Lund University
Department of Political Science

STVK02
VT2020
Tutor: Petrus Olander

# Closeness and turnout

## The 2016 Georgian parliamentary election

Matilda Wilhelmsson

# Abstract

In this thesis I examine how two variables: closeness and the TP-share (two-party-share) of votes impact turnout in elections. I use data from the 2016 Georgian parliamentary election and its two-rounds of election in the majoritarian districts. This field of research poses a wide variety of methodological approaches. One concerns the definition of the dependent variable, turnout. Instead of only using the turnout in the districts in the 2nd round I look at the change in turnout in the districts between the 1st round and the 2nd round of elections and use statistical analysis to test the impact of the two variables. I find that closeness and the two-party share of votes have a significant positive effect on turnout. Previous studies have all used their own different ways to define the closeness variable. Therefore, I also test hypotheses connected to the choice of closeness-measurement. I establish that the mpm (multi-party-margin) and the tpm (two-party-margin) are best suited for explaining the closeness-turnout hypothesis.

*Key words*: turnout, closeness, two-party-share of votes, statistical analysis, Georgia
Words: 9479

# Table of contents

# 1 Introduction

The first theory that predicted a relationship between closeness and turnout was presented by Downs in 1957. Downs created a mathematical model on the costs and benefits of voting for the rational voter. The basic idea was that the benefit of voting would be greater for the voter if the voter had the chance of being pivotal (changing the election results). This was more likely to happen if the election was close and every vote seemed to count. Thus, in a close election, voters would benefit more from voting and therefore, the turnout would be higher.

The research on closeness in elections and its impact on turnout is an interesting field of study since the results produced has been contradictory. It is still debated if these variations in results can be blamed on methodological differences and thus it needs further testing. One way in which previous studies have differed from each other is in how they have defined the dependent variable, turnout. Many studies have used the closeness of previous elections and tested its effect on turnout (Fauvelle-Aymar & François 2006). Others have used the ex-post measurement of turnout, using the actual closeness of the same election they measured the turnout from (Matsusaka & Palda 1993). Studies using the ex-post measurement of turnout have been criticized on the basis that the measurement might pose an endogeneity problem (Fauvelle-Aymar & François 2006, p.473). However, using closeness data from a previous election might be a bad choice. If the time between the elections is too great some other variables might interfere with the results (Fauvelle-Aymar & François 2006, p.473).

Looking at an election with two rounds helps us avoid these problems. Therefore, in this thesis, data from the 2016 Georgian parliamentary election will be used to test the relationship between closeness and turnout. In many of the majoritarian districts a second round was needed to appoint a winner (Central Election Commission of Georgia 2016a, p. 7). The second round was held only three weeks later than the first election. Previous studies have also used two-round elections to

test the effects of closeness on turnout (Fauvelle-Aymar & François 2006; Simonovits 2012; De Paola & Scoppa 2014), but they differ both in methodology and in their results, therefore more studies need to be conducted to confirm their results.

Previous studies have also differed in how they have constructed the measurement of closeness. Some studies have used the percentage margin (Simonovits 2012; De Paola & Scoppa 2014) and others have also included the raw vote margin (Matsusaka & Palda 1993; Fauvelle-Aymar & François 2006). Some studies consider all parties in an electoral race (Fauvelle-Aymar & François 2006) and others focus on the competition between the two leading parties (Matsusaka & Palda; Simonovits 2012; De Paola & Scoppa 2014). Therefore, these different measurements need to be subjected to further testing to determine which is the most appropriate to explain the relationship between closeness and turnout.

To accurately predict the relationship between closeness and turnout, other variables that might impact turnout also need to be considered. An array of different variables has been tested previously, some with a higher success rate than others (Cancela & Geys 2016). In this thesis, I will control for some of these effects by altering the dependent variable. I will further explain how I do this in chapter 3. Two variables that are still relevant to control for are: the TP-share (two-party-share) of votes and campaign spending and mobilization. I will further introduce these variables in chapter 2.

I will dedicate the rest of chapter 1 to stating the purpose and the research question of this thesis. Chapter 2 begins with an introduction of previous research while I simultaneously present my hypotheses. Chapter 2 ends with a brief introduction to the Georgian election system. Chapter 3 will in detail explain how I have reasoned when deciding the methodology for this thesis. A third hypothesis concerning methodology is presented. Lastly in chapter 3, the origin of the data used to test the hypotheses is presented. In chapter 4 the results from the statistical analysis are revealed and the hypotheses are evaluated. A further discussion of the results follows in chapter 5 where they are compared to the results of previous studies.

Chapter 6 contains the conclusion. I end this paper with suggesting ideas for future research in chapter 7.

## 1.1 Purpose

The main purpose of this thesis is to investigate a possible relationship between closeness and turnout. I will do so by using the advantage of Georgia's 2016 parliamentary election and its two rounds of election in the majoritarian districts. To isolate the effect between closeness and turnout I will also control for the TP-share of votes and the impact it may have on turnout. I also aim to make contributions to the methodological research in this area by testing different measurements of closeness to see which measurement is the most suitable for predicting the effects of closeness on turnout.

## 1.2 Research question

To narrow the focus for this thesis I propose the following as my research question:

Does closeness in elections affect voter turnout?

# 2 Theory

I begin this theory chapter with presenting the previous research on turnout and its determinants. The studies vary in both their methodological and theoretical aspects thus making this field diverse and disputed. To help me answer the research question I propose two hypotheses in this chapter that later will be subjected to testing. Lastly, a short introduction of the Georgian election system is given, which is the case I have chosen to study in this thesis.

## 2.1 Previous theory on the determinants of turnout and presentation of hypotheses

Under this section I discuss previous theory and relate it to the hypotheses I have chosen to test in this thesis. I also explain why I will not include variables that have been stated as important by previous research.

### 2.1.1 Closeness in elections

Before reviewing the history of this researching field, it is time to define the concept of closeness. In this thesis closeness is defined as: *the margin (percentage, ratio, or raw votes) between the two parties with the most votes in an election* (Fauvelle-Aymar & François 2006, pp. 474-476). As will be explained further in chapter 3, it is important to note that the closeness variable takes on a small value when the margin between the parties is small and takes on a higher value when the distance is greater. Therefore, a high closeness value means that the margin between the two parties is in fact not small (and close) but rather large. In chapter 3 we will also explore some possible variations within the closeness concept and how they are measured in different ways.

