Mobile Internet access and political outcomes: Evidence from South Africa

Dante Donati[†]
Universitat Pompeu Fabra

This version: September 2017[*]

Abstract

Does 3G Internet coverage affect political participation and individuals' voting behavior in developing countries? In this paper I provide an empirical answer to this question looking at the South African municipal elections results between 2011 and 2016. I conduct two complementary analyses. In the first part, I exploit the temporal and geographical variation in 3G Internet coverage to estimate its causal impact on voter turnout and vote shares of the major South African parties, ANC, DA and EFF, respectively. In order to mitigate concerns on potential endogeneity, I use a high-resolution newly constructed dataset along with a Diff-in-Diff estimation, and perform a series of robustness checks. In the second part, I develop an extensive analysis of the potential mechanisms that may drive the observed relationships. A triple difference estimator is used to assess the role of the Internet in providing information on the financial administration of the municipal money. Finally, I carry out a heterogeneity analysis across sub-samples of the original dataset to examine which segments of the population are more likely to be affected by Internet coverage.

The main findings show that Internet access did not significantly affect voter turnout, yet it caused a reduction in the vote share of the ruling ANC party by almost 6.4 pp. The second major party, the DA, has gained from the Internet arrival. Its share rose by approximately 2.2 pp. Moreover, among localities with high level of corruption-related expenditure, a unitary increase in coverage causes a reduction in the votes for the ANC party of almost 1.2 pp. Finally, evidence from the heterogeneity analysis points out the information role the Internet plays in more isolated localities.

Keywords: Corruption, Media, Mobile Internet, Municipal elections, Political outcomes.

[†]I thank Ruben Durante, Ruben Enikolopov, Nicola Gennaioli, Eliana La Ferrara, David Nagy, Maria Petrova, Giacomo Ponzetto, Nandan Rao, Marta Reynal-Querol and Alessandro Tarozzi for their comments and useful insights. I acknowledge financial support from LEAP Bocconi (Laboratory for Effective Anti-Poverty Policies) for the purchasing of data on mobile Internet coverage.

[*]This is a preliminary version of the paper. I am currently working on it to improve the analysis with additional information.

1. Introduction

In the past few decades, Information and Communications Technology (ICT) has transformed the world. By connecting people and places, ICT has played a vital role in the national, regional and global development, and it holds significant promise for the future (World Bank, 2006). At the same time, the Internet can also be seen as a powerful political tool that allows people to access additional/new information, to overcome collective action and coordination problems, and foster political change. Indeed, some scholars (McChesney, 2007) have argued that the world is experiencing a "communication revolution", an idea that relates to the presumed democratizing role of the Internet, which relax media ownership rules in the traditional media sector. Furthermore, more recently the Internet has also led to the birth of a novel and particularly popular news platform, that is social media. The main novelty social media bring to the society is the horizontal flow of information between individual users. The information role of social media might also be amplified by the possibility of being constantly "on line", thanks to the expansion of the 3G mobile Internet accompanied by the growing usage of the smartphones.

Can high-speed mobile Internet access affect political participation and political outcomes in developing countries? This essays seeks to answer this question looking at the mobile Internet expansion that took place between 2011 and 2016 in South Africa, a country ranked as 39th out of 180 in terms of freedom of press [1]. The country is also peculiar from the political standpoint: the current ruling party, the African National Congress (ANC), has dominated the scene since 1994. Only in the more recent elections its rivals experienced a notably increase in popularity.

Identifying the causal impact of mobile Internet access on political outcomes is challenging. In particular, compared to the literature on TV and radio coverage, in this case it is harder to exploit technological features of the transmitters to find an exogenous source of variation. Therefore, in the absence of a valid instrument, this paper attempts to tackle endogeneity issues related to the Internet coverage adopting a Diff-in-Diff estimation. In particular, I exploit the gradual temporal and geographical expansion of 3G Internet coverage along with the change in political outcomes across 10,000 voting districts between the 2011 and 2016 Municipal Elections. To conduct the research, I rely on a newly-constructed dataset containing geo-referenced information at the voting district-year level on Internet coverage, political outcomes, population density, night light as a proxy for GDP, and additional information on infrastructure, geography and irregular municipal expenditure.

The empirical strategy comes with some limitations. Specifically, coverage is far from random, yet it reflects the result of private operators' decision problem on if and where to install the technology. In this process, demand side factors play a fundamental role (Buys at al., 2009). Therefore, a central identification assumption must hold in order for my estimates to be unbiased: conditional on observables – mainly GDP and population density – and on voting district and time fixed effects, Internet coverage is exogenous. The paper carry out some robustness checks to provide some evidence in favor of this assumption. Among other things, I exploit Municipal Election results in 2000 and 2006 to perform placebo estimates.

The findings of my analysis show that Internet coverage did not significantly affect

[1] 2016 World Press Freedom Index.

voter turnout. At the same time, they demonstrate that Internet access can be detrimental for the popularity of the ruling party. In fact, a unitary increase in Internet coverage led to a drop in the share of votes for the ANC by approximately 6.4 pp. At the same time, there is evidence showing that the main ANC opponent, the Democratic Alliance (DA), benefited from Internet coverage by roughly 2.2 pp. Finally, less robust evidence based on predicted outcomes shows that a new party formed in 2013, the Economic Freedom Fighters, did not benefit from Internet penetration.

The paper also develops an extensive analysis of the potential mechanisms that may drive the observed relationships. A triple difference estimator has been applied to understand the role of the Internet in providing information on the financial administration of the municipal money. This analysis points out that among localities with high level of corruption-related expenditure, a unitary increase in coverage causes a reduction in the votes for the ANC party of almost 1.2 pp. Moreover, a heterogeneity analysis across sub-samples of the original dataset has been conducted to examine which segments of the population are more likely to be affected by Internet coverage. Overall, this analysis points out the information role the Internet plays in more isolated localities.

My paper relates to the literature in the political economy of media that has tried to analyze the potential impact of the diffusion of broadband internet and digital ICTs on various forms of political participation and mobilization. Most of these works focus on developed economies, and particularly on landline Internet.

For instance, the pioneering work by Falk et al. (2014) studies Internet and voting behavior in Germany. The authors address endogeneity in Internet availability by exploiting technological peculiarities of the preexisting voice telephony network that hinder the roll-out of fixed-line broadband Internet. Their results suggest that Internet had a negative impact on turnout and, at the same time, its availability crowded out TV viewership but not newspaper readership. A more recent paper by Campante et al. (2016) addresses the phenomenon from a broader perspective, better emphasizing the potential underlying mechanisms. In particular, the paper sheds lights on the existing trade-off between the information and entertainment role of the web, looking not only at the short but also at the long run. Their findings support the idea that access to Internet provides exit opportunities for voters dissatisfied with mainstream politics: turnout decreases in the short term. However, the Internet is consequently used as a political tool to reach out and recruit these individuals by the recently formed 5 Star Movement.

Since my paper focuses on Local Municipalities it also relates to the work by Gavazza et al. (2016), which look at the impact of Internet penetration on local policies in the UK. The authors argue that the effect of the Internet in displacing traditional media should be larger for local elections, as local newspaper seems to suffer the most from the substitution effect. In addition, their descriptive evidence also shows that only 11% of the people use internet to get information on an MP, local councilor or politician. Also, the Information/Entertainment usage varies across age, education and socio-economic status. The empirical evidence relies on rainfall as an instrument for Internet availability. The authors show that turnout, local expenditure and taxes are negatively affected by Internet penetration. Most interestingly, heterogeneity analysis displays that the negative effect on turnout is significant only for low-educated or low-aged localities. This is consistent with the idea that the Internet affected political outcomes only for those individuals that mainly utilize it as a source of entertainment.

The previous works focus only on developed economies. In addition, their findings are

very context-related, being the result of particular election outcomes, in a specific point in time and in countries with a particular background. One would expect the Internet to impact differently depending on these factors. Therefore, external validity of these estimates is low, and relying on them to explain phenomena in underdeveloped or developing economies would be wrong. Overall, empirical evidence on these countries is limited. A recent paper by Miner (2015) studies the impact of landline internet penetration in Malaysia on pro-government vote. Using an IV approach, the author shows that areas with higher internet penetration experienced higher turnout and lower share of votes for the ruling coalition.

My work builds on the presented literature to the extent that it also focuses on how the expansion of the Internet affects political outcomes. The main difference is that I consider mobile technology as opposed to landline Internet. Under this aspect my work also relates to the study of Manacorda and Tesei (2015), which find support for the liberation technology argument that digital mobile ICT fosters political mobilization. Moreover, by focusing on a developing country, my paper also seeks to fill the existing gap by providing at least new suggestive evidence on political outcomes in the African context. Finally, I also provide a comprehensive picture on the potential underlying mechanisms.

The remainder of this paper is structured as follows. Section 2 briefly describes the political background and Section 3 provides descriptive evidence on the Internet and media markets. Section 4 explains the empirical strategy and Section 5 presents the data. Finally, Section 6 shows the results and Section 7 draws the conclusions.

2. Political background

Since the end of apartheid in 1994 the African National Congress (ANC) has dominated South Africa's politics. The ANC is the ruling party in the national legislature, as well as in eight of the nine provinces. The ANC received 62.15%[2] of the vote during the 2014 general election. The party is led by Jacob Zuma, who has served as President of South Africa since May 9, 2009. The main challenger to the ANC's rule is the Democratic Alliance, which received 22.23% of the vote in the 2014 election. The newly formed Economic Freedom Fighters (EFF), led by expelled ANC Youth League leader Julius Malema, contested its first municipal election since its formation in 2013 and received 6.35% of the vote in the general election.

Local government in South Africa consists of municipalities of various types. The largest metropolitan areas are governed by metropolitan municipalities, while the rest of the country is divided into district municipalities, each of which consists of several local municipalities. In 2016, there were 8 metropolitan municipalities, 44 district municipalities and 205 local municipalities. The councils of metropolitan and local municipalities are elected by a system of mixed-member proportional representation every 5 years. The following table depicts the overall results of the currently major 3 parties in the last 3 municipal ballots.

[2]This and the following numbers come from the Independent Electoral Commission (IEC) of South Africa, http://www.elections.org.za

Table 1: South African municipal election results

Party name	2006	2011	2016
African National Congress (ANC)	64.8%	62.0%	54%
Democratic Alliance (DA)	16.2%	23.9%	26.2%
Economic Freedom Fighters (EFF)	Formed in 2013		7.9%

Source: IEC of South Africa

The streaking feature that emerges from the figures is the gradual decline in the ANC vote share, which has been accompanied by the DA increase in popularity. In particular, between 2011 and 2016 ANC lost approximately 8 percentage points (pp), while DA gained 2pp. Establishing any trend for the newly formed EFF is difficult, yet this party receive almost 8% in the last municipal elections.

Understanding what might have driven the weakening in ANC popularity is difficult. A possible explanation might relate to the various corruption scandals that have emerged in the last years. For instance, a major campaign issue during the 2016 election was corruption within the ANC, in particular President Jacob Zuma's relationship with the Gupta family[3] and funding for the construction of his home at Nkandla. At the same time, after the elections new scandals came out[4]. In January 2017 the ANC was taken to court by a South African public relations expert, TV and radio personality Sihle Bolani for some work done during the elections. According to court papers filed in the High Court in Johannesburg, the ANC planned to spend R50 million (almost $3.8 million) on a covert campaign targeting opposition parties in the 2016 local government elections. In particular, a covert team, initially known as the War Room, intended to "disempower DA and EFF campaigns" and set a pro-ANC agenda using a range of media, without revealing the ANC's hand. Apparently, one of the most widely adopted strategy behind the fake campaign was the use of fake posters, such as the one targeting EFF electorate depicted in Figure 8 in the appendix.

On the one hand, if the fake news and disinformation campaign worked properly, then the ANC would have observed a positive return at the ballots. On the other hand, the disclosure of corruption-related scandals would imply a negative effect on the popularity of the leading party. If the decline in ANC vote share were only driven by these scandals, then one should notice large fall in the aggregate vote share also in previous elections. In fact, ANC-related corruption cases were used to come out during previous campaigns as well. Thus, why did the ANC loose so much in the last round?

A plausible explanation is the increase in Internet access and use of social media that occurred between 2011 and 2016. In particular, this paper argues that individuals in covered areas had the possibility of accessing new and relatively unbiased information through the Internet. This, in turns, should impact on their political participation and voting preferences. The potential mechanism behind this assumption would entail that Internet access helped voters in both 1. accessing information on ANC corruption scandals and 2. realizing that some news about competitors were fake. This mechanism implies that in areas with higher Internet coverage people decided to punish the ruling party and

[3] An Indian-South African business family which owns a business empire spanning computer equipment, media and mining.

[4] For instance, see http://amabhungane.co.za/article/2017-01-24-inside-the-ancs-black-ops-election-campaign

vote for alternative candidates. Therefore, we should observe a negative relation between Internet coverage and ANC vote share.

3. The Internet in South Africa

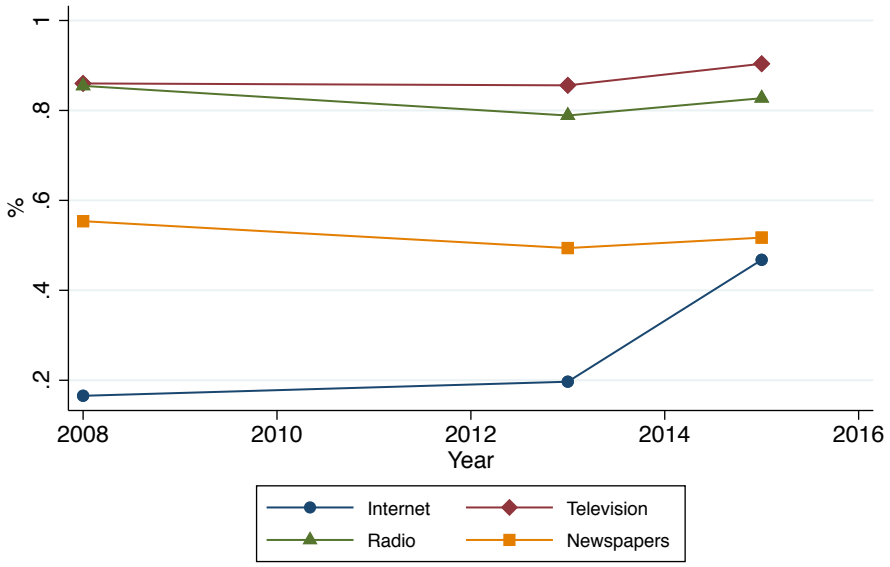
In terms of Internet access, South Africa is one of the most technologically advanced countries on the African continent. Nevertheless, there is still large spatial variation in Internet coverage within the country, with overall urban areas enjoying more connectivity than rural places.

