Erasmus School of Economics

Bachelor Thesis Economics & Business Economics

Behavioral & Health Economics

Smoking behavior and Voting behavior during the Dutch elections in 2017

A research on the difference in voting behavior between smokers and non-smokers

Author: Sami Gamil

Student ID: 482618

Supervisor: Ms. E.F. de Weerd

Second assessor: Ms. W.Q. Toh

Date final version: 14-07-2021

Erasmus University Rotterdam
Making Minds Matter

Abstract

This study researches the differences in voting behavioral patterns between smokers and non-smokers. Cross-sectional data from the Dutch national elections in 2017 are used to research this topic. The data has been retrieved from the LISS panel. By means of Ordinary least squares regressions and logistic regressions, three hypotheses have been tested. No significant differences in voter turnout are found. Likewise, no differences are found in the political preferences based on the political spectrum are found. However, smokers are more likely to vote for parties that do not actively advocate against the use of tobacco. Though, this result cannot be solely allocated to the tobacco standpoints of the parties, since these parties have other political standpoints that might affect the voting behavior.

Table of Contents

1 Introduction

When people have to make a decision, you expect them to choose what is best for them. Especially decisions regarding health. Then, why do people tend to make decisions that do not optimize their health? Making non-maximizing health decisions is called risky health behavior. Risky health behaviors are one of the major causes of deaths that can be prevented (Cawley & Ruhm, 2011). These behaviors include drinking alcohol, smoking, using drugs, and unhealthy diets. According to Cawley & Ruhm (2011), people who have risky health behavior tend to be time-invariant. This means that time preferences are not consistent over time. They are strongly present-biased and thus, the benefit of the risky health behavior outweighs the cost of the bad effect on health in the future. Characteristics of these people might explain the time preferences of these people compared to the non-users of risky health behaviors.

Smoking is one of the risky health behaviors and tobacco is seen as one of the most deadly products (Chaloupka & Warner, 2000). Therefore, tobacco regulation is a priority in a lot of countries, since it leads to a lot of avoidable deaths. And with success, The worldwide use of tobacco has diminished over the last few centuries. This is mostly due to the newly obtained information regarding the negative health effects of smoking and the policies implemented for the regulation of tobacco. According to Willemsen (2011), the number of smokers in the Netherlands peaked in the 1950s and 1960s. 60% of the Dutch people smoked during this period. During that time, more information about the negative health impacts came to light. As a consequence, the number of smokers started to drop (Bruggink, 2013). According to the Trimbos Instituut (2020) and CBS (2021), 20.2% of people living in the Netherlands that are aged 18 or older smoke cigarettes and/or shag in 2020, compared to 14% in the United States (Centers for Disease Control and Prevention, 2021). This is the lowest percentage measured since the peak in 1958. Six years before, in 2014, the percentage was 25.7%, so the group of smokers has decreased by roughly 20% relative to 2014. Since 1989, the percentage of smokers has gradually decreased every year across all levels of education, gender and age groups, but at a different rate (Bruggink, 2013).

The Dutch government has implanted a national plan to minimalize the use of tobacco in the Netherlands. The main goal is to create a non-smoking generation by 2040 (Nationaal Preventieakkoord, 2018). Some of the measures that the government will take are the increase of excises starting from 2020. In 2023 excises on cigarettes will increase even more. Also, it will be prohibited to smoke on the grounds of sports clubs by 2025. Another big measure is the decrease of selling points of tobacco. These are just some of the potential targets for the (near) future. In order for these plans to go through, the government must feel supported by the Dutch citizens. The group

that is targeted by this plan to minimalize tobacco use, is the group of smokers. In order to effectively execute this plan, the behavior of smokers should be analyzed. Previous studies show that smokers are less politically active and are less likely to participate in social activities than non-smokers (Albright et al., 2015). I am interested in the reason why smokers differ from non-smokers in this aspect and the differences in political preferences between smokers and non-smokers. In order to target smokers for anti-smoking campaigns, their voting behavior and political preferences should be studied. In this way, the policymakers can accurately determine smokers' political preferences. Political parties in the Netherlands are somewhat divided in their opinions on the regulation of tobacco. They either make a big point of tackling this issue, whereas other parties think smoking is a personal subject, which should not be influenced by political interventions (Willemsen, 2011). This study is interested in finding differences between smokers and non-smokers and how it affects the political participation and preference differences between these groups.

In this research, these matters will be brought to light, by performing an empirical study. The research question that has been formulated is:

To what extent do smokers differ from non-smokers in political participation and political preference?

This research question will be answered with the help of three sub-questions:

1. What is the difference in political participation between smokers and non-smokers?

2. How do the political preferences of smokers differ from non-smokers?

3. Do smokers have a higher chance of voting on political parties who do not actively advocate against the use of tobacco and, thus, vote on basis of their own benefit?

This study can help policymaking by the government by exploring the voting behavior of the smoking population. It is important to maximize the voter turnout to make the group as representative as possible. Otherwise, democracy fails (Patterson 2002; Wattenberg, 2002). As previously mentioned, Albright et al. (2015) showed that smokers are less likely to vote than non-smokers. If the government can determine the groups that tend to have a lower voter turnout than average, a targeted campaign can be launched to get those groups to the polling station.

Furthermore, this study gives insight into whether groups tend to vote on basis of their own benefit or interest. Specifically, whether smokers tend to have a higher chance than non-smokers to vote on a political party that does not actively advocate against tobacco.

Voting behavior and risky health behavior are not often mentioned together in previous literature. A few studies have done research on smoking and voting behavior These studies have only researched

if smokers have a lower voter turnout than non-smokers in the United States. This study is the first study that compares the voter turnout of smokers and non-smokers in The Netherlands. Also, this is the first study that examines the differences in political preferences.

This study adds to the literature about the behavior of risky health users about social participation and political participation. Moreover, it contributes to the literature about self-interested voting.

2 Theoretical framework

The theoretical framework consists of two parts, which will be the basis of the theory behind the three sub-questions. First of all, the previous literature regarding the subjects of this research will be discussed. After that, the context of Dutch politics and the political spectrum are discussed.

2.1 Literature

To gain insight into the characteristics of smokers and non-smokers, the difference in characteristics of smokers and non-smokers will be discussed. Next, the different behavior patterns of smokers compared to non-smokers will be discussed based on previous articles and papers, which could help explain the differences in voting behavior. This paper is interested in finding differences in voting patterns between smokers and non-smokers. Therefore, literature about the voting behavior of smokers and non-smokers will be discussed. After that, the literature behind selfish voting is discussed to answer the third sub-question. On basis of these literature sections, the three sub-questions will be hypothesized and these hypotheses will be elucidated.

2.1.1 Differences in characteristics between smokers and non-smokers

This section will describe the general differences between smokers and non-smokers. For the regression analysis, it is useful to use these characteristics that lead to a higher chance of smoking. This could prevent omitted variable bias, which gives an inaccurate coefficient for the effect of smoking on the dependent variables.

There are numerous existing papers that examine smoking behavior and what key characteristics affect the possibility to start smoking. In 1990, Escobedo et al. have found that some sociodemographic characteristics influence the initiation of smoking. They have used the past smoking experiences of 18- to 35-year-olds to analyze their characteristics during the time of smoking initiation. The characteristics used are age, sex, race, and educational attainment. The data they used was retrieved from two different surveys which interviewed different races during 1987. They looked at people who had smoked more than 100 cigarettes in their life and were interested in the initiation age of smoking. The results show that the initiation rate for every age is higher for males than females for almost every race. Only white males and females have almost the same initiation rate. The ages at which most people start smoking are from 15 to 19 years old, the initiation rate gradually decreases with older ages. According to Escobedo et al. (1990), educational attainment also has an effect on initiation. People who graduated high school have a lower initiation rate than people who have lower education for every age of initiation. Age, sex, race, and educational attainment all affect the initiation rate of smoking.

Another study has analyzed the effect of sociodemographic characteristics. This study analyzed current smoking statuses, rather than smoking initiation. Ponniah & Bloomfield (2008) conducted a study in New Zealand to analyze smoking rates. Existing data from 2006 were used for the analysis. The results from this study are broadly in line with the study done by Escobedo et al. (1990). Ethnicity affected the smoking rate, however, the ethnicities in New Zealand are not entirely relevant for this paper, since its analyzes the smoking prevalence in the Netherlands. More importantly, age and gender showed differences in smoking rates. Males show a 2% higher smoking rate. The age group 20-24 showed the highest smoking rate where the rate gradually decreased with increasing age after the age of 24. In addition, Ponniah & Bloomfield (2008) examined the effect of the level of deprivation. The prevalence of smokers is higher in areas where people earn less money and where people are unemployed. Concluding, in addition to the effect of sex and age, this paper adds employment status and income as factors that impact smoking rates.

Sterling & Weinkam (1976) have conducted a study using data from 1970. This study goes deeper into the effect of employment status. This study shows that the type of employment impacts smoking rates. Different effects are found for the white race compared to the black race. White males show higher smoking prevalence when they are blue-collar workers compared to white-collar workers. Blue-collar workers perform manual labor and white-collar workers perform professional work and often involves an administrative setting. The study found that employed women tend to have a higher smoking rate amongst them than women who perform housework. However, the data and results may be outdated, since the study was conducted in 1970.

Lastly, Klein et al. (2013) found an effect of having a partner on the chance of smoking. Having a partner and living together reduces the chance of smoking. Moreover, the chance of having a partner also increases the chance of smoking cessation, but only if the couple is living together. This effect is supported by Kaprio & Koskenvuo (1988), who did a study that studied determinants of smoking cessation. Most of the determinants are individual aspects and characteristics. One determinant that will be pointed out is having a partner. Having a partner increased the chance of quitting smoking. This supports the idea that having a partner reduces the chance of smoking in general.

