CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

The Impact of Vaccinations on the Development of Covid-19 Pandemic

Master's thesis

Author: Bc. Vít Kulhánek

Study program: Economics and Finance

Supervisor: Mgr. Roman Kalabiška

Year of defense: 2023

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, April 25, 2023

Vit Kulhanek

Abstract

This thesis aims to examine the effect of vaccination on the development of the Covid-19 pandemic. The three key variables are used as dependent variables: the number of new cases, new deaths, and hospitalization. The dataset containing numerous countries and capturing periods from 2020 to 2022 was obtained, therefore a panel data estimator was employed. Moreover, the Czech Republic and Israel were selected for deeper investigation, and their data were filtered from the dataset. The data structure changed from panel data to time series, so OLS regression was selected as an appropriate method. In all models, vaccination variables and also several others were included in lags because a time gap is necessary to increase individuals' immunity in the case of the vaccine. Last but not least, the excess deaths analysis is created and focuses on investigating excess deaths caused primarily or secondarily by the Covid-19 pandemic. Furthermore, it predicts the amount of money not paid in the form of pensions till 2030 for the elderly who are included in the excess deaths. Finally, it compares this amount of money with the expenditures associated with vaccine purchases.

JEL Classification	C01, C23, I10, I31
Keywords	Covid-19, vaccination, panel data, time series data
Title	The Impact of Vaccinations on the Development of Covid-19 Pandemic

Abstrakt

Cílem této práce bylo zkoumat vliv vakcinace na vývoj epidemie Covid-19. Závislými proměnnými byly vybrány tři klíčové faktory. Jedná se o počet nových případů, úmrtí a množství hospitalizací. Data byla získána pro velké množství států a zachycují období od roku 2020 do roku 2022, proto byl použit estimátor pro panelová data. Následně byly blíže zkoumány vybrané země, a to Česká republika a Izrael. Data pro tyto dva státy byla získána z celosvětového datasetu, což změnilo strukturu dat z panelového typu na časové řady. OLS regrese byla vybrána jako vhodná metoda pro práci s daty jednotlivých států. Proměnné zachycující vakcinace a několik dalších byly ve všech modelech zahrnuty se zpožděními, protože je potřeba určité časové období, aby se v případě vakcinace vybudovala imunita jednotlivců. V neposlední řadě je vytvořená analýza nadměrných úmrtí, která zkoumá přebytečnou smrtnost způsobenou primárně nebo sekundárně pandemií Covid-19. Analýza predikuje množství peněz, které nejsou vyplaceny formou penze důchodcům zahrnutým v přebytečných úmrtí do roku 2030. Na závěr je srovnáváno množství nevyplacených peněz do roku 2030 s náklady, které byly použity na nákup vakcín proti Covidu-19.

Klasifikace JEL	C01, C23, I10, I31
Klíčová slova	Covid-19, vakcinace, panelová data, data časových řad
Název práce	Vliv vakcinace na pandemii Covid-19

Acknowledgments

I would like to express my gratitude to my supervisor Mgr. Roman Kalabiška for his guidance, time and advice. I would also like to offer my special thanks to my family and friends for their endless support throughout my study.

Typeset in LaTeXusing the IES Thesis Template.

Bibliographic Record

Kulhanek, Vit: *The Impact of Vaccinations on the Development of Covid-19 Pandemic*. Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 101. Advisor: Mgr. Roman Kalabiška

Contents

List of Tables

List of Figures

Acronyms

CZK Czech Koruna

CZSO Czech Statistical Office

FD First Difference

GMM Generalized Method of Moments

HAC Heteroskedasticity and Autocorrelation Consistent

OLS Ordinary Least Squares

SE Standard Errors

VIF Variance Inflation Factor

WHO World Health Organization

Master's Thesis Proposal

Author	Bc. Vít Kulhánek
Supervisor	Mgr. Roman Kalabiška
Proposed topic	The Impact of Vaccinations on the Development of Covid-19 Pandemic

Motivation Covid-19 is a term which was not known three years ago. However, now it is one of the most repeated words and everyone can hear about it on daily basis. The Covid-19 pandemic has had an impact on everybody and we had to get used to living with it. Consequently, there are ongoing attempts to offset and put an end to its negative repercussions. One possible way how to return to the life everyone was used to living appears to be vaccination. The main goal of vaccine is to reduce the number of hospitalizations and deaths by increasing individuals' immunity. Moreover, it could also reduce the number of transmissions. The aim of this thesis is to explore whether vaccines have the expected results.
To the best of our knowledge, there are a lot of medical papers related to the effect of vaccination. However, there are not many data analyses focusing on similar topics because data related to this issue were not available. There are many variables which have been updated daily since the beginning of the pandemic, nevertheless it takes some time to develop vaccines against this virus. Vaccinations have started in 2021 but their number has increased rapidly since the beginning of the second quarter of 2021.

Hypotheses

Hypothesis #1: The vaccine has significant and slowing effect on the development of Covid-19 pandemic.

Hypothesis #2: There are other significant factors reducing the pandemic such as weather, restrictions.

Hypothesis #3: There is potential reduction of mandatory social expenses on government's budget which can partially compensate the extra costs related to Covid-19 in the long run.

Methodology For the purpose of analysis, I will use several data sources. Firstly, it is necessary to determine that I will be working with the panel data capturing variables for many countries all over the world. The data capturing the information of Covid-19 will be downloaded from Our World in Data website which contains daily updated observations. I will use the observations aggregated into greater units such as weeks to capture trend of development of variables. Aggregation should reduce the distortion of observations caused by weekends testing and vaccination which were always the lowest of the week. Furthermore, dataset will be scaled by the total population to ensure higher cross-country comparability. Additional data will be taken from Weatherbase website which contains average temperature in countries. Since I will be working with panel data containing many countries, lagged dependent variable will probably also be included, therefore the Generalized Method of Moments will be most likely used in this part of thesis.
Next section will be devoted to a particular country, and I would like to focus on the Czech Republic and carry out a deeper analysis containing different types of restrictions (levels of lockdowns). This section will be build based on the cross-sectional type of data which needs OLS or other type of econometric regression. It will provide the answer to the second hypothesis.
Furthermore, I would like to create an excess deaths analysis of deaths related to Covid-19. According to official numbers, there are more than 40 000 people who have died from this virus in the Czech Republic. Based on the mortality data from the Czech Statistical Office (further as CZSO), I would like to model excess quantity of pensioners who passed away during the Covid-19 pandemic and evaluate the effect on the amount of money that government pays to the elderly. Based on this model, I would like to predict whether government's pension budget reduction caused by Covid-19 deaths could partially compensate for the increased pandemic spending in the long run.

Expected Contribution The results of the proposed work should provide evidence whether vaccination has a reducing effect on the pandemic and if there are additional factors, such as weather, which might affect further expansion. There are not many complex analyses containing data from countries all over the world related to a similar topic. The thesis could be used for further analyses explaining for instance situation in a particular country and its comparison with the existing results.
The focus on the Czech Republic is the main contribution of this thesis. It should provide the answer whether nationwide restrictions had a slowing effect on Covid-19

development. Moreover, excess deaths analysis looks for potential surpluses in fiscal policy and its comparison to excess Covid-19 spending.
Last but not least, I will test whether countries with higher expenditures observe greater reduction of number of positively tested people or a decrease in deaths caused by Covid-19.
The results might be used in future to determine how much each percentage of vaccinated people could reduce the numbers of new Covid-19 cases because the vaccines are effective only for a certain period before an additional vaccine dose is necessary in order to maintain protection against the virus.

Outline

1. Introduction
2. Literature review
3. Data description
4. Methodology
5. Analytical part
6. Case study of particular countries
7. Excess deaths analysis
8. Interpretation of results
9. Conclusion
10. References / Bibliography

Core bibliography

Li, A. Y., T. C. Hannah, J. R. Durbin, N. Dreher, F. M. McAuley, N. F. Marayati, Z. Spiera, M. Ali, A. Gometz, J. Kostman, T. F. Choudhri (2020): "Multivariate analysis of factors affecting covid-19 case and death rate in u.s. counties: The significant effects of black race and temperature." medRxiv.

Toharudin, T., R. S. Pontoh, R. E. Caraka, S. Zahroh, P. Kendogo, N. Sijabat, M. D. P. Sari, P. U. Gio, M. Basyuni, B. Pardamean (2021): "National vaccination and local intervention impacts on covid-19 cases." Sustainability 13(15).

Velasco, J. M., W.-C. Tseng, C.-L. Chang (2021): "Factors affecting the cases and deaths of COVID-19 victims." Int J Environ Res Public Health 18(2).

M. Şahin (2020): "Impact of weather on COVID-19 pandemic in Turkey." Sci Total Environ

Alagoz O, Sethi AK, Patterson BW, Churpek M, Alhanaee G, Scaria E, et al. (2021): "The impact of vaccination to control COVID-19 burden in the United States: A simulation modeling approach." PLoS ONE 16(7): e0254456.

Chen, X., Huang, H., Ju, J. et al. (2022): "Impact of vaccination on the COVID-19 pandemic in U.S. states." Sci Rep 12, 1554.
https://doi.org/10.1038/s41598-022-05498-z

Amaia Calderón-Larrañaga, A., Serhiy Dekhtyar, S., Vetrano, D. L., Tom Bellander, T., Fratiglioni, L. (2020): "COVID-19: risk accumulation among biologically and socially vulnerable older populations." Ageing Research Reviews, Volume 63. https://doi.org/10.1016/j.arr.2020.101149.

WHO. (2022): "Global excess deaths associated with COVID-19, January 2020–December 2021." https://www.who.int/data/stories/global-excess-deaths-associated-with-covid-19-january-2020-december-2021

Watson, O. J. et al. (2022): "Global impact of the first year of COVID-19 vaccination: a mathematical modelling study." The Lancet Infectious Diseases, Volume 22, Issue 9, 1293 - 1302 https://doi.org/10.1016/S1473-3099(22)00320-6

Livingston, EH, Malani, PN, Creech CB. (2021): "The Johnson & Johnson Vaccine for COVID-19. "JAMA. 325(15):1575. doi:10.1001/jama.2021.2927

Kayano, T., Sasanami, M. et al. (2022): "Number of averted COVID-19 cases and deaths attributable to reduced risk in vaccinated individuals in Japan". The Lancet, Volume 28. https://doi.org/10.1016/j.lanwpc.2022.100571

Chapter 1

Introduction

Covid-19 is a term which was not known three years ago. However, now it is one of the most repeated words everyone hears about on a daily basis. SARS-CoV-2 virus, also known as Covid-19, caused the most recent pandemic that has had an impact on the population, and everybody had to get used to living with it. Consequently, there were attempts to mitigate the development of the Covid-19 pandemic and eliminate public restrictions. One of the most effective ways to terminate the pandemic and to return to the life everybody was used to living appears to be vaccination. The main goal of vaccines is to reduce the number of new confirmed cases, hospitalizations, and deaths by improving individuals' immunity. The number of new cases is reduced because there is an expectation to lessen transmissions (World Health Organization 2022). The thesis provides a complex analysis using data from numerous countries all over the globe. The main aim is to explore whether the introduction of vaccines correlated with the aforementioned effects on the macro level and estimate what changes in the number of cases, hospitalizations and deaths may be attributed to the vaccines. There were vaccination campaigns in the vast majority of countries worldwide, and thus vaccines should serve as a fundamental tool for the pandemic elimination. Public restrictions were the additional instrument used by governments to slow down the pandemic expansion. This factor is included in the form of the stringency index. Furthermore, world analysis is extended by a deeper inspection of two selected countries in order to provide additional information about vaccination. Last but not least, an analysis investigating excess deaths caused by the pandemic in the Czech Republic was created. The main goal of this analysis is to evaluate the amount of money not paid in pension till 2030. The structure of the thesis is as follows. Chapter 2 is dedicated to the review of existing literature. Chapter 3 consists of data description and utilization. Chapter 4 is devoted to the selection of methodology. Chapter 5 presents the empirical results. Chapters are further divided into world analysis and individual country analysis. Chapter 6 provides an excess deaths analysis focusing on numerous deaths caused by the pandemic. Finally, the conclusion is presented in Chapter 7.

Chapter 2

Literature Review

Ever since its outbreak, the Covid-19 pandemic has been analysed by numerous studies. This comes as no surprise since the pandemic has affected almost all individuals in multiple aspects. Understandably, many papers focus on the effects the virus itself has had on society. Nevertheless, the impact of vaccination has not yet been thoroughly analysed. Thus, it constitutes the primary focus of this thesis, along with the data analysis of factors that could have had a significant effect on the development of the pandemic. The fact that several studies examine the determinants of Covid-19 without considering vaccines might be attributed to a delay caused by a need to spend a certain amount of time on the creation of vaccines and testing before the commencement of official public vaccination. The first part of the literature review is dedicated to papers analysing the determinants of the Covid-19 pandemic before the formal start of vaccination (i.e., December 2020). Subsequently, the second part of this section is devoted to vaccination-related papers. Interestingly, several of these studies predict how many lives have been saved but do not analyse the effect of vaccination on the rate of hospitalization.

