Navigating the Platforms: the Effect of Party's Ideology on Mobile Internet's Impact on Elections in Europe

Etienne Madinier

Working paper: November 18, 2023

Abstract

Mobile Internet has changed election outcomes in Western Europe. We investigate the role that party ideology plays for the impact of broadband mobile technology on the outcomes of first-round elections in lower chambers across Europe. We combine election results with two types of data: (i) at the constituency level, we utilize the share of the population that has access to mobile Internet, and (ii) at the party level, we describe the political platform using either results from political scientists' surveys or the content of the parties' manifestos. Our analysis yields four main findings. First, candidates arguing to represent the middle-class interests garnered fewer votes in areas with widespread broadband mobile Internet. In contrast, those who fueled the polarization within the same middle-class along the lines of wealth and education received more votes. This decline in middle-class identification left room for new non-economic identification processes, potentially benefiting communitarian parties. In line with this, our second result exposes a positive effect of broadband mobile Internet on eurosceptic parties. However, holding euroscepticism constant, there is no additional effect of mobile broadband on parties that advocate traditionalism or hostility toward migrants. Third, candidates championing for the rural population also increased their vote share in places with broadband mobile Internet, as the latter was simultaneously marginalized by legacy media and largely instrumentalized in the political debate. Finally, calling for state funding of mass culture in their manifesto decreased the vote share of the candidates in places with broadband mobile Internet as it offers a convenient access to culture and entertainment.

1 Introduction

The political implications of broadband Internet have become increasingly salient and discussed in public debate, especially since its expansion in the mid-2000s. But the public perception of these consequences was not unequivocal, combining, for example, some enthusiasm for potential grassroots democratic movements and defiance from malicious manipulation and authoritarian abuses. For instance, in 2006, Time magazine named Persons of the Year people who anonymously contributed user-generated content to websites under the rewarding etiquette 'You.' But in 2017

already, The Economist was disenchanted, headlining: 'Social media's threat to democracy.' In the meantime, a similar disenchantment occurred regarding the consequences outside of Western countries. Social media were initially acclaimed as a democratizing force in light of their purported role in the Arab Spring; however, the praise eventually waned. A few years later, concerns regarding the role of Facebook in relation to ethnic cleansing in Myanmar were raised.[1]

The US presidential elections provide another example of this ambiguity. In 2008, the Western liberal press unanimously celebrated: newly elected President Obama had won, in part, thanks to his online campaign. Eight years later, those same newspapers were equally unanimous, yet somehow less thrilled, if not accusatory: Trump had also been elected, in part due to what had transpired online during the campaign.[2] In less than 10 years, the internet had propelled two candidates with significantly different political platforms.[3] At best, it appeared that the internet rewarded successful political entrepreneurs who had simply mastered this tool better than their opponents, irrespective of their ideologies. At worst, the initial promotion of universalist values was now yielding unintended consequences.

Scholars have also shown interest in this subject. A recent multi-country analysis conducted in Europe by Manacorda et al. (2023) reveals a positive relationship between the availability of 3G and support for a specific ideological category of candidates. The study employs various alternative definitions of what the researchers refer to as "communitarian" parties, which can be based on

[1] See for instance:

"Social Media Made the Arab Spring, But Couldn't Save It" (Wired, January 26 2016, https://www.wired.com/2016/01/social-media-made-the-arab-spring-but-couldnt-save-it/, consulted 08/01/2023).

"Comment Internet a fait les "printemps arabes"" (Le Monde, October 14 2017, https://www.lemonde.fr/idees/article/2017/10/14/comment-internet-a-fait-les-printemps-arabes_5201063_3232.html, consulted 08/01/2023).

"A Genocide Incited on Facebook, With Posts From Myanmar's Military" (The New York Times, October 15 2018, https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html, consulted 08/01/2023).

"How Facebook's Rise Fueled Chaos and Confusion in Myanmar" (Wired, July 6 2018, https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/, consulted 08/01/2023).

[2] See for instance:

"The first election the internet won" (The Guardian, November 05 2008, https://www.theguardian.com/world/deadlineusa/2008/nov/05/uselections20082, consulted 07/16/2023).

"How Obama's Internet Campaign Changed Politics" (The New York Times, November 07 2008, https://archive.nytimes.com/bits.blogs.nytimes.com/2008/11/07/how-obamas-internet-campaign-changed-politics/, consulted 07/16/2023).

"Propelled by Internet, Barack Obama Wins Presidency" (Wired, November 11 2008, https://www.wired.com/2008/11/propelled-by-in/, consulted 07/16/2023).

"Here's How Facebook Actually Won Trump the Presidency" (Wired, November 15 2016, https://www.wired.com/2016/11/facebook-won-trump-election-not-just-fake-news/, consulted 07/16/2023).

"How the Internet Is Threatening Our Freedom" (Politico, May 18 2016)https://www.politico.com/magazine/story/2016/05/2016-election-internet-campaign-facts-digital-new-media-213899/, consulted 07/16/2023).

[3] For instance, Barack Obama would later support one of the largest healthcare system expansions of coverage since World War II (the Affordable Care Act, also known as Obamacare), while Donald Trump repeatedly promised to "repeal and replace" it.

factors such as their endorsement of restrictive migration policies, opposition to EU integration, or their left-right political orientation. The research demonstrates that the most extreme candidates within these dimensions garnered higher vote shares when voters had access to broadband mobile internet. Because the political indicators used in the study are correlated, this robustly indicates that a specific segment of the political spectrum (i.e., "communitarian" platforms) benefited from the expansion of broadband mobile internet. However, due to this correlation, the study falls short of providing insights into whether some dimensions were more influential than others in shaping this increase in vote share.

This paper adopts a complementary approach to Manacorda et al. (2023), with the aim of determining which types of politicians benefited from the proliferation of mobile internet in elections within Western democracies. By framing the research question as a treatment heterogeneity exercise, we estimate the correlation between the impact of broadband mobile internet and various political dimensions that characterize candidates' political platforms. This approach enables us to go beyond multi-faceted labels like "communitarian" or "populist" and allows for a comparison of the significance and influence of each individual dimension.

In our study, we employ a generalized random forests algorithm (Athey et al., 2019) to assess the impact of 3G and advanced technologies ($3G^{+}$) at the constituency level on candidates' vote shares during the first round of parliamentary elections in Europe spanning from 1990 to 2018. To gauge political positions, we combine data from surveys conducted by political scientists with the text encoding of party manifestos. Our analysis yields four main findings. First, candidates advocating for middle-class interests received fewer votes in regions with widespread broadband mobile internet. In contrast, those who exacerbated polarization within the middle class along the lines of wealth and education garnered more votes. This decline in middle-class identification created space for new non-economic identification processes, potentially benefiting communitarian parties. In alignment with the first finding, our second result highlights a positive effect of broadband mobile internet on eurosceptic parties. However, when holding euroscepticism constant, there is no additional impact of mobile broadband on parties advocating traditionalism or displaying hostility toward migrants. Third, candidates championing the interests of rural populations also experienced an increase in their vote share in areas with broadband mobile internet. This was due to the simultaneous marginalization of rural issues by legacy media and their extensive instrumentalization in political discourse. Finally, calling for state funding of mass culture in their manifesto decreased the vote share of the candidates in places with broadband mobile Internet, as it offers convenient access to culture and entertainment.

This paper contributes to several strands of the political economy literature: (a) the political effects of the internet; (b) the structure of ideological cleavages in political competition; and (c) multiparty econometrics.

Zhuravskaya et al. (2020) provides a comprehensive review of the growing body of literature on

the political effects of the internet and social media. In this context, two papers are particularly relevant to our current work: Guriev et al. (2021) and Manacorda et al. (2023), both of which conduct multi-country analyses to explore the effects of mobile internet on election outcomes. Guriev et al. (2021) demonstrate that the advent of 3G technology resulted in lower vote shares for incumbent parties and higher vote shares for anti-establishment populist opposition parties in Europe. Meanwhile, Manacorda et al. (2023) reveals that the same technology led to increased voter support for communitarian parties. In our study, we aim to move beyond these broad categorizations and delve into the specific dimensions that influence the impact of mobile broadband internet.

The classification of political positions and understanding their relationships has been a long-standing challenge in political science. Early work by Laver and Hunt (1992) provided empirical evidence that politics is indeed multidimensional, based on a sample of twenty countries. Additionally, Poole and Rosenthal (2011) emphasized the significance and empirical power of the left-right axis in explaining the voting behavior of representatives. However, it's worth noting that other academic studies have highlighted how shocks or external factors can lead to the emergence of new dimensions in politics. For instance, Bonomi et al. (2021) demonstrated how economic shocks can impact voters' social identification, potentially reducing their demand for redistribution while intensifying conflicts along other social dimensions. Our objective in this study is to document how the diffusion of the internet has interplayed with political conflict.

This paper contributes to this literature in several ways. First, we document what we call a decline in the middle-class consensus, meaning that candidates who use the middle class as a positive figure and target of their discourse/platform are actually negatively affected by the availability of mobile broadband Internet. The technology has simultaneously increased polarization on economic issues that matter most to the middle class: education and business-oriented policies. Second, we confirm previous results showing that anti-EU parties increased their vote shares due to $3G^+$, but we do not find a positive and significant relationship regarding hostility toward migration or support for traditional values. Third, we provide evidence that parties advocating for support of the rural population benefited from technology diffusion. Finally, we show that the availability of broadband mobile Internet diminished people's support for state funding of leisure, culture, and mass media.

Another contribution of this paper is methodological. To our knowledge, we are the first to apply the Generalized Random Forest algorithm (Athey et al., 2019) to study multiparty elections. The estimation of the effect of broadband mobile internet on electoral outcomes is framed as a heterogeneity problem where the treatment effect is described as a function of the candidates' characteristics. We argue that this approach conveys several desirable properties for such exercises and believe it is a promising new conceptualization and empirical tool for political economy. In particular, we explore how this procedure relates to the other multiparty econometric methods of

Katz and King (1999) and Tomz et al. (2002).

We utilize two different data sources to describe the political space: the Chapel Hill Expert Survey (Jolly et al., 2022), which is based on a survey of political science experts, and the Comparative Manifesto Project (Lehmann et al., 2022). Besides the robustness exercise facilitated by this approach, our results foster interest in the content of parties' manifestos as a means to describe political ideologies and communication. Using textual data to measure ideology has been recently employed for instance by Gentzkow et al. (2019) to analyze trends in the partisanship of congressional speeches in the US from 1873 to 2016, highlighting a sharp increase in the early 1990s. Their measure is based on the ease with which an observer could infer a congressperson's party from a single utterance. An other example by Enke (2020) focuses on another single-axis measure (universalist vs. communal moral values) derived from campaign documents.

Our approach differs in several key aspects. Firstly, we aim to establish a relevant classification of politicians in response to a specific shock ($3G^+$) rather than a general classification of similar ideologies. In otherwords, we seek to group candidates whose ideologies and treatment effect (of mobile Internet) are similar. Secondly, in both cases mentioned, the researchers begin with an ideological partition, such as Democrats vs. Republicans or universalists vs. communalists, and then utilize high-dimensional text data to construct their single-axis measure. In our paper, we start with an already dimensionality-reduced dataset but maintain the ideological space as multidimensional throughout the study. We also conduct an additional variable selection procedure in our empirical strategy to prevent over-fitting in our final regressions.

The rest of this paper is organized as follows. Section 2 provides some background information on our treatment (i.e., $3G^+$) and the elections we are focusing on. Section 3 introduces and summarizes our data sources. Section 4 presents our empirical strategy. Section 5 details our results, and Section 6 concludes.

2 Background

In this section, we begin by offering detailed background information on our treatment, i.e., $3G^+$. We provide a technical specification of this technology, trace its development across Europe, and highlight the key changes it introduced. Next, we present the institutional setting of our observational study and examine the broad political trends that occurred during our time period. Finally, we give some context on the current academic debates related to the political dimensions exposed by our results.

2.1 The spread of broadband mobile Internet

The main focus of this paper is wireless mobile technologies, which aim at transferring data faster over telecommunication networks, improving the quality of service for users, and also offering new

possibilities. The 3G technology was initially developed between the 1980s and early 2000s, the years during which the commercial launch of the technology will take place. Starting in 2010, the technology will be gradually replaced by 4G, a new technology that increases bandwidth speed again. 3G was the first cellular technology to allow one to browse the internet on one's cellphone, download pictures, videos, and use IP telephony. 4G technology, with its higher feasible data traffic, introduced, for instance, the possibility to watch TV on a cellphone.

The development of the infrastructure and market was mostly driven by the private sector but largely accompanied by the governments (OECD, 2021).[4] The development of the technology was also dependent on the willingness of other actors, such as municipalities, housing corporations, and citizen's cooperatives (Lemstra, 2017). Access to the technology had direct local economic impacts, such as enhancing access to foreign goods (Malgouyres et al., 2021).

Overall, the features provided by these technologies were crucial to the development of Web 2.0 with the explosion of user-generated content and the emergence of social media, allowing the Internet to become a major source of entertainment and information. It did not only create a new media, but also led legacy media to have a presence online and use online content as a source of information for their own content (Hatte et al., 2021). The cost structure of information production and consumption was dramatically affected by the existence of the Internet (Cagé et al., 2020). There was also disintermediation between citizens and politicians, deeply affecting political communication.

2.2 Parliamentary Elections in Europe (1990-2020)

The parliamentary systems in Europe exhibit a rich diversity of regimes, reflecting historical, cultural, and political factors specific to each country. Most of the countries in this study are parliamentary regimes, but some, such as France or Poland, are rather referred to as semi-presidential.[5] Our paper focuses on the first round of lower chamber elections. These elections are often local and feature many candidates running, especially when their voting system includes some proportional representation flavor.[6]

[4]There are two main reasons for this: (1) there was a need for large investments to make the technology available over each country's territory, (2) the telecommunications sector has always been regarded as strategic and required some state intervention and regulation. The development of the market was sustained both on the supply and demand sides. For instance, governments have extended investment possibilities (either by offering preferential loans or relaxing barriers to foreign investment). Demand could be stimulated by direct intervention on prices, but also by the increasing offer of services (including governmental ones such as administrative procedures) available through this technology.

[5]Additionally, there is a majority of bicameral systems with a lower and upper house, but some countries, such as Portugal, have a unicameral system.

[6]The diverse parliamentary regimes in Europe are accompanied by various voting systems that determine how citizens' preferences are translated into representative outcomes. These systems include the first-past-the-post system (e.g., UK), proportional representation (e.g., Spain) with several variations, majority runoff (e.g., France), and a mixture of such systems (e.g., Germany).

Elections in Europe between 1990 and 2020 have undergone significant political transformations. The 1980s marked a turning point in Europe as the continent witnessed a notable rise in nationalism. The process of economic globalization, which intensified during the period, led to economic dislocation, job insecurity, and growing inequalities. Nationalism found fertile ground among those who felt marginalized by globalization, as it offered a sense of belonging and protection from perceived threats to economic stability and cultural identity. The deepening of supranational institutions, and particularly the process of European integration, triggered a backlash against perceived encroachments on national sovereignty. The influx of immigrants from within the European Union, former colonies, as well as refugees from conflict zones created a new context for nationalist discourse. The latter is now concerned about cultural assimilation and perceived threats to national identity, fueling fears of social disintegration.

In the late period, the effects of the 2008 crisis and the subsequent austerity measures reshaped political cleavages and led to increased support for non-mainstream parties, particularly populist ones (Algan et al., 2017; Guriev and Papaioannou, 2022). However, populism is a complex notion to comprehend (Badie et al., 2018), as it may correspond more to a political style than a distinct ideology.[7] Two major aspects of contemporary populism are: (i) anti-elite rhetoric, and (ii) similar to nationalism, a strong emphasis on national sovereignty, which materializes as euroscepticism in the EU context. Populists also heavily criticize intermediary institutions, including traditional media, which they argue capture the will of the people. Finally, their discourse is characterized by a significant emotional charge.

It is right-wing populism that has predominantly experienced electoral success in Europe, but left-wing populism has also played a significant role in countries such as France, Spain, and Greece. Moreover, some candidates do not easily fit into this binary left-right classification. Both left and right-wing populism include elements of anti-capitalism. However, the latter often focuses on criticizing financial globalization, which is seen as a form of stateless capitalism, but frequently adopts an anti-taxation discourse. A substantial portion of right-wing populism not only supports national sovereignty but also embraces ethno-nationalism. On the other hand, left-wing populists, without relying on ethnic conceptions, attempt to mobilize the labor class as the people against the elite as a way to reignite class conflict.

[7]The term initially appeared around 1870 and referred to a Russian revolutionary movement (*Narodnichestvo*), advocating agrarian socialism with an idealization of the peasant community as the essence of the "people," opposed to urban centers and the tsarist state. A distinguishing feature of modern populism, absent from *Narodnichestvo*, is its embodiment by a charismatic leader. An early occurrence of such a political movement appeared in the late 1880s in France, centered around General Boulanger, who criticized the inefficiency of political parties and claimed to transcend the left-right cleavage. Another example was the rise of the Tea Party in the US, initiated by popular protests, especially by western farmers ("grangers"). Compared to the other two movements, the Tea Party did not directly attack democratic values, and their conception of the American people also involved an ethnic dimension. The second wave of populism took place between the two world wars, mostly in Europe and Latin America. A third wave occurred during the decolonization process and the Cold War, mostly in Latin America, Africa, and Asia.

3 Data sources and main variables

In this section, we describe our main variables of interest and their sources. The data combine 3 main types of information: European parties' political platforms, their vote share in lower chamber elections at the constituency level, and mobile broadband internet access also measured at the constituency level. Summary statistics of all variables are provided in Tables A1 to A3.

3.1 Elections

The main outcome of this study is the vote share of a specific party (or set of parties) obtained during the first round of lower chamber elections in a given constituency. The data used for analysis are sourced from the CLEA election archive project (Kollman et al., 2019). For each candidate in each constituency, the dataset includes information about their party affiliation (if any), the number of registered voters, the number of valid votes, and the number of votes they received. The geographical boundaries of the constituencies are derived from the related GeoReferenced Electoral Districts Datasets (Kollman et al., 2019).

The data cover 94 elections taking place in 15 countries from 1990 to 2018. Figure 2 shows the timeline of these elections. The map of the 1530 constituencies is displayed in Figure 1. The sample includes 9042 vote casts.

3.2 Political platforms

We proxy for the candidates' ideology by their political affiliation, as provided in the CLEA database. Subsequently, we match our data with the Chapel Hill Expert Survey (CHES) (Jolly et al., 2022). This survey analyzes the opinions of political scientists on various topics, including party ideologies. From their responses, 0 to 10 scores are derived along several political dimensions for each party to capture their political platform. Our focus is on 8 dimensions: economy (left-right), hostility toward migrants, hostility toward the European Union (EU), support for law & order, support toward the rural population, anti-environment policies, traditionalism, and anti-elite rhetoric. We chose those variables because of their importance in the definition surrounding populist or communitarian platforms.[8] Summary statistics are shown in Table A1, while the correlations between each dimension are depicted in Figure 3. Throughout the paper, this measure of political platforms is referred to as survey-based ideology.

Alternatively, we measure the parties' political platform using the content of their manifestos, from the Comparative Manifesto Project (Lehmann et al., 2022). In this database, party manifestos are divided into quasi-sentences, which are subsequently manually coded based on predefined categories. All variables indicate the share of quasi-sentences in their respective categories as a fraction of the overall number of allocated codes per document. The variables fall into 7 categories:

[8] Traditionalism in the sense defined by Hooghe et al. (2002) along the so-called GALTAN axis.

external relations (e.g., anti-imperialism, favorable mentions of particular countries), freedom and democracy (e.g., human rights, constitutionalism), political system (e.g., corruption, centralization), economy (e.g., protectionism, economic growth paradigm), welfare and quality of life (e.g., environmental protection, welfare state expansion), fabric of society (e.g., traditional morality, multiculturalism), and social groups (e.g., favorable references to labor groups, the middle class). The list of the 55 variables can be found in Table A2. Throughout the paper, we refer to this measure as the manifesto-based ideology.

Using the two separate measures allows us to assess the robustness of our approach. We also think their differences represent some advantages. The manifesto data, for instance, can be more easily manipulated by the candidates compared to expert opinions. For instance, a party may have a strong opinion on the environment but may deliberately choose to communicate about other issues. We could therefore think of expert surveys as a raw measure of the full ideological platform, while manifesto data weighs each dimension by the emphasis chosen by the party. The CMP data also offer a much finer grained description of the parties' positions and seems more data driven.

We also match those candidates to the classification of populism used in Guriev et al. (2021). Summary statistics for each political block we consider can be found in Table A3.

Over our time period, the data shows a steady decrease in voter turnout. As seen in Figure A1, the initial average turnout level was about 80% in 1990, and it declined to less than 65% almost 30 years later. The trend of populist vote in Figure A2 is the opposite, with an increase from roughly 13% to 18%. However, this increase is not driven by right-wing candidates, for whom the trend is, if anything, decreasing. Both the top half and top decile of anti-elite parties initially experienced a decline in their vote share until the mid-2000s. After that, the trend reversed, bringing the top decile back to its initial 1990 level of support and even higher for the top half.

3.3 Broadband mobile internet

To measure access to broadband mobile Internet, we utilize digital maps of $3G^+$ access from 2006 to 2018, which are provided by Collins Bartholomew's Mobile Coverage Explorer.[9] The original map is a binary grid of 1 km $\times$ 1 km cells. To create a continuous broadband mobile access measure ranging from 0 to 1, we calculate the weighted average of the grid cells lying within each constituency. The weighting is based on the population in each grid cell. For our analysis, we consider a constituency as treated if the continuous indicator previously described is greater than 0.5. In other words, a constituency is regarded as treated if at least half of its population has access to the technology. The full distribution of the share of the constituency that is treated can be found in Figure A6.

[9]These maps are created by combining data submitted by operators to the GSM Association, which represents the interests of mobile network operators across the world.

The deployment of this technology over time is illustrated in Figures 4, A3, A4 and A5. The technology became available in 2006 and rapidly spread across Europe during the first development phase until 2011. With the vast majority of populated places covered, the technology continued to spread during the 2010s to reach more isolated citizens. By the end of the period, all countries except Cyprus had achieved a coverage rate above 90%. Summary statistics for both the continuous and dichotomous measures can be found in Table A3.

3.4 Controls

The population density data come from the Gridded Population of the World (GPW) data (CIESIN, 2018), a 30 arc-second resolution estimate of human population density based on counts consistent with national censuses and population registers with respect to relative spatial distribution. First, it is used as a weight for the measure of access to broadband mobile internet (see Section 3.3). We also compute the average population density at the constituency level to proxy for urbanization. Following Henderson et al. (2011, 2012), we proxy regional development by nighttime light (Li et al., 2020) which will also be included as a control in the specifications. The two variables are summarized in Table A3.