There are also some differences between smokers and non-smokers in the Netherlands observed in population statistics. According to the Trimbos Instituut (2020), men are more likely to smoke than women in 2020. 22.8% of males reported smoking compared to 17.7% of females. Also, educational attainment plays a role in the smoking rate. High educated people smoke less than low and intermediate educated people. Finally, also age is an indicator of smoking rates. People aged from 18-29 years old show the highest smoking rates, which gradually decrease after the age of 29. All of

these statistics are in line with the previously mentioned literature. Age, sex and educational attainment show the same effect in the literature and statistics of the Trimbos Institute (2017).

2.1.2 Differences in general behavior between smokers and non-smokers

This section describes the differences between smokers and non-smokers in social activities or other habits. This section could be useful to determine any mediating reasons for the differences in political participation and preferences.

A study in 2003, done by Lindström et al. has studied the relationship between social activities and smoking. In particular, it compared three groups of baseline smokers that either continued, reduced, or stopped smoking. These groups were compared to the reference population that existed of intermittent smokers and non-smokers. Logistic regression was used to find a difference between the population groups in terms of social participation. The results show that the group that kept being daily smokers had much lower rates of participating in social events compared to the reference population. These social events consist of study circles, demonstrations, meetings of organizations for example. Also, more informal events had lower participation rates amongst the daily smokers. These events are visiting a theatre, sports events or parties. Thus, (daily) smokers are less likely to participate in social events and are less involved in communities in general.

Another study, done in the United States, has also done research on social engagement among smokers (Rutten et al., 2011). This study had found similar results as Lindström et al. (2003) regarding social participation. Smokers are less likely to participate in social events than non-smokers. Furthermore, this study also is interested in trust among smokers. The results show that smokers are less likely to trust information sources than non-smokers. However, this study does not imply any causal relationships. Both directions of causality can be plausible for the effect of smoking on mistrust and social engagement.

2.1.3 Differences in voting behavior between smokers and non-smokers

Previous literature on the relationship between voting behavior and smoking behavior specifically is limited. There have been few papers that have researched the correlation between smoking and voting. Albright et al. (2015) have found that being a smoker reduces the chance of voting in the elections. This research has been conducted in Colorado. The study based their study on the findings of Lindström et al. (2003) and Lindström et al. (2000), which found that smokers are less active in social activities and are more likely to be disconnected from society. The study uses logistic regression to find the associations between smoking and voting. The study uses several control variables that this study will take into account as well. However, one control variable Albright uses, is perceived health status. This study will refrain from using this variable, since health may partially

mediate the effect of smoking on voting, thus decreasing the effect of smoking. This will be explained more clearly in the Data & Method section.

Denny & Doyle (2007) researched the effect of self-reported health on voting and the effect of smoking on voting. They used a probit model to estimate the outcomes. The results were consistent with the study of Albright (2015) in terms of smoking. Smokers tend to have a lower probability of voting in the elections. Also, self-reported health had a negative correlation with the probability of voting. Thus, individuals with poor health are less likely to vote. This confirms the assumption that health mediates the effect of smoking on voting. Mattila et al. (2013) also find that poor health decreases the chance of voting in the elections.

Hersch et al. (2004) have executed more specific research on the regulation of smoking. This study examines the voters' preferences for smoking restrictions in public places at the state-level in the United States. This study's findings are also consistent with the result that smokers are less likely to vote in the elections. However, this study goes more into detail about the preferences of smokers. Smokers are less likely than non-smokers to be in favor of smoking restrictions in public places.

2.1.4 Self-interested voting

Several previous studies have researched the effect of self-interest and selfishness on voting behavior. Elinder et al. (2008) did a study on pocketbook voting. This is voting behavior that benefits a person the most financially. This study was conducted in Sweden using data from the National elections of 1994 and 1998. The sample consisted of the treatment group which was parents with young children. The control group was parents with older children. A political party proposed reductions in financial support to parents with young children in 1994. The results showed that the treatment group's vote share on that party was 12.5% than the control group. Four years later, in 1998, the same party promised to cap the child care fees which should benefit the treatment group, this resulted in an increase in votes from the treatment group. The treatment group's vote share on that party was 13.5% higher than the control group. This analysis shows clear evidence on the prospective pocketbook voting.

Additionally, DeScioli et al. (2020), found in their experiment with real payoffs that the individuals often chose for the policy that maximized their own payoff instead of the policy maximizing the society's payoff. Furthermore, Acemoglu & Robinson (2006) stated that voters are in favor of the tax rates that maximize their own wealth according to the Static model.

2.1.5 Hypotheses

Based on the previous literature the following hypotheses will be made.

Hypothesis 1: Smokers are less likely to have voted in the Dutch elections of 2017 than non-smokers.

The articles of Albright et al. (2015), Denny & Doyle (2007) and Hersch et al. (2004) all found that smokers significantly have a lower probability of voting than non-smokers. Based on these findings it is expected that smokers in the Netherlands are less likely to have voted in the Dutch elections of 2017 than non-smokers.

Hypothesis 2a: Smokers are less likely to vote 'left-wing' than non-smokers in the Dutch national elections of 2017.

Smokers tend to be less likely to participate in social events than non-smokers. Generally, smokers are less involved in any kind of community (Lindström et al., 2003). Not being a part of social groups or attending social events leads to a more individualistic approach, which can be an aspect of a right-winged approach. Right-winged parties support individualism and accept inequality and unequal chances. Left-winged parties, on the contrary, strive for equality and value society higher than the individual. Due to the social disengagement of smokers, the expectation is that smokers tend to be more right-winged and thus, less likely to vote left-wing than non-smokers in the Dutch national elections of 2017.

Hypothesis 2b: Smokers are less likely to vote progressive than non-smokers in the Dutch national elections of 2017.

This hypothesis is based on a similar realm of thought as with hypothesis 2a. Progressivism is defined as a society that is responsible for each other and the society itself is more important than the individual itself. The common good is valued highly in progressivism and innovation is required to enhance society as a whole. Again, due to the lack of communities and societies among smokers, the expectation is that smokers tend to be more conservative and are less likely to vote progressive than non-smokers in the Dutch national elections of 2017.

Hypothesis 3: Smokers are less likely to vote for parties that are specifically in favor of the regulation of tobacco in the Dutch national elections of 2017.

According to the discussed papers in section 2.1.4, voters tend to act self-interested regarding voting. Especially in terms of financial wealth. Elinder et al. (2008) and DeScioli et al. (2020) both found clear evidence of pocketbook voting in both their studies. Tobacco regulation largely consists of an increase in excises. Consequently, voting against tobacco regulation may be financially beneficial to

smokers. Additionally, smokers that vote self-interested are against any form of tobacco regulation given their preferences.

2.2 Research Context

The political spectrum and the Dutch political system will be explained in the *Research Context* section of this theoretical framework. The political spectrum will be used to categorize the political preferences of the individuals. The Dutch political system will be briefly explained by describing the election process and the framework of the parliament. After that, the national elections of 2017 are summarized.

2.2.1 Political spectrum

To divide the Dutch political parties into groups based on their believes and standpoints the political spectrum is often used in the Netherlands. The political spectrum divides parties in two ways. Left-wing/right-wing and progressive conservative. Figure 1 shows a self-constructed typical structure of the political spectrum.
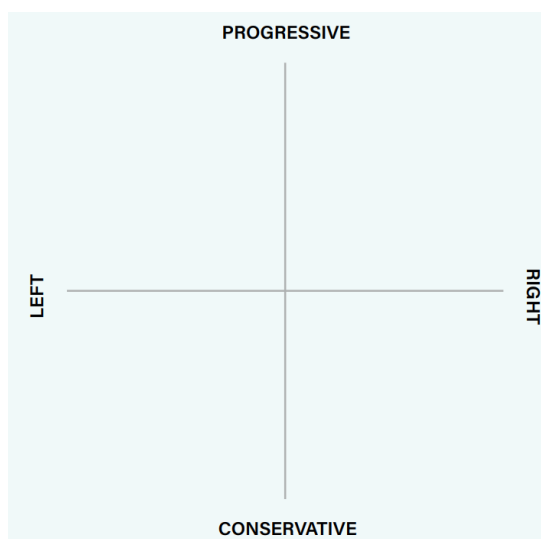
Figure 1 political spectrum

On the y-axis, the parties are divided between progressive and conservative on the x-axis, the parties are divided between left-wing and right-wing.

The political spectrum divides parties in two different ways. Left-wing/right-wing and progressive/conservative. Left-wing is described as striving for equality. The left-wing right-wing distinction is generally based on economical matters like equality in income and the role of the government in the economy (Eisinga, 1997). The political parties that are left-wing strive for equal rights and equal chances (Bobbio, 1996). This is translated to the society by giving everyone access to higher education, basic insurance, and leveling taxes structures. The right-wing parties justify inequality. Bobbio (1996) shows two theories for justifying inequality: People are born equal, but

society and environment shape them into unequal persons (Rousseau, 1992). The other theory is from Nietzsche (1973), who states that men are born equal, which is good for the balance in society. In short, the left-wing favors equality and equal chances. Right-wing legitimizes inequality and sees wealth as a consequence of hard work and dedication, while the lazy and uninspired people are poor (Jahn, 2011). These definitions of the left-wing and right-wing can be supported by the article of NOS (2017), who defined left-wing as a form of socialism, which states that all decisions are made in favor of the society instead of self-interest. This leads to equality with equal chances. Right-wing is defined as a form of libertarianism, which is more individualistic. The Libertarians seek freedom and little interference by the government in the market. Money is earned by yourself and can be spent on yourself.

The distinction between Progressive and Conservative is, broadly, the best source for policymaking. Conservatives highly value the Constitution and private power. Progressives strive for more equality and do not value the Constitution as high as the progressives. Change in society is supported. The conservatives are not likely to deviate from the Constitution, according to West (1990). Conservative in general context means being averse to change or innovation and keeping traditions intact. by way of contrast is progressive the opposite, progressivism supports innovation and requires innovation to enhance the society.

