CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Vojtěch Válek

The impact of COVID-19 on election outcomes: The case of the Czech Republic

Bachelor thesis

Prague 2022

Author: Vojtěch Válek

Supervisor: Mgr. Barbara Pertold-Gebicka, M.A., Ph.D.

Academic Year: 2021/2022

Bibliographic note

VALEK, Vojtech. *The impact of COVID-19 on election outcomes: The case of the Czech Republic* Prague 2020. 58 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor Mgr. Barbara Pertold-Gebicka, M.A., Ph.D.

Abstract

This thesis examines the relationship between the COVID-19 pandemic and the election results in the 2021 parliamentary elections in the Czech Republic. Voter turnout together with the results of several political parties is analyzed at the district level. First, the election results of 2021 are examined separately using OLS regression, heteroskedasticity is not detected in the model. A panel dataset is then created, which also includes election results from 2013 and 2017. Based on the results of the Hausman test, the fixed effects method is used to analyze the panel data, and the first-differencing method is used as well for comparison. Pairs of years are then analyzed with the use of first-differencing, when the results are first estimated without 2013, and then without 2017. The results of the analysis mainly show a significant relationship between the number of infected people and the results of individual political parties. While for political party ANO this relationship is negative, it is rather positive for the other political parties. The second dependent variable of interest is the number of deaths related to COVID-19. In this case, the results of the analysis do not indicate a significant connection with the election results.

Keywords

COVID-19, election results, voter turnout, ordinary least squares, fixed effects, random effects, first-differencing, Czech Republic

Abstrakt

Tato práce zkoumá vztah mezi pandemií Covidu-19 a volebními výsledky parlamentních voleb 2021 v České republice. Volební účast společně s výsledky několika politických stran je analyzována na úrovni okresů. Nejprve jsou zvláště otestovány volební výsledky roku 2021 pomocí OLS regrese, heteroskedasticita není v modelu detekována. Poté je vytvořen panelový dataset, který zahrnuje volební výsledky z let 2013 a 2017. Na základě výsledků Hausmanova testu je použita metoda fixních efektů k analýze panelových dat, metoda první diferenciace je použita pro srovnání také. Dvojice let jsou poté analyzovány za použitá první diferenciace, kdy výsledky jsou nejprve odhadnuty bez roku 2013 a poté bez roku 2017. Výsledky analýzy především ukazují signifikantní vztah mezi počtem infikovaných lidí a výsledky jednotlivých politický stran. Zatímco pro stranu ANO je tento vztah negativní, je spíše pozitivní pro ostatní politické strany. Druhá zkoumaná závislá proměnná je počet úmrtí souvisejících s Covidem-19. V tomto případě výsledky analýzy neindikují signifikantní vztah s výsledky voleb.

Klíčová slova

Covid-19, volební výsledky, volební účast, metoda nejmenších čtverců, fixní efekty, náhodné efekty, první diferenciace, Česká republika

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes

Prague, August 2, 2022

Vojtech Valek

Acknowledgment

I would like to express a deep gratitude to my supervisor Mgr. Barbara Pertold-Gebicka, M.A., Ph.D., who provided me with valuable insights and helped me shape this thesis.

Bachelor's thesis proposal

Author	Vojtěch Válek
Supervisor	Mgr. Barbara Pertold-Gebicka, M.A., Ph.D.
Proposed topic	The impact of COVID-19 on election outcomes: The case of the Czech Republic

Motivation

The main research question I intend to study is how the coronavirus pandemic has affected the outcome of parliamentary election in 2021 in the Czech Republic.

It is possible that some parts of the Czech Republic have been affected by COVID-19 more than others. Therefore, the analysis will be conducted at the level of the districts of the Czech Republic, and at this level I will examine the results of some selected political parties in parliamentary election in 2021 and how these results were affected by the coronavirus pandemic.

COVID-19, which first appeared in 2019 in China, grew in the following months into a global pandemic that significantly changed life as we knew it. The pandemic affected basically everything and everyone, regardless of where a person comes from, what job they have, what social status they have, and so on. The fact that COVID-19 has become an important political topic in all countries of the world follows from this, as it was the governments and politicians of individual countries who had to implement all kinds of measures to fight the disease.

The question therefore arises as to how the coronavirus has affected voters' electoral preferences. Of course, the answer to such a question can vary greatly from country to country, as each country has dealt (and is still dealing) with the COVID-19 pandemic in its own way. Some countries managed it better than others, implemented different measures, or implemented them with different intensity, and so on. The COVID-19 pandemic has become a political issue, McKee et al. (2021) states that populist parties are generally the ones who benefit the most from the coronavirus pandemic. On the other hand, according to Baccini et al. (2021), Donald Trump (who could certainly be described as a populist politician) was harmed by the coronavirus pandemic and would have won the presidential election in 2020 if it weren't for it. Thus, I want to find out what effect the pandemic had on the election results in the Czech Republic and which political parties it helped and which it harmed.

To find answers to my questions, I will analyze district-level election data from the 2021 parliamentary elections and compare them with the results of elections from earlier years, specifically from 2013 and 2017.

Methodology

In my thesis, I will analyze the election results of several political parties in the Czech Republic, such as ANO 2011 or the SPOLU coalition. I will look for answers to questions such as: What is the relationship between the number of victims of COVID-19 and the election results in individual districts of our country? or What is the relationship between the number of infected and the election results? I will use regression analysis to estimate these relationships.

I will work with data from the parliamentary elections of 2021, 2017 and 2013, specifically I will be interested in the results of individual political parties and voter turnout. These data will be compared with COVID-19 statistics, specifically with the number of deaths and the number of infected people. Other independent variables will be mainly socio-economic indicators, such as unemployment or the age composition of the population. All the mentioned data will be obtained from the Czech Statistical Office, except for the coronavirus statistics, which will be obtained from the database of the Ministry of Health of the Czech Republic. Since the data for the key explanatory variable, the number of COVID-19 deaths, are only available at the level of districts (not municipalities), the analysis will be carried out at that level (a total of 76 districts + the capital Prague).

I will develop an econometric model to answer the desired questions.

First, I will separately analyze the election results from 2021 using the ordinary least squares method. However, such results may not be very accurate, as there is a high probability of the presence of endogeneity. Therefore, the analysis of panel data will follow, when I will analyze together the years 2021, 2017 and 2013. Here the fixed effects, random effects and first differencing models will be used.

Expected Contribution

Since COVID-19 has only been around for a few years, its effects in many areas are still unclear. A lot of studies related to COVID-19 have already been done even in its relatively short existence, which is completely logical, if we consider how unprecedented event this is. Over time, there are increasingly more studies dealing with the effects of the coronavirus pandemic on politics, I believe that such studies are of great importance. My work should therefore contribute to the still under-researched relationship between COVID-19 and election results. As far as I know, such a detailed study has not yet been carried out in the Czech Republic. The event, which was probably most studied by experts in this sense was the already mentioned presidential election in America in 2020. Johnson (2020) stated at the very beginning of the pandemic in 2020 that COVID-19 could significantly change the politics of the United States. Parzuchowski et al. (2021), dealt with the specific results of the elections. Among other things, according to this study support for the Democrats increased compared to the previous elections, especially in counties with a lower mortality rate.

In the Czech Republic, several worth mentioning studies have been carried out dealing with the relationship between socio-demographic indicators and electoral preferences of voters. One such a study was carried out by, for example, Černý (2019) and since I will use many socio-demographic variables in my work as well, I will also refer to such studies. Moreover, as I will also use the election results from 2017 in my analysis, this paper should definitely be included.

Outline

1. Introduction
2. Literature review
3. Data and methodology
4. Results
5. Conclusion

Core bibliography

McKee, M., Gugushvili, A., Koltai, J., Stuckler, D. (2021). Are populist leaders creating the conditions for the spread of COVID-19? Comment on" A scoping review of populist radical right parties' influence on welfare policy and its implications for population health in Europe". International journal of health policy and management, 10(8), 511.

Baccini, L., Brodeur, A., Weymouth, S. (2021). The COVID-19 pandemic and the 2020 US presidential election. Journal of Population Economics, 34(2), 739-767.

Johnson, A. F., Pollock, W., Rauhaus, B. (2020). Mass casualty event scenarios and political shifts: 2020 election outcomes and the US COVID-19 pandemic. Administrative Theory and Praxis, 42(2), 249-264.

Parzuchowski, A. S., Peters, A. T., Johnson-Sasso, C. P., Rydland, K. J., Feinglass, J. M. (2021). County-level association of COVID-19 mortality with 2020 United States presidential voting. Public Health, 198, 114-117.

Černý, J. (2019). Population Characteristics of Voters: Evidence from the Czech Parliamentary Election.