South African Internet market is mainly dominated by three private Internet service providers (ISPs) with more or less homogenous market shares (Figure 9 in the appendix). The country has experienced a sharp increase in Internet usage since 2008: users as percentage of the population were around 10% in 2008, while they were more than 50% in 2014 (Figure 10a in the appendix) . At the same time, fixed broadband Internet subscribers per 100 people remain below 6 (Figure 10b in the appendix), and they mostly live in large cities where landline is available. Therefore, the rest (and most) of Internet users rely on mobile technologies. As a matter of fact, if we look at the number of mobile cellular subscriptions per 100 people (Figure 10c in the appendix) we can observe that after 2008 people started subscribing for a second line. As expected, the temporal variation in the usage was accompanied by a significant spatial expansion of mobile Internet coverage between 2008 and 2015 (Figure 11 in the appendix).

According to the 2017 annual report from We Are Social (https://wearesocial.com), an independent agency that monitors internet activity combining data from various qualified sources, almost 26.8 million people (50%) were active Internet users in January 2016. 13 million of them (24%) were active social media users, and 10 million (19%) were active mobile social users (Figure 12 in the appendix). The majority (92%) of South Africa's adult population owns a phone whether dumb, feature or smartphone. Smartphones represent the biggest share of this majority at 60%. Only 18% owns a laptop or a desktop computer, and 7% a tablet. The average South African spends just under 5 hours a day online (assuming time spent online at work is also counted). Average daily time spent on the Internet via a mobile phone is about 3 hours, and the time spent on any social media is just less than 3 hours. Note how time spent on social media and watching TV are almost exactly the same amount. This may be due to what the digital world calls using a "second screen", that is, tweeting, Whatsapping or Facebooking etc., while watching something on TV. Figure 12 in the appendix also reports the top ten most popular social media among South Africans, with WhatsApp and Facebook holding to the top two spots. Most of Facebook's users in South Africa are between the ages of 20 and 29 (41%). Senior citizens above the age of 60 account for 7% of users.

Previous literature (such as Gavazza et. al 2016) has documented the existence of a substitution effect of the Internet arrival on the use of traditional media in order to access information. Did this happen in South Africa? Figure 1 below is based on the elaboration of Afrobarometer data and helps us answering this question. Regular users are defined as those individuals that use the media as a source of information at least a few times a week. As expected, Internet usage rose between 2008 and and 2015, and it sharply increased between 2013 and 2015. Meanwhile, the Internet did not crowd out traditional media. Newspapers readership only marginally decreased, while the use of radio and TV remained almost constant over the last years.

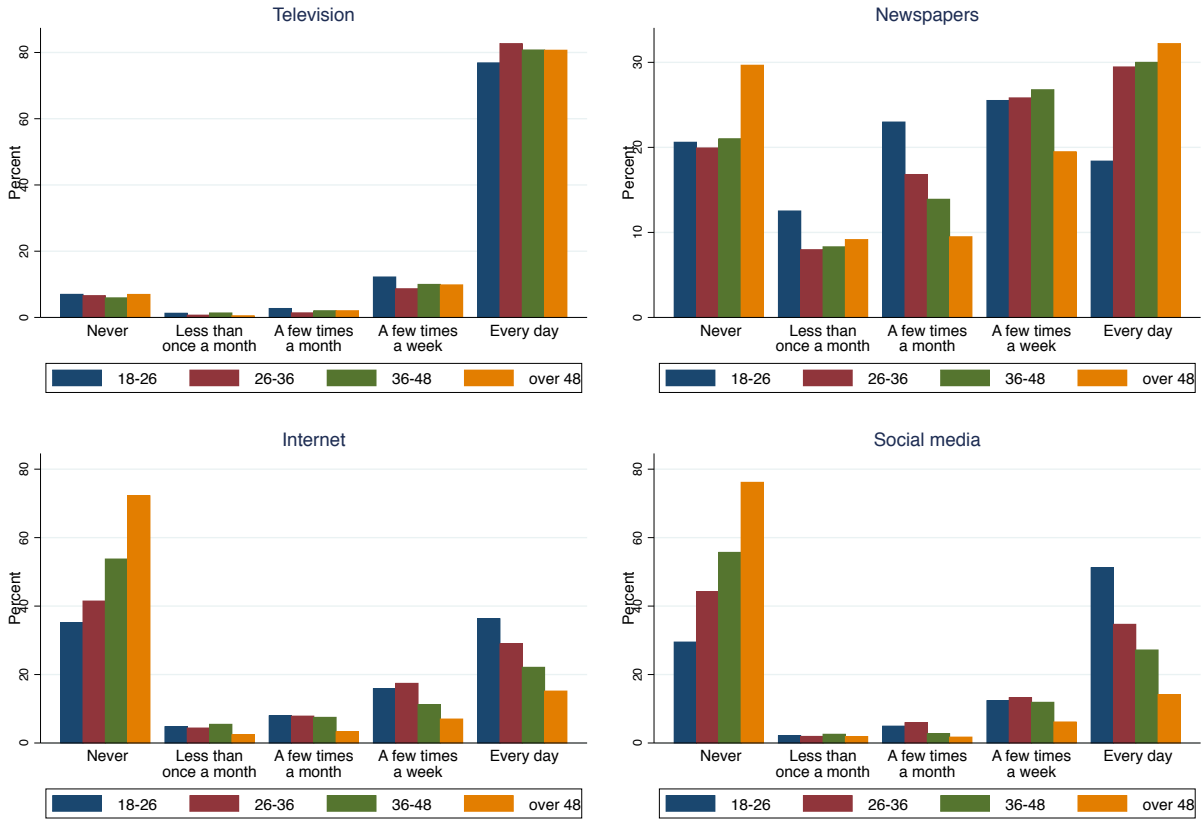
Figure 1: Regular users by type of media



Notes: Author's own calculations based on Afrobarometer.

Is there heterogeneity in the use of these media across different age categories? Figure 2 below shows this point in year 2015.

Figure 2: Use of media across age groups



Notes: Author's own calculations based on Afrobarometer.

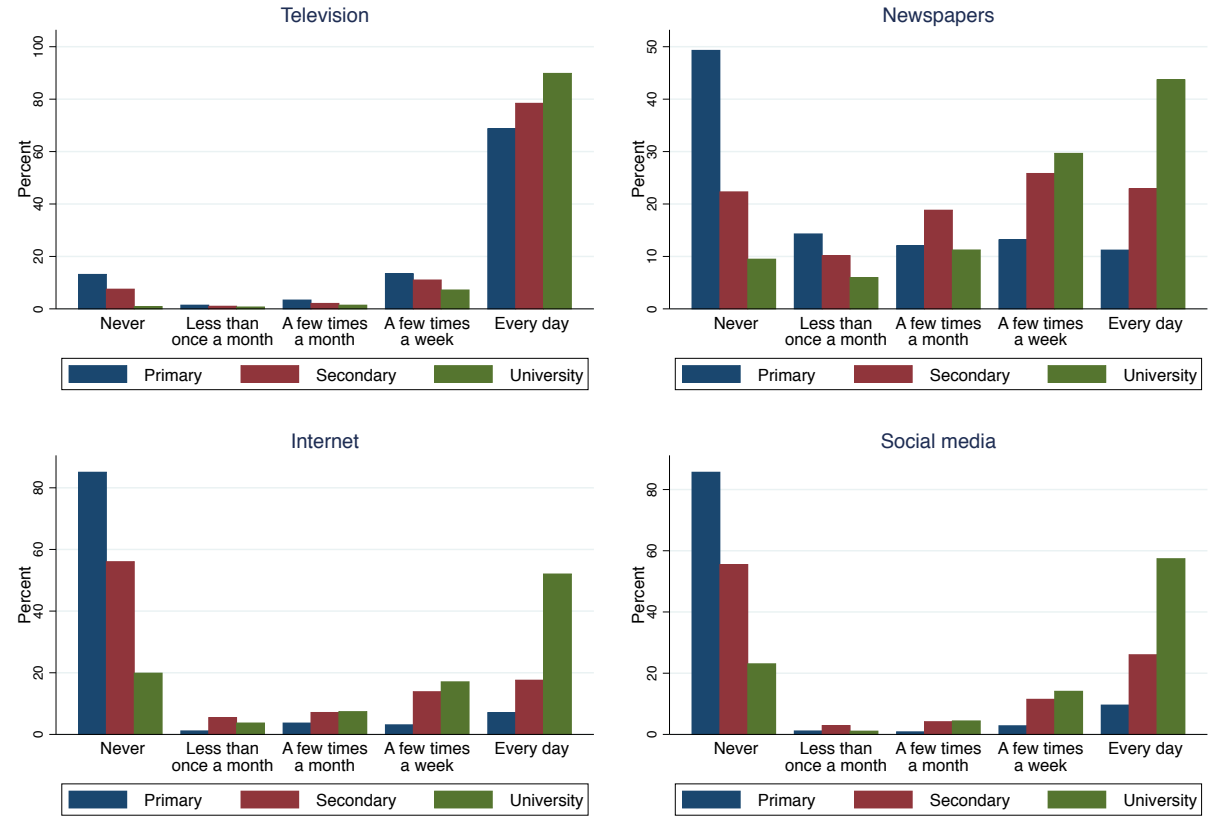
For each type of media (radio is excluded for the sake of conciseness, yet its statistics are very similar to the ones on TV) the charts depicts the frequency of its usage by age group (age quartiles). As expected, the use of television does not vary substantially across

different age categories, with approximately 80% of the people in each age group watching TV to get information on a daily basis. The situation slightly differs for newspapers. Here we can see that relatively old people are more likely to read newspapers every day and, at the same time, they are also more likely not to read them at all. Apparently, as people get older their choice on readership gets more polarized.

As on the Internet and social media usage the picture is remarkably different. Here the reader might observe that only relatively young people use the Internet or the social media as source of information on a regular basis. Most importantly a large fraction of old people (almost 80%) never access information through to this type of media. These charts suggest that we should expect to find a larger effect of Internet coverage in those places with higher proportion of youth, as they represent the segment of the population that most likely utilizes this type of media.

What about heterogeneity across education levels? Figure 3 below reports the frequency in media usage by different education categories.

Figure 3: Use of media across educational groups



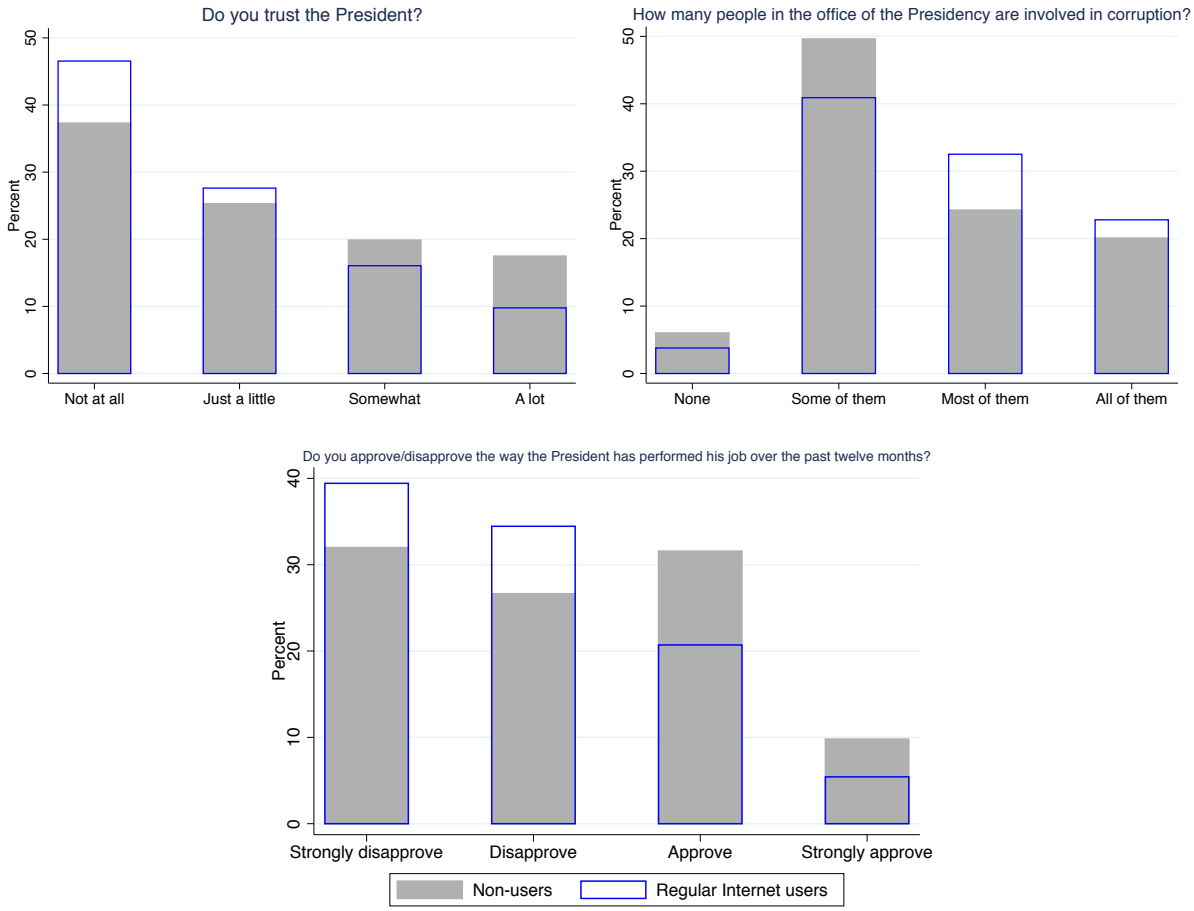
Notes: Author's own calculations based on Afrobarometer.