4 Empirical strategy

In this section, we present two identification strategies to understand the effect of broadband mobile Internet availability and how it can be described in terms of political platforms. Our main assumption is that people vote at least partially according to their ideological views, such that changes in support for a given value by individuals actually translate into changes in the party's vote share accordingly. Therefore, we chose to focus on the first round of lower chamber elections, as we believe they are more likely to incentivize expressive voting.[10]

4.1 Difference-in-difference approach

The first approach we employ is to categorize parties into consistent political blocs based on certain dimensions of their political platform (e.g., anti-EU parties, right-wing populists). Subsequently, we conduct separate estimations using the following difference-in-differences (diff-in-diff) specification at the constituency × year level for each political bloc:

$$Vote_{it} = \beta \cdot 3G_{it}^{+} + \alpha_i + \tau_{country(i)t} + \gamma \cdot X_{it} + \epsilon_{it} \tag{1}$$

where i indexes constituencies, and t indexes years. The dependent variable $Vote_{it}$ is the political bloc's vote share. $3G_{it}^{+}$ is the treatment status, and β is the main coefficient of interest. α_i is a constituency fixed effect controlling for any non-time-varying characteristics of the constituency.

[10] See, for instance, (Pons and Tricaud, 2018).

This could be, for example, the country's voting system, whether the constituency is part of the administrative capital, whether it lies by the sea, etc. $\tau_{country(i)t}$ is a country $\times$ year fixed effect (i.e., election fixed effect). It controls for any national level and time varying characteristics such as inflation, the ruling party, etc.[11] X_{it} is a vector of additional constituency's level controls, namely the log of the constituency population density and the average nighttime light. Standard errors are clustered at the constituency level.

By definition, $Vote_{it}$ is observed only if at least one party of the bloc decided to run in the constituency, which could bias the results (Katz and King, 1999). As a robustness exercise, we estimate the coefficient(s) of interest, including those non-contested districts, imputing for the non-contested districts a null vote share. We also present the results of an event study.[12] And because the set-up is actually a staggered diff-in-diff, we compute the estimator suggested by Dube et al. (2022).

This diff-in-diff approach in this multiparty elections setup has some limitations. Firstly, it requires assumptions on which political blocs are relevant to consider for estimating the effect, which may eventually influence whether any significant effect is found by the researcher.[13] Secondly, because it does not explicitly deal with the dimensions along which candidates are grouped, it prevents us from taking full advantage of the rich variations initially offered by the data.

4.2 Random forest approach

To overcome these limitations, we utilize the General Random Forest (GRF) algorithm (Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019). As the algorithm is particularly suited to estimate heterogeneous treatment effects and uncover which variables drive the heterogeneity, it allows us to extract more information from the electoral data and adds robustness to our results. This alternative empirical strategy consists of two main steps: (1) the treatment effect estimation and (2) the test for heterogeneous effects along political dimensions. Appendix C presents the intuition of the algorithm and how it compares with the diff-in-diff approach, a deeper analysis of how the GRF can be used to overcome many econometric challenges when analyzing multiparty elections, and the detailed implementation of the algorithm in this study.

The treatment effect is estimated at the party $\times$ constituency $\times$ year level according to the following non parametric equation:

$$Votes_{kit} = f(X_{it}, Platform_k) + \beta(Platform_k) \cdot 3G_{it}^{+} \tag{2}$$

[11] These fixed effects are estimated separately for each type of candidate, such that all of the mentioned characteristics could matter differently for each political bloc.

[12] For each constituency, the outcome is only observed in some years, while the treatment is observed annually. Therefore lags and forwards of the treatment can be defined either at the year or election levels, and I show results based on both definitions.

[13] See appendix C

where k indexes the political parties and $Platform_k$ is party k's political platform. The output is a vector $\beta(\widehat{Platform_k})_{it}$ of the Conditional Average Treatment Effect (CATE). How to use fixed effects in machine learning is an active area of research, and we decided to use the sufficient representation approach (Johannemann et al., 2019). In practice, we include time and country level average of the population and wealth in X_{it} to account for the country-level dynamics. We cannot include constituency-level fixed effects directly, but the estimation procedure still accounts for the difference in probability of being treated across units based on observable characteristics (see Appendix C for more details).

Then, in order to test for heterogeneity along specific political dimensions, following Athey and Wager (2019) and Semenova and Chernozhukov (2021), we estimate the best linear projection of the CATE, according to the following specification:

$$\beta(\widehat{Platform_k})_{it} = \eta \cdot Platform_k + \zeta_{political\ family(k)} + \epsilon_{kit} \tag{3}$$

where $\zeta_{political\ family(k)}$ are 11 political families fixed effects as defined in the CHES data. This way, we can estimate the correlation between the CATE and specific political dimensions within those ideologically homogeneous political families. The model we use also allows for municipality level and time heterogeneity, i.e., the CATE β is actually not only a function of $Platform_k$ but of t as well as X_{it} too. We also use two separate datasets (survey and manifesto-based ideologies) to see whether the algorithm gives consistent output.

Interpreting the coefficients of equation (3) is not straightforward. For instance, a positive coefficient indicates that an increase in the corresponding ideological dimension will result in a larger treatment effect but does not inform on the sign itself. However, it still means that candidates with high values along this dimension will obtain more votes in places with $3G^+$ compared to places without (but they could very well lose votes in both cases). For the sake of interpretation, we normalize the vector $Platform^k$. In the case of the survey-based ideology, each dimension varies between -1 and 1 given the symmetrical nature of the axis (e.g., left-right) such that 0 would correspond to the centrist position. For the manifesto based ideology, each dimension varies between 0 and 1 since dimensions are already signed. We can finally compare these coefficients to the constant term and/or the fixed effects to understand the effective sign of the treatment effect for a given platform.

5 Results

5.1 Differences-in-differences evidence

Before presenting our main results, we apply the diff-in-diff approach presented in Section 4.1 and ensure that our data replicates previous findings from the literature. Firstly, we demonstrate that broadband mobile internet had, if anything, a negative effect on turnout, rather in the second

election after having access to $3G^+$. Secondly, consistent with Manacorda et al. (2023), we show a positive effect on the vote share of eurosceptic parties, those hostile toward migrants, traditionalists, or the most economically liberal parties. However, given that these political stances tend to be correlated, this finding should not be entirely surprising. Additionally, we find in line with Guriev et al. (2021) a similar pattern for right-wing populist parties, but this cannot be solely explained by their anti-elite rhetoric. Main results are shown in Tables 1 and 2, completed by Table A5. Further details can be found in Appendix B.

5.2 Random forest evidence

In this section, we provide a comprehensive analysis of the treatment effect using the GRF algorithm. We demonstrate that the algorithm successfully found heterogeneity based on certain political dimensions, and discuss their relative importance. Then we study how those dimensions correlate with the treatment effect by estimating Equation (3). Finally, we integrate these findings with stylized facts from external survey data and explore their connections to the existing social and political science literature.

5.2.1 Treatment-effect estimation

Forest fit Table 3 assesses whether the forest succeeded in finding any heterogeneity in the treatment effect and evaluates the calibration of the model.[14] Both the survey and manifesto data analyses reveal that the forest achieved significant heterogeneity. However, the calibration on the survey data appears to be somewhat better compared to the manifesto data. We see two potential explanations. First, it could be that better calibration would be achieved if we performed variable selection on this data set and fitted a new random forest. Second, if voters are able to distinguish between the actual political platform and the manifesto content, using the latter may actually hurt the estimation because of measurement errors.[15] Figure 5 presents the distribution of the CATE for both ideology measures separately. The correlation coefficient between the two distributions is 0.450 .

Variables' importance The two separate analyses of survey and manifesto data to represent political platforms give pretty consistent results in terms of variable importance. First, economic

[14]Following Athey and Wager (2019) and Chernozhukov et al. (2018), the CATE is fitted as a linear function of the out-of-bag forest estimate. The mean prediction absorbs the average treatment effect, and the coefficient of the differential prediction can be interpreted as a measure of the quality of the estimates of treatment heterogeneity. Practically, if the coefficient is 1 then the treatment heterogeneity estimates are well calibrated; if it is significant and positive, then at least it could be evidence of a useful association between the estimate and its corresponding out-of-bag prediction, and be interpreted as the forest succeeding in finding heterogeneity.

[15]At the end of this section, we regress the treatment effect on both survey and manifesto-based ideology and observe that all except one variable of the survey-based ideology remains significant while most of the manifesto-based ideology does. This suggests that manifesto data indeed possesses explanatory power.

issues are shown to be important in both cases. However, the classical left-right axis is not the only relevant cleavage. Both the position toward the EU and the support toward rural populations are found to be relevant in the two cases.

Figures 6 and 7 display the variables importance for the survey and manifesto-based ideologies, respectively.[16] The red line on each graph represents the value each variable would have if all variables were equally important.[17] It is also useful to compare the group of candidate-level variables with the group of constituency-level ones. The treatment effect is likely to be heterogeneous with respect to both levels, but it is plausible that one source of heterogeneity dominates the other. Indeed, we would expect the most important candidate characteristics to have higher importance than the constituency's characteristics. Both approaches regarding the average variable importance and the comparison of parties versus constituency variables are remarkably consistent, i.e., the most important constituency's characteristic does have an importance close to the expected average (dashed red line). Additionally, it is rather wealth than population density that seems to explain most of the constituency-level heterogeneity.

On the one hand, the most important variable when we use survey-based ideology is the left-right economic axis, closely followed by the candidates' support for the rural population. Then come constituency wealth, the candidate's opposition to the EU, and their leaning toward traditionalist views, all three with comparable levels of importance. The remaining six variables have an importance below the expected average (i.e., the red line). The four most important variables in the manifesto ideology, in decreasing order, are as follows: (i) Need for state funding of cultural and leisure facilities, including arts and sports, (ii) general favorable references to the middle class, (iii) favorable mentions of the European Community/Union, and (iv) market regulation, which supports policies designed to create a fair and open economic market.[18]

On the other hand, for the manifesto-based ideology, 17 variables have an importance higher than the average, and 19 variables have an importance higher than the most important constituency-level characteristic. Because this model includes many more variables than the previous one, we conduct a second exercise to isolate the most important drivers of the CATE. Before properly estimating equation (3), we perform a LASSO analysis, already restricted to the 19 most important variables according to Figure 7. The coefficients' paths are shown in Figure 8, where the coefficient estimates are plotted as a function of the LASSO penalty parameter. In this graph, the most important variables should have strictly positive or negative values even for the highest

[16] There is no unique way to measure how important each variable is in each model, so I follow the method suggested by Athey and Wager (2019). In a given tree, the importance of a variable is proportional to the number of splits regarding this variable, each split being weighted by the inverse of its depth.

[17] This value is suggested by Athey and Wager (2019); Basu et al. (2018) to perform variable selection.

[18] According to the codebook, this category may include calls for increased consumer protection, increasing economic competition by preventing monopolies and other actions disrupting the functioning of the market, and defending small businesses against the disruptive powers of big businesses, among others. A separate category named "economic planning" exists in the data.

values of the penalty parameter. In Figure 8, one can distinguish two groups of variables: (a) the top 8 variables in orange, which all have a positive size starting from log-penalty parameters of at least 9, and (b) the remaining 13 variables, which gradually become non-null as the penalty decreases. The leading group of variables includes, in decreasing order: (i) Favorable references to the middle class, (ii) State funding of cultural and leisure facilities, (iii) Supply-side oriented economic policies, (iv) Favorable mentions of the European Community/Union in general, (v) Market regulation, (vi) Specific policies in favor of agriculture and farmers, (vii) Civic-mindedness (e.g., appeals for national solidarity and the need for society to see itself as united), and (viii) the need to expand and/or improve educational provision.

For the remainder of this paper, we focus on the four most important variables from the survey-based ideology and the latter eight variables from the manifesto-based ideology as our variables of interest.

5.2.2 The treatment effect as a function of political platforms

In this section, we first present all the specifications we estimate to describe the correlation between political platforms and the effect of broadband mobile Internet. We then detail the empirical results themselves and their interpretation.

Specifications Table 4 displays the estimates of equation (3) using the best-linear projection estimator (Semenova and Chernozhukov, 2021), with the survey-based variables as independent variables. The dependent variable, i.e., the CATE was itself fitted on the survey data, except in column (4) where it was fitted on the manifesto data. This is done to cross-check the consistency of the results across the two databases. In column (1), we include only the most important variables (those above the average variable importance) as control variables. In column (2), we additionally control for the other ideological dimensions. Column (3) includes political family fixed effects, as defined in the CHES data. Finally, column (5) weights the observations by the average vote share a given party obtained in the data. This is done to give more importance to parties that are more likely to be part of the government compared to small outsider parties.

Table 5 performs a similar analysis of the CATE but is now focused on the results obtained using the manifesto data. The specification of column (1) only includes the 8 most important variables, as described in the previous LASSO analysis. Column (2) controls for the additional 10 other dimensions, which are more important than constituency-level heterogeneity (see Figure 7). Column (3) introduces political family fixed effects; column (4) uses the CATE fitted on the survey data as a dependent variable; and column (5) weights observations as in Table 4.

Figures 9 and 10 summarize our findings by showing our preferred estimates (respectively Table 4 column (3) and 5 column (3)) and comparing them with the constant term to help us with the interpretation of the overall effect of $3G^+$. In Figure 9 the constant term for the survey-based ideology is precisely zero, such that we can directly interpret the sign and magnitude of a given

coefficient as the change in vote for a fictional candidate, which would hold median positions on all dimensions except for the one related to the coefficients, along which they would hold the most extreme position. In Figure 10 the constant term for the manifesto-based ideology is negative but very small compared to all those significantly different from zero. Therefore, we neglect it in our interpretations and again directly interpret the sign and magnitude of a given coefficient as the change in vote for a fictional candidate, but that now would solely campaign about the dimension related to the coefficient.

The purpose of Table 6 is to explore the relationship between the treatment effect and the candidates' political platforms, represented by both the survey and manifesto data. The purpose is twofold: first, we assess the robustness of the results of Tables 4 and 5, second, we compare whether one measure is a better predictor of the CATE than the other.[19] Columns (1) and (2) control for political family fixed effects with the CATE from the survey-based and manifesto-based ideologies, respectively. Columns (3) and (4) use the same weighting scheme as column (5) in Tables 4 and 5.

Finally, Table 7 focuses on constituency-level heterogeneity, both regarding population density and wealth. Using the same specification as Table 5 column (3), the sample is now split between constituencies that are below and above median regarding the two constituency characteristics. Columns (1) and (2) display the results for sparsely vs. densely populated constituencies; columns (3) and (4) compare poor and rich places; and columns (5) and (6) compare poor and rich places, but now the median is defined at the national level. The correlation between being sparsely populated and below median wealth in the data is 0.060 and -0.019 when the latter is defined at the national level.

We now present and interpret our main empirical results.

The fall of the middle class consensus We first show that the left-right gradient poorly explains the effect of mobile broadband Internet. Then, based on the correlation between political platforms and the effect of broadband internet, we argue that $3G^+$ rather exacerbated the polarization within the European middle class. Indeed, access to online information and political content could fragment the middle class by exposing different segments with conflicting preferences to distinct political platforms related to their economic and educational status.

When the treatment effect is fitted to the survey data, the most important variable is the left-right economic axis (Figure 6). However, its associated coefficient is only significantly different from zero (favoring the left) when comparing candidates within their political families, as seen in Table 4 columns (3) and (4). The effect is not precisely estimated when candidates are weighted by their political importance, as shown in Table 4 column (5). This could happen, for instance, if

[19]While this analysis is too short to definitively determine which approach strictly dominates the other, we believe it is a promising path toward a proper combination of measures based on communication strategies used by politicians with less manipulable data, such as expert surveys.

the overall relationship between the CATE and the left-right axis is U-shaped.

Conversely, when fitted to the manifesto data, the second most important variable according to the LASSO results, and with the larger effect in Table 5, is the positive reference to the middle class, as shown in Figure 7. Two other selected variables are explicitly related to economic policies: market regulation and supply-side-oriented policies (Figure 7). Both variables have a positive correlation with the CATE, with the relationship being stronger for supply-side-oriented policies, as shown in Table 5.

The negative correlation between references to the middle class or civic mindedness and the effect of broadband mobile internet appears to be more prominent in the poorest constituencies (Table 7 columns (3) and (5)). This suggests that it is the lower middle class that has become less inclined to subscribe to the unifying vision of society presented by politicians. Both supply-side oriented policies and state funding of education show a positive correlation with the treatment effect; however, the positive relationship with market regulation is more evident in sparsely populated and rich constituencies (Table 7 columns (1), (4), and (6)).

Figure 12 illustrates country-level heterogeneity using a similar specification to Table 5 column (1) but controlling for political family fixed effects. The strongest negative relationship between positive references to the middle class and the effect of mobile internet is observed in Ireland, a country with "a dramatic collapse in levels of trust in government and the political system in the aftermath of the [2008] economic crisis" (Nolan et al., 2014) that popularized the term "middle class squeeze" (Hugrée et al., 2017). Conversely, the only three countries where the correlation is positive are Estonia, Portugal, and Romania, three of the poorest countries in the sample. In these countries, the middle-class narrative is more likely to be presented as a promise rather than a fiction used to encompass a vast majority of the voters.

The middle class has historically been the most and even abusively self-identified social class in Europe. As described by Hugrée et al. (2017) in their study of inequalities and social classes in Europe, the term "middle class" does not solely refer to an economic status but also encompasses a broad range of socio-cultural practices. By appealing to the middle class, parties may aim to garner support from a large and diverse segment of the population while also downplaying potential class conflict. But the European middle class is not a homogeneous group and can be internally divided based on two main factors: education and economic capital. On the one hand, there are more educated individuals who are often associated with working in the public sector and voting for left-wing parties. On the other hand, there are middle-class individuals with economic capital who tend to support politicians advocating for business-oriented policies. Keeping this in mind, Tables 5 and 7 suggest that mobile broadband internet made this cleavage more salient. Indeed, in addition to the negative coefficient for the appeal to the middle class, both supply-side oriented policies and market regulation reflect the demands of the economically endowed middle class, while state funding of education would satisfy the more educated stratum. All three of these variables correlate

positively with the CATE. This finding is in line with the idea that the previously self-identifying middle class is experiencing a polarization around its internal cleavages related to their economic and educational status. The Internet would therefore amplify the long-term trend documented by Gethin et al. (2021).[20]

One potential explanation for the unsuccessful appeal to the middle class on the Internet is the highly segmented audience and its tendency to spread outrageous content, especially compared to legacy media such as TV. The latter has historically been a privileged medium for promoting middle-class values (Ferrara et al., 2012).[21] But on the Internet, the incentives are quite different, and one of the most efficient ways to increase user interactions is with outrageous content (Guadagno et al., 2013).

The European Social Survey results in Figure 11 indicate that people with broadband internet access tend to spend more time reading newspapers or listening to the radio, but they watch substantially less TV. This suggests that broadband Internet has provided a better substitute for TV compared to other media. Additionally, looking at Figure A8, we observe a monotonic positive relationship between income and propensity to watch the news online initially. However, over time, the lowest income category's consumption of internet news increases to a level similar to that of the richest category. This trend is consistent with the idea that the Internet may polarize economic and political issues, as individuals may prefer like-minded news sources (Gentzkow and Shapiro, 2010).

Overall, the media economy of the Internet versus TV has led voters to be more exposed to a divided picture of society, where the middle class no longer plays its cohesive role but is fragmented by its economic and social heterogeneity. This interpretation is supported by the negative (although not always significant) correlation of the CATE with an overall civic-mindedness discourse that presents a unified society (Table 5). The decrease in political identification with this consensual group may have either fueled class conflict or left room for identification tied to non-economic criteria, favoring tribalism and support for communitarian parties, as documented by Manacorda et al. (2023).

Opposition to the EU The third most important variable in the survey-based ideology is the position toward EU and a similar variable is picked up in the manifesto-based ideology (Figures 6

[20] They show how between 1948 and 2020, the vote for leftist parties has gradually become associated with higher-educated voters, giving rise in the 2010s to a disconnection between the effects of income and education on the vote in Western democracies. And they provide evidence that the reversal of this education cleavage is linked to the emergence of a new "socio-cultural" axis of political competition.

[21] Besides its potential political impacts, the main reason why TV news may seek to present a positive and peaceful narrative about society is economic. As the CEO of TF1, one of the largest TV networks in France, explained in his book, "But from a business perspective, let's be realistic: basically, TF1's job is to help Coca-Cola, for example, sell its product. But for an advertising message to be perceived, the viewer's brain has to be available. The aim of our programs is to make them available: in other words, to entertain them, to relax them, to prepare them between two messages." (*Les dirigeants face au changement, Éditions du Huitième jour, 2004*)

and 7). The more favorable of the European Union a candidate is, the fewer votes they will obtain in places with $3G^+$ compared to places without. This is also supported by our diff-in-diff analysis (see Section B). This is consistent with Manacorda et al. (2023) who show an increase in the vote share of eurosceptic candidates once a municipality accesses mobile Internet. When estimated at the country level, this negative relationship is mostly driven by Italy, the UK, and France (Figure 13).

Tradition and immigration In Table 4, both traditionalism and hostility toward migrants have either a non-significant or negative relationship with the effect of mobile Internet and are dropped by the variable selection procedure for the manifesto-based ideology. This may appear to contradict the findings of Manacorda et al. (2023), who reported a positive effect of mobile Internet on the vote share of candidates hostile toward ethnic minorities, advocating restrictive immigration policies, and holding traditionalist views. However, upon examining Figure 3, we observe that traditionalism is highly correlated with hostility toward migrants, and despite the link being weaker, it is the second most correlated dimension with euroscepticism. Therefore, if we run diff-in-diff estimates as in Table 1 Panel A, and observe that eurosceptic candidates benefited from $3G^+$, it is not entirely surprising that we also find a positive effect for traditionalists and anti-immigration candidates since they will tend to be the same candidates (Table 1 Panel B and C). This illustrates one advantage of the GRF algorithm, which overcomes this issue by accounting simultaneously for several dimensions and comparing their respective explanatory power during the estimation.

Support for the Rural World Focusing on survey-based ideology, the second most important variable is the support of the rural population, while for manifesto-based ideology, support for agriculture and farmers is also selected (Figures 6 and 7).[22] In both cases, the variables have a positive and significant effect, particularly in sparsely populated and economically disadvantaged constituencies (Tables 4, 5, and 7). This result is also supported by our diff-in-diff analysis (Tables 1 qnd 2).

Portrayed as a positive figure of the native labor class, forgotten and deserving of help, this archetype can be contrasted with either the urban disconnected elite or the unemployed youth with a migration family background (Coquard, 2016). Historically conservative, farmers are now also targeted by far-right movements in Europe.[23]. However, they have become a major focus for

[22]Although related, the two variables do not capture exactly the same idea. In the EU, farmers account for 4.4% of total employment, while 28.8% of the population live in rural areas (`https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=388145`, accessed 07/17/2023). However, farmers are overrepresented in the public/media discourse related to rural areas, as explained by the sociologist Benoit Coquard (see, for instance, "Le complexe rural" May 29, 2023, `https://www.radiofrance.fr/franceculture/podcasts/lsd-la-serie-documentaire/grandir-a-la-ferme-9003388`, consulted 08/07/2023).