2.1 Studies not considering vaccination

Velasco *et al.* (2021) created a cross-country examination of factors that could have a significant effect on the number of Covid-19 deaths and cases. The authors collected data from 141 countries that have been hit by this disease. Data capture the period from the beginning of the outbreak until the end of 2020 and for the methodology, an OLS regression was chosen. The authors use two main models for their analysis. The first one is devoted to the total Covid-19 cases, and the second one is dedicated to the total Covid-19 deaths. Independent variables consist of total tests, the country's median age, rural population, population density, GDP per capita, average temperature, and average rainfall. The model identifies the number of tests and average temperature as significant variables positively affecting the number of deaths and cases. The vaccination data were unavailable at that time, and thus they cannot be taken into consideration. Moreover, the study compares the model results based on the population criterion, which divides countries into two categories, i.e., high-population countries whose population exceeds the limit of 10 million people and countries with a population lower than the stated limit. As a result of the model using the number of cases of Covid-19 as the dependent variable, there are some differences. For instance, the independent variable of the rural population bears significance only for the low-population areas, whereas the temperature is essential merely for high-population areas. Results with Covid-19 deaths as the dependent variable show the importance of the rural population, and the median age population density only for low-population areas. There is quite a perceptible distinction of the R-squared between the models. The model containing low and high population areas differ by 0.206 for the cases regression and 0.253 for the death regression. Unfortunately, no table or list containing the countries included in the dataset is provided. Nevertheless, it is essential to mention some issues identified in this study that may weaken the trustworthiness of the results. The problem is related to the data structure, which has been gathered from many different countries, and therefore it is a cross-sectional study. In that situation, it is necessary to pay attention to the potential presence of unobserved heterogeneity that was not taken into consideration. Each country has its specific factors, such as different policies in the form of restrictions that were enacted as direct and fast protection against the development of the pandemic. The study does not address this problem and thus brings endogeneity into the model. Due to this issue, the OLS model appears to be inconsistent. The analysis should be redone using an appropriate panel data methodology.

Li *et al.* (2020) was examining the Covid-19 pandemic in U.S. counties. The dataset covers the period from the pandemic's beginning to the 14^{th} of April 2020. They defined positive cases and deaths as dependent variables. Since they were investigating the first several months of Covid-19 presence, they set restrictions that omit counties with less than 50 cases per 100 000 people or counties where the first case appeared in the last 3 weeks before the end of the dataset period. These restrictions were used for the model considering positive cases as the dependent variable. Within the observed period, the pandemic was not present in most of the included counties - from a total of 3143 counties, only 661 were left in the dataset. There were restrictions for the death analysis that excluded counties with less than 10 deaths per 100 000 people or the first death appearing in the last 2 weeks before the end of the examining period. As a result, only 217 counties were included. In the paper, two models were used. The authors created a sequential regression for examining the effect of temperature and race on confirmed positive cases and deaths. Furthermore, they used logistic regression to investigate differences between the lowest and the highest quartiles for positive cases. This method cannot be applied to death analysis because of an inadequate number of observations, and therefore they used linear regression. The sequential model consists of 4 step models where all variables from the previous model are added to the next step. Model 1 was created as univariate. Model 2 added population density, GDP per capita, tests per 100 000 people, and several others. Model 3 continues with demographic variables such as the structure of the population, proportion of African American community, and female/male percentage of the population. Model 4 finishes with comorbidities variables such as obesity, diabetes, etc. The authors pointed out that African Americans were associated with higher mortality and more positive cases. Nevertheless, it is not mentioned that the authors could assign every positive case to a particular race. Thus, they can only say that counties with a higher percentage of African Americans have the aforementioned effects. Moreover, the study suggests that the sufficiency of vitamin D could reduce the effect of Covid-19. The claim is supported by evidence from their models mentioned hereinabove which prove that higher temperature is associated with fewer Covid-19 cases. However, it has been found that it does not affect deaths. The variable capturing temperature was calculated as a mean of temperatures between the last day of the dataset and the temperature 10 days prior to the confirmation of the first case in a particular county. Thus, it does not reflect the variability of weather over the period.

Şahin (2020) investigates whether the weather affects the quantity of new Covid-19 cases. The paper aims to determine the relationship between weather variables and the Covid-19 pandemic. Şahin focuses on the situation in Turkey from the first announced positive case till the 5th of April 2020. The dataset captures nine Turkish cities with the highest concentration of confirmed cases of Covid-19. There are more variables providing weather information included in his analysis, namely temperature, dew point, humidity, and wind speed. All of them were provided in daily measures. Moreover, the population size was taken into consideration. The study includes Spearman's correlation test to detect potential relations. Each variable was tested at the time of Covid-19 confirmation and then 3, 7, and 14 days before it. As a result, the highest positive correlation was identified for more populated areas, meaning that more crowded areas are connected to greater Covid-19 concentration. The author finds that temperature is negatively related to the number of new cases. In order to investigate the seasonality effect, it would be beneficial to repeat the study and analyse recent data, given that Covid-19 has been present for more than 3 years.

2.2 Studies analysing vaccination

Toharudin *et al.* (2021) were investigating whether intensive public restrictions had a reducing effect on the Covid-19 pandemic in Jakarta. Moreover, they inspect the impact of vaccination on the development of the pandemic. The study uses data from the 1st of March 2020 to the 18th of March 2021. However, only less than 5 million Indonesians had received a dose of vaccine by the end of the observed period, thus only a low percentage (0.7 %) of the Indonesian population was vaccinated. The dataset includes a daily updated number of new cases, recovered cases, and the number of deaths caused by Covid-19 in the capital of Indonesia, Jakarta. The city was chosen because it represents the area that was affected the most. Their analysis consists of two different types of models. The first one uses the Bayesian structural time series models with previously mentioned variables. The result of restrictions is correlated with an almost immediate and significant reduction in the number of new cases. The authors presented the effect of vaccination as desired. Still, it is a vague argument because although they first assign the downward trend of new Covid-19 cases to the vaccination effect, they subsequently explain that the number of new cases is decreasing because of the restrictions. Furthermore, it is highly speculative to argue that 0.7 % of vaccinated people could reduce the pandemic, and therefore they have a lack of vaccination data to conclude about the effect of vaccines. The authors recommend further examination that will include a more extended period of time. An artificial neural network model is used in the second modelling section. The target of this section is to forecast the development of the Covid-19 pandemic in the next 7 days. The nnetar: Neural

Network Time Series Forecasts function in the R program was used, and as a result, the confidence interval of expected numbers of particular variables was presented. In conclusion, it is mentioned that the forecast was accurate only for the next day due to the data fluctuation. The number of tests undertaken is always higher during the working days and there is a perceptible drop at weekends. Moreover, the number of tests is not the same every day because it is mainly determined by the number of people that come to test themselves. It would be beneficial to redo the analysis with the current data.
Chen *et al.* (2022) tried to measure the effect of vaccination on Covid-19 development in the USA. They collected data from October 2020 to the 7th of March 2021 and merged them into a weekly aggregation. Website "Our World in Data" was selected as a source of data and is further described in Chapter 3: Data Description because the thesis is working with this source. However, vaccination started in December 2021, thus there were not many people vaccinated in the observed period. Vaccination rate which contains people vaccinated with at least one dose and people fully vaccinated followed by the number of tests, hospitalizations, and total cases of Covid-19 were collected. Dependent variables – Covid-19 cases and hospitalizations were transformed into a one-week lag. This thesis transforms independent variables into lags because more lags with different delays are incorporated into models. They also added policies restricting social contact, temperature, and snow depth. This section of the analysis is modelled using the OLS regressions with the total cases and hospitalizations as dependent variables. Vaccination reduced the growth of both dependent variables. They do not comment on any other mentioned variables' results. The authors ran the tests even with the daily observations and ended up with similar results. Furthermore, logarithms were used for analysing variables and its result also led to a reduction in the pandemic. Investigation of the variables' distribution is provided in this thesis by histograms and logarithmic transformation is used as a tool for skewness correction. Another part of the study is dedicated to herd immunity with the Susceptible-Infected-Recovered model which should predict when herd immunity will be achieved based on the average vaccination level during a certain period. The model predicts that this immunity should be reached during July 2021 at a level of 60.2 % vaccinated population in the USA. They modified this model by the different intensities of vaccination, vaccine effectiveness, and vaccine hesitancy which refers to people who for various reasons refuse vaccination. As a result, they found out that vaccination has a significant effect whose marginal impact on the number of

hospitalizations and total cases in the USA decrease as the vaccination rate goes up. Herd immunity which can be achieved in the population is predicted based on a specific model. However, the authors mentioned the distinction of results as many factors are difficult to forecast. There is a problem of missing data that is commented on in the study. Missing data on recovery and mortality were present in 22 states and they fixed it by filling in the median value computed from the rest of the 29 states.

Kayano *et al.* (2022) examined the effect of vaccination against Covid-19 in Japan. The dataset consists of total positive Covid-19 cases, deaths, and vaccination that included information about age (six age groups) and sex. They collected data from the beginning of March to the end of November 2021. Almost 75 % of people were vaccinated in Japan at the end of November 2021. The dataset also included information about the number of doses for individuals. Fortunately, Japanese public websites contain information about the vaccination status of positively tested people, and thus it is easier to detect whether an individual has at least one dose of vaccine. They aim to model how many cases and deaths were spared as a result of the vaccination program. They calculated daily incidence for unvaccinated people and people with at least one dose further divided based on age group and sex. Moreover, 95 % confidence intervals were computed using bootstrapping with 1000 repetitions. As a result of their study, they claim a reduction of nearly 30 % of total cases and almost 70 % of deaths. Unfortunately, restrictions are not included in the modelling, although it is mentioned that this variable has a crucial effect on the number of prevented cases and spared deaths. They also claim that the highest number of averted cases and deaths are apparent for the age group above 65 years. Although the most averted deaths are for the oldest, it is still the most affected group of people. Surprisingly, even though they found out that vaccination prevents more females from being positively tested, females also seem to have a higher number of deaths.

Watson *et al.* (2022) created a study analysing the number of averted deaths as a result of vaccination. Their dataset contains 185 countries and focuses on the first year of vaccination campaigns, i.e., the year 2021. They used the transmission model to predict how the Covid-19 pandemic would have developed without vaccination. They defined a population-based SEIRS model with age-structured patterns and used not only the official numbers of deaths caused by Covid-19 but also examined the excess mortality in every country. The authors discover that the official number of deaths caused by Covid-19 is underestimating the actual extent of the pandemic. There might be many factors that caused death, although the individuals were not positively tested, and thus they are not officially recorded in Covid-19 deaths. For instance, restrictions led to solitude and promoted fear in people with health issues that even worsened their condition. The main part of their thesis aimed to predict the number of averted deaths due to vaccination. They used Metropolis-Hastings Markov Chain Monte Carlo-based sampling scheme and used the output for reproductive number estimation (represented by letter R). Ethnicity was not taken into consideration. They had a problem with excess deaths data availability, and thus they needed to use model-based estimates instead which can represent a bias in their data and consequently in the number of averted deaths. The most challenging part is associated with the prediction of deaths without vaccination. The authors took 100 repetitions of estimated R distribution and vaccine efficacy for every country and created a situation of no vaccines in the analysing period, therefore the development of the R trend is the same as before vaccination. Moreover, they calculated the total number of averted cases as the official quantity of deaths subtracted by the number of deaths based on the R trend prediction in the situation when the vaccine is not existing. The authors did not adjust the R trend without vaccines because it is complicated to predict how governments and population would have behaved. Furthermore, they predict the trend of the pandemic based on historical data that do not precisely capture the exact future development. Moreover, based on the previous comments, there might be a bias in the total averted deaths. Their analysis concludes that 14.4 million people were saved based on the official reports of deaths, and even 19.8 million in the case of excess deaths reports.

Rustagi *et al.* (2021) focus on analysing the effect of vaccination in Asian countries. They collected data for a total of 48 countries from the 24$^{\text{th}}$ of February 2020 to the 26$^{\text{th}}$ of September 2021. The data consist of variables as follows: total cases, total deaths, and amount of vaccination doses on a daily updated basis. The authors grouped daily observations into monthly units which are usable, and the same aggregation will be used in this thesis. They decided to use 3 different methodologies in their study. They conducted linear regression, polynomial regression, and support vector machine approach. Each of these methodologies consists of 4 models, running total cases against the number of people with 1 dose of vaccine and then total cases against people with 2 doses of vaccine. Two additional models were constructed by replacing total cases with total deaths and using the same independent variables. OLS models were defined as univariate resulting in quite a low R-squared (from 0.04 to almost 0.21). The OLS regression does not provide sufficient outputs, and therefore as a next step, polynomial regression was defined using up to a quartic degree. Polynomial regression should explain more profound interconnections of variables. As a result, they claim that there is a decrease in total deaths and cases after the first dose of vaccine, however with one additional dose, the reduction is even more significant, up to 75 %. The support vector machine algorithm could be understood as an advanced method of OLS and polynomial regression. This method is used for making predictions and projecting the accuracy of a model. For all models, a square root mean error was calculated and it is used as a measure of the accuracy of predictions. The support vector machine is considered the best model as it shows the smallest square root mean error, thus guarantees the highest accuracy of results. Unfortunately, they did not perform any model that would include both the first and the second dose of vaccine into one model to show how much variability it could detect.

