SOFIA UNIVERSITY "ST. KLIMENT OHRIDSKI"

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

DEPARTMENT "STATISTICS AND ECONOMETRICS"

Applications of data analysis algorithms

Boyko Amarov

ABSTRACT OF DOCTORAL THESIS

For obtaining the educational and scientific degree "Doctor" in the professional field 3.8. Economics (Doctoral program "Analytical research on data" (Data Science))

Scientific advisor: Prof. Ivan Ivanov, Dr. Sc.
Scientific consultant: Assoc. Prof. Boyan Lomev, PhD

Sofia, 2023

The dissertation was discussed at the Extended Departmental Council of the department for "Statistics and Econometrics" at the Sofia University "St. Kliment Ohridski", Faculty of Economics and Business Administration on 20.03.2023. The author of the thesis is a part-time PhD student at the same department, according to the decision of the Faculty Council of the Faculty of Economics and Business Administration, protocol No and Order No RaO 20-1187 of 17.07.2018 of the rector of the university.

Contents

1 General characteristics of the thesis

1.1 Relevance of the topic

Widespread access to the Internet enables the provision of a wide range of government electronic services (e-services) that can significantly facilitate the interactions between citizens and government bodies and reduce the costs associated with these services. Instead of waiting in queues at government or municipal offices, citizens can submit applications, such as tax returns, from their homes or mobile phones. In the years of the pandemic caused by COVID-19, e-services have proved vital to enable schools and universities to continue operating despite lockdowns by switching to online forms of teaching. In other areas, the e-government services also helped reduce face-to-face meetings.

Despite the significant benefits of using government e-services, the uptake has been slower than expected (Carter & Weerakkody 2008). This is particularly true for Bulgaria, which lags behind most of the European Union (EU) countries in the use of government e-services, despite having a good Internet infrastructure (Eurostat 2021). Research on the diffusion of government e-services points to three main barriers to their use. To use e-services, citizens need a computer or a mobile phone and a reliable internet connection. Access to government websites or mobile applications is only a prerequisite for effective use of e-services. In order to make full use of them, citizens must also possess a degree of skills for working with information and communication technologies (ICT). Several studies in the area of government e-services point to weak ICT skills among some citizens as a significant barrier to greater use of e-services (Bélanger & Carter 2009; van Deursen & van Dijk 2014). The use of e-government services has been studied in a number of countries (Distel & Ogonek 2016), including Bulgaria's neighbors – Romania (Colesca & Dobrica 2008), Turkey (Kurfalı, Arifoğlu, Tokdemir, *et al.* 2017) and Greece (Voutinioti 2013).

The present doctoral thesis contributes to filling this gap by systematically analyzing the usage of government e-services in Bulgaria based on data from a survey of internet users in Bulgaria in 2021 (Chapter 1).

One important area where new technologies can reduce time and cost is voting technology in elections. Replacing the traditional paper ballots with electronic voting machines (EVM) in Bulgaria's July and November 2021 general elections has sparked a heated public debate.

The discussion on the benefits and drawbacks of EMV voting in Bulgaria essentially re-

peats arguments from similar debates in the Netherlands (Jacobs, Bart & Pieters 2009) and Germany (Volkamer 2010). The main arguments for introducing machine voting and automatic vote counting include faster reporting of election results and a more limited scope for human error in counting ballots. An additional advantage of machine voting is automatic ballot verification, preventing voters from casting blank, erroneous, or spoiled ballots (invalid ballots). The invalid ballots, as well as votes for "None of the Above" (NOTA), are discarded from the election result (uncounted votes). The latter advantage may be significant in Bulgaria, where in the last two elections for municipal councils in 2015 and 2019, uncounted votes exceeded 15 percent of all votes.

For this dissertation, the term "invalid ballots" includes all ballots that are not counted in the final election result and are not NOTA votes.

Introducing compulsory machine voting after April 2021 coincided with a sharp decline in turnout in Bulgaria from 50.6 percent in April 2021 to 42.2 percent (July) and 40.2 percent in November. The decline in voter turnout continued in the following year's elections, reaching a level below 40 percent in October 2022 (39.4 percent).

Low turnout can call into question the legitimacy of elected political representatives, and when turnout declines unevenly across different groups in society, groups with lower turnout remain politically underrepresented (Lijphart 1997). This observation forms one of the main arguments against compulsory EVM voting. If specific groups of voters do not go to the polls because of concerns about the operation of EVM or because they have doubts about their vote being counted correctly, this behavior will lead to a selective decline in voter turnout, disenfranchising these citizens. These concerns find support in research that shows differences in turnout when using different voting and vote-counting technologies, including electronic voting machines (Card & Moretti 2007; Dandoy 2021; Roseman & Stephenson 2005). This dissertation examines three aspects of machine voting in Bulgaria.

In Chapter 3, the thesis examines three forms of non-participation (invalid voting, NOTA voting, and abstention) during the local elections in 2019, when the share of uncounted votes reached 18 percent. The shares of these forms of non-participation are modeled within a multilevel regression framework with socioeconomic and demographic, institutional, and political characteristics as explanatory variables. The analysis uses data with different levels of aggregation – municipality, locality, and polling station. The analysis shows evidence of the protest character of some of the invalid votes. This result motivates the analysis in Chapter 4, which explores whether the option of voting with a machine voting is associated with lower levels of uncounted ballots. This analysis uses

data from the 2019 European Parliament (EP) and April 2021 parliamentary elections. The statistical model reveals a tendency towards a redistribution from invalid ballots to NOTA and votes for marginal (small parties with only a slight chance to obtain a mandate in the elections) parties when using a machine, but also a positive net effect on the probability of valid votes for non-marginal parties.

Chapter 5 of the thesis analyses two aspects of introducing EMV voting in Bulgaria. The study's first aim is to assess the selectivity of machine voting by examining associations between the propensity to use an EMV and voters' socioeconomic and demographic characteristics. In addition, the study assesses the effect of previous experience with machine voting on the propensity to use EMV. The analysis employs data from the April 2021 general election, the 2019 EP election at the polling station, and the 2011 census data at the locality and municipality levels.

The dissertation's final chapter examines the variation in turnout differences in the July and November 2021 parliamentary elections compared to April 2021 by socioeconomic and voter demographic characteristics. The analysis uses the proportion of voters who voluntarily chose to vote with a machine in the April 2021 election to measure voter willingness to adopt the new technology. The study uses polling station-level data at the three general elections in 2021 and models the difference in turnout within a two-stage linear regression model. The statistical model shows a weak positive association between the proportion who voluntarily used machine voting and turnout differences.

1.2   Purpose, object, subject, and scope of the study

The first goal of the thesis is to study the characteristics of users and non-users of e-government services in Bulgaria. The second goal of the dissertation is to examine the use of electronic voting machines in elections in Bulgaria and the effect of their introduction on voter turnout and uncounted vote shares.

Regarding the first goal, the object of the dissertation is Internet users in Bulgaria. Regarding the second goal, the object of this thesis is the parliamentary elections in Bulgaria in April, July, and November 2021, the EP elections, and the local elections in 2019.

Regarding the first goal, the subject of this dissertation is the associations between socioeconomic characteristics, the level of information and communication technology (ICT) skills of Internet users in Bulgaria, and their propensity to use government e-services.

The subject of the dissertation regarding the second goal is the associations between socioeconomic and demographic characteristics, institutional and political characteristics

of municipalities, localities, and polling stations, and forms of non-participation in elections (non-voting, invalid voting, NOTA voting, and voting for a marginal party).

1.3   Research tasks

1. Estimate the associations between socioeconomic characteristics, ICT skills, and the propensity of Internet users in Bulgaria to use e-government services in three areas: administration, education, and healthcare.

2. Estimate the associations between probabilities of non-participation in the 2019 local elections, socioeconomic characteristics at the municipality and locality level, and institutional and political characteristics at the municipality (electoral district) level.

3. Estimate the effect of EVM availability and usage at the polling station in the 2019 EP and April 2021 general election on the proportions of uncounted votes and marginal parties' votes.

4. Model the propensity to vote with an EVM during the 2019 EP and April 2021 general elections according to socioeconomic and political characteristics of localities and municipalities and political preferences at the polling station level

5. Model the change in turnout in Bulgaria's July and November 2021 parliamentary elections depending on the proportion of voters voluntarily voting by machine in the April 2021 elections.

1.4   Summary of methodology

All studies in this dissertation use generalized multilevel Bayesian regression models.

While the e-government services research uses individual-level data, all election-related studies rely on aggregate polling station-level data and municipal and locality-level data. The analyses related to the Bulgarian elections employ data that is publicly available from the Central Electoral Commission, the National Statistical Institute and the national employment agency.

1.5 Limitations

In the first chapter of the dissertation, the main limitation of the analyses is imposed by the design of the study, which does not differentiate between the use of *electronic* government services and traditional forms of their use.

The analyses of elections in Bulgaria use data aggregated at the polling station, locality, and municipality level. For this reason, the statistical models do not allow direct inferences about individual voter behavior (Piantadosi, Byar & Green 1988).

1.6 Structure of the dissertation

The thesis includes an introduction, six chapters, a conclusion, a bibliography, and one appendix (161 pages). The dissertation contains 30 figures and 30 tables.

The first chapter of the dissertation presents the Bayesian approach to generalized regression models, the primary analytic tool employed in the following five chapters.

The second chapter examines the usage of e-government services among Internet users in Bulgaria depending on their socioeconomic characteristics, online shopping habits, and level of ICT skills.

The third chapter presents the main theoretical approaches explaining variation in turnout and uncounted votes that underlie the analyses in the following three chapters. In addition, the chapter presents an analysis of the variation in uncounted ballot shares in the 2019 local elections in Bulgaria according to the socioeconomic characteristics of municipalities and localities, as well as institutional and political factors.

The fourth chapter examines the effect of the introduction of electronic voting machines on the shares of people who abstained, voted with an invalid ballot, voted with NOTA, or voted for a marginal party. The analysis uses data from the 2019 EP and April 2021 general elections.

The fifth chapter analyses differences in the propensity to use electronic voting machines in the April 2021 and 2019 EP elections by socioeconomic and demographic characteristics of municipalities and localities. In addition, the analysis examines the association between the propensity to use the machines and previous experience with machine voting at the polling station level in the April 2021 parliamentary elections.

The final, sixth chapter examines the decline in turnout in the July and November 2021

parliamentary elections relative to turnout in April of that year as a function of the proportion who volunteered to vote by machine.

The dissertation concludes with a summary of the main results, a discussion of the limitations of the analyses, and suggestions for future research.