Contents

Acronyms

ANO political party

CSSD Czech Social Democratic Party, political party

KSCM Communist Party of Bohemia and Moravia, political party

MLR Chief Executive Officer

OLS Customer Lifetime Value

SPD Freedom and Direct Democracy, political party

SPOLU political coalition

Chapter 1

Introduction

The disease COVID-19, which first appeared in China in 2019, spread almost all over the world within a few months, and on March 11, 2020, the WHO declared a pandemic. The WHO reported that 1,813,188 people died of COVID-19 worldwide in 2020 alone, with the true number likely to be much higher. The previous way of life has changed radically for most people, as various preventive measures have been introduced. People could not travel, could not meet each other, and in some cases could not even leave the house.

COVID-19 has affected almost everything and everyone. As Donthu & Gustafsson (2020) says, the whole society has been affected and the economic consequences are and will be huge, adding that something like this could happen again in the future, so we should be prepared.

The economy stagnated, people lost their jobs, people radically reduced their financial spending (Chetty et al. (2020)). The governments of individual countries had to react to the situation. Some succeeded, some failed, as McKee et al. (2021) states, the populist leaders underestimated the whole situation and did not respond adequately, and as Malhotra & Kuo (2008) says, who in his study dealt with the responsibility of politicians in dealing with a natural disaster, if voters get enough relevant information, they are able to form an opinion that corresponds to reality and then make an informed decision at the polls.

During most of the COVID-19 crisis, at least in the first two years or so, the government of the Czech Republic was led by the political party ANO. As Ferejohn (1986) mentions, voters largely decide who to vote for based on the performance of the ruling party while in office.

The literature dealing with the impact of COVID-19 on election results is only emerging, although many studies have already been published and some of them are described in more detail in the literature review chapter. As for the Czech Republic, this work should be one of the first studies dealing with this phenomenon.

This work tries to determine if there is a connection between COVID-19 and the results of the parliamentary elections in the Czech Republic in 2021. Specifically, it tries to estimate what effect the number of deaths from COVID-19 and the number of infected people had. The results of the analysis are achieved with the use of econometric methods. Specifically, these methods are ordinary least squares regression, fixed effects estimation, random effects estimation and first-differencing estimation.

The thesis has the following structure. First, the literature used in this work is presented, which clarifies the choice of variables. The following chapter presents the data that is the basis for the final analysis. Then the individual models that are used to estimate the results are described. Finally, the results of the analysis are presented and described together with the conclusion of this thesis.

Chapter 2

Literature Review

There are many factors influencing voter preferences, and people in different parts of the world based their decision who to vote for on different reasons. Given that COVID-19 appeared only a few years ago, for obvious reasons there is not yet much professional literature that deals with this topic, that is the impact of the coronavirus pandemic on election results. Moreover, in every democratic country, elections are held once in a certain period of time (mostly once every 4 to 5 years), thus there are many countries in which significant elections (whether parliamentary, presidential, or any other) have not been held since the beginning of the COVID-19 pandemic. Over time, there will undoubtedly be more studies dealing with this issue. However, there are already numerous published scholarly papers and studies that examine the impact of the coronavirus on election results and voter behavior. In this chapter, these papers will be described together with other literature examining more general factors affecting voter preferences.

2.1 General Factors of Voter Behaviour

Antunes (2010) summarizes 3 basic theoretical models that are based on older scholarly studies and based on which people decide who to vote for during elections. The first is a sociological model that has its origins in the work of Lazarsfeld (1944), who described how people in America made decisions during the presidential elections in 1940. According to this model, an individual makes decisions based on social factors. His study led to the conclusion that the social groups in which the given individual lives are a crucial factor determining voters' electoral preferences and that, on the contrary, the effect of the media in the pre-election competition of individual candidates for citizens' votes is negligible.

Another, psychological model, states that voters make decisions primarily on the basis of their own party identification. This does not necessarily mean a person's formal membership in a political party, or other direct involvement in the functioning of a given party, but it is a situation where a person simply develops some kind of relationship with a given political party, good or bad, as described by Campbell et al. (1980).

The third and last is the rational model described by Downs (1957), which states that voters make decisions based on self-interest in order to maximize their own utility.

These theories are often linked, that is voters can make decisions based on several factors and, in principle, any factor affecting voters could be included in one of the listed categories. By moving from the models to these specific factors, it is possible to get a more concrete picture of how such an ordinary voter makes a decision.

In the past, it was common, perhaps more so than today, to divide society into classes, and it was common to make use of the assumption that people in the same class have the same voting preferences and that different classes have different voting preferences. As Manza & Brooks (2008) points out, the working class usually voted for socialist parties, while the upper classes voted for conservative parties. Although, as Boschken et al. (2001) mentions, the importance of class has changed over the years and is no longer as important as it once was, social stratification is still an important factor. Therefore, the essence of economic models of voting behaviour, although redefined over the years, still applies and group of people with the same characteristics tend to vote similarly, as Manza & Brooks shows on an example of the income-based model according to which low-income voters in the United States have supported the Democrats persistently for a long time. He also mentions that even the division into occupational categories is still relevant and gives evidence from America indicating that while people in occupational groups still vote similarly, the voting preferences of entire occupational groups have changed over the years, with for example self-employed people shifting from centrists over the years to strong supporters of the Republicans.

Many studies have confirmed that the education of voters has a significant influence on election results, not only in the sense of who they vote for, but also whether they come to the polls at all.

Regarding voter turnout, Hoskins et al. (2008) finds a positive relationship between education and voter turnout in his study, when he presents the result in a sample of 19 European countries that the probability of a person going to the polls increases by 0.9 percentage points with each year of education. A positive relationship between these variables is also found by Dee (2004), who analyzes the American environment, or by Kolstad & Wiig (2016), who, on the contrary, focused on developing countries, specifically Tanzania.

The level of achieved education also seems to have an influence when deciding which political party the voter will support. Werts et al. (2013) analyzed a dataset of 18 European countries and concluded, among other things, that less educated people are more likely to vote for radical parties on the right of the political spectrum. Furthermore, his findings also show that unemployed people or manual workers often vote for these political parties. One of the 18 countries analyzed is Slovakia, which was the focus of Kotrč (2017) and Dusková (2021) in their research papers, who analyzed the election results of extreme right-wing and populist parties, respectively. Their results, which were carried out at the level of municipalities and districts, basically coincide with those of Werts et al.. According to Kotrč, the analyzed right-wing party was mainly supported by people from municipalities with higher unemployment, or young people and people with less education. Dusková and her results suggests that, in addition to the mentioned unemployment, low wages or unskilled workers also contribute to the support of populist parties. Moreover, she also discovered that the largest ethnic minority in Slovakia largely votes for other than populist parties.

According to many scholars, another important factor influencing voters' preferences is religion. Smidt et al. (2010) says that religion has historically played a very important role in American politics and mentions Campbell (2007) and his book where the author describes the significant role of religion in the 2004 presidential election and which religious groups favored which candidates.

With regard to the Czech Republic, Černỳ (2019) focused on the last Czech parliamentary elections, which took place in 2017. The analysis was carried out at the level of municipalities and examined all nine political parties that entered the Chamber of Deputies. According to his conclusions, the political parties ANO, SPD and KSCM enjoyed greater support in parts of the Czech Republic with higher unemployment rate or a lower proportion of university-educated people. This is in line with the findings from the already mentioned studies in this chapter, as these political parties, that is ANO, SPD and KSCM, are either on the right or left end of the political spectrum, or could be characterized as populist parties. Conversely, characteristically different political parties, such as ODS or TOP 09, did well in municipalities with opposite attributes.

2.2 Elections In The Era of COVID-19 and Empirical Approach

Most of the papers published so far dealing with the impact of the COVID-19 on election results has focused on the US presidential elections from the end of 2020. This is logical, since it is probably the most politically watched regularly recurring event, and at the same time, quite a lot of time has passed since then , so that the effects of the coronavirus on elections could be properly analyzed.

Already before the election itself, Johnson et al. (2020) tried to predict what COVID-19 would mean for the results of the presidential candidates. He obtained the COVID-19 statistics from predicted outcome models, that is he worked with estimates of deaths related to COVID-19. He chose age as the key variable, since his main data were the share of individual age groups in the population of American states and the shares of Republican and Democratic voters from the 2016 election within each age group. Using rather simple mathematical calculations, he came to the conclusion that due to the reduction in the number of elderly voters in the so-called swing states (that is, states that tend to be decisive for election results), significant political changes may occur, as it would be mainly the Republicans, the party of the then president Donald Trump, who would lose their voters.

Parzuchowski et al. (2021) has already researched the post-election results, he too used the number of deaths related to COVID-19 as the only coronavirus statistic, claiming that the infection rate is an unreliable indicator due to the uneven availability of tests. Thus, he tried to estimate the impact of coronavirus-related deaths on the election results while controlling for a large number of economic, demographic, but also health variables. The states were divided into quartiles and then a population-weighted linear regression was used for the change in votes received by a given political party between 2016 and 2020. Among other things, he found that compared to the lowest mortality quartile, the one with the highest mortality was associated with the smaller increase in Democratic vote share when comparing 2020 election to the ones four years earlier.