As mentioned in the introduction the idea that closeness is related to turnout originates from Downs (1957) who developed a rational voter theory. This theory was transformed into a mathematical function consisting of variables Downs theorized would impact the "gain" a voter would get from voting in an election. One of the variables in this function was the voter's chance of being a pivotal voter. According to Downs, being a pivotal voter, meaning having the chance of switching the outcome of the election, would increase the voter's benefit of voting. Though this chance would realistically be very small since most elections on a bigger scale are not tied by only a few votes. Some decades later Matsusaka and Palda (1993) used Downs theoretical foundation to test the closeness-turnout hypothesis. In their study they call this the "Downsian Closeness Hypothesis" (DCH). Matsusaka and Palda (1993) put the theory to test by examining the impact of the closeness of the electoral race in the 1979 and the 1980 Canadian national election on the turnout. Their results confirmed that close elections (chance of being the pivotal voter) indeed led to higher turnout (since voters are benefiting more from voting). These results however changed when they instead of looking at aggregated data, looked at data on the individual level. Then they found no statistical evidence that the DCH should be true and they concluded their study with claiming that the aggregated data suffered from an ecological fallacy and therefore the results in favor of the DCH were likely false.

Since the study on Canadian elections by Matsusaka & Pada the DCH has been tested by several studies (Fauvelle-Aymar & François 2006; Simonovits 2012; De Paola & Scoppa 2014), all reaching the same conclusion: closeness does impact turnout. These studies all used aggregated district data. Cancela and Geys (2016) did a meta-analysis of the existing literature and on the success rate of tested hypotheses. They found that the success rate of the DCH in national elections is only around 68% of all studies conducted to test it (Cancela & Geys 2016, p. 270). To try and bring more clarity into this field of research I will subject the DCH to testing:

*Hypothesis 1: Turnout is higher in close elections.*

### 2.1.2 Two-party share of votes and turnout

Besides the already introduced variables, another variable has been subjected to testing in recent studies. Simonovits (2012) and De Paola & Scoppa (2014) looked at how the share of people voting for a party that did not proceed to the second round affected turnout. Their hypotheses were that the higher the third-party share of voters in the first round, the greater the decrease in turnout in the second round. This is because people who voted for another party that did not make it to the second round rather stay at home in the second round and not vote at all than changing party preferences and voting for any of the two leading parties. This could be seen as a democratic issue in countries who have some form of majoritarian election system where people do not participate in the second round of voting.

To investigate if there is any truth to the third-party hypothesis, I will include this in my analysis. I do this by adding a variable called TP-share of votes (two party-share) which adds the shares of votes received by the two leading parties in the first round of voting:

*Hypothesis 2: Turnout is higher in elections where the TP-share of votes is higher.*

### 2.1.3 Campaign mobilization and spending

While the main hypothesis in the previously mentioned studies concerns the effect of closeness on turnout, several of them investigate the importance of another variable, namely campaign mobilization and spending. Campaign mobilization can include phoning voters, putting up posters, holding rallies and the spending tells how much money has been spent on these activities.

Cox & Munger (1989) looked at both closeness and campaign mobilization and spending in gubernational and House elections in the US 1982. They theorized that closeness in electoral races are likely to spur campaign mobilization and spending because politicians and other elites feel the outcome of the race is unsure. Increased mobilization and spending engage more voters and leads to a higher turnout in the

election. They predicted that when this campaign-variable is included in analysis, it would erase the effect of closeness so that the closeness-turnout relationship would become so small it was trivial. They did find great support for the effect of campaign mobilization and spending. "In the 1982 congressional elections, an extra dollar in per capita spending by Senate and gubernatorial candidates increased House turnout by 1.6% and 1.1%, respectively" (Cox & Munger 1989, p. 226). However, the results from their analysis showed that even though mobilization and spending had a great impact on turnout, closeness still also had an independent statistically significant impact on turnout (Cox & Munger 1989, pp. 218, 226).

Shachar & Nalebuff (1999) conducted a similar study to Cox & Munger and further developed the closeness-turnout model. Their model explains in detail how closeness leads to an increase in campaign mobilization and spending which then in turns impacts turnout. The relationship between closeness and turnout is portrayed as indirect. They first tested the two independent variables separately and found that both campaign mobilization and spending and closeness was significant when tested against turnout. Then they put these two (and other relevant variables) into their model and found statistical evidence supporting their theory.

In their meta-analysis of the existing turnout-literature, Cancela & Geys (2016) also included a variable on campaign expenditures. Campaign expenditures have a high success rate, as around 85% of studies including the variable found it to be significantly impacting turnout in national elections (Cancela & Geys 2016, p. 270). This means that it is an important variable to include when looking at the closeness-turnout hypothesis. Unfortunately, no district data on campaign spending and mobilization is available for the 2016 Georgian parliamentary election and therefore I cannot include this variable in this thesis. I acknowledge that this is a flaw and that the absence of this variable might impact my results, making them less accurate. However, previous studies that have included both a variable on closeness and a variable on campaign spending and mobilization have found that both variables have an independent impact on turnout (Cox & Munger 1989; Matsusaka & Palda 1993; Fauvelle-Aymar & François 2006). Two other studies (Simonovits 2012; De Paola & Scoppa 2014) did not include the campaign variable in their studies and therefore no studies so far suggest that the inclusion of a variable

of campaign spending and mobilization would erase the impact closeness has on turnout. Rather than to fully dismiss results in this thesis one should be vary of the uncertainty of the exact magnitude of the closeness variable and interpret it carefully.

### 2.1.4 Other determinants of turnout

Studies looking into turnout often include a wide variety of variables, from those mentioned above, social, and economic variables are often included in the analysis. A sample of variables that have been included are: religion, native citizens, minority/majority language speakers, population growth (Matsusaka & Palda 1993), level of education (Matsusaka & Palda 1993; Fauvelle-Aymar & François 2006), age, occupation (Fauvelle-Aymar & François 2006), turnout in previous election, turnout in first round of elections (Simonovits 2012), municipality fixed effects, candidate fixed effects (De Paola & Scoppa 2014). The most used is probably electorate size. Previous studies have tested the hypothesis that a larger electorate ought to produce a lower turnout (Matsusaka & Palda 1993; Simonovits 2012). The reasoning is that a larger electorate makes a tie or a close race less likely and therefore a voter is less likely to have the pivotal vote which reduces the voter's incentives to vote (Matsusaka & Palda 1993, p. 863; Simonovits 2012, p. 368).