2.2.2 Politics in the Netherlands

The Netherlands is a parliamentary democracy, this means that people, 18 years or older, who have Dutch citizenship are able to vote in the elections. In general, every four years there is a national election for the house of Representatives (Tweede Kamer, 2021). The House of Representatives consists of 150 members that are chosen by the public. Together with the Senate, the House of Representatives forms the Dutch Parliament. The Dutch Parliament is responsible for making new laws in cooperation with the government. The government consists of elected members and the king. Each member in the House of Representatives represents a political party. Voters cast their vote on a political party in these national elections. Voters can also choose to cast a blank vote, which is a vote cast on none of the eligible people. It is mostly done as a protest against the available options. Afterward, it becomes clear how many votes each party has obtained, and thus, how many seats in the House of Representatives they have obtained. There is a total of 150 seats. After the elections, parties will work together to form a coalition, which must have a majority of the seats in the House of Representatives, so a minimum of 76 seats is required. The remaining parties will form the opposition. The Coalition determines the Cabinet, which consists of the Prime Minister, the other ministers, and the State Secretaries (Tweede Kamer, 2021).

In this research, the focus will be on the national elections of the House of Parliament in 2017. the main goal is to find out if there are any differences between smokers and non-smokers regarding voter turnout and political preferences. This will be examined by looking at the results of the national elections of 2017. There are many political parties that did not get a seat in the House of Representatives in 2017. To keep the data relevant and clear, only the political parties that managed to get a seat will be taken into account in this research. In table 1, the results from the elections in March 2017 are shown. It shows that there is a total of 13 different parties that have obtained a seat in the House of Representatives.

Political Party	Seats	Votes	Votes (%)
VVD	33	2,238,351	21.29
PVV	20	1,372,941	13.06
CDA	19	1,301,796	12.38
D66	19	1,285,819	12.23
GL	14	959,600	9.13
SP	14	955,633	9.09
PvdA	9	599,699	5.70
CU	5	356,271	3.39
PvdD	5	335,214	3.19
50Plus	4	327,131	3.11
SGP	3	218,950	2.08
DENK	3	216,147	2.06
FvD	2	187,162	1.78
Other parties	-	161,327	1.51
Turnout	**150**	**81.9%**	**100**

Table 1 Results national elections 2017 the Netherlands (Source: Kiesraad Databank Verkiezingen (2017))

3 Data

To obtain an answer to the research question and the corresponding sub-questions, empirical research will be executed. For this study, the data from the LISS panel will be used. The LISS panel (Longitudinal Internet Studies for the Social Sciences) consists of 7,500 individuals that get paid to complete questionnaires every month. Based on a true probability model, the households are drawn from the population register in The Netherlands. The individuals are thus all Dutch citizens. The data consists of several studies. The general background information of the individuals is gathered by a single-wave questionnaire, which can be updated every month. The LISS panel also has 10 Core Studies, which are longitudinal studies that are carried out every year. These Core Studies each have their specific theme.

Two Core Studies will be used in this research, *Health* and *Politics & Values.* Each study has been carried each year, this means that the studies have several waves. For this study, voting behavior during the elections in 2017 will be analyzed. Therefore, the data about voter participation and preferences will be retrieved from the 10th wave of the *Politics & Values* study. The reason the 10th wave has been chosen, is that it is the first wave after the national elections. The 9th wave has been administered before the elections. Political preferences are mostly formed around the time of the elections. Thus, a survey before the elections does not contain information about the upcoming elections. The 9th wave only contains information about the previous elections that were held in 2012. Therefore, to gather data of the elections of 2017, a survey administered after the elections is needed. This questionnaire from the 10th wave contains information about the most recent elections in March 2017. This is the first questionnaire about *Politics and Values* after the national elections in March 2017. This wave was administered from December 2017 until February 2018. This questionnaire consists of three parts. The interest lies in the voter turnout and the political preferences of the voters during the elections in March 2017. The answers to these questions are found in the second part of the questionnaire, which has been administered in January 2018. From this study, variables that will be retrieved are if the person has voted and for what party the individual has voted

The *Health* Core Study provides data on the smoking statistics of the individuals in the sample. This study also is administered each year. The 9th wave of the *Health* Study is used to retrieve the data. This study was executed from November 2016 until December 2016. This questionnaire was administered three months before the elections. Political parties released their election campaigns in late 2016 and during these months until March 2017, the individuals form their opinions on the released election campaigns. Consequently, their political preference for the upcoming elections will

be formed. During this period, the survey was completed about the smoking statistics. Since we are interested in the decision-making and political activity of smokers. The 9th wave is used. The only assumption that will be made is that individuals who reported to be smokers during November and December 2017, during the administration of the 9th wave, are assumed to still be smokers during the elections in March 2017.

The background information about the individuals will be retrieved from the Background Study. This questionnaire can be updated by every individual updated each month in case any changes affect the background variables. The statistics about smoking dates from November 2017. Consequently, the updated version of the background variables from November 2017 will be used as the source for the background data. Variables about marital status, age, gender, occupation, education, income, and ethnicity will be retrieved. These variables will serve as control variables in this research.

In this research, we are interested in all citizens in the Netherlands who are eligible to vote. Accordingly, observations younger than 18 years and/or not eligible to vote have been removed from the sample. Furthermore, people who forgot if they voted are removed. 2,027 observations have been deleted that were under the age of 18. This is 17.4% of the total sample. The number of observations that remain is 9,612. After that, 214 out of 5,903 observations, that answered the question about voting in the 2017 elections, that were not eligible to vote or forget if they voted have been removed. The observations that have been removed were 3.6% of the total observations.

4 Method

This section consists of two parts: first, the variables will be described and the statistics of each variable will be shown in a table. Then, a correlation table will be shown which includes every independent variable, to assure that there is no multicollinearity. The second part will describe the regression methods. The reason for choosing the regression methods will be explained, as well as a check on every assumption that this regression method requires.

4.1 Variables

4.1.1 Dependent variables

This section describes the dependent variables for each hypothesis. There are four dependent variables in total. For hypotheses 2a, 2b, and 3, everyone who voted blank, on a party without a seat in the House of Representatives or did not vote has been removed from the sample.

Voter turnout

The dependent variable of the first hypothesis is the voter turnout of the Dutch elections in 2017. Hypothesis 1 is stated below:

Hypothesis 1: Smokers are less likely to have voted in the Dutch elections of 2017 than non-smokers.

The variable is called *Voter* and is a binary variable which that have the value 0 (= non-voter) or 1 (= voter). This variable is retrieved from the LISS panel, Health & Values 10th wave. The respondents were asked whether they have voted in the most recent elections (March 2017).

Political preference

The dependent variable of the second hypothesis is political preference. Hypotheses 2a and 2b are stated below:

Hypothesis 2a: Smokers are less likely to vote 'left-wing' than non-smokers in the Dutch national elections of 2017.

Hypothesis 2b: Smokers are less likely to vote progressive than non-smokers in the Dutch national elections of 2017.

The political preference will be measured in two different ways. Left-Right and Progressive-Conservative. Figure 2 shows the political spectrum for the 13 parties that were voted into the House of Representatives in 2017, where Progressive-Conservative is on the Y-axis and Left-Right is on the X-axis.

Figure 2 Dutch political parties in the political spectrum (Source: Boersema, 2021)

The political parties will be divided into a left-wing/right-wing group and a progressive/conservative group based on figure 2. The first and second column of table 2 shows how the parties have been categorized. The spectrum is based on the newspaper article by Boersema (2021) in the Dutch newspaper Trouw.

The parties will also be categorized based on their standpoint regarding tobacco regulation. The parties present their political agenda a few months before the elections. The political agenda is the set of plans of the party for the period after the elections until the next elections. Therefore, the political agendas of the elections in 2017 are used to determine the standpoint against tobacco regulation. The parties will be divided into two groups. The first group is active tobacco regulation and the other group is passive/non-tobacco regulation. the standpoints of the 13 parties are stated[1] in table A of the appendix. Column 3 of table 2 shows how the parties have been categorized.

---

[1] Translated from Dutch to English

Political party	Left-wing/Right-wing	Progressive/Conservative	Active tobacco regulation/passive or no tobacco regulation
VVD	Right-wing	Conservative	passive or no tobacco regulation
PVV	Right-wing	Conservative	passive or no tobacco regulation
CDA	Right-wing	Conservative	Active tobacco regulation
D66	Right-wing	Progressive	Active tobacco regulation
GL	Left-wing	Progressive	Active tobacco regulation
SP	Left-wing	Progressive	passive or no tobacco regulation
PvdA	Left-wing	Progressive	passive or no tobacco regulation
CU	Left-wing	Conservative	Active tobacco regulation
PvdD	Left-wing	Progressive	passive or no tobacco regulation
50Plus	Left-wing	Progressive	passive or no tobacco regulation
SGP	Right-wing	Conservative	passive or no tobacco regulation
DENK	Left-wing	Progressive	passive or no tobacco regulation
FvD	Right-wing	Conservative	passive or no tobacco regulation

Table 2 political parties and their characteristics

For hypothesis 2a, the dependent variable is a binary variable called *Left* that indicates whether the person has voted for a party classified as right or left. The variable has 2 values, 0 (= right) and 1 (= left). This variable has been transformed to this binary variable with the help of the LISS panel variable that indicates for which party the individual has voted. All of the 13 parties have been classified as left or right so that that only 2 groups remain.