A considerably more thorough analysis of the post-election results was carried out in study by Baccini et al. (2021), who used the number of infected people in addition to the number of deaths to estimate the effect of the coronavirus on the election results. He performed the analysis at the district level using two methods. Firstly, Baccini et al. used the OLS method, where his dependent variable was the difference in Trump's results in 2020 and 2016, and the independent COVID-19 variables were supplemented with many other explanatory variables describing the characteristics of the population. Secondly, he used the 2SLS method where he instrumented the COVID-19 statistics with the share of meat-processing workers. His results clearly show that Democratic candidate Joe Biden benefited significantly from the pandemic situation and that if it weren't for COVID-19, Trump would most likely have won the election.

The impact of COVID-19 on voter turnout was addressed in study by Fernandez-Navia et al. (2021), who analyzed regional elections in Spain from the middle of 2020. According to his results, municipalities with more infected people had up to 5.1 percentage points lower voter turnout. He also found out that people from these municipalities tend to vote for nationalist parties.

Chapter 3

Data

This chapter will present the data and variables that were used in the analysis along with some basic descriptive statistics. The chapter is divided into a description of dependent and independent variables. The selection of independent variables is mainly based on the overview of used literature and thus on already conducted studies that dealt with this or a similar topic.

3.1 Dependent Variables

In total, we have six independent variables, the first one is voter turnout and the other five are the shares of votes received by the five political entities.

Turnout: Voter turnout in the parliamentary elections expressed as a percentage.

Party Share: The share of votes received by a given political party or coalition in parliamentary elections, expressed as a percentage.

The political entities here are four political parties and one coalition. Political parties ANO, SPD, CSSD and KSCM and coalition SPOLU. Although only SPOLU, ANO, SPD and the PIRATI a STAROSTOVE coalition entered the Chamber of Deputies in 2021, I decided to also include the political parties CSSD and KSCM in the analyzes, as they are traditional Czech political parties that have always entered the Chamber of Deputies in previous years. There could be a problem with coalitions when comparing data between 2021, 2017 and 2013, since these coalitions did not exist in 2013 and 2017, and the political parties forming the SPOLU coalition and the PIRATI a STAROSTOVE coalition in these years competed for votes separately. I decided to remove the PIRATI a STAROSTOVE coalition from the analysis, because in 2013 STAROSTOVE ran on the TOP 09 candidate list, which is other political party, and thus there would be a problem in obtaining the necessary data. I kept the SPOLU coalition in the analysis and for the years 2013 and 2017 I only added the votes of the individual political parties that make up the SPOLU coalition together, that is the parties ODS, KDU-CSL and TOP 09. Thus, although the SPOLU coalition did not exist in 2013 and 2017, in my dataset is listed even for these years.

All election data, that is both voter turnout and individual party results for each election year, were obtained from the Czech Statistical Office. Basic descriptive statistics of all dependent and independent variables are shown in Table 3.1. The statistics clearly show which parties were successful in the elections and which were not - on one hand SPOLU or ANO with big gains, on the other the CSSD and KSCM, which did not cross the five percent threshold and thus did not get into the Chamber of Deputies. An interesting and at first confusing figure is the comparison of means of the SPOLU coalition and ANO. Although the SPOLU coalition won the elections, with a total official gain of 27.79 percent, and thus defeated ANO with a gain of 27.12 percent, these descriptive statistics show different values and ANO has even a higher mean than SPOLU. A possible cause of the difference between these values and the official statistics may be the votes of voters from abroad, which were not included in this analysis.

3.2 Independent Variables

A total of 11 independent variables were initially selected, two related to COVID-19, and nine other variables, representing economic and socio-demographic characteristics of the population, all these variables can be seen in Table 3.1. The selected independent variables are generally based on a review of pro-

Table 3.1: Summary statistics of all variables

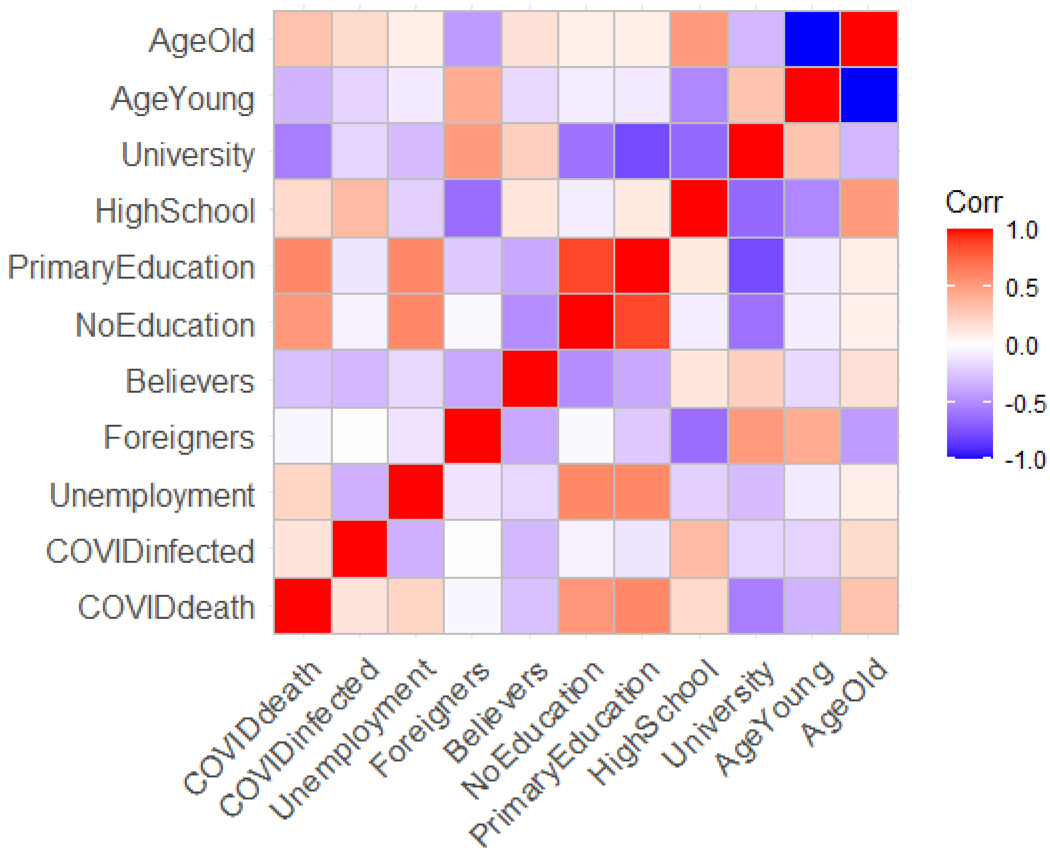
Type	Variable	Min	Median	Mean	Max	Std.Dev.
Dependent	Turnout	52.71	65.50	64.86	74.75	4.52
	SPOLU	15.37	26.40	25.62	40.02	4.98
	ANO	17.46	28.18	28.86	39.11	4.56
	SPD	4.59	10.18	10.34	15.50	2.26
	CSSD	2.82	4.78	4.81	8.59	1.07
	KSCM	1.99	4.10	3.98	6.37	0.85
Independent	COVIDdeath	13.44	30.89	31.29	65.60	8.01
	COVIDinfected	125.9	162.3	163.5	201.6	18.32
	Unemployment	1.28	3.20	3.48	8.47	1.27
..	Foreigners	1.21	3.60	4.56	18.52	2.99
	Believers	9.53	19.98	20.92	45.89	8.42
	NoEducation	0.31	0.53	0.58	1.26	0.21
	PrimaryEducation	6.72	10.96	11.24	15.61	1.60
	HighSchool	43.07	54.40	53.69	57.44	2.52
	University	6.99	12.16	13.01	31.07	4.24
	AgeYoung	76.66	78.94	79.14	84.72	1.41
	AgeOld	15.28	21.06	20.86	23.34	1.41

Source: Made by author based on available data

fessional literature, however some other variables that, according to some studies, would be suitable for this research could not be used due to the unavailability of data. For example, as Černỳ (2019) discovered, the number of distraints has a noticeable negative effect on voter turnout. Although his study also focused on the Czech Republic, data for 2021 is not available.

Statistics capturing the distribution of people according to individual occupations, which according to Manza & Brooks (2008) is also a relevant indicator, are also not captured due to the unavailability of such data at the district level. At the district level, the average salary is also not available, which could also be an interesting variable. For example, Baccini et al. (2021) also works with factors such as government bans and restrictions, however, given that basically the same measures were in place throughout

Figure 3.1: Correlation between independent variables



the Czech Republic at the same time, the use of such data in our analysis does not make much sense.