I will not include any of these variables and I have two reasons for doing so. The first one being that there simply is a shortage of this information on district level in Georgia (except electorate size). The second reason is connected to the dependent variable primarily used in this thesis, namely the change in turnout in each district. The districts are not being compared to each other, but the difference in turnout in the first and second round of elections in the same district is in focus. Thus, I can control for possible variation in these variables that may occur between districts. I will further discuss the implications of using this method in chapter 3.

## 2.2 Elections in Georgia

The Georgian election system is a mixed system that combines a national proportional vote with a majoritarian vote in the electoral districts (Central Election Commission of Georgia 2016b, p. 2). There is a total of 150 members in parliament out of which 73 are elected in the majoritarian districts and 77 from a national proportional closed party list vote. Elections are held every four years and parties and blocs must get a minimum of five percent of the proportional vote to qualify for a seat. In the majoritarian districts, a party must receive over half of the votes cast (50% of the votes) to win the district. If no party succeeds with this in the first round, a second round is held between the two leading parties from the first round. In this thesis I take advantage of the data generated by the two rounds of election in the majoritarian districts to test my hypotheses.

In 2016 the first round of elections was held 8th of November and a second round was needed to be held the 30th of November in 50 out of 73 majoritarian districts to appoint a winning candidate (Central Election Commission of Georgia 2016a, p. 7). Two parties, the Georgian Dream party, and the United National Movement, are frequently in the top two (ElectionsPortal 2016a; ElectionsPortal 2016b).

# 3 Method

I start this method chapter by explaining how I will define and measure the dependent variable. Then I also explain the definitions of the independent variables and how I measure them. I propose a third hypothesis related to the choice of methodology. By testing this hypothesis, I hope to bring more clarity into my answer of the research question. After this follows a section on statistical analysis and how it will be implemented to test the hypotheses. Lastly, I dedicate a section to the data used in the analysis.

## 3.1 Turnout: variation between districts or variation within districts

As with any scientific research, how one defines and measures the chosen variables has a great impact on the results and conclusion. Therefore, I will specify how I have chosen to define and measure the dependent variable, turnout. Two recurrent approaches in previous research have been either to measure the variation in turnout between districts (Matsusaka & Palda 1993; Fauvelle-Aymar & François 2006; De Paola & Scoppa 2014) or to measure the change in turnout between the first and second round of elections in each district (Fauvelle-Aymar & François 2006; Simonovits 2012).

The simplest way to define the turnout variable is to use the percentage-turnout from each district and compare them directly with each other. However, this method does not take in to account other differences between the districts that could have an impact on turnout. Therefore, studies using this direct comparison should be sure to control for variables that could potentially impact levels of turnout. Some control-variables that have been used in previous studies where listed in chapter 2.

Instead of using the turnout directly some studies have chosen to instead look at the change in turnout between the first and second round of election. This is simply done by subtracting the turnout from the first round from the turnout in the second round to get the change in turnout:

$$(turnout_{2nd\ round} - turnout_{1st\ round} = \Delta\ turnout)$$

This method can of course only be applied to elections with two rounds. Looking at the change in turnout within the same district has two implications. District-specific factors that stay constant between the two elections will be automatically controlled for and does not need to be included in the analysis. This enables a closer focus on the closeness and TP-share of votes variables and their relationship to turnout. However, when one uses this method, the opportunity is lost to control for other interesting variables that differs between the districts and might have an impact on turnout-levels such as education and unemployment. Therefore, the choice between the two dependent variables is a choice between a more concentrated study on a few chosen variables or a study where the importance of many variables can be examined at the same time.

Fauvelle-Aymar & François (2006) applied both methods with the argument that while the between district variation-method allowed them to include a range of other variables, using the change in turnout made it possible to only look at the specific relationship between closeness and turnout. I will apply the change in turnout as dependent variable. The reason that change in turnout is applied in this study is, as mentioned in chapter 2, because of the lack of appropriate regional Georgian data when it comes to previously tested control variables. To make the results more reliable I have chosen to do a more concentrated study. When more data becomes available, a complimentary study including more variables and using the turnout as the dependent variable should be performed.

## 3.2 Measuring closeness

In this section I focus on explaining the methodological aspects of the independent variable, closeness, which is used to test the main hypothesis in this study, the closeness-turnout hypothesis. Closeness might not at a first glance appear as an especially difficult concept and variable. However, there are two main considerations that must be taken before choosing exactly how one defines and applies closeness in a scientific study. Firstly, one must choose whether to use the ex-post closeness, recorded from the same election as the turnout, or the ex-ante closeness, recorded from a previous election. Secondly, when one has decided which data to use, one must define exactly how one measures closeness in elections. Previous studies have challenged each other over how to define and implement the closeness variable and therefore, I will first present their arguments before presenting what methodological practices I have chosen to use in this thesis.

### 3.2.1 Using ex-post or ex-ante closeness

Before I present the methodological variations and arguments in connection to the closeness variable, a terminological explanation of ex-post and ex-ante closeness is in place. When I use the term ex-post closeness, I am referring to closeness data that has been recorded from the same election as the turnout data was recorded (Fauvelle-Aymar & François 2006, p. 473). The use of ex-post closeness is most common and a necessity if one analyses an election with only one round of election. Turnout is measured after the election takes place and so is the closeness. The voters are expected to have a rational sense of how close the election is going to be and therefore their expected closeness should be the same as the actual closeness on the election day. The ex-ante closeness is instead the closeness recorded from a previous election (Fauvelle-Aymar & François 2006, pp. 473-474). This election functions as a prediction of how close the next election will be. Thus, closeness and turnout are not recorded from the same election as with the ex-post closeness but closeness from one election and then from the following election, turnout is measured.

Matsusaka and Palda (1993) used both ex-post and ex-ante closeness to test the DCH and to see if the results varied between the two measurements. They

concluded that there was no real difference in the results from applying both closeness-variations (Matsusaka & Palda 1993, pp. 866-867). In a later study by Fauvelle-Aymar & François (2006) the use of ex-post closeness was criticised. Fauvelle-Aymar & François argued that since closeness and turnout is recorded from the same election it causes a severe endogeneity problem. Ex-post closeness influences participation while at the same time it (ex-post closeness) also is a function of the dependent variable of participation (Fauvelle-Aymar & François 2006, p. 473). In the end, you have both ex-post closeness on the independent and the dependent sides of the relationship. In the same study, Fauvelle-Aymar & François (2006, p. 473) also criticized the use of ex-ante closeness in cases where the first election took place a long time before the second one. They argued that if the elections where too far apart, some variables or events might have impacted the turnout so that the closeness in the previous election might no longer be an accurate estimation of the closeness on the election day.