The dependent variable of hypothesis 2b is also a binary variable, this variable is labeled *Progressive.* The binary variables can have the values 0 (= conservative) and 1 (= progressive). The transformation of this variable is almost identical to the transformation of the dependent variable of hypothesis 1. The same variable from the LISS panel has been used to indicate the party preference of every respondent. For this hypothesis, however, the 13 parties have been classified as progressive or conservative instead of left-wing or right-wing. Hence, the voters' preferences have been divided into two groups.

Self-interested voting

The final sub-question focuses on the selfish behavior of smokers regarding tobacco regulation.

The hypothesis is stated below:

Hypothesis 3: Smokers are less likely to vote for parties that are specifically in favor of the regulation of tobacco in the Dutch national elections of 2017.

The dependent variable is a binary variable '*Pro_Regulation*' This variable is somewhat similar to the dependent variables in hypotheses 2a and 2b. The two values are 0 (= anti-regulation) and 1(=Pro_Regulation). This variable has been transformed into a binary variable with the help of the same process as in hypotheses 2a and 2b. The political parties are divided into an active tobacco regulation group and a passive/no tobacco regulation group.

4.1.2 Independent variable

Smoker

The main independent variable is the same for all hypotheses. This is a binary variable called *Smoker*, which indicates whether the individual is a smoker (= 1) or a non-smoker (= 0). This variable is composed by combining two variables from the LISS panel. The first variable is a binary variable that shows whether a respondent has ever smoked or has never smoked. The second variable is also a binary variable. This question was only asked to the respondents who answered '1' in the previous question. The follow-up question is "do you still smoke now", the two possible answers are "Yes" and "No, I stopped.". The variable *smoker* has the value 0 when the respondent reported to have never smoked or had quit smoking. Value 1 is when the respondent reported to have smoked and is still smoking today.

4.1.3 Control variables

Control variables are used to estimate the effect of smoking on the dependent variables as accurately as possible. Control variables increase the internal validity by minimizing the number of confounders and omitted variables. In this estimation, there are other factors that influence smoking behavior and voting behavior. Therefore, control variables that affect both the independent and dependent variables will be added into the estimation. The theoretical framework has explained which variables are necessary to include in the regression analyses. A set of control variables will be used based on literature that shows these variables could be associated with both smoking and voting behavior. The control variables that will be used in the estimations are age, gender, income, education, occupation, and marital status.

Age

The variable for age has directly been used from the age variable in the LISS panel from the background variables. This is a continuous variable. The name of this variable is *Age.*

Gender

The variable that determines gender is a binary variable called *Male.* this variable has been copied from the LISS panel data. Value 1=male and value 2=female.

Income

Several income variables are listed in the background variables study of the LISS panel. Four variables have been examined via Stata. The four variables from the LISS panel are gross income, imputed gross income, net income, and imputed net income. The imputed net incomes are an estimate based on the gross income if net income is not entered and vice versa. The variable with the least missing variables and outliers has been chosen as the income variable in this research, to keep the sample as large and accurate as possible. This variable is the imputed net income variable. The imputed net income will be used for this research and is renamed to *net_income.* To be able to interpret the coefficient of income better, the income will be standardized by dividing it by 1000. This variable is a continuous variable.

Education

The variable *education* is a transformation of the education variable from the LISS panel. The LISS panel divides the levels of education into six groups: primary school, vmbo (intermediate secondary education, US: junior high school), havo/vwo (higher secondary education/preparatory university education, US: senior high school), mbo (intermediate vocational education, US: junior college), hbo (higher vocational education, US: college) and wo (university).

In this research, these six groups will be formed into three groups according to the CBS (Central Bureau of Statistics Netherlands): low educated, intermediate educated, and high educated. These groups are also used in the yearly report about smoking statistics of the Trimbos Institute (2020). These are the most used categories in the Netherlands. The low educated group consists of primary school and vmbo. The intermediate educated people have finished havo/vwo or mbo. Finally, the high educated group consists of hbo and wo. These categories have been turned into separate dummies for the logistic regression, with Low educated as the reference category.

Occupation

The occupation variable from the LISS panel contained 14 categories. This has been reduced to six categories in this research to generalize the results and categories. Because all observations with an age below have been removed, the 14th category *Too young to have an occupation* is automatically removed. The remaining 13 occupation categories have been generalized into the following six categories: *unemployed*, *retired, employed, homemaker, student,* and *unable to work/disabled.* A homemaker is a person who manages the household. Usually, the partner is employed. All unpaid activities have been labeled as unemployed. The variable *occupation* is a categorical variable and all

of these categories have been turned into separate dummies where unemployed is the reference category.

Partner

The partner variable indicates whether the individual lives with a partner, married or unmarried. This variable is copied from the LISS panel background variables study. No transformations have been made. The variable *partner* is a binary variable with the values 0 (= no partner) or 1 (= partner).

4.1.4 Descriptive statistics

Table 3 shows the descriptive statistics of all the variables used in the regressions. The table shows that the voter turnout of the entire sample was 88.7%. The sample consists of people eligible to vote. The entire voter turnout in the elections of 2017 in the Netherlands was 81.9% (Kiesraad, 2017). Thus, the percentage of people that voted is higher in this sample than the actual percentage in the population. The statistics also show that the ratio left/right and progressive/conservative is almost 0.5, which means that the sample is nearly equally divided. In the national elections, the left/right ratio was 0.375 and the progressive/conservative ratio was 0.463Table 3 shows that there are no outliers in the two continuous variables *age* and *income*. This shows that the removal of the outliers in *income* has been executed successfully. The table also shows that 14.8% percent of the observations have indicated that they smoke. According to the Trimbos Institute (2017), the percentage of smokers was 23.1%, this is a significant difference that might affect the external validity of this research. The percentage of left-wing voters in the sample is 12% higher in the sample than in the population (Kiesraad, 2017) For progressive voters the difference is 4.6%. The population means could not be found for every variable. For the income variable net median income is used as population mean.

Variable	Obs	Mean sample	Mean population	Std. Dev.	Min	Max
Voter	5370	.887	.819	.316	0	1
Progressive	4488	.509	.463	.500	0	1
Left	4488	.495	.375	.500	0	1
Smoker	4604	.155	.231	.362	0	1
Male	8664	.486	.494	.500	0	1
Partner	8664	.712	-	.453	0	1
Net income	8664	1562.04	2066.04	1103.546	0	11200
Age	8664	48.971	42.2	18.221	18	104
Low educated	8638	.273	.280	.445	0	1
Int educated	8638	.365	.410	.482	0	1
High educated	8638	.362	.290	.481	0	1
Unemployed	8664	.061	.036	.240	0	1
Retired	8664	.200	.183	.400	0	1
Employed	8664	.538	-	.499	0	1
Homemaker	8664	.071	-	.256	0	1
Student	8664	.092	-	.290	0	1
Unable	8664	.038	-	.191	0	1

Table 3 descriptive statistics (Source Column 4, Mean population: Trimbos & CBS)

Table B of the appendix shows the correlation matrix of all the independent variables. No variables are very highly correlated with each other. The highest significant correlation coefficients are mostly caused by the occupation dummies. The highest coefficient is 0.648 for *age* and *retired*. This is not a surprisingly high coefficient since the standard retirement age is 67 is the Netherlands. Thus, the higher the age, the higher the chance of retirement. The same is the case with *student* and *age.* The coefficient is -0.465. Which indicates a negative correlation. Also *income* and *employed* are somewhat correlated with each other, with a coefficient of 0.408. However, none of the correlation coefficients are high enough to cause problems in the regression by causing multicollinearity. This will be further explained when testing the logistic regression assumptions.

4.2 Method of analysis

This section will describe how the hypotheses will be tested. The decision behind the methods of analysis and the methods will be explained. Two methods will be discussed, Ordinary least squares (OLS) regression and logistic regression.

4.2.1 Ordinary least squares regression

OLS is a linear regression model that predicts the linear relationship between two or more variables. OLS finds the constant and coefficient by minimizing the squared errors. The errors are squared to remove the negative signs and to increase weight of larger differences. An OLS regression will be done for all hypotheses. For all hypotheses, the set of control and independent variables will be the same. The left side, the independent variable, differs for each hypothesis.

The following regressions will be executed:

$$Y_i = \beta_o + \beta_1 Smoker_i + \beta_2 Partner_i + \beta_3 Male_i + \beta_4 LowEducated_i$$
$$+ \beta_5 IntermediateEducated_i + \beta_6 HighEducated_i + \beta_7 Retired_i$$
$$+ \beta_8 Employed_i + \beta_9 Homemaker_i + \beta_{10} Student_i + \beta_{11} Unable_i + \beta_{12} Age_i$$
$$+ \beta_{13} NetIncome_i + \mathcal{E}_i$$

Where Y_i is the outcome variable for all hypotheses. Thus Y_i is *Voter* for hypothesis 1, *Left* for hypothesis 2a, *Progressive* for hypothesis 2b and *ProRegulation* for hypothesis 3.

The outcome variable is a binary variable that represents the probability of that event happening. In a linear regression model, the outcome variable has to be continuous. When doing a linear regression, the model sees a binary outcome variable as a continuous outcome variable, which is not the case. The 0 and the 1 of a binary variable do not indicate a rank or increase. This means that the linear model can produce negative values and values above 1. When estimating a probability, those values are not interpretable. OLS also assumes homoskedasticity, which is also automatically violated with a binary outcome variable. Therefore a linear regression is not suitable when the dependent variable is binary. Consequently, a logistic regression will be executed to overcome this issue. The logistic regression will be explained in section 4.2.2.

4.2.2 Logistic regression

A logistic regression will be used to test the hypotheses. A logistic regression is an appropriate regression analysis when the dependent variable is binary. It is used to predict the effect of the independent variable(s) on a binary outcome variable. The logistic model generates a probability of a certain outcome. The outcome to test the first hypothesis is voter/non-voter. A person with certain

characteristics may have a higher probability of voting than a person with other characteristics. The logistic model predicts these probabilities, as well as their sign and significance.