Before running any regressions, it is advisable to examine the interrelationships between the independent variables, specifically the correlation between them. This can prevent the problem of multicollinearity, which is a situation when there are two or more independent variables in the regression model that are highly correlated with each other. Using such variables could lead to inaccurate results. Therefore, the intercorrelations of all variables are calculated, the Figure 3.1 shows these intercorrelations.

There are several strong correlations between the variables. As can be seen from the figure, it is primarily *AgeYoung* and *AgeOld*, where the correlation coefficient is actully equal to -1, which means a perfect negative linear relationship (Wooldridge, 2013, p. 739). This is a completely logical relationship, because *AgeYoung* represents people under 65 years of age, and on the contrary, the variable *AgeOld* contains all people who have reached at least 65th years of age, so if a person belongs to one group, he does not automatically belong to the other. This fact should have been revealed before the actual selection of these variables. Another strong correlation, this time with a value of 0.87, is found between *NoEducation* and *PrimaryEducation*.

Although there is no generally recognized threshold above which the correlation would be considered too large, and from which the results of the analysis would certainly be distorted, the problematic absolute value of the correlation coefficient is usually set around 0.7 or 0.8. I decided to choose a threshold of 0.8. Thus, all values of the correlation coefficient, which are mathematically represented as

$$\mathrm{Corr}(X, Y) \equiv \frac{\mathrm{Cov}(X, Y)}{\mathrm{sd}(X) \cdot \mathrm{sd}(Y)}$$

that are greater than 0.8 in absolute value are taken as too large and the corresponding pair of independent variables have a problematic relationship. Therefore, one variable from such a pair must be excluded from the model. Apart from the already mentioned two relationships with too high value of correlation coefficient, no other problematic pairs were found. The variables *AgeYoung* and *NoEducation* were excluded from the analysis. Since the proportion of old people could be an important factor as it is the group most at risk from the COVID-19, it was kept in the analysis at the expense of *AgeYoung*. In the case of *NoEducation*, this variable was selected from the pair, as the share of people with no education is generally very small.

All data on independent variables, except for the COVID-19 data, was obtained from the Czech Statistical Office. Except for unemployment, all values were obtained in absolute values and subsequently recalculated to the share of the population from the given district. Data for the variable *Believers* and all variables related to education were taken from census data. Thus, for the year 2021, new data are available from the same year's census, but for 2013 and 2017, the same data from the 2011 census were used (how this problem was solved in the analysis is explained in Chapter 5. All other economic and socio-demographic variables are measured on an annual basis.

As for the coronavirus data, it was obtained from the database of the Ministry of Health of the Czech Republic. The variable *COVIDdeaths* represents the cumulative number of deaths from COVID-19 in a given district from the beginning of the pandemic to October 7, 2021, that is the day before the elections, calculated per 10,000 inhabitants. Similarly, *COVIDinfected* represents the cumulative number of infected people in a given district from the beginning of the pandemic to October 7, 2021, calculated per 1,000 inhabitants.

The final list of independent variables is following:

COVIDdeaths: The cumulative number of deaths from COVID-19 per 10,000 inhabitants in a given district,

COVIDinfected: The cumulative number of infected with COVID-19 per 1,000 inhabitants in a given district,

Unemployment: Unemployment rate in a given district,

Foreigners: Share of foreigners in the population in a given district,

Believers: Share of believers in the population in a given district,

PrimaryEducation: Share of people with primary education in a given district,

HighSchool: Share of people with high school education in a given district,

University: Share of people with university education in a given district,

AgeOld: Share of people who are at least 65 years old in a given district.

Chapter 4

Methodology

This chapter describes the methods used to analyze the data described in the previous chapter. Each method is described in general along with relevant tests verifying the fulfillment of its assumptions. Firstly, a simple ordinary least squares (OLS) regression is used to analyze the cross-sectional data for 2021 and estimate their implications for election results in that year. This is followed by more complex methods of data analysis, which also include data from 2013 and 2017. Specifically, to analyze such panel data, the methods of fixed effects, random effects and first-differencing are used.

4.1 Cross-sectional Data Estimation Method

4.1.1 OLS regression

OLS estimation is used to find the basic relationships between individual variables. Because data from only one time period is analyzed here, it is appropriate to use OLS regression. By using this method, we estimate the coefficients of individual independent variables, thanks to which it will be possible to determine not only how strongly the independent variables affect the dependent variable, but also, for example, if there is a positive or negative relationship between them.

The general model, on which our final model is based, is represented by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + u, \quad (4.1)$$

where β_0 symbolizes intercept, $\beta_1, ..., \beta_k$ are the parameters of the respective independent variables and u is the error term. This is the general multiple linear regression (MLR) model on which our final model (specified later) is based, such analysis allows multiple observed factors to affect dependent variable y (Wooldridge, 2013, p. 71)

To achieve consistent results in the form of correct estimates of $\beta_0, ...\beta_k$ coefficients, the MLR model needs to meet several assumptions.

Assumption MLR.1 - Linear in parameters

The equation (4.1) represents the model in the population, with $\beta_0, ..., \beta_k$ unknown parameters and unobserved error u.

Assumption MLR.2 - Random sampling

Following the model in assumption MLR.1, we have a random sample of n observations, $(x_i 1, x_i 2, ..., x_i k, y_i) : i = 1, 2, ..., n$.

Assumption MLR. 3 - No perfect collinearity

No independent variable is constant, and there are no exact linear relationships among them.

Assumption MLR. 4 - Zero conditional mean

It holds that $E(u|x_1, x_2, ..., x_k) = 0$. That is the expected value of the error u equals zero for any values of the independent variables.

Assumption MLR. 5 - Homoskedasticity

It holds that $Var(u|x_1, ..., x_k) = \sigma^2$. That is the variance of error u is the same for any values of explanatory variables.

When assumptions MLR.1 through MLR.4 are met, then the resulting $\hat{\beta}$ estimate is an unbiased estimate of the population parameters. That is $E(\hat{\beta}_j) = \beta_j$, for $j = 0, 1, ..., k$, and for any values of β_j (Wooldridge, 2013, p. 83-87).

However, in most cases it is possible to find quite a few unbiased estimators of β_j, therefore the assumption MLR. 5 must be included. If the assumptions MLR.1 to MLR.5 are fulfilled, then the estimator β_j is the best linear unbiased estimator. This is summarized by the Gauss-Markov theorem.

Gauss-Markov theorem

Under Assumption MLR.1 through MLR.5, $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ are the best linear unbiased estimators of $\beta_0, \beta_1, ..., \beta_k$, respectively.

Including the MLR.5 assumption ensures that the resulting OLS estimator has the smallest variance under these assumptions (Wooldridge, 2013, p. 101-102).

Although we expect that our model meets the aforementioned assumptions, certain shortcomings and possible problems must also be taken into account. The model will almost certainly suffer from endogeneity, as there are definitely unobserved characteristics of the districts (that is the error term u) that are correlated with the variables included in the model. The presence of endogeneity (and thus the violation of the assumption MLR.4) is exactly the reason why first-differencing, fixed effects and random effects models are subsequently used, as they are able to deal with this problem. Another potential problem is the possible presence of heteroskedasticity and multicollinearity in the model. Therefore, appropriate tests are introduced later in this chapter.

4.2 Panel Data Estimation Methods

4.2.1 Fixed effects

Fixed effects can be used to estimate panel data, as the so-called fixed effects transformation removes the unobserved (fixed) effect and time-constant independent variables before the actual estimation. The fixed effects transformation takes the original model, or rather its equation, and subtracts from it the equation of the model with the values of the variables being their averages over time. The transformation is shown by the following equations.

$$y_{it} - \bar{y}_i = \beta_0 - \beta_0 + \beta_1(x_{it} - \bar{x_i}) + ... + \beta_k(x_{it} - \bar{x_i}) + a_i - a_i + u_{it} - \bar{u_i}, \quad (4.2)$$

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{it} + ... + \beta_k \tilde{x}_{it} + \tilde{u}_{it}, \quad (4.3)$$

where $\tilde{y}_{it} = y_{it} - \bar{y}_i$.

The same holds for $\tilde{x}_i$ and $\tilde{u}_i$.

Finally, the equation (4.3) can be written as:

$$\tilde{y_{it}} = \beta_1 \tilde{x}_{it1} + ... + \beta_k \tilde{x}_{itk} + \tilde{u}_{it} \quad (4.4)$$

So the results are the time-demeaned data and the equation also shows that the unobserved effect a_i has disappeared. It disappeared because it is not dependent on time and therefore the average over time is still equal to a_i as shown by equation (4.5). Getting rid of the unobserved effect a_i means that the equation (4.4) can be estimated using pooled OLS (Wooldridge, 2013, p. 484-485).

$$\bar{a_i} = \frac{1}{T} T a_i = a_i \quad (4.5)$$

By solving the problem of endogeneity, the assumption of strict exogeneity is introduced. If all independent variables meet this assumption, that is if holds that $E(u_{it}|X_i, a_i) = 0$ for every t, in the model represented by the equation (4.6), along with the assumptions of random sampling and no perfect linear relationships between the independent variables, then the fixed effects estimator is unbiased. The fixed effects estimation has a few more assumptions that are somehow of an extended version of the assumptions for the OLS estimator. These are homoskedasticity of the errors and no serial correlation, meaning that the idiosyncratic errors are not correlated. If these assumptions are met together with the previous four assumptions, then the resulting fixed effects estimator can be labeled as the best linear unbiased estimator. (Wooldridge, 2013, p. 509).

$$y_{it} = \beta_1 x_{it1} + ... + \beta_k x_{itk} + a_i + u_{it} \tag{4.6}$$

4.2.2 Random effects

Fixed effects estimation is used to remove the unobserved effect of a_i, however, if a_i is uncorrelated with all the independent variables included in the model, then using fixed effects does not lead to an efficient estimator and in such a case it is better to use random effects estimation (Wooldridge, 2013, p. 492).