Jain *et al.* (2022) focused their study on analysing the effect of Covid-19 vaccination against the Omicron variant of the virus. They examined the impact across 32 European countries from the 13$^{\text{th}}$ of October 2021 to the 1$^{\text{st}}$ of January 2022. The authors chose the longitudinal fixed effects Poisson regression model. Fixed effects were selected in order to deal with unobserved effects that are different for each country but stable over time. Their dataset was based on daily updated numbers of new Covid-19 cases per one million people. Moreover, vaccine coverage was included, as well as the stringency index which combines 9 indicators, such as home restrictions, school and workplace closures, etc. Firstly, they investigated the impact of full vaccination on total cases per million people. Furthermore, vaccination and the stringency index were added in two lags, each representing one week. 14-day lag is incorporated into Country Level Analysis, while a one-month delay is tested in World Analysis. They found a highly significant result that a percentage increase in full vaccination implies a decrease in total cases by almost 17 %. Total cases decrease by roughly 2 % with one-unit improvement in the stringency index. One of the study's limitations is the focus on European countries which cannot describe the whole world effect of vaccination on the Omicron variant. There might be a bias based on the official numbers of detected Covid-19 cases which may not capture all people who got infected by Covid-19 because they might not have had symptoms, and thus did not test themselves. Another model was defined to capture the effect of the so-called booster vaccination which is an additional dose of vaccine. This vaccination dose was mainly focused on the elderly to provide them with extra protection against the harmful effects of Covid-19. The authors argue that they did not find a significant relationship between the booster vaccine and the reduction in total cases that were examined in the second model used in the study. However, a problem might arise in a situation if the number of observations is insufficient to provide reliable results. This comes down to the fact that the authors did not mention whether their infections data were split by a variant of the virus or they had access only to the number of Omicron cases or whether they worked with the amount of all positively tested people in the period when the variant Omicron was mainly detected. Arguably, using total cases of Covid-19 for evaluation of the effect of vaccination against the Omicron variant might not accurately describe the desired effect which was intended to be investigated.

Chapter 3

Data Description

3.1 World Analysis

For the purpose of the analysis, a large dataset was downloaded from the website "Our World in Data". The dataset was created by Ritchie *et al.* (2020) and includes various variables that capture Covid-19 information. It is based on daily observations of more than 200 countries all over the globe, with the first observations collected at the beginning of January 2020. The essential variables consist of, *inter alia*, the number of new cases of Covid-19, the number of tests used for detection of the virus' presence, the number of people newly and fully vaccinated, total population that will be used for rescaling the variables, the number of deaths and hospitalizations. The number of hospitalizations as well as the number of deaths are variables related to Covid-19 hospitalizations and deaths that were diagnosed to positively tested people. Moreover, the stringency index representing the level of restriction in society is also included and will be described later.

The dataset contains a lot of missing values, therefore the daily observations were aggregated into monthly intervals in order to reduce the measurement error and to eliminate the multidimensionality of the dataset. Monthly observations also reduce the volatility of new tests variable, which is desirable given the noise in daily data. Although the dataset contains many variables scaled by a million people, i.e., new deaths or hospitalized patients per million, it was decided to rather scale variables by total population and to transform them per thousand people to ensure the same scales for all variables and avoid negligible values of variables when per million people scaling is used.

Weatherbase (2022) is an additional source of data that will be used in the study. It contains information about average monthly temperatures in more than 260 countries. As described in the literature review, several studies have considered average temperature as an essential element influencing the number of positively tested people and the total development of the Covid-19 pandemic. However, the drawbacks in each study were pointed out, and thus a final statement about the significance of this variable is desired. Consequently, average temperature was incorporated into models as a control variable. The scraping method was chosen as an appropriate tool for obtaining the values. The Weatherbase website contains historical average temperature values for several cities in every country, and therefore monthly temperatures in every country were calculated as the average of all cities within a country for each month. Finally, the downloaded dataset was merged with the previously mentioned data obtained from the website "Our World in Data".

The main goal of the first part of the thesis is to examine the effect of vaccination on the number of deaths, hospitalizations, and new cases of Covid-19 using data from numerous countries all over the globe. All three mentioned variables are transformed into per thousand people format and will be used as the dependent variables in models. Unfortunately, the data do not include the number of new hospitalizations. The dataset provides only the number of people in hospitals on a given day, and therefore the increase of 100 people in hospitals does not mean that only a hundred new people were admitted thereto. Many people could have been discharged from hospitals, therefore the actual number of new inpatients might be higher. Thus, the variable incorporated into the models was computed as an average of hospitalized people in a particular month. The most important independent variables are the number of newly vaccinated people per thousand in a specific month and the number of fully vaccinated per thousand inhabitants. These two variables were created from the total number of people vaccinated and people fully vaccinated. The goal of vaccination is to reduce the quantity of all three aforementioned dependent variables by increasing individuals' immunity, and thus the expectation is to find out negative and significant relation between vaccination and dependent variables. Nevertheless, the analysis would clarify whether even the first dose of vaccine reduces the dependent variables or if people need to be fully vaccinated to observe a reduction in the development of Covid-19.

Three variables will be incorporated into the models as control variables to ensure the results' robustness. Şahin (2020) and Li *et al.* (2020) consider temperature as a significant element influencing the development of the Covid-19 pandemic. Therefore, average temperature was added to the model with the expectation of reducing the pandemic's development in all models. Furthermore, the number of new tests per thousand people was added as an additional control variable and it is anticipated to be positively related because the confirmation of Covid-19 is proved only by a positive test. Naturally, if no testing occurs, the presence of the virus cannot be revealed. Last but not least, the stringency index is considered an important factor affecting the spread of the Covid-19 pandemic. This index can be described as a measurement of the strictness of government policies. It is a variable that contains nine different measures that influence the development of Covid-19, specifically school closures, cancellation of public events, workplace closures, closure of public transport, restrictions on public gatherings, stay-at-home requirements, public information campaigns, international travel controls, and restrictions on internal movements, as mentioned by Roser (2021). A certain score is assigned to each of these elements and then the mean score of all metrics is calculated. The value of the stringency index varies within the range of 0 to 100, where 100 represents the strictest restrictions.

Most of the variables were reported as a daily new quantity of tests or utilized vaccines and its aggregation into monthly intervals was simple. The stringency index was provided for every day and the transformation was done by computing an average of all days within a month for each country. Since the index range has a limited interval, distortion by the aggregation is not expected. However, the stringency index is adjusted based on the development of the pandemic, mainly by the number of new cases and its trend which demonstrates the pandemics' expansion, and therefore it needs to be incorporated into the model in a lag that should indicate the effect of public restrictions.

The main target is to analyse the effect of vaccination on the development of the Covid-19 pandemic, however, the dataset used for the examination contains missing observations for the vaccination data due to the unavailability of the vaccines. One of the causes of the missing data is that the official start of vaccination took place in December 2020, as a result of which all earlier observations were disregarded. It is expected that the vaccine does not increase immunity immediately after its implementation, and therefore lags in observations might provide a significant result. The data after this date also contained some missing values, however, for the purpose of this part, it is assumed that they are distributed randomly, and thus are left unchanged. If the assumption is incorrect and the data are unavailable due to a specific characteristic at a country level, it will be resolved in the analysis. Unfortunately, the number of new tests provided is not collected after the end of June 2022, and thus the final dataset captures the period from October 2020 to June 2022.

Data pre-processing reveals many missing observations for the hospitalization variable. The reason is that only 31 countries have meaningful hospitalization data. As a result, there will be two different datasets in the analysis. One will capture all mentioned data and the second one will exclude the hospitalization variable to encompass data from more countries. Table A.1 summarizes all countries that will be used for the hospitalization model and the majority of nations are located in Europe which is caused mainly by the quality of data. On the other hand, Table A.2 contains a list of all countries in the dataset that shows 124 countries that remain in the dataset after hospitalization exclusion. There are countries from Europe, South America, Asia, and Africa that provide enough diversity in data.

Table 3.1 placed below provides the descriptive statistics for the most essential variables used in World Analysis. Based on the summary of the stringency index, it is visible that in society overall, during the Covid-19 pandemic, the restrictions were quite strict. Average temperature values lie between -20.8 degrees of Celsius and a maximum of 36.1 degrees, thus the range is extensive and might provide a significant result. There were, on average, almost 7 new cases per thousand people in each country every month. Nevertheless, the median per thousand people is much lower. According to the difference between the mean and median, the expectation is to observe a lot of outliers with high numbers which is confirmed firstly by the histograms and further by the maximum value. New deaths, tests, and vaccinations show the same relation between mean and median, and thus they all need to be corrected.

Regarding hospitalizations, all other statistics apart from the maximum are below 1, thus it is apparent that only a small portion of people who were positively tested needed to visit a hospital. The lowest values of descriptive statistics measures are connected to new deaths. Figure B.1, Figure B.2, and Figure B.3 located in Appendix B show histograms that present distributions of relevant variables that will be included in modelling. All the histograms except the stringency index and average temperature are highly skewed to the right, and therefore a logarithmic transformation is implemented as an appropriate and essential tool because it can be used to adjust the skewed distribution in order to provide distribution closer to the Gaussian normal distribution.

Logarithms reduce the weight given to extreme values. Furthermore, it was necessary to add a small constant to all observations with a minimum equal to zero which were highly skewed and transformation into a logarithm was undertaken. The constant had to be added to avoid unfeasible outcomes due to computation logarithms of zero.

Table 3.1: Descriptive statistics - world data

	Mean	St. Dev.	Min	Median	Max
Stringency Index	45.979	19.601	0.000	45.644	94.800
Average Temperature	19.262	9.255	−20.801	22.997	36.143
New Cases	6.936	18.684	0.000	1.000	397.264
New Vaccinated	36.011	51.986	0.000	16.012	784.484
New Vaccinated Fully	35.959	51.266	0.000	15.434	572.746
New Tests	102.071	250.450	0.002	30.626	3,731.906
New Deaths	0.051	0.099	0.000	0.010	2.142
Hospitalizations	0.192	0.195	0.0005	0.124	1.369

Note: All variables apart from Stringency Index and Average Temperature are in per thousand people terms

Moreover, a correlation matrix was created in Table A.3 located in Appendix A in order to investigate the bivariate relationships among variables. All variables were included in one table, although some variables will be used as dependent variables, and thus their correlation is not a problem. This is the case for hospitalization and new deaths. Unsurprisingly, the correlation between the number of new cases and the number of new tests is relatively high which is caused by the nature of the variables. One cannot be considered a person with Covid-19 until they receive a positive test result. Similar relation can be seen between new cases and new deaths compared to hospitalization, which is caused by the nature of the variables. People who are captured in new deaths are those who had to be considered as Covid-19 positive, therefore it is a portion of new cases variable. Similarly, hospitalization captures the amount of hospitalized people who are also included in new cases variable. Variables capturing the vaccination are correlated, which is nevertheless caused by the fact that one cannot be fully vaccinated unless they have the first dose of the vaccine. Moreover, vaccination variables are not significantly correlated with any dependent variable, which is not surprising because vaccination is not expected to have an immediate effect but it will increase individuals' immunity in a short time period. According to the official information (WHO Collaborating Centre for Vaccine Safety 2021), it is expected that it takes at least 2 weeks to build immunity after the vaccination. The relation between average temperature and dependent variables is negative, as expected. There is not a high pairwise correlation among independent variables and thus multicollinearity should not be an issue in the analysis.

3.2 Country Level Analysis

The second part of the thesis is dedicated to a specific country investigation and its comparison to World Analysis examined in the first part. Two countries will be discussed. The selection of the first particular country was based on the subjective choice and with continuity to an excess deaths analysis that will be studied in the third section discussed later in the thesis, and thus the Czech Republic was chosen. Moreover, a country with a fast vaccination campaign, i.e., Israel, will be selected because such an investigation could provide different results to which extent vaccination reduced the development of the Covid-19 pandemic. The situation in the country with quickly vaccinated people is expected to be much more favourable compared to World Analysis. Evidence of the fast vaccination campaign in Israel is presented in Figure B.4 located in Appendix B. It captures the comparison of vaccination progress between the Czech Republic and Israel. The graph indicates the evolution of people vaccinated by at least one dose of vaccine (black line) as well as fully vaccinated people (red line). The datasets for the separated analyses of the Czech Republic and Israel will be filtered from World Analysis to ensure the comparability of the results. Using the aggregated monthly intervals for the specific country investigation would not be efficient because in that case only 18 observations would be examined, therefore daily observations will be more beneficial.

The restriction for the data period is done in a similar way as already mentioned in World Analysis. Furthermore, since average temperature data were collected on a monthly basis, it was necessary to find a different source of data. Website National Oceanic and Atmospheric Administration (2023) includes weather information. Based on the country selection, observations were downloaded for the Czech Republic and Israel. More specifically, Klementinum weather measurements were selected for the Czech Republic and Jerusalem center for Israel. The dataset includes many weather variables, however only temperature was filtered. There are multiple observations for every day, therefore daily averages were created. Datasets contain several unclear measurements that distort the values, and thus these observations were omitted. Finally, average temperature datasets were merged with daily observations' datasets based on the date specification. Investigation of vaccination's effect has to be based on lags because vaccination needs some time to enhance individuals' immunity fully, thus lags of 14, and 30 days were created for new and fully vaccinated variables. Two weeks and one-month delay were chosen according to WHO Collaborating Centre for Vaccine Safety (2021) which set two weeks as the minimum time that a vaccine needs to build immunity. A month and two months of delays for vaccination variables were created in World Analysis. Furthermore, lags for the stringency index with delays of 14 and 30 days are also made because an increase in the index needs some time to exhibit its effect.