When running a logistic regression, the coefficients of the variable are not as easy to interpret as with a linear regression. In general, the logistic regression can only estimate the marginal effect of a coefficient keeping all other variables constant at a reasonable value. Therefore, only the sign will be interpreted, the magnitude of the coefficient can only be measured for one specific individual with certain characteristics at a time. A coefficient of a variable in a linear regression simply shows how much the outcome variable changes when the independent variable increases by one. When running a linear regression with a binary outcome, the linear function represents a probability. The linear regression model can be transformed into a logistic regression model by using the logistic function.

The logistic regressions that will be executed are as follows:

$$Pr(Y_i) = \frac{1}{1 + e^{-(\beta_o + \beta_1 Smoker_{it} + \cdots + \beta_{13} NetIncome_{it} + e_{it})}}$$

Where Y_i is the outcome variable for all hypotheses. Thus Y_i is Voter for hypothesis 1, Left for hypothesis 2a, *Progressive* for hypothesis 2b and *ProRegulation* for hypothesis 3.

A logistic regression requires 6 assumptions (Laerd Statistics):

1. The dependent variable should consist of two categorical, independent groups.
2. There are two or more independent variables, which should be measured at the continuous or nominal level.
3. Independence of observations.
4. The data must not show multicollinearity.
5. There needs to be a linear relationship between any continuous independent variables and the logit transformation of the dependent variable.
6. There should be no significant outliers, high leverage points or highly influential points.

For the regression to be internally valid, these six assumptions should be met. The first assumption is already met since the dependent variable is a binary variable for every hypothesis. The second assumption is also easy to check. All variables in the regression are either continuous or nominal. The categorical variables do not have an intrinsic order.

The third assumption requires more thought. Independence of assumptions means that each observation is independent and not related to one another. There is no way to test this assumption. However, the sample consists of all independent individuals who have filled in the questionnaires. The only way an individual could affect another individual is by living in the same household. I expect

this result to be marginal and thus, will not affect the validity of the research. Therefore, I assume that assumption 3 is met.

The fourth assumption, the data must not show multicollinearity, means that the independent variables should not be extremely correlated with each other. This means that the Pearson correlation coefficient should not exceed an absolute value of 0.8 (Franke, 2010). Table B shows the correlation coefficients of every combination of independent variables, the highest correlation coefficient is 0.648 which is lower than 0.800. Therefore, the data does not show multicollinearity. Assumption 4 is met.

Assumption 5 says the relationship between the continuous independent variables and the logit transformation of the dependent variable is linear. The Box-Tidwell approach was used to check this assumption. The Box-Tidwell approach transforms the continuous variables into an interaction of the log transformation of the variable(x) and the variable x itself. In this regression, the variables *age* and *Income* are the only continuous variables. The new generated variables are *lnincome* and *lnage.*

These new variables are added to the original regression. If these variables showed a significant coefficient, which means a p-value lower than 0.05, these transformation additions made a better prediction. If this is the case, the fifth assumption is not met. Table C of the appendix shows the logistic regression results with the log transformations of all hypotheses. The stars of significance are shown to easily read off the significance. If the coefficient of the log transformations has two or three stars, the regression does not meet assumption 5. Column 1 of table C shows the results for hypothesis 1. None of the log transformations of *age* and *net_income* have a significant coefficient. This is also the case for hypothesis 3, which is shown in column 4 of table C. for hypotheses 2a and 2b however, *ln_age* shows a significance level of 99%. This means that the log transformation of *age* has made a better prediction of the model than *age.* Therefore, assumption 5 is not met for hypotheses 2a and 2b. On the contrary, hypotheses 1 and 3 do meet the requirements of assumption 5.

The sixth assumption states that there are no outliers significant outliers. This assumption can be checked by checking table 1. Because there are only two continuous variables. These have to be checked, the (binary) dummy variables are ruled out of outliers. The minimum and maximum value of *age* are 18 and 104 respectively and shows no outliers. Graph I in the appendix shows visually that outliers are non-existent. *Income* shows no outliers either according to graph II in the appendix. The minimum income is 0 and the maximum is 11,200, which is realistic for the Dutch population. Assumption 6 is met.

5 Results

This chapter will present the results of the regression analyses that have been described in the previous chapter. The results consist of 3 hypotheses that all have been tested through two estimation methods: OLS and logistic regression.

5.1 Hypothesis 1

The first hypothesis is tested through an OLS regression and a logistic regression. The hypothesis is stated below.

Hypothesis 1: Smokers are less likely to have voted in the Dutch elections of 2017 than non-smokers.

The result of the logistic regression is shown in the first column of table 4. No significant effect can be found for the effect of being a smoker on voting. This means that there is no significant difference between smokers and non-smokers on voter turnout ceteris paribus. The coefficient is negative which means that in this sample, being a smoker decreases the probability of having voted in the Dutch national elections of 2017. This result is in line with the expectations, however, since this coefficient is not significant, hypothesis 1 cannot be accepted.

Having a partner significantly increases the probability of voting on a 1% significance level ceteris paribus, the coefficient shows a positive significant value which means a higher probability of the outcome variable *Voter*. Also, intermediate and high educated also have a significant coefficient on a 1% significance level when all other variables stay constant. Thus, the level of education has a significant effect on voter turnout. The coefficients of *Intermediate education* and *High education* are both positive (0.513 & 1.264 respectively), meaning they both increase the probability of voting relative to *low education* where *High education* has a bigger effect on voting than *Intermediate education* ceteris paribus*. Age* also has a positive significant coefficient which indicates that an increase in age increases the probability of voting significantly on a 1% significance level. Students have a significantly higher chance of voting than people that are unemployed on a 1% significance level. Males showed a lower probability of voting compared to females in this sample. However, this coefficient is not significant, thus an actual difference between males and females cannot be assumed. All results are ceteris paribus.

Column 1 in table 5 shows the OLS regressions, which can be used to estimate the magnitude of effects. Having a partner increases the probability of voting by 6.6% and is significant on a 1% significance level. Intermediate and high educated people have a 5.3% and 11% higher voting percentage than low educated people respectively ceteris paribus. This result is also significant on a

1% significance level. Students have a 7.1% higher voter turnout than unemployed people ceteris paribus on a 1% significance level.

VARIABLES	(1) Voter	(2) Left	(3) Progressive	(4) Pro regulation
Smoker	-0.064	-0.159	0.123	-0.557***
	(0.135)	(0.101)	(0.101)	(0.109)
Partner	0.655***	-0.230***	-0.466***	0.099
	(0.111)	(0.081)	(0.081)	(0.082)
Male	-0.087	-0.068	-0.155*	-0.242***
	(0.117)	(0.080)	(0.080)	(0.081)
Int educated	0.513***	0.194**	0.373***	0.166*
	(0.127)	(0.096)	(0.096)	(0.100)
High educated	1.264***	0.321***	0.741***	0.662***
	(0.161)	(0.102)	(0.102)	(0.103)
Retired	-0.076	0.077	-0.166	0.137
	(0.262)	(0.171)	(0.172)	(0.175)
Employed	-0.141	-0.037	-0.099	-0.091
	(0.233)	(0.165)	(0.166)	(0.169)
Homemaker	-0.381	-0.111	-0.442**	-0.089
	(0.273)	(0.201)	(0.202)	(0.206)
Student	0.692**	0.098	0.440*	0.762***
	(0.315)	(0.252)	(0.257)	(0.257)
Unable	-0.488*	0.290	0.104	-0.172
	(0.288)	(0.236)	(0.237)	(0.248)
Age	0.033***	0.023***	0.007**	-0.003
	(0.005)	(0.004)	(0.003)	(0.004)
Net income	0.164**	-0.183***	-0.218***	-0.020
	(0.080)	(0.046)	(0.047)	(0.044)
Constant	-0.726**	-0.932***	0.045	-0.347
	(0.347)	(0.270)	(0.268)	(0.274)
Observations	4,004	3,369	3,369	3,369

Standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Table 4 logistic regression analyses

5.2 Hypothesis 2a

Hypothesis 2 consists of the exact same independent and control variables as hypothesis 1. The outcome variable is the dummy variable *Left-wing.* The hypothesis states that smokers are less likely to vote left-wing.

Hypothesis 2a: Smokers are less likely to vote 'left-wing' than non-smokers in the Dutch national elections of 2017.

Column 2 of table 4 shows that there is no significant difference in political preference in terms of left-wing/right-wing between smokers and non-smokers. The coefficient of *Smoker* is -0.159 which indicates a negative effect of *smoker* on the probability of voting *Left-wing.* Though, the coefficient is not significant. This means that no significant difference in political preference of left-wing/right-wing between smokers and non-smokers can be concluded. Hypothesis 2a is not supported by this result.

Control variables that show a significant coefficient are *Partner,* both education categories, *age,* and *Net Income.* All coefficients are significant on a 1% significance level except for *Intermediate education*, which is significant on a 5% significance level. This means that having a partner, higher age and a high income all affect the probability of voting left-wing. Having a partner and a higher income decrease the probability of voting left-wing. An increase in age has a positive effect on the probability of voting left-wing. Also intermediate and high educated people show a higher probability of voting left-wing than low educated people. The other control variables do not have a significant coefficient. Thus an effect of these variables cannot be assumed.

The OLS results in column 2 of table 5 show that having a partner decreases the probability of voting left-wing with 5.4% ceteris paribus on a 1% significance level. The results also show that intermediate and high educated have a 4.6% on 5% significance level and a 7.5% on a 1% significance level higher probability of voting left-wing than low educated people. Lastly, people with an income 1000 euros higher than another individual are 4.2% less likely to vote on a left-wing party. All results are ceteris paribus.

5.3 Hypothesis 2b

Hypothesis 2b has been executed the same way as hypothesis 2b. The only difference is the outcome variable, which is *Progressive* instead of *Left-wing.* Column 3 of table 4 shows the results of the logit regression for hypothesis 2b.