Considering the previous research and the methodological debate on the area I have chosen to use the ex-ante closeness. To avoid, to the greatest extent possible, that the two elections are too far apart, thus, inviting the possibility of other events disturbing the analysis, I have chosen to take advantage of the Georgian election system and its majoritarian districts where two rounds of election are often required. In the 2016 election, the second election took place only three weeks after the first election (Central Election Commission of Georgia 2016a, p.7). By using ex-ante closeness and taking advantage of the short time between the elections in the Georgian majoritarian districts I aim to make my results as reliable as possible.

### 3.2.2 Different measurements of closeness

We have now discussed and decided from which elections different data should be collected to avoid methodological errors. Before we start using the data for closeness there is one more aspect to consider. Previous studies have constructed and used different closeness measurements to test similar hypotheses. Thus, there is a risk that variations in results are not due to a variation in closeness but rather due to the way each study has defined and measured closeness. Besides contributing

to existing theory in this field of study, my aim is also to make methodological contributions. Therefore, I will test the four most common ways to measure closeness in elections.

I draw inspiration from the study by Fauvelle-Aymar & François (2006). They conducted a thorough test of which closeness measurement is the most appropriate to use when testing the closeness-turnout hypothesis. They grouped these measurements into those that only consider the margin between the two leading parties and those who looks at the multi-party competition, considering all parties in the district election rather than just the two leading parties (Fauvelle-Aymar & François 2006, p. 475). They found that the two-party measurements generally did better since most of the multi-party measurements failed to show statistical significance when testing the closeness-turnout hypothesis (Fauvelle-Aymar & Francois 2006, p. 484-486). In this thesis I will include only those measurements that Fauvelle-Aymar & François (2006) found to show significance, that is the: multi-party margin (mpm), raw-vote margin (rvm), two-party ratio (tpr) and the two-party margin (tpm). These measurements all focus on the margin between the two leading parties in a race. This could potentially be problematic in an election where there are more than two parties succeeding to the second round of election. Since we assume that the two leading parties will be leading in the second round too, the measurements might not correctly predict the actual closeness in the second round if the voters choose to vote for a third party. Again, we can avoid this methodological problem by taking advantage of the Georgian election system. Since the two leading parties are the only ones succeeding to the second round, we do not have to worry about increased support for third parties. Therefore, using measurements focusing on the two leading parties should function well in this case.

The first measurement of closeness is perhaps the most intuitive one. The multi-party margin (mpm) generates the percentage difference between the two leading parties, by dividing their difference in number of votes by the total number of votes cast for all the competing parties.

$$\boldsymbol{mpm} = \frac{(\boldsymbol{v1}-\boldsymbol{v2})}{(\sum_{i=1}^{c} \boldsymbol{v}_{ij})}$$

The number presented is the same percentage-difference that is likely to be shown to the public during the election coverage and therefore it is likely that it guides people in their decision whether to vote or not in the second round.

The raw vote margin (rvm) generates the raw difference in number of votes between the two leading parties.

$$\boldsymbol{rvm} = (\boldsymbol{v1} - \boldsymbol{v2})$$

Just as the mpm, the rvm measurement is also likely to be available to the public and guide them in their estimation of closeness. However, this measurement is problematic when used to compare districts if they are of slightly different size and are used to different levels of closeness. While a difference in 5000 votes might be a lot for one district considering its circumstances, it might not be for another one.

The last two measurements focus exclusively on the performance of the two leading parties and compare them in relation to each other. The two-party ratio (tpr) only looks at the ratio between the two leading parties, ignoring the rest of the competitors.

$$\boldsymbol{tpr} = (\frac{\boldsymbol{v1}}{\boldsymbol{v2}})$$

The two-party margin (tpm) only looks at the percentage difference between the two leading parties by dividing their difference in number of votes with their combined number of votes.

$$\boldsymbol{tpm} = \frac{(\boldsymbol{v1}-\boldsymbol{v2})}{(\boldsymbol{v1}+\boldsymbol{v2})}$$

The tpm is very similar to the mpm, only the denominator is different. In an election where two parties get almost all the votes, the tpm and the mpm will generate similar results since the value of the denominator will be almost the same. To test if the four measurements produces different results when used to test the turnout-closeness hypothesis, I propose a third hypothesis:

*Hypothesis 3: The mpm is the best measurement to predict the closeness-turnout hypothesis. The rvm will be the second best to use, followed by the tpr and tpm.*

To clarify once more, the mpm will be best at predicting the DCH since this is the number often used in election coverage, it is likely to guide voters in making the decision whether to vote or not, and it is comparable between districts. The rvm is also a number that is commonly available during election campaigns, but a raw number might not be comparable between districts if they vary in size (number of voters) and normally have different levels of turnout. The tpr and tpm are measurements that are not normally available to the public and therefore they do not directly guide the voters in the decision making. However, in elections where two parties are getting almost all the votes, the value produced by the tpm will be very similar to the one produced by the mpm. Hence my prediction that the mpm will perform the best, rvm second best and the tpr and tpm will perform the worst.

## 3.3 The TP-share of votes

Besides testing different measurements of closeness as independent variables I will also, as mentioned in chapter 2, test if the TP-share (two-party-share) of votes has any effect on turnout. The TP-share of votes is, as the name suggests, the two leading parties´ (in each district) combined percentage share of the votes in the first round of elections. I mainly draw inspiration from the study by De Paola & Scoppa (2014) who also constructed the TP-share of votes variable in this way to test its impact on turnout. Other studies have instead used the share of votes gained by other parties than the two leading parties (Simonovits 2012). These two seemingly different variables are just two different ways of testing the same hypothesis and will generate the same results if applied to the same data.

## 3.4 Statistical analysis

To test my hypotheses, I will use statistical analysis and more precisely OLS-regression. Since I have data from 50 Georgian districts, statistical analysis is the best suited method since it will minimize time and work but also allow us to take

advantage of the straightforward way of testing hypotheses using probability methods. An introduction to the concept of OLS-regressions is in place:

> *Ordinary least squares (OLS) regression is a statistical method of analysis that estimates the relationship between one or more independent variables and a dependent variable; the method estimates the relationship by minimizing the sum of the squares in the difference between the observed and predicted values of the dependent variable configured as a straight line* (Encyclopedia 2020).