Table 3.2 and Table 3.3 below provide descriptive statistics for both investigated countries. The stringency index does not differ among tables, only the median value is slightly higher in Israel. New cases statistics do not show different values except the maximum, which is almost five times larger in Israel's case, however it might be caused by more tests provided which is confirmed by the median of new tests. Both variables considering vaccination show similar values in the tables with an exception in maximum that is two times smaller in the Czechia table. The mortality rate presented by new deaths variable is greater in the Czech Republic. Based on the mean value, there are five times more deaths in Czechia. The maximum value differs by 0.024 per thousand people. Based on the statistics from the Worldometer (2023) website, the official number of deaths caused by Covid-19 in the Czech Republic was 42 149 till the end of 2022. According to the dataset value, Czechia's population equals 10 493 990, whereas the population of Israel is approximately 9 449 000. However, Israel reported 12 037 deaths caused by Covid-19 till the end of 2022. Hospitalizations per thousand people are at least two times smaller in Israel which comes hand in hand with evidence of deaths. People that need to be hospitalized are expected to have more severe conditions and the percentage of deaths is increasing. Last but not least, average temperature shows more stable temperature conditions in Israel, whereas Czechia temperature has a wider range with negative values during winter, therefore average temperature might be a more important factor for the analysis of the Czech Republic.

Table 3.2: Descriptive statistics - Czechia

Statistic	Mean	St. Dev.	Min	Median	Max
Stringency Index	46.106	20.081	14.810	42.260	81.480
New Cases	0.575	0.787	0.000	0.251	5.468
New Vaccinated	1.222	1.703	0.0001	0.513	8.008
New Fully Vaccinated	1.212	1.847	0.000	0.495	8.677
New Tests	8.741	7.542	0.115	6.822	40.702
New Deaths	0.005	0.006	0.000	0.003	0.038
Hospitalizations	0.268	0.260	0.002	0.196	0.910
Average Temperature	9.580	7.602	−9.210	8.019	29.129

Note: All variables apart from Stringency Index and Average Temperature are in per thousand people terms

Table 3.3: Descriptive statistics - Israel

Statistic	Mean	St. Dev.	Min	Median	Max
Stringency Index	45.424	19.134	14.810	48.965	87.040
New Cases	0.716	1.858	0.000	0.260	25.748
New Vaccinated	1.290	2.650	0.000	0.302	16.783
New Fully Vaccinated	1.181	2.574	0.000	0.239	17.405
New Tests	8.466	7.379	0.572	7.116	47.369
New Deaths	0.001	0.002	0.000	0.001	0.014
Hospitalizations	0.101	0.084	0.007	0.078	0.393
Average Temperature	16.928	6.803	1.619	17.343	30.934

Note: All variables apart from Stringency Index and Average Temperature are in per thousand people terms

There are several figures containing histograms of variables using Czechia and Israeli data (see Figure B.5, Figure B.6, Figure B.7, Figure B.8, Figure B.9, and Figure B.10). Like in World Analysis, the distribution of variables is investigated in order to eliminate skewness which is corrected by logarithmic transformation. The stringency index and average temperature do not need to be transformed into logarithms in either country. All other variables were highly skewed and thus transformation was necessary with one exception, i.e., the number of tests per thousand in Israel that was left without correction.

Separated data of Israel and Czechia contain only several unavailable observations in the dataset and thus it is expected that they are randomly distributed, therefore no adjustments are required to correct the missing data.

Figure B.11 and Figure B.12 in Appendix B were created for comparison of hospitalizations and new cases with average temperature in time. It is visible that in the case of Czechia a decreasing trend of the dependent variable is associated with increasing values of average temperature and vice versa. It responds to the changes of seasons which are evident in the geographic location of Czechia. The mentioned trend is also noticeable in Israel's graphs, however an exception is present in the second half of 2021. There is an increase in the number of cases as well as hospitalizations, although average temperature is reaching the maximum of its values. Figure B.13 and Figure B.14 in Appendix B provide a comparison of population vaccination percentage against new cases and hospitalizations. It is more difficult to distinguish the effect of vaccination because its primary purpose is to lessen the development of the variables, however the peak in 2022 is apparent in both countries regardless of the percentage of vaccinated people. It could be caused by a new variant of Covid-19 and the number of new cases in a situation without the vaccination campaign cannot be predicted.

Chapter 4

Methodology

4.1 World Analysis

The main aim of the thesis is to investigate the impact of the number of newly and fully vaccinated people on dependent variables (number of new cases, quantity of deaths, and hospitalization). The analysis deals with a cross-country dataset consisting of several time periods, and therefore is working with panel data. Consequently, including many different countries' observations leads to heterogeneity. It is necessary to consider unobserved heterogeneity caused by culture, education, geographical position, and there might be a lot of other factors that could influence the dependent variables and possibly be correlated with independent variables. It is essential to use a methodology that can deal with unobserved heterogeneity and is suitable for panel data. Random effects and fixed effects are estimators that are usable in this situation. The Hausman test is created to distinguish which of these estimators is better. The null hypothesis definition is that both methods are consistent, and the rejection means that only the within estimator is consistent. Table A.4, Table A.5, and Table A.6 contain several tests that are decisive for the selection of appropriate methodology. Each table is related to one of the three models according to the dependent variables. Hausman test results rejected the null hypothesis in all these tables and showed that only the fixed effects are consistent, therefore the results also reveal evidence for unobserved heterogeneity in a static model. The static model refers to a model where the lag of the dependent variable is not included.

Moreover, it is essential to define tests for the presence of serial correlation in the static models. Wooldridge test for AR(1) errors in FE panel models and

Wooldridge first-difference-based test for AR(1) errors in first differenced panel models were used as tests. The latter should be considered as superior in the case of serially correlated errors, although the serial correlation is not always differenced away. Table A.4 and Table A.5 show that both tests rejected the null hypotheses of no serial correlation, and therefore the first-difference estimator is considered as a baseline. Moreover, the first-difference regression is also preferred for the hospitalization model because the null hypothesis cannot be rejected for the Wooldridge first difference test.

Breusch-Pagan tests are provided in the tables to verify the assumption of homoskedasticity. Although the tests result in the rejection of the null hypothesis of homoskedasticity and serial correlation is present, there is a tool that enables their restoration called heteroskedasticity and autocorrelation consistent (further as HAC) robust standard errors.

It is necessary to consider that the dependent variables might depend on past values. For instance, new cases refer to transmission problems and thus it is expected that past values are essential for the present value of the variable. Hospitalizations and the number of new deaths are not that simple to distinguish. Hospitalization refers to severe conditions of Covid-19 which are based on the quality of individuals' immunity and might not influence others. A similar relation is intuitively expected for the number of deaths. Nevertheless, performing tests answers the intuition. If the dependence on earlier observation is present, then the static model may have an issue with omitted variable bias which is problematic.

The incorporation of lagged dependent variables leads to movement from static models to the area of dynamic panel data estimators. Implementing these lags results in the inconsistency of classical panel data models, such as the first difference or within estimator due to the Nickell bias (Nickell 1981). Instrumentation could be used as a tool to retrieve consistency, nevertheless finding external instruments that satisfy conditions of exogeneity and validity is nearly impossible.

Arellano & Bond (1991) find a solution to this problem in the form of the Difference Generalized Method of Moments estimator (further used as Difference GMM). The model works as follows. Firstly, data transformation using the first differencing is done to remove the unobserved heterogeneity. In the next step, internal instruments are utilized and the adjusted equation is estimated using GMM. Instruments are included in the model in the form of lagged levels of variables. Under the no serial correlation assumption these instruments are exogenous and the level of moment conditions are then derived based on the instrument exogeneity assumption. After all, the endogeneity caused by the lagged dependent variable and any other endogenous variables is solved. Difference GMM estimator is suitable for "short" panel data and in the presence of fixed effects according to Roodman (2009a) which is satisfied in this analysis. Furthermore, the two-step GMM corrects the heteroskedasticity and ensures robustness. In case that the dependent variable is highly persistent, there might be a problem with the Difference GMM estimator which suffers from weak instruments. However, Arellano & Bover (1995) and Blundell & Bond (1998) developed an additional method called the System Generalized Method of Moments estimator (further as System GMM). This model is built on the Difference GMM and brings more moment conditions. It introduces an estimation of the equation instrumented by lagged first differences. These moment conditions should solve the issue of weak instruments.

In order to validate internal instruments, the Arrelano-Bond test and the Hansen-Sargan test are defined for models. The first one mentioned measures the serial correlation in the differenced errors, whereas the latter is used to test overidentifying restrictions. The Arellano-Bond test has a specification that the level of the serial correlation test depends on the number of lags created. The third order of the serial correlation is tested when two lags of the dependent variable are included. The number of lags is increasing till (p-1) order when the null hypothesis of no serial correlation testing p-th order cannot be rejected. Moreover, instrumental proliferation mentioned by Roodman (2009b) needs to be considered and the number of instruments has to be supervised because increasing the number of instruments raises the p-value of the Hansen-Sargan test which should not be high. Finally, the Hausman test is provided to determine whether variables are influenced by measurement error.

The Difference GMM estimator is therefore formulated for each model and is considered superior to the first difference model if the lagged dependent variable is significant. The last test (Arrelano-Bond) in Table A.4 refers to the dynamic model with one lag of new cases dependent variable and tests for the second-order serial correlation. This test results in the rejection of no serial correlation in the new cases model, and therefore the dependent variable needs to be included in the second lag. The results of this particular model are presented in Table 5.1. The Difference GMM estimator was formulated using various specifications. At the beginning, it is necessary to verify the exogeneity of instruments using the p-value of the Hansen-Sargan test which is equal to 0.29. As a result, it is not possible to reject the null hypothesis of instrument exogeneity and at the same time, the p-value is not high enough to have instrument proliferation. The null hypothesis of no serial correlation in errors cannot be rejected using the Arellano-Bond test, thus it also confirms the exogeneity of instruments and the appropriate number of lags. The System GMM estimator was also defined when the Difference GMM satisfies the required conditions. However, after the System GMM formulation, the first lag of the dependent variable was not persistent. Therefore, the System GMM was not the best method for the situation. Persistency is expected when the coefficient of the lagged dependent variable is close to one, therefore the Difference GMM is considered to be the best method for new cases as the dependent variable.

Moving to the dynamic model and using the Difference GMM estimator does not provide the expected results for the model of the new death. The estimation results in Table 5.2 show that the Hansen-Sargan test has a p-value equal to 0.15 which is enough not to reject instrument exogeneity. Instrument proliferation is unaffected since the p-value is not close to 1. The Difference GMM estimator included in the third column of the table uses only the dependent variable as an instrument. Moreover, the Arellano-Bond test cannot reject the null hypothesis of no serial correlation in the Difference GMM estimator. The problem appears after including the first lag of the dependent variable because the lag is not considered a significant variable which could bring speculation about whether the Difference GMM is preferred. In this case, the first difference estimator is considered the best.

The Difference GMM estimator for the hospitalization model was formulated with the first lag of the dependent variable as an instrument, however the Arellano-Bond test rejects the null hypothesis of no serial correlation in the Difference GMM model. Furthermore, Table 5.3 - Estimation results of hospitalization model using world data provides the estimates of the Difference GMM estimator in column number (3). There is a problem with perfect proliferation because the p-value of the Hansen-Sargan test is equal to 1, therefore the Difference GMM does not seem to be an adequate estimator for this type of model. Thus, similarly to the new deaths model, the first difference estimator is preferred.

Formulation of models is mainly driven by the target of the thesis, therefore two lags of newly and fully vaccinated are incorporated. The stringency index is included in a one-month lag. The additional lag of the stringency index was

insignificant, thus it was not included in models because it is not the primary examined variable (comments in Subsection 5.1.2). New cases variable was included in two lags because it was necessary for the new cases model creation. Moreover, average temperature and new tests were included without lags.
Models for new deaths and hospitalization are identical because it is expected that these two variables will be influenced by the same factors. The only difference compared to the first model mentioned is the inclusion of new cases variable in the present time which replaces new tests variable.
As a result, the following models were formulated:
Model for new cases variable:

$$\begin{aligned}\log(New\_cases_{it}) =& \beta_1 \log(New\_cases_{i(t-1)}) + \beta_2 \log(New\_cases_{i(t-2)}) + \\ &+ \beta_3 \log(New\_vaccinated_{i(t-1)}) + \\ &+ \beta_4 \log(New\_vaccinated_{i(t-2)}) + \\ &+ \beta_5 \log(New\_fully\_vaccinated_{i(t-1)}) + \\ &+ \beta_6 \log(New\_fully\_vaccinated_{i(t-2)}) + \\ &+ \beta_7 \log(Stringency\_index_{i(t-1)}) + \beta_8 \log(New\_tests_{it}) + \\ &+ \beta_9 Average\_temperature_{it} + Month_t + a_i + \epsilon_{it}\end{aligned}$$

where ϵ_{it} is the idiosyncratic error, a_i is the time-invariant unobserved heterogeneity, and $Month_t$ represents a vector of dummy variables for every month reflecting the time effects. As mentioned, it is expected that vaccinations affect the dependent variable with a delay because the vaccine needs some time to become effective, whereas the contemporaneous effect is assumed for average temperature and the number of tests provided. Description of variables in the two additional models is the same as in the first model mentioned.