Hypothesis 2b: Smokers are less likely to vote progressive than non-smokers in the Dutch national elections of 2017.

The coefficient of *Smoking* shows that smokers are more likely to vote progressive than non-smokers. This is against the expectation. However, the coefficient is not significant on a 5% significance level. A difference in political preference in terms of voting progressive/conservative between smokers and non-smokers cannot be concluded, thus hypothesis 2b will be rejected.

The control variables show the same effect as column 2 for hypothesis 2a. The only differences are the significance levels and *Homemaker.* The significance levels remain significant at least on a significance level of 5%. *Homemaker* significantly differs from *Unemployed*. The coefficient of homemakers is significantly negative which implies that homemakers are less likely to vote progressive than unemployed people. The effect of OLS is also in line with hypothesis 2a, however, the magnitude of all the significant coefficients is bigger. Homemakers are 10.7% less likely to vote progressive than unemployed people ceteris paribus on a 1% significance level.

VARIABLES	(1) Voter	(2) Left	(3) Progressive	(4) Pro regulation
Smoker	-0.010	-0.038	0.030	-0.122***
	(0.014)	(0.024)	(0.024)	(0.024)
Partner	0.066***	-0.054***	-0.111***	0.023
	(0.011)	(0.019)	(0.019)	(0.019)
Male	-0.013	-0.017	-0.039**	-0.056***
	(0.011)	(0.019)	(0.019)	(0.019)
Int educated	0.053***	0.046**	0.090***	0.037
	(0.013)	(0.023)	(0.023)	(0.023)
High educated	0.110***	0.075***	0.177***	0.156***
	(0.014)	(0.024)	(0.024)	(0.024)
Retired	-0.009	0.018	-0.041	0.032
	(0.024)	(0.041)	(0.041)	(0.041)
Employed	-0.000	-0.012	-0.026	-0.021
	(0.022)	(0.040)	(0.040)	(0.039)
Homemaker	-0.034	-0.026	-0.107**	-0.021
	(0.027)	(0.048)	(0.049)	(0.048)
Student	0.071**	0.022	0.103*	0.182***
	(0.033)	(0.060)	(0.061)	(0.060)
Unable	-0.058*	0.070	0.025	-0.036
	(0.031)	(0.057)	(0.057)	(0.056)
Age	0.003***	0.006***	0.002**	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)
Net income	0.011*	-0.042***	-0.050***	-0.005
	(0.006)	(0.010)	(0.010)	(0.010)
Constant	0.597***	0.277***	0.511***	0.418***
	(0.036)	(0.064)	(0.065)	(0.064)
Observations	4,004	3,369	3,369	3,369
R-squared	0.050	0.045	0.037	0.041

*** p<0.01, ** p<0.05, * p<0.1 Standard errors in parentheses

Table 5 OLS regression analyses

5.4 Hypothesis 3

Hypothesis 3 uses the same independent and control variables as the previous hypotheses. The outcome variable is the dummy variable *Pro regulation.* The hypothesis states that smokers are less likely to vote in favor of the regulation of tobacco.

Hypothesis 3: Smokers are less likely to vote for parties that are specifically in favor of the regulation of tobacco in the Dutch national elections of 2017.

Column 4 in table 4 shows that smoking has a significant effect on the probability of voting on a pro-regulation party. The coefficient is -0.557 which means that being a smoker significantly decreases the probability on a 1% significance level ceteris paribus. The magnitude of the coefficient can be interpreted when all other variables have a reasonable value. I will interpret the result for 42-year-old male with an income of 2,066 euros, he is employed and is intermediately educated. This is average and common based on table 1. With the help of the regression formula, smokers are 11.63% less likely to vote for a pro-regulation party, this is shown in Appendix B.

Therefore, I find results that support hypothesis 3. Males significantly vote less for pro-regulation parties on a 1% significance level ceteris paribus. The results also show that students and highly educated people have a higher probability of voting for pro-regulation parties than unemployed and low educated people respectively ceteris paribus. These results are significant on a 1% significance level.

The OLS regression analysis in column 4 of table 5 shows that smokers are 12.2% less likely to vote for a pro-regulation party than non-smokers. In addition, males are 5.6% less likely to vote for a pro-regulation party than females. Finally, high educated people are 15.6% more likely to vote for a pro-regulation party than low educated people and students 18.2% more likely than unemployed people. All results are on a 1% significance level and ceteris paribus.

Figure 3 & figure 4 show a coefficient plot that shows which independent and control variables have a significant effect on the outcome variable. The dot is the coefficient and the corresponding horizontal lines are the 95% confidence intervals. When the lines do not touch the x=0 red line. This means that the variable has a significant effect. When those corresponding coefficients have a negative x-value, the variable has a significant negative impact on the probability of the outcome variable. When the coefficient has a positive x-value and the confidence interval line does not touch the x=0 line. The corresponding variable increases the probability of the outcome variable. For all hypotheses, the coefficient plots are shown in figure 3 & figure 4. Figure 3 represents the coefficient plot of the OLS regressions and figure 4 represents the coefficient plot of the logistic regressions.
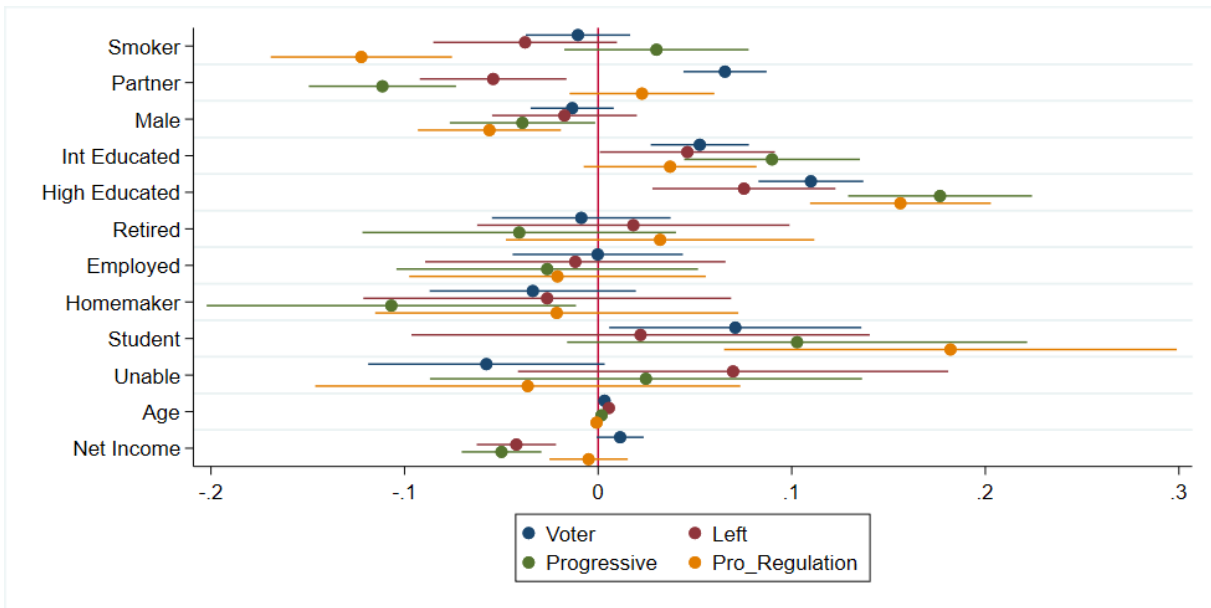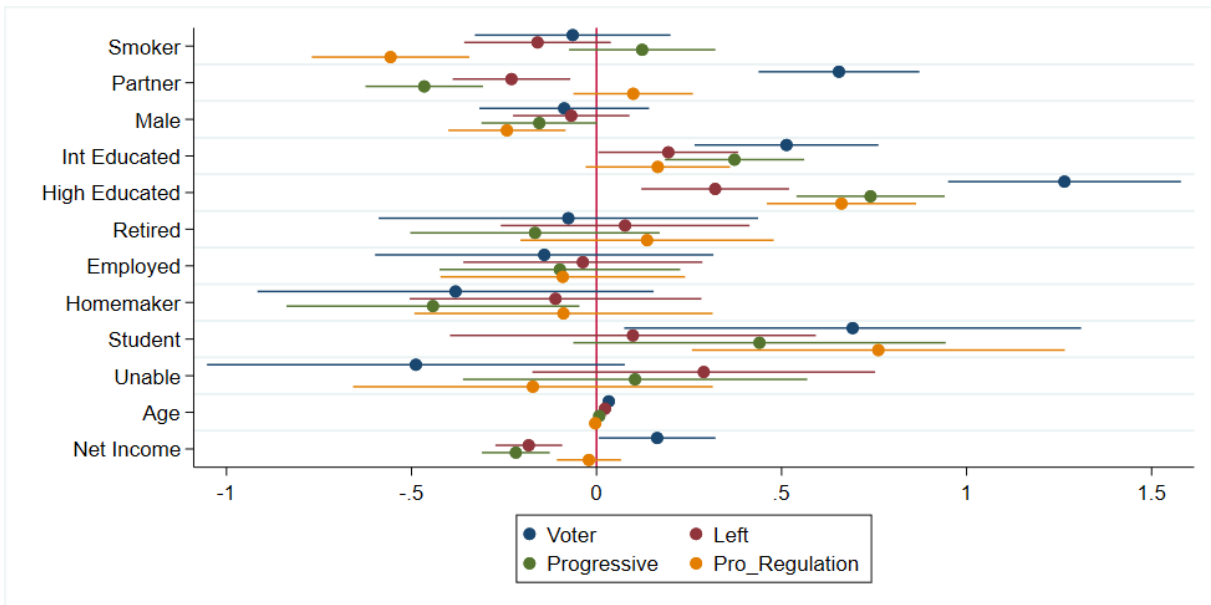
Figure 3 coefficient plot OLS regressions

Figure 4 coefficient plot logistic regressions

6 Conclusion

In this research voting and smoking are the main interests. Especially, the effect of smoking on voting behavior. To find an answer to this question, data from the 2017 national elections in the Netherlands have been used to research the relation between smoking and voting behavior. Voting behavior had been divided into several subjects: Voter turnout, political preference, and self-interested voting. The research question that has been formulated is:

To what extent do smokers differ from non-smokers in political participation and political preference?