Information through television is accessible from all the groups in a pretty similar way. However, for newspaper the situation is different, with the majority of university graduates reading newspapers every day. By contrast, almost 50% of low-educated individuals never read newspapers. Similar patters appear when we look at Internet and social media usage across different education groups. In particular, above 80% of the people with primary education never use these media. Meanwhile, almost 50% of highly-educated individuals use them every day. The reasons explaining these patterns are not revealed by this simple analysis, yet one may think that higher education is associated with higher usage through

income. In other words, only richer (i.e. highly educated) individuals can afford to buy the necessary technology (smartphone, laptop, Internet subscription, etc.). Overall, these figures might imply that the expected effect of Internet coverage is larger in places where the average level of education is also higher.

The previous figures show that a non-negligible proportion of individuals, especially if young and educated, were used to access information through the Internet and social media in 2015. If this new source of information actually played a role in revealing to the people about corruption scandals inside the ruling party, then one might expect to observe heterogeneous opinions towards the current ANC leader and President across Internet users and those who never use it. Figure 4 below shows the answers to three questions on the President and his office for two distinct categories: regular Internet users (those that use the media as a source of information at least a few times a week) and non-regular users.

Figure 4: Opinions towards the presidency by Internet usage



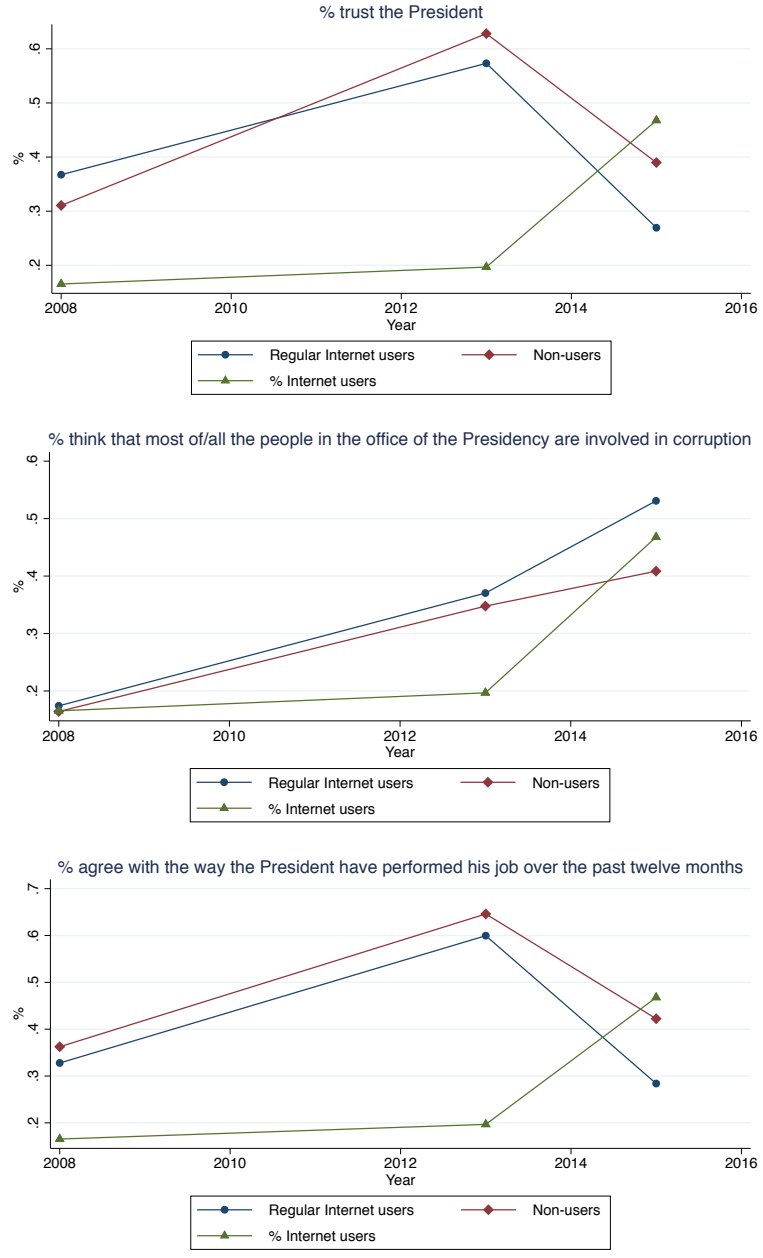
As expected, respondents who use the Internet regularly are less likely to trust the President and more likely to think that most of the people in his office are involved in corruption and to strongly disapprove the way the President performed his job in the last year. Similar results are obtained when considering other questions on MP and local governors. The negative correlation between Internet usage and ANC/President popularity holds when controlling for the use of other media, the level of education and the age of the respondents.[5] Finally, Figure 13 in the appendix shows the answers to these

[5]Results from various OLS regressions are not reported here for the sake of conciseness.

questions for regular newspapers readers and non-readers. The differences in the answers across the two groups are remarkably smaller than those reported in Figure 4.

The correlational evidence reported above might still be spurious. In other words, the negative correlation can be driven by unobserved factors that are likely to affect respondents' opinion and their Internet usage. Because of this possible issue, looking at past trends may help. Hence, to provide additional evidence in line with the idea that the Internet could actually play an important role in the political scenario, I depict the evolution of the answers to the previous three questions over the last years.

Figure 5: Trends in the opinions towards the presidency by Internet usage



Notes: Author's own calculations based on Afrobarometer.

For this purpose I aggregate the previously depicted possible answers and create a binary indicator. Figure 5 above shows the temporal patterns along with the trend in the percentage of Internet users. Trust on the President was higher for Internet users in

2008. However, despite its increase, in 2013 for these people it was lower compared to non-Internet users. Most importantly, in 2015 the gap between the two groups was even larger. The percentage of people that think the President is involved in corruption constantly increased over time, and its growth was similar across the two groups before the large Internet expansion. In fact, after 2013 this number rose remarkably more for Internet users than non-users. A similar pattern is visible in the last graph, where we observe that the drop in the President's popularity is higher for Internet-users than non-users.

The descriptive evidence I reported in this section partially sheds light on the potential mechanisms that may drive the causal relationship between Internet coverage and political outcomes. In particuar, it might be useful as a starting guidance for the interpretation of the potential findings of this paper and to perform the heterogeneity analysis.

4. Empirical strategy

4.1 Baseline specification

I seek to identify the causal impact of Internet penetration on political outcomes. For this purpose, I will exploit the temporal and spatial variation in coverage and ballot results that occurred over the period 2011-2016. However, Internet allocation is far from random, yet it reflects market-based calculations and profit optimization choices performed by the three private ISPs. Endogeneity issues are perhaps less severe than when considering landline Internet, being the mobile technology more difficult to restrain in a given precise area by its electromagnetic nature. However, the decisions on the region a private operator aims at targeting and the place in which the transmitter is located are endogenous.

Therefore, I draw from a World Bank report by Buys at al. (2009) which studies the determinants of disparities in cell phone coverage in Sub-Saharan Africa in order to understand the predictors of mobile internet coverage. The report highlights that both demand and supply side factors play a role. Among other things, it shows that places with larger potential market size (as measured by per capita income), lower elevation and smoother terrain characteristics are significant determinants of better coverage.

Taking this study into account, the empirical strategy of this paper is based on a Diff-in-Diff approach that exploits the high-resolution features of the data at the voting district level or unions of them. The main specification I estimate is the following:

$$y_{it} = \beta_0 + \beta_1 C_{it} + \beta_2 Post_{it} + \beta_3 C_{it} * Post_{it} + \beta_4 X_{it} + \beta_5 W_i * Post_{it} + \mu_i + \varepsilon_{it} \qquad (1)$$

Where i is the index for voting districts or unions of them, and t is the year level time index, such that $t = 2011, 2016$. C stands for $Coverage$, which is the mean Internet coverage of each locality and $Post$ is a dummy variable equal to 1 in year 2016. X contains time variant controls such as (the log of) luminosity and population density, while W is a vector of time constant variables measured at the baseline (2011) also including 5^{th} degree polynomials of (the log of) luminosity and population density in 2011. Finally, μ represents voting district (VD) fixed effects. Variable y is alternatively turnout and the vote shares of ANC, DA and EFF parties.

Specification reported in (1) allows me to account for various sources of potential endogenity. In particular, voting district fixed effects account for time invariant unobserved

factors that affect the outcome and may also be correlated with Internet coverage. The year dummy capture instead the time trend in the outcomes that is common to all localities. Finally, interacting baseline controls with the time dummy I allow demand and supply side factors to have different impacts on the outcome over time. I cluster standard errors at the level of Local Municipality (234 clusters), to account for both cross-sectional and temporal correlation in the errors.

In order for β_3 to consistently identify the Average Treatment Effect of Internet coverage on the outcomes of interest the following identification assumption is necessary. Conditional on demand side factors (mainly GDP and population density), on supply side (geographical) features, and on voting district and time fixed effects, Internet coverage is exogenous.

Although in principle I could use observations at 3 points in time (2006-2011-2015) I prefer to concentrate my analysis on the last two electoral rounds. The main reason is that Internet coverage, conditional on the usual observables, is more likely to be exogenous in this period than between 2006-2011, i.e. the period of its first installation.

In order to provide some evidence in favor to the identification assumption I perform some tests. The first one is a standard AET (Altonji, Elder and Taber 2005) test. This approach uses the degree of selection on observed variables as a guide to the degree of selection on the unobservables. Specifically, in my context I regress the outcome variables on the component of *Coverage* that is explained by geographical and economic observables. If the coefficient on predicted coverage is not statistically significant, then this might suggest that the "exogenous" component of it actually matters and drives the potential results.

The second way to provide evidence in favor of the conditional exogeneity assumption is to carry out placebo estimates, regressing electoral outcomes before 2006 on Internet coverage after 2011. In particular, to conduct this test I look at political outcomes between the 2000 and 2006 municipal elections. Results from such a procedure are in favor of the exogeneity assumption if future *Coverage* does not significantly predict any political outcome over that period.

Finally, as an additional visual robustness check for my estimates I calculate the predicted counterfactual distribution of the variable of interest in the absence of Internet coverage. I will also show figures reporting the trends of the outcome variables over time for different quartiles of *Coverage* to inspect potential parallel trends before the Internet penetration.

4.2 Testing the mechanisms

In order to obtain some suggestive evidence on which potential mechanisms could drive the causal relationship between Internet and political outcomes I develop an extensive heterogeneity analysis. The basic idea behind this analysis is to draw from the descriptive evidence of section 3 along with the insights from the political scenario described in section 2 and then examine if and how Internet coverage plays a differential role across various socio-economic conditions and combinations of them. This analysis is conducted in two different and somehow complementary ways.

The first one is based on a triple-difference estimation (DDD). This has a twofold purpose. On the one hand, a DDD estimation provides a further tool to tackle potential

endogeneity of Internet coverage. On the other hand, it gives additional insights on the information role of the Internet in localities where the probability of corruption scandals is higher. Specifically, the DDD strategy takes advantage of the temporal and spatial variation in *Coverage* along with the variation in wasteful, irregular and unauthorized expenditure per capita across municipalities. In fact, localities where this expenditure is higher should be more likely to experience corruption scandals. If the Internet is used as a source of information, then one would expect to find significant effects of *Coverage* in municipalities with a high level of irregular expenditure per capita and where connectivity is good. To examine this channel I estimate the following regression:

$$\begin{aligned} y_{imt} = \quad & \beta_0 + \beta_1 C_{imt} + \beta_2 Post_{imt} + \beta_3 C_{imt} * Post_{imt} + \beta_4 Exp_m * Post_{imt} + \\ & \beta_5 C_{imt} * Exp_m + \beta_6 C_{imt} * Post_{imt} * Exp_m + \theta X_{it} + \rho W_i * Post_{imt} + \mu_i + \varepsilon_{imt} \end{aligned} \tag{2}$$

where the usual notation applies, and m represents the Local Municipality the observation i belongs to. Exp_m is the sum of wasteful, irregular and unauthorized audited expenditure per capita of municipality m in 2014, that is the last year for which the information is available. Notice that standalone Exp_m is omitted because of the inclusion of VD fixed effects. In this case vector W also includes the total expenditure of each municipality in 2014 to control for differences in the level of development across municipalities. The main coefficient of interest is β_6, the one associated to the triple-interaction term. According to the descriptive evidence presented in section 3 this coefficient should be negative when Y is the ANC share, reflecting the voters' disapproval with the President's politics through the electoral process.

The second strategy to study the potential mechanisms is simpler and it also involves the use of data on socio-economic variables from the 2011 Census. In particular, I split the original sample in various sub-samples according to observable and possibly relevant characteristics measured at baseline (2011). These include income and population density to assess potential heterogeneity across rural and urban areas; distance from main cities, average terrain roughness and the percentage ownership of traditional media (TV and Radio) to study differences between more and less "isolated" places. The idea is that circulation of information should be lower in more isolated localities, hence one may expect a higher impact of the Internet arrival there. In addition, I look at heterogeneity effects of coverage across municipalities with high and low average level of education. In this case, according to section 2 one should expect to observe a larger impact in localities with higher average years of schooling, being highly-educated individuals those who use the Internet more regularly. Finally, I consider the proportion of young people (individuals aged between 18 and 35) and look at heterogeneity effects across municipalities with larger and smaller youth population. In line with the descriptive evidence, places with more youths should respond more to the Internet arrival, as these people are the most regular Internet users. For each of the identified sub-samples (and combination of them), I run regression (1) and (2) to understand which characteristics are more suitable to explain the observed relationship.