Model for the number of new deaths is as follows:

$$
\begin{aligned}
\log(New\_deaths_{it}) = &\beta_1 \log(New\_cases_{it}) + \beta_2 \log(New\_cases_{i(t-1)}) + \\
&+ \beta_3 \log(New\_cases_{i(t-2)}) + \\
&+ \beta_4 \log(New\_vaccinated_{i(t-1)}) + \\
&+ \beta_5 \log(New\_vaccinated_{i(t-2)}) + \\
&+ \beta_6 \log(New\_fully\_vaccinated_{i(t-1)}) + \\
&+ \beta_7 \log(New\_fully\_vaccinated_{i(t-2)}) + \\
&+ \beta_8 \log(Stringency\_index_{i(t-1)}) + \\
&+ \beta_9 Average\_temperature_{it} + Month_t + a_i + \epsilon_{it}
\end{aligned}
$$

Last but not least, model for the number of hospitalizations is as follows:

$$
\begin{aligned}
\log(Hospitalization_{it}) = &\beta_1 \log(New\_cases_{it}) + \beta_2 \log(New\_cases_{i(t-1)}) + \\
&+ \beta_3 \log(New\_cases_{i(t-2)}) + \\
&+ \beta_4 \log(New\_vaccinated_{i(t-1)}) + \\
&+ \beta_5 \log(New\_vaccinated_{i(t-2)}) + \\
&+ \beta_6 \log(New\_fully\_vaccinated_{i(t-1)}) + \\
&+ \beta_7 \log(New\_fully\_vaccinated_{i(t-2)}) + \\
&+ \beta_8 \log(Stringency\_index_{i(t-1)}) + \\
&+ \beta_9 Average\_temperature_{it} + Month_t + a_i + \epsilon_{it}
\end{aligned}
$$

4.2 Country Level Analysis

The aforementioned methodology is appropriate for the first part of the thesis, which works with panel data. However, in the second part, only data for a particular country are selected and therefore panel data estimators are no longer applicable. In this section, a time series data structure is present because observations are collected at daily intervals for only one country and therefore less complex method could be chosen for examination. The Ordinary Least Squared (further as OLS) estimator is an appropriate model for the research question. OLS is feasible because linearity in parameters and random sampling is expected. OLS has more assumptions whose satisfaction is required to become the unbiased estimator. One of the assumptions is homoskedasticity which refers to the constant variance of errors and its violation called heteroskedasticity could be tested by the Breusch-Pagan test. The serial correlation was also tested using the Durbin-Watson test which has the null hypothesis that the correlation of disturbances is equal to 0.

Table A.7 and Table A.8 contain results of Breusch-Pagan tests for both countries that will be investigated. All models rejected the null hypotheses, and thus the assumption of homoskedasticity is violated. The Durbin-Watson test rejected the null hypothesis of no serial correlation in all models. Although the tests resulted in the rejection of the null hypotheses of homoskedasticity and serial correlation, there are HAC robust standard errors that ensure the assumption of homoskedasticity and no serial correlation. Furthermore, a test for no perfect collinearity assumption is required as well. It states that no independent variable is a constant or perfect linear combination of the others. Variance Inflation Factor (further as VIF) is an appropriate tool for examining the presence of collinearity with a cut-off set to 10 (Wooldridge 2016), therefore variables exceeding the limit violate the assumption of no perfect collinearity. Table A.9, Table A.10, Table A.11, Table A.12, Table A.13, and Table A.14 contain results of VIF test in all models for both countries. A higher value (slightly above 8 at the maximum) is associated with the new vaccinated variable in the lag of 30 days which is caused by the nature of the connection between the lags of the same variable. The only variable exceeding the limit of 10 is the stringency index, however its value is not greater than 13 in any of the models. It also corresponds to the nature of the variable that the lag of 14 days and the lag of 30 days are closely connected, therefore no adjustments are required.

Similarly to World Analysis, three models were formulated. Since the data are collected on a daily basis, it was necessary to transform lags for different time periods. Delays were created for seven days, two weeks, and a month. New cases used 7-day and 14-day lags because it is expected to influence dependent variables in a shorter period that cannot be captured in World Analysis. WHO Collaborating Centre for Vaccine Safety (2021) mentions that 14 days are necessary for the organism to build immunity against the virus, therefore 14-day and also 30-day delay are used for vaccination variables. The stringency index is included in the same lags as vaccination because it needs at least the same time to show the expected result of this variable. Many variants of models were constructed and the final versions are provided on the next page.

$$\begin{aligned}\log(New\_cases_t) =& \beta_1 \log(New\_tests_t) + \beta_2 \log(New\_vaccinated_{(t-14)})+ \\ &+ \beta_3 \log(New\_vaccinated_{(t-30)})+ \\ &+ \beta_4 \log(New\_fully\_vaccinated_{(t-14)})+ \\ &+ \beta_5 \log(New\_fully\_vaccinated_{(t-30)})+ \\ &+ \beta_6 \log(Stringency\_index_{(t-14)})+ \\ &+ \beta_7 \log(Stringency\_index_{(t-30)})+ \\ &+ \beta_8 Average\_temperature_t + \epsilon_t\end{aligned}$$

$$\begin{aligned}\log(New\_deaths_t) =& \beta_1 \log(New\_cases_t) + \beta_2 \log(New\_cases_{t-7})+ \\ &+ \beta_3 \log(New\_cases_{t-14})+ \\ &+ \beta_4 \log(New\_vaccinated_{(t-14)})+ \\ &+ \beta_5 \log(New\_vaccinated_{(t-30)})+ \\ &+ \beta_6 \log(New\_fully\_vaccinated_{(t-14)})+ \\ &+ \beta_7 \log(New\_fully\_vaccinated_{(t-30)})+ \\ &+ \beta_8 \log(Stringency\_index_{(t-14)})+ \\ &+ \beta_9 \log(Stringency\_index_{(t-30)})+ \\ &+ \beta_{10} Average\_temperature_t + \epsilon_t\end{aligned}$$

$$\begin{aligned}\log(Hospitalization_t) =& \beta_1 \log(New\_cases_t) + \beta_2 \log(New\_cases_{t-7})+ \\ &+ \beta_3 \log(New\_cases_{t-14})+ \\ &+ \beta_4 \log(New\_vaccinated_{(t-14)})+ \\ &+ \beta_5 \log(New\_vaccinated_{(t-30)})+ \\ &+ \beta_6 \log(New\_fully\_vaccinated_{(t-14)})+ \\ &+ \beta_7 \log(New\_fully\_vaccinated_{(t-30)})+ \\ &+ \beta_8 \log(Stringency\_index_{(t-14)})+ \\ &+ \beta_9 \log(Stringency\_index_{(t-30)})+ \\ &+ \beta_{10} Average\_temperature_t + \epsilon_t\end{aligned}$$

where ϵ_t represents the error term.

Chapter 5

Empirical Results

This chapter consists of three sections. The first one is dedicated to World Analysis that is further divided into subsections of individual models. Detailed description in each subsection is followed by a summary and comparison of models. The second section is devoted to individual countries and each country is divided according to the models and their summaries are also provided. Finally, all models are compared in Section 5.3.

5.1 World Analysis

According to the models formulated in Section 4.1 World Analysis, the hospitalization model independent variables are the same as in the model of new deaths. This similarity is based on the expectation that these two models deal mainly with severe cases that mostly concern people with certain comorbidities. Serious progress of Covid-19 often resulted in hospitalization. Unfortunately, the chance of death rises when a person is hospitalized which confirms the hypothesis that these two models are closely connected. On the other hand, the new cases model does not distinguish between the severity of Covid-19 progress because it focuses on the identification of Covid-19 presence.

5.1.1 New Cases Model

Several tests were formulated in Table A.4: Specification of tests in order to decide which method will be the best. The Difference GMM estimator was selected for the new cases model because the lag of the dependent variable included in the model is considered significant. Moreover, two-step GMM was formulated because it corrects the heteroskedasticity and ensures the robustness of results.

The results of this particular model are presented in Table 5.1 below. The table with results contains two columns, (1) and (2) both using the Difference GMM estimator. In the first column, two lags of the dependent variable were used and only new cases variable was used as an instrument assuming that other variables are exogenous. However, in the second column instrumentation using new cases, new tests and both vaccination variables were implemented because these variables might be subjected to measurement error. The columns were compared using the Hausman test with the null hypothesis that both models are consistent. Having both models consistent would show that instrumenting only reduced efficiency, nevertheless, the p-value of the test is negligible and the null hypothesis is therefore rejected. As a result, endogeneity appears to be present and instrumentation is necessary. To conclude, the second column is considered the most trustworthy.

Table 5.1 provides the estimation of results. The interpretation of results will be based on the second column containing instrumentation. Only the dependent variable's first lag is considered significant with a positive coefficient as was expected. It refers to the transmission problem that more positively tested people a month ago increases the current numbers.

The results reveal that a one-percentage increase in the number of newly vaccinated people per thousand in the previous month should decrease the number of new cases per thousand in the current month by 0.277 %. Any other vaccination variable was not considered significant. From the author's point of view, having at least one dose of vaccine seems to be enough to decrease the transmission of the virus, and thus reduce the number of new cases. Furthermore, vaccination appears to be a fast defence against the development of new positive cases. On the other hand, being fully vaccinated is expected to have a greater effect on the number of deaths. The second lag of full vaccination variable is significant and has a reducing effect only in the model with the dependent variable instrumentation.

Unsurprisingly, new tests variable has high significance and a one-percentage increase in tests results in 1.727 % more positively detected cases. New tests substitute the effect of new cases in the present time that is used in other models, therefore its positive significance was expected. Average temperature is insignificant and has the coefficient value close to zero. The possible explanation of average temperature insignificance could be the inclusion of lags of

the dependent variable and its variation. Moreover, there are countries with stable average temperatures during the whole year, which were however also affected by the Covid-19 pandemic.
In the case of the stringency index, a log-level relation needs to be taken into account. The stringency index reflects its purpose and increasing the public restriction by one point in the previous month decreases the number of new cases by 1.1 % in the current month. To conclude, the estimated results confirm expectations.

Table 5.1: Estimation results - new cases model using world data

	Dependent variable:	
	$\log(New\_cases_{it})$	
	D-GMM	D-GMM
	(1)	(2)
$\log(New\_cases_{i(t-1)})$	0.579^{***}	0.368^{***}
	(0.089)	(0.120)
$\log(New\_cases_{i(t-2)})$	-0.250^{***}	-0.013
	(0.076)	(0.090)
$\log(New\_vaccinated_{i(t-1)})$	-0.126^{***}	-0.277^{**}
	(0.037)	(0.113)
$\log(New\_vaccinated_{i(t-2)})$	0.049	0.007
	(0.036)	(0.057)
$\log(New\_fully\_vaccinated_{i(t-1)})$	-0.043	0.073
	(0.048)	(0.130)
$\log(New\_fully\_vaccinated_{i(t-2)})$	-0.087^{**}	-0.040
	(0.036)	(0.082)
$\log(New\_tests_{it})$	0.898^{***}	1.727^{***}
	(0.133)	(0.544)
$Average\_temperature_{it}$	-0.020	0.0003
	(0.012)	(0.019)
$Stringency\_index_{i(t-1)}$	-0.007	-0.011^{**}
	(0.005)	(0.005)
No. of countries	124	124
No. of instruments	97	77
Hansen-Sargan test (p-value)	0.35	0.29
Arellano-Bond test (p-value)	0.54	0.48
Hausman test (p-value)	-	0

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses

5.1.2 New Deaths Model

This model-creating process is similar to the previous Subsection 5.1.1. Table A.5 called Specification of tests was already commented in Chapter 4. Almost all test results are the same as for the new cases model, however moving to the dynamic model and using the Difference GMM estimator does not provide the expected results because the lag of the dependent variable is not significant. Therefore, the first difference estimator is considered to be the best one. The estimation results in Table 5.2 contain three columns, (1), (2), and (3). The third column containing the Difference GMM estimator has already been commented on. The first column provides results using the first difference estimator, nevertheless it was presented that the model suffers from heteroskedasticity and autocorrelation, and therefore the robust standard errors were defined to ensure homoskedasticity and correct autocorrelation. The result of the correction is presented in the second column which is considered the most plausible. The R-squared of the new deaths model is close to 0.41 which shows that the model fits the data well. Furthermore, the Adjusted R-squared value does not differ from the R-squared which is additional evidence of the trustworthiness of the formulated model.

The coefficients provided in Table 5.2 below show that the number of current new cases as well as the number of cases in the previous month have a significant effect on the number of deaths in the current month. A one-percentage increase in the number of new cases per thousand people raises the number of deaths per thousand people by 0.743 %. One-month lag of new cases per thousand people has a smaller effect on deaths but is still positive and equal to 0.407 %. Fortunately, the number of new cases and the number of new deaths are not proportional.