To keep in mind when using statistical analysis to study questions in social science is that while we have the advantage of testing large amounts of data, it will not give us the possibility to conduct a deeper qualitative study that can discuss the casual mechanism in depth. In this study I test how closeness impacts turnout. Previous studies have found support for a relationship between closeness and turnout using statistical analysis. However, by using this method we cannot explain and prove the exact casual mechanism as it is stated in the DCH, that people vote because their vote might be pivotal, but we can test if more voters participate in close electoral races. Thus, statistical analyses can be benefitted from a complementary qualitative analysis examining the casual mechanism to its core.

However, the aim of this thesis is to test if closeness and the TP-share of votes impacts turnout. The method of statistical analysis and OLS-regression has been used by previous studies conducted on this field of research. Thus, it is in my interest to use a similar method and make the results in this thesis comparable to the results of previous studies.

### 3.4.1 Coefficients

The coefficients generated by the OLS-regression give us vital information about the steepness and direction of the slope (Teorell & Svensson 2007, pp. 167-168). By looking at these coefficients we will know how many units y (turnout) increases when x increases with one unit.

$$T = Cx + TPx$$

Here I have assumed that the model is linear. This assumption builds on the models used by previous studies. T represents the turnout, C the closeness coefficient and TP the TP-share of votes. Each coefficient is followed by the variable x which represents the increase in closeness and TP-share of votes, respectively. Together the coefficients on the right side of the equation are used to explain the turnout, on the left side. However, we cannot interpret the effect of our independent variables on the turnout by only looking at the coefficients generated by the regression.

### 3.4.2 Significance and adjusted R-square

To help us interpret the coefficients generated by the regression we have the help of the significance value. The significance is calculated with the help of the regression and tells us how unlikely the results we have would be if our hypothesis is wrong (Teorell & Svensson 2007, p. 140). We have to be careful, the significance does not tell us if our hypothesis is correct but shows us how likely it is that our data just shows a deviation from the rest of the data (Teorell & Svensson 2007, p. 146). Within social science the conventional significance level is 5 percent and often levels of 10, 5 and 1 percent are used to test hypotheses (Teorell & Svensson 2007, pp. 140-141). This means that our results have a 10, 5 or 1 percent of probability of just being random deviations and not a proof of a relationship between variables x and y. Using three different significance levels lets us see what difference in conclusion we could have depending on the significance level, with the 1 percent level our conclusion would be more careful than using the 10 percent level (Teorell & Svensson 2007, p. 145-146). Important to note is that a hypothesis that passes the 1 percent level is not necessarily truer or more proven than a hypothesis that passes a 10 percent level, but it is statistically more unlikely that the first hypothesis is incorrect. Studies in this field (Matsusaka & Palda 1993; Fauvelle-Aymar & François 2006; Simonovits 2012; De Paola & Scoppa 2014) have previously used significant levels of 10, 5 and 1 percent and to make the results from this study as comparable as possible to the previous research the same levels

should be used. Therefore, I will use the significance levels of 10, 5 and 1 percent to test my hypotheses.

Another variable that will be generated by the OLS-regression is the R-square for every individual variable or model. The value of the R-square can vary between 0 and 1 and is a measurement of how much of the variance in the regression that has been explained by applying the chosen independent variables (Teorell & Svensson 2007, p. 175). If our R-square is 0, we could not explain any of the variance in turnout with our chosen variables. If the R-square is 1 it means that the independent variables we applied are responsible for all variance experienced by the dependent variable. It would mean that closeness and the TP-share of votes are the only variables impacting turnout within the districts. This is of course unlikely to happen. However, the R-square is not suitable for analysing multi-variable regressions since its value tends to increase automatically every time a new variable is added, even though the variable might not have an effect on the dependent variable (Teorell & Svensson 2007, p. 201). Instead I will use the adjusted R-square, which only considers the real effect a variable has. The adjusted R-square is also generated by the regression and has been preferred before by many studies in this field.

### 3.4.3 Interpreting variables

Before interpreting the significance and the R-square (Teorell & Svensson 2007, pp. 175, 177, 193-197) reminds us we must keep five things in mind: spuriousness, indirect effects, covariation, time order and cross-study comparison. Even if the R-square and the significance would suggest a relationship between our independent and dependent variables there is a possibility this might not be the case.

Spuriousness is a serious problem when using OLS-regression to test hypotheses. The chosen independent variables might seem to explain variance in the dependent one, but the relationship could be a coincidence and another variable that we have not included in our analysis could be responsible for the variance (Teorell & Svensson 2007, p. 193). However, if we control for this other variable the spurious relationship between the independent and dependent variable will disappear.

Therefore, I include the TP-share of votes in the hope of eliminating the risk of spuriousness.

Secondly, one must be wary of indirect effects coming from other variables. Shachar and Nalebuff (1999) claimed that the relationship between closeness and turnout was indirect since increased closeness would lead to more spending and mobilization thus leading to a higher turnout. There is a risk that this could be the case in my study as well but as I previously argued, the studies who included a variable on spending and mobilization still found a significant direct relationship between closeness and turnout.

Another reason as to why it is important to carefully choose which variables to include in a model is covariation. Covariation happens when two of the independent variables tend to exist simultaneously thus making it hard to distinguish their effects on the dependent variable (Teorell & Svensson 2007, p. 196-197). One example is the independent variables of electorate size and the rvm. As mentioned earlier in chapter 2, smaller electorates are more likely to have a smaller margin counted in raw votes than the larger ones. To avoid the disturbing effects of covariation on the results it is important to include these variables in a model and control for their joint effects on turnout (Teorell & Svensson 2007, p. 197). This makes it possible for us to see their individual effects on turnout without the covariation.

One problem that might seem simple but can sometimes prove hard to solve is how we determine the order in which the dependent and the independent variables took place. If we are unlucky it could be that the dependent variable occurred first and then impacted the independent variables. The adjusted R-square would be the same, but our casual mechanism would be the opposite. Luckily for this study, it is quite easy to distinguish a certain order in which variables impacted each other. Since we collect the data from two elections, separate in time, we can easily determine which variable appeared first. We collect closeness data from the first election and turnout data 3 weeks later from the second election.

Lastly, when we interpret the R-square we must compare it to the R-squares of previous studies. Without the comparison we cannot know if our results are worthy

of interest or weaker than those from previous studies. The number itself does not tell us much. From these three recommendations we can draw a conclusion. It is very important that we establish a casual mechanism, a theory, that is well thought through before we start interpreting the values generated by the OLS-regression.