[23]Politico, June 23, 2023, `https://www.politico.eu/article/why-eu-conservatives-epp-are-targeting-the-farmers-vote-elections-2024-eu-parliament/`, consulted 08/07/2023

the entire political spectrum due to their intrinsic connection to environmental issues. In a context where legacy media may have a metropolitan bias (Chandler and Munday, 2016), traditional media are disregarded by the rural population, and therefore the Internet could become its privileged medium. As an anecdotal illustration, a survey conducted in France in 2021 found that approximately 9% (resp. 47%) of people from Paris, 18% (55%) in urban municipalities, and 24% (58%) in rural municipalities expressed anger (resp. suspicion) toward the media.[24]. These two aspects make the Internet a fertile environment for political platforms targeting rural constituencies or referring to agriculture.

Leisure, culture and mass media In the estimates based on manifestos, the coefficient related to the support for state funding of cultural and leisure facilities, including arts (e.g., museums, art galleries, libraries, etc.), cultural mass media, and sport, is consistently negative and significant across all specifications (Tables 5 to 7). This finding is quite interesting as it suggests that broadband Internet not only changed political communication and news coverage but also induced major changes in the overall cultural, leisure, and entertainment consumption patterns of voters and their socialization. The Internet remains widely used for purposes other than political information. For instance, Figure A7 shows the main reasons why people use the Internet based on the same French survey as before. Approximately 90% of people use social media to socialize with friends and relatives, between 77% in 2012 and 82% in 2015 for entertainment, while 54% and 71% respectively, mention following the news. This suggests that broadband Internet serves as a substitute for cultural and leisure consumption as well.

The fact that TV was largely oriented toward entertainment could also explain why mobile broadband would have a greater impact on this specific media consumption rather than on newspapers or radio. Broadband Internet has made relatively cheap access to mass culture possible, and the development of cultural content on the platform initially had very little support from the state.[25] Therefore, the negative impact on candidates supporting public funding for these activities is likely to be driven by a cost-benefit analysis of consumers and taxpayers. The availability of broadband Internet offers an alternative and convenient way to access cultural and entertainment content, leading to reduced demand for state-funded facilities in these areas.

Other dimensions The last three remaining variables in the survey-based ideology are: (i) support for anti-environmental policies; (ii) support for law and order; and (iii) anti-elite rhetoric. Anti-environmental views are positively and significantly correlated with the treatment effect. But neither (ii) nor (iii) are significantly related to it in any of the specifications.

[24]IFOP: "Le regard des Français sur les médias et l'information", June 2021 (`https://www.ifop.com/wp-content/uploads/2021/06/118202-Resultats.pdf`, consulted 08/07/2023)

[25]The spread of broadband Internet also coincides with the generalization of consumers paying a subscription for unlimited access rather than a price proportional to their consumption.

Manifestos compared to expert survey data When the treatment effect is regressed on both the survey and manifesto data (Table 6), all coefficients for the variables derived from the manifestos remain significant, while all variables from the survey data become non-significant, except for anti-environmental views. This suggests that variations in the treatment effect are better explained by the manifesto-based ideology, and makes the case that the corpus of texts issued by political actors themselves is a promising path toward a better understanding of which cleavages are crucial in a given study.

6 Conclusion

Mobile Internet has changed election outcomes in Western Europe. We investigate the impact of broadband mobile technology on the outcomes of first-round elections in lower chambers across Europe. We combine election results with two types of data: (i) at the constituency level, we utilize the share of the population that has access to mobile Internet, and (ii) at the party level, we describe the political platform using either results from political scientists' surveys or the content of the parties' manifestos.

From a methodological standpoint, we employ the Generalized Random Forest (GRF) algorithm to examine multiparty elections and estimate the treatment effect at the constituency level as a function of the political platforms of various parties. This approach enables us to pinpoint the most significant cleavages and issues that influence the effect of broadband Internet on political results, providing a more nuanced and accurate picture of the complex interactions between technology and politics in the digital age. We contend that this multi-dimensional approach offers many desirable properties and can apply to many other constituency level shocks and candidate characteristics.

Our analysis yields four main findings. First, candidates arguing to represent middle-class interests garnered fewer votes in areas with widespread broadband mobile Internet. In contrast, those who fueled the polarization within the same middle class along the lines of wealth and education received more votes. This decline in middle-class identification left room for new non-economic identification processes, potentially benefiting communitarian parties. In line with this, our second result exposes a positive effect of broadband mobile Internet on eurosceptic parties. However, holding euroscepticism constant, there is no additional effect of mobile broadband on parties that advocate traditionalism or hostility toward migrants. Third, candidates championing for the rural population also increased their vote share in places with broadband mobile Internet, as the latter was simultaneously marginalized by legacy media and largely instrumentalized in the political debate. Finally, calling for state funding of mass culture in their manifesto decreased the vote share of the candidates in places with broadband mobile Internet, which offers convenient access to culture and entertainment. We suggest that it can be directly linked to specific characteristics of the Internet and its role as a substitute for traditional media like TV. In particular, the Internet's ability to disintermediate and segment political communication, its inclination to promote cleaving

content, and its influence on the entertainment and cultural markets.

Our findings are specific to European lower chamber elections and may not necessarily apply to other geographic areas or different types of elections. But one advantage of studying the first round of legislative elections is that it favors expressive voting, which helps candidate results reflect more accurately voters' preferences.

The impact of the Internet extends beyond its ties to political discourse. While we have observed its role in disseminating political content, it is also largely used for socialization and entertainment. Investigating the effects of these interactions between political and non-political actors presents a compelling pathway to comprehending the multifaceted changes introduced by this technology in our daily lives. As the Internet continues to shape and reshape our communication patterns, behaviors, and preferences, it is likely that new and unforeseen developments will arise in the coming years. This underscores the importance of ongoing research to track and understand the evolving relationship between technology and society.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological) 44*(2), 139–160.

Algan, Y., S. Guriev, E. Papaioannou, and E. Passari (2017). The european trust crisis and the rise of populism. *Brookings papers on economic activity 2017*(2), 309–400.

Athey, S. and G. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*(27), 7353–7360.

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Athey, S. and S. Wager (2019). Estimating treatment effects with causal forests: An application. *Observational Studies 5*(2), 37–51.

Badie, B., D. Vidal, et al. (2018). *Le retour des populismes: L'état du monde 2019*. La découverte.

Basu, S., K. Kumbier, J. B. Brown, and B. Yu (2018). Iterative random forests to discover predictive and stable high-order interactions. *Proceedings of the National Academy of Sciences 115*(8), 1943–1948.

Bonomi, G., N. Gennaioli, and G. Tabellini (2021). Identity, beliefs, and political conflict. *The Quarterly Journal of Economics 136*(4), 2371–2411.

Brodeur, A., N. Cook, and A. Heyes (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review 110*(11), 3634–3660.

Cagé, J., N. Hervé, and M.-L. Viaud (2020). The production of information in an online world. *The Review of economic studies 87*(5), 2126–2164.

Chandler, D. and R. Munday (2016). *A dictionary of media and communication*. OUP Oxford.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

CIESIN (2018). Gridded population of the world, version 4 (gpwv4): Population density adjusted to match 2015 revision un wpp country totals, revision 11.

Coquard, B. (2016). «paris? jamais de la vie» goûts et dégoûts territoriaux chez les jeunes ruraux de classes populaires. *Savoir/Agir* (3), 39–45.

Dube, A., D. Girardi, O. Jorda, and A. Taylor (2022). A local projections approach to difference-in-differences event studies. *NBER working paper*.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Ferrara, E. L., A. Chong, and S. Duryea (2012). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics 4*(4), 1–31.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica 87*(4), 1307–1340.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2021, 10). Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020*. *The Quarterly Journal of Economics 137*(1), 1–48.

Guadagno, R. E., D. M. Rempala, S. Murphy, and B. M. Okdie (2013). What makes a video go viral? an analysis of emotional contagion and internet memes. *Computers in Human Behavior 29*(6), 2312–2319.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021, 11). 3G Internet and Confidence in Government. *The Quarterly Journal of Economics 136*(4), 2533–2613.

Guriev, S. and E. Papaioannou (2022). The political economy of populism. *Journal of Economic Literature 60*(3), 753–832.

Hatte, S., E. Madinier, and E. Zhuravskaya (2021). Reading twitter in the newsroom: Web 2.0 and traditional-media reporting of conflicts. *Available at SSRN 3845739*.

Henderson, J. V., A. Storeygard, and D. N. Weil (2012). Measuring economic growth from outer space. *American economic review 102*(2), 994–1028.

Henderson, V., A. Storeygard, and D. N. Weil (2011). A bright idea for measuring economic growth. *American Economic Review 101*(3), 194–199.

Heywood, A. (2021). *Political ideologies: An introduction*. Bloomsbury Publishing.

Hooghe, L., G. Marks, and C. J. Wilson (2002). Does left/right structure party positions on european integration? *Comparative Political Studies 35*(8), 965–989.

Hugrée, C., É. Penissat, and A. Spire (2017). *Les classes sociales en Europe. Tableau des nouvelles inégalités sur le vieux continent*. Agone.

Johannemann, J., V. Hadad, S. Athey, and S. Wager (2019). Sufficient representations for categorical variables. *arXiv preprint arXiv:1908.09874*.

Jolly, S., R. Bakker, L. Hooghe, G. Marks, J. Polk, J. Rovny, M. Steenbergen, and M. A. Vachudova (2022). Chapel hill expert survey trend file, 1999–2019. *Electoral Studies 75*, 102420.

Katz, J. N. and G. King (1999). A statistical model for multiparty electoral data. *American Political Science Review 93*(1), 15–32.

Kitschelt, H. (1994). *The transformation of European social democracy*. cambridge university press.

Kollman, K., A. Hicken, D. Caramani, D. Backer, and D. Lublin (2019). Constituency-level elections archive. version 14, december 16 2020.

Kollman, K., A. Hicken, D. Caramani, D. Backer, D. Lublin, J. Selway, and F. Vasselai (2019). Georeferenced electoral districts datasets.

Laver, M. and W. B. Hunt (1992). Policy and party competition.

Lehmann, P., T. Burst, T. Matthieß, S. Regel, A. Volkens, B. Weßels, and L. Zehnter (2022). The manifesto data collection. manifesto project (mrg/cmp/marpor). version 2022a.

Lemstra, W. (2017). Broadband developments in europe: A retrospective review of the determinants of supply and demand. *telecomunicacions, regulació, Unió Europea, polítiques*.

Li, X., Y. Zhou, M. Zhao, and X. Zhao (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific data 7*(1), 168.

Malgouyres, C., T. Mayer, and C. Mazet-Sonilhac (2021). Technology-induced trade shocks? evidence from broadband expansion in france. *Journal of International Economics 133*, 103520.

Manacorda, M., G. Tabellini, and A. Tesei (2023). Mobile internet and the rise of communitarian politics.

Nolan, B., C. T. Whelan, E. Calvert, T. Fahey, D. Healy, A. Mulcahy, B. Maître, M. Norris, I. O'Donnell, and N. Winston (2014, 01). 346Ireland: Inequality and Its Impacts in Boom and Bust. In *Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences*. Oxford University Press.

OECD (2021). Implementation and usage of the oecd recommendation on broadband development. (318).

Pons, V. and C. Tricaud (2018). Expressive voting and its cost: Evidence from runoffs with two or three candidates. *Econometrica 86*(5), 1621–1649.

Poole, K. T. and H. L. Rosenthal (2011). *Ideology and congress*, Volume 1. Transaction Publishers.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, 931–954.

Roemer, J. E., W. Lee, and K. Van der Straeten (2007). *Racism, xenophobia, and distribution: Multi-issue politics in advanced democracies*. Harvard University Press.

Roemer, J. E. and J. E. Roemer (2009). *Political competition: Theory and applications*. Harvard University Press.

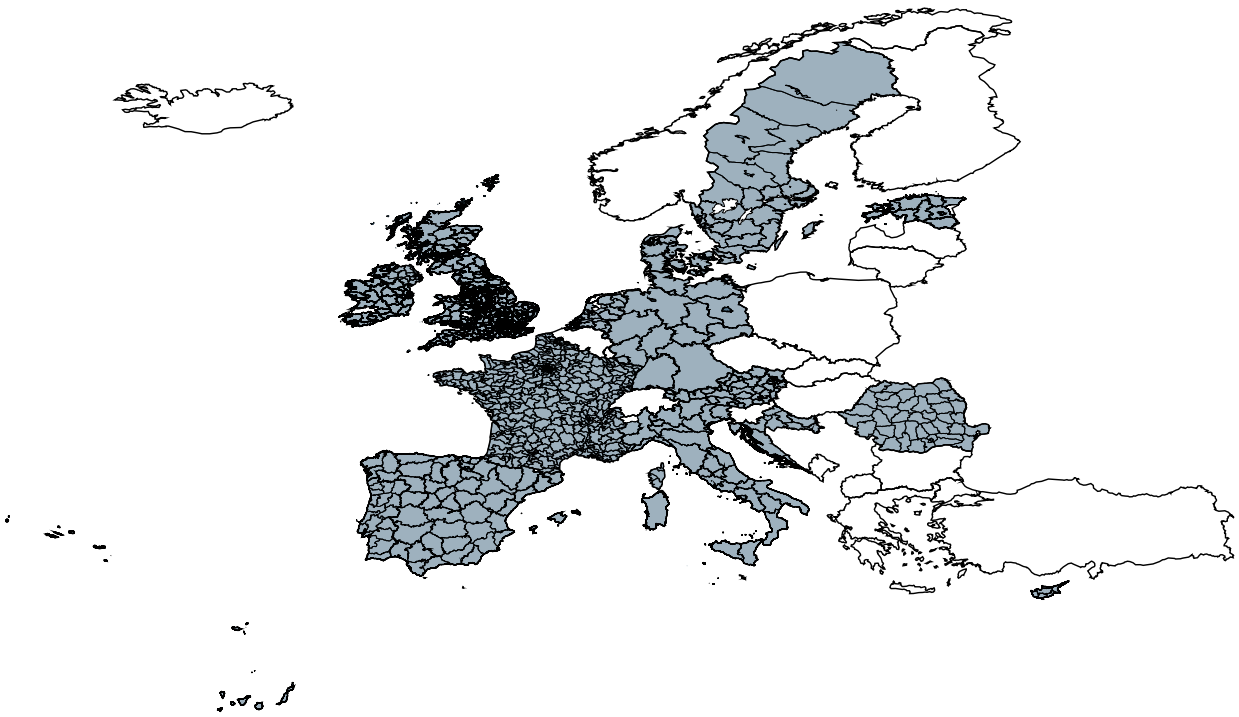
Semenova, V. and V. Chernozhukov (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal 24*(2), 264–289.

Tomz, M., J. A. Tucker, and J. Wittenberg (2002). An easy and accurate regression model for multiparty electoral data. *Political Analysis 10*(1), 66–83.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

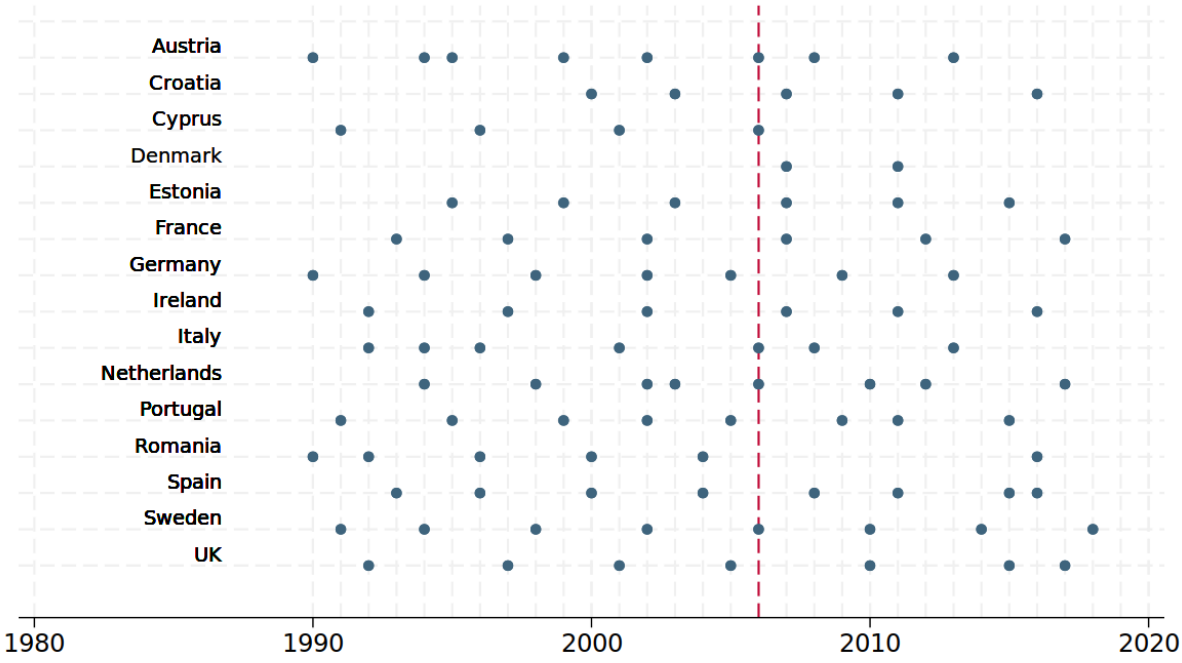
Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political effects of the internet and social media. *Annual review of economics 12*, 415–438.

Figure 1: Sample of constituencies



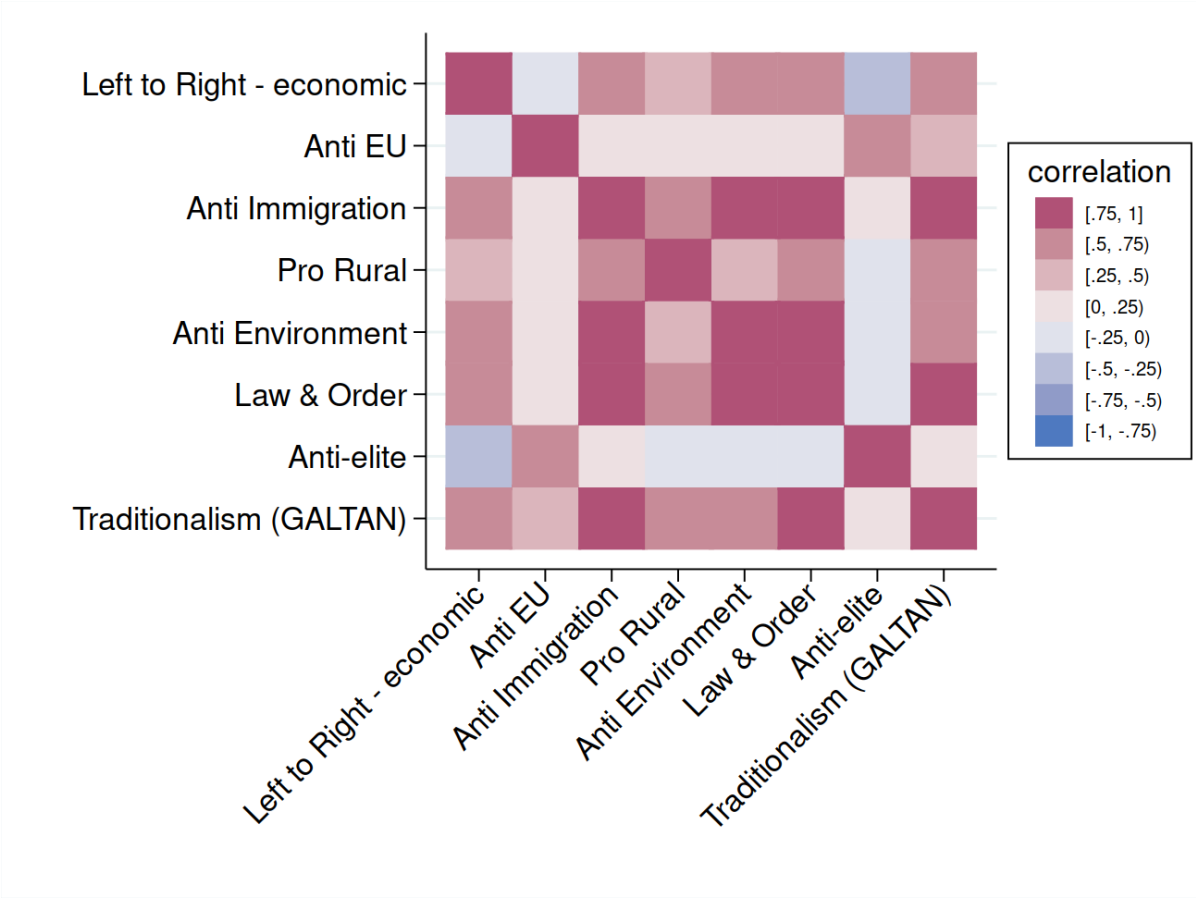
Source: GeoReferenced Electoral Districts Datasets

Figure 2: Sample of elections



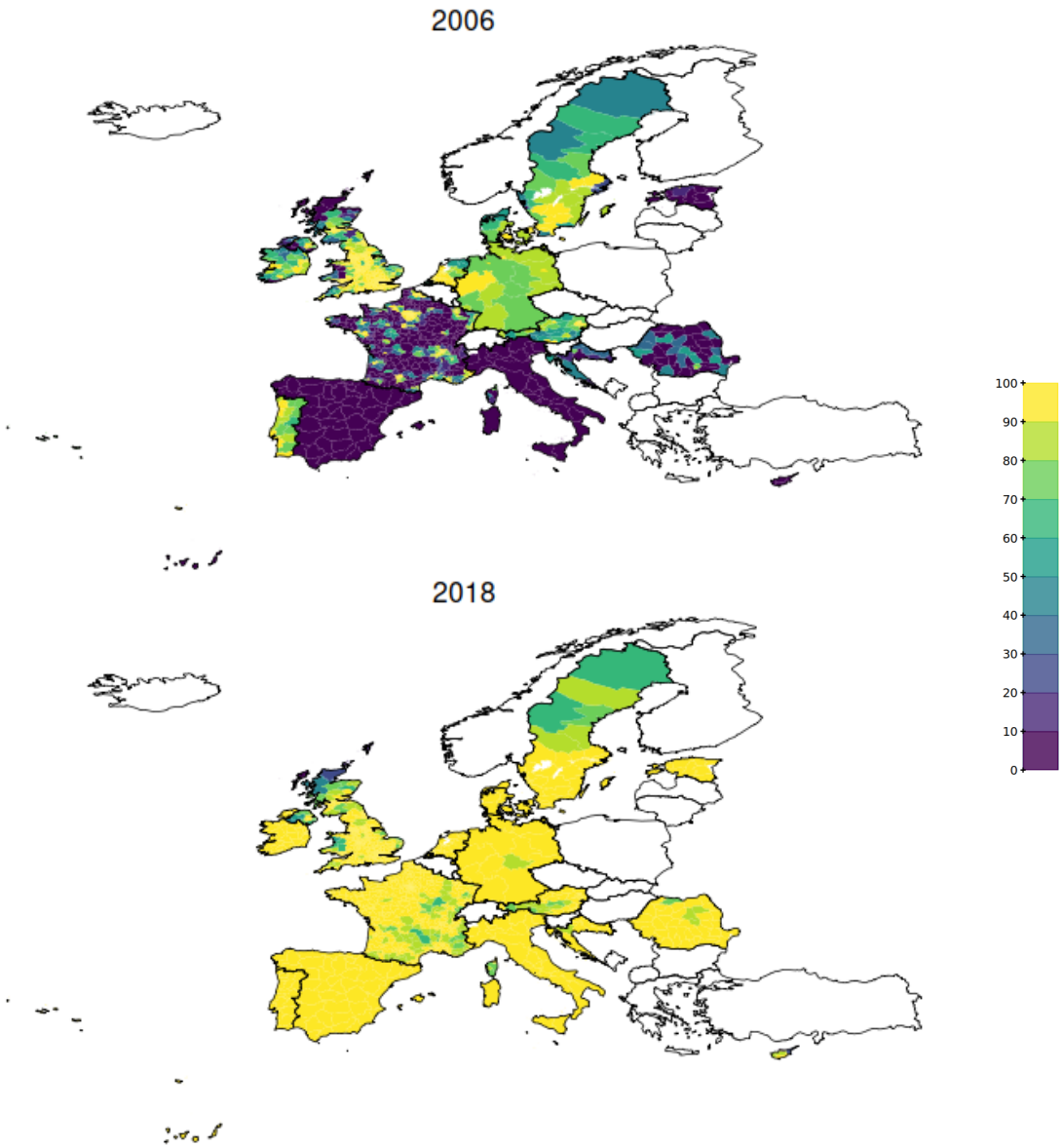
Source: CLEA election archive project

Figure 3: Correlation heat map between ideological dimensions (CHES)