In other words, to be worthwhile to use random effects, it should holds that for t = 1, 2, ..., T and j = 1, 2, ..., k

$$Cov(x_{itj}, a_i) = 0 \tag{4.7}$$

The equation (4.7) can be assumed to hold (that is, the covariance between the unobserved effect a_i and any independent variable is zero) if it is reasonable to believe that all factors relevant to the determination of the dependent variable have been included in the model. That is almost certainly not the case in our example, as there are probably unobserved effects that distort the final model. However, it is reasonable to use random effects also under the assumption that the effect of a_i is relatively small compared to the other variables included in the model.

Therefore, this method is also used in the analysis and its results are included for possible comparison with the results of fixed effects.

To estimate the coefficients β by random effects, the composite error term $v_{it} = a_i + u_{it}$ is formed. The initial equation then has the following form:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + ... + \beta_k x_{itk} + v_{it} \tag{4.8}$$

Unlike in fixed effects estimation, the intercept is included here, as it is necessary to justify the assumption that unobserved effect a_i has a zero mean. By introducing the composite error v_{it}, the unobserved effect a_i, which is itself independent of time, became part of the composite error at each time period. Therefore, it now holds that v_{it} is (positively) serial correlated in time and the equation (4.9) holds (Wooldridge, 2013, p. 492-493).

$$Corr(v_{it}, v_{is}) = \frac{{\sigma_a}^2}{({\sigma_a}^2 + {\sigma_u}^2)}, \tag{4.9}$$

where t and s are different time periods and ${\sigma_a}^2 = Var(a_i)$, ${\sigma_u}^2 = Var(u_it)$.

The transformation under random effects is similar to that under fixed effects, although there are differences. The following equation represents this transformation.

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta_1(x_{it1} - \theta\bar{x}_{i1}) + ... + \beta_k(x_{itk} - \theta\bar{x}_{ik}) + (v_{it} - \theta\bar{v}_i), \tag{4.10}$$

where θ is defined as $1 - [\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)]^{1/2}$ taking values from 0 to 1. Thus the random effects transformation subtracts only portion of time averages of the variables, while transformation under fixed effects subtracts the whole time averages. Such data used in the equation (4.10) is then appropriately referred to as quasi-demeaned. Using the random effects transformation, the problem of serial correlation has been solved, and similarly to fixed effects, it is now possible to use pooled OLS to estimate the equation (4.10) (Wooldridge, 2013, p. 493).

The assumptions for the estimator under random effects are very similar to those for fixed effects, but there are two important differences, namely that the expected value and variance of the unobserved effect a_i is constant. That is $\mathrm{E}(a_i \mid \mathbf{X}_i) = \beta_0$ and $\mathrm{Var}(a_i \mid \mathbf{X}_i) = \sigma_a^2$. Under these assumption, plus the assumptions of fixed effect estimator, the random effects estimator is consistent and asymptotically normally distributed (Wooldridge, 2013, p. 510).

4.3 First-differencing

First-differencing is a method that can be represented by the following equation:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + (u_{i2} - u_{i1}), \tag{4.11}$$

for $t = 2$.

The equation 4.11 was created by subtracting the 4.12 equation from the 4.13 equation.

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1} \tag{4.12}$$

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2} \tag{4.13}$$

These two equations represent observations from two different time periods, and thus in the resulting equation 4.11, which can be further simplified to

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i, \tag{4.14}$$

the variables are equal to the differences between the two observed time periods. The important thing is that by subtracting the equations, the unobserved effect a_i disappeared (Wooldridge, 2013, p. 461).

The first-differencing method is mainly used when working with panel data of two time periods, however, it can be used for $t > 2$ as well (Wooldridge, 2013, p. 490).

4.4 Testing the Assumptions

Several tests verifying the correctness of our models and non-violation of the assumptions of the chosen models must be performed. Specifically, these tests are used to assess multicollinearity, heteroskedasticity and serial correlation in the models.

Although the mutual correlation of variables was tested in the Chapter 3, after which problematic variables were removed, and which theoretically should prevent the issue of multicollinearity, it is advisable to test the model itself for the presence of multicollinearity with an appropriate test. A *Variance Inflation Factor* (VIF) is used for this purpose. The VIF is calculated for each variable and, as the term in the variation of the β estimate, is a function of R^2 for a given coefficient j and $\mathrm{VIF}_j = 1/\left(1 - R_j^2\right)$ applies. It is usually stated that if the VIF value is more than 10, then the model suffers from multicollinearity (Wooldridge, 2013, p. 98).

To test the presence of heteroskedasticity, the Breusch-Pagan test is utilized. Assuming a normal distribution of the disturbances, the Breusch-Pagan test is performed with the null hypothesis of homoskedastic errors (Breusch & Pagan, 1979). The null hypothesis is rejected when the p-value is less than 0.05, meaning the presence of heteroskedasticity in the model, in which case heteroskedasticity-robust standard errors should be used (Wooldridge, 2013, p. 277). If the p-value is large enough (that is greater than 0.05), then the model is fine and heteroskedasticity is not present in it.

Rather for orientation purposes, the autocorrelation is calculated using the Breusch-Godfrey test. For orientation purposes means that in this case it is calculated only on non-coronavirus variables, because for logical reasons, there are no COVID-19 statistics for the years 2013 and 2017 and because cumulative values for the COVID-19 variables are used, that is only one numerical value for each district is used. Although, for example, a certain serial correlation could be expected for the variable *COVIDinfected*, since, for instance, if the number of infected people was high last month, it is quite possible that there will be a high number of infected people this month as well.

Finally, the Hausman test is used to determine which of the methods, fixed or random effects, is more suitable for testing the given model. The null hypothesis in the Hausman test is that the unobserved effect a_i is uncorrelated with the independent variables in the model. If the null hypothesis is rejected at the p-value < 0.05, it means correlation between them and it is more appropriate to use the fixed effects method.

Chapter 5

Empirical Results

The empirical results of the chosen models are presented in this chapter together with their description and the description of the results of individual tests. Overall, several analyzes are performed, first the year 2021 is analyzed separately using OLS regression. Subsequently, a panel of data for the years 2013, 2017 and 2021 is analyzed using fixed effects or random effects, and first-differencing. Finally, panel data are analyzed for pairs of years, first 2017 and 2021 and then 2013 and 2021 using first-differencing.

5.1 OLS Regression for 2021

When testing multicollinearity using VIF, the value of the *University* variable was equal to 18.46, which significantly exceeded the set threshold of VIF = 10. This variable already had relatively high correlation values in the case of testing mutual correlations in the Chapter 3. However, it did not exceed the set limit of the correlation coefficient and was therefore kept in the analysis after the initial sorting of the variables, and also because it is generally an important variable in the analysis of such a topic as factors influencing decision making of voters. However, the VIF value is too high for this variable to be retained in the final model, thus it is dropped from it. After discarding *University* and recalculating the VIF values, everything is fine and no other variable needs to be removed.

By performing the Breusch-Pagan test on the model with all dependent variables in turn, the presence of heteroskedasticity was not detected. That is, all p-values were higher than 0.05 and thus the null hypothesis of homoskedastic errors cannot be rejected, our model does not suffer from heteroskedasticity, and there is no need to use robust standard errors[1].

Table 5.1 shows the results of the OLS regression for the year 2021. The coronavirus variables *COVIDdeath* and *COVIDinfected* have different effects on turnout, the former having a negative effect, while the latter having a positive effect. The opposite effects are also discovered regarding the results of political parties. They have the same effect in only one case, namely the SPOLU coalition, on which both have a negative effect. However, all relationships between the two COVID-19 variables and the dependent variables are non-significant, except in one case. The number of infected people had a negative effect on KSCM's electoral gains. According to the estimate, which is significant at the 10 % significance level, if the number of infected people in a given district increases by 100 people per 1,000 inhabitants, the KSCM will lose 2.3 percentage points of votes. This negative relationship is not so surprising, as the KSCM is mainly supported by the elderly, and if there are many infected people around them, then out of concern for their health, they might not go to the polls and prefer to stay at home. Of the other factors, *PrimaryEducation* or the *Unemployment* played a significant role in the 2021 parliamentary elections. All models (that is, analyzing any of the dependent variable) show a high R^2 value. The largest R^2 has a model with *Turnout* as a dependent variable, where the selected variables capture almost 85 variations in the results. A good result is also that all F statistics are significant, and thus the models are overall significant, which supports the correct selection of independent variables.