1.7 Contributions

The dissertation contributes by applying statistical modeling in five areas:

1. It systematically analyzes the propensity to use e-government services in administration, education, and healthcare according to socioeconomic characteristics, level of ICT skills, and experience with Internet technologies. This part of the dissertation represents the first systematic study of the demand for government e-services in Bulgaria.

2. It systematically analyzes the propensity of voter non-participation (abstention, invalid voting, and NOTA voting) in the 2019 local elections according to municipalities' and localities' socioeconomic and demographic characteristics. In addition, the dissertation examines the propensity of non-participation according to institutional and political characteristics at the municipality and polling station levels. This part of the study contributes to the scholarly literature on invalid voting and turnout research.

3. It explores the effect of the availability of an EVM on the propensity to abstain, vote invalidly, vote with NOTA, and vote for marginal parties in the 2019 EP elections. In addition, the dissertation explores the association between the propensity for these four forms of voter behavior and the proportion of voters using machine voting in the 2019 EP elections and the April 2021 general election. This part of the thesis contributes to research on adopting electronic voting technologies in elections and unconventional voter behavior.

4. It explores the propensity to use electronic voting machines in the 2019 EP and April 2021 parliamentary elections according to socioeconomic and demographic characteristics measured at the municipality and locality levels. In addition, the dissertation examines the effect of experience with machine voting on the propensity to vote by machine. This part of the dissertation contributes to research on the digital transformation of voting technologies.

5. Examines differences in turnout in the November and July 2021 elections relative to the April 2021 elections as a function of the proportion of voters who volunteered to vote by machine in April 2021, as well as as a function of political preferences at the polling station level and socioeconomic and demographic characteristics at the municipality and locality level. The last part of the dissertation contributes to research on voter turnout and voting technology types.

2 Main content of the dissertation

The following references to figures, tables, and formulae refer to the dissertation, not the present thesis summary.

2.1 Introduction

The introduction motivates this dissertation's research and summarizes the methodology, main results, and limitations.

2.2 Chapter 1: Bayesian regression models

The first chapter briefly describes Bayesian multilevel regression models, which find application in the remaining five chapters of the dissertation.

2.3 Chapter 2: Electronic Government Services

The second chapter of the dissertation analyses the usage of government e-services by Internet users in Bulgaria and the differences in their use according to the level of digital technology skills, education, and socioeconomic factors. The analysis covers three broad government and municipal e-services categories: administrative services and education and healthcare-related services. The study uses data from a survey of internet users aged 15+ in Bulgaria collected between June and August 2021. Variation in the propensity to use the services in these three categories is modeled within a hierarchical logistic regression model incorporating the socioeconomic characteristics of the respondents, as well as varying effects at the regional level. Respondents with high education and good ICT skills have a higher predicted probability of using government e-services. The model shows no differences in behavior between men and women regarding the likelihood of

using e-services, except for school-related e-services, which women are more likely to use.

2.3.1 Methodology

The analysis of public e-services uses data from a survey of internet users in Bulgaria aged 15 or older, collected between June and August 2021. The interviews were conducted face-to-face by trained interviewers. The main questionnaire contained questions on the use of e-government services over the past twelve months – 1 039 of respondents stated that they had used at least one government e-service during this period (users). Other 385 respondents did not use any e-service during the same period (non-users). The study includes an additional survey of non-users of government e-services, concentrated in large cities in Bulgaria. The survey of internet users in Bulgaria was carried out within project KP-06-H45/3/30.11.2020 of the Scientific Research Fund "Identification of citizens' attitudes and assessment on access, quality and use of e-public services."

The respondents provided information on their use of government e-services in three broad areas: administration, education, and healthcare. Administrative services include e-renewal of identity documents, driver's licenses, e-filing of tax returns, payment of taxes and fines, receiving e-payments from government institutions such as pensions, social security, and honoraria, and inquiries about one's social security status. Administrative services also include purely informative services such as searching for vacancies or looking for information on applying for social benefits in online government sources.

Healthcare-related services include online health information searches, online inquiries about one's health status, and services allowing citizens to schedule appointments with their personal GP and consult medical professionals online. In education, e-services include access to school and university online courses, e-learning materials, access to exam grades, online enrolment in schools and universities, study schedules, and electronic payment of fees at educational institutions.

In the field of education, the questionnaire includes questions on the use of e-services by public schools (providers of primary and secondary education), universities (providers of higher education), NGOs, private educational institutions, and administrative services related to the affixing of apostilles and the issuing of electronic copies of diplomas and other certificates. Electronic services provided by schools, universities as well as private educational institutions include online e-learning courses, online examinations

and counseling, access to assessments, electronic school registers and electronic fee payments to educational institutions.

In addition to the use of government e-services, respondents provide information on socioeconomic and demographic characteristics: gender (male, female) occupational status (retired, employed, unemployed, or student) level of education (primary, secondary, or tertiary) type of locality in which they usually live (city, small town, or village) monthly income in BGN measured in six groups Respondents rate their ICT skills on a five-point scale from low to high and also indicate how often they shop online: never, rarely (up to once or twice a year), or often (every month). Due to the uneven distribution of ICT skills, this variable has been grouped into three levels for analysis: low, medium, and high. Table 2.1 shows the number of respondents for all characteristics used in the analysis. The distribution of users and non-users of e-services is shown in Table A.1 in the Appendix.

In the sample of users, 65.54 percent of respondents used at least one of the administrative services, 42.06 percent at least one of the education-related services, and 42.06 percent at least one of the services, health-related, and 12.42 percent of users have used services from all three categories in the past year. Figure 2.1 shows the proportions of users for the most frequently accessed services in the three areas.

The likelihood of using e-services was modeled within hierarchical logistic regression models (Model 2.1 in the thesis). The models contain fixed effects of socioeconomic and demographic characteristics (gender, education, type of residence, income), level of ICT skills, and frequency of online shopping. All models include normally distributed varying effects at the administrative region level (28 regions). Regional effects account for possible unmeasured differences in the propensity to use e-services. Examples include ethnic composition, cultural differences, differences in the supply and promotion of municipal services, and different citizen experiences with local government, education, and health institutions that could be related to the likelihood of using e-services. In addition, regional effects partially account for the sample design, as non-users were selected among residents of the largest cities in Bulgaria.

Because education services are primarily targeted at young people and because of the de facto mandatory nature of online learning for a significant number of students during the pandemic, the analysis of consumption of these services splits the data into two parts: students and economically active respondents. Adults who have already completed their education are more likely to engage with school-related services when helping with their children's education. Because the age of respondents is not available in the data, the oc-

cupational status of respondents serves as a proxy measure for age. The student sample contains (n = 159) observations and economically active respondents (employed and unemployed) total 1 091. A separate model is estimated for each of these two samples. The two models differ in the independent variables used, as the survey of Internet users does not collect income information from students and retirees. Retired respondents are excluded from this analysis as all types of e-services considered are primarily targeted at younger people. Only 3 of the 114 retired respondents in the sample used services offered by schools, and none used services from the other three provider types.

2.3.2 Results

Table 2.2 in the dissertation shows the posterior means and standard deviations of the fixed effects in the logistic regression model for service use from the three broad categories: administration, education, and health. Positive coefficients indicate a higher predicted probability of use, and coefficients with a 95% central credible interval that does not contain zero are highlighted in boldface.

Regarding differences in the use of e-services between men and women, the model shows no evidence of a greater propensity of men to use e-government services compared to women for administrative and health-related services, and the results are similar to the findings of Bélanger & Carter (2009) and Inkinen, Merisalo & Makkonen (2018). At the same time, women have a higher predicted probability of using education-related e-services than men.

E-services can significantly reduce the costs for citizens when they have to interact with the administration, and this is more relevant for residents of small settlements far from large administrative centers, as the costs in money and travel time are relatively higher for them. However, the model shows no differences in the likelihood of using government e-services between citizens living in large cities or rural areas, and residents of small towns show a lower likelihood of using e-health services.

One possible explanation for the lack of difference between cities and villages is lower awareness of the availability of these services outside large cities. Another possible explanation is a potentially more difficult access to fast Internet in some of Bulgaria's villages and small towns, complicating access to e-services. These results must be considered cautiously as the survey design includes non-users of e-services in large cities. For this reason, the difference in the probability of using e-services between large cities and rural areas is likely larger than the current model suggests.

Since the model does not directly include the age of respondents, occupational status captures age differences, with students being the youngest and retirees the oldest. Unsurprisingly, the model shows that students are more likely to use e-learning services than employees, as most education-related services are relevant to younger people. An additional explanation for this effect is the shift to remote learning methods during the COVID-19 pandemic, which forced students and pupils to switch to online forms of learning. Students are also less likely to use electronic administrative services. A possible explanation for this result is that most of these services are mainly relevant to economically active individuals. The same explanation applies to the low predicted likelihood of retirees using education-related services. An unexpected result is the predicted higher likelihood of retirees using administrative services than employed persons. An analysis of the robustness of the model indicated that this result was a consequence of the survey design, which did not collect income data on retired respondents and students. Estimating the model without the income variable (not shown) revealed a lower propensity of retired individuals to interact with administrative services.

The model shows a higher likelihood of using administrative services for middle- and high-income citizens and a lower likelihood for the lowest- and highest-income citizens. While the model reveals no systematic variation in probabilities of use across income groups for healthcare-related services, respondents with the highest incomes also have the lowest predicted probability of consuming education services.

During the COVID-19 pandemic, visits to health facilities and crowded places were associated with a higher risk of infection. In this regard, e-health services provided an opportunity to avoid in-person visits. As older people consume healthcare services to a greater extent than other age groups and their risk of severe COVID-19 is higher than younger people, the author of this paper expected the effect to be that pensioners would be more likely to use healthcare e-services. The model shows only weak evidence of such an effect on the overall use of these services. Although the coefficient on the retiree indicator is positive, the posterior probability that it is greater than zero is only 0.84.

Although the sample consists only of users with Internet experience, the model shows substantial variation in the probability of using all three types of services depending on the respondents' education and ICT skill level. Low levels of education and weak ICT skills are associated with a low probability of using all three types of services, supporting the hypothesis of an existing digital divide in society regarding adequate access to e-government services. Unlike the results of Bélanger & Carter (2009), which found no association between respondents' previous experience with commercial Internet services

such as online shopping and use of government e-services, the model shows a strong association, with respondents using the Internet for shopping frequently also having a higher predicted probability of using services from all three categories. A possible common factor behind more frequent online shopping and use of government e-services is these users' trust in Internet technology in general.