5. Data and descriptive statistics

5.1 Dataset construction and sources

The research is based on a newly-assembled dataset containing time-varying and geo-referenced information on political outcomes (turnout and vote share of the main parties), and the spatial mean of 3G Internet coverage, economic development (as measured by luminosity) and population density at the voting district level or unions of them. In addition, each observation contains time constant information on average elevation (m) and roughness (m^2), area (km), a dummy variable indicating the presence of a major road, and the distance (km) from the closest city with population over 1 million people.

The construction of the dataset involved two steps.

Firstly, the creation of new geographical units of analysis. In fact, data on political outcomes come at a very disaggregated level, that is the voting district. However, the boundaries and the number of these districts change over time. In 2006 there were 18,872 voting districts, 20,857 in 2011 and 22,612 in 2016. Hence, the first step was to create a stable geographically and time invariant unit of observation. This has two purposes: on the one hand it allows comparability of electoral outcomes over time, on the other hand it solves the problem of endogenous change in the district boundaries. To create new units of observation I used a Python algorithm that combine neighboring and mutable voting districts until it reaches a stable aggregation. Therefore, the output of the algorithm gives the smallest possible unions of voting districts whose borders are constant between 2006 and 2016. In particular, the algorithm created 10,073 stable clusters. I replicate a similar procedure to create stable observational units between 2000 and 2006, that is the period I will use to perform placebo tests. In this case the algorithm created 8,330 observations.

Secondly, for each of these units I calculated zonal statistics (mean and standard deviation) of the above mentioned variables using the Python GIS toolbox. In particular, the following list contains details on the sources I used.

Administrative data on political outcomes for municipal elections come from the Independent Electoral Commission (IEC) of South Africa. These data are freely available on the IEC website (www.elections.org.za). However, only the most recent shape-files are available online. I obtained the comprehensive set of shape-files for all years since 2000 directly from the IEC office. Voting data contains information on the total number of potential voters, the actual number of those who actually voted, and the number of votes each party got in each voting district. For the newly formed party EFF data before 2016 are missing. In order to include this party in the analysis I computed the predicted share of votes this party could have received in 2011 using observables. In particular, I used the estimated coefficients from regressing the actual share of votes in 2016 on the observables as main predictors for the outcome in 2011. In addition, I used the number of potential voters as a proxy for population for each voting district.

Data on Internet coverage come from Collins Mobile Coverage Explorer, a web based roaming coverage map service made available through Collins Bartholomew's partnership with the GSMA. Using the latest mapping technology, the company combines up-to-date world base maps with unique mobile network coverage data provided by operators from around the world. These maps are then delivered to network operators to help them tell their users where they can use their phones when abroad. The data that have been licensed collate, for all years between 2007 and 2015 (but with the exception of 2010), the most recent submission during that year from all member operators in each country.

The dataset comes in GIS vector format and for each country provides 2G, 3G and 4G coverage, separately: each pixel has value 1 if covered, 0 otherwise. In South Africa the geographical precision varies from year to year, with a maximum pixel size of 1km by 1km (up to almost 200m x 200m in the most recent version). I exploit only 3G coverage data since there is practically no variation in 2G or 4G technologies between 2009 and 2015: almost all places had 2G already before 2009, almost no place had 4G in 2015 (nor before). To proxy coverage in 2016 I use Internet coverage in 2015, as this depicts the situation up to December of that year. Measurement error associated to this approximation should be small. In fact, municipal elections in 2016 were held on the 3rd of August. The procedure assumes that 3G coverage did not change abruptly in the months right before elections. If it did, my estimate would represent a lower-bound for the actual effect.

Luminosity (nigh-light) is used as a proxy for economic development. These data are collected by the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) satellite program and is maintained and processed by the National Oceanic and Atmospheric Administration (NGDC 2010, Baugh et al. 2009). They are available for download in GIS vector format at http://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html. I use data until 2013, which is the last available year. Each pixel (1 square kilometer) in the luminosity data is assigned a digital number (DN) representing its luminosity. The DNs are integers that range from 0 to 63. The higher this number is, the greater the economic activity in the pixel is. The reader may look at Pinkovskiy and Sala-i-Martin (2016) for a recent application of this dataset as a proxy for GDP in Africa. Unfortunately, being these data available only up to 2013 I had to predict nigh-light for 2015/16. I applied a version of linear interpolation by assuming for each observation a constant growth between 2013 and 2015, using the observed growth experienced between 2011 and 2013.

Data on education, monthly income, TV/radio/cellphone ownership and age distribution at the municipality level come from the 2011 population Census. These are freely available from the official website of Statistics South Africa (http://www.statssa.gov.za). In addition, for each municipality I also retrieve information on the total expenditure[6] and on the level of wasteful, irregular and unauthorized expenditure[7] available on the audited financial statements produced by the Department of National Treasury (https://municipaldata.treasury.g In this case I consider data for 2014, being this the most recent year for which the information is available.

Data on elevation and ruggedness (as measured by the standard deviation of the elevation) at similar resolution come from the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). The dataset is hosted by the Earth Resources Observation and Science (EROS) Center and is freely available at https://topotools.cr.usgs.gov/gmted_viewer/. Moreover, the shapefile containing information on the main roads comes from the ESRI archives. Finally, information on the land use (i.e. cropland vs. pasture area) and urbanization are drawn from the EarthStat database. This contains figures on agricultural activities measured in 2000 and constructed by combining agricultural inventory data and satellite-derived land cover data. For these geographical variables I compute the spatial mean for each voting district.

[6]Statement of Financial Performance: how a municipality has spent money and received income.

[7]Specific expenditure amounts from audited financial results, recorded in the notes to the annual financial statements of each municipality.

The analysis is conducted after cleaning the dataset and considering only observations which farer than 10 km from the closest main city[8], and such that their population density in 2011 was lower than the 90th percentile. Hence, overall 1424 observations are dropped. This is mainly done in order to avoid contagion effects due to the possible, yet fairly negligible expansion of the landline broadband Internet. In fact, excluding the major urban agglomeration mitigates potential confounding effects that may create biased results. Nevertheless, including these observations does not substantially alter the magnitude of the estimated coefficients.

5.1 Descriptive statistics

Table 2 provides the descriptive statistics for the final sample. The time-varying variables are displayed for the years 2006, 2011 and 2016. Notice that the table also reports the statistics for some variables measured at the municipality level in 2011. In this case, the total number of observations corresponds to the number of municipalities, i.e. 234. Finally, Figure 6 below shows the geographical expansion of the Internet coverage between 2011 and 2016 for my units of analysis.

Table 2: Descriptive statistics

Variable	Year	N. of obs.	Mean	Median	St. Dev.	Min	Max
Turnout	2006	8736	0.55	0.55	0.15	0	1
	2011	8736	0.58	0.59	0.10	0	1
	2016	8736	0.57	0.57	0.10	0	1
ANC share	2006	8530	0.65	0.75	0.29	0	1
	2011	8721	0.63	0.71	0.28	0	1
	2016	8703	0.60	0.66	0.26	0	1
DA share	2006	8530	0.11	0.02	0.21	0	1
	2011	8721	0.16	0.03	0.25	0	1
	2016	8703	0.18	0.05	0.26	0	1
EFF share	2006*	8497	0.08	0.07	0.05	-0.05	0.34
	2011*	8688	0.07	0.07	0.05	-0.09	0.32
	2016	8703	0.06	0.03	0.07	0	0.62
Coverage	2006	8736	0.16	0	0.35	0	1
	2011	8736	0.28	0	0.41	0	1
	2016	8736	0.62	0.68	0.36	0	1
Luminosity	2006	8736	11.70	2.14	19.03	0	63
	2011	8736	13.23	3.00	20.45	0	63
	2016	8736	14.72	5.42	16.61	0	63
Pop. Density	2006	8736	313.64	39.82	645.24	0	5253.54
	2011	8736	352.10	43.16	695.18	0	3302.79
	2016	8736	388.10	47.47	760.90	0	4665.39
Road		8736	0.21	0	0.41	0	1
Urban area		8736	0.33	0.28	0.29	0	1
Elevation (m)		8736	887.62	906.02	503.89	1.32	2485.74
Roughness (m2)		8736	57.55	39.52	56.79	0	505.00
Distance city (Km)		8736	247.20	233.78	160.80	10.01	798.88
Area (Km2)		8736	140.41	15.86	663.37	0.10	36373.35

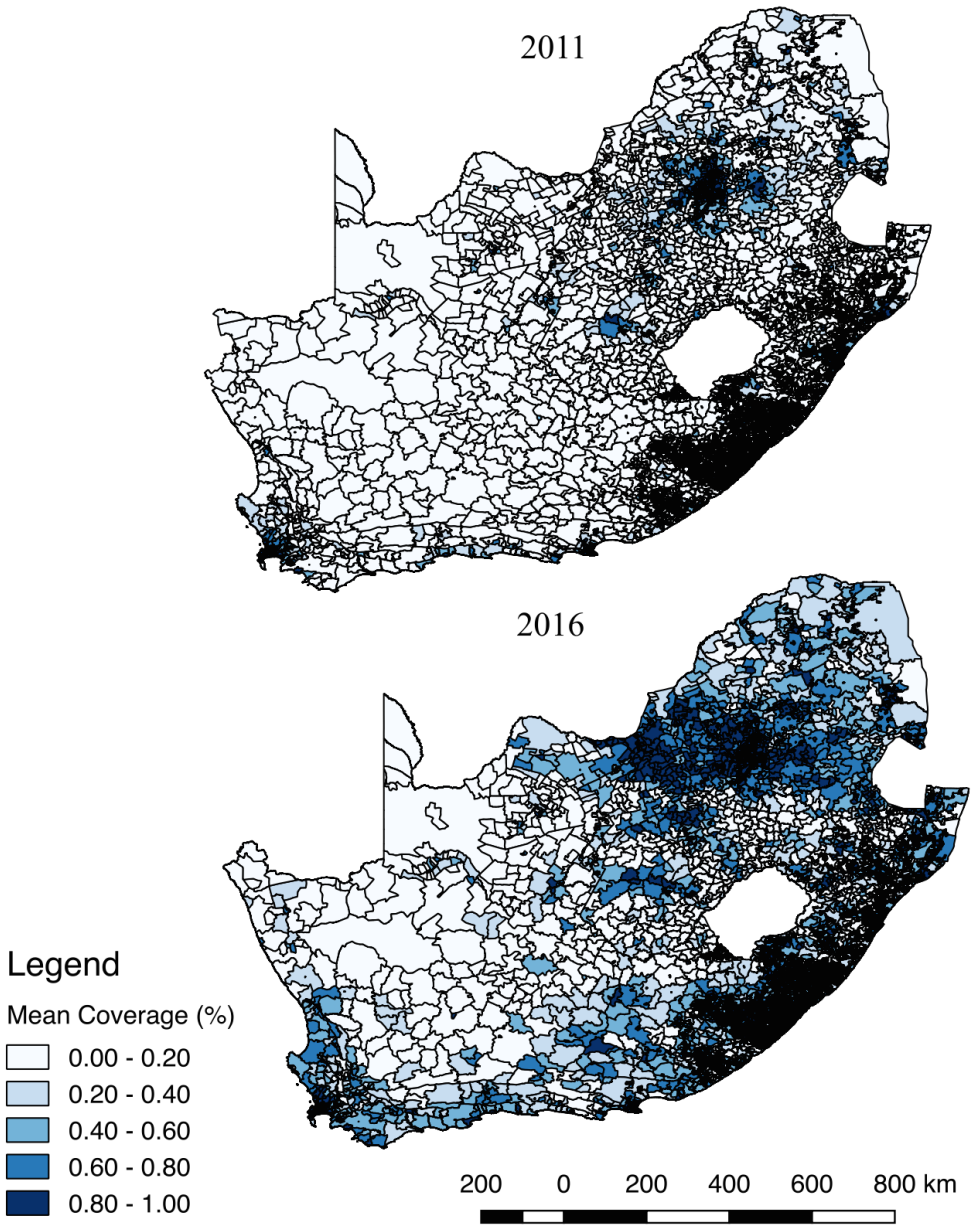
[8]I consider cities with more than 1 million inhabitants. These are Cape Town, Durban, Johannesburg and Pretoria.

Table 2 continued: DESCRIPTIVE STATISTICS

Population	2011	234	221200	101922	521230	6984	4434084
Share of youth	2011	234	0.31	0.31	0.04	0.23	0.42
Education	2011	234	6.59	6.47	0.83	4.67	8.84
Has radio	2011	234	0.64	0.65	0.08	0.25	0.80
Has TV	2011	234	0.68	0.71	0.13	0.28	0.88
Has cellphone	2011	234	0.84	0.85	0.06	0.60	0.95
Has Internet	2011	234	0.26	0.25	0.08	0.12	0.51
PC Monthly income (ZAR)	2011	234	2221	2025	1140	795	6956
PC Corruption Exp. (ZAR)	2011	234	1546	368	5191	0	65062
PC Total Exp. (ZAR)	2011	234	3405	3348	2217	418	10257

*Notes: * indicates that these variables have been estimated. Ownership of radio, TV and cellphone and access to Internet are measured as a fraction out of the total number of households. PC stands for per capita values.*

Figure 6: MOBILE INTERNET PENETRATION



6. Results

6.1 Main findings

This section describes the main findings of this paper. The first dependent variable I consider is voter turnout. A priori it would be difficult to understand if and how Internet coverage in South Africa affected this variable. Columns (1-2) in Table 3 show that places that faced a better Internet coverage between 2011 and 2016 experienced a significant positive variation in voter turnout.