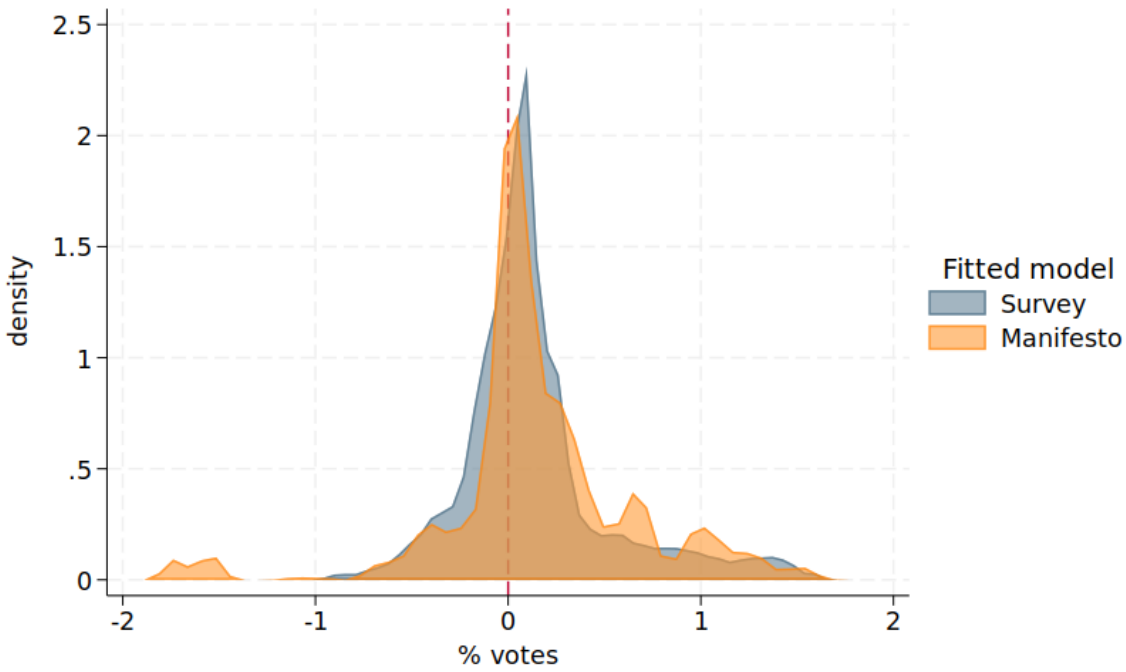
Source: CHapel Hill Expert Survey (1999-2019)

Figure 4: Maps of broadband mobile internet deployment



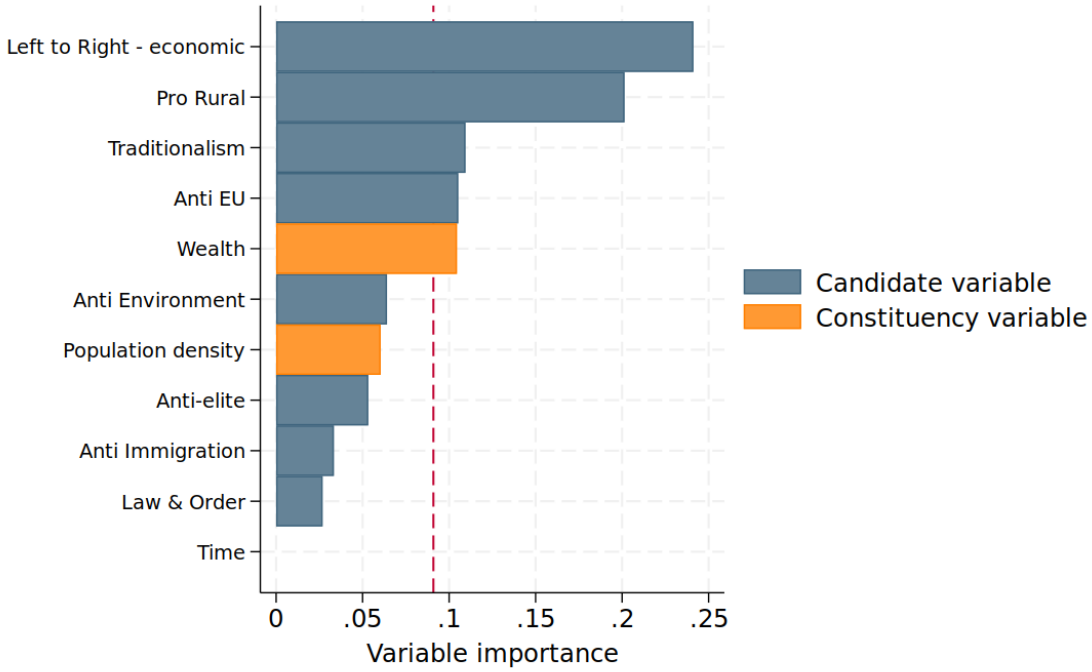
Source: Collins Bartholomew's Mobile Coverage Explorer

Figure 5: Distribution of estimated CATE



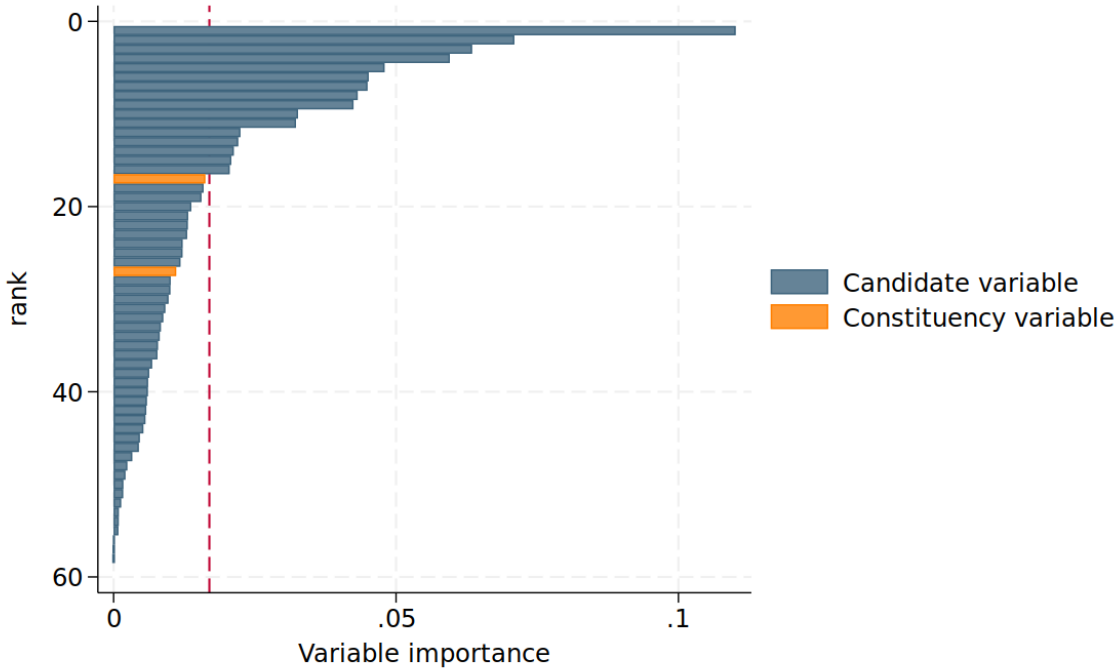
Notes: The figure presents the distribution of CATEs estimated by the GRF on survey-based and manifesto-based ideology. The unit of analysis is constituency × election × party.

Figure 6: Variable importance (survey-based ideology)



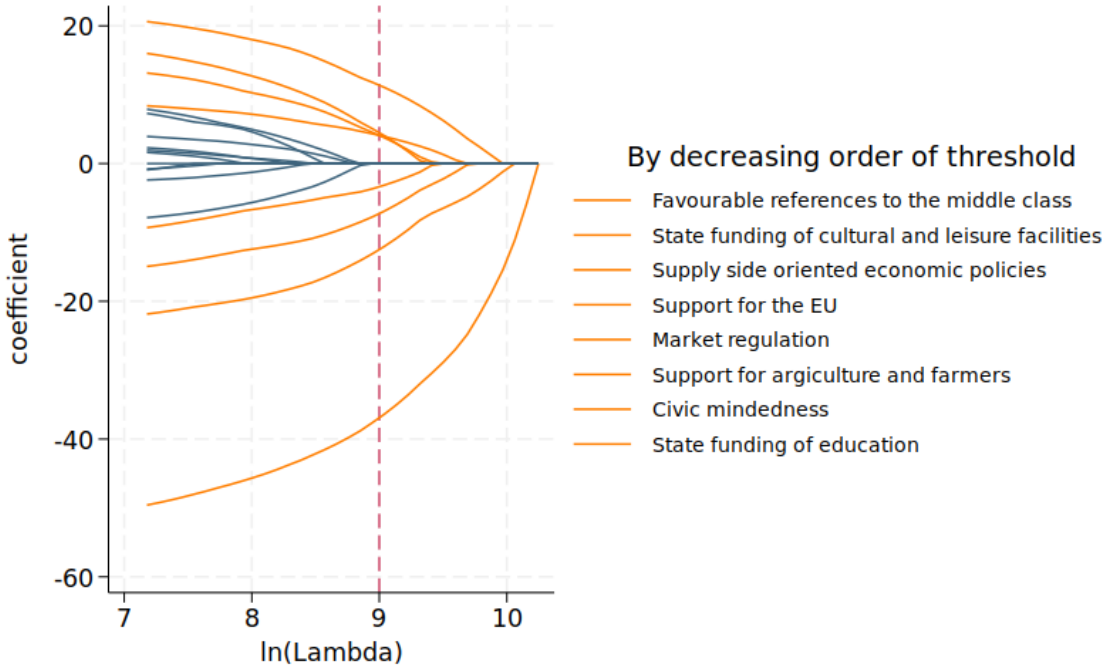
Notes: The figure presents the variable importance for the forest. In a given tree the importance of a variable is proportional to the number of splits regarding this variable, each split being weighted by the inverse of its depth. The red dashed line represent the value if all variables were equally important.

Figure 7: Variable importance (manifesto-based ideology)



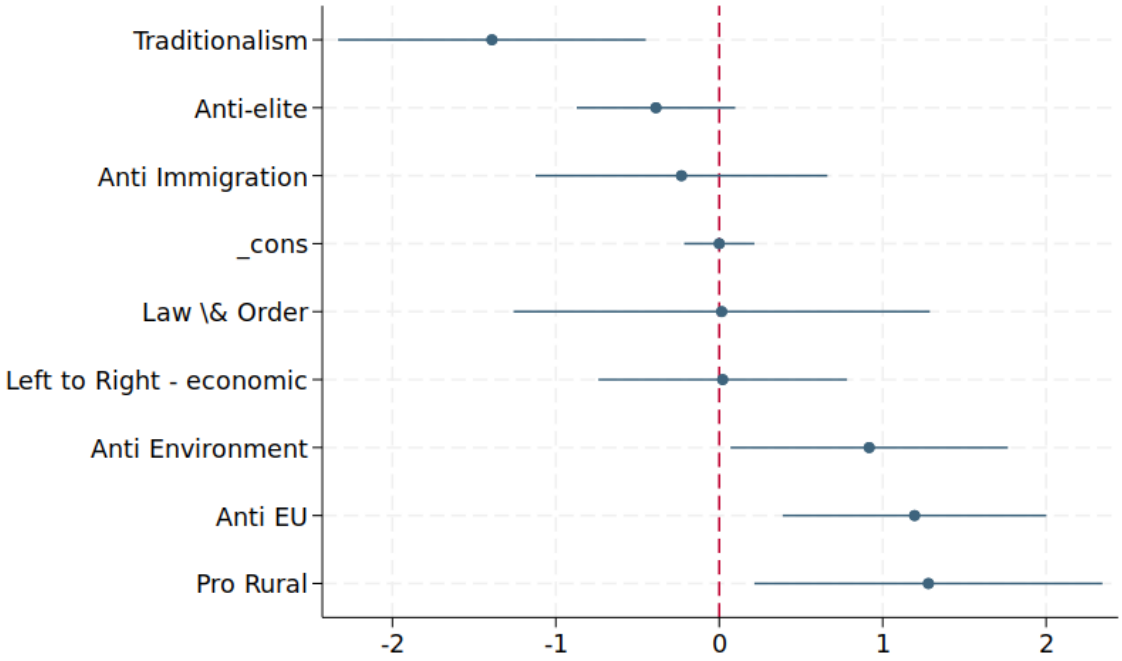
Notes: The figure presents the variable importance for the forest. In a given tree the importance of a variable is proportional to the number of splits regarding this variable, each split being weighted by the inverse of its depth. The red dashed line represent the value if all variables were equally important.

Figure 8: LASSO coefficient path (manifesto-based ideology)



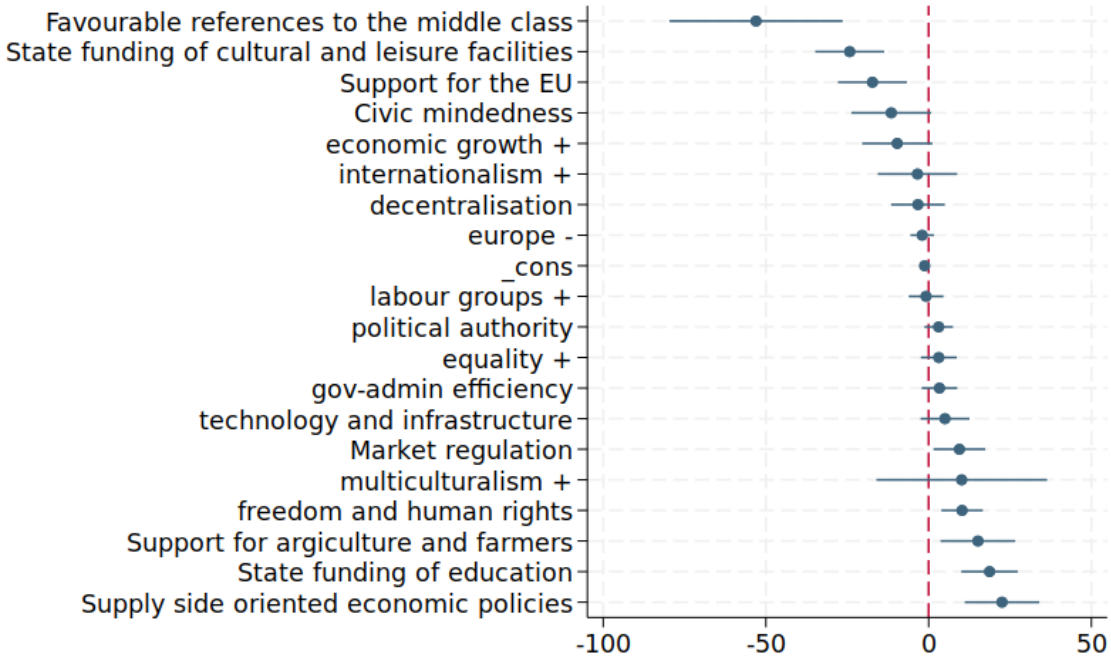
Notes: The figure presents the estimated coefficients of the CATE regressed on manifesto-based ideology variables, as a function of the penalization parameter. The red dashed line isolates the most important variables.

Figure 9: Best Linear Projection (survey-based ideology)



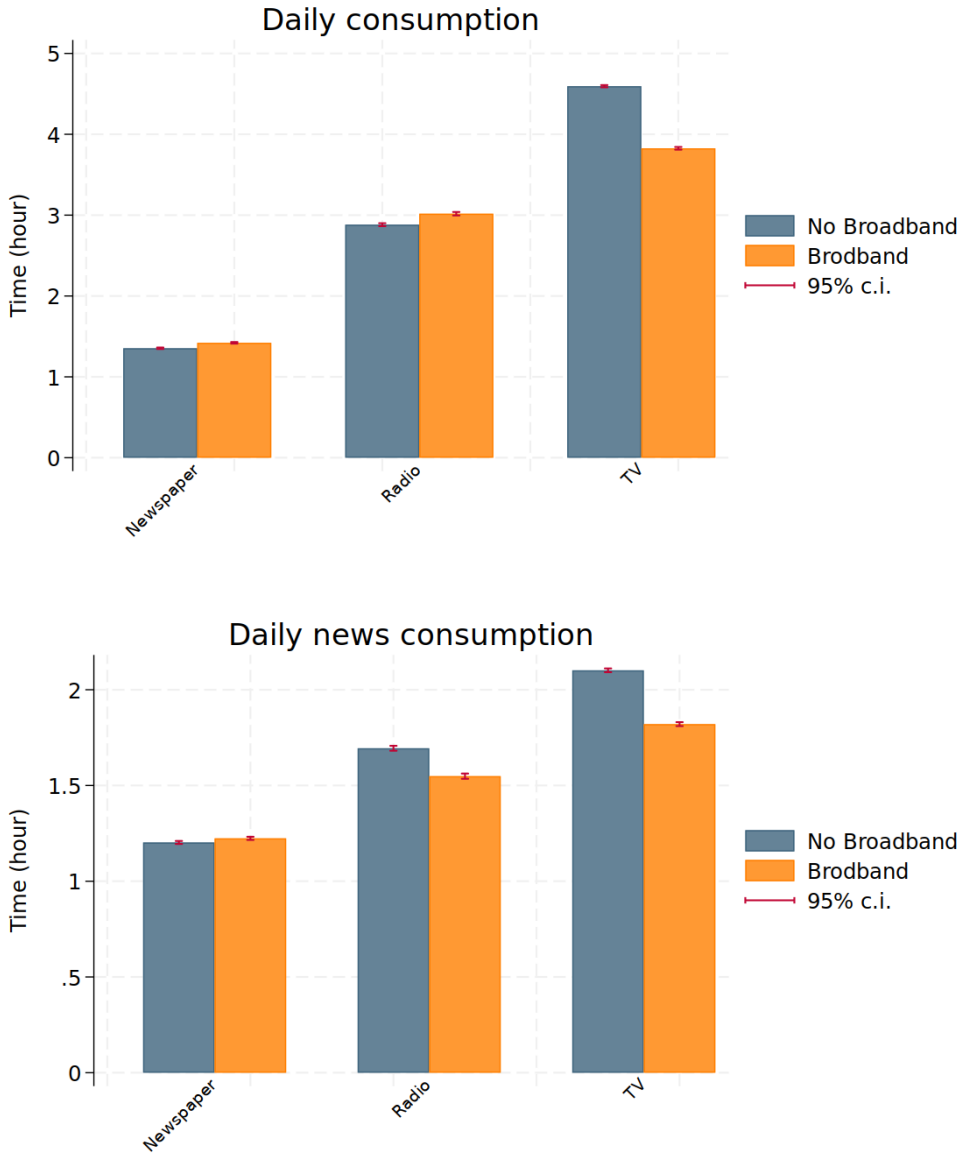
Notes: The figure presents the estimated coefficients of the CATE regressed on survey-based ideology variables, using the same specification as column (3) of Table 4. The unit of analysis is constituency × election × party. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure 10: Best Linear Projection (manifesto-based ideology)



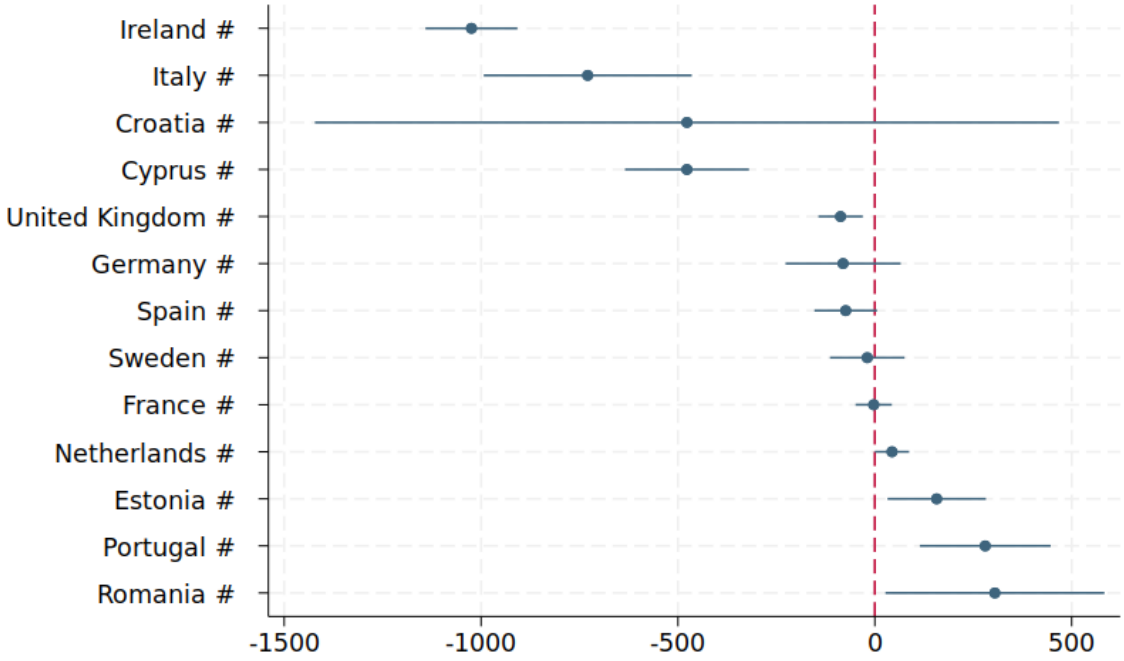
Notes: The figure presents the estimated coefficients of the CATE regressed on manifesto-based ideology variables, using the same specification as column (3) of Table 4. The unit of analysis is constituency × election × party. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure 11: Media consumption in Europe depending on Internet access