Full vaccination is not considered significant for a one-month lag nor a two-month lag, however the first lag has a negative coefficient. According to the results, full vaccination does not provide additional protection for individuals against death. Being vaccinated is considered significant only in the second lag. 30 days do not seem to be enough to increase individuals' immunity and have a reducing effect on the number of deaths. Two-month lag of new vaccination is significant at a 10% level and a one-percentage increase in new vaccinated decreases the number of deaths per thousand by 0.093 %.

Similarly to the new cases model in Subsection 5.1.1, average temperature is not considered a significant variable, nevertheless, it has a negative coefficient.

The explanation could be the same as for the previous model. Last but not least, the stringency index does not show a significant estimate, however it has at least the negative sign of coefficient. There was an attempt to include the second lag of the stringency index, however it was not considered a significant variable. Furthermore, the second lag of the index decreases the Adjusted R-squared and therefore it was not included in the model.

Table 5.2: Estimation results - new deaths model using world data

	Dependent variable:		
	$\log(New\_deaths_{it})$		
	FD	*Robust SE*	*D-GMM*
	(1)	(2)	(3)
$\log(New\_deaths_{i(t-1)})$			0.028
			(0.103)
$\log(New\_cases_{it})$	0.743***	0.743***	0.496***
	(0.024)	(0.031)	(0.039)
$\log(New\_cases_{i(t-1)})$	0.407***	0.407***	0.543***
	(0.035)	(0.042)	(0.081)
$\log(New\_cases_{i(t-2)})$	0.057	0.057	−0.028
	(0.035)	(0.044)	(0.059)
$\log(New\_vaccinated_{i(t-1)})$	0.057	0.057	−0.070**
	(0.044)	(0.052)	(0.032)
$\log(New\_vaccinated_{i(t-2)})$	−0.093**	−0.093*	0.006
	(0.044)	(0.053)	(0.029)
$\log(New\_fully\_vaccinated_{i(t-1)})$	−0.037	−0.037	0.043
	(0.046)	(0.056)	(0.039)
$\log(New\_fully\_vaccinated_{i(t-2)})$	0.016	0.016	−0.031
	(0.044)	(0.045)	(0.033)
$Average\_temperature_{it}$	−0.020	−0.020	−0.002
	(0.017)	(0.014)	(0.010)
$Stringency\_index_{i(t-1)}$	−0.002	−0.002	−0.008**
	(0.006)	(0.007)	(0.004)
Constant	−0.236***	−0.236***	
	(0.053)	(0.039)	
Observations	1,736		124
R^2	0.414		
Adjusted R^2	0.411		
F Statistic	135.606*** (df = 9; 1726)		
Hansen-Sargan test (p-value)			0.15
Arellano-Bond test (p-value)			0.82
Note:			*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses

5.1.3 Hospitalization Model

As mentioned at the beginning of Section 5.1, it is expected that the hospitalization and new deaths models are closely related. It is necessary to keep in mind that the hospitalization model works with a different dataset that contains 31 countries and the list of them is provided in Table A.1. The reduction is caused by the unavailability of Covid-19 hospitalizations data in many countries.
Table 5.3 - Estimation results of hospitalization model using world data located below provides the estimates of the Difference GMM estimator in column number (3). The Difference GMM estimator was formulated with the first lag of the dependent variable as an instrument, however there is a problem with perfect proliferation because the p-value of the Hansen-Sargan test is equal to 1. Thus, similarly to Subsection 5.1.2 New Deaths Model, the first difference estimator is the best method for this model because it deals with the unobserved heterogeneity and does not reject the assumption of no serial correlation.
Table 5.3 of results contains two more columns, (1) and (2), where the first column provides the estimates using the first difference estimator. There is again the problem with heteroskedasticity. The results after correction are considered the most plausible and are presented in the second column. The R-squared is above 0.65 which is only 1 % more compared to the Adjusted R-squared, therefore the model is well-fitted.
The model's results are presented in the already mentioned Table 5.3. The main focus is on the description of the estimates located in the second column. Starting with new cases variable which was considered significant in the current month, in one-month delay and also with a two-month lag. As expected, the current number of new cases does not have such a high coefficient as the number of new cases with a one-month lag. It is understandable because severe conditions that lead people to hospitals could be revealed after some time. One-percentage increase in the number of current new cases per thousand raises the hospitalizations by 0.053 %, compared to the boost equal to 0.45 % caused by a one-percentage increase of new cases in the previous month. Moreover, a one-percentage increase in new cases two months ago decreases the number of hospitalizations in the present month by 0.259 %. Intuitively, it makes sense because most people would stay in the hospital for several weeks and then recover or pass away within that time.
Full vaccination is considered significant in both lags with negative coefficients. One-percentage increase in full vaccination reduces the number of people in

hospitals by 0.098 % in case of a month-lag and 0.064 % in case of a two-month lag. It supports the hypothesis that an additional dose of vaccine is essential for further protection against the severe conditions of the virus. Being vaccinated is also considered significant in both lags, however only the first lag has a negative sign and a one-percentage increase in the vaccination per thousand in the previous month decreases the hospitalization in the current month by 0.087 %.

Average temperature is significant at a 1% significance level and has a reducing effect. One-degree increase in average temperature reduces the number of hospitalizations by 5.7 % because it is included in a log-level relation. Average temperature is significant for the first time and it is probably caused by the reduction in the number of countries in the dataset which now mostly consists of countries from Europe where the fluctuation of temperatures is apparent.

Table 5.3: Estimation results - hospitalization model using world data

	Dependent variable:		
	$\log(Hospitalization_{it})$		
	FD	*Robust SE*	*D-GMM*
	(1)	(2)	(3)
$\log(Hospitalization_{i(t-1)})$			-1.234^{***}
			(0.378)
$\log(New\_cases_{it})$	0.053^{***}	0.053^{*}	0.433^{***}
	(0.013)	(0.030)	(0.042)
$\log(New\_cases_{i(t-1)})$	0.450^{***}	0.450^{***}	0.911^{***}
	(0.027)	(0.036)	(0.179)
$\log(New\_cases_{i(t-2)})$	-0.259^{***}	-0.259^{***}	0.313^{***}
	(0.026)	(0.028)	(0.113)
$\log(New\_vaccinated_{i(t-1)})$	-0.087^{**}	-0.087^{**}	0.058
	(0.037)	(0.036)	(0.045)
$\log(New\_vaccinated_{i(t-2)})$	0.101^{**}	0.101^{*}	-0.062
	(0.045)	(0.055)	(0.053)
$\log(New\_fully\_vaccinated_{i(t-1)})$	-0.098^{**}	-0.098^{**}	-0.025
	(0.044)	(0.038)	(0.029)
$\log(New\_fully\_vaccinated_{i(t-2)})$	-0.064^{**}	-0.064^{*}	-0.007
	(0.031)	(0.034)	(0.026)
$Average\_temperature_{it}$	-0.057^{***}	-0.057^{***}	0.022^{*}
	(0.008)	(0.010)	(0.013)
$Stringency\_index_{i(t-1)}$	0.007^{*}	0.007^{*}	0.008^{*}
	(0.004)	(0.004)	(0.005)
Constant	-0.045	-0.045^{*}	
	(0.028)	(0.027)	
Observations	364		31
R^2	0.656		
Adjusted R^2	0.647		
F Statistic	74.907^{***} (df = 9; 354)		
Hansen-Sargan test (p-value)			1
Arellano-Bond test (p-value)			0.005
Note:	*p<0.1; **p<0.05; ***p<0.01		
	Standard errors in parentheses		

5.1.4 Summary of World Models

Table 5.4 was created to show the comparison of all models that were commented above. The most trustworthy model for each dependent variable was included. The table's main purpose is to distinguish the effects of individual variables among the models. There is new tests variable in the new cases model that substitutes the effect of new cases in the present time because it cannot be included, since it is the dependent variable in the model. It is apparent that new cases variable is important in the present and also in the first lag for all dependent variables. It is obvious because this variable mainly determines the three key statistics. Fortunately, new cases variable is not proportional to new deaths nor hospitalizations. Furthermore, the second lag of the new cases variable is significant for the hospitalization model with a negative coefficient. Understandably, most people would stay in the hospital for several weeks and then recover or pass away within that time.

Full vaccination is significant only for the hospitalization model in both lags with negative coefficients. The reason might be the number of countries in the dataset. The hospitalization dataset was reduced to only 31 mainly European countries with usable data. New vaccinated variable has negative coefficients for all models in World Analysis. The only difference is the significance of the lag. The first lag is important among all models, however the model of new death has the second lag significance with a negative value. According to the various countries in the dataset, the two-month lag of new vaccinated is more correlated with new deaths than full vaccination.

Average temperature has a negative value for the new deaths and the hospitalization model, however only the latter is found significant. It may be again due to the number of countries in the hospitalization dataset. The stringency index is found significant for new cases and hospitalization models, nevertheless it differs in the coefficient sign. A higher level of the stringency index restricts social contacts, and thus should reduce the number of transmissions which is supported by the negative coefficient in the new cases model.

Table 5.4: Comparison of world analysis models

	Dependent variable:		
	$\log(New\_cases_{it})$	$\log(New\_deaths_{it})$	$\log(Hospitalization_{it})$
	(1)	(2)	(3)
$\log(New\_tests_{it})$	1.727^{***}		
	(0.544)		
$\log(New\_cases_{it})$		0.743^{***}	0.053^{*}
		(0.031)	(0.030)
$\log(New\_cases_{i(t-1)})$	0.368^{***}	0.407^{***}	0.450^{***}
	(0.120)	(0.042)	(0.036)
$\log(New\_cases_{i(t-2)})$	-0.013	0.057	-0.259^{***}
	(0.090)	(0.044)	(0.028)
$\log(New\_vaccinated_{i(t-1)})$	-0.277^{**}	0.057	-0.087^{**}
	(0.113)	(0.052)	(0.036)
$\log(New\_vaccinated_{i(t-2)})$	0.007	-0.093^{*}	0.101^{*}
	(0.057)	(0.053)	(0.055)
$\log(New\_fully\_vaccinated_{i(t-1)})$	0.073	-0.037	-0.098^{**}
	(0.130)	(0.056)	(0.038)
$\log(New\_fully\_vaccinated_{i(t-2)})$	-0.040	0.016	-0.064^{*}
	(0.082)	(0.045)	(0.034)
$Average\_temperature_{it}$	0.0003	-0.020	-0.057^{***}
	(0.019)	(0.014)	(0.010)
$Stringency\_index_{i(t-1)}$	-0.011^{**}	-0.002	0.007^{*}
	(0.005)	(0.007)	(0.004)
Constant		-0.236^{***}	-0.045^{*}
		(0.039)	(0.027)
Note:			*p<0.1; **p<0.05; ***p<0.01
			Standard errors in parentheses

5.2 Country Level Analysis

This section is divided into Czechia and Israel analysis. Both countries have the models formulated evenly to be comparable. According to equations formulated in Section 4.2, new deaths and hospitalization models have the same independent variables in the equation.

5.2.1 Czechia Analysis

5.2.1.1 New Cases Model

Table A.15: Estimation results of new cases model – Czechia contains the results of OLS in the first column and the model adjusted by HAC robust standard errors in the second column that is considered the most plausible.
New tests are included with an expectation of high positive significance. According to the model, a one-percentage increase in the new tests per thousand people raises the number of new cases per thousand people by 0.473 %. Full vaccination does not affect the number of new cases in the 7-day lag nor 14-day lag significantly. The reason is that the main decrease follows the first dose of vaccine. Compared to the similar model in Table 5.1 in World Analysis, new vaccinated variable in 30-day lag is also considered significant at a 5% significance level with a negative relation to the dependent variable. Although the 14-day version has a reducing effect under OLS, it is not significant after HAC correction. One-percentage increase in new vaccinated per thousand people 30 days ago decreases the number of new cases by 0.163 %.
Average temperature has a strong reducing effect on the dependent variable and a one-point increase in the variable decreases the dependent variable by 13.4 % because it is included in a log-level relation. Figure B.11 supports this evidence because higher average temperature in Czechia is associated with fewer new cases. The stringency index has significant results only in the 14-day lag, however this time interval can still include the period before the manifestation of symptoms, and thus its results are unsurprisingly positive.

5.2.1.2 New Deaths Model

Table A.16: Estimation results of deaths model – Czechia provides two columns of results – OLS and after robust SE. New cases are significant in all variants. The number of new cases per thousand people is positively related to the number of deaths. One-percentage increase in the new cases increases deaths by 0.713 %. On the other hand, the same increment 7 and 14 days ago decreases deaths by 0.081 % and 0.097 %, respectively.

Vaccination of at least one dose does not provide adequate protection against the worst consequence of Covid-19. Both lags of 14 and 30 days show positive coefficients in the OLS regression, nevertheless, only firstly mentioned lag is significant and positive after HAC adjustment. Full vaccination changed the sign of the coefficient and decreases the number of deaths. One-percentage increase in full vaccination per thousand people 14 days ago reduces the number of deaths by 0.072 % and a 30-day delay reduces the dependent variable by 0.089 % but the latter lag is no longer significant after HAC correction. Thus, further protection in the form of an additional dose of vaccine is supported in this model.