Table 3: Diff-in-diff estimates of the impact of mobile internet coverage on political outcomes

		Turnout		**ANC share**	**DA share**	**EFF share**
	(1)	(2)	(3)	(4)	(5)	(5)
Coverage	-0.029***	-0.026***	-0.009	0.022	-0.006	0.016**
	(0.008)	(0.008)	(0.008)	(0.021)	(0.008)	(0.007)
Post (=1 if year is 2016)	-0.016***	-0.017***	0.049	-0.014	0.170***	-0.056
	(0.005)	(0.006)	(0.035)	(0.086)	(0.059)	(0.035)
Coverage • Post	0.027***	0.022***	0.012	-0.064**	0.022**	-0.014*
	(0.008)	(0.008)	(0.010)	(0.029)	(0.010)	(0.008)
log(Luminosity)		-0.001	0.000	-0.024	0.003	0.006
		(0.004)	(0.006)	(0.022)	(0.007)	(0.006)
log(Pop. Density)		0.043***	0.037**	0.073***	-0.030*	0.027**
		(0.014)	(0.015)	(0.025)	(0.016)	(0.013)
log(Luminosity) • Post			0.039	-0.135*	-0.016	0.019
			(0.030)	(0.075)	(0.031)	(0.027)
log(Pop. Density) • Post			0.011	-0.010	-0.025	0.083***
			(0.039)	(0.058)	(0.051)	(0.032)
Road • Post			0.010**	0.008	-0.003	-0.006*
			(0.004)	(0.008)	(0.005)	(0.003)
Urban area • Post			-0.004	0.016	-0.015**	0.004
			(0.008)	(0.016)	(0.007)	(0.006)
log(Elevation) • Post			-0.001	-0.012*	-0.010**	0.003
			(0.003)	(0.007)	(0.004)	(0.002)
log(Rughness) • Post			-0.001	0.013**	0.009***	-0.004
			(0.003)	(0.006)	(0.003)	(0.003)
log(Distance city) • Post			-0.008**	-0.000	-0.007*	0.000
			(0.003)	(0.009)	(0.004)	(0.003)
log(Area)• Post			-0.002	-0.011**	-0.004*	0.004
			(0.002)	(0.004)	(0.002)	(0.002)
Turnout				0.084	-0.108***	-0.066**
				(0.058)	(0.031)	(0.029)
Turnout • Post				0.195***	-0.052*	-0.092***
				(0.059)	(0.032)	(0.030)
VD Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	17,472	17,472	17,472	17,424	17,424	17,391
Adjusted R2	0.661	0.662	0.670	0.791	0.944	0.329

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses are clustered at the Local Municipality level (234 clusters). Up to the 5th order polynomial terms for (the log of) Luminosity and Pop. Density interacted with Post are included among the controls. VD stands for Voting District or union of them.*

In particular, the coefficient of interest (Coverage • Post) is positive and strongly significant in both specifications. It suggests that a unitary increase in coverage – i.e.

moving from 0 to 100% of the area covered by the Internet – causes turnout to go up by approximately 1.7 percentage points (pp), on average. This is almost 3% of the mean of the dependent variable at the baseline (2011). However, when the full set of controls is included into the equation the magnitude of the coefficient drops and its significance vanishes. In the most conservative specification reported in column 3 there seem to be no evidence supporting any causal relationship between Internet coverage and turnout. In fact, note that the coefficient is no more statistically significant, and this is largely due to the reduction in its magnitude compared to columns (1) and (2) following the inclusion of important control variables. Fixed effects at the voting district level (or unions of them) are included in all specifications.

Column (4) displays the results for the vote share of the ANC party. For conciseness I only report results for the most conservative and complete specification. The coefficient of interest is negative and significant at the 5% significance level. This finding is in line with my expectations. In particular, it suggests that a unitary coverage expansion causes a decrease in popularity of the ruling party of approximately 6.4 pp, which is 10% of the mean of the dependent variable at the baseline. Alternatively, a one-standard deviation increase in coverage causes a drop in ANC vote share of about 2.6 pp, that is 4% of its mean at the baseline.

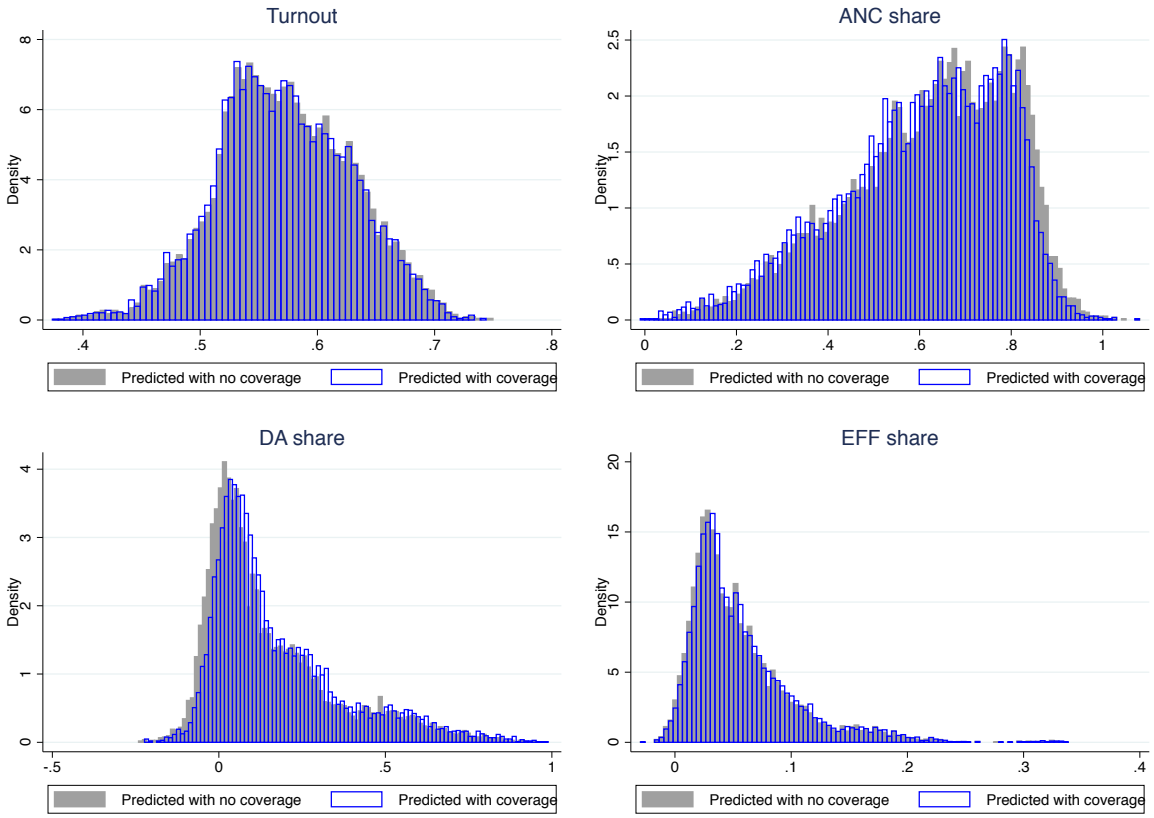
The magnitude of the estimated coefficient is therefore quite impressive, and it seems to support the hypothesized mechanism. In other words, it can be that more voters in localities covered by the Internet decided to punish the incumbent party because of the type of information they received thorough the web. Although this specification is not informative on the exact mechanism behind this relationship, it still provides supporting evidence for the role the Internet played in the last municipal elections. The following analysis will try to analyze to what extent the additional information on corruption scandals that voters received through the web might have driven the results.

Finally, the last two columns of Table 3 display results for the major opponent parties. In particular, column (5) shows that a unitary increase in coverage causes the DA share of votes to rise by 2.2 pp, which is almost 14% of the mean of dependent variable at the baseline. This is in line with the idea that some of the ex ANC supporters decided to reward the party that denounced corruption scandals because they came to know about them through the Internet. Surprisingly, column (6) suggests that neither the EFF party benefited from the Internet, although the coefficient is only significant at the 10% significance level. In particular, its share of votes drops by 1.4 pp as a result of a unitary increase in coverage. However, this result is likely to be imprecise because of the noise introduced by estimating the potential outcome of the EFF in 2011. Therefore, one should be careful in interpreting the causal relationship between the Internet and the EFF share of votes.

6.2 Robustness

Before the examination of the potential mechanisms I perform some additional exercises to prove the robustness of the applied methodology and the respective estimates. As a visual check on the actual impact of coverage on political outcomes I compute the counterfactual distributions in the absence of coverage in year 2016. These capture how the overall vote share/turnout distribution would look like if the Internet did not expand. The figure below shows the results.

Figure 7: Actual and counterfactual predicted distributions



The blue density functions represent the predicted scenario under Internet coverage, while the gray ones depict the hypothetical case without coverage (counterfactual distribution), that is what would have happened if coverage was absent. The histograms are in line with the previous results. In particular, there is no significant difference in the actual and counterfactual predicted distributions when we look at voter turnout and EFF vote share. In fact, the histograms perfectly overlap. By contrast, for the ANC the predicted counterfactual shows that the party would get more votes in a scenario with no Internet penetration. In fact, the gray distribution shifts to the right with respect to the blue one. At the same time, the DA party would loose votes if Internet coverage was absent.

To convince the reader that the estimated coefficients actually capture causal effects placebo tests are extremely useful. In particular, the idea behind these tests is to regress the outcome variables observed in years before the introduction of the mobile technology on Internet coverage in year 2011 and 2016. If coverage is exogenous, then one should find no effects in the results of the test. The optimal time-window to conduct the test is to consider the outcome variables in the period 2000-2006. In fact, in this period no variation in mobile Internet occurred (apart from the main cities, which are anyway excluded from the sample). Columns (1-3) of the following table display the results. The reader may notice that Internet penetration is uncorrelated with political outcomes between 2000 and 2006. In fact, the coefficients on the interaction terms are largely insignificant for ANC and DA vote share. As on turnout, the negative relation is weak and perhaps driven by the large overall drop in political participation between 2000 and 2006. Overall, placebo estimates are in favor of the conditional exogeneity assumption.

Table 4: Placebo estimates

	Turnout 2000-2006 (1)	ANC share 2000-2006 (2)	DA share 2000-2006 (3)	Turnout 2006-2011 (4)	ANC share 2006-2011 (5)	DA share 2006-2011 (6)
Placebo Coverage	0.033	0.006	0.024**	-0.006	0.047**	-0.036***
	(0.021)	(0.028)	(0.011)	(0.013)	(0.020)	(0.010)
Post	-0.241***	0.171	-0.007	-0.068	0.027	0.218**
	(0.085)	(0.110)	(0.059)	(0.056)	(0.116)	(0.107)
Placebo Coverage • Post	-0.036*	-0.030	-0.010	-0.012	-0.113***	0.056***
	(0.020)	(0.030)	(0.012)	(0.017)	(0.024)	(0.012)
log(Luminosity)	0.016	0.016	0.013**	0.033***	0.015	0.017***
	(0.016)	(0.024)	(0.006)	(0.012)	(0.022)	(0.006)
log(Pop. Density)	-0.032***	0.034**	-0.020***	-0.049***	0.059***	-0.038***
	(0.012)	(0.014)	(0.007)	(0.018)	(0.021)	(0.013)
Controls • Post	Yes	Yes	Yes	Yes	Yes	Yes
VD Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	11,642	11,344	11,344	17,472	17,251	17,251
Adjusted R2	0.595	0.815	0.893	0.569	0.798	0.926

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses are clustered at the Local Municipality level (234 clusters). Placebo Coverage is Internet coverage between 2011 and 2016. In columns (1-3) Post is = 1 if year is 2006, while in columns (4-6) Post is = 1 if year is 2011. Up to the 5th order polynomial terms for (the log of) Luminosity and Pop. Density interacted with Post are included among the controls. VD stands for Voting District or union of them.*

As a comparison tool, I also conduct placebo tests using outcomes between 2006-2011. However, as Figure 13 in the appendix shows, actual Internet penetration significantly expanded over this period. Therefore, one should not expect to find insignificant results, being the expansion of this technology highly correlated over space and time[9]. Hence, results of the test performed over this period would be misleading.

A visual inspection of the trends in the outcome variables across different quartiles of 2016 Internet coverage might be useful to spot potential parallel patterns. Figure 14 in the appendix serves this purpose. It reports unconditional means of turnout, ANC and DA vote share by quartile of coverage between 2000 and 2016.[10]

Finally, I perform a standard AET test. This strategy is useful in cases in which doubt remains about the exogeneity of the treatment variable. The approach uses the degree of selection on observables as a guide to the degree of selection on the unobservables. In particular, it involves two steps. Firstly, I regress Internet coverage in 2016 on a bunch of potentially relevant predictors and then I calculate its fitted values. Secondly, I regress political outcomes in 2016 on the predicted coverage to assess the extent to which its plausibly endogenous component may affect these outcomes. The following table shows these steps.

[9]The correlation coefficient between coverage in 2011 and coverage in 2016 is .70

[10]Note that two different samples have been used to generate these figures. One sample has been used for the period 2000-2006, and another one for the period 2006-2016. Additional work will be carry out to create a unique sample.