Table 2.3 in the dissertation presents a summary of the dispersion parameters of the regional effects. All three standard deviations have 95 percent credible intervals with a lower bound far from zero, indicating regional heterogeneity in using e-services, even after controlling for the effects of other variables. Heterogeneity is highest for health-related services. Figure 2.2 in the dissertation visualizes the averages of the posterior effects and shows that residents in the Varna and Burgas regions, including two of the largest cities in Bulgaria, had a higher propensity to use e-health services than other regions. This result is remarkable as the non-user sample is concentrated in the country's major cities, including Varna and Burgas. The correlations between regional effects across service types (Table 2.3) are positive but small, and their 95 percent credible intervals include zero. This means that respondents living in a region with a high propensity to adopt one type of service are not necessarily more likely to use the other types of services.

2.3.2.1 Administrative services Table 2.4 in the dissertation summarizes the fixed effects in the consumption pattern of the most commonly used administrative services: criminal record issuance, tax return filing, tax, and local fee payment, and electronic payment of fines. As with the general analysis of e-service use in the previous section, the pattern for individual administrative services shows a lower predicted likelihood of service use by respondents with no online shopping experience and a low level of digital skills. The exceptions are paying fines online and receiving pensions and fees electronically.

While the model results for the broad service groups show a lower propensity to use them for younger respondents (students), the detailed analysis of administrative services shows a higher predicted likelihood for students to use an e-criminal record. A possible explanation for this result is the entry into the employment of upper-year students for which they need the criminal record certificate. The predicted probability of receiving electronic payments is higher for respondents of retirement age, as most pensioners in Bulgaria receive their pensions by bank transfer. Compared to employed respondents, the model shows a lower predicted probability of using administrative services for unem-

ployed respondents, with the effects being strongest for receiving electronic payments and services related to income tax and local taxes and fees.

Respondents' monthly income and level of education are associated with the likelihood of using electronic tax-related services. The model also shows a difference in behavior between men and women, with women having a lower predicted probability of paying fines online than men. The latter result is likely to reflect differences in the frequency of receiving fines between men and women.

2.3.2.2 Education Table 2.5 summarizes the a posteriori distribution of the fixed effects of the logistic regression models for the economically active population and the use of e-services in education. For three of the four types of services examined in the model, the coefficients show no significant differences in the probability of using e-services between males and females. The exception is the higher propensity of women to access school e-services compared to men. A possible explanation for this effect is the traditional distribution of roles in a considerable proportion of Bulgarian families, where there is a tendency for women to be more involved in child-rearing than men, including concerning children's schooling.

The model shows no significant differences in predicted probabilities of e-service use by respondents' income, except for university e-services, where high income is associated with low use probability. Since the questionnaire does not distinguish between part-time and full-time employment and age is absent in the model, this effect may capture differences in respondent age. Younger people who are still at university are more likely to earn relatively low wages and are more likely to work part-time than other workers. The propensity to use administrative and private educational e-services does not vary significantly with respondents' income. Also, the results show no evidence of a relationship between occupational status and the propensity to use e-services.

At the same time, low levels of education are associated with a lower propensity to use all types of e-services. Confidence intervals for the coefficients on the indicator for primary education include zero for all four types of services, but this is related to the low number of respondents (14) in this category, resulting in high posterior standard errors. Beyond the association between education and propensity to use services, the model predicts a lower likelihood of using all four service types for individuals with low self-rated digital skills. In the case of private educational e-services, the 95 percent confidence interval includes zero, but the a posteriori probability that the coefficient is less than zero is high (0.97). The frequency of online shopping used to measure respondents' experience

with commercial e-services is positively associated with the propensity to use all types of services except those provided by private providers.

For private e-services, the model shows evidence of a rural-urban divide. Those living in rural areas and small towns have a lower predicted probability of using these services. This effect is likely underestimated due to the concentration of nonusers in Bulgaria's large cities. One possible explanation for this effect is that private education providers tend to be concentrated in Bulgaria's large cities. The model shows no evidence of an existing city/rural divide for other types of e-services, but this result should be considered carefully as the effect is likely underestimated.

For all four types of e-services, the regional effects indicate the existence of heterogeneity across regions in terms of propensity to use all four types of services even after the model has accounted for the effect of the other variables (Table A.3 in the Appendix). The posterior standard deviations range from a mean of 0.86 (school services) to 1.00 (administrative services), and all standard deviations have 95 percent confidence intervals with a lower bound greater than 0.54. The uneven distribution of educational institutions across regions may partially explain this difference, as most universities and private institutions are located in larger cities. Local factors, such as differences in the promotion of e-services in schools and local governments, are likely to influence the propensity of Internet users to use these services. Another factor behind regional heterogeneity may be partly the concentration of the sample of nonusers in large cities.

The two correlation coefficients with 95 percent confidence intervals, including zero, are between the regional effects of school and administrative services and university and administrative services. All other correlations are positive and range from moderate (university and administrative services: 0.47) to strong (administrative and private services: 0.84), indicating that regions with a high propensity to use one of the services also tend to exhibit a high propensity to use the other types. Estimates of regional effects can serve as a basis for further analyses of differences across regions and help identify local success factors. An example might be the Burgas region, which includes the fourth largest city in the country and shows one of the highest propensities to adopt educational e-services, despite the survey design.

In contrast to the findings for economically active respondents, the model for the student sample (Table 2.6 in the thesis) shows no evidence of differences between men and women for any of the service types. The high coefficients for levels of education reflect the interest of students with primary education in services provided in secondary education. In contrast, students with secondary education are interested in accessing univer-

sity e-services. Another difference with the results for economically active respondents is the association between ICT skill levels.

A sensitivity analysis of the results reveals that this is a consequence of the small number of university-educated learners in the sample (8 respondents), all of whom highly rate their digital skills. Similar to the results for economically active individuals, students living in rural areas show a lower propensity to use private e-services. Students without experience in e-shopping are less likely to use all types of e-services considered in the model. The model also shows no significant regional heterogeneity.

2.3.2.3 Healthcare In the area of e-government services in health care, Table 2.7 in the dissertation summarizes the fixed effects in the regression model for four types of services – online searches for healthcare-related information, reviewing one's health status, booking an appointment with a personal doctor or specialist, and consulting a doctor online. For most of the independent variables, the findings from the general model (Table 2.2 in the thesis) find confirmation in the model for individual types of health services.

The exception is the variable for respondents' professional status. Compared to employed respondents, the model shows a higher predicted probability of obtaining health information online and having an online consultation with a doctor for retired respondents.

Table A.4 in the Appendix summarizes the posterior distributions of the regional effects. As in the previous analyses, the pattern shows strong regional heterogeneity in the propensity to use e-services in education, as well as a positive correlation between the regional effects.

2.4 Chapter 3: Forms of Non-participation in the local elections 2019

This chapter of the thesis builds on the results in Amarov (2021b). It examines the 2019 municipal council elections and the two rounds of municipal mayoral elections, which are the first local elections in Bulgaria in which the electoral law allows voters to explicitly indicate that they do not support any of the candidates (NOTA). Few countries allow such an option on the ballot, including Ukraine, Belarus, the US state of Nevada, India, Bangladesh, Spain, and Indonesia. For this reason, there is relatively little research on its usage – (Alvarez, Kiewiet & Núñez 2018; Damore, Waters & Bowler 2012) show a pronounced protest nature of NOTA voting, but also indications of low awareness of voters

using this option. This chapter contributes to the research on invalid (blank, spoilt, or spoiled ballots) and NOTA voting.

Because the secrecy of the vote prevents voter behavior from being directly observable, empirical studies of invalid voting have relied primarily on direct voter surveys (Cohen 2017; Driscoll & Nelson 2014; Singh 2017) or on analyses of aggregate data summarized at various levels: States, electoral or administrative districts and polling stations (Fossati & Martinez i Coma 2020; Power & Garand 2007; Uggla 2008).

Direct surveys allow for the study of individual-level differences in voter behavior and the variation in invalid voting across a wide range of respondents' socioeconomic and demographic characteristics, political engagement and awareness, and attitudes toward political parties. Despite these advantages, direct surveys suffer from two main disadvantages. First, they cannot capture invalid voting caused by unintentional voter errors. On the other hand, direct polls potentially suffer from over-reporting of voter participation (Belli, Traugott & Beckmann 2001) and valid voting (Dejaeghere & Vanhoutte 2015).

In addition to enabling analyses of unintentional voter behavior, aggregate data analysis has the advantage that, in many cases, election data are available at different levels of aggregation that can be linked to data on socioeconomic characteristics at a regional level. This study uses data from the Central Electoral Commission (CEC) in Bulgaria, which provides detailed polling station-level data on the number of registered voters, the number of invalid and NAP votes, and the number of votes for parties and candidates who appeared in the elections. These data allow for the analysis of the variation of invalid and NPV votes with respect to characteristics of the political and institutional environment, such as the contestability of elections, the degree of political fragmentation, the number of seats in a constituency (municipality), and the number of simultaneous elections (Aldashev & Mastrobuoni 2016).

To allow for a broader analysis of associations between forms of non-voting and voter characteristics, the present study links CEC data with information on the age, education, and ethnic structure of localities in Bulgaria available from the 2011 census. In addition, the analysis uses municipality-level data from the employment agency on the level of registered unemployment and poverty rate estimates provided by the National Institute of Statistics (NIS).

The analysis uses a two-level multinomial regression model for the three forms of non-participation: abstention, invalid voting or NOTA voting. The main results of the statistical modeling show a higher tendency towards protest (NOTA) voting in localities with a high concentration of educated residents and a lower predicted probability of invalid

voting in these localities. The analysis supports more pronounced protest voting in constituencies with low competition between political parties and candidates. The model also examines the relationship between the shares of invalid ballots in contested districts and districts with a clear winner, explored by Aldashev & Mastrobuoni (2016) using data from Italian parliamentary elections. The results indicate a higher probability of invalid balloting in polling stations with a large vote difference between the leading and the second candidate/party, but only in contested constituencies.

2.4.1 Methodology

According to the CEC, a total of 3 077 684 people or 49.8 percent of registered voters voted in the October 27, 2019 municial councils election. The voting took place in all 265 municipalities in 11 845 polling stations located in 4 210 localities. Elections for municipal mayors were held in the same polling places, and the number of citizens who voted was 3 085 199 or 49.8 percent of all registered voters. These data do not include 39 mobile polling stations or polling stations in jails and hospitals. The total number of voters in excluded precincts is 1 635. One polling station (Dedintsi village) is also excluded from the analysis as there is no data for the locality in the 2011 census, which is used in the analysis.