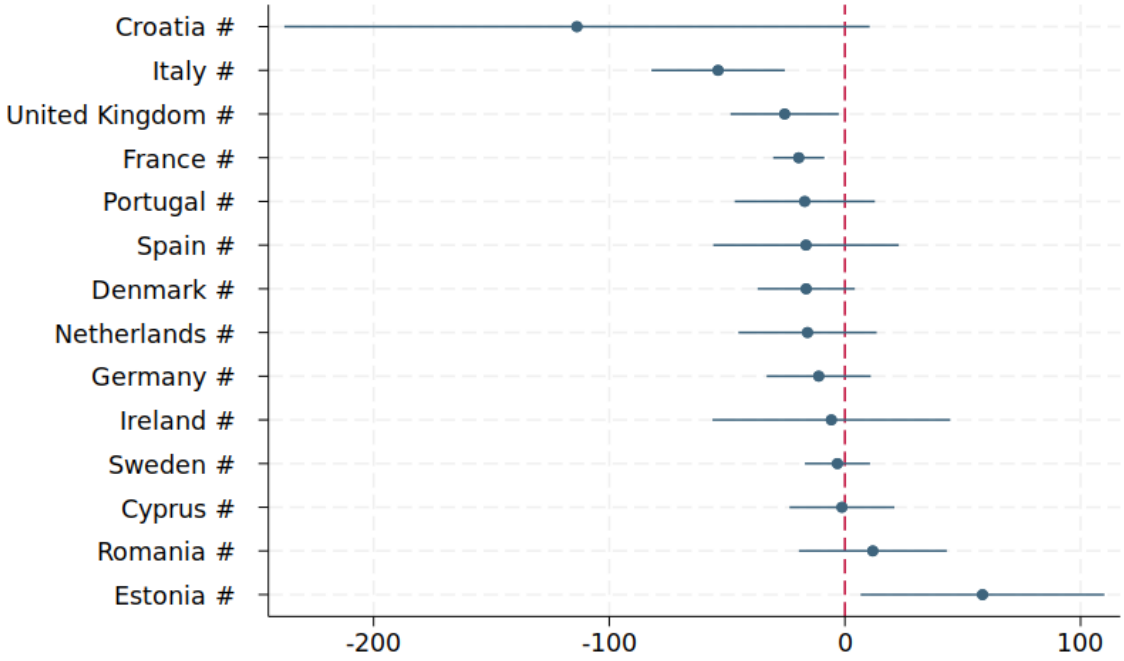
Source: ESS survey (2002, 2006, 2010, 2016, 2020)

Figure 12: Country level correlations between favorable mentions of the middle class and the CATE



Notes: The figure presents country-level heterogeneity using a similar specification to Table 5 column (1) but controlling for political family fixed effects and interacting the ideological dimension of interest with country fixed effects. The unit of analysis is constituency × election × party. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure 13: Country level correlations between favorable references to the EU and the CATE



Notes: The figure presents country-level heterogeneity using a similar specification to Table 5 column (1) but controlling for political family fixed effects and interacting the ideological dimension of interest with country fixed effects. The unit of analysis is constituency × election × party. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Table 1: Diff. in Diff. - effect of broadband mobile internet political blocs voteshares

Panel A					
Dependent variable:	Anti EU vote share (top-decile)				
3G+	-1.554***				
	(0.438)				
before 2010 × 3G+		-1.994***	-2.440***	-2.007***	-1.616***
		(0.436)	(0.530)	(0.436)	(0.365)
after 2010 × 3G+		6.808***	7.026***	6.781***	2.473***
		(1.058)	(1.016)	(1.067)	(0.695)
Observations	6,162	6,162	6,162	6,126	8,531
R-squared	0.847	0.847	0.861	0.847	0.876
Mean of dependent variable	10.24	10.24	10.24	10.28	7.43
Panel B					
Dependent variable:	Anti immigration vote share (top-decile)				
3G+	-1.381***				
	(0.409)				
before 2010 × 3G+		-1.601***	-2.363***	-1.604***	-1.334***
		(0.423)	(0.520)	(0.422)	(0.351)
after 2010 × 3G+		2.304**	2.713**	2.297**	1.226**
		(1.064)	(1.298)	(1.064)	(0.517)
Observations	6,270	6,270	6,270	6,270	8,531
R-squared	0.869	0.869	0.886	0.869	0.888
Mean of dependent variable	9.82	9.82	9.82	9.82	7.24
Panel C					
Dependent variable:	Traditionalist vote share (top-decile)				
3G+	-1.500***				
	(0.406)				
before 2010 × 3G+		-1.623***	-2.452***	-1.625***	-1.364***
		(0.418)	(0.511)	(0.416)	(0.347)
after 2010 × 3G+		1.646*	1.269**	1.644*	0.340
		(0.861)	(0.616)	(0.862)	(0.223)
Observations	4,272	4,272	4,272	4,238	8,520
R-squared	0.840	0.840	0.867	0.839	0.916
Mean of dependent variable	10.98	10.98	10.98	11.04	5.62
Constituency FE	✓	✓	✓	✓	✓
Country × Year FE	✓	✓	✓	✓	✓
Population			✓		
Income (Nightlight)				✓	
Including non-contested constit.					✓

Notes: The table presents the effect of $3G^+$ on the share of voters obtained by given political bloc in each panel. OLS estimates. The unit of analysis is constituency × election. In column (1) we only control for constituency's and election fixed effect. In column (2), we switch to the time varying model. In columns (3) and (4) we respectively control for population density and wealth. Column (5) impute a null vote share to parties that do not run in a given constituency. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Diff. in Diff. - effect of broadband mobile internet political blocs voteshares (continued)

Panel D					
Dependent variable:	Right populist vote share				
3G+	-0.949***				
	(0.268)				
before 2010 × 3G+		-1.166***	-2.116***	-1.169***	-0.987***
		(0.268)	(0.359)	(0.269)	(0.226)
after 2010 × 3G+		2.682***	2.918**	2.676***	2.199*
		(1.000)	(1.249)	(1.002)	(1.142)
Observations	6,615	6,615	6,615	6,599	8,582
R-squared	0.886	0.886	0.900	0.886	0.885
Mean of dependent variable	9.90	9.90	9.90	9.92	7.66
Panel E					
Dependent variable:	Anti elite vote share (top-decile)				
3G+	-1.695***				
	(0.305)				
before 2010 × 3G+		-1.881***	-2.272***	-1.889***	-1.495***
		(0.305)	(0.405)	(0.305)	(0.251)
after 2010 × 3G+		1.635	1.572	1.594	0.238
		(1.383)	(1.665)	(1.371)	(0.739)
Observations	6,038	6,038	6,038	6,038	8,537
R-squared	0.888	0.888	0.895	0.888	0.898
Mean of dependent variable	10.78	10.78	10.78	10.78	7.69
Panel F					
Dependent variable:	Pro rural vote share (top-decile)				
3G+	0.057				
	(0.210)				
before 2010 × 3G+		-0.008	-0.043	-0.011	-0.523*
		(0.203)	(0.240)	(0.205)	(0.273)
after 2010 × 3G+		4.893***	5.464***	4.905***	-0.070
		(1.416)	(1.770)	(1.415)	(0.207)
Observations	1,513	1,513	1,513	1,495	8,434
R-squared	0.988	0.988	0.990	0.988	0.968
Mean of dependent variable	9.19	9.19	9.19	9.24	1.73
Constituency FE	✓	✓	✓	✓	✓
Country × Year FE	✓	✓	✓	✓	✓
Population			✓		
Income (Nightlight)				✓	
Including non-contested constit.					✓

Notes: The table presents the effect of $3G^+$ on the share of voters obtained by given political bloc in each panel. OLS estimates. The unit of analysis is constituency × election. In column (1) we only control for constituency's and election fixed effect. In column (2), we switch to the time varying model. In columns (3) and (4) we respectively control for population density and wealth. Column (5) impute a null vote share to parties that do not run in a given constituency. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Calibration of random forests

	coeff	s.e.	t-stat	p-value	N
Data:		Survey - CHES			
mean pred.	0.99	0.35	2.80	0.00	38362
diff pred.	1.24	0.33	3.72	0.00	
Data:		Manifestos - CMP			
mean pred.	0.93	0.33	2.80	0.00	38362
diff pred.	2.19	0.23	9.63	0.00	

Notes: Following Athey and Wager (2019) and Chernozhukov et al. (2018), the CATE is fitted as a linear function of the the out-of-bag forest estimate. The mean prediction absorbs the average treatment effect, and the coefficient of the differential prediction can be interpreted as a measure of the quality of the estimates of treatment heterogeneity. Practically, if the coefficient is 1 then the treatment heterogeneity estimates are well calibrated; if it is significant and positive then at least it could be evidence of a useful association between the estimate and its corresponding out-of-bag prediction, and be interpreted as the forest succeeding in finding heterogeneity. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Best linear projection of the CATE against the political platform - Survey-based ideology

Dependent variable:	Conditional Average Treatment Effect				
	(1)	(2)	(3)	(4)	(5)
Left to Right - economic	0.278	0.021	-0.891**	-0.853*	-1.083
	(0.255)	(0.388)	(0.454)	(0.454)	(0.798)
Pro Rural	0.956**	1.280**	0.874	0.950*	1.362
	(0.459)	(0.543)	(0.553)	(0.553)	(0.846)
Anti EU	0.735***	1.195***	1.295***	1.342***	1.662**
	(0.281)	(0.411)	(0.478)	(0.478)	(0.699)
Traditionalism	-0.870**	-1.391***	-1.467**	-1.478**	-2.477**
	(0.352)	(0.480)	(0.587)	(0.587)	(1.124)
Anti Immigration		-0.231	-0.696	-0.756	-2.886***
		(0.455)	(0.651)	(0.651)	(1.023)
Anti Environment		0.917**	1.852***	1.887***	5.788***
		(0.433)	(0.621)	(0.620)	(1.191)
Law & Order		0.015	0.345	0.263	1.028
		(0.649)	(0.601)	(0.601)	(1.029)
Anti-elite		-0.387	-0.341	-0.352	0.442
		(0.247)	(0.329)	(0.330)	(0.571)
Observations	38,362	38,362	38,362	38,362	38,362
Mean of dependent variable	0.14	0.14	0.14	0.15	0.14
SD of dependent variable	0.39	0.39	0.39	0.51	0.39
Fitted model	Survey	Survey	Survey	Manifesto	Survey
Political family FE			✓	✓	✓
Weights					✓

Notes: The table presents the best-linear projection estimator (Semenova and Chernozhukov, 2021) of CATEs regressed on survey-based ideology variables. The unit of analysis is constituency × election × party. Standard errors are clustered at the constituency level. In column (1), we include only the most important variables (those above the average variable importance) as control variables. In column (2), we additionally control for the other ideological dimensions. Columns (3) includes political family fixed effects, as defined in the CHES data. Column (4) is similar to Column (3) but the CATE was fitted on the manifesto data. Finally, column (5) weights the observations by the average vote share a given party obtained in the data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Best linear projection of the CATE against political platform - Manifesto-based ideology

Dependent variable:	Conditional Average Treatment Effect				
	(1)	(2)	(3)	(4)	(5)
Favourable references to the middle class	-50.090***	-53.028***	-46.639***	-50.126***	-62.095***
	(12.333)	(13.579)	(12.711)	(12.723)	(15.512)
State funding of cultural and leisure facilities	-23.857***	-24.250***	-23.709***	-24.576***	-43.629***
	(4.450)	(5.389)	(6.336)	(6.341)	(8.709)
Supply side oriented economic policies	19.351***	22.555***	23.462***	23.290***	36.766***
	(4.841)	(5.835)	(6.202)	(6.197)	(9.001)
Support for the EU	-16.312***	-17.268***	-17.119***	-17.110***	-27.061***
	(4.993)	(5.402)	(5.301)	(5.304)	(6.290)
Market regulation	8.311**	9.456**	9.497**	8.459**	5.946
	(3.499)	(4.038)	(4.296)	(4.299)	(5.171)
Support for argiculture and farmers	12.406**	15.161***	17.094***	16.763***	31.388***
	(5.490)	(5.862)	(6.360)	(6.362)	(9.054)
Civic mindedness	-9.038	-11.481*	-14.384**	-15.158**	-17.825*
	(6.047)	(6.249)	(6.927)	(6.939)	(9.658)
State funding of education	15.303***	18.732***	17.888***	19.085***	23.976***
	(4.037)	(4.429)	(5.643)	(5.656)	(8.322)
Observations	30,812	30,812	30,812	30,812	30,812
Mean of dependent variable	0.18	0.18	0.18	0.15	0.18
SD of dependent variable	0.57	0.57	0.57	0.43	0.57
Fitted model	Manifesto	Manifesto	Manifesto	Survey	Manifesto
Other dimensions		✓	✓	✓	✓
Political family FE			✓	✓	✓
Weights					✓

Notes: The table presents the best-linear projection estimator (Semenova and Chernozhukov, 2021) of CATEs regressed on manifesto-based ideology variables. The unit of analysis is constituency × election × party. Standard errors are clustered at the constituency level. In column (1), we include only the most important variables (those above the average variable importance) as control variables. In column (2), we additionally control for the other ideological dimensions. Columns (3) includes political family fixed effects, as defined in the CHES data. Column (4) is similar to Column (3) but the CATE was fitted on the survey data. Finally, column (5) weights the observations by the average vote share a given party obtained in the data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Best linear projection of the CATE against the political platform - Mixed model

Dependent variable:	Conditional Average Treatment Effect			
	(1)	(2)	(3)	(4)
Favourable references to the middle class	-52.455***	-49.386***	-79.955***	-74.833***
	(13.053)	(13.043)	(17.468)	(17.462)
State funding of cultural and leisure facilities	-23.027***	-22.187***	-41.337***	-40.005***
	(5.613)	(5.612)	(7.089)	(7.096)
Supply side oriented economic policies	20.248***	20.454***	29.049***	29.653***
	(5.388)	(5.391)	(6.837)	(6.843)
Support for the EU	-15.498***	-15.418***	-23.433***	-23.088***
	(5.665)	(5.665)	(6.578)	(6.580)
Market regulation	8.079**	8.907**	6.016	6.826
	(3.639)	(3.635)	(4.420)	(4.419)
Support for argiculture and farmers	15.361***	15.525***	25.670***	25.509***
	(5.840)	(5.846)	(8.463)	(8.464)
Civic mindedness	-11.661*	-10.474	-21.787*	-19.822*
	(6.910)	(6.904)	(11.255)	(11.253)
State funding of education	16.252***	15.164***	22.831***	21.129***
	(4.302)	(4.291)	(5.864)	(5.858)
Left to Right - economic	-1.501**	-1.499**	-1.496	-1.564
	(0.703)	(0.703)	(1.078)	(1.079)
Anti EU	1.084	1.133	1.829*	1.879*
	(0.741)	(0.740)	(1.061)	(1.061)
Anti Immigration	0.888	0.802	1.602	1.471
	(0.942)	(0.942)	(1.225)	(1.224)
Pro Rural	0.290	0.371	-0.900	-0.790
	(1.037)	(1.035)	(1.261)	(1.258)
Anti Environment	1.779*	1.875*	3.305**	3.486**
	(1.050)	(1.049)	(1.399)	(1.397)
Law & Order	-0.622	-0.685	-2.306	-2.309
	(1.195)	(1.195)	(1.518)	(1.516)
Anti-elite	-0.206	-0.221	-0.824	-0.849
	(0.579)	(0.579)	(0.926)	(0.926)
Traditionalism	-1.301	-1.394	0.178	-0.049
	(1.370)	(1.368)	(1.834)	(1.831)
Observations	30,812	30,812	30,812	30,812
Mean of dependent variable	0.15	0.18	0.15	0.18
SD of dependent variable	0.43	0.57	0.43	0.57
Fitted model	Survey	Manifesto	Survey	Manifesto
Political family FE	✓	✓	✓	✓
Weights			✓	✓

Notes: The table presents the best-linear projection estimator (Semenova and Chernozhukov, 2021) of CATEs regressed both on survey-based and manifesto-based ideology variables. The unit of analysis is constituency × election × party. Standard errors are clustered at the constituency level. Columns (1) and (2) control for political family fixed effects with the CATE from the survey-based and manifesto-based ideologies respectively. Columns (3) and (4) additionally weight the observations by the average vote share a given party obtained in the data. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

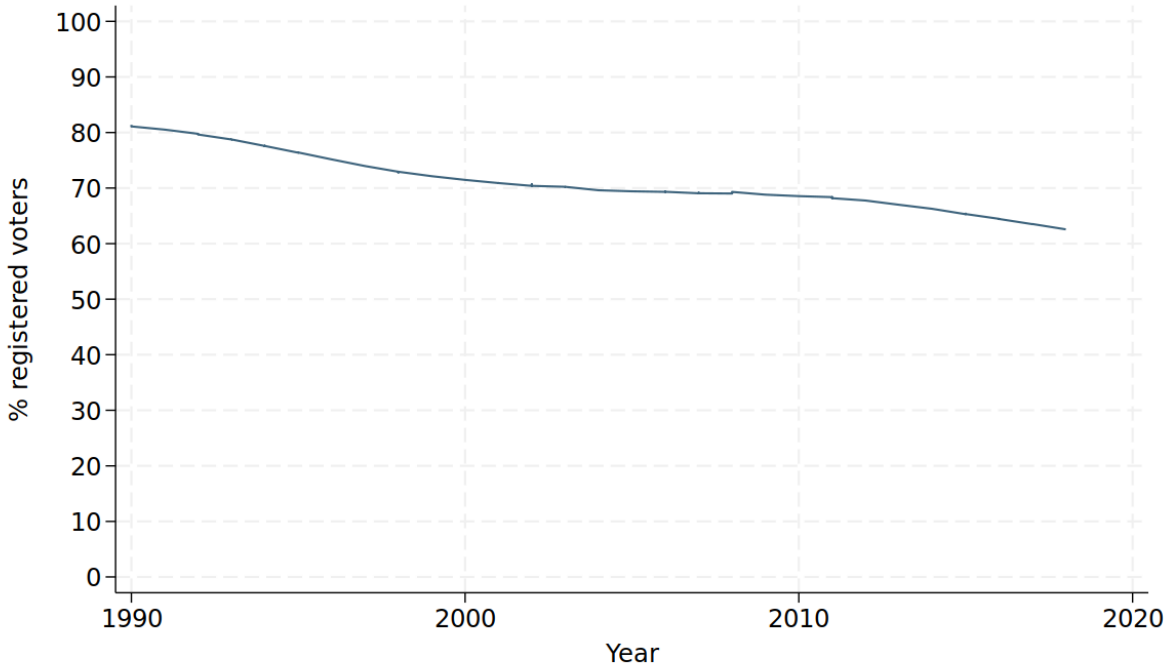
Table 7: Best linear projection of the CATE against the political platform - constituency level heterogeneity

Dependent variable:	Conditional Average Treatment Effect					
	(1)	(2)	(3)	(4)	(5)	(6)
State funding of cultural and leisure facilities	-28.415***	-19.183*	-17.540**	-32.447***	-22.101**	-26.039***
	(6.979)	(10.922)	(7.722)	(10.457)	(9.124)	(8.911)
Market regulation	11.215*	9.523	2.148	16.395**	2.481	15.589**
	(5.873)	(6.032)	(5.617)	(6.382)	(5.869)	(6.055)
Favourable references to the middle class	-50.734***	-54.498***	-53.590***	-33.971	-55.460***	-37.682*
	(17.399)	(19.509)	(16.915)	(23.029)	(18.594)	(20.557)
Support for the EU	-15.170**	-23.026***	-19.815***	-15.456*	-22.779***	-10.859
	(6.874)	(8.569)	(7.127)	(8.450)	(7.651)	(7.722)
Supply side oriented economic policies	23.757***	22.098**	24.348***	22.482*	26.357***	18.391*
	(7.819)	(10.763)	(7.376)	(12.303)	(8.451)	(10.098)
State funding of education	16.338***	21.769**	10.278	26.980***	13.571*	21.941***
	(6.045)	(10.413)	(6.659)	(10.421)	(7.975)	(8.409)
Support for argiculture and farmers	18.990**	10.242	20.417***	13.471	26.743***	8.225
	(8.422)	(10.659)	(7.830)	(11.715)	(9.230)	(9.446)
Civic mindedness	-12.188	-17.150	-14.258*	-10.662	-25.482**	-2.780
	(7.559)	(13.652)	(7.996)	(13.938)	(10.068)	(10.378)
Anti Environment	0.894	-0.978	-0.116	0.956	-0.495	1.420*
	(0.813)	(0.854)	(0.711)	(1.099)	(0.854)	(0.839)
Observations	15,351	15,461	15,479	15,058	15,232	15,305
Mean of dependent variable	0.16	0.19	0.17	0.19	0.17	0.19
SD of dependent variable	0.47	0.65	0.53	0.61	0.57	0.57
Fitted model	Manifesto	Manifesto	Manifesto	Manifesto	Manifesto	Manifesto
Political family FE	✓	✓	✓	✓	✓	✓
Other dimensions	✓	✓	✓	✓	✓	✓
Sample	Sparse	Dense	Poor	Rich	Poor (rel.)	Rich (rel.)