Table 5: AET test results

	Turnout	**ANC**	**DA**	**Coverage**	**Turnout**	**ANC**	**DA**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Predicted Coverage					0.007	-0.013	-0.045
					(0.015)	(0.042)	(0.037)
Coverage	-0.003	-0.030**	0.044***				
	(0.005)	(0.014)	(0.009)				
log(Luminosity)	-0.025***	-0.094***	0.095***	0.133***			
	(0.004)	(0.012)	(0.014)	(0.012)			
log(Pop. Density)	0.009***	0.042***	-0.051***	0.033***			
	(0.002)	(0.005)	(0.006)	(0.004)			
Road	0.006*	-0.043***	0.050***	0.054***			
	(0.004)	(0.010)	(0.010)	(0.010)			
Urban	-0.004	-0.002	0.036*	-0.042**			
	(0.007)	(0.020)	(0.019)	(0.020)			
log(Elevation)	-0.008	0.010	-0.029	0.016			
	(0.007)	(0.014)	(0.020)	(0.013)			
log(Roughness)	0.013***	0.001	0.004	-0.064***			
	(0.003)	(0.007)	(0.008)	(0.009)			
log(Distance City)	-0.034**	0.172***	-0.197***	-0.012			
	(0.017)	(0.047)	(0.056)	(0.027)			
log(Area)	-0.007***	0.004	-0.007	0.000			
	(0.002)	(0.004)	(0.005)	(0.005)			
VD Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	8,736	8,703	8,703	8,736	8,736	8,703	8,703
Adjusted R2	0.287	0.482	0.572	0.618	0.268	0.436	0.507

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses are clustered at the Local Municipality level (234 clusters). Observations refer to year 2016. VD stands for Voting District or union of them.*

Columns (1-3) show the correlation between political outcomes and coverage in 2016, controlling for the usual observables. Then, column (4) depicts the potential economic and geographical determinants of Internet penetration. I use the estimated coefficients from this column to compute the predicted coverage. Finally, columns (5-7) report the results of the AET procedure. The coefficients on predicted coverage are not statistically significant for any political outcomes. This suggest that the meaningful correlation between coverage and ANC/DA shares depicted in columns (2-3) is mainly driven by the plausibly "exogenous" component of coverage. Overall, these results foster the findings of Table 3.

6.3 Mechanisms

Results of section 6.1 are partially in line with the findings of Miner (2016) for Malaysia. Specifically, he finds that Internet access seems to jeopardize the reputation and weaken the popularity of the ruling parties in developing countries.

Why do we see these effects? Does this mean that traditional media are biased towards the incumbent? Answering these questions is quite challenging, and it requires knowledge about substitution effects among different media, on what people exactly do with the new technology, and which segments of the population are mostly affected by it. Suggestive evidence from section 3 shows that in 2016 around 39% of the population utilizes the Internet as a source of information on a regular basis. Also, approximately 47% of the population regularly uses the Internet for any kind of purpose. This means that, among

regular Internet users, almost 83% of them use the Internet also to get some news. This share is quite impressive and suggests that for many people in South Africa the web is not only a source of entertainment, rather its informative power is large. At the same time, there seems to be little or insignificant substitution of traditional media with the new ones. Therefore, the information that Internet users get from the web add up to the one coming from traditional sources.

The main findings of section 6.1 are in line with the idea that this additional information is, somehow, damaging the reputation of the ANC party. Specifically, we have seen that Internet users are more likely to think that most of the people in the office of the President are involved in corruption. Is it really possible that the ANC party lost votes because a segment of the population was informed about the overall unsatisfactory administration and corruption scandals?

In order to provide some causal evidence on this I provide results from a triple difference estimation described by equation (2) where I exploit the difference in corruption-related expenditure per capita across municipalities as an additional source of variation. In places where this illegal expenditure is higher, voters are more likely to be exposed to corruption scandals, especially if they are covered by the Internet. That is, among localities with similar corruption level, those with coverage should exhibit an additional drop in the ANC vote share, being their inhabitants more likely to know about the scandals. Table 6 provides the results.

Table 6: Triple difference estimates

	All municipalities				**Municipalities where ANC was ruling**			
	Turnout	**ANC**	**DA**	**EFF**	**Turnout**	**ANC**	**DA**	**EFF**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Coverage	0.009	-0.036	-0.007	0.026**	0.006	-0.038	-0.011	0.029**
	(0.011)	(0.035)	(0.014)	(0.013)	(0.011)	(0.038)	(0.013)	(0.014)
Post	0.057	0.233*	0.164**	-0.024	0.059	0.232*	0.187**	-0.024
	(0.050)	(0.128)	(0.070)	(0.042)	(0.052)	(0.138)	(0.074)	(0.044)
log(Exp) • Post	-0.001	0.009*	-0.000	-0.000	-0.001	0.008	-0.002	-0.000
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)
Coverage • log(Exp)	-0.004*	0.010*	0.000	-0.002	-0.003	0.010	0.001	-0.003
	(0.002)	(0.006)	(0.003)	(0.002)	(0.002)	(0.006)	(0.003)	(0.003)
Coverage • Post	-0.003	0.012	0.017	-0.028**	0.001	0.010	0.017	-0.031**
	(0.013)	(0.044)	(0.016)	(0.014)	(0.013)	(0.047)	(0.016)	(0.015)
Coverage • log(Exp) • Post	0.003	-0.012**	0.001	0.003	0.002	-0.011*	0.001	0.004
	(0.002)	(0.006)	(0.003)	(0.003)	(0.002)	(0.006)	(0.002)	(0.003)
log(Total Exp) • Post	0.001	-0.041***	0.001	-0.002	0.001	-0.040***	0.002	-0.004
	(0.005)	(0.014)	(0.005)	(0.004)	(0.005)	(0.015)	(0.004)	(0.004)
log(Luminosity)	0.001	-0.020	0.002	0.007	0.001	-0.022	0.001	0.007
	(0.007)	(0.020)	(0.006)	(0.006)	(0.007)	(0.021)	(0.006)	(0.006)
log(Pop. Density)	0.037**	0.063**	-0.029*	0.027**	0.040***	0.065**	-0.034**	0.028**
	(0.015)	(0.025)	(0.015)	(0.013)	(0.015)	(0.026)	(0.015)	(0.014)
Controls • Post	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
VD Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of observations	17,472	17,424	17,424	17,391	16,414	16,381	16,381	16,349
Adjusted R2	0.670	0.795	0.944	0.331	0.666	0.761	0.934	0.326

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses are clustered at the Local Municipality level (234 clusters). Up to the 5th order polynomial terms for (the log of) Luminosity and Pop. Density interacted with Post are included among the controls. VD stands for Voting District or union of them. Exp is the sum of wasteful, irregular and unauthorized audited expenditure per capita in 2014. Total Exp is the overall public expenditure per capita of the municipality in 2014.*

The figures support the hypothesized channel. Column (2) reports the results for the

ANC vote share. The coefficient on the triple-interaction term (Coverage • log(Exp) • Post) is negative and statistically significant at the 5% significance level. Its magnitude suggests that, among localities with high level of corruption-related expenditure, a unitary increase in coverage causes a reduction in the votes for the ANC party of almost 1.2 pp, which is almost 2% of the mean of the dependent variable at the baseline. If we exclusively focus on municipalities where the ANC was ruling (which are anyway the majority) the magnitude of the coefficient does not significantly vary. At the same time, other outcomes seem not to be affected. Quite surprisingly, Internet coverage does not affect the DA vote share.

Finally, in order to provide additional evidence on the mechanisms behind the observed relations, I conduct an analysis of the heterogeneity using different sub-samples of the original sample. In particular, I consider the following dimensions: income (luminosity), population density, terrain roughness, distance from main cities, share of traditional media ownership, average years of schooling and share of young people. The first two variables are a proxy for general well-being, while the second pair measure geographic isolation. Finally, the last three variables capture potential heterogeneous effects across places with different level of news circulation and regular Internet users. I also consider some combinations of these dimensions.

For each of these variables, I use their median at the baseline in order to create two categories. Table 7 provides the results for this analysis.

Table 7: Heterogeneity analysis in the baseline specification

	Turnout				ANC				DA			
	Coverage • Post	S.e.	N. of obs.	Adj. R2	Coverage • Post	S.e.	N. of obs.	Adj. R2	Coverage • Post	S.e.	N. of obs.	Adj. R2
Poor	-0.008	(0.025)	8,718	0.641	-0.074	(0.064)	8,709	0.709	0.032	(0.027)	8,709	0.874
Rich	0.026**	(0.010)	8,754	0.701	-0.077**	(0.031)	8,715	0.853	0.013	(0.009)	8,715	0.970
Low density	-0.002	(0.017)	8,736	0.645	-0.069*	(0.040)	8,723	0.744	0.035*	(0.021)	8,723	0.898
High density	0.024**	(0.010)	8,736	0.699	-0.068**	(0.033)	8,701	0.832	0.013	(0.009)	8,701	0.971
Rough	0.018	(0.016)	8,736	0.628	-0.101**	(0.047)	8,718	0.742	0.028	(0.018)	8,718	0.907
Not rough	0.010	(0.012)	8,736	0.707	-0.035	(0.031)	8,706	0.836	0.011	(0.011)	8,706	0.964
Far	-0.001	(0.017)	8,736	0.645	-0.022	(0.051)	8,731	0.727	0.017	(0.017)	8,731	0.918
Not far	0.026**	(0.012)	8,736	0.685	-0.083**	(0.035)	8,693	0.831	0.029**	(0.014)	8,693	0.961
High youth share	0.018	(0.013)	8,162	0.745	-0.071*	(0.041)	8,132	0.863	0.011	(0.013)	8,132	0.962
Low youth share	0.010	(0.018)	9,310	0.596	-0.019	(0.044)	9,292	0.692	0.012	(0.014)	9,292	0.878
High educated	0.017	(0.016)	7,172	0.723	-0.078*	(0.041)	7,134	0.892	0.027*	(0.015)	7,134	0.962
Low educated	0.012	(0.015)	10,300	0.625	-0.025	(0.039)	10,290	0.669	0.012	(0.014)	10,290	0.825
Rich and High educated	0.024	(0.019)	5,512	0.707	-0.092*	(0.052)	5,482	0.908	0.019	(0.014)	5,482	0.975
Rich and Low educated	0.030**	(0.014)	3,242	0.684	-0.036	(0.040)	3,233	0.643	-0.002	(0.012)	3,233	0.814
Poor and High educated	-0.000	(0.038)	1,660	0.736	-0.098	(0.071)	1,652	0.812	0.081	(0.052)	1,652	0.882
Poor and Low educated	-0.012	(0.031)	7,058	0.600	-0.048	(0.086)	7,057	0.680	0.017	(0.027)	7,057	0.830
Radio	0.024	(0.016)	7,996	0.657	-0.075*	(0.039)	7,956	0.872	0.003	(0.014)	7,956	0.961
No Radio	0.001	(0.012)	9,476	0.678	-0.039	(0.040)	9,468	0.656	0.033***	(0.013)	9,468	0.882
TV	0.010	(0.019)	7,462	0.689	-0.064	(0.040)	7,424	0.885	0.012	(0.015)	7,424	0.961
No TV	0.016	(0.012)	10,010	0.657	-0.043	(0.042)	10,000	0.673	0.018	(0.013)	10,000	0.837
Cellphone	0.020	(0.015)	7,792	0.723	-0.083**	(0.038)	7,760	0.868	0.013	(0.013)	7,760	0.967
No Cellphone	0.007	(0.017)	9,680	0.624	-0.021	(0.041)	9,664	0.677	0.010	(0.015)	9,664	0.890

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Each row is a separate regression. Standard errors in parentheses are clustered at the Local Municipality level. All regressions include VD Fixed Effects and all controls reported in Table 3.*

For each dependent variable, each row represents a specific regression, of which only the coefficient of interest it reported. The table shows that Internet coverage affected turnout in localities with higher income, more densely populated and closer to the main cities. The magnitude of the coefficient is fairly stable across these subsamples. Following section 3, one would also expect to find some impacts in places where the shares of young[11] and highly educated individuals is higher. However, there is no evidence on this for turnout. Being the correlation between average education and income usually very high, I decide to consider combinations of these two variables to explore the relative importance of schooling in places with high and low income level. Quite surprisingly, rich but less educated localities experienced an increase in turnout due to Internet penetration.

If we consider the ANC share of votes, the heterogeneity analysis becomes more interesting. In this case, the ruling party seems to have lost many votes in places with high income, closer to large cities but more geographically isolated in terms of terrain characteristics *(Rough)*, as a consequence of Internet arrival. In addition, localities with many youth or where the population is on average more educated seem to drive the results. Particularly, among rich municipalities, those with higher average years of schooling experienced a large and significant shift of votes. Most importantly, there seems to be no role for education among poor localities. Despite the difference in magnitudes, coefficient for poor and highly educated places is insignificant. Finally, in localities where radio and cellphone ownership rates are higher the ANC party significantly lost more votes after the Internet expanded.

The DA exhibits similar patterns. However, in this case the party gained votes in places where radio ownership was lower, suggesting that the Internet played a role where individuals were used to be less exposed to the circulation of information through traditional media.

Overall, these results seem to suggest that income plays a fundamental role: in order to access the technology (subscribing for the Internet bundle or purchasing a mobile phone) individuals must be sufficiently rich. To better examine the importance of education, media and geographic isolation for the access to web-based information on corruption scandals I run the heterogeneity analysis on the triple difference specification. In this case I only focus on rich municipalities. Results are displayed in Table 8 below, where only the coefficient on the triple interaction term is reported.

The numbers point out that among rich localities with higher corruption-related expenditure, those with lower average education, more geographically isolated and where information circulation through traditional media was presumably lower experienced a significant decline in ANC popularity as a consequence of Internet penetration. The estimated coefficients are stable across subsamples and seem to point out the fundamental information role that the web played for those voters residing in geographically isolated localities, where the use of traditional media was generally low. Therefore, according to this scenario, Internet coverage seem to be relevant for the provision of new web-based information in places where this source could not substitute out the use of other media. At the same time, in places where voters were exposed to a larger amount of information through traditional media, the arrival of the Internet did not influenced the incumbent's vote share.

[11] I consider the share of individuals aged 18 - 35.