[1] Robust standar errors are still provided in the parentheses in Table 5.1.

Table 5.1: OLS Regression for 2021

	Dependent variable:					
	Turnout	SPOLU	ANO	SPD	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)	(6)
COVIDdeath	−0.042	−0.041	0.028	0.024	0.022	0.011
	(0.040)	(0.058)	(0.053)	(0.031)	(0.015)	(0.013)
COVIDinfected	0.011	−0.011	−0.002	−0.009	−0.003	−0.023***
	(0.016)	(0.023)	(0.021)	(0.012)	(0.006)	(0.005)
Unemployment	−0.921***	−1.199***	0.827**	0.196	0.133	0.055
	(0.252)	(0.366)	(0.334)	(0.192)	(0.096)	(0.079)
Foreigners	−0.180	0.177	−0.199	−0.202**	0.068	0.021
	(0.122)	(0.177)	(0.162)	(0.093)	(0.046)	(0.038)
Believers	0.026	0.087	−0.047	0.021	0.040***	−0.018
	(0.037)	(0.054)	(0.049)	(0.028)	(0.014)	(0.012)
PrimaryEducation	−1.861***	−1.493***	1.631***	0.799***	−0.289***	0.038
	(0.256)	(0.372)	(0.340)	(0.195)	(0.098)	(0.080)
HighSchool	0.286**	−0.311	0.011	−0.099	0.316***	0.268***
	(0.143)	(0.208)	(0.190)	(0.109)	(0.055)	(0.045)
AgeOld	−0.456**	−0.059	0.342	0.152	−0.025	−0.025
	(0.205)	(0.298)	(0.272)	(0.156)	(0.078)	(0.064)
Constant	82.906***	65.053***	1.344	3.999	−10.174***	−6.868**
	(8.375)	(12.188)	(11.130)	(6.378)	(3.194)	(2.628)
Observations	77	77	77	77	77	77
R^2	0.843	0.723	0.725	0.633	0.586	0.558
Adjusted R^2	0.824	0.691	0.693	0.590	0.537	0.506
Residual Std. Error (df = 68)	1.903	2.769	2.529	1.449	0.726	0.597
F Statistic (df = 8; 68)	45.629***	22.235***	22.445***	14.685***	12.020***	10.734***

Note: *p<0.1; **p<0.05; ***p<0.01

5.2 Panel data results for 2013, 2017 and 2021

This section analyzes panel data containing all three election years (that is 2013, 2017 and 2021).The models in this section does not use data that comes from the 2011 census, which are the variables *Believers* and *PrimaryEducation*, *HighSchool* and *University* (the variable *NoEducation* has already been dropped from the analysis earlier). The dependent variable *SPD* is also removed, as this political party did not yet exist in 2013.

5.2.1 Fixed and Random Effects models

Heteroskedasticity is present in the model, since the Breusch-Pagan test revealed very low p-values, thus robust standard errors[2] must be used.

The implementation of the Hausman test determined the fixed effects model as the more appropriate method, as the p-value is very low. Thus, these panel data are analyzed with the use of fixed effects, the results of the analysis are shown in the 5.2 (results using random effects for comparison are presented in the Table 7.1 in appendix).

The fixed effects model with such a limited number of variables shows relatively many significant results. Of the independent variables of our main interest, it is the cumulative number of infected people *COVIDinfected* that influenced the election results more , as it is significant at the 1 % level for all dependent variables except KSCM. The number of infected most significantly influenced the election results of ANO, where it had a negative effect. If 100 more people per 1,000 inhabitants were infected in the district, it meant a loss of 6.3 percentage points for ANO. It had the positive effect on other parties, the SPOLU coalition and CSSD benefited from a higher infection rate among the population with 1 % significance level. The SPOLU coalition has a positive and significant relationship even with the *COVIDdeath* variable. If 100 more people died per 10,000, then SPOLU obtained 6.9 percentage points of more votes. Given that this model contains independent variables that are with few exceptions all statistically signific-

[2]Robust standard errors are in the parenthesis in Table 5.2.

ant, a high R^2 value can also be expected. The results confirm this, because for most independent variables this value exceeds 0.9, which is a really high number. The F Statistic again demonstrates overall significance.

Table 5.2: Fixed Effects Estimation, All Years

	Dependent variable:				
	Turnout	SPOLU	ANO	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)
COVIDdeath	0.005	0.069**	0.020	0.013	−0.028
	(0.013)	(0.029)	(0.039)	(0.039)	(0.023)
COVIDinfected	0.021***	0.036***	−0.063***	0.026***	0.006
	(0.003)	(0.006)	(0.008)	(0.008)	(0.004)
Unemployment	0.047	0.041	−0.530***	0.876***	0.415***
	(0.063)	(0.134)	(0.184)	(0.183)	(0.106)
Foreigners	0.391***	−0.640***	0.869***	−1.512***	−0.747***
	(0.092)	(0.196)	(0.269)	(0.268)	(0.155)
AgeOld	0.542***	−1.475***	4.243***	−3.908***	−2.504***
	(0.133)	(0.282)	(0.388)	(0.387)	(0.224)
Observations	231	231	231	231	231
R^2	0.951	0.764	0.910	0.946	0.965
Adjusted R^2	0.925	0.635	0.860	0.917	0.945
F Statistic (df = 5; 149)	583.053***	96.276***	299.785***	524.724***	811.997***

Note: *p<0.1; **p<0.05; ***p<0.01

5.2.2 First-differencing

Although for t greater than 2, which is this example, since we have observations from 3 different years, fixed effects method is usually used rather than first-differencing method, it is not always necessarily the case that first-differencing is less appropriate. The Breusch-Godfrey test to detect serial correlation in the model can also help us decide which method to use. Its application reveals a certain degree of serial correlation in the model, as the p-value is very low, less than 0.05, and thus the null hypothesis of no correlation of idiosyncratic errors of individual variables over time is rejected, and in such a case, the first-differencing method appears to be more suitable (Wooldridge, 2013, p. 490). This is the reason why the first-differencing method is also applied to these panel data.

Applying VIF revealed one variable with value greater than 10, but this is the COVID-19 variable *COVIDinfected*, which cannot be excluded from the model as it is a variable of our main interest. Based on Breusch-Pagan results, the presence of heteroskedasticity was detected in the model[3].

Regression results are displayed in Table 5.3. The overall significance of the model is again satisfied due to the significant values of the F Statistic. The R^2 is also still high for all independent variables, but compared to the fixed-effects estimation, it has decreased a bit overall, so in the case of the fixed-effects estimation of this panel data, the independent variables in the model explain the dependent variables a bit better. Other indicators have also changed. The effect of *COVIDdeath* on ANO results is now negative, whereas it was positive for fixed effects. *COVIDdeath* became statistically significant at the 5 % level for the dependent variable *Turnout*, on the contrary, the number of deaths no longer significantly affects the results of the SPOLU coalition. The number of infected people, that is *COVIDinfected*, still remains the defining coronavirus statistic, now statistically affecting the KSCM as well. The impact of this variable on individual dependent variables, in the sense of positivity and negativity, remained the same, and

[3]Robust standar errors are in the parenthesis in Table 5.3.

the magnitude of the coefficients did not change much either. Of the other independent variables, the variable *Foreigners* completely lost its significance, and almost completely also the variable *Unemployment*, which is now significant only for one dependent variable, voter turnout. The *AgeOld* variable remained relatively significant, and the values of some coefficients also changed noticeably. For instance, for the CSSD political party, the value of the coefficient decreased from -3.908 to -0.778. The difference between these values is -3.13, meaning if the share of people over 65 years old increased by 1 percentage point, then the CSSD in the given district lost roughly 3.13 percentage points less of share of votes than in the case of fixed effects estimation.