Average temperature is also considered as a variable minimizing the number of deaths. It is included using a log-level relation, and thus increment of one degree of Celsius in average temperature reduced deaths by 4.8 %. Similarly to subsubsection 5.2.1.1 New Cases Model, 14-day delay in the stringency index is not enough to have a reducing effect on the number of deaths, however with 30-day delay the stringency index changes the relation and reduces the number of deaths by 1.5 %. The stringency index is also included in a log-level relation. This evidence supports the goal of higher public restrictions.

5.2.1.3 Hospitalization Model

Table A.17: Estimation results of hospitalization model – Czechia provides the results of the last model. New cases and a 7-day delay of the same variable are considered significant with a positive effect on the number of hospitalizations. A one-percentage rise in new cases at the current time per thousand people increases hospitalizations by 0.25 %. The 7-day lag of the same variable has a lower coefficient equal to 0.078 %. 30-day lag is not considered significant, however it changes the sign of the coefficient to be negative.

New vaccinated variable shows to be significant with a positive coefficient equal to 0.122 % for the first lag and insignificant for the second lag. On the other hand, being fully vaccinated has a reducing effect on the dependent variable. A one-percentage increase in full vaccination per thousand people decreases hospitalization by 0.078 % for a 14-day lag and 0.172 % for a 30-day delay, however only the second lag of 30 days is significant at a 10% significance level after HAC correction. The importance of the second dose of vaccine is therefore supported.

Average temperature and the stringency index are included in a log-level relation, therefore a one-point increase in average temperature reduces the hospitalization number by 10.6 %. Figure B.12 provides further evidence that average temperature highly influences the number of hospitalizations. The stringency index with 14-day delay has a positive effect on the dependent variable which estimates that a one-point increase in the stringency index raises hospitalization by 3.4 %, however the delay of 30 days changes the sign and has a negative effect on hospitalization that is equal to 2.8 % with a one-point increment.

5.2.1.4 Summary of Czechia Models

This section is devoted to a summary of variables in Czechia models. Table 5.5 below compares all three already discussed models. Models are included in the version after HAC correction. The R-squared of Czechia models is slightly above 0.57 in the new cases model and slightly above 0.8 in the two other models. Moreover, differences in R-squares between the individual models are highest (equal to 0.006) when using new cases as the dependent variable. Therefore, it is confirmed that the models are formulated properly and all models are trustworthy.

New tests variable in the new cases model substitutes the effect of new cases variable in the other two models. These variables have a straight effect on the dependent variable and positive significance is apparent as expected.

New vaccinated variable is significant among all models, yet the delay in the effect differs. So does, surprisingly, the effect's direction. Only the first lag is positive and significant for deaths and hospitalisations, whereas a slowing effect is visible for the 30-day lag in the new cases model. Full vaccination has negative signs in both lags for all models, however it is considered significant only in the 14-day lag for the model of new deaths and in the 30-day lag for hospitalization as the dependent variable. The first goal of vaccination is to reduce the number of transmissions that is satisfied with the first dose of vaccine. The second goal of vaccination is to create protection against severe progress and according to (2) and (3) it is provided by an additional dose of vaccine.

Average temperature is an important factor across all models. It has a strong reducing effect on dependent variables. Average temperature has the least prominent effect on the number of deaths which is almost three times smaller compared to the new cases model. Figure B.11 and Figure B.12 supported importance of average temperature. The 14-day delay in the stringency index does not provide enough time to show the expected effect of the variable, however 30-day lag has a negative coefficient for new deaths and hospitalizations models, while an estimate near 0 for the new cases model. To sum it up, new deaths and hospitalization models are closely related, which is confirmed by the signs and significant levels of variables, whereas variables influencing the development of new cases are slightly different.

Table 5.5: Comparison of models - Czechia

	Dependent variable:		
	$\log(New\_cases_t)$	$\log(New\_deaths_t)$	$\log(Hospitalizations_t)$
	(1)	(2)	(3)
$\log(New\_tests_t)$	0.473***		
	(0.084)		
$\log(New\_cases_t)$		0.713***	0.250***
		(0.037)	(0.061)
$\log(New\_cases_{(t-7)})$		−0.081***	0.078**
		(0.031)	(0.037)
$\log(New\_cases_{(t-14)})$		−0.097***	−0.005
		(0.035)	(0.042)
$\log(New\_vaccinated_{(t-14)})$	−0.112	0.129***	0.122**
	(0.071)	(0.046)	(0.061)
$\log(New\_vaccinated_{(t-30)})$	−0.163**	0.101	0.134
	(0.077)	(0.076)	(0.109)
$\log(New\_fully\_vaccinated_{(t-14)})$	−0.044	−0.072**	−0.078
	(0.035)	(0.036)	(0.057)
$\log(New\_fully\_vaccinated_{(t-30)})$	−0.061	−0.089	−0.172*
	(0.079)	(0.064)	(0.097)
$Average\_temperature_t$	−0.134***	−0.048***	−0.106***
	(0.019)	(0.011)	(0.016)
$Stringency\_index_{(t-14)}$	0.024*	0.032***	0.034***
	(0.013)	(0.007)	(0.012)
$Stringency\_index_{(t-30)}$	0.002	−0.015**	−0.028**
	(0.011)	(0.007)	(0.013)
Constant	−2.971***	−5.621***	−0.908*
	(0.704)	(0.369)	(0.507)
Note:			*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses

5.2.2 Israel Analysis

5.2.2.1 New Cases Model

Table A.18: Estimation results of new cases model – Israel summarises the model with new cases as the dependent variable. The table has two columns, (1) with OLS results and (2) after HAC robust standard errors correction which is considered the most trustworthy. One-percentage increase in new tests raises the number of new cases by 6.3 % because it is included in a log-level relation, however, it is not considered significant after HAC correction.
New vaccination has negative signs in both lags and coefficients equal to 0.334 % in the case of the 14-day lag and 0.577 % is the effect of the 30-day delay. Full vaccination is an insignificant factor in this model which is similar to the same model formulated for the Czech Republic. The main goal of vaccination on the number of new cases is to reduce transmissions. According to all three models having the number of new cases as the dependent variable, the first dose of vaccine seems to have the desired effect.
Average temperature is another significant variable with a reducing effect. One additional degree of Celsius decreases the number of new cases by 5.3 % because it is included in a log-level relation. This effect is further supported by Figure B.11. The stringency index was found to be significant only for the 30-day delay, however it has a positive coefficient.

5.2.2.2 New Deaths Model

Results of the model with new deaths variable as the dependent variable are provided in Table A.19 which contains two columns: one for OLS and the second after HAC robust SE. The only significant variant of new cases variable is the present time whose coefficient is equal to 0.367 % for a one-percentage increase in the independent variable. Although it is insignificant, new cases variable with a 14-day delay has a negative coefficient.
New vaccinated variable is significant only for the 30-day delay before HAC correction. The second column of results shows the insignificance of new vaccination. Furthermore, full vaccination is also insignificant, nevertheless it has negative relation for a 30-day delay equal to 0.068 % for a one-percentage increment.
Average temperature is insignificant, however it has a negative sign of the estimate. The stringency index is significant at a 1% significance level for both included variants. The one with a 14-day delay has a positive coefficient,

whereas the variable with a 30-day lag is negatively related to the dependent variable. 14 days seem insufficient to show the proper effect of the index, however the interval of one month is sufficient. A one-point increase in the stringency index 30 days ago decreases the number of deaths by 3.7 %.

5.2.2.3 Hospitalization Model

Results are provided in Table A.20 with the identical structure as in the aforementioned models. All delays of new cases variable are considered significant and have a positive sign except the 14-day lag. The current new cases variable increases hospitalization by 0.147 %, whereas a one-percentage increment in the new case's 7-day lag raises the hospitalizations by 0.079 %.
New vaccination has a reducing effect for the first lag at the 1% significance level. One-percentage increase in the new vaccination variable per thousand people 14 days ago decreases the number of hospitalizations per thousand by 0.114 % and ceteris paribus only with 30 days lag resulted in a decrease of 0.048 % but not significant. Full vaccination is insignificant for a 14-day lag as well as for a 30-day lag, nevertheless the latter has a negative sign of the estimate. This might be caused by the fast vaccination campaign supported by Figure B.4.
Average temperature is an important reducing factor shortening the number of hospitalizations by 2 % for each degree of Celsius increment (see also Figure B.12). The stringency index with a 14-day delay positively affects the number of hospitalizations. On the other hand, a one-point increase in the stringency index 30 days ago decreases hospitalizations by 1 %, nevertheless this variable is insignificant after HAC application.

5.2.2.4 Summary of Israel Models

Table 5.6 below compares all three formulated models to determine whether there are differences. Only models after HAC correction are included. The R-squared of the new cases model of Israel is 0.37 which is smaller compared to the new cases model of Czechia, however it does not cause any problem because its Adjusted R-squared is equal to 0.36 which is close to the R-squared. The deaths model has the R-squared close to 0.5 and the last model has a value near 0.8. Both have the Adjusted R-squared values close to the R-squared, therefore all models are considered well-fitted. New tests and new cases independent variables capture the information about the detection of Covid-19, thus the expectation of positive coefficients is confirmed. Moreover, the 7-day lag of new cases still significantly affects the number of people in hospitals. Vaccination of at least one dose is considered significant and negatively related to both lags in the new cases model and in the first lag for the hospitalization model. Full vaccination is not significant in any of the models. Average temperature is substantial only in the new cases and hospitalization models where it shows a reducing effect (supported by Figure B.11 and Figure B.12). The stringency index with a 14-day delay does not provide a reducing effect, however 30-day lag ensures a reduction in the new deaths and hospitalization models which supports that public restriction influences the dependent variables.

Table 5.6: Comparison of models - Israel

	Dependent variable:		
	$\log(New\_cases_t)$	$\log(New\_deaths_t)$	$\log(Hospitalization_t)$
	(1)	(2)	(3)
New_tests_t	0.063		
	(0.050)		
$\log(New\_cases_t)$		0.367^{***}	0.147^{***}
		(0.038)	(0.041)
$\log(New\_cases_{(t-7)})$		0.035	0.079^{***}
		(0.034)	(0.016)
$\log(New\_cases_{(t-14)})$		-0.045	0.032
		(0.038)	(0.029)
$\log(New\_vaccinated_{(t-14)})$	-0.334^{**}	0.068	-0.114^{***}
	(0.148)	(0.047)	(0.043)
$\log(New\_vaccinated_{(t-30)})$	-0.577^{***}	0.100	-0.048
	(0.175)	(0.070)	(0.046)
$\log(New\_fully\_vaccinated_{(t-14)})$	0.035	0.026	0.033
	(0.052)	(0.023)	(0.022)
$\log(New\_fully\_vaccinated_{(t-30)})$	0.194	-0.068	-0.006
	(0.130)	(0.051)	(0.034)
$Average\_temperature_t$	-0.053^{*}	-0.003	-0.020^{*}
	(0.030)	(0.011)	(0.010)
$Stringency\_index_{(t-14)}$	0.019	0.048^{***}	0.034^{***}
	(0.034)	(0.010)	(0.009)
$Stringency\_index_{(t-30)}$	0.064^{*}	-0.037^{***}	-0.010
	(0.036)	(0.010)	(0.007)
Constant	-6.350^{***}	-6.833^{***}	-3.081^{***}
	(1.585)	(0.604)	(0.535)
Note:			*p<0.1; **p<0.05; ***p<0.01 Standard errors in parentheses

5.3 Comparison of Models

Although the effects of variables within individual countries were commented, it would be more transparent to see whether there are differences between all models that were included in the analyses. Table 5.7 below contains the summary of all models. It is divided into nine columns and always three columns are related to the same dependent variable. The first triplet shows the number of new cases as the dependent variable. The second triplet deals with the number of deaths and the last one is related to hospitalizations. Columns (1), (4), and (7) are the results of World Analysis. (2), (5), and (8) refer to the Czechia models. The last three are connected to Israel Analysis.
New tests results support the hypothesis of a positive and significant effect. The same relation is apparent for the number of new cases in the current time with the exception of new tests in Israel's new cases model that is insignificant. Naturally, there would not be any pandemic without testing and detection of positive cases.
World Analysis shows positive effects of the first lag of new cases variable in all models. Followed by a negative coefficient in the second lag for the hospitalization model. The first lags are also important for individual countries with negative coefficients only in the new deaths model. Therefore new cases lags are also crucial factors.
Vaccination with one dose has negative estimates in all models with new cases as the dependent variable. Full vaccination is not considered significant in all lags when referring to new cases models. It supports the hypothesis that the first dose of vaccine reduces the number of transmissions.
Concerning death models, new vaccination has a negative coefficient only in the lag of 60 days in World Analysis. It might be the reason why full vaccination is not significant for this world model. On the other hand, the individual analyses show different results. Full vaccination significantly and negatively affects the number of deaths only in Czechia.
There are differences in signs of vaccination estimates when talking about hospitalization models. The first vaccine has a significant and diminishing estimate only for Israel in the first lag and in the 30-day lag concerning World Analysis. Full vaccination is insignificant for Israel. On the other hand, Czechia and World Analysis saw the opposite effect and full vaccination seems to be associated with a reduction in the dependent variable. This might be caused by the fast vaccination campaign in Israel at the beginning.