Table 8: Heterogeneity analysis in the DDD specification

	Turnout				ANC				DA			
	Coverage • log(Exp) • Post	S.e.	N. of obs.	Adj. R2	Coverage • log(Exp) • Post	S.e.	N. of obs.	Adj. R2	Coverage • log(Exp) • Post	S.e.	N. of obs.	Adj. R2
Rich and High educated	0.001	(0.004)	2,422	0.647	0.004	(0.014)	5,482	0.910	-0.001	(0.005)	2,413	0.958
Rich and Low educated	-0.001	(0.004)	6,332	0.718	-0.020**	(0.010)	3,233	0.649	-0.002	(0.004)	6,302	0.973
Rich and Rough	0.001	(0.004)	2,422	0.647	-0.022*	(0.012)	2,413	0.793	-0.001	(0.005)	2,413	0.958
Rich and Not rough	-0.001	(0.004)	6,332	0.718	-0.003	(0.010)	6,302	0.879	-0.002	(0.004)	6,302	0.973
Rich and Radio	-0.000	(0.005)	5,364	0.677	-0.001	(0.012)	5,332	0.908	-0.002	(0.005)	5,332	0.974
Rich and No Radio	0.000	(0.005)	3,390	0.719	-0.024**	(0.011)	3,383	0.678	-0.002	(0.002)	3,383	0.935
Rich and TV	-0.005	(0.006)	5,434	0.710	0.002	(0.014)	5,404	0.906	0.003	(0.004)	5,404	0.974
Rich and No TV	0.001	(0.005)	3,320	0.683	-0.021**	(0.010)	3,311	0.675	-0.003	(0.004)	3,311	0.882

*Notes: *** p<0.01, ** p<0.05, * p<0.1 Each row is a separate regression. Standard errors in parentheses are clustered at the Local Municipality level. All regressions include VD Fixed Effects and all controls reported in Table 6.*

7. Conclusions

This paper has provided preliminary evidence on the causal impact of mobile Internet coverage on political outcomes exploiting the last South African municipal elections. A high-resolution dataset and a Diff-in-Diff specification have been used to mitigate concerns on potential endogeneity. The main findings show that Internet coverage did not significantly affect turnout. However, it caused a reduction in the vote share of the ruling ANC party of almost 6.4 pp. The second major party, the DA, has gained from the Internet arrival. Its share rose by approximately 2.2 pp. Finally, less robust evidence suggests that the newly formed EFF group did not benefit from Internet penetration. A series of robustness checks have been conducted to foster the identification strategy and the precision of the estimates. Overall, the results of these tests are in favor of the underlying identification assumptions.

The second part of the paper develops a comprehensive analysis on the potential mechanisms that may drive the observed relationships. A triple difference estimator has been applied to understand the role of the Internet in providing information on the financial administration of the municipal money. This analysis point out that among localities with high level of corruption-related expenditure, a unitary increase in coverage causes a reduction in the votes for the ANC party of almost 1.2 pp. Moreover, a heterogeneity analysis across sub-samples of the original dataset has been conducted to examine which segments of the population are more likely to be afected by Internet coverage. Evidence from this investigation is mixed. Overall, it shows that the ruling party lost more votes in richer and geographically isolated localities, and where the shares of youth and educated people are larger. Finally, the heterogeneity analysis of the triple difference specification demonstrates that the Internet played a fundamental information role towards corruption in localities that are isolated in terms of both geographical characteristics and news circulation through traditional media.

Future work should try to address additional questions. Firstly, understanding how politicians react to the new technology can be useful to better examine which mechanisms are driving the observed relationship. Secondly, providing evidence on additional outcomes such as public policies and the occurrence of protests, riots and political manifestations might help in supporting the findings of this paper. Lastly, additional effort

is needed to prove the conditional exogeneity of Internet coverage. Exploiting potential technological constraints in the expansion of the network may provide a reliable way to tackle the problem. In particular, the ideal alternative would be to find a valid instrument for coverage. Research on this is in progress.

Appendix

Figure 8: A FAKE EFF POSTER

Figure 9: ISPs MARKET SHARES (BUSINESSTECH)

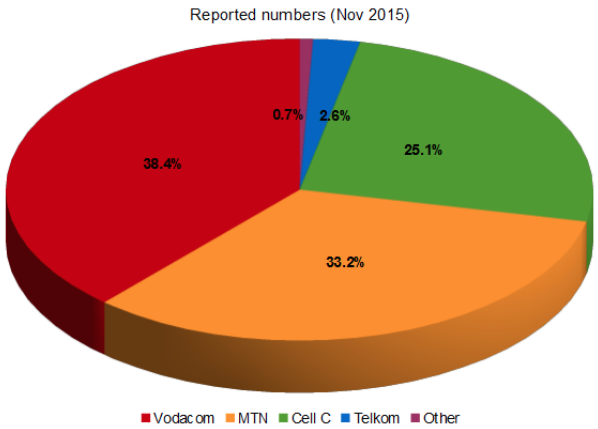
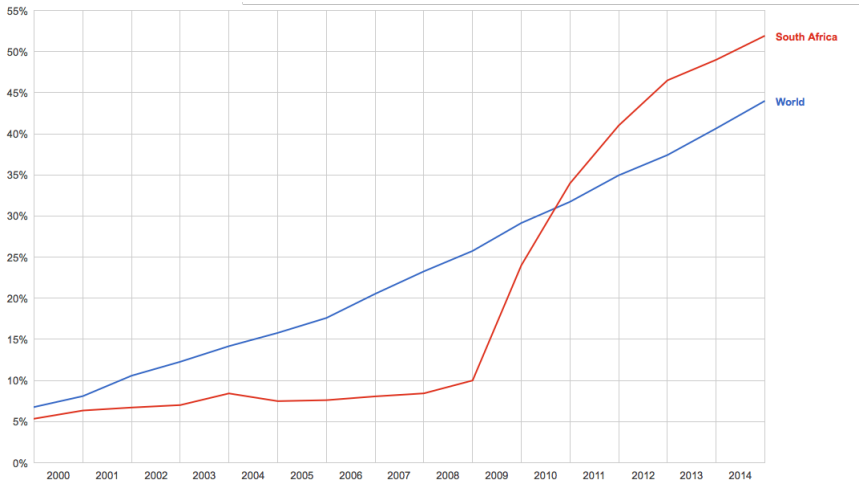
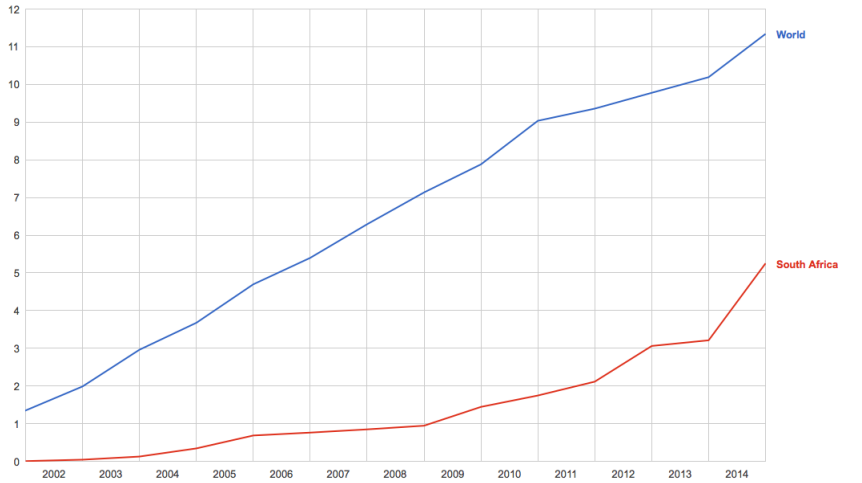


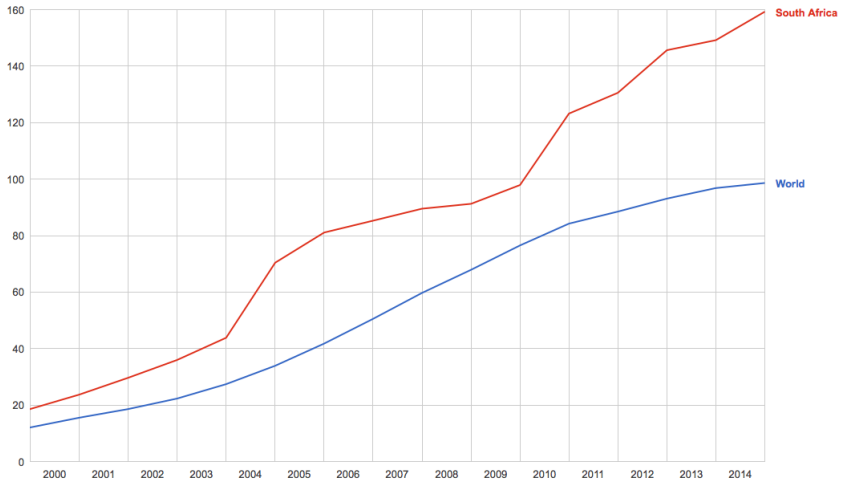
Figure 10: Trends over the last 15 years (World Bank)



(a) Internet users as percentage of population



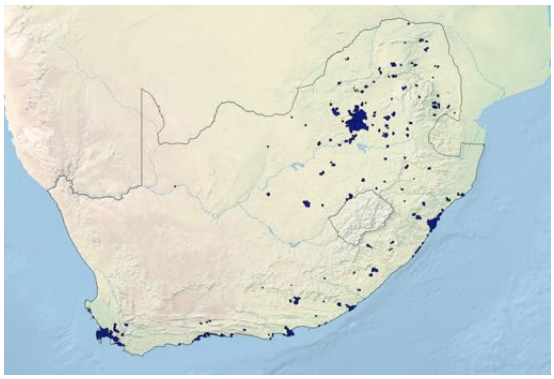
(b) Fixed broadband internet subscribers



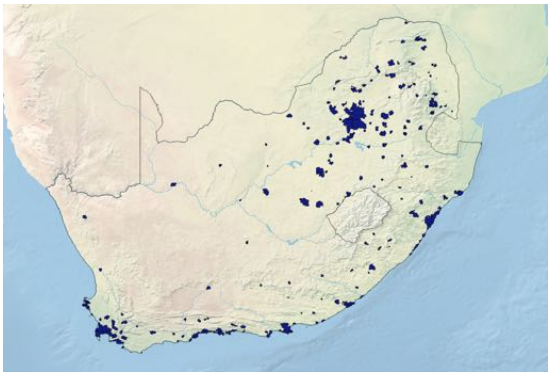
(c) Mobile cellular subscriptions per 100 people

Figure 11: Mobile internet coverage 2008-2014 (Collins Bartholomew)