The final election results included 82.0 percent of all votes in the municipal council election and 92.9 percent in the mayoral election. In the run-off election, the votes counted amounted to 95.2 percent of all votes. Figure 3.1 in the dissertation shows the distribution of uncounted vote shares for the three elections.

Of all ballots excluded from the final result, the NOTA votes represent 16.3 percent of uncounted votes in the municipal council elections and 41.6 percent in the mayoral elections. In contrast to the first round of mayoral elections, where NOTA had a lower share of uncounted ballots, NOTA accounts for the majority of uncounted ballots in the second round. The share of NOTA voters remained largely unchanged between the two rounds of the mayoral election – 3.0 percent in the first round and 3.1 percent in the second round. At the same time, the proportion of invalid ballots was significantly lower in the second round (1.7 percent) than in the first round (4.1 percent).

The literature review (Section 3.2 in the dissertation) points to a selection of independent variables from three broad groups that can explain the variation in the shares of the three forms of nonparticipation. These are socio-economic and demographic characteristics and variables describing the political environment and institutional framework

of elections.

Of the institutional characteristics, the number of simultaneous elections, the number of seats, and the degree of disproportionality of constituencies (for municipal council elections) are relevant to local elections in Bulgaria. The other institutional variable relevant for the first round of local elections is the number of elections held simultaneously. In local elections in Bulgaria, this number is determined by the number of inhabitants in the locality: citizens in localities with less than 600 inhabitants elect a municipal council and a municipal mayor. At the same time, citizens in larger localities also elect a mayor (or district mayors in Sofia, Plovdiv, and Varna).

The number of seats in the municipal councils is determined by law according to the population of the municipality in 2007. It varies between 11 in municipalities up to 5 000 inhabitants and 61 in the Capital City Municipality. The (Gallagher 1991) index is the most commonly used measure of disproportionality between votes and seats received. This index is a modified version of the root mean square deviation between the proportion of votes received and the proportion of seats received.

The index equals zero when there is complete correspondence between vote shares and mandate shares and increases with the degree of disproportionality. In the elections for municipal councils, the indicator ranges between 0.01 and 0.19, with an average value of 0.06 (Table A.7 in the Appendix).

There is no uniform approach to operationalizing the abstract idea of electoral competition in the literature. The most commonly used are the difference between the ex-post results of the top two political parties/candidates (margin of victory) (Uggla 2008; Aldashev & Mastrobuoni 2016) or the vote share of the largest party (Fatke & Heinsohn 2017). Both approaches rest on the assumption that voters' expectations of the closeness of elections are close to realized outcomes. This study follows the approach in Aldashev & Mastrobuoni (2016) and includes in the model the margin of victory at the municipality (constituency) level, the margin of victory at the polling station level, and an interaction effect between them.

Regarding the operationalization of the level of political fragmentation, there are two commonly used approaches. While some authors use the number of parties/candidates that took part in the election (Aldashev & Mastrobuoni 2016), others use the effective number of parties/candidates (Laakso & Taagepera 1979) based on the actual votes received by the parties. A preliminary examination of the data showed a better fit of the model when using the absolute number of parties or candidates, as well as a more straightforward interpretation of the results, as the effective number of parties is highly

correlated with the margin of victory, a key explanatory variable in the model.

The Heinsohn (2018) highlights the limited choice between political representatives as one of the drivers of invalid voting in French presidential elections. The model for the first round of mayoral elections includes the share of voters who voted for political parties without a mayoral candidate of their own. For the second round of mayoral elections, a natural measure of this restriction is the percentage of voters who voted for candidates who dropped out in the first round. In the model for the municipal elections, a variable measuring the share of voters who voted on the same day for mayoral candidates without support from political parties also competing in the municipal council elections is included.

To measure the degree of mobilization by political elites, (Fossati & Martinez i Coma 2020) use the share of ballots with a marked preference for a candidate. The present study uses a similar operationalization but counts only preferences for candidates other than the party list leader since, in the 2019 local elections, ballots without a marked preference automatically counted as a preference for the list leader. A sensitivity analysis of the models shows no significant difference in using the share of all preferences.

A literature review shows a variety of practices regarding the choice of socio-demographic variables, especially in analyses of aggregate data. A commonly used measure of the degree of economic development is gross domestic product (GDP) or its change in comparative analyses at the state or large constituency level (Fossati & Martinez i Coma 2020; Kostadinova 2003). Instead of GDP per capita, Fatke & Heinsohn (2017) use tax receipts at the municipality level as a measure of regional economic development. In Bulgaria, GDP data is publicly available at the regional level but not for individual municipalities or localities. Preliminary analysis showed no relationship between GDP per capita, GDP growth over time in the last mandate, and the forms of nonparticipation. One possible reason for this is the high level of aggregation (28 districts). At municipality level NSI provides estimates of the shares of residents with incomes below the poverty line. The National Institute of Statistics provides poverty rate estimates based on the 2011 Census and the 2012 EU-SILC (Ryustem, Dimitrova, Kostova, *et al.* 2018). The unemployment rate is another measure of economic development used in a number of studies of non-voting and invalid voting (Uggla 2008; Fatke & Heinsohn 2017). For Bulgaria, data on registered unemployed are available at the municipality level from the national employment agency.

In an analysis of a presidential election in Indonesia, Fossati & Martinez i Coma (2020) used an index of ethnic homogeneity, while in other studies, the authors included shares

of different minority groups. A preliminary analysis of the data showed no significant non-linear relationship between ethnic group shares and forms of nonparticipation. For this reason, the present study uses the shares of the two major ethnic groups – ethnic Turks (8.8 percent of the total population) and Roma (4.9 percent) from 2011 census data.

The 2011 Census also provides information at the locality (town/village) level on ethnic, age, and educational composition. As a measure of citizens' education level, the model uses the proportion of residents with at least a high school diploma at the locality level.

For each of the three choices, the probabilities of not voting, not voting, or NOTA voting are modeled within a two-stage binomial regression model with varying constants at the municipality level. In addition, the regression coefficients for the share of preferences used at the polling station level also vary across municipalities. A full description of the model is contained in Model 3.1 in the dissertation.

2.4.2 Results

The posterior distributions of the fixed effects in regression models are summarized by their means and 95 percent central credible intervals in Table 3.1 for the municipal council elections and Tables 3.2 and 3.3 for the first and second rounds of the municipal mayoral elections. The results show associations between the forms of nonparticipation and variables from all three sets of characteristics – institutional, political, and societal.

2.4.2.1 Institutional and political characteristics The models for all three elections show a complex association between the strength of political competition, abstention, invalid voting, and NOTA voting. All three models contain an interaction effect between the margins of victory at the polling station and municipality (electoral district) levels. The posterior mean of the coefficient of this interaction effect is negative in all models except in the equation for abstention in the second round of local elections, where the mean is positive. However, the coefficient is small and close to zero in absolute value. The predicted probabilities depending on the margin of victory at the polling station level are plotted in Figure 3.3 for the first round of local elections, Figure 3.4 for the second round, and Figure 3.5 for the municipal council election. In all figures, the predicted probabilities are plotted for two levels of political competition at the municipal level – high (first quartile of the margin of victory distribution) and low (third quartile).

For the second round of local elections, the model shows a trend toward more frequent invalid voting for high margins of victory (low competition) at the polling station for

all levels of competition at the municipal level. However, the predicted probability increases more rapidly in highly contested municipalities. In the first round of municipal mayoral elections, the probability of an invalid vote also increases with an increasing margin of victory at the polling station level, but only in municipalities with a low level of competition. In municipalities with a high level of competition, the model shows a trend towards a lower probability of an invalid vote with a higher margin of victory at the polling station level. The results for municipal council elections are similar to those for the first round of mayoral elections. The current results differ from the findings of Aldashev & Mastrobuoni (2016), who find declining observation-level proportions of invalid votes with a margin of victory in polling stations with low competition. One possible reason for this is the different levels of data aggregation in the two studies.

The predicted probability for NOTA voting declines with the margin of victory at the polling station level in all three elections, with the decline being stronger in municipalities with low levels of municipal-level competition. The models show no systematic variation in turnout by the level of competition at the polling station level except in the municipal council elections. In these elections, the predicted probability of abstention grows with decreasing competition, and the growth is stronger in municipalities with high political competition.

The regression model for the first round of mayoral elections includes an interaction effect between the margin of victory at the municipal level and an indicator for the presence of a runoff in the same election, which identifies municipalities where the winner achieved less than half the votes. The predicted probabilities for the forms of nonparticipation in Figure 3.6 show the probability of invalid and NPV voting increasing with the margin of victory in municipalities with a first-round winner. For mayoral elections decided in a second round, the model shows no association between the level of competition and NPV voting and a weak effect of decreasing invalid voting.

In municipal council elections, the model shows a lower predicted probability of invalid voting for a larger number of political parties participating in the electoral race while revealing no systematic variation in the probabilities of abstention and NOTA voting. This result is consistent with using invalid voting to protest against limited political choices. In contrast, in the first round of mayoral elections, invalid voting shows no systematic variation with respect to the number of candidates. However, the probability of NOTA voting decreases with the number of candidates.

Contrary to expectations, polling stations with a high proportion of voters who voted for non-politically endorsed mayoral candidates show a higher probability of invalid

votes and a lower probability of NOTA votes in the municipal council elections. At the same time, in the first round of local elections, the model shows a trend towards a lower probability of invalid voting for a high share of voters who supported parties without their own candidates for mayor.

The model for the second round of mayoral elections shows strong evidence of a higher probability of all forms of nonparticipation, especially of NOTA, with a high proportion voting for candidates who dropped out in the first round, a result close to the findings of Heinsohn (2018).

The model shows a trend towards higher turnout, a lower predicted probability of NOTA voting (in council elections), and higher turnout in polling stations with a high proportion of voters using preferential voting in the council elections. This result holds for the municipal council elections and the first round of mayoral elections and partially supports the findings of Fossati & Martinez i Coma (2020). At the same time, the standard deviations of the municipality-by-municipality variation in the coefficients on preferential voting have posterior distributions far from zero and point to considerable heterogeneity in the strength and direction of this association across municipalities (Tables 3.4 and 3.5 and Figure A.1).

The correlation coefficients between the municipality-level constants and the coefficients on the share of preferences used are negative and close to -0.9 in all models (Tables 3.7 and 3.8). The negative correlation shows a tendency towards a strong negative association between forms of nonparticipation and preference use in municipalities with a high propensity for nonparticipation.