Notes: The table presents the best-linear projection estimator (Semenova and Chernozhukov, 2021) of CATEs regressed on manifesto-based ideology variables for specific types of constituencies. The unit of analysis is constituency × election × party. Standard errors are clustered at the constituency level. Using the same specification as Table 5 column (3), the sample is now split between constituencies which are below and above median regarding the two constituency characteristics. Columns (1) and (2) display the results for the sparsely vs. densely populated constituencies, columns (3) and (4) compare poor and rich places, columns (5) and (6) compare poor and rich places but now the median is defined at the national level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

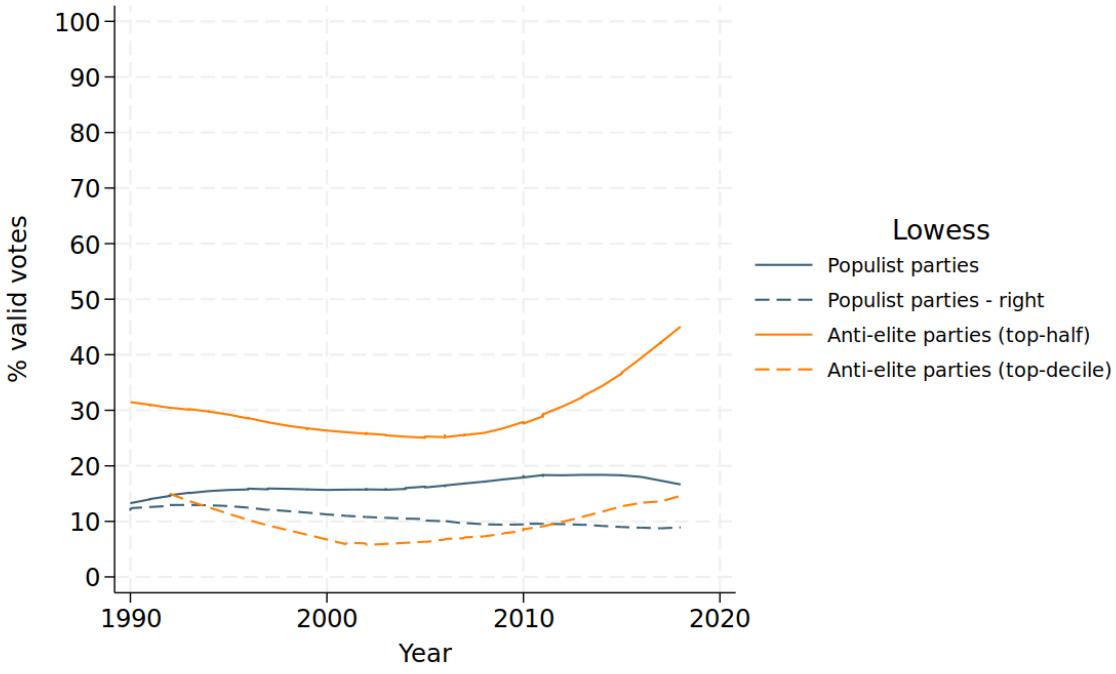
A Online appendix

Figure A1: Turnout over time



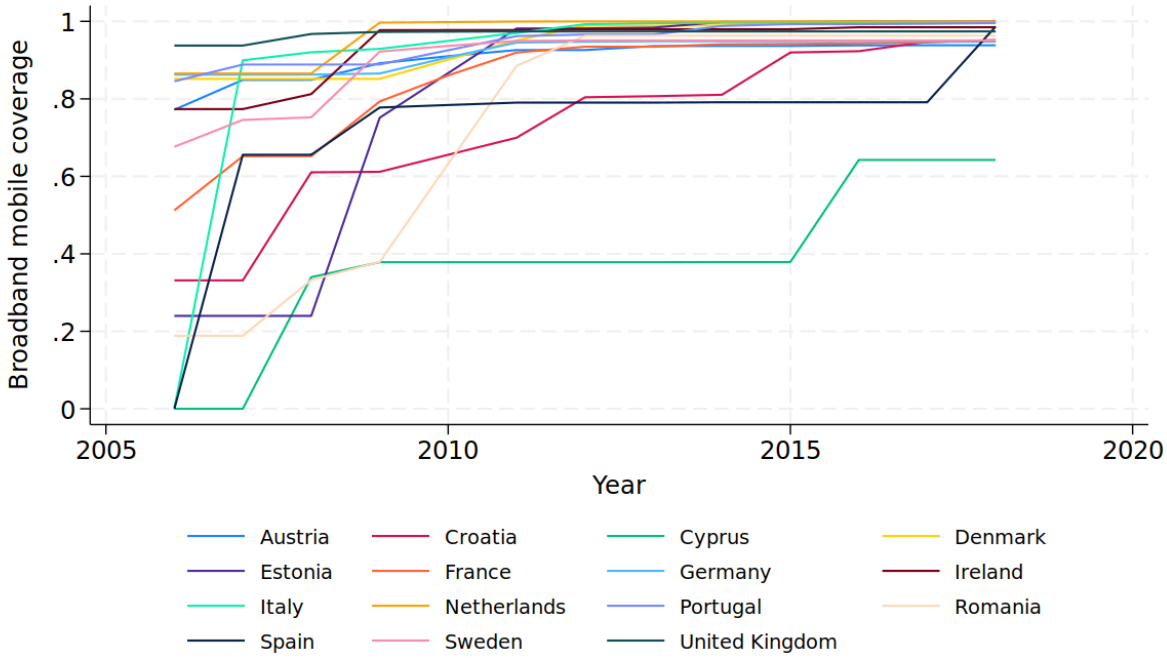
Source: Collins Bartholomew's Mobile Coverage Explorer

Figure A2: Populist and anti-elite vote over time



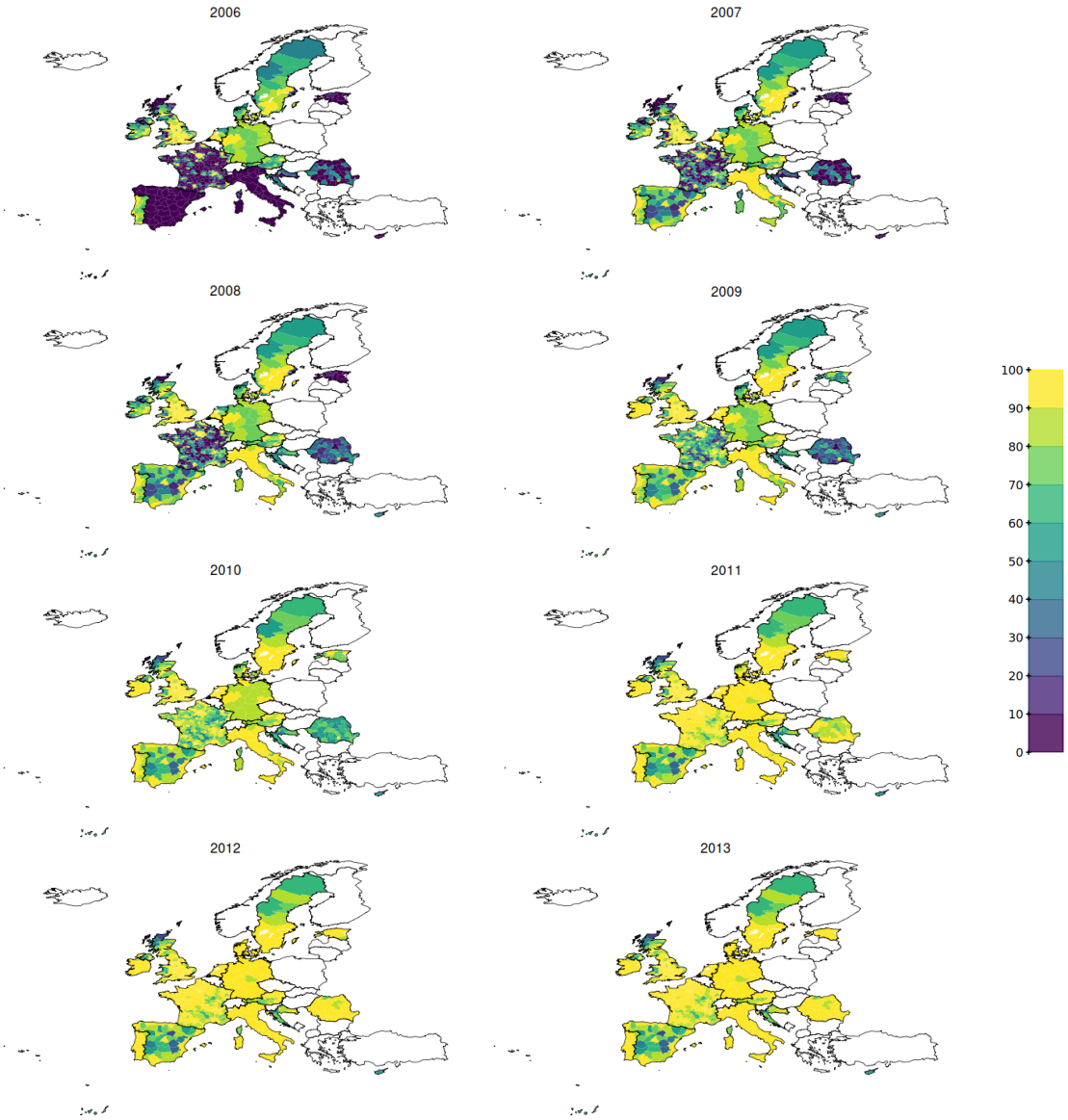
Source: CLEA election archive project

Figure A3: Broadband mobile internet deployment



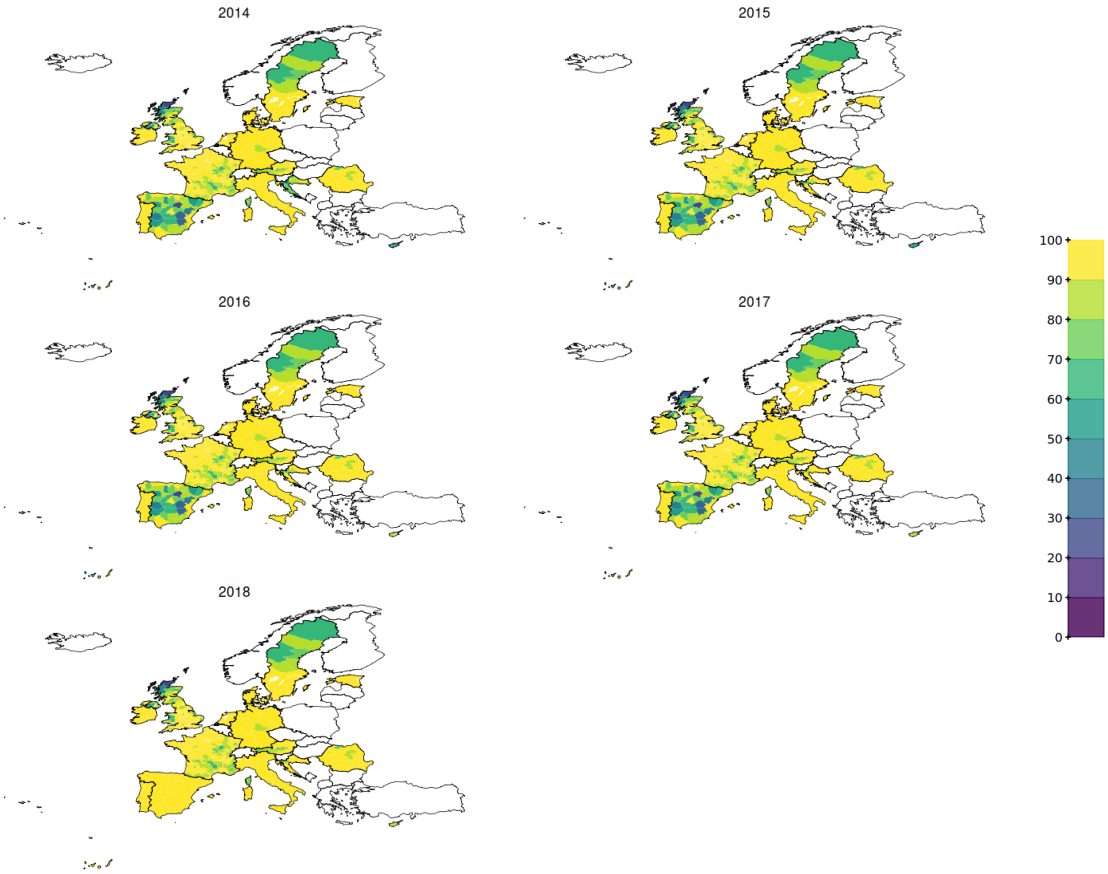
Source: CLEA election archive project

Figure A4: Maps of broadband mobile internet deployment



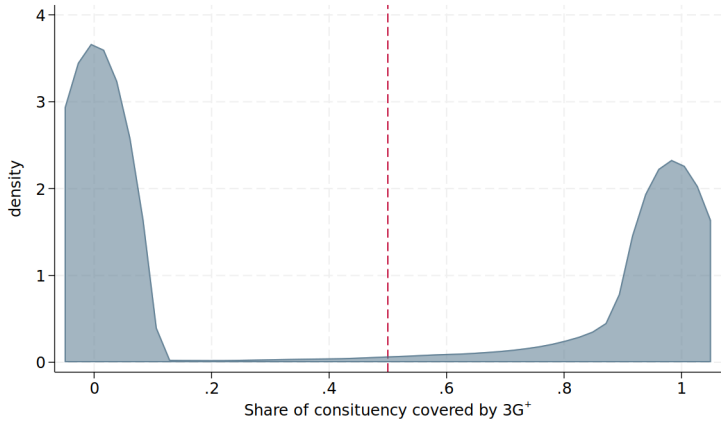
Source: Collins Bartholomew's Mobile Coverage Explorer

Figure A5: Maps of broadband mobile internet deployment (continued)



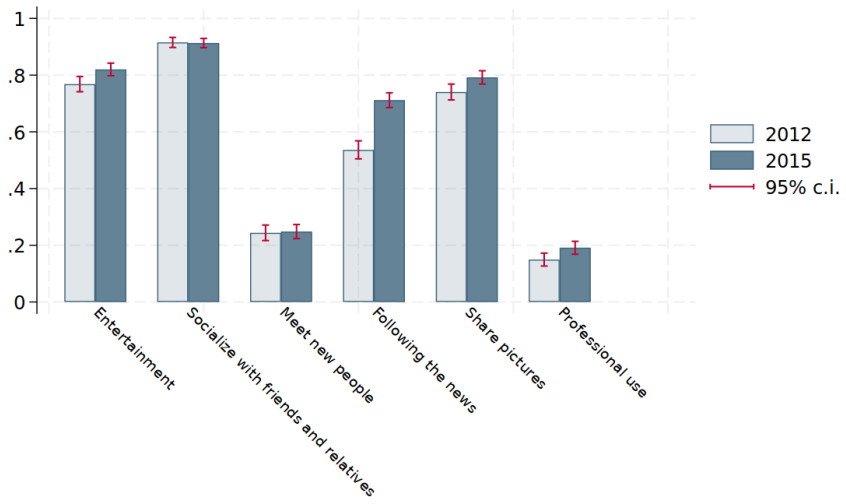
Source: Collins Bartholomew's Mobile Coverage Explorer

Figure A6: Distribution of the treatment continuous measure



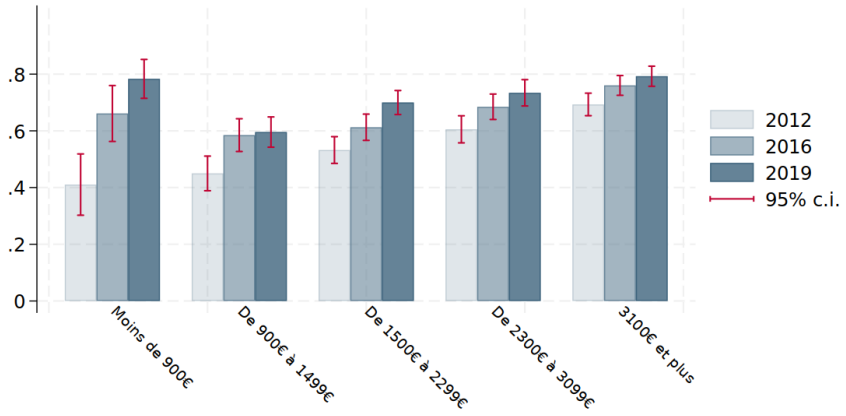
Source: Collins Bartholomew's Mobile Coverage Explorer for $3G+$ and Gridded Population of the World (GPW) data for the weights.

Figure A7: Use of social media (France)



Source: ARCEP - *"Baromètre du numérique"*

Figure A8: Propensity to follow the news online over the past 12 month, depending on income bracket (France)



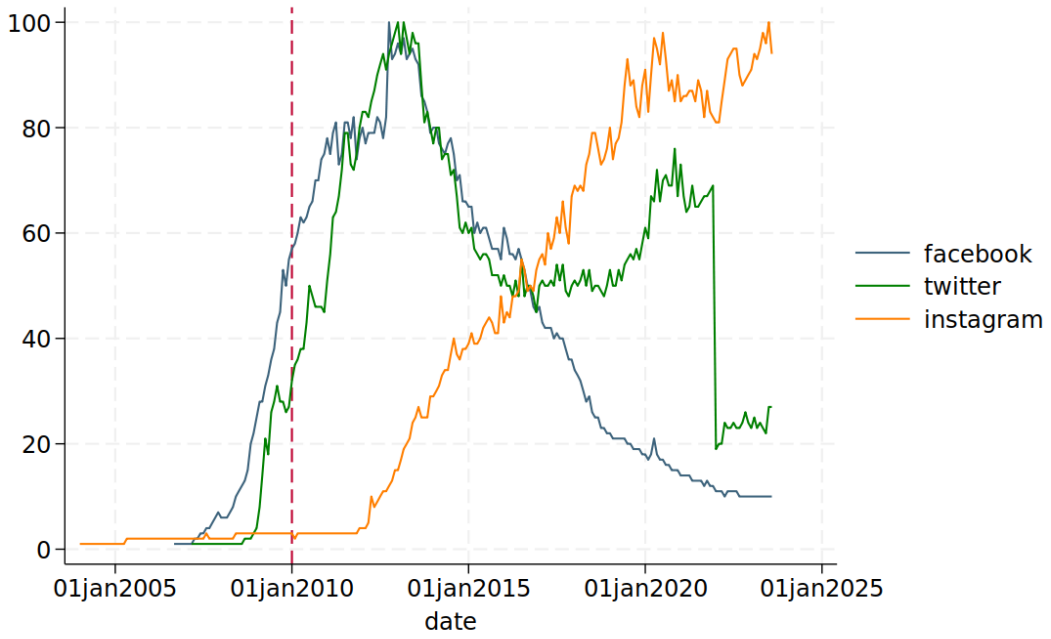
Source: ARCEP - *"Baromètre du numérique"*

Table A1: Summary statistics of parties political platforms - CHES

	Mean	Median	SD	Min	Max	N
Left to Right - economic	4.82	2.15	0.21	4.80	9.33	679
Anti EU	5.14	1.66	3.00	4.42	8.97	679
Anti Immigration	5.17	2.39	0.50	4.92	10.00	618
Pro Rural	4.76	1.62	1.25	4.57	10.00	614
Anti Environment	4.85	2.02	0.29	5.00	9.31	599
Law & Order	5.08	2.38	0.80	4.90	9.92	618
Anti-elite	4.07	2.63	0.77	3.15	10.00	554
Traditionalism	4.97	2.47	0.29	4.53	9.82	677

Notes: all dimensions are represented along a 0 to 10 scale.

Figure A9: Google trends of social media



Source: Google Trends

B Diff-in-diff approach

In this section we detail the results of the diff-in-diff approach presented in Section 4.1. After showing that if anything $3G^+$ decreased turnout, we inspect to which extent our data replicate results from Manacorda et al. (2023) and Guriev et al. (2021). We finally show the results for candidates specifically supporting the rural world, as the relevance of this dimension was raised by our analysis.

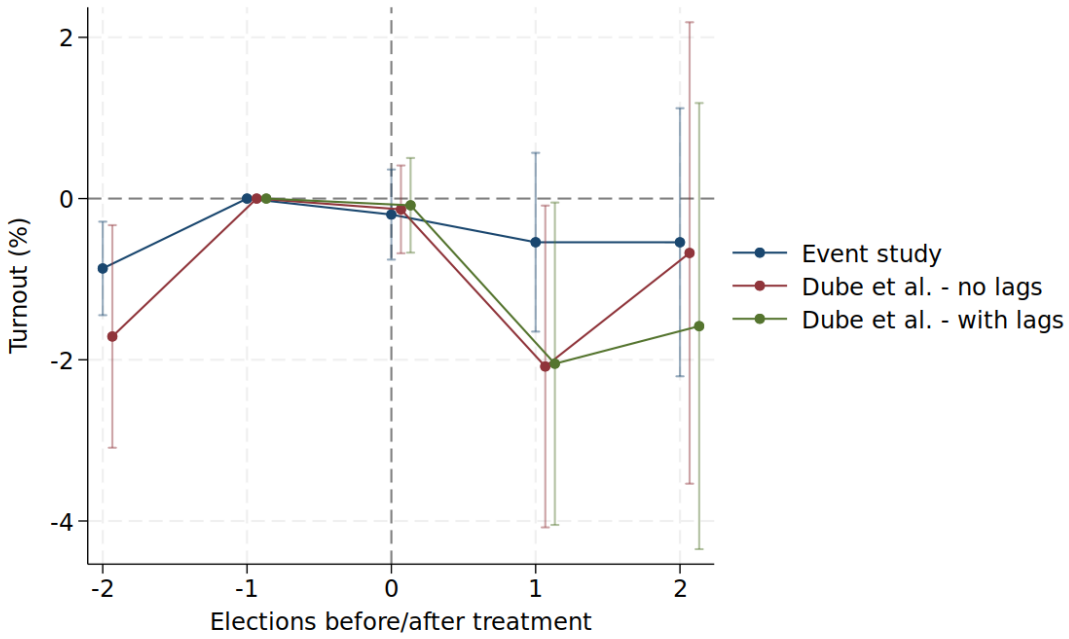
In our estimates, additionally to equation 1, we consider a model with time varying treatment effect. Indeed besides the increased data traffic allowed by $3G^+$ itself, another significant transformation has affected internet browsing during our time period: the emergence of social media. Although these two phenomena are not entirely independent, one can argue that the new interactions introduced by social media might induce different effects compared to the previous content found online (Zhuravskaya et al., 2020), even once mobile broadband Internet was introduced. Therefore, we also consider the following specification:[26]

$$Vote_{it} = (\beta^{early} \cdot \mathbb{1}_{t<2010} + \beta^{late} \cdot \mathbb{1}_{t\geq 2010}) \cdot 3G^+_{it} + \alpha_i + \tau_{country(i)t} + \gamma \cdot X_{it} + \epsilon_{it} \quad (4)$$

We also consider three alternative estimators: (i) a simple event study where the effect is estimated as an interaction between the treatment status and the number of elections since or

[26]The choice of 2010 as a threshold is significant because it coincides with the year when Facebook reached a milestone by achieving half of its highest level of user activity or engagement (Figure A9).

Figure A10: Event-study: effect of broadband mobile internet on turnout



Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

before the constituency is treated; (ii) the corresponding staggered diff-in-diff estimator suggested by Dube et al. (2022) and (iii) the same estimator as (ii) but now controlling for lagged dynamic of the outcomes.

In Table A4, we study whether mobile broadband Internet affected turnout. In column (1) we only control for consituency's and election fixed effect. In column (2), we switch to the time varying model. In columns (3) and (4) we respectively control for population density and wealth. None of the coefficients, whatever our specification is, is significantly different from zero. Looking at the alternate estimators in Figure A10, we see a negative effect for the first lag of the treatment, significant only for the Dube et al. (2022) estimator.

Tables 1, 2 and A5 have the same specification as Table A4 from column (1) to (4); and column (5) impute a null vote share to parties that do not run in a given constituency.

In Table 1 Panels A to C and Table A5 panel A, we examine whether our data replicates the findings from Manacorda et al. (2023). We focus on candidates in the top decile along different dimensions, as detailed in Panels A to C. These panels correspond respectively to candidates most eurosceptic, hostile toward immigration, traditionalist, and economically right-wing. In comparison to Manacorda et al. (2023), we find a positive and significant effect for each of these groups only in the post-2010 period. We show our alternative estimators in Figure A11 and A12 and find positive but not always significant effect for the first lag of the treatment, which will mechanically consider

observations later in time.