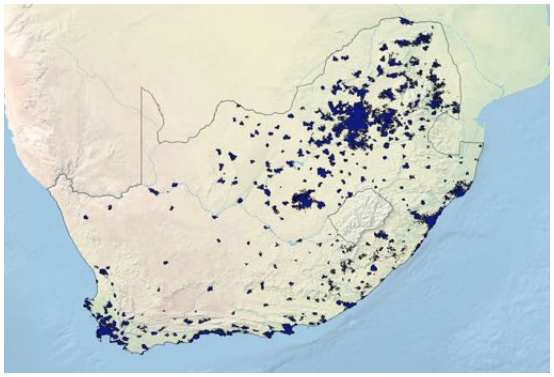
2008



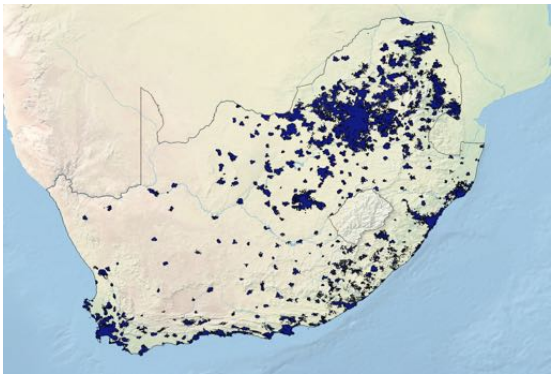
2009



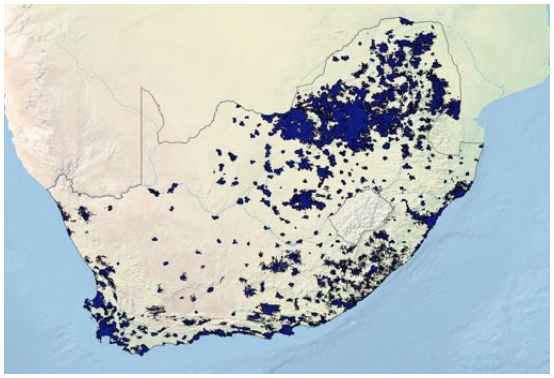
2011



2012



2013



2014

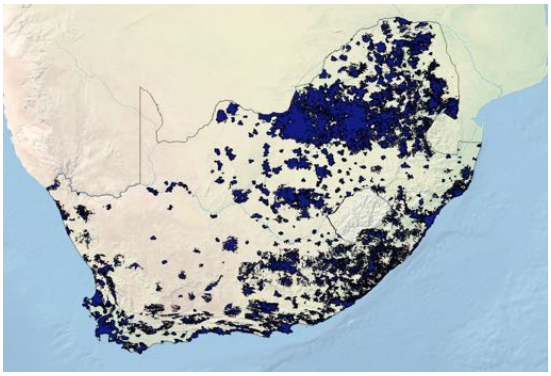
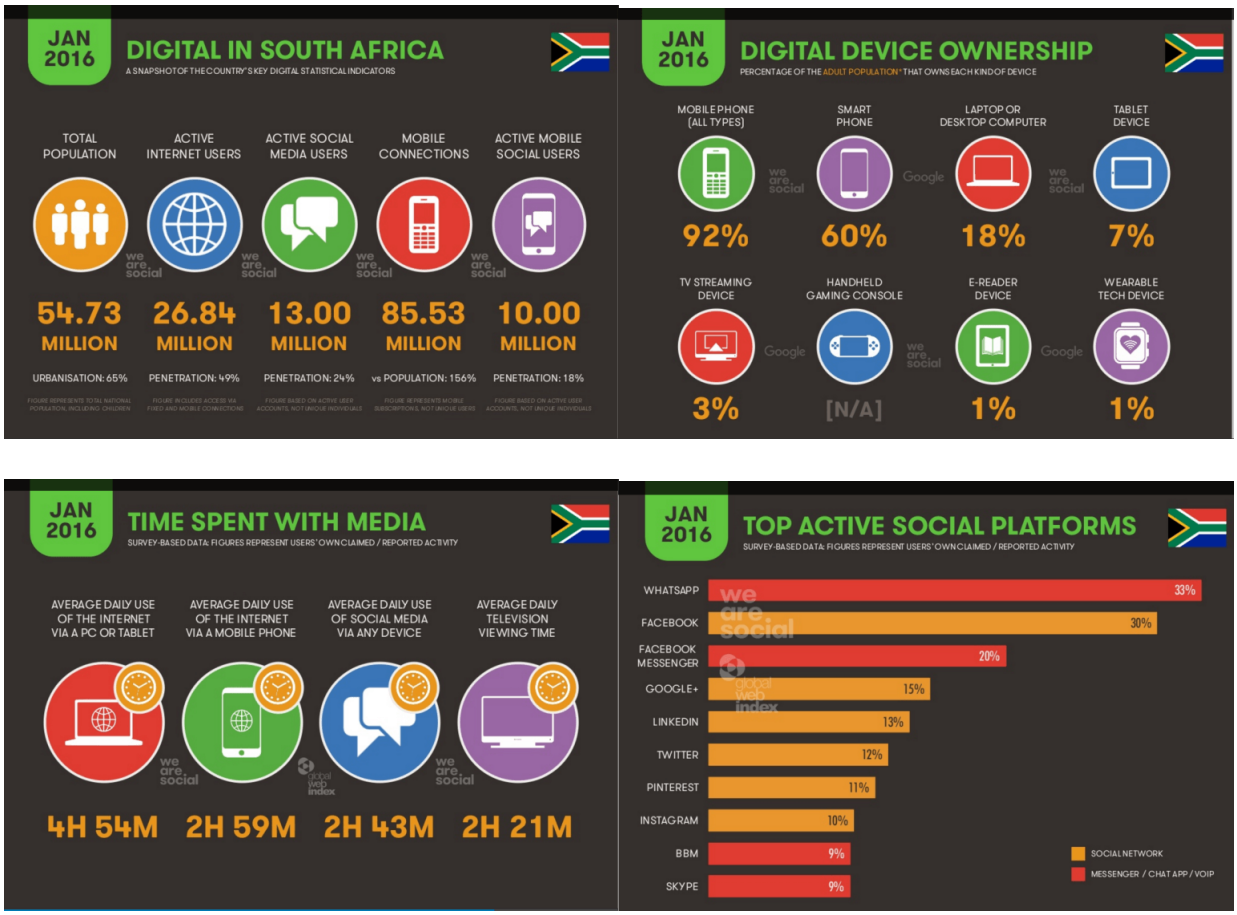


Figure 12: Media users in South Africa (We are social)



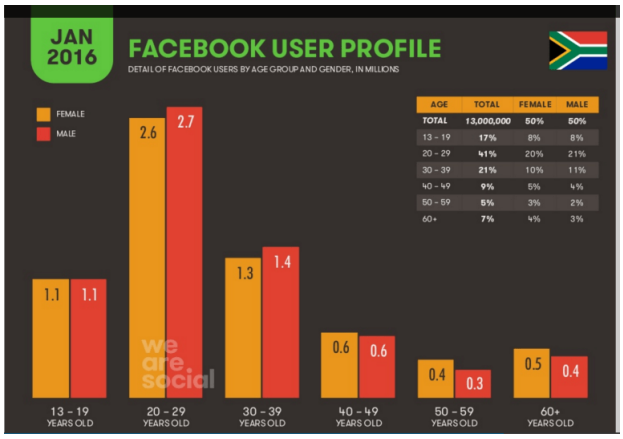
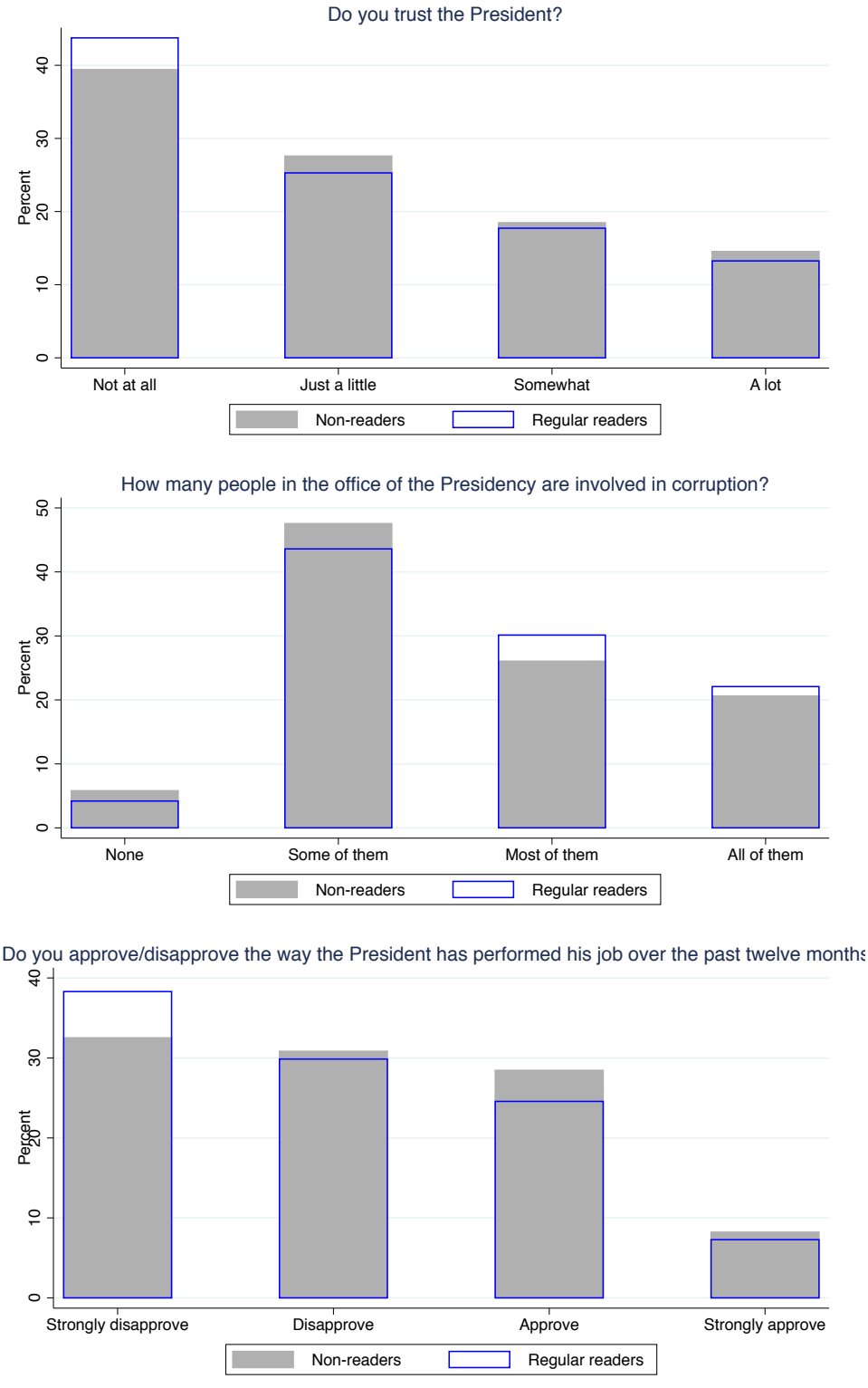
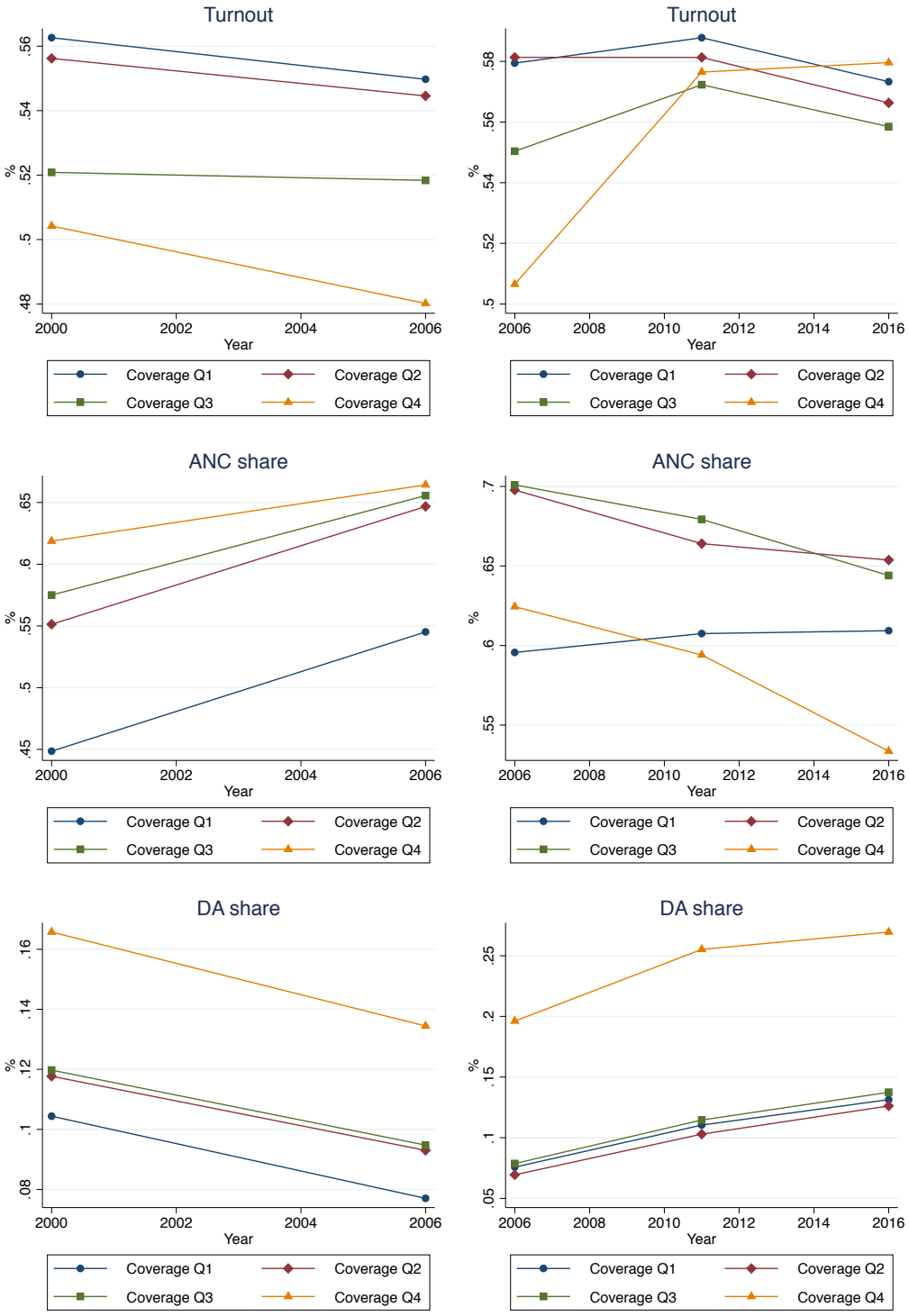


Figure 13: Opinions towards the presidency by newspapers readership



Notes: Author's own calculations based on Afrobarometer.

Figure 14: Trends in political outcomes by quartile of Internet coverage



References

[1] Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of political Economy, 113(1), 151-184.

[2] Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

[3] Buys, P., Dasgupta, S., Thomas, T. S., & Wheeler, D. (2009). Determinants of a digital divide in Sub-Saharan Africa: A spatial econometric analysis of cell phone coverage. World Development, 37(9), 1494-1505.

[4] Campante, F. R., Durante, R., & Sobbrio, F. (2013). Politics 2.0: The multifaceted effect of broadband internet on political participation (No. w19029). National Bureau of Economic Research.

[5] DellaVigna, S., & Kaplan, E. (2007). The Fox News effect: Media bias and voting. The Quarterly Journal of Economics, 122(3), 1187-1234.

[6] Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and political persuasion: Evidence from Russia. The American Economic Review, 101(7), 3253-3285.

[7] Falck, O., Gold, R., & Heblich, S. (2014). E-lections: Voting Behavior and the Internet. The American Economic Review, 104(7), 2238-2265.

[8] Gavazza, A., Nardotto, M., & Valletti, T. M. (2015). Internet and politics: Evidence from UK local elections and local government policies.

[9] Grace, J., Kenny, C., Zhen-Wei Qiang, C. (2004). Information and communication technologies and broad-based development: A partial review of evidence. World Bank Working Paper No. 12, World Bank, Washington, DC.

[10] La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap operas and fertility: Evidence from Brazil. American Economic Journal: Applied Economics, 4(4), 1-31.

[11] Manacorda, M., & Tesei, A. (2016). Liberation technology: mobile phones and political mobilization in Africa.

[12] Miner, L. (2015). The unintended consequences of Internet diffusion: Evidence from Malaysia. Journal of Public Economics, 132, 66-78.

[13] Odendaal, N. (2014). Space matters: the relational power of mobile technologies. urbe. Revista Brasileira de Gestão Urbana, 6(1), 31-45.

[14] Odendaal, N. (2011, June). The spaces between: ICT and marginalization in the South African city. In Proceedings of the 5th International Conference on Communities and Technologies (pp. 150-158). ACM.

[15] Petrova, M., Sen, A., & Yildirim, P. (2016). Social Media and Political Donations: New Technology and Incumbency Advantage in the United States.

[16] World Bank (2006). Information and communications for development 2006: Global trends and policies. Washington, DC: World Bank.

[17] World Bank (2008). Africa infrastructure country diagnostic: Information and communications technology in sub-Saharan Africa – A sector review. Mimeo.