### 3.4.4 Different closeness models

As mentioned above, the best ways to test my hypotheses is by including them in a model and putting them through an OLS-regression. One problem is that one aim with this thesis is to test different variations of closeness variables/measurements (the mpm, rvm, tpr, tpm). These measurements cannot be tested at the same time in the same model. They measure the same thing or very similar things thus, their effects on turnout will interfere and distort the results. This phenomenon is called multicollinearity (Teorell & Svensson 2007, pp. 211). Thus, it requires four different models where we test the different closeness measurements separately in combination with the variable on the TP-share of votes.

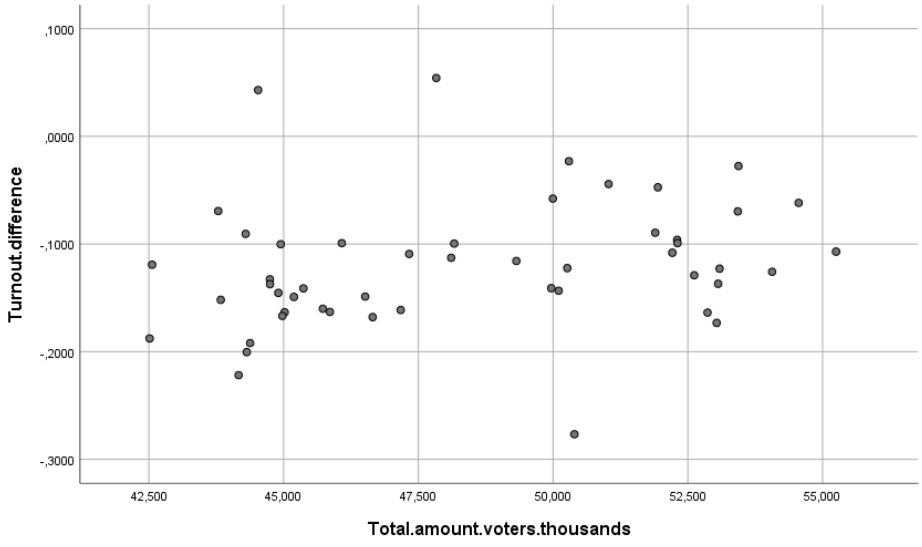
## 3.5 Data from the 2016 Georgian Parliamentary election: Two-round elections in the majoritarian districts

I use data from electionsportal.ge which collects data from elections in Georgia, both national and local elections are included. The website was set up by the National Democratic Institute (NDI) and is funded by the Swedish development agency (SIDA) (ElectionsPortal 2020a). The website uses voter list data and elections data collected by the Central Election Commission of Georgia (CEC) and the information on electoral geographic boundaries comes from the Caucasus Research Resource Centre (CRRC) (ElectionsPortal 2020b). Data is available both for the first round of majoritarian district election (ElectionsPortal 2016a) and for the second round (ElectionsPortal 2016b). It includes data on voter turnout, election results and information about the top two contenders in each of the 73 districts. The data is displayed in form of an interactive map and in a list.

# 4 Results and analysis

In this section I present the results from the OLS-regression and examine my hypotheses. Before moving on to the results I want to clarify two things concerning the terminology used. Firstly, as previously mentioned, the term closeness can be confusing. When I say that the closeness variable increases, what I am really saying is that the margin between the parties is increasing. This means that the parties are in fact getting less close. The second clarification is needed since in the 2016 Georgian election, turnout decreased between the first and the second round in 48 out of 50 districts who had two rounds of election (Figure 1). Therefore, when I use the terms increase and decrease to explain the results, it really means decrease less and decrease more. But since I compare the districts relative to each other I will use increase and decrease in turnout to explain the general results. This terminology is also easier to understand and makes the results easier to compare with those of previous studies.

**Figure 1.** Change in turnout/turnout difference (percent) between the first and second round of elections in the Georgian majoritarian districts plotted against district size.




Note: Data: electionsportal.ge & Central Election Commission of Georgia (CEC).

Looking at the results in Table 1 we can see that all the closeness measurements are significant and have the expected negative sign. In Model 1 an increase with 10% in the mpm (percentage margin) yields a -4.46% decrease in turnout. The rvm shows us that when the margin between the two leading parties increases with 1,000 votes in the first round, turnout decreases with -1.7% in the second round. Since tpr represents the ratio between the two leading parties it is a bit trickier to interpret the results. A 1-unit increase in ratio between the parties leads to an -8.6% decrease in turnout in the second round. Ex. If the biggest party gets 30% of the votes and the second party gets 10% of the votes it gives us a ratio of 3/1=3, a 1-unit increase would happen if the biggest party instead got 40% and the second 10% thus resulting in a ratio of 4/1=4. Lastly, the tpm, which is similar to the mpm but only focuses on the two leading parties, shows a -3.2% decrease in turnout when the tpm increases by 10%.

The variable on the TP-share of votes is significant across all four models and shows the expected positive sign. The variables´ effect on turnout ranges from lowest a 1.82% increase in turnout (Model 4) when the TP-share of votes increases with 10% to a 2.83% increase in turnout in Model 2. Thus, less people vote in the second round when the closeness margin increases, and more people vote when the two leading parties share of votes of is high in the first round.

**Table 1.** The impact of closeness on turnout (change in turnout), testing 4 models, including TP-share of votes as an independent variable.

| Dependent variable: | Model 1 - Δturnout | Model 2 – Δturnout | Model 3 - Δturnout | Model 4 - Δturnout |
|---|---|---|---|---|
| Constant | -.234*** | -.260*** | -.131 | -.178** |
| | (.001) | (.001) | (.148) | (.029) |
| Closeness mpm | -.446*** | | | |
| | (.000) | | | |
| Closeness rvm (1,000) | | -.017*** | | |
| | | (.000) | | |
| Closeness tpr | | | -.086*** | |
| | | | (.000) | |
| Closeness tpm | | | | -.320*** |
| | | | | (.000) |
| TP-share of votes | .257*** | .283*** | .212** | .182* |
| | (.004) | (.002) | (.027) | (.062) |
| R-Square | 0.508 | 0.472 | 0.490 | 0.501 |
| Adjusted R-square | 0.487 | 0.450 | 0.469 | 0.480 |
| N | 50 | 50 | 50 | 50 |

Note: N= number of districts/samples. *$p < .10$. **$p < .05$. ***$p < .01$. Data: electionsportal.ge