Table 5.3: First-differencing Estimation, All Years

	Dependent variable:				
	Turnout	SPOLU	ANO	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)
COVIDdeath	0.028**	0.014	−0.024	0.006	−0.024
	(0.012)	(0.028)	(0.036)	(0.025)	(0.017)
COVIDinfected	0.010***	0.057***	−0.067***	0.058***	0.016***
	(0.003)	(0.006)	(0.008)	(0.006)	(0.004)
Unemployment	0.192***	−0.157	−0.225	0.176	0.127
	(0.061)	(0.143)	(0.181)	(0.126)	(0.087)
Foreigners	−0.062	0.131	0.332	0.106	−0.129
	(0.117)	(0.272)	(0.345)	(0.241)	(0.166)
AgeOld	−0.251	−0.226	3.219***	−0.778**	−1.418***
	(0.184)	(0.428)	(0.543)	(0.378)	(0.260)
Constant	2.671***	−4.219***	4.123***	−11.254***	−4.128***
	(0.436)	(1.014)	(1.286)	(0.896)	(0.617)
Observations	154	154	154	154	154
R^2	0.829	0.863	0.906	0.917	0.702
Adjusted R^2	0.823	0.858	0.902	0.914	0.692
F Statistic (df = 5; 148)	143.674***	185.688***	283.763***	326.081***	69.842***
Note:					*p<0.1; **p<0.05; ***p<0.01

5.3 Comparison of 2021 with individual previous years separately

If $t = 2$, as in this case, the fixed effects method and first-differencing are the same thing, and both methods will give identical results, but it is more straightforward to use first-differencing. (Wooldridge, 2013, p. 490). Now that the years 2013 and 2017 are no longer being analyzed simultaneously, we can use the 2011 census data for them and thus include the variables *Believers*, *PrimaryEducation* and *HighSchool*.

5.3.1 First-differencing for years 2021 and 2017

The model for these years does not detect the presence of multicollinearity, as the highest VIF value is 5.353689 for the variable *University*, so in this case this variable can be included in the analysis. However, the model in case of some dependent variables has very low p-values, lower than 0.05, meaning the presence of heteroskedasticity. Therefore, robust standard errors are utilized[4].

The results of the COVID-19 variables *COVIDdeath* and *COVIDinfected* in the Table 5.4 show that the pandemic had a different impact on the elections and electoral gains of individual parties. Both variables had a negative, but very low and statistically insignificant effect on voter turnout. On the contrary, both had a positive impact on the SPOLU coalition and the SPD party, however, with the exception of the *COVIDdeath* effect on SPOLU's results, remained insignificant. Both of these variables were significant only for CSSD. Interesting is that *COVIDdeath* had a negative effect, while *COVIDinfected* had a positive effect. A district with 100 more deaths per 10,000 inhabitants from COVID-19 reduced the share of votes received by the CSSD party by 4.4 percentage points on average compared to the 2017 election. The number of infected people had a similar effect on ANO's electoral gains. In this case, the variable *University*, which is the most significant of the non-coronavirus independent variables, entered the model for

[4] Robust standard errors are in the parenthesis in Table 5.4.

the first time. R^2 values are noticeably lower than in the case of previous models. However, in most cases, the independent variables in this model still explain over 40 percent of the outcomes of the dependent variables. In the case of Turnout, R^2 is equal to 0.279, and in this case the results could be described as debatable. Overall, the model is statistically significant at the 1 % level, according to the resulting F Statistic.

Table 5.4: First-differencing Estimation, Years 2021-2017

	Dependent variable:					
	Turnout	SPOLU	ANO	SPD	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)	(6)
COVIDdeath	−0.005	0.017	0.012	0.008	−0.044***	−0.007
	(0.011)	(0.021)	(0.027)	(0.013)	(0.013)	(0.012)
COVIDinfected	−0.004	0.016**	−0.040***	0.008	0.025***	0.026***
	(0.004)	(0.008)	(0.010)	(0.005)	(0.005)	(0.005)
Unemployment	0.128	0.297	−0.176	0.165	0.229	−0.073
	(0.135)	(0.269)	(0.338)	(0.162)	(0.170)	(0.157)
Foreigners	−0.041	0.298	0.125	0.146	0.109	−0.094
	(0.104)	(0.207)	(0.260)	(0.125)	(0.130)	(0.120)
Believers	0.175*	−0.442**	0.052	−0.092	−0.062	0.121
	(0.091)	(0.182)	(0.229)	(0.110)	(0.115)	(0.106)
PrimaryEducation	−0.072	0.310	−0.315	0.211*	0.227*	0.073
	(0.103)	(0.205)	(0.257)	(0.123)	(0.129)	(0.119)
HighSchool	0.103*	0.316***	−0.069	0.049	0.072	−0.079
	(0.058)	(0.116)	(0.146)	(0.070)	(0.073)	(0.067)
University	−0.243**	0.418**	−0.396*	−0.286**	0.156	0.441***
	(0.093)	(0.186)	(0.234)	(0.112)	(0.117)	(0.108)
AgeOld	−0.343	0.261	0.076	0.221	1.273***	0.555*
	(0.246)	(0.489)	(0.615)	(0.295)	(0.309)	(0.285)
Constant	6.842***	2.196	3.655	−0.311	−6.881***	−11.426***
	(1.278)	(2.540)	(3.195)	(1.532)	(1.604)	(1.479)
Observations	77	77	77	77	77	77
R^2	0.279	0.446	0.439	0.417	0.492	0.623
Adjusted R^2	0.183	0.372	0.363	0.339	0.424	0.572
F Statistic (df = 9; 67)	2.886***	5.994***	5.817***	5.334***	7.222***	12.281***

Note: *p<0.1; **p<0.05; ***p<0.01

5.3.2 First-differencing for years 2021 and 2013

Once again the dependent variable *SPD* is removed from the model, since this political party did not yet exist in 2013.

As in the case of the 2021 and 2017 comparison, all VIF values are below 10 now as well, so there is no need to remove any of the variables. Some p-values are again below 0.05, indicating the presence of heteroscedasticity and therefore the application of robust standard errors[5].

Similar to the comparison between 2021 and 2017, in this case too, *COVIDdeath* is significant only for the results of the CSSD party and affects it in a negative way. The variable *COVIDinfected* is again much more important, in this case it is significant for all dependent variables, although in the case of voter turnout it is only at the 10 % level. Compared to the results of the 2013 elections, the number of infected people harmed both entities that entered the Chamber of Deputies in the 2021 elections, that is the negative effect of the variable *COVIDinfected* is in the case of SPOLU coalition and ANO. On the contrary, this variable has a positive relationship with the parties CSSD and KSCM. The strongest effect of *COVIDinfected* was on the ANO party, where the estimate is equal to -0.088, and thus in the district with an increase in the number of infected people by 100 per 1,000 inhabitants, the ANO party lost an average of 8.8 percentage points. As in the comparison of 2021 and 2017, in this case too the *University* variable is the most significant from the other variables, which therefore indicates the importance of this factor in the voters' decision-making. Voter turnout increases significantly with high school attainment, as shown by the estimate of the *HighSchool* variable equal to 0.659 at the 1 % level. Value of R^2 is high enough to consider the models plausible, with the exception of the dependent variables *Turnout* and *SPOLU*, where it does not reach high values. In the case of a model with a dependent variable *SPOLU*, the model is not overall significant as shown by the F Statistic, meaning that the selected independent variables do not explain the dependent variable *SPOLU* and

[5]Robust standard errors are in the parenthesis in Table 5.5.

the results of such a test cannot be used. For the other dependent variables, the model is overall significant.

Table 5.5: First-differencing Estimation, Years 2021-2013

	Dependent variable:				
	Turnout	SPOLU	ANO	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)
COVIDdeath	−0.014	0.034	0.077	−0.128***	−0.016
	(0.019)	(0.036)	(0.054)	(0.035)	(0.030)
COVIDinfected	0.014*	−0.041***	−0.088***	0.074***	0.050***
	(0.007)	(0.013)	(0.020)	(0.013)	(0.011)
Unemployment	0.094	−0.193	0.009	0.133	0.122
	(0.083)	(0.158)	(0.235)	(0.155)	(0.132)
Foreigners	0.195	−0.232	0.121	0.359	−0.219
	(0.126)	(0.240)	(0.358)	(0.236)	(0.202)
Believers	0.120	−0.315	0.418	−0.030	0.401
	(0.164)	(0.313)	(0.467)	(0.308)	(0.263)
PrimaryEducation	0.017	−0.228	−0.906*	1.406***	−0.390
	(0.179)	(0.341)	(0.507)	(0.334)	(0.286)
HighSchool	0.659***	−0.228	−0.766	0.731**	−0.432
	(0.179)	(0.341)	(0.508)	(0.335)	(0.287)
University	0.283	−0.167	−2.877***	1.130***	1.423***
	(0.214)	(0.409)	(0.608)	(0.401)	(0.343)
AgeOld	−0.206	−0.844*	−0.319	1.439***	0.734*
	(0.257)	(0.489)	(0.729)	(0.480)	(0.411)
Constant	4.918*	9.147*	29.751***	−26.749***	−31.402***
	(2.595)	(4.945)	(7.363)	(4.853)	(4.154)
Observations	77	77	77	77	77
R^2	0.292	0.188	0.666	0.626	0.644
Adjusted R^2	0.197	0.079	0.622	0.576	0.597
F Statistic (df = 9; 67)	3.067***	1.723	14.876***	12.479***	13.489***

Note: *p<0.1; **p<0.05; ***p<0.01

Chapter 6

Conclusion

This work investigated the effects of COVID-19 on the results of the parliamentary elections in the Czech Republic in 2021. The analysis was carried out at the level of the districts of the Czech Republic, and Prague was also included. The dependent variables were voter turnout and the share of votes received by 5 political entities (1 coalition and 4 political parties). The key independent variables were the cumulative number of deaths per 10,000 inhabitants and the cumulative number of infected people per 1,000 inhabitants. Numerous economic and socio-demographic variables were also included, such as unemployment, level of education, or the proportion of old people in society.

