latest local elections. For the IV estimates in column 4-9, Anderson Rubin weak instrument-robust 95% confidence intervals are reported in squared brackets. Standard errors are clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.