The correlation coefficients between the municipality-level constants for the different equations are positive, consistent with the hypothesis that voters view all forms of nonparticipation as partially exchangeable forms of behavior (Tables 3.7, 3.8, and 3.9). At the locality level, the correlations between the constants are also positive, albeit lower.

A significant proportion of the variation in the probabilities of all forms of nonparticipation in elections is associated with the number of residents in the locality. Figures 3.8 and A.2 (in the Appendix) visualize the predicted probabilities according to the number of inhabitants in the locality at the mean of all other model variables. At both choices, the probability of a valid vote decreases with the size of the locality, but the decrease is stronger in localities with more than 600 inhabitants. This decrease is mainly due to lower turnout in larger settlements. While in small localities, both types of uncounted votes increase with the number of inhabitants of the settlement, in settlements with more than 600 inhabitants, the probabilities of these two forms of nonparticipation de-

crease with population.

The coefficients on the share of educated residents show similar effects for all three types of elections. While the predicted probabilities of invalid voting are lower in localities with a higher share of educated citizens, the predicted probabilities of protest voting are higher. This result suggests a higher willingness of educated citizens to protest vote (NOTA) and that at least some invalid ballots result from errors that low-educated voters make more often. A high concentration of educated residents is also associated with higher turnout in all three elections.

2.5   Chapter 4: Invalid ballot papers and machine voting

The first objective of the fourth chapter of the dissertation is to examine the effect of machine voting on uncounted votes (Chapter 3) using polling station-level data from the 2019 EP elections. In these elections, around 25.7 percent of voters voted by machine in the machine-equipped polling stations. Because machines are allocated to polling stations in a quasi-random fashion, the 2019 EP elections allow us to estimate the effect of machine availability on uncounted votes and votes for marginal parties. The analysis uses a two-stage multinomial regression model for the probabilities of not voting at the polling station level. The model also includes independent variables describing socioeconomic and demographic characteristics at the municipality and locality levels potentially associated with voter behavior in elections (Chapter 3).

A second objective of the analysis is to examine the variation in the types of uncounted and marginal party votes relative to the percentage of voters who chose to use a machine. This association is modeled within a two-stage multinomial regression model using data from machine polling stations in the 2019 European Parliamentary and April 2021 general elections.

2.5.1   Methodology

This study uses data from the 2019 EP elections, where 2 064 192 voters (32.9 turnout rates) voted in 11 873 polling stations. EVMs were deployed in 2 999 of these polling stations, and voters could decide whether to use the machine or vote by paper ballot.

In the literature, there is no unified approach to operationalizing the abstract idea of a marginal party. While Uggla (2008) considers non-parliamentary parties, Desai & Lee (2021) test different thresholds between 2.5 and 10 percent, reaching similar conclusions

in all scenarios. In this study, parties or coalitions that achieved less than two percent of the total votes cast are considered marginal. By comparison, the smallest (according to the election result) party to receive a seat in the EP in these elections was Democratic Bulgaria (DB), which received 6.06 percent of all votes. A sensitivity analysis of the results showed the robustness of the main findings to changes in the definition of a marginal party, for example, using half-a-percent or one-percent thresholds.

Figure 4.2 shows the distribution of uncounted votes and votes for marginal parties by type of polling station. The proportion of invalid ballots is lower in polling stations with machine voting, while the share of NOTA votes is almost unchanged. At the same time, the share of votes for marginal parties is significantly higher in machine polling stations. Overall for all machine polling stations, the percentage of votes cast for non-marginal parties is 83.0, while for paper ballot-only polling stations, this share is almost identical: 83.1 percent.

This observation supports the hypothesis that voters who choose to vote with a machine that does not allow blank or spoiled ballots tend to cast NOTA or marginal party votes as a substitute for blank or spoiled ballots.

The fixed effects in the model include independent variables potentially associated with voter behavior, motivated in Section 3.3 of the dissertation: the number of residents in a locality according to reports of National Statistical Institute, shares of residents aged under 30 and over 65, the shares of ethnic Turks and Roma, and the shares of educated residents (secondary or higher education) according to the 2011 census data. At the municipality level, the share of unemployed (employment agency) and estimates of the shares of residents with incomes below the poverty line are included. In addition, the model includes control variables at the polling station level for the shares of the main political parties participating in the elections: GERB, BSP, MRF, and Democratic Bulgaria, because parties' different attitudes towards machine voting are potentially related to the extent of machine use. The model does not include the institutional and political variables discussed in Chapter 3, as they are irrelevant to the EP elections. All numerical variables are standardized to a mean of zero and a standard deviation equal to one.

Following the methodology in Chapter 3, the models include varying effects at the municipality and locality level, which capture unmeasured heterogeneity due to common attitudes, local political mobilization, and potential differences in electoral organization, such as the preparation of polling stations. The model for the 2019 EP elections includes a variable indicating the presence of a machine in the polling station and coefficients varying across municipalities for this variable. The second model for the same election

is estimated only on EP election data from polling stations with machines and includes a spline function with a thin plate basis (Wood 2003) for the fraction using machine voting instead of an indicator of machine presence. The same specification is followed in the model for the April 2021 parliamentary election, estimated with the data for polling stations with machine voting.

2.5.2 Results

Table 4.1 summarizes the posterior distribution of the regression model's fixed coefficients. Figure 4.3 shows the differences in the predicted probabilities of the forms of nonparticipation and 95 percent credible intervals according to the presence or absence of a machine in the section. The predicted probability of an invalid vote is significantly lower in sections with machines present, and the credible intervals do not overlap. At the same time, the probability of voting for marginal parties is higher in these polling stations. The differences in voting probabilities with NOTA are smaller than in polling stations without machines. The net effect on the probability of a valid vote for non-marginal parties is positive, with a posterior probability of being greater than zero equal to approximately 95 percent. This result is similar to the findings of Desai & Lee (2021), who find a trend towards redistribution from invalid voting to explicit forms of protest voting when machine voting is introduced.

The results for the control variables in the model are similar to those for local elections in Section 3.4. The concentration of educated residents in the locality is the independent numerical variable with the largest coefficient in the equations for NOTA and invalid ballots. A higher proportion of educated residents is associated with a higher predicted probability of NOTA votes and marginal parties. As in the analysis of the 2019 local elections, the model shows no evidence of an association between the share of educated residents and turnout.

As in the 2019 local elections, polling stations in larger localities show a higher probability of nonparticipation. Votes for marginal parties are more likely in large localities than in small ones. At the same time, the variation in proportions of NOTA and invalid ballots does not appear to be related to locality size.

Contrary to results from previous studies showing lower voter turnout among younger citizens (Blais & Dobrzynska 1998; Blais 2006), the concentration of residents under the age of 30 does not indicate a lower predicted probability of nonparticipation in elections. The results for the age structure of localities are also similar to those from local elections

in the same year. The concentration of elderly residents (over 65) in the locality is associated with a higher predicted probability of invalid voting, a possible indicator of difficulties in handling paper ballots by older voters. In addition, the model shows a small positive effect on the probability of voting for marginal parties in localities with a higher proportion of older residents.

The proportion of Roma among residents at the locality level shows no association with nonparticipation in elections. The posterior mean of the coefficient in the model for invalid voting is positive, with positive bounds at the 95 percent credible interval, but the effect is small. For the ethnic Turkish population, the results indicate higher turnout and lower probabilities for all forms of nonparticipation, as do the analysis results of the 2019 local elections.

Unlike other studies that find a positive association between unemployment and invalid/protest voting, the model shows lower probabilities for all forms of nonparticipation in municipalities with higher unemployment rates. Holding unemployment rates (and the other variables in the model) fixed, polling stations in municipalities with a higher proportion of people below the poverty line have a higher predicted probability of an invalid vote. This result is close to the findings from the study of blank ballots in Ontario, Canada elections (Galatas 2008).

Like local election results, the model shows evidence of higher turnout and lower probability of invalid voting in polling stations with high proportions of preferential voting. At the same time, the coefficients on the fraction of ballots with marked preferences are virtually zero in the marginal and NOTA voting equations.

The model in Table 4.1 only includes an indicator for the presence of a machine in the polling station but does not account for how many voters used the machine. In order to analyze the association between the fraction who voted by machine and the probabilities of nonparticipation, a separate model is fitted only for polling stations with a machine. The model contains the same independent variables as the original model but examines only four types of voting: votes for a conventional party, votes for a marginal party, and invalid and NPV votes, with votes for a non-marginal party serving as the reference category. Instead of an indicator for the presence of an EVM in the polling station, the model includes polling station-level turnout and a smooth function (cubic regression spline) of the percentage who voted with an EVM.

Figure 4.4 visualizes the predicted probabilities of the four types of voting as a function of the fraction of EVM votes. While the predicted probability of an invalid vote decreases monotonically for all values of the percent voting by machine, the share of NOTA votes

reaches a maximum at about 11 percent EVM votes. After that, it remains roughly constant and even shows a decreasing trend. For voting for marginal parties, the model also shows a sharp increase in the probability of voting for these parties, up to about 11 percent of those voting by EVM, and at higher EVM, shares increase more slowly. The posterior probability of marginal voting at 75 percent machine use being more than 0.02 higher relative to the probability of marginal voting at 0.25 percent machine use is only 0.74. Valid voting for non-marginal parties appears significantly higher in polling stations where almost all voters voted by EVM. This result should be viewed cautiously as machine voting use by more than 75 percent of voters occurred in only one percent of the polling stations.

The same model applied to data from the April 2021 general election shows a similar result, with an initial strong decrease in invalid ballots at the expense of an increasing likelihood of NOTA and marginal parties' votes (Figure 4.5). In contrast to the 2019 EP elections, the average probability of voting for marginal parties is about twice as low. A possible explanation for this result is a lower propensity of voters to support marginal parties in important elections. The fixed coefficients in the regression model are presented in Table A.9 in the Appendix.

2.6 Chapter 5: Selectivity of electronic voting machines usage

This chapter of the thesis continues the research in Amarov (2021a) and analyses two aspects of the introduction of machine voting in Bulgaria. The first objective of the study is to assess the selectivity of electronic machine voting by examining associations between the propensity to use machine voting and the socioeconomic and demographic characteristics of voters. In addition, the study assesses the effect of previous experience with machine voting on the propensity to use machines. The analysis uses data from the April 2021 general election and the 2019 EP election at the polling station and 2011 census data at the locality and municipality levels.

2.6.1 Methodology

At the April 2021 general election in Bulgaria, a total of 3 136 129 voters voted in 11 950 polling stations across the country. In 9 396 polling stations with more than three hundred registered voters, voters can vote by paper ballot or by machine.