Several different econometric models were used to estimate the results. First, the elections were analyzed using OLS regression, which showed a negative relationship between the number of infected and the election results of KSCM, possible explanation of that would be that KSCM voters (who are mainly older people) did not go to the polls due to fear of infection. Subsequently, panel data were analyzed, when election results from the 2013 and 2017 parliamentary elections were added. Some variables had to be removed from this panel dataset with all election years due to insufficient or complete unavailability of data. Specifically, these were the election results of the SPD party, the proportion of religious people, and all the variables related to education.

Fixed effects, random effects and first-differencing methods were used to analyze such data. Firstly, based on the Hausman test, the fixed effects model was determined to be more appropriate than the random effects model. Its results showed a significant effect of the number of infected people on all dependent variables with the exception of the KSCM, which was positive in the case of voter turnout and electoral gains of the SPOLU coalition and CSSD. Conversely, the number of infected people had a negative impact on the ANO results. This can be understood, for example, in the way that citizens could blame the ANO government party for the bad coronavirus situation, a similar explanation was offered by Baccini et al. (2021) in the case of the US presidential election. Subsequently, these panel data were analyzed using the first-differencing model, basically confirmed most of the fixed effects results. According to the model, the number of infected people had a significant effect on all dependent variables, while the only negative effect was towards ANO.

Finally, first-differencing estimation was performed on the panel data containing the years 2021 and 2017 and the panel data containing the years 2021 and 2013. In this case, the excluded variables were returned to the model (with the exception of SPD, which was only returned to the model comparing the years 2021 and 2017) .

After removing 2013, there were minor changes in the results. The number of infected slightly lost its significance, but it still remained significant for most of the dependent variables. Moreover, still with a negative effect on the results of ANO and a positive effect on the results of other political parties. The number of deaths related to COVID-19 significantly affected only the electoral gains of CSSD, in a negative way.

The removal of 2017 basically only confirmed the findings so far, that is also in this case the number of infected people significantly influenced all dependent variables, in this case not only the results of ANO, but also the results of SPOLU coalition were negatively affected. However, the model with the SPOLU coalition was not overall significant (this was the only case of a non-significant model).

By comparing the results of all models, it can be concluded that the cumulative number of infected significantly influenced the results of the elections to the Chamber of Deputies in 2021. It did not have too much influence on the turnout, although even here we can see a rather positive, albeit very small, effect. However, it had a great influence on the election results of political parties. On the one hand, it clearly harmed the ANO party, on the other hand, it helped most of the other parties. The number of deaths had rather negligible results and most estimates came out insignificant, with the exception of the CSSD, which it rather harmed. In general, it can be concluded that COVID-19 had a mobilizing effect and motivated people to vote, rather than the opposite.

I hope that this work could contribute to the emerging literature on the impact of COVID-19 on electoral outcomes and politics as such.

Bibliography

Antunes, R. (2010), 'Theoretical models of voting behaviour', *Exedra* **4**(1), 145–70.

Baccini, L., Brodeur, A. & Weymouth, S. (2021), 'The covid-19 pandemic and the 2020 us presidential election', *Journal of Population Economics* **34**(2), 739–767.

Boschken, H. L., Goldthorpe, J. H., Hout, M., Houtman, D., Manza, J., Nieuwbeerta, P., Pakulski, J., Rempel, M. & Weakliem, D. L. (2001), *The breakdown of class politics: A debate on post-industrial stratification*, Woodrow Wilson Center Press.

Breusch, T. S. & Pagan, A. R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica: Journal of the econometric society* pp. 1287–1294.

Campbell, A., Converse, P. E., Miller, W. E. & Stokes, D. E. (1980), *The american voter*, University of Chicago Press.

Campbell, D. E. (2007), *A matter of faith: Religion in the 2004 presidential election*, Brookings Institution Press.

Černỳ, J. (2019), 'Population characteristics of voters: Evidence from the czech parliamentary election'.

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M. & Team, T. O. I. (2020), *How did COVID-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data*, Vol. 27431, National Bureau of Economic Research Cambridge, MA.

Dee, T. S. (2004), 'Are there civic returns to education?', *Journal of public economics* **88**(9-10), 1697–1720.

Donthu, N. & Gustafsson, A. (2020), 'Effects of covid-19 on business and research'.

Downs, A. (1957), 'An economic theory of political action in a democracy', *Journal of political economy* **65**(2), 135–150.

Dusková, N. (2021), 'Do less developed regions vote populistic parties? the case of slovakia'.

Ferejohn, J. (1986), 'Incumbent performance and electoral control', *Public choice* pp. 5–25.

Fernandez-Navia, T., Polo-Muro, E. & Tercero-Lucas, D. (2021), 'Too afraid to vote? the effects of covid-19 on voting behaviour', *European Journal of Political Economy* **69**, 102012.

Hoskins, B., D'hombres, B. & Campbell, J. (2008), 'Does formal education have an impact on active citizenship behaviour?', *European educational research journal* **7**(3), 386–402.

Johnson, A. F., Pollock, W. & Rauhaus, B. (2020), 'Mass casualty event scenarios and political shifts: 2020 election outcomes and the us covid-19 pandemic', *Administrative Theory & Praxis* **42**(2), 249–264.

Kolstad, I. & Wiig, A. (2016), 'Education and electoral participation: reported versus actual voting behaviour', *Applied Economics Letters* **23**(13), 908–911.

Kotrč, M. (2017), 'The rise of extreme right: Evidence from the slovak parliamentary election'.

Lazarsfeld, P. F. (1944), 'The controversy over detailed interviews—an offer for negotiation', *Public opinion quarterly* **8**(1), 38–60.

Malhotra, N. & Kuo, A. G. (2008), 'Attributing blame: The public's response to hurricane katrina', *The Journal of Politics* **70**(1), 120–135.

Manza, J. & Brooks, C. (2008), 'Class and politics', *Social class: How does it work* pp. 201–231.

McKee, M., Gugushvili, A., Koltai, J. & Stuckler, D. (2021), 'Are populist leaders creating the conditions for the spread of covid-19? comment on" a scoping review of populist radical right parties' influence on welfare policy and its implications for population health in europe"', *International journal of health policy and management* **10**(8), 511.

Parzuchowski, A. S., Peters, A. T., Johnson-Sasso, C. P., Rydland, K. J. & Feinglass, J. M. (2021), 'County-level association of covid-19 mortality with 2020 united states presidential voting', *Public Health* **198**, 114–117.

Smidt, C., den Dulk, K., Froehle, B., Penning, J., Monsma, S. & Koopman, D. (2010), *The disappearing God gap?: religion in the 2008 presidential election*, Oxford University Press.

Werts, H., Scheepers, P. & Lubbers, M. (2013), 'Euro-scepticism and radical right-wing voting in europe, 2002–2008: Social cleavages, socio-political attitudes and contextual characteristics determining voting for the radical right', *European Union Politics* **14**(2), 183–205.

Wooldridge, J. M. (2013), 'Introductory econometrics: A modern approach (5th ed.)'.

Chapter 7

Appendix

Table 7.1: Random Effects Estimation, All Years

	Dependent variable:				
	Turnout	SPOLU	ANO	CSSD	KSCM
	(1)	(2)	(3)	(4)	(5)
COVIDdeath	−0.003	0.042	0.166***	−0.058	−0.046*
	(0.018)	(0.035)	(0.057)	(0.053)	(0.026)
COVIDinfected	0.026***	0.035***	−0.060***	0.001	−0.006
	(0.003)	(0.007)	(0.011)	(0.010)	(0.005)
Unemployment	−0.147**	0.001	−1.063***	1.495***	0.956***
	(0.073)	(0.126)	(0.151)	(0.120)	(0.074)
Foreigners	0.149	−0.073	0.130	−0.421***	−0.373***
	(0.099)	(0.157)	(0.164)	(0.118)	(0.085)
AgeOld	0.206	−1.415***	2.107***	−1.288***	−1.158***
	(0.152)	(0.256)	(0.284)	(0.211)	(0.144)
Constant	56.147***	48.326***	−7.457	30.226***	29.044***
	(3.301)	(5.532)	(6.085)	(4.464)	(3.100)
Observations	231	231	231	231	231
R^2	0.876	0.530	0.652	0.777	0.917
Adjusted R^2	0.874	0.520	0.644	0.772	0.915
F Statistic	1,596.244***	254.207***	420.766***	785.888***	2,496.221***

Note: *p<0.1; **p<0.05; ***p<0.01