The closeness-turnout hypothesis, hypothesis 1, is confirmed. Independent of how one chooses to measure turnout the variable still reaches significance and the coefficients are all negative, as it should be according to the hypothesis, and their magnitudes implies a solid effect on turnout. When the electoral margin increases, turnout decreases. Hypothesis 2 is also confirmed, the TP-share of votes-variable is significant across all four models with the coefficients showing a positive sign, in line with the hypothesis, and their magnitudes confirming a substantial effect. Hypothesis 3 however, could only be partially confirmed. The mpm had the highest adjusted R-square at 0.487, followed by the tpm at 0.480. The reason as to why the tpm did well might be because of a high TP-share in the first election in most districts. In chapter 3 I explained how a higher TP-share would make the mpm and the tpm to function more and more as the same measurement. This is because if the two leading parties gets all the votes, the denominator in both measurements would be the same value. Since the denominator is the only thing differing between the measurements, they would become the same measurement. Using the data on the TP-share I already have and calculating a mean value shows that the mean TP-share

in the districts is 74.15%. This high mean could be the reason for why the tpm did well in this study. The rvm performed worst and only scored an adjusted R-square at 0.450. This might be because the rvm measures the margin in number of votes and not percentages. Since the electoral districts vary in size (number of registered voter) a percentage measurement might perform better.

# 5 Discussion and comparison with previous research

The results generated by this study have shown to be significant and are showing the expected signs (hypothesis 1 and 2). We can read the magnitude of the variables´ effect on turnout by looking at the coefficients and we can value the model's capacity to explain the variation in turnout by looking at the adjusted R-square values. However, we cannot fully evaluate the results in Table 1 without comparing them to the results of previous studies. I have previously introduced several studies which have produced results in this field of research. While all the results and studies are highly relevant for understanding variations in turnout, I have chosen to compare my results only directly with two of these studies: Fauvelle-Aymar & François 2006 and Simonovits 2012. These studies both used the change in turnout as the dependent variable in their models and they include the same (partially) variables I have used in my models. Thus, the comparison will avoid some methodological differences that otherwise might lead to a variation in results.

Fauvelle-Aymar & François (2006) used both the turnout (pp. 482-483) and the change in turnout (pp. 487-488) to test their hypotheses. They tested the four different closeness measurements: mpm, rvm, tpr and tpm. They also included spending and other variables into their models and therefore their models are not identical to mine. All four closeness measurements in their study reached significance. A 10% increase in the mpm resulted in a -1.36% decrease in turnout in the second round. For the rvm, a 1,000 votes increase led to a -2.94% decrease in turnout. A 1-unit increase in the tpr (ratio measurement) resulted in a -2.708% decrease in turnout and for the tpm a 10% increase meant a decrease in turnout of an astonishing -8.553%. The adjusted R-square values for the respective models were: mpm 0.469, rvm 0.477, tpm 0.470, tpr 0.471. These results are quite the opposite to the ones generated in this thesis. In my study, the models of the mpm

and the tpm had the highest adjusted R-square while the rvm and the tpr had lower values. The magnitude of the mpm is over twice as big (-4.46%) in this thesis than in the study of Fauvelle-Aymar & François. The tpr also showed to have a substantially larger effect in my study (-8.6%) than it did in their study (-2.708%). I did not find that the tpm influenced turnout of the same magnitude (only -3.25%) as it did in their study (-8.5535). The rvm showed an impact of similar magnitude in both studies. Looking at the adjusted R-square values one can see that the value range is approximately the same in both studies (0.450-0.490) with models 1 and 4 in my study performing slightly better than the models used by Fauvelle-Aymar & François. This is interesting since Fauvelle-Aymar & François included a variable on campaign spending in their models which according to previous theory should have an impact on turnout. I am satisfied to have reached these values of the adjusted R-square without having included this variable.

Simonovits only included one of the four closeness measurements in his study, the mpm (2012, p. 369). He did, however, also include the TP-share of votes in his model. He found that if the mpm increased with 10% that it would result in a 2% decrease in the turnout of the second round. Instead of using the two-party share of votes as his variable, he used the opposite, the third-party share of votes. An increase in the third party-share will lead to a decrease in the two-party-share of the same proportion, thus, making them directly comparable. He found a 10% increase in the third-party share in the first vote to lead to a -0.4% decrease in the second vote. This is the same as a 10% increase in my variable, the TP-share of votes, and would lead to a 0.4% increase in turnout. However, this variable did not reach significance in his model thus, he ruled out its impact on turnout. I found the TP-share of votes to both be significant across all models and to have a substantially larger effect on turnout. A 10% increase in the TP-share of votes leads to an increase of 1.8-2.8% in turnout (a decrease of 10% would lead to a decrease of the same magnitude). I used the data generated by his model 3 which generated a R-square value of 0.853. Since he did not present any value on the adjusted R-square I cannot directly compare this value to my values but since his R-square is significantly higher than my R-square values I suspect that his adjusted R-square would be higher than in my models. He did not include campaign mobilization and spending thus, leaving its potential impacts on his model unknown.

To summarize this comparison of my results I found that the effects of the mpm showed significantly greater magnitude in my study than in previous studies. Compared to Fauvelle-Aymar & François (2006) who included the variable on campaign spending my models still had similar (and slightly higher) adjusted R-square values. I did find the TP-share of votes to be significant and its effects to be of greater magnitude than when it was previously applied by Simonovits (2012).

# 6 Conclusion

In the beginning of this thesis I introduced the following research question: Does closeness in elections affect voter turnout? To help the answering of this question I proposed three different hypotheses. Firstly, closer elections should have higher turnout. Secondly, in elections were the TP-share of votes is higher, turnout should be higher. Thirdly, the mpm and the rvm should be best at predicting the relationship between closeness and turnout. I took advantage of the 2016 Georgian parliamentary election and its two-rounds of voting in the majoritarian districts. Thus, I could avoid common methodological errors and use the closeness in the first round and the turnout from the second round in my analysis.

The first hypothesis, which is also the most central one, concerns the relationship between closeness and turnout. Since there has been great variation in previous methodology, I tested four different measurements/variables of closeness. I found that all the closeness variables were significant and had an impact on turnout. For the second hypothesis, the impact of the TP-share (two-party-share) of votes in the first round of election on turnout in the second round were found to be significant across all four models. The third hypothesis could not be wholly confirmed. The mpm(multi-party-margin) and the tpm(two-party-margin) were best at predicting the relationship between closeness and turnout. This answers the research question. Closeness does affect turnout. Independent of what measurement is used and even when the TP-share of votes is controlled for. A 10% increase in the closeness margin (mpm) in the first round leads to a decrease of -4.46% in turnout in the second round of election.