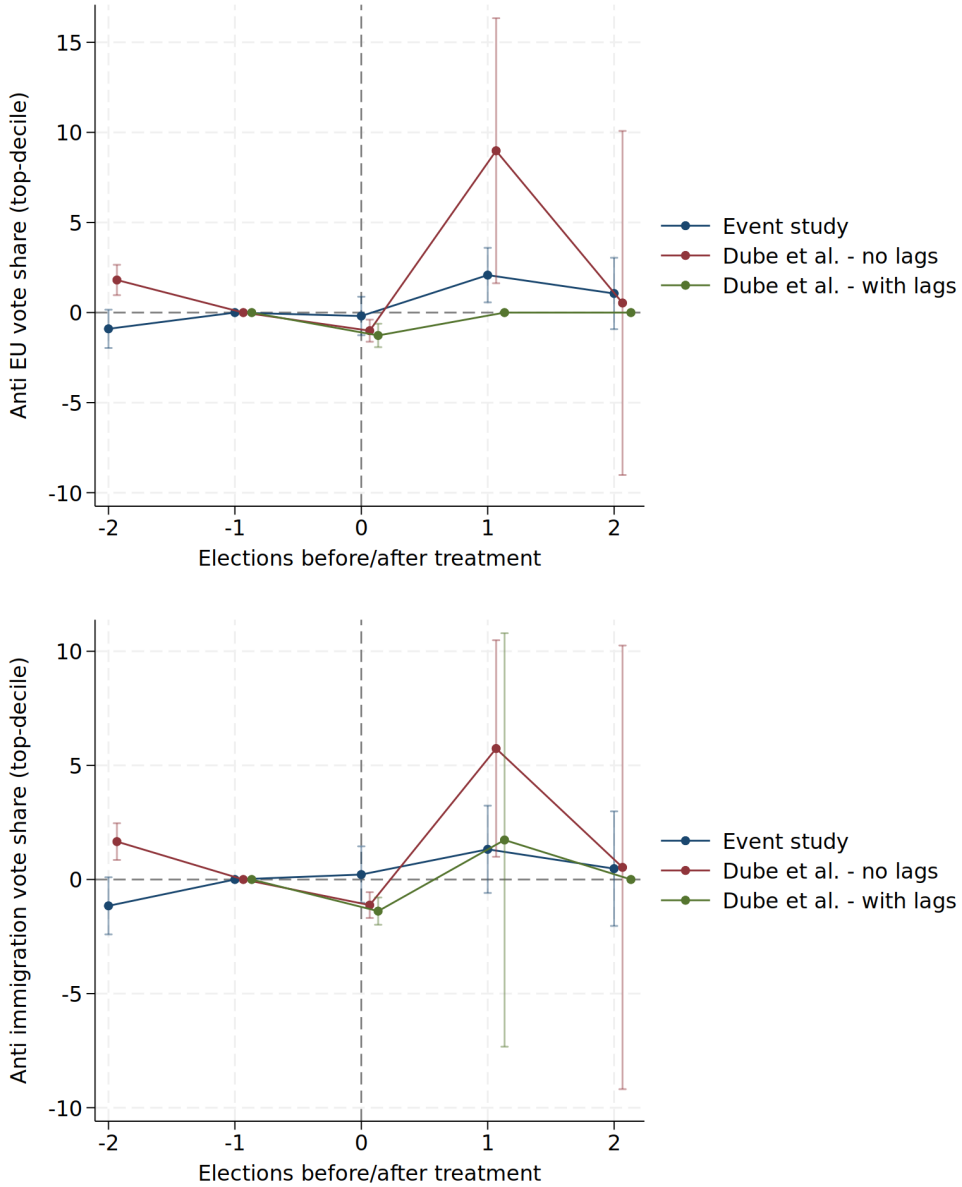
One plausible explanation for this discrepancy with Manacorda et al. (2023) is that that the late period is also the one during which the randomness of the treatment is also higher (the early period being a time when constituencies gaining access to broadband had the highest wealth and population); and they utilize an instrument for their estimation and therefore are able to have unbiased estimates for the pre-2010 period. Therefore, it is here hard to disentangle between proper identification and time varying effects.

In Panels D and E of Table 2 and Table A5 Panels B and C, we investigate the impact of mobile broadband Internet on the populist vote using different definitions. Panel A examines the vote for any candidate classified as populist according to the Guriev et al. (2021) classification that we were able to match in our sample. Panel B focuses exclusively on right-wing candidates of the latter set. To further narrow our analysis, we assess the influence of the anti-elite discourse of the candidates. Specifically, we consider the top half and top decile of candidates in terms of their anti-elite rhetoric in Panels C and D, respectively. Across these different definitions, we generally observe a negative effect on populist votes. The event and staggered estimators also suggest no effect on general populist vote (Figure A13). However, in the late period, we do find a positive and statistically significant effect for right-wing populist votes in Table 2 Panel D, and is positive and significant for the first lag of the treatment in two out of three estimators in Figure A13.

When we focus solely on candidates employing anti-elite rhetoric, the effect remains negative both in Table 2 panel D and Figure A14 except for the first lag in the top decile of anti elite candidates. It could be important to consider that lower chamber elections might not be the context in which populist candidates perform best. Contemporary populism often revolves around a charismatic leader and a stance against intermediary institutions. This might explain the absence of stronger positive effects on populist votes in our analysis.

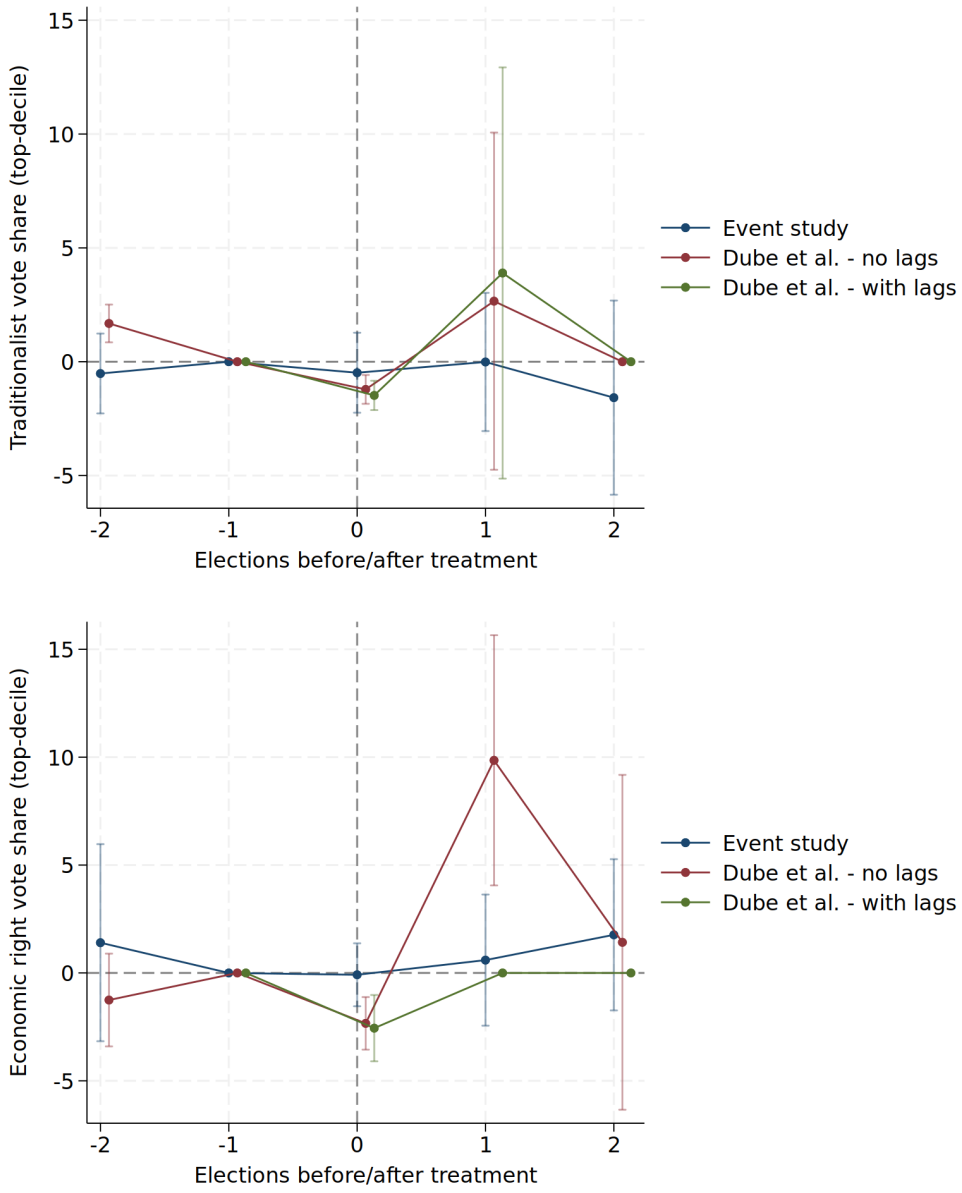
In Table 2 panel F and Table A5 panel D, we study the effect of mobile broadband Internet on the vote for candidate supportive of the rural population, based again on the top half or top decile respectively of the candidates alignment in panel A and B. Consistent with the GRF analysis in Section 5.2.2 we see a positive and significant correlation between $3G^+$ and the pro rural candidates vote share. This is consitent with other estimators shown in Figure A15.

Figure A11: Event-study: effect of broadband mobile internet on communitarian vote



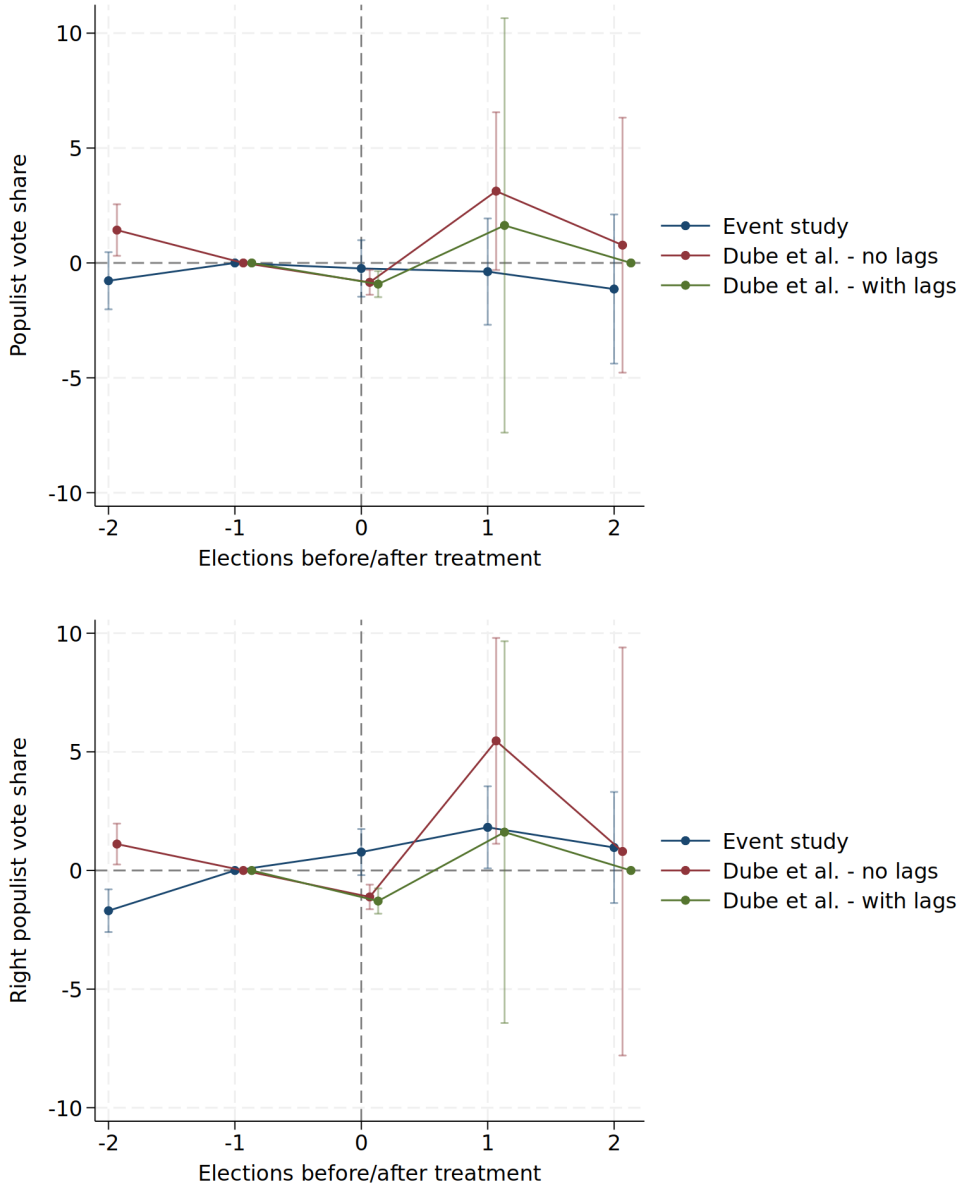
Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure A12: Event-study: effect of broadband mobile internet on communitarian vote (continued)



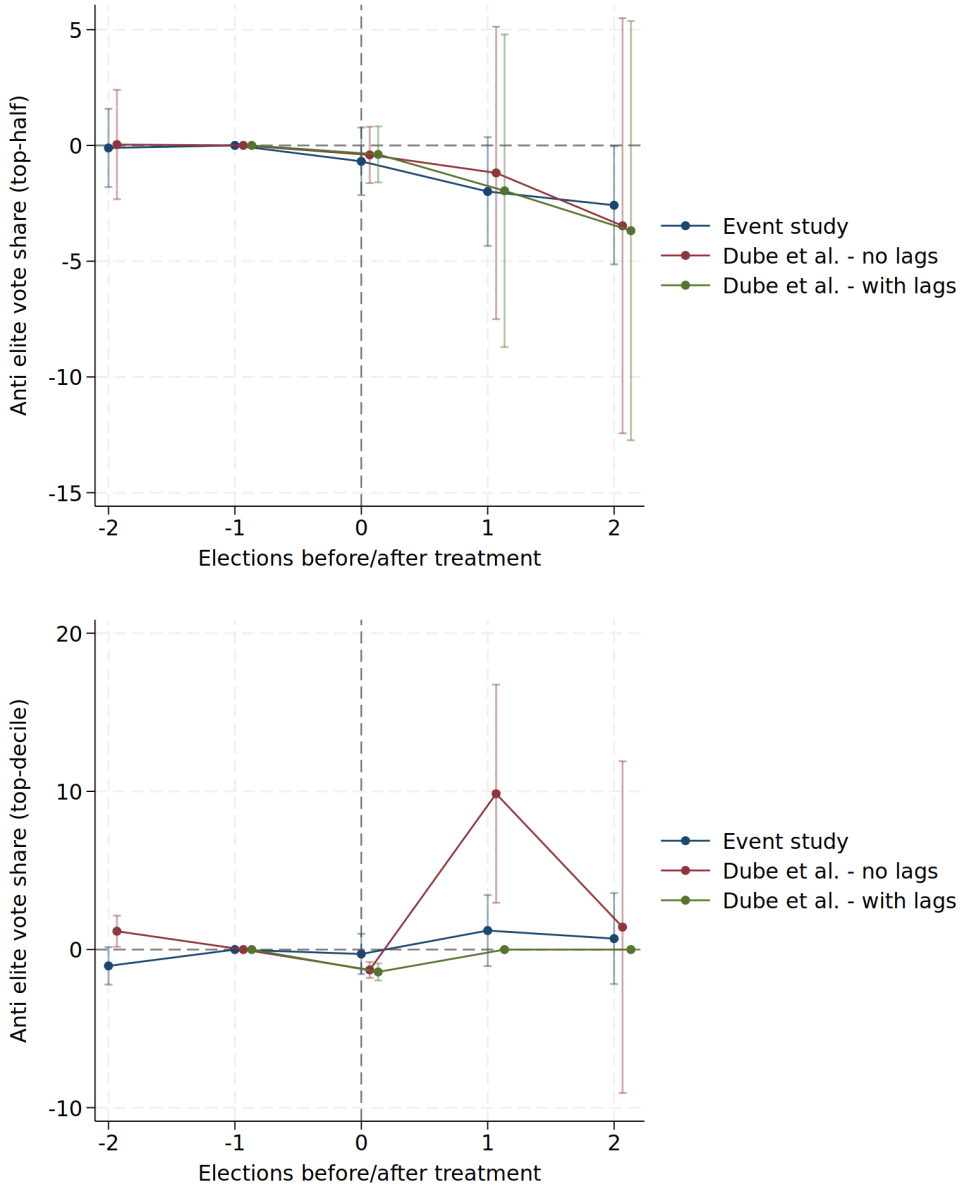
Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure A13: Event-study: effect of broadband mobile internet on populist vote



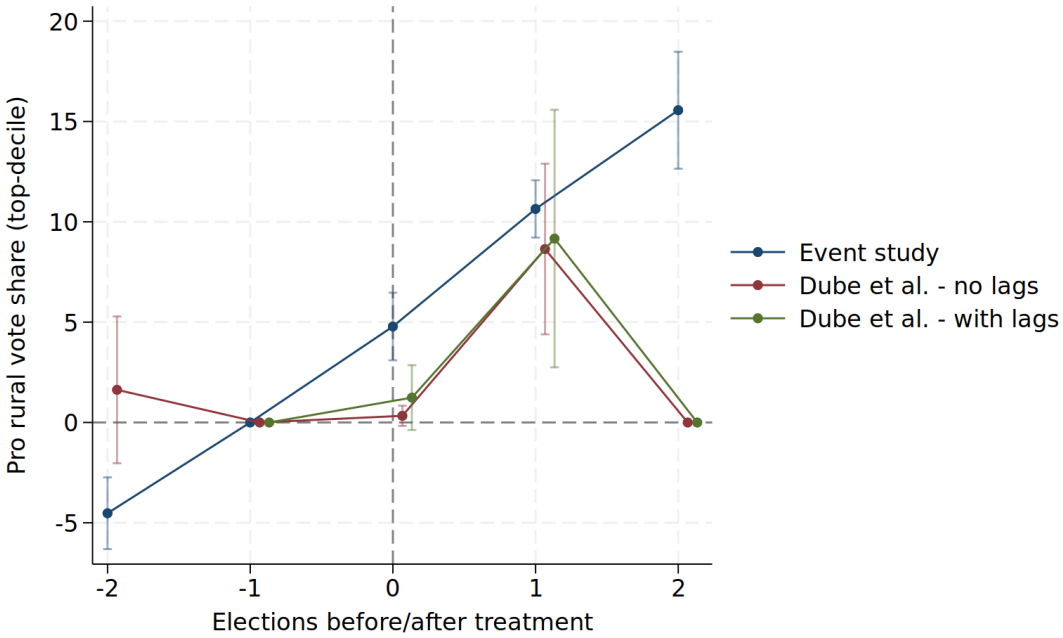
Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure A14: Event-study: effect of broadband mobile internet on anti elite vote



Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Figure A15: Event-study: effect of broadband mobile internet on pro rural vote



Notes: The Figure presents the event study and the staggered diff-in-diff estimators of Dube et al. (2022) with and without lags of the outcomes. The unit of analysis is constituency × election. Point estimates and 90% confidence intervals are presented. Standard errors are clustered at the constituency level.

Table A2: Summary statistics of parties political platforms - CMP

	Mean	Median	SD	Min	Max	N
foreign special +	0.30	0.59	0.00	0.00	4.72	506
foreign special -	0.04	0.19	0.00	0.00	2.14	506
anti-imperalism	0.09	0.30	0.00	0.00	2.71	506
military +	1.12	1.61	0.00	0.44	11.82	506
military -	0.47	0.86	0.00	0.05	6.84	506
peace	0.45	0.71	0.00	0.21	6.29	506
internationalism +	2.48	1.90	0.00	2.20	10.14	506
europe +	2.25	2.33	0.00	1.73	15.75	506
internationalism -	0.20	0.67	0.00	0.00	7.38	506
europe -	0.88	2.41	0.00	0.00	33.89	506
freedom & human rights	2.50	2.98	0.00	1.55	27.59	506
democracy	2.89	2.61	0.00	2.29	19.44	506
constitution +	0.33	0.78	0.00	0.00	8.05	506
constitution -	0.18	0.72	0.00	0.00	9.36	506
decentralisation	3.37	5.20	0.00	1.70	43.10	506
centralisation	0.12	0.41	0.00	0.00	5.56	506
gov-admin efficiency	4.53	4.31	0.00	3.47	25.96	506
political corruption	0.74	1.60	0.00	0.15	21.11	506
political authority	3.25	4.58	0.00	1.78	31.27	506
free market economy	1.52	2.61	0.00	0.55	25.35	506
incentives	2.73	2.60	0.00	2.09	16.19	506
market regulation	2.65	2.68	0.00	2.04	23.23	506
economic planning	0.47	1.14	0.00	0.00	9.52	506
corporatism/mixed economy	0.21	0.39	0.00	0.00	2.66	506
protectionism +	0.23	1.06	0.00	0.00	19.17	506
protectionism -	0.17	0.48	0.00	0.00	7.21	506
economic goals	2.19	2.35	0.00	1.37	10.08	506
keynesian demand management	0.30	1.10	0.00	0.00	11.79	506
economic growth +	1.85	3.04	0.00	0.95	41.48	506
technology & infrastructure	5.76	3.79	0.00	5.47	20.59	506
controlled economy	0.65	1.45	0.00	0.00	12.24	506
nationalisation	0.42	0.98	0.00	0.00	8.45	506
economic orthodoxy	1.73	2.78	0.00	0.80	20.00	506
marxist analysis +	0.13	1.48	0.00	0.00	32.39	506
anti-growth economy +	1.45	2.61	0.00	0.26	19.65	506
environmentalism +	6.48	7.22	0.00	4.65	62.11	506
culture +	2.79	2.53	0.00	2.34	17.32	506
equality +	5.36	4.34	0.00	4.44	26.56	506
welfare +	8.03	5.00	0.00	7.25	34.75	506
welfare -	0.51	1.08	0.00	0.00	9.95	506
education +	4.87	3.20	0.00	4.58	30.00	506
education -	0.04	0.16	0.00	0.00	2.21	506
national way of life +	1.64	3.85	0.00	0.37	57.33	506
national way of life -	0.22	0.64	0.00	0.00	4.13	506
traditional morality +	1.52	3.32	0.00	0.31	25.26	506
traditional morality -	0.39	0.83	0.00	0.00	6.94	506
law and order +	4.15	3.69	0.00	3.50	23.26	506
civic mindedness +	1.57	1.84	0.00	0.95	10.03	506
multiculturalism +	0.85	1.60	0.00	0.24	13.37	506
multiculturalism -	0.66	1.91	0.00	0.00	16.18	506
labour groups +	2.88	3.19	0.00	1.96	27.78	506
labour groups -	0.06	0.23	0.00	0.00	2.42	506
agriculture +	2.51	3.49	0.00	1.85	55.00	506
middle class and prof. groups	0.41	0.69	0.00	0.08	5.40	506
minority groups	1.32	1.74	0.00	0.73	11.60	506
non-economic demographic groups	2.62	2.76	0.00	1.85	20.39	506

Notes: the variables indicate the share (in %) of quasi-sentences in the respective category as a fraction of the overall number of allocated codes per document.