This research question has been answered with the help of three hypotheses. These hypotheses have been formed with the help of the literature and previous researches on this topic. To test these hypotheses, an OLS regression and a logistic regression have been executed. The logistic regression has been executed because an OLS estimation is not compatible with binary outcome variables, and as a result, the coefficients cannot be interpreted correctly.

For hypotheses 1, 2a & 2b, I do not find the support of the hypotheses on a 5% significance level. There is no difference between smokers and non-smokers regarding voter turnout in the Dutch national elections of 2017. Neither is there a difference between smokers and non-smokers regarding the political preferences on the political spectrum. This result is unexpected, according to the previous literature, smokers have a significantly lower voter turnout than non-smokers. This effect could not be found in this research. This could be because the sample was not representative, as the voter turnout and smoker rates differ a lot from the general population. Furthermore, smokers were expected to significantly vote less for progressive and left-winged parties, this is based on the literature that states that smokers are socially more disengaged and disconnected. Which results in a more individualistic lifestyle without many communities to be a part of. This effect was not found in this research.

Hypothesis 3, on the other hand, can be accepted. The coefficient of *smoker* is significant on a 1% significance level. The coefficient is negative, which means that smokers are less likely to vote on parties that have an active tobacco regulation policy. This result is in line with the predicted outcome that smokers are self-interested voters regarding tobacco regulation. The potential benefit for society does not outweigh the benefit of the individual to not be restrained regarding tobacco use. This is in line with the literature about self-interested voting. This is also in line with Hersch et al. (2004), who found that smokers are less likely to vote in favor of smoking restrictions than non-smokers.

In conclusion, smokers differ from non-smokers only in their preference regarding tobacco regulation. Based on this research, differences in political preference on basis of the political agenda cannot be assumed. Neither can be assumed that smokers and non-smokers differ in political participation, meaning that the voter turnout does not differ in these two groups. Based on previous researches, however, smokers tend to have a lower probability of voting than non-smokers.

7 Discussion

No conclusion can be drawn for the difference in voter turnout between smokers and non-smokers, which is caused by the insignificant coefficients. This may have several explanations. First of all, there may genuinely be no differences between smokers and non-smokers. However, this is unlikely, since several previous studies have shown that smokers have a lower voter turnout than non-smokers. Another explanation is the representativeness of the sample. The sample showed a smoker percentage of 15.5%. In contrast, the population sample is 23.1%. Moreover, the voter turnout in the sample is 88.7%, while the voter turnout of the entire population is 81.9%.

The same can be concluded For hypothesis 2a regarding the representativeness. The ratio left/right in the sample is 0.495 whereas the ratio in the population is 0.375. This is a difference of 12%. Hypotheses 2a and 2b cannot be supported by the results. A reason can be that there is no difference in political preference regarding the political spectrum, in this case, the reasoning of the hypotheses is incorrect. Another reason is the representativeness, as just mentioned.

Support for the third hypothesis is found in this research. Smokers in this sample are less likely to vote for parties that actively advocate against tobacco regulation ceteris paribus. However, political parties' election campaigns consist of many more standpoints than tobacco regulation. Tobacco regulation is just a small part of the entire campaign. Consequently, there are unobserved similarities between the political parties in the Pro regulation category apart from tobacco regulation that could have influenced the voters' preferences for these parties. Thus, I cannot fully support hypothesis 3.

The control variables were limited because of the available data. The LISS panel did not have a variable for ethnicity, which has been shown to affect smoking rates. Therefore, there could be omitted variable bias. Finally, the fifth assumption of logistic regression did not hold for hypotheses 2a and 2b, which makes the outcome less reliable.

For future research, it will be interested to isolate an election on tobacco policies to find the difference between smokers and non-smoker. This solves the problem of hypothesis 3 in this research and eliminates external factors. Moreover, smoking is one of the risky health behaviors. It would be interesting to broaden the research using other risky health behaviors, such as drugs users and (heavy) drinkers. The Netherlands has a special and complicated soft drug policy, thus research on the difference between soft drug users and non-users could be interesting regarding drugs regulation policymaking.

Reference list

50Plus (2016). *Verkiezingsprogramma* 2017-2021. Retrieved from
https://50pluspartij.nl/images/PDFs/50PLUS_verkiezingsprogramma_2017-2021_DEFINITIEF.pdf

Acemoglu, D., & Robinson, J. A. (2006). *Economic origins of dictatorship and democracy*. Cambridge
University Press.

Albright, K., Hood, N., Ma, M., & Levinson, A. H. (2015). Smoking and (not) voting: the negative
relationship between a health-risk behavior and political participation in Colorado. *Nicotine &
Tobacco Research*, *18*(3), 371-376.

Bobbio, N. (1996). *Left and right: The significance of a political distinction*. University of Chicago
Press.

Boersema, W. (2021, February 17). Politieke partijen schuiven massaal op naar links. *Trouw.*
https://www.trouw.nl/politiek/politieke-partijen-schuiven-massaal-op-naar-links~b184d362/

Bruggink, J. W. (2013). Ontwikkelingen in het aandeel rokers in Nederland sinds 1989. *Tijdschrift voor
gezondheidswetenschappen*, *91*(4), 234-240.

CBS (2021). Aantal volwassen rokers in 2020 gedaald. Retrieved from https://www.cbs.nl/nl-
nl/nieuws/2021/10/aantal-volwassen-rokers-in-2020-gedaald

CDA (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from
https://d14uo0i7wmc99w.cloudfront.net/Afdelingen/Friesland/Weststellingwerf/documenten/CDA
%20landelijk%20verkiezingsprogramma%202017-2021.pdf

Centers for Disease Control and Prevention (n.d.). Current Cigarette Smoking Among Adults in the
United States. Retrieved from
https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm

Chaloupka, F. J., & Warner, K. E. (2000). The economics of smoking. *Handbook of health
economics*, *1*, 1539-1627.

ChristenUnie (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from
https://www.christenunie.nl/l/library/download/10071522/verkiezingsprogramma+2017-
2021+christenunie.pdf?redirected=1

D66 (2016). *Verkiezingsprogramma 2017-2021*. Retrieved from
https://dnpprepo.ub.rug.nl/10864/1/D66_vp_TK2017_def.pdf

DENK (2016). *Verkiezingsprogramma 2017-2021*. Retrieved from https://www.bewegingdenk.nl/wp-content/uploads/2020/11/Verkiezingsprogramma_DENK_2017-2021.pdf

Denny, K. J., & Doyle, O. M. (2007). "… Take up thy bed, and vote" Measuring the relationship between voting behaviour and indicators of health. *The European Journal of Public Health*, *17*(4), 400-401.

DeScioli, P., Cho, B., Bokemper, S., & Delton, A. W. (2020). Selfish and Cooperative Voting: Can the Majority Restrain Themselves?. *Political Behavior*, *42*(1), 261-283.

Eisinga, R. N., Franses, P., & Scheepers, P. L. H. (1997). Trends in de links-rechts oriëntatie van de Nederlandse kiezers 1978-1995.

Elinder, M., Jordahl, H., & Poutvaara, P. (2008). Selfish and prospective: Theory and evidence of pocketbook voting.

Escobedo, L. G., Anda, R. F., Smith, P. F., Remington, P. L., & Mast, E. E. (1990). Sociodemographic characteristics of cigarette smoking initiation in the United States: implications for smoking prevention policy. *Jama*, *264*(12), 1550-1555.

Franke, G. R. (2010). Multicollinearity. *Wiley international encyclopedia of marketing*.

FvD (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from https://dnpprepo.ub.rug.nl/10938/1/FvD_verkprogTK2017.pdf

Groenlinks (2016). *Verkiezingsprogramma 2017-2021*. Retrieved from https://deformatiewijzer.nl/wp-content/uploads/2017/02/Verkiezingsprogramma-GroenLinks-2017-20213.pdf

Hersch, J., Rossi, A. F. D., & Viscusi, W. K. (2004). Voter preferences and state regulation of smoking. *Economic Inquiry*, *42*(3), 455-468.

Jahn, D. (2011). Conceptualizing Left and Right in comparative politics: Towards a deductive approach. *Party Politics*, *17*(6), 745-765.

Kaprio, J., & Koskenvuo, M. (1988). A prospective study of psychological and socioeconomic characteristics, health behavior and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers. *Journal of clinical epidemiology*, *41*(2), 139-150.

Kiesraad (2017). Officiële uitslag Tweede Kamerverkiezing 15 maart 2017. Kiesraad. Retrieved June 10, 2021, from https://www.kiesraad.nl/actueel/nieuws/2017/03/20/officiele-uitslag-tweede-kamerverkiezing-15-maart-2017

Kiesraad (2017). Tweede Kamer 15 maart 2017. Kiesraad. Retrieved May 25, 2021, from https://www.verkiezingsuitslagen.nl/verkiezingen/detail/TK20170315

Klein, T., Rapp, I., & Schneider, B. (2013). The influence of couples' living arrangements on smoking habits and body weight. *Comparative Population Studies*, *38*(3).

Laerd Statistics (n.d.). Binomial Logistic Regression Analysis using Stata https://statistics.laerd.com/stata-tutorials/binomial-logistic-regression-using-stata.php

Lindström, M., Hanson, B. S., Östergren, P. O., & Berglund, G. (2000). Socioeconomic differences in smoking cessation: the role of social participation. *Scandinavian journal of public health*, *28*(3), 200-208.