# 7 Future research

In this thesis I have primarily focused on the relationship between closeness and turnout. I have also investigated the relationship between the TP-share of votes and turnout. I have partly limited my analysis on purpose to make it focused and easy to interpret. However, if future research has access to more data and is not limited by the scope of a bachelor's thesis, I have a few suggestions on how to continue the research in this field.

### 7.1.1 Campaign mobilization and spending

I have already thoroughly introduced this variable and the theory behind it in chapter 2, but I also want to finish this thesis by once more stating its importance. Previous research has found a significant impact from campaign mobilization and spending on turnout. Some (Cox & Munger 1989) view this variable as independently affecting turnout while others (Shachar & Nalebuff 1999) introduced a causal mechanism where increased closeness spurred elites and politicians to mobilize and spend more thus, increasing the turnout. This variable could not be included in this thesis due to lack of detailed data on election spending in the Georgian majoritarian districts during the 2016 parliamentary election. It is probable that this variable, if included, would impact, and change the results of this thesis. Thus, research and collecting data on campaign mobilization and spending on the regional level in Georgia should be the next step to develop this study. However, previous studies that have included this variable also found that closeness had an independent and significant effect on turnout but if this data becomes available in the future, researchers should include it in their analysis.

### 7.1.2 Between district analysis and inclusion of more variables

In chapter 3 I explain why I have chosen to look at the change in turnout within the same districts rather than just comparing the turnout between the districts. One reason is that this allows me to automatically control for several variables that might vary between the districts without having to include them in the analysis. This was useful for me since not enough data on these variables is available on district level in Georgia. Thus, research and data on these variables should be collected in the Georgian districts to allow for a deeper analysis. If data becomes available future research should not avoid these variables but include them in the analysis and test their impact on closeness. I ended chapter 2 with briefly introducing a few other variables that have been included in previous studies, some of them proven to have a significant impact on closeness. Including these variables would give us even more knowledge on the whole set of variables that affect turnout. However, since the data is not currently available, and the scope of this thesis is limited the variables were not included in this study.

### 7.1.3 Include data from upcoming election

In this study, data from 50 majoritarian districts in Georgia were used in an OLS-regression. To make the results more reliable future research could incorporate the data generated by the next Georgian parliamentary election which will be held in October 2020 (NDI 2020). I would have used data from the previous parliamentary election in 2012 but no runoffs were held in the majoritarian districts that year (ElectionsPortal 2012). Since then the election system was changed. In 2012 candidates in the majoritarian districts only needed 30% of the vote to win the district but before the 2016 election that threshold was raised to 50% (Agenda.ge 2016). This made second rounds more common in the district elections.

# 8 References


Agenda.ge, 2015. Major electoral reform ahead in Georgia. [Electronic] Available:https://agenda.ge/en/news/2015/1254 Downloaded: 2020-08-16

Cancela, Joao & Benny Geys, 2016. "Explaining voter turnout: A meta-analysis of national and subnational elections", *Electoral Studies*, Vol. 42, pp. 264-275.

Central Election Commission of Georgia, 2016a. Report on the Parliamentary Elections of Georgia October 8, 2016 (pdf). [Electronic] Available:https://cesko.ge/res/docs/AnnualReport2016-CeskoEng.pdf Downloaded: 2020-08-05

Central Election Commission of Georgia, 2016b. Parliamentary Elections of Georgia, 8 October 2016 Information Sheet (pdf). [Electronic] Available:https://cesko.ge/res/docs/InformationSheet06.10.2016.pdf Downloaded: 2020-08-08

Cox, Gary. W & Michael C. Munger, 1989. "Closeness, Expenditures, and Turnout in the 1982 U.S. House Elections", *The American Political Science Review*, Vol. 83, No. 1, pp. 217-231.

De Paola, Maria & Vincenzo Scoppa, 2014. "The impact of closeness on electoral participation exploiting the Italian double ballot system", *Public Choice*, Vol. 160, No. 3, pp. 467-479.

Downs, A., 1957. "An economic theory of democracy". New York: Harper & R

ElectionsPortal, 2012. 2012 Parliamentary – Majoritarian. [Electronic] Available:https://data.electionsportal.ge/en/event_type/3/event/32/shape/65134/shape_type/1?data_type=official Downloaded: 2020-08-16

ElectionsPortal, 2016a. 2016 Parliamentary – Majoritarian. [Electronic] Available:https://data.electionsportal.ge/en/event_type/3/event/46/shape/89486/shape_type/1?data_type=official Downloaded: 2020-08-05

ElectionsPortal, 2016b. 2016 Parliamentary – Majoritarian Runoff. [Electronic] Available:https://data.electionsportal.ge/en/event_type/3/event/50/shape/89486/shape_type/1?data_type=official Downloaded: 2020-08-05

ElectionsPortal, 2020a. About Project. [Electronic] Available:https://data.electionsportal.ge/en Downloaded: 2020-08-05

ElectionsPortal, 2020b. Data Source. [Electronic] Available:https://data.electionsportal.ge/en Downloaded: 2020-08-05

Encyclopedia, 2020. Ordinary Least Squares Regression. [Electronic] Available:https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/ordinary-least-squares-regression Downloaded: 2020-08-05

Fauvelle-Aymar, Christine & Abel François, 2006. ”The Impact of Closeness on Turnout: An Empirical Relation Based on a Study of a Two-Round Ballot”, *Public Choice,* Vol. 127, No. 3, pp. 461-491.

Matsusaka, John. G & Filip Palda, 1993. ”The Downsian Voter Meets the Ecological Fallacy”, *Public Choice,* Vol. 77, No.4, pp. 855-878.

NDI (National Democratic Institute), 2020. 2020 Georgia Parliamentary Election. [Electronic] Available:https://www.ndi.org/2020-georgia-parliamentary-election Downloaded: 2020-08-10

Shachar, Ron & Barry Nalebuff, 1999. “Follow the Leader: Theory and Evidence on Political Participation”, *The American Economic Review*, Vol. 89, No. 3, pp. 525-547.

Simonovits, Gábor, 2012. “Competition and turnout revisited: The importance of measuring expected closeness accurately”, *Electoral Studies*, Vol. 31, No. 2, pp. 364-371.

Teorell, Jan & Torsten Svensson, 2007. ”Att fråga och att svara”. Liber: Stockholm