During election day (4.4.2021), machine voting was suspended in two of the 31 voting districts – Veliko Tarnovo and Dobrich, due to software errors in the machines. For this

reason, 559 machine polling stations with a total of 159 550 voters are excluded from the analysis. The remaining 8 837 machine polling stations had a total of 2 776 091 voters or 93.8 of all voters. Of the voters who voted at machine polling stations, a total of 770 002 or 27.7 percent cast their vote by machine.

The average (by polling station) share who voted by machine is 26.4 percent, and the distribution of shares by precinct shows a strong variation between 0.0 and 100.0 percent. In the central 50 percent of the sections, the proportion varies between 13.9 and 38.1 percent (Figure 5.1). The distribution of shares shows a pronounced bimodality with a large number of polling stations with very low machine use – in 11.3 percent of machine polling stations, less than 3 percent of votes were cast by machine.

The average machine usage rate is similar to experimental machine voting in the 2016 presidential and 2019 EP elections. In the presidential election, on average, 27.4 percent of voters in machine polling stations took advantage of this opportunity.

In the 2019 EP election, 3 000 of total 11 873 polling stations are equipped with machines and 616 994 voters vote in these stations (29.9 percent of all voters). Of the voters at machine polling stations, 25.7 percent choose to cast their vote by machine. The average proportion of those voting by machine is 26.0, with the proportions of machine use for the central half of polling stations varying between 11.7 and 37.7 percent (Figure 5.1).

The literature review in Section 5.2 points to socioeconomic and demographic variables potentially associated with preferences for choice technology: age, level of education, degree of urbanization, and ethnic minority membership. This analysis uses the same operationalizations as the invalid voting study in Chapter 3.

Like the model in Chapter 3, the regression model for the probability of machine voting includes fixed effects for socioeconomic and demographic characteristics. In addition to these, we include the party preferences of voters at the polling station level relative to the four main parties and coalitions appearing in both elections: the GERB, the BSP, the MRF, and the DB. The addition of these variables aims to account for differences in the attitudes of supporters of these parties toward machine voting (Alvarez, Levin & Li 2018). Because machine voting preferences are potentially related to voters' political education and culture, the model also includes the percentage of ballots with a marked preference as a proxy measure (Santana Pereira & Costa Lobo 2020).

The second objective of this analysis aims to examine the association between previous experience with machine voting and the propensity to machine vote in future elections. To this end, the model for the 2021 election contains an indicator for the presence of machine voting in the 2019 EP election in the same polling station.

The municipality- and locality-level varying effects aim to account for unmeasured heterogeneity in attitudes at the local level and serve as municipality- and locality-level estimates of preferences for machine voting that are adjusted for differences in socioeconomic, demographic, and political characteristics (the fixed effects in the model). Because election administration can have a significant impact on how voters perceive new voting technology (Alvarez, Katz & Pomares 2011; Alvarez, Levin & Li 2018), these estimates can be used to identify problems in election organization at the municipality and locality level.

A preliminary analysis of the shares of votes cast by machine in April 2021 at the municipality level shows substantial variation in the regression coefficients for the percentage of votes cast by preference and the indicator for machine use in 2019. For this reason, the model allows the coefficients for the share of preferences used and the availability of machine voting in 2019 to vary across municipalities.

All numerical independent variables in the model are standardized to zero mean and standard deviation equal to one. The model uses weakly informative normal a priori distributions centered at zero and with standard deviations equal to one. For mean values of the numerical independent variables, the model expects machine utilization ranging between 4 and 73 percent with 95 percent probability.

2.6.2 Results

Table 5.1 summarizes the posterior distribution of the fixed effects in the models. Both models show high Bayesian R^2 (Gelman, Goodrich, Gabry, *et al.* 2019) values of 0.68 (EP 2019) and 0.83 (April 2021 general election).

The posterior means of the coefficients of locality size and share of educated residents are positive for both models, indicating a higher predicted probability of machine voting in large settlements and in cities/villages with a higher concentration of educated residents. These results are similar to the findings of Roseman & Stephenson (2005) and support the hypothesis of a higher propensity to use new technologies by educated citizens in urbanized regions, including new voting technologies (Alvarez, Katz & Pomares 2011). These results do not rule out the hypothesis that some highly educated citizens may be critical of the lower transparency of machine voting and have stronger concerns about the secrecy of their votes. However, they show that at the aggregate level in Bulgaria, the net association between the probability of voting by machine, education, and the size of the locality is positive.

The posterior means of the coefficients on education and number of inhabitants per locality in the model for the 2019 EP elections match those from the 2021 parliamentary elections. In the EP 2019 model, the posterior credible intervals are wider, partly reflecting the smaller number of observations, while the credible intervals in the 2021 regular parliamentary election have strictly positive bounds.

The coefficients for the concentration of Roma and ethnic Turks in the locality have negative a posteriori means and confidence intervals far from zero, with the coefficient for the concentration of ethnic Turks being larger. This result indicates a tendency of the ethnic Turkish population to avoid machine voting, even when the model controls for the effect of the share of votes cast for the MRF. This party traditionally receives strong support in areas with ethnic Turkish populations.

Voters' political preferences at the polling station level are strongly associated with machine voting, with the strongest association observed between the share voting for the MRF (Figure 5.2). The predicted probability of voting by machine in a polling station with no votes for the MRF is 0.33 (at the means of the numerical independent variables and with no previous experience of machine voting), while this probability is 0.26 with 10 percent of votes cast for the MRF. In polling stations with a majority for the MRF, the predicted probability is less than 0.1. The model also shows evidence of a lower probability of machine voting in polling stations with strong support for GERB and BSP, but the effect is significantly smaller. Polling stations with high support for DB show a higher probability of machine voting, partly explained by positive messages from this coalition on machine voting. The results for differences in the predicted probability shares of parties should be viewed with a degree of caution, as in addition to party preferences, these shares may also account for structural differences at the polling station level that are not included in the model.

Since the shares of valid votes of all electoral participants must sum to one hundred percent, the political party shares in the models are negatively correlated. A sensitivity analysis of the model shows the robustness of the results when the shares of individual parties/coalitions are excluded from the models.

One argument against compulsory machine voting is the alleged difficulty of the elderly to use the machines. The model shows a weak negative association between the concentration of elderly residents (65+) in a locality and the predicted probability of machine voting (Figure 5.3). This result should be viewed with caution, as in addition to the age of residents in the locality, this variable may reflect structural characteristics that are not included in the model.

The effect of previous experience with machine voting is positive, with a posterior mean for the entire population of municipalities of 0.19 and a 95 percent confidence interval between 0.12 and 0.25. At the same time, the association between the presence of machine voting experience and the probability of voting in the next election varies significantly across municipalities in Bulgaria (Figure A.3). While for most municipalities, the effect of having machine experience is associated with a higher propensity to vote by machine in the April 2021 general election, this association is even negative in some municipalities. This result may inform further research, particularly on aspects of election organization that potentially discouraged some voters from voting by machine (Alvarez, Levin & Li 2018).

As well as the effect of having experience with machine voting, the coefficients on the percentage of ballots with a marked preference varied considerably between municipalities in Bulgaria in both the 2021 and 2019 parliamentary elections. A.4 in the Appendix shows the averages of the coefficients by the municipality in the 2019 parliamentary elections.

While in some municipalities, more widespread use of preferential voting is associated with a stronger propensity to use machine voting, this association is negative in other municipalities. This result may inform subsequent research on the nature of voter mobilization in different municipalities.

2.7 Chapter 6: Turnout and machine voting in 2021

The dissertation's final chapter examines the variation in turnout differences in the July and November 2021 general elections compared to April 2021 across socioeconomic and voter demographic characteristics. The study uses polling station-level data from the three 2021 general elections in Bulgaria and models the difference in turnout within a two-stage linear regression model. The analysis uses the proportion who voted by machine in the April 2021 election, when voters are free to choose the form of voting, as a measure of voter willingness to adopt the new technology.

2.7.1 Methodology

In the April 2021 general election, 11 950 polling stations were equipped with a machine. These polling stations (with at least three hundred registered voters) represent about 79 percent of all polling stations and include about 93 percent of the country's voters.

For this study, we compare polling stations from all three elections (April, July, and November 2021). The three elections take place over a short period, and there are relatively few changes in polling stations, allowing polling stations to be linked into a panel including a total of 8 257 polling stations with data for all three elections. This number does not include the polling stations in Dobrich and Veliko Tarnovo, since in the April 2021 elections, the machine voting there was suspended during the election day for technical reasons.

The distribution of differences in turnout in July and November compared to April 2021 at the polling station level is depicted in Figure 6.2. In polling stations with machine voting in all three elections, turnout was, on average, -8.9 percentage points lower in July compared to April and by an average of -11.8 percentage points in November compared to April. By comparison, turnout declined, albeit less in polling stations without machine voting. In these polling stations, turnout fell by -4.7 and -6.7 percentage points in July and November, respectively, compared to April.

An initial analysis of the April 2021 electronic machine use and turnout differences in the next two elections shows a positive association (Figure 6.3) between the turnout difference and the share of voters who voted by machine in the first election.

This study aims to measure the association between voters' propensity to vote by machine as a proxy for adopting the new voting technology at the polling station level. In the event that voter reluctance to vote by machine is associated with voter attrition in the next two elections, we expect to observe a positive association between the proportion who voted by machine in April and the turnout gap in the next election.

Turnout differences are modeled within a two-stage linear regression model with coefficients varying at the municipality and locality levels. Due to the presence of a significant number of polling stations with unusually large differences in turnout (over 50 percentage points difference), the model uses a t-distribution as a likelihood function. A full description of the model is found in Model 6.1 in the dissertation.

2.7.2 Results

Table 6.1 shows the posterior means and 95 percent central credible intervals for the fixed effects of the regression models. Because of the standardization of the target and independent variables, the regression coefficients show the expected difference in standard deviations of the target variables given a difference of one standard deviation of the independent variable with the other independent variables fixed, including the varying

effects in the model.

The main variable in the model is the proportion of machine votes in the April 2021 election. Its coefficient has positive posterior means with credible intervals bounded away from zero. This is evidence of a positive association between the propensity to use machines in voluntary elections and the turnout gap. A difference of one standard deviation in the machine votes in April 2021 (1 sd = 15 percentage points) is associated with an expected increase of only 0.06 standard deviations in the turnout difference (1 sd = 11 percentage points) for the November election. The hypothesis that precincts with more than a 60 percentage point (4 standard deviations) difference in machine use have an expected turnout difference of less than three percentage points has an a posteriori probability virtually equal to 1 (July) and 0.78 (November).