Lindström, M., Isacsson, S. O., & Elmståhl, S. (2003). Impact of different aspects of social participation and social capital on smoking cessation among daily smokers: a longitudinal study. *Tobacco control*, *12*(3), 274-281.

Mattila, M., Söderlund, P., Wass, H., & Rapeli, L. (2013). Healthy voting: The effect of self-reported health on turnout in 30 countries. *Electoral Studies*, *32*(4), 886-891.

Ministerie van Volksgezondheid, Welzijn en Sport (2018). *Nationaal Preventieakkoord.* Retrieved from https://www.rijksoverheid.nl/onderwerpen/gezondheid-en-preventie/documenten/convenanten/2018/11/23/nationaal-preventieakkoord

NOS (2017, March 9). Politiek gaat altijd over links en rechts. Maar bestaat dat nog wel?. Retrieved from https://nos.nl/op3/artikel/2162205-politiek-gaat-altijd-over-links-en-rechts-maar-bestaat-dat-nog-wel

Patterson, T. E. (2002). The Vanishing Voter: Why Are the Voting Booths So Empty?. *National Civic Review*, *91*(4), 367-377.

Ponniah, S., & Bloomfield, A. (2008). Sociodemographic characteristics of New Zealand adult smokers, ex-smokers, and non-smokers: results from the 2006 Census. *The New Zealand Medical Journal (Online)*, *121*(1284).

PvdA (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from https://dnpprepo.ub.rug.nl/10867/19/PvdA-Verkiezingsprogramma-2017.pdf

PvdD (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from https://www.partijvoordedieren.nl/data/files/2016/10/PvdDVerkiezingsprogramaTweedeKamerverkiezingen2017-a2e7a068.pdf

PVV (2016). *Concept Verkiezingsprogramma 2017-2021.* Retrieved from
https://dnpprepo.ub.rug.nl/628/12/PVV2017-2021con.pdf

Rutten, L. J. F., Blake, K., Hesse, B. W., & Ackerson, L. K. (2011). Isolated and skeptical: social engagement and trust in information sources among smokers. *Journal of Cancer Education*, *26*(3), 465-473.

SGP (2016). *Verkiezingsprogramma 2017-2021*. Retrieved from
https://dnpprepo.ub.rug.nl/10869/1/Verkiezingsprogramma2017-2021SGP.pdf

SP (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from
https://www.sp.nl/sites/default/files/pak_de_macht.pdf

Sterling, T. D., & Weinkam, J. J. (1976). Smoking characteristics by type of employment. *Journal of occupational medicine.: official publication of the Industrial Medical Association*, *18*(11), 743-754.

Trimbos (2020). Cijfers roken. Trimbos-instituut. Retrieved from
https://www.trimbos.nl/kennis/cijfers/cijfers-roken

Tweede Kamer. (n.d.). Democracy in the Netherlands. Retrieved from
https://www.houseofrepresentatives.nl/how-parliament-works/democracy-netherlands

Tweede Kamer. (n.d.). House of representatives. Retrieved from
https://www.houseofrepresentatives.nl/house-representatives

Tweede Kamer. (n.d.). The Cabinet. Retrieved from https://www.houseofrepresentatives.nl/cabinet

VVD (2016). *Verkiezingsprogramma 2017-2021.* Retrieved from
https://vvd.nl/content/uploads/2016/11/vvd_verkiezingsprogramma_pages.pdf

Wattenberg, M. P. (2002). *Where have all the voters gone?*. Harvard University Press.

West, R. (1990). Progressive and conservative constitutionalism. *Michigan Law Review*, *88*(4), 641-721.

Willemsen, M. (2011). *Roken in Nederland: de keerzijde van tolerantie*. Maastricht: Maastricht University.

Appendix

Party	Statement on tobacco regulation in election campaign in 2017
VVD	"To inform the youth about the negative effects of tobacco, they are able to make the right decision about this matter." (VVD, 2016, p. 62)
PVV	Does not mention tobacco regulation in their campaign, in 2012 they advocated against the excise duties on tobacco. (PVV, 2016)
CDA	"The government should discourage tobacco use effectively, by enhancing the excise duties on tobacco." (CDA, 2016, p. 74)
D66	"The next generation should be a tobacco free generation. This should happen through informing the youth, regulating the supply and regulating the marketing." (D66, 2016, p. 101)
GL	"All terraces should be non-smoking areas and addicting additions to cigarettes should be banned." (Groenlinks, 2016, p. 38)
SP	"Raise awareness on how a tobacco addiction starts on lower and middle school." (SP, 2016, p. 33)
PvdA	"Raise awareness on the effects of smoking. We'll work together with insurance companies on preventive health care regarding quitting smoking." (PvdA, 2016, p. 51)
CU	"The places where tobacco can be bought will be reduced to specialty shops only." (ChristenUnie, 2016, p. 50)
PvdD	"Help with quitting smoking stays in the basic insurance package." (PvdD, 2016, p. 25)
50PLUS	"The government should actively and clearly campaign to discourage use of tobacco." (50Plus, 2016, p. 9)
SGP	"Campaigns against every form of addiction, starting from the age of 12. Strongly against alcohol advertisements and strongly against soft drugs." (SGP, 2016, p. 22)
DENK	Does not mention tobacco, alcohol or soft drugs in their campaign. (DENK, 2016)
FvD	Does not mention tobacco and alcohol in their campaign. Advocates for legalization of soft drugs and prevention of drug addiction. (FvD, 2016)

Table A Political parties' statements regarding tobacco regulation in their election campaign of 2017

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) smoker	1.000													
(2) male	0.040***	1.000												
(3) partner	-0.084***	0.057***	1.000											
(4) net_income	-0.045***	0.313***	0.008	1.000										
(5) age	-0.053***	0.012	-0.037***	0.182***	1.000									
(6) low_educated	0.062***	-0.030***	-0.018*	-0.244***	0.232***	1.000								
(7) int_educated	0.059***	0.011	-0.003	-0.145***	-0.170***	-0.464***	1.000							
(8) high_educated	-0.115***	0.017*	0.019*	0.372***	-0.044***	-0.461***	-0.572***	1.000						
(9) unemployed	0.070***	-0.015	-0.073***	-0.145***	0.020*	0.061***	-0.019*	-0.037***	1.000					
(10) retired	-0.082***	0.062***	-0.085***	0.054***	0.648***	0.162***	-0.110***	-0.039***	-0.128***	1.000				
(11) employed	0.019	0.087***	0.128***	0.408***	-0.328***	-0.237***	0.011	0.208***	-0.275***	-0.539***	1.000			
(12) homemaker	-0.017	-0.238***	0.094***	-0.263***	0.113***	0.132***	-0.017*	-0.105***	-0.070***	-0.138***	-0.297***	1.000		
(13) student	-0.005	0.005	-0.091***	-0.371***	-0.465***	-0.028**	0.162***	-0.137***	-0.081***	-0.159***	-0.344***	-0.088***	1.000	
(14) unable	0.080***	-0.025**	-0.051***	-0.082***	0.028***	0.067***	0.003	-0.065***	-0.051***	-0.099***	-0.214***	-0.055***	-0.063***	1.000

*** p<0.01, ** p<0.05, * p<0.1

Table B matrix of correlations

	(1)	(2)	(3)	(4)
VARIABLES	Voter	Left	Progressive	Pro_regulation
Smoker	-0.163	-0.163	0.077	-0.558***
	(0.141)	(0.106)	(0.106)	(0.114)
Partner	0.640***	-0.280***	-0.468***	0.117
	(0.118)	(0.085)	(0.085)	(0.087)
Male	-0.029	0.013	-0.167**	-0.271***
	(0.125)	(0.085)	(0.084)	(0.086)
Low_Educated	-1.215***	-0.402***	-0.721***	-0.698***
	(0.173)	(0.109)	(0.109)	(0.111)
Int_Educated	-0.721***	-0.210**	-0.423***	-0.486***
	(0.153)	(0.092)	(0.093)	(0.093)
Omitted High_Educated	-	-	-	-
Unemployed	0.569*	-0.189	0.008	0.190
	(0.316)	(0.250)	(0.252)	(0.264)
Retired	0.471*	0.064	-0.123	0.227
	(0.284)	(0.220)	(0.221)	(0.233)
Employed	0.406*	-0.174	-0.206	0.123
	(0.234)	(0.203)	(0.204)	(0.217)
Homemaker	0.026	0.007	-0.213	-0.001
	(0.320)	(0.267)	(0.266)	(0.279)
Student	1.377***	0.154	0.977***	0.637*
	(0.455)	(0.352)	(0.366)	(0.358)
Omitted Unable	-	-	-	-
Age	0.068	0.262***	0.297***	-0.154*
	(0.114)	(0.081)	(0.080)	(0.079)
Net_Income	-0.002	-0.002**	0.002*	-0.000
	(0.002)	(0.001)	(0.001)	(0.001)
ln_income	0.000	0.000**	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
ln_Age	-0.007	-0.049***	-0.059***	0.031*
	(0.024)	(0.017)	(0.016)	(0.016)
Constant	0.058	-2.213***	-2.237***	1.541*
	(1.158)	(0.850)	(0.844)	(0.828)
Observations	3,655	3,112	3,112	3,112

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table C Box-Tidwell approach for all hypotheses

Appendix B

$$Pr(ProRegulation_i) = \frac{1}{1 + e^{-(-0.347\ -0.041-0.12-0.091+0.166-0.242+0.099)}} = 35.98\%$$

$$Pr(ProRegulation_i) = \frac{1}{1 + e^{-(-0.347\ -0.041-0.12-0.091+0.166-0.242+0.099-0.557)}} = 24.35\%$$