Table A3: Summary statistics

	Mean	Median	SD	Min	Max	N
OUTCOMES						
Turnout (%)	57.35	24.53	0.34	64.84	100.00	7,866
Populist vote share	11.88	10.28	0.00	9.27	65.88	7,251
Right populist vote share	9.85	7.92	0.00	8.04	62.03	6,678
Anti elite vote share (top-half)	64.74	34.13	0.33	66.13	100.00	8,568
Anti elite vote share (top-decile)	10.74	9.14	0.00	8.22	68.29	6,123
Economic right vote share (top-decile)	9.68	9.52	0.00	5.03	55.79	3,839
Anti EU vote share (top-decile)	10.19	7.86	0.00	8.45	71.76	6,228
Anti immigration vote share (top-decile)	9.76	7.26	0.00	8.25	70.27	6,342
Traditionalist vote share (top-decile)	10.43	6.98	0.00	9.49	70.27	4,600
Pro environment vote share (top-decile)	4.52	4.97	0.18	3.15	52.26	4,738
Pro rural vote share (top-decile)	8.65	12.57	0.00	1.80	67.15	1,691
TREATMENT						
3G (average)	0.44	0.00	0.48	0.00	1.00	9,058
3G (dummy)	0.46	0.00	0.50	0.00	1.00	9,058
CONTROLS						
Income [Nightlight-mean]	22.10	20.02	0.00	12.48	63.00	8,966
log(1+Population density)	5.89	1.78	0.84	5.68	10.49	9,058

Notes: Turnout is expressed in % of registered voters, all other outcomes are expressed in % of valid votes.

Table A4: Diff-in-Diff: effect of broadband mobile internet on turnout

Dependent variable:	Turnout (%)			
	(1)	(2)	(3)	(4)
3G+	-0.192			
	(0.150)			
before 2010 × 3G+		-0.103	-0.215	-0.100
		(0.135)	(0.169)	(0.136)
after 2010 × 3G+		-0.792	0.986	-0.799
		(0.657)	(0.827)	(0.654)
Observations	7,848	7,848	7,848	7,794
R-squared	0.992	0.992	0.992	0.992
Mean of dependent variable	57.35	57.35	57.35	57.21
Constituency FE	✓	✓	✓	✓
Country × Year FE	✓	✓	✓	✓
Population			✓	
Wealth				✓

Notes: The table presents the effect of $3G^+$ on turnout. OLS estimates. The unit of analysis is constituency × election. In column (1) we only control for constituency's and election fixed effect. In column (2), we switch to the time varying model. In columns (3) and (4) we respectively control for population density and wealth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Diff. in Diff. - additional results

Panel A Dependent variable:	Economic right vote share (top-decile)				
3G+	-0.256				
	(1.014)				
before 2010 × 3G+		-1.919***	-0.554	-2.017***	0.022
		(0.700)	(0.516)	(0.721)	(0.205)
after 2010 × 3G+		5.642***	6.703***	5.592***	-0.070
		(1.884)	(2.426)	(1.887)	(1.101)
Observations	3,370	3,370	3,370	3,316	8,595
R-squared	0.935	0.935	0.939	0.935	0.863
Mean of dependent variable	8.65	8.65	8.65	8.60	4.32
Constituency FE	✓	✓	✓	✓	✓
Country × Year FE	✓	✓	✓	✓	✓
Population			✓		
Income (Nightlight)				✓	
Including non-contested constit.					✓
Panel B					
Dependent variable:	Populist vote share				
3G+	-1.195***				
	(0.298)				
before 2010 × 3G+		-1.092***	-1.610***	-1.094***	-1.305***
		(0.310)	(0.430)	(0.310)	(0.249)
after 2010 × 3G+		-2.910*	-2.248	-2.910*	-0.548
		(1.641)	(1.893)	(1.639)	(1.675)
Observations	7,208	7,208	7,208	7,168	8,689
R-squared	0.911	0.911	0.918	0.911	0.897
Mean of dependent variable	11.92	11.92	11.92	11.95	9.90
Panel C					
Dependent variable:	Anti elite vote share (top-half)				
3G+	-1.138				
	(0.728)				
before 2010 × 3G+		-0.697	-0.749	-0.698	-0.708
		(0.763)	(0.881)	(0.763)	(0.755)
after 2010 × 3G+		-4.872***	-3.583*	-4.894***	-4.223***
		(1.775)	(1.987)	(1.777)	(1.547)
Observations	8,544	8,544	8,544	8,500	8,959
R-squared	0.949	0.950	0.950	0.949	0.956
Mean of dependent variable	64.84	64.84	64.84	65.10	61.84
Panel D					
Dependent variable:	Pro rural vote share (top-half)				
3G+	0.206				
	(0.787)				
before 2010 × 3G+		0.393	-0.964	0.398	0.290
		(0.850)	(1.035)	(0.851)	(0.852)
after 2010 × 3G+		-1.157	-0.422	-1.147	1.309
		(1.641)	(1.826)	(1.639)	(2.405)
Observations	8,926	8,926	8,926	8,866	9,001
R-squared	0.866	0.866	0.869	0.865	0.870
Mean of dependent variable	48.48	48.48	48.48	48.54	48.08

Notes: The table presents the effect of $3G^+$ on the share of voters obtained by given political bloc in each panel. OLS estimates. The unit of analysis is constituency × election. In column (1) we only control for constituency's and election fixed effect. In column (2), we switch to the time varying model. In columns (3) and (4) we respectively control for population density and wealth. Column (5) impute a null vote share to parties that do not run in a given constituency. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C GRF and multiparty elections econometrics

C.1 The issue

When studying electoral data in multi-party systems, which are prevalent in the vast majority of modern democracies, the decision of whether and how to group candidates becomes a crucial consideration for researchers studying these political regimes. Classifying political positions in relation to one another is a long-standing issue in political science. It has been a subject of research in its own right and is also an implicit concern in many empirical studies.

For instance the left-right axis dates back to seating arrangements in the French parliament after the French Revolution but political scientists have since suggested various representation of the political space, often with more than one dimension. One frequently used model with two dimensions is the Cartesian product of left vs. right economic policies along with authoritarian vs. libertarian values (Kitschelt, 1994; Heywood, 2021).

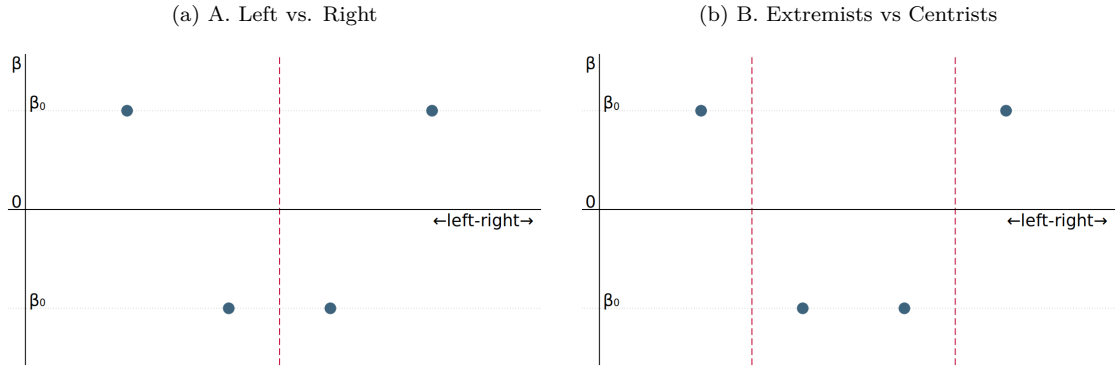
Our work is especially related to the literature focusing on econometrics techniques to estimate treatment effects on parties vote share. Two notable works were conducted by Katz and King (1999) and Tomz et al. (2002). The first built upon the analysis of compositional data (Aitchison, 1982) to allow the empirical modelling to satisfy core properties of electoral data. The latter offered a method less computationally demanding, more tractable model and easier to scale to a large number of parties. Our approach also relate the the empirical study of multi-issue politics in modern democracies. Notably, Roemer et al. (2007) use party-unanimity Nash equilibrium (Roemer and Roemer, 2009) to document the interplay between racism and redistribution issues in political competition.

We now present the intuition for two empirical issues we are addressing by applying the GRF (Athey et al., 2019) to study multiparty elections.

Consider a scenario in which a researcher aims to estimate the effects of a specific treatment in a political context featuring four parties ordered along the left-right axis. Let the true treatment effects for each party be denoted as $(\beta_1, \beta_2, \beta_3, \beta_4)$ and the order of the tuple corresponds to the party's positioning on the left-right axis. Also assume $(\beta_1, \beta_2, \beta_3, \beta_4) = (\beta_0, -\beta_0, -\beta_0, \beta_0)$, where $\beta_0 > 0$. The crucial point arises when the researcher chooses how to model the political space for estimating treatment effects. If they opts for a Left vs. Right approach, the estimated effects would be $(0, 0)$, while an Extremists vs. Centrist parties approach would yield estimated effects of $(2\beta_0, -2\beta_0)$. Both stories are valid, yet they convey different narratives and mostly rely on the researcher's prior beliefs or choices. This idea is illustrated in Figure C1.

Consider another scenario where the four candidates can now be ordered along the left-right axis, and a libertarian vs authoritarian axis. Also assume that there is a positive correlation between the two dimensions, but that the treatment effect is only a function of the first one. If for some reason the researcher had a different prior and solely focused on the second dimension,

Figure C1: Two different partitions used to estimate heterogeneous treatment effect



it is likely that they will find a significant effect anyway. And if they conduct their analysis for the two cleavages separately, it could be hard to compare the two results and understand whether some dimension mattered the most. We could even imagine new scenarios with more dimensions where some have a positive and others a negative relationship with the treatment effect and the empirical challenges would then come always more numerous.

We frame our empirical strategy as an heterogeneity/partitioning exercise to overcome this challenge. The main underlying assumption is that in Figure C1, panel B is the preferred. However, this assumption does not hold universally, as obtaining a null result can be informative in itself and may hold significant interest for both academics and society. But we acknowledge that researchers do have prior beliefs and that cherry-picking and p-hacking is a real issue in economics (Brodeur et al., 2020). We believe it is more advantageous to transparently address this aspect, rather than candidly ignoring it. To illustrate this, let us consider a scenario where each candidate is characterized by only three dichotomous variables (e.g., left-right, populism, pro-environment), resulting in eight distinct types. In such a case, these candidates can already be organized into a 877 different partitions.

C.2 The algorithm

Formal setup Given a constituency level variable of interest ($3G^+$ in our study), the researcher wants to find out what is the effect of this variable on electoral outcomes and has no prior on which cleavage of the multi-dimensional political competition is suited to describe this effect. Formally, our main underlying assumption is that the researcher's interest is to find how to group candidates who experience similar effects, or equivalently to maximize treatment heterogeneity across groups. The second underlying assumption is that the treatment effect can be described by some candidates observable characteristics.

Formally:

- N constituencies denoted by $i \in \{1, \ldots, N\}$

- K candidates denoted by $k \in \{1, \ldots, K\}$
- $W_i \in \{0, 1\}$ is the treatment status of constituency i
- $Y_{ik} \in [0, 1]$ is the vote share obtained by candidate k in constituency i
- $C_k \in \mathcal{C}$ is the vector of observable characteristics of candidate k
- $X_i \in \mathcal{X}$ is the vector of observable characteristics of constituency i

In the simplest scenario, the researcher's focus is solely on the treatment effect β^{k_0} of variable W on the outcome Y for a specific candidate k_0. The standard approach would then be to estimate:

$$Y_{ik_0} = \beta^{k_0} \cdot W_i + \gamma^{k_0} \cdot X_i + \epsilon_{ik} \tag{5}$$

However, in certain contexts, the researcher might aim to estimate a treatment effect for a group of candidates. For example, when investigating the voting behavior for unaffiliated candidates, there could be multiple such candidates in a given constituency. Let θ_0 represent the set of candidates of interest, referred to as a candidate type. The model equation in this case would be:

$$\sum_{k=1}^{K} \mathbb{1}_{k \in \theta_0} \cdot Y_{ik} = \beta^{\theta_0} \cdot W_i + \gamma^{\theta_0} \cdot X_i + \epsilon_{ik} \tag{6}$$

and we would have

$$\beta^{\theta_0} = \sum_{k=1}^{K} \mathbb{1}_{k \in \theta_0} \cdot \beta^k \tag{7}$$

Now the example of affiliated vs. unaffiliated candidates is a binary partition of candidates. In this case, the effect of the vote share is entirely characterized by the estimate of β^{θ_0}. However, in multiparty election settings, researchers often incorporate more than two classes, such as left, center, and right. In such cases, comprehensively capturing the treatment effect requires at least estimating the effect for an additional candidate type, denoted as θ_1. This introduces certain empirical challenges, as the error terms in equation (6) for estimating β^{θ_0} and β^{θ_1} are correlated for instance. Furthermore, researchers need to make a pre-determined decision regarding the relevant political types for their study, raising the challenges we presented in the previous section.

Building on the existing literature on multiparty econometrics, we now propose to overcome the last issue by framing it as an heterogeneity exercise. We assume that the researcher interest is to find types that will cover the entire set of candidates, and present the maximum treatment effect heterogeneity across types.

Partitions and trees Let $\mathcal{C}$ be the definition set of vector C_k. A type of candidate θ is a subset of $\mathcal{C}$. Let $\Theta = \{\theta_1, \ldots, \theta_P\}$ be a given partition of $\mathcal{C}$.[27] The elementary bloc of the procedure is

[27] $\cup_{p=1}^{P} \theta_p = \mathcal{C}$ and $\forall i, j \; i \neq j \Rightarrow \theta_i \cap \theta_j = \emptyset$

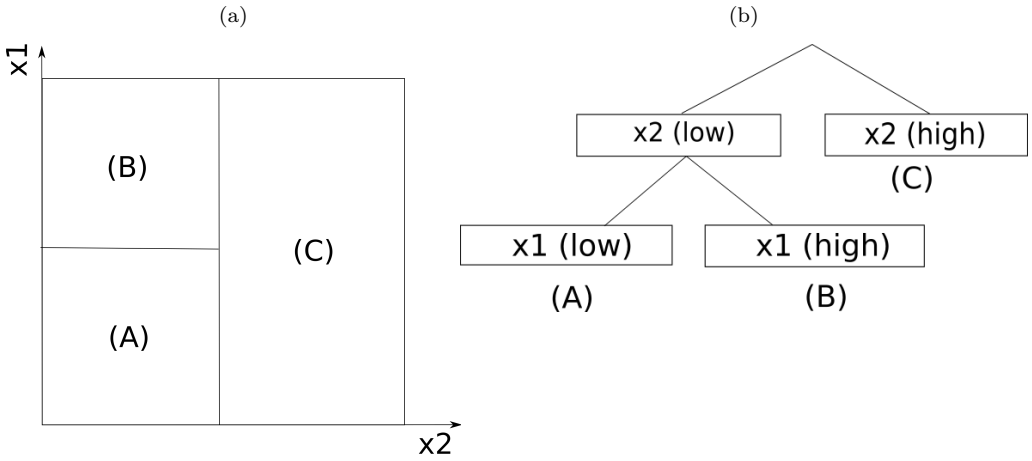
to find Θ such that it maximizes the heterogeneity of treatment across types θ_p. This is done by applying Athey and Imbens (2016), relying on two main ideas: recursive partioning and honest splitting. The first refeer on how the best partition is sought: the algorithm finds which split along one of $\mathcal{C}$'s dimensions will maximize treatment heterogeneity and then proceed recursively in the two resulting subsets. The latter to the fact that the splits and final estimates are computed using two separate subsets to obtain unbiased estimates.

Let $l(\theta, w)$ be the set of observations such that $C_k \in \theta$ and $W_i = w$, i.e. the set of observations corresponding to candidates of type θ and with treatment status w. Then for a given type of candidate θ, the estimated treatment effect is estimated as the difference between the average outcome among treated and control observations.

$$\beta^{\theta} = \frac{\sum_{i \in l(\theta,1)} Y_{ik}}{\#l(\theta, 1)} - \frac{\sum_{i \in l(\theta,0)}}{\#l(\theta, 0)} \tag{8}$$

So for a given partition Θ, the treatment effect is estimated as a step function of variables in $\mathcal{C}$. The partitions resulting from such procedure can be directly represented as trees, as illustrated in Figure C2.

Figure C2: A two dimensional partition and its binary tree representation



The advantage of this procedure is that now the partition is not fixed *ex-ante* by the researcher, but rather an output of the estimation procedure. But trees are known to be quite noisy. This can be tackled by fitting numerous trees and then averaging the output, i.e. fitting a forest. For the full estimation algorithm, we follow the generalized random forest by Athey et al. (2019). The full procedure ends up being a non parametrical estimation of the following equation:

$$Y_{ik} = f(X_i, C_k) + \beta(C_k) \cdot W_i \tag{9}$$

This underlines that we are now interested in finding the dimensions which better explain the

heterogeneity of the treatment β. The reader should definitely refer at least to Athey and Imbens (2016) and Athey et al. (2019) to fully understand the algorithm.[28]

C.3 GRF implementation and properties for multiparty elections

Clustering A specific feature of electoral data is that outcomes of different candidates in a given constituency are directly related. If one vote share decreases, there must be at least another vote share increasing. To address this issue, we closely follow Tomz et al. (2002) who suggested to use Seemingly Unrelated Regressions (SUR) to account for the correlations in the equation terms between political blocs. We mimic this clustering structure with standard errors clustered at the constituency level. Indeed, as we study a constituency level treatment, SUR is equivalent to run a single regression with stacked data and coefficients interacted with each political bloc separately with standard errors clustered at the constituency level.

Partially contested districts An other empirical challenged faced by researchers in multiparty elections is partially contested districts. Indeed, especially in local elections it could be that not all types of candidates would run in every constituency. One solution offered by Tomz et al. (2002) is to estimate separate models for each competition pattern. But this approach becomes hard to implement if there are too many parties and therefore even more potential competitions arrangements. For the GRF method, as random forests are perfectly suited to handle high level of interactions, we propose to use a parsimonious yet flexible approach. By simply hot encoding each party's absence in a given constituency, the treatment effect could depend on any competition configuration, and possibly differently for each type of candidate.

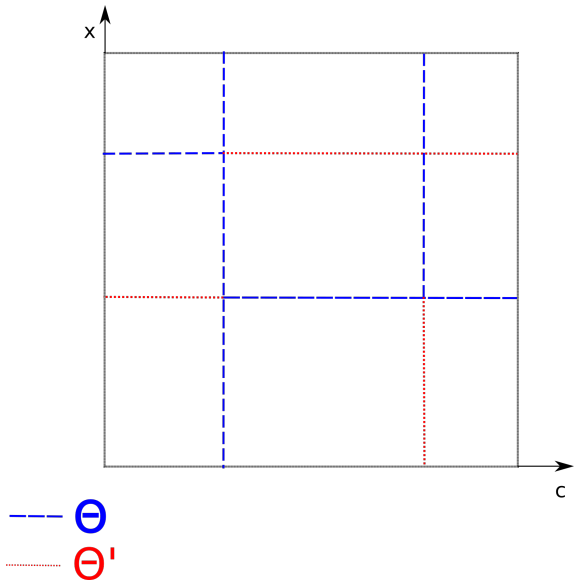
Turnout, log ratios A last property satisfied by vote shares data is that if all candidates are considered, all vote shares should be bounded between 0 and 1, and eventually sum to 1 all together. Katz and King (1999) explore this aspect and propose an econometric framework that also incorporate the link between turnout and actual vote shares obtained by the candidates, but their method is hard to implement for more than three political parties. The main idea is to consider the vector $(Y_1, \ldots, Y_{K-1})$ and enforce the constraint $Y_K = 1 - \sum_1^{K-1} Y_k$. They also consider the transformed outcome $ln\left(\frac{Y_k}{Y_K}\right)$. This approach could be adopted with the GRF algorithm by considering votes in terms of registered voters and regarding abstention as the candidate K for instance. We decided in this study not to do so, because we believe it would have make the interpretation of the coefficients of equation (3) more confusing when discovering the approach. We hope further research will reconcile the two approaches to strengthen the methodology.

[28] Athey et al. (2019) also insist on the link with matching.

Orthogonalization The authors of the GRF empirically found crucial in observational studies to account for features that affect treatment effects and those that affect treatment propensities.[29] In GRF, this is done in a first step based on "orthogonalization" (Robinson, 1988) where the level of outcomes and the treatment propensities are estimated respectively based on C and X, and X alone.

Constituency heterogeneity We can easily introduce constituency level heterogeneity by simply allowing the procedure two split also on dimensions of $\mathcal{X}$, definition set of X. In such case, the resulting trees are now partitions of $\mathcal{C} \times \mathcal{X}$. One can always construct a refinement of such partitions as the Cartesian product of two partitions of respectively $\mathcal{C}$ and $\mathcal{X}$. This is illustrated in Figure C3. Practically, we used in our study constituency level heterogeneity as a benchmark for variable selection (Section 5.2.1), based on the heuristic that the treatment effect is likely to be heterogeneous have two level of heterogeneity (candidate and constituency level), but it is plausible that one source of heterogeneity dominates the other. Indeed, we would expect the most important candidate characteristics to have higher importance than the constituency's characteristics.

Figure C3: A partition Θ of $\mathcal{C} \times \mathcal{X}$ and its refinement Θ' as the Cartesian product of two partitions of $\mathcal{C}$ and $\mathcal{X}$



Other properties Generalized random forest offer many other possibilities that could be useful for multiparty econometrics. For instance, the treatment is not required to be dichotomous. It is also feasible to use instrument variables. In our presentation, the estimated effect in a given tree is a step function with constant values within each leaf but this can be relaxed, for instance by using linear regressions within those leafs (bringing the approach even closer to Tomz et al. (2002)).

[29] `https://grf-labs.github.io/grf/REFERENCE.html`

C.4 Implementation in our study

In this section, we describe the exact implementation we use of the GRF algorithm, using the `grf` package from `https://grf-labs.github.io`.

- We compute the first estimate the outcome $\hat{Votes}_{kit}$ with a `regression_forest` fitted both on the candidates platforms and the constituencies characteristics, clustered at the constituency level and with `min.node.size=100`
- Similarly we estimate the treatment propensity $3\hat{G}_{it}^{+}$ with a `regression_forest` fitted solely on the constituencies characteristics, clustered at the constituency level and with `min.node.size=100`.
- Finally, we fit the `causal_forest` on both the candidates platforms and the constituencies characteristics, using our two previous estimates for the parameters `Y.hat` and `W.hat`. Again we cluster at the constituency level and set `min.node.size=100`.

C.5 Conclusion

Identifying some limitations in the standard approach to multiparty econometrics when studying the effect of a local variable at the constituency level, we proposed to apply the Generalized Random Forest Algorithm Athey et al. (2019). The setting being framed as an heterogeneity exercise, the algorithm has the advantage to uncover which political cleavage shape the effect of a constituency level treatment. By framing our approach as an heterogeneity, this algorithm offers the advantage of revealing the specific political divisions that influence the effect of a treatment at the constituency level. We demonstrated how this approach aligns with other empirical challenges that have been previously identified in the study of electoral data. We are optimistic that this methodology will find increasing utility, even if it may require some adjustments, ultimately becoming a valuable and practical tool for researchers in the field of political economics.