The posterior means of the coefficients on the control variables in the regression model show a significant variation of turnout differences with respect to voters' policy preferences by April 2021. The decline in voter turnout is smaller in urban areas than in rural areas, with larger localities showing a trend toward smaller declines in November but not in the July election. A possible explanation for the lack of association of July elections is seasonal migration related to summer vacations.

The decline in turnout tends to be stronger in polling stations with high support for the ruling (at the time of the April elections) GERB party. One possible explanation for this lower turnout is potential disillusionment among supporters of this party as it failed to form a government despite winning the April 2021 elections. A second possible explanation for this effect is a higher distrust among GERB voters of machine voting, which is not measured in the share who used machine voting in April. Although statistically significant, this effect is also small.

In contrast, the differences in turnout are smaller in districts with higher support for two opposition parties: the BSP and the DB, which may indicate mobilization among their supporters. For the DB, this effect was stronger in the November elections.

Polling stations in settlements with a high concentration of educated residents show a trend towards a smaller decline in turnout in the July elections. However, the model shows about twice as small a coefficient for the next parliamentary elections. The decline in turnout is lower in localities with a larger share of the ethnic Turkish population. In contrast, the share of Roma at the locality level shows no association with the change in turnout.

Despite concerns that older citizens would refuse to go to vote if they had to use a machine compulsorily, the model shows no evidence of a relationship between the con-

centration of older (65+) residents and turnout differences in either election. A small negative effect is observed in the November election when localities with higher concentrations of younger residents show a slight trend toward higher declines in turnout.

The share of residents below the poverty line and the unemployment rate shows no evidence of systematic variation in turnout differences.

The overall explanatory power of the model is low, with a Bayesian R^2 of 0.11 (July) and 0.22 (November), indicating a low predictive power of the model. A comparison of the posterior distribution and the distribution of turnout differences in Figures A.5 and A.6 also shows a better fit to the November election data than to the July election data. The model overestimates the probability of extremely large negative turnout declines in both elections after April 2021.

3 Conclusions and contributions

This dissertation examines two aspects of societal transformation in Bulgaria within the digital revolution that began in the mid-20th century and led to the widespread adoption of electronic services after the 1990s.

The first is the introduction of e-government services and their use by Internet users in Bulgaria. The quality and functionality of the services offered have been the subject of public debate, but user demand for them has not been systematically studied. This dissertation fills this gap. Based on a survey of Internet users in Bulgaria, the dissertation investigates their propensity to use e-government services in administration, education, and health care depending on socioeconomic characteristics, ICT skills, and experience with Internet technologies.

Statistical modeling results indicate a strong association between the likelihood of using e-services according to their ICT skills and experience with Internet technologies. Internet users with low levels of ICT skills, low levels of education, and residents of small settlements show a lower propensity to use government e-services. These results describe a barrier to the use of e-services for some citizens who lack the necessary skills to benefit from the convenience of these services. Bulgaria is no exception to other countries in this respect (Bélanger & Carter 2009; van Deursen & van Dijk 2014; Hsu, Huang, Kinsman, *et al.* 2005).

Another result of the statistical analysis is estimates of the propensity of internet users to use e-services in the 28 administrative regions of the country. While the administration

can produce its estimates of the frequency of use of different services according to the users' location, this study provides estimates that consider the occupational status, ICT skills level, and education of users at the individual level. These estimates can target information and education campaigns to users in areas with a low propensity to consume government e-services. At the same time, they can serve as a basis for future research on the causes of regional heterogeneity.

The analysis draws attention to potential consumer problems with two health services – booking an appointment with a GP or specialist and checking health status. For two other services – online health information retrieval and online consultation with a doctor – the oldest users show a higher propensity to use them than younger users. The pattern does not show the same difference for the first two services, which is surprising since senior citizens consume health services more frequently than younger citizens, whether online or through traditional means.

The survey design imposes a major limitation on studying government e-service consumption among Internet users. Respondents answer the question of how often they use e-services, but it remains unclear from the data how often they consume government and municipal services in general (online or offline).

The second aspect of digital transformation looks at the introduction of electronic voting machines and election results reporting in Bulgaria between 2019 and 2022. The use of electronic voting machines sparked a widespread and intense public debate in 2021 and 2022, when the electoral code was amended twice – once in 2021 to make machine voting compulsory in most polling stations and again in late 2022 to allow voters to choose between machine voting and traditional paper ballots. One of the main arguments in the debate against compulsory e-voting in Bulgaria points to the sharp drop in voter turnout between the April 2021 parliamentary elections and the July and November elections of the same year and attributes this drop specifically to the use of electronic machines.

An analysis of the differences in turnout between July and April 2021 and between November and April 2021 shows a trend toward a lower decline in turnout in polling stations and a higher proportion of voluntary machine voting in April. At the same time, the model explains only a small part of the variation in turnout differences, and the effect is weak and insufficient to explain the decline in turnout.

Another aspect of the study of electronic voting machines examines the selectivity of their use among different socioeconomic and demographic groups, as well as differences in propensity to machine voting according to voters' political preferences. The analysis reveals a trend towards more frequent machine voting in localities with higher

proportions of educated residents, low proportions of ethnic minorities, and high support for Democratic Bulgaria (a political coalition); this result is between both the 2019 EP elections and the 2020 parliamentary elections. Voluntary use of the machines is less frequent in localities with a high concentration of elderly residents, but the effect is stronger in the parliamentary elections than for the EP.

In addition, an analysis of machine voting in April 2021 shows a higher likelihood of machine use in polling stations that had a machine in the EP election two years earlier. The analysis also shows that the relationship between experience with a machine in the polling station and propensity to machine vote is complex and varies strongly between municipalities. This result could serve as a starting point for future research on experiences with electronic machines and the reasons for variation across municipalities.

One of the arguments for and against mandatory electronic machine voting in elections is the assumption that machine voting would improve the representativeness of elections by removing the possibility of blank and erroneous ballots. An analysis of this argument uses data from the 2019 EP elections and the April 2021 parliamentary elections. In the EP elections, only a third of polling stations with more than three hundred registered voters were equipped with a machine. The allocation of machines to polling stations is not related to specific characteristics of polling stations and can be seen as a quasi-experimental design. A comparison of the proportions of invalid ballots, NOTA votes, and votes for marginal political parties between machine- and paper-only polling stations show a tendency of redistribution from invalid ballots to NOTA and marginal parties in machine-enabled polling stations. Despite the trend towards redistribution, the statistical model shows a net positive effect of machine availability on the probability of a valid vote for a non-marginal party.

Since the discussion for and against machine voting is largely related to the nature of invalid voting, the dissertation also includes an analysis of abstention, invalid, and NOTA voting in the 2019 local elections, when the share of uncounted ballots reached a record 18 percent. The analysis looks at three forms of nonparticipation in the 2019 local elections – non-voting, invalid voting, and NOTA voting. Statistical models reveal complex associations between forms of nonparticipation and political competition at the electoral district (municipality) and polling station levels. In municipalities with strong competition between political parties/mayoral candidates, the probability of invalid voting increases with decreasing competition at the polling station level. At the same time, the predicted probability of NOTA voting decreases with weakening competition at the polling station level.

One consistent result in the analyses of invalid and NOTA voting is that localities with high concentrations of educated residents are more likely to see NOTA votes and fewer invalid ballots. Although the data do not allow causal relationships to be inferred, this result is consistent with hypotheses of protest voting by educated and politically engaged citizens and unintentional errors by poorly educated voters (McAllister & Makkai 1993; Lundell & Högström 2021).

In addition, the analysis of local elections shows significant variation in the propensity for different forms of nonparticipation across municipalities in the country and a positive correlation between municipality-level effects and different forms of nonparticipation. The positive correlation indicates a tendency for voters to view the three forms of nonparticipation as mutually exchangeable.

Another study result is the association between the use of preferences and forms of invalid voting. The model finds evidence of a negative association between high shares of preferential voting and nonparticipation, while this association varies strongly across municipalities.

A major limitation of all analyses using aggregate data is that they do not allow direct inferences about individual voter behavior (Piantadosi, Byar & Green 1988). However, the results can stimulate further discussion and deeper investigation of voter experiences with machine voting and unconventional voter behavior in elections.

3.1   Publications related to the thesis

1. Amarov, B. and Netov, N. (2022) Usage of electronic public services in Bulgaria, *Journal of International Business Research and Marketing*, 7, pp. 36–41. Available at: https://doi.org/10.18775/jibrm.1849-8558.2015.73.3004.

2. Amarov, B. and Netov, N. (2023) Usage of electronic education services in Bulgaria, *International Journal of Management Science and Business Administration*, 9(2), pp. 30–36. Available at: https://doi.org/10.18775/ijmsba.1849-5664-5419.2014.92.1003.

3. Amarov, B. (2023) Electronic Voting Machines and Turnout in the Bulgarian 2021 Parliamentary Elections, in Societal Transformations and Sustainable Development with Respect to Environment in the Post COVID-19 Digital Era. First Annual Transform4Europe PhD Conference, 8-9 December 2021, Sofia, Bulgaria: St. Kliment Ohridski University Press, pp. 41–48.

3.2 Conferences

1. Amarov, B. (2021a) Adoption of electronic voting machines in the Bulgarian 2021 parliamentary elections. *In: the twenty-second faculty of economics and business administration, Sofia University "St. Kliment Ohridski" annual conference: 'Environmental, social and governance challenges for recovery and resilience'*, November 26-27, 2021.

2. Amarov, B. (2021b) Electronic voting machines and turnout in the Bulgarian 2021 parliamentary elections*, in Societal transformations and sustainable development with respect to environment in the post COVID-19 digital era.* 8th-9th December 2021.

3. Amarov, B. and Netoff, N. (2022a) 'Usage of electronic education services in Bulgaria', in 87th international scientific conference on economic and social development. Svistov, Bulgaria.

4. Amarov, B. and Netoff, N. (2022b) 'Usage of electronic public services in Bulgaria', in 87th international scientific conference on economic and social development. Svistov.

5. Amarov, B. (2021) 'Multilevel modeling of invalid voting rates in the Bulgarian local elections 2015 and 2019', in Applied modeling in economics, finance and social sciences, 28 June 28 - July 2, 2021. Sozopol, Bulgaria.

References

1. Aldashev, G. & Mastrobuoni, G. (2016). Invalid ballots and electoral competition. *Political Science Research and Methods*. 7 (2), 289–310. https://doi.org/10.1017/psrm.2016.36.

2. Alvarez, R.M., Katz, G. & Pomares, J. (2011). The Impact of New Technologies on Voter Confidence in Latin America: Evidence from E-Voting Experiments in Argentina and Colombia. *Journal of Information Technology & Politics*. 8 (2), 199–217. https://doi.org/10.1080/19331681.2011.559739.

3. Alvarez, R.M., Kiewiet, D.R. & Núñez, L. (2018). A taxonomy of protest voting. *Annual Review of Political Science*. 21 (1), 135–154. https://doi.org/10.1146/annurev-polisci-050517-120425.

4. Alvarez, R.M., Levin, I. & Li, Y. (2018). Fraud, convenience, and e-voting: How voting experience shapes opinions about voting technology. *Journal of Information Technology & Politics*. 15 (2), 94–105. https://doi.org/10.1080/19331681.2018.1460288.

5. Amarov, B. (2021a). Adoption of electronic voting machines in the Bulgarian 2021 parliamentary elections. In: *The twenty-second faculty of economics and business administration, Sofia University 'St. Kliment Ohridski' annual conference: "Environmental, social and governance challenges for recovery and resilience" ", November 26-27, 2021*. 2021

6. Amarov, B. (2021b). Multilevel modeling of invalid voting rates in the Bulgarian local elections 2015 and 2019. In: *Applied modeling in economics, finance and social sciences, 28 june -2 july 2021*. 2021 Sozopol, Bulgaria.

7. Bélanger, F. & Carter, L. (2009). The impact of the digital divide on e-government use. *Communications of the ACM*. 52 (4), 132–135. https://doi.org/10.1145/1498765.1498801.

8. Belli, R.F., Traugott, M.W. & Beckmann, M.N. (2001). What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics*. 17 (4), 479–498.

9. Blais, A. (2006). What affects voter turnout? *Annual Review of Political Science*. 9 (1), 111–125. https://doi.org/10.1146/annurev.polisci.9.070204.105121.

10. Blais, A. & Dobrzynska, A. (1998). Turnout in electoral democracies. *European Journal of Political Research*. 33 (2), 239–261. https://doi.org/10.1111/1475-6765.00382.

11. Card, D. & Moretti, E. (2007). Does voting technology affect election outcomes? Touch-screen voting and the 2004 presidential election. *The Review of Economics and Statistics*. 89 (4), 660–673. https://doi.org/10.1162/rest.89.4.660.

12. Carter, L. & Weerakkody, V. (2008). E-government adoption: A cultural comparison.

Information Systems Frontiers. 10 (4), 473–482. https://doi.org/10.1007/s10796-008-9103-6.

13. Cohen, M.J. (2017). Protesting via the null ballot: An assessment of the decision to cast an invalid vote in latin america. *Political Behavior*. 40 (2), 395–414. https://doi.org/10.1007/s11109-017-9405-9.

14. Colesca, S.E. & Dobrica, L. (2008). Adoption and use of E-Government services: The case of Romania. *Journal of Applied Research and Technology*. 6 (03). https://doi.org/10.22201/icat.16656423.2008.6.03.526.

15. Damore, D.F., Waters, M.M. & Bowler, S. (2012). Unhappy, uninformed, or uninterested? Understanding 'None of the above' voting. *Political Research Quarterly*. 65 (4), 895–907. https://doi.org/10.1177/1065912911424286.

16. Dandoy, R. (2021). An Analysis of Electronic Voting in Belgium. In: Didier Caluwaerts & Min Reuchamps (eds.). *Belgian Exceptionalism.* First. Routledge. pp. 44–58. https://doi.org/10.4324/9781003104643-5.

17. Dejaeghere, Y. & Vanhoutte, B. (2015). Virtuous villages and sinful cities? A spatial analysis into the effects of community characteristics on turnout and blank/invalid voting in local elections in Belgium 2006–2012. *Acta Politica.* 51 (1), 80–101. https://doi.org/10.1057/ap.2014.38.

18. Desai, Z. & Lee, A. (2021). Technology and protest: The political effects of electronic voting in India. *Political Science Research and Methods*. 9 (2), 398–413. https://doi.org/10.1017/psrm.2019.51.

19. Distel, B. & Ogonek, N. (2016). To Adopt Or Not To Adopt: A Literature Review On Barriers To Citizens'adoption Of E-government Services. In: *Twenty-Fourth European Conference on Information Systems (ECIS).* 2016 İstanbul,Turkey.

20. Driscoll, A. & Nelson, M.J. (2014). Ignorance or opposition? Blank and spoiled votes in low-information, highly politicized environments. *Political Research Quarterly*. 67 (3), 547–561. https://doi.org/10.1177/1065912914524634.

21. Eurostat (2021). *Digital economy and society statistics - households and individuals*

22. Fatke, M. & Heinsohn, T. (2017). Invalid voting in german constituencies. *German Politics*. 26 (2), 273–291. https://doi.org/10.1080/09644008.2016.1194398.

23. Fossati, D. & Martinez i Coma, F. (2020). Exploring citizen turnout and invalid voting in Indonesia: Two sides of the same coin? *Contemporary Politics.* 26 (2), 125–146. https://doi.org/10.1080/13569775.2019.1690210.

24. Galatas, S. (2008). "None of the Above?" casting blank ballots in ontario provincial elections. *Politics & Policy*. 36 (3), 448–473. https://doi.org/10.1111/j.1747-1346.2008.00116.x.

25. Gallagher, M. (1991). Proportionality, disproportionality and electoral systems. *Electoral Studies*. 10 (1), 33–51. https://doi.org/10.1016/0261-3794(91)90004-C.

26. Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. (2019). R-squared for bayesian regression models. *The American Statistician*. 73 (3), 307–309. https://doi.org/10.1080/00031305.2018.1549100.

27. Heinsohn, T. (2018). Casting an invalid vote at the 2017 French presidential election. *French Politics*. 16 (4), 383–399. https://doi.org/10.1057/s41253-018-0077-6.

28. Hsu, J., Huang, J., Kinsman, J., Fireman, B., Miller, R., Selby, J., et al. (2005). Use of e-Health Services between 1999 and 2002: A Growing Digital Divide. *Journal of the American Medical Informatics Association*. 12 (2), 164–171. https://doi.org/10.1197/jamia.M1672.

29. Inkinen, T., Merisalo, M. & Makkonen, T. (2018). Variations in the adoption and willingness to use e-services in three differentiated urban areas. *European Planning Studies*. 26 (5), 950–968. https://doi.org/10.1080/09654313.2018.1448756.

30. Jacobs, Bart & Pieters, W. (2009). Electronic voting in the Netherlands: From early adoption to early abolishment. In: Alessandro Aldini, Gilles Barthe, & Roberto Gorrieri (eds.). *Foundations of security analysis and design V: FOSAD 2007/2008/2009 tutorial lectures*. Lecture notes in computer science. Berlin ; New York, Springer. pp. 121–144. https://doi.org/10.1007/978-3-642-03829-7_4.

31. Kostadinova, T. (2003). Voter turnout dynamics in post-Communist Europe. *European Journal of Political Research*. 42 (6), 741–759. https://doi.org/10.1111/1475-6765.00102.

32. Kurfalı, M., Arifoğlu, A., Tokdemir, G. & Paçin, Y. (2017). Adoption of e-government services in Turkey. *Computers in Human Behavior*. 66168–178. https://doi.org/10.1016/j.chb.2016.09.041.

33. Laakso, M. & Taagepera, R. (1979). 'Effective' number of parties: A measure with application to west europe. *Comparative Political Studies*. 12 (1), 3–27. https://doi.org/10.1177/001041407901200101.

34. Lijphart, A. (1997). Unequal participation: Democracy's unresolved dilemma presidential address, american political science association, 1996. *American Political Science Review*. 91 (1), 1–14. https://doi.org/10.2307/2952255.

35. Lundell, K. & Högström, J. (2021). Institutions or the societal setting? Explaining invalid voting in local elections in sweden. *Scandinavian Political Studies*. 44 (3), 280–298. https://doi.org/10.1111/1467-9477.12199.

36. McAllister, I. & Makkai, T. (1993). Institutions, society or protest? Explaining invalid votes in Australian elections. *Electoral Studies*. 12 (1), 23–40. https://doi.org/10.1016/0261-3794(93)90004-4.

37. Piantadosi, S., Byar, D.P. & Green, S.B. (1988). The ecological fallacy. *American journal of epidemiology*. 127 (5), 893–904.

38. Power, T.J. & Garand, J.C. (2007). Determinants of invalid voting in Latin America. *Electoral Studies*. 26 (2), 432–444. https://doi.org/10.1016/j.electstud.2006.11.001.

39. Roseman, G.H. & Stephenson, E.F. (2005). The effect of voting technology on voter turnout: Do computers scare the elderly? *Public Choice*. 123 (1-2), 39–47. https://doi.org/10.1007/s11127-005-3993-3.

40. Ryustem, R., Dimitrova, D., Kostova, M., Corral, P. & Azevedo, J.P. (2018). *Poverty mapping in the republic of bulgaria*

41. Santana Pereira, J. & Costa Lobo, M. (2020). What explains preferential voting? A field experiment in Portugal. *Análise Social*. 4–26 Páginas. https://doi.org/10.31447/AS00032573.2020234.01.

42. Singh, S.P. (2017). Politically Unengaged, Distrusting, and Disaffected Individuals Drive the Link Between Compulsory Voting and Invalid Balloting. *Political Science Research and Methods*. 7 (1), 107–123. https://doi.org/10.1017/psrm.2017.11.

43. Uggla, F. (2008). Incompetence, alienation, or calculation?: Explaining levels of invalid ballots and extra-parliamentary votes. *Comparative Political Studies*. 41 (8), 1141–1164. https://doi.org/10.1177/0010414007301702.

44. van Deursen, A.J.A.M. & van Dijk, J.A.G.M. (2014). The digital divide shifts to differences in usage. *New Media & Society*. 16 (3), 507–526. https://doi.org/10.1177/1461444813487959.

45. Volkamer, M. (2010). Electronic Voting in Germany. In: Serge Gutwirth, Yves Poullet, & Paul De Hert (eds.). *Data Protection in a Profiled World*. Dordrecht, Springer Netherlands. pp. 177–189. https://doi.org/10.1007/978-90-481-8865-9_10.

46. Voutinioti, A. (2013). Determinants of User Adoption of e-Government Services in Greece and the Role of Citizen Service Centres. *Procedia Technology*. 8238–244. https://doi.org/10.1016/j.protcy.2013.11.033.

47. Wood, S.N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 65 (1), 95–114. https://doi.org/10.1111/1467-9868.00374.