Average temperature is an important reduction factor in all models concerning individual countries. The problem with the insignificance of the variable in the case of the world new cases and the new deaths model is the variety of countries included in the dataset. There could be many countries with stable temperatures during the whole year, however Covid-19 was not eliminated by this condition. This assumption might be confirmed by the significance of the variable in the world hospitalization model, where the number of countries is reduced due to the unavailability of data and mainly European countries with fluctuating average temperatures are left. Individual countries analyses have a great reduction relation between average temperature and the dependent variables (see Figure B.11 and Figure B.12). To conclude, average temperature has a strong power in influencing the development of the Covid-19 pandemic.
Last but not least, the stringency index was included in the models capturing the level of public restrictions. Lags of 14 days do not reduce the dependent variables because there might not be enough time to show the expected effect. Therefore, the stringency index was also included in the one-month delay and it reveals a reducing impact on hospitalization and deaths in both countries. World Analysis shows a negative coefficient only in the new cases model. To sum it up, the stringency index seems to have the desired effect on the development of the pandemic.

Table 5.7: Comparison of all models

	Dependent variable:								
	$\log(New\_cases_t)$			$\log(New\_deaths_t)$			$\log(Hospitalization_t)$		
	World	Czechia	Israel	World	Czechia	Israel	World	Czechia	Israel
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\log(New\_tests_t)$	1.727^{***}	0.473^{***}							
	(0.544)	(0.084)							
New_tests_t			0.063						
			(0.050)						
$\log(New\_cases_t)$				0.743^{***}	0.713^{***}	0.367^{***}	0.053^{*}	0.250^{***}	0.147^{***}
				(0.031)	(0.037)	(0.038)	(0.030)	(0.061)	(0.041)
$\log(New\_cases_{(t-7)})$					-0.081^{***}	0.035		0.078^{**}	0.079^{***}
					(0.031)	(0.034)		(0.037)	(0.016)
$\log(New\_cases_{(t-14)})$					-0.097^{***}	-0.045		-0.005	0.032
					(0.035)	(0.038)		(0.042)	(0.029)
$\log(New\_cases_{t-30})$	0.368^{***}			0.407^{***}			0.450^{***}		
	(0.120)			(0.042)			(0.036)		
$\log(New\_cases_{t-60})$	-0.013			0.057			-0.259^{***}		
	(0.090)			(0.044)			(0.028)		
$\log(New\_vaccinated_{(t-14)})$		-0.112	-0.334^{**}		0.129^{***}	0.068		0.122^{**}	-0.114^{***}
		(0.071)	(0.148)		(0.046)	(0.047)		(0.061)	(0.043)
$\log(New\_vaccinated_{(t-30)})$	-0.277^{**}	-0.163^{**}	-0.577^{***}	0.057	0.101	0.100	-0.087^{**}	0.134	-0.048
	(0.113)	(0.077)	(0.175)	(0.052)	(0.076)	(0.070)	(0.036)	(0.109)	(0.046)
$\log(New\_vaccinated_{(t-60)})$	0.007			-0.093^{*}			0.101^{*}		
	(0.057)			(0.053)			(0.055)		
$\log(New\_fully\_vaccinated_{(t-14)})$		-0.044	0.035		-0.072^{**}	0.026		-0.078	0.033
		(0.035)	(0.052)		(0.036)	(0.023)		(0.057)	(0.022)
$\log(New\_fully\_vaccinated_{(t-30)})$	0.073	-0.061	0.194	-0.037	-0.089	-0.068	-0.098^{**}	-0.172^{*}	-0.006
	(0.130)	(0.079)	(0.130)	(0.056)	(0.064)	(0.051)	(0.038)	(0.097)	(0.034)
$\log(New\_fully\_vaccinated_{(t-60)})$	-0.040			0.016			-0.064^{*}		
	(0.082)			(0.045)			(0.034)		
$Average\_temperature_t$	0.0003	-0.134^{***}	-0.053^{*}	-0.020	-0.048^{***}	-0.003	-0.057^{***}	-0.106^{***}	-0.020^{*}
	(0.019)	(0.019)	(0.030)	(0.014)	(0.011)	(0.011)	(0.010)	(0.016)	(0.010)
$Stringency\_index_{(t-14)}$		0.024^{*}	0.019		0.032^{***}	0.048^{***}		0.034^{***}	0.034^{***}
		(0.013)	(0.034)		(0.007)	(0.010)		(0.012)	(0.009)
$Stringency\_index_{(t-30)}$	-0.011^{**}	0.002	0.064^{*}	-0.002	-0.015^{**}	-0.037^{***}	0.007^{*}	-0.028^{**}	-0.010
	(0.005)	(0.011)	(0.036)	(0.007)	(0.007)	(0.010)	(0.004)	(0.013)	(0.007)
Constant		-2.971^{***}	-6.350^{***}		-5.621^{***}	-6.833^{***}	-0.045^{*}	-0.908^{*}	-3.081^{***}
		(0.704)	(1.585)		(0.369)	(0.604)	(0.027)	(0.507)	(0.535)

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors in parentheses

Chapter 6

Excess Deaths Analysis

Excess deaths analysis is the third part of the thesis and its main goal is to focus on the most severe impact of the Covid-19 pandemic, i.e., numerous deaths. In this section, only the Czech Republic will be examined. The main aim of this section is to predict the amount of money not paid in pensions till the end of 2030 as a consequence of excess deaths caused by the pandemic.
The structure of this section is as follows. Firstly, identify the excess mortality caused by the pandemic with a focus on older people who are not working full time, however the government provides them monthly payments in the form of pensions. Secondly, life expectations for the specific age groups will be investigated. Then, the average pension will be presented. After that, the calculation will be formulated to estimate the amount of money that will remain in the government's pension budget until the end of 2030. Moreover, the cost of vaccines will be investigated and will be used as an expenditure that could be compensated by the amount of money that will not be paid to pensioners.

6.1 Death Statistics

The Czech Statistical Office (2022) (further as CZSO) measures the number of deaths in the Czech Republic and provides statistics for every week from the year 2012. Furthermore, the dataset is divided into subsections based on the age groups of the population. There are six groups starting with children of 0-14 years of age. The following two groups consisting of 15-44 years old people and 45-64-year-old individuals could be seen as the main productive years of life. Nevertheless, the most important age groups for the purpose of the analysis are remaining, i.e., 65-74, 75-84, and the last group of 85+ years of age. The last three mentioned groups are supposed to be the most vulnerable to the negative symptoms of Covid-19. The hypothesis is to find out that the elderly represent the greatest portion of total deaths, and therefore these groups will be used for the exploration of excess mortality caused primarily or secondarily by Covid-19.

The Ministry of Health of the Czech Republic (2022) reported 42 149 official deaths caused by Covid-19 as of the 30^{th} of December 2022. However, the excess deaths analysis will be based on the total number of deaths reported by CZSO because many people might not pass away primarily from Covid-19 but due to other factors connected with the pandemic. Fear caused by the mass media and solitude as a result of public restrictions could affect mortality, however these deaths cannot be included in the official deaths caused by Covid-19. Moreover, the Ministry of Health does not divide the statistics about Covid-19 deaths into age groups. Their data consist of daily reported deaths at the time the virus was detected in the organism and therefore it is not suitable for the analysis.

Based on the downloaded data from CZSO, Figure B.15 was created. The graph shows the number of deaths in each age group from 2012 to 2022. Table A.21 contains the values used for the figure creation. The first two groups depicting people under 45 years old do not present visible differences. Moreover, the values are close to 0 for the youngest group and near 3 000 for the group 15 to 44 years old. Green colour representing group 45-64 shows a decreasing trend in the graph with an increment in 2021 and almost the same value in 2022, however everyone in this group cannot be considered as a pensioner and therefore this group will not be used in the analysis. The rest of the groups show an apparent increase in the number of deaths and these groups will be further examined. Deaths in the next group 65-74 were rising during the years with a peak in 2021 followed by a slight reduction in the last observed year. 75-84 years old people observed the most visible increase out of all groups. The rise was not that noticeable in 2020, although it was equal to 2 000 more deaths compared to the previous year. However, the number of deaths grew rapidly in the next two years. The last age group 85+ represented by the purple line also rose with a boost in the last year.

The problem is with determining the excess deaths because it is not known what the number of deaths in the years from 2020 to 2022 in a situation without the Covid-19 pandemic would be. The estimate of deaths in the mentioned year could be forecasted based on the values from the previous years. For this purpose, the Excel function forecast is used to estimate mortality in the last

3 years based on the statistics presented in Table A.21. The function creates linear regression using 3 inputs. It takes the column containing the number of deaths as the y-value and the x-value reflects the time dimension. The third input is the year that should be forecasted. It is necessary to predict the years 2020 to 2022, and therefore the observations used in the forecast capture years 2012 to 2019. The same period is used for all 3 years of prediction.
The following Table A.22 compares the forecasted values and real values during the Covid-19 pandemic. The last two rows compare the sum of values forecasted and real. The total difference is equal to 46 452 deaths which are about 10 % higher than the official number of deaths caused by Covid-19, i.e. 42 149 by the end of 2022. The difference in the values will be used in the computation. The excess of deaths is depicted in Table A.23 and it was calculated from Table A.22 by subtracting the real number of deaths from the forecasted value. Surprisingly, there are more predicted deaths in 2020 than the real number for the two groups which is the reason why there are negative values. However, the values are increasing in the following years.

6.2 Life Expectancy

CZSO (2022) has an additional table that is usable and important for the analysis. It measures the life expectancy of people in a form that is appropriate for the investigation. The file reports the life expectancy for every year at the time of birth and then for age groups 5, 10, . . . till 105. Moreover, it is separated for men and women. The file contains life expectancies from 1912 to 2021, from which only the last three rows (2019, 2020, 2021) were selected. Unfortunately, values for the year 2022 are not available at the time of writing this thesis. The values between 2020 and 2021 do not differ extensively, therefore it is expected that the same will follow in 2022. Thus, it was decided to use the values from 2021 also for the following year. Data about life expectancy are usually reported at the time of birth and represent the quality-of-life measure, however it would not make sense that all people would die at a certain age at the latest.
Table A.24 provides life expectancies based on the data from CZSO. A detailed description of the methodology that is used for the determination of the values is not provided, however CZSO is considered a trustworthy source of data and information about the Czech Republic. The age groups were filtered to contain values only for people older than 65 years old. The first step in the

data transformation was to merge certain columns and to create an average of life expectancies. Group 65-74 consists of averages of ages 65, 70, and 75. The next group 75-84 was calculated based on values from 75, 80, and 85 years old people. The last group was calculated based on the rest of the provided values. These calculations were provided separately for men and women and the result is presented in Table A.25. Moreover, the table contains one more section called Average which is computed as an average of men's and women's life expectancies and this value will be further used in the calculations. The average of men and women is constructed because it is impossible to distinguish the sex of individuals in the excess deaths statistics. Unsurprisingly, the life expectancy for women is higher than for men. Using averages may undervalue the real condition, however the author's approach is to have a rather undervalued estimate than having the highest value with a very low percentage of possibility. Since the investigation is extending until 2030, life expectancy above 10 years is out of scope of the analysis.

6.3 Pensions

Czech Social Security Administration (2022) describes the conditions required to get a pension in the Czech Republic. For men, only the age matters, while women's retirement age is also adjusted by the number of children raised. Based on effective law, the retirement age is gradually changing and will soon reach 65 years for all cohorts. For example, people born in 1947 had a pension age equal to 62. According to the age groups included in the dataset, everyone in the groups is above 65 years old and therefore considered a person receiving a pension.

The Ministry of Labour and Social Affairs (2023) created a report about the increase in pensions in the year 2023 and included a graph showing the development of average pensions in the Czech Republic from the year 2010. Table A.26 depicts the average pensions in CZK. It is apparent that pensions have been rising rapidly in recent years. The increase is 17 % in 2022 compared to 2021 and an additional 12 % in 2023. Since the beginning of the pandemic, the average pension has risen by 39.4 % due to high inflation which is expected to subside in the following years. For the estimation of unpaid pensions in the future, it is necessary to forecast the average pensions till 2030 which is the last year of investigation.

Increase in pensions is influenced by inflation. Czech National Bank (2023) created a prognosis about the future inflation rate in the Czech Republic. Although inflation is at a level of 14.8 % at the time the excess deaths analysis is created, it is predicted to almost reach the 2 % inflation target at the beginning of 2024. However, the 70 % confidence interval indicates the inflation rate to be slightly below 5 %. The estimated inflation is used for forecasting pensions. The increase in average pensions is therefore predicted to be 4 % in 2024, then 3 % in the next two years, and the next four years are set to be equal to 2 %. Predictions are also included in Table A.26.

6.4 Vaccines' Expenses

Many types of expenses are associated with the Covid-19 pandemic, such as increased budgets for healthcare operating at or above capacity, or subsidies for entrepreneurs. These costs should be difficult to measure at a detailed level. Therefore, expenditure that could be measured at a precise amount is sought. Vaccines should serve as a direct protection against the development of Covid-19 and also should provide protection against the severe process of the virus that could result in death. The cost of vaccines' purchases could serve as a good measure, therefore money not paid in pensions in the long run could be compared to the vaccines' investments.

The Ministry of Health of the Czech Republic (2023) provides a list of its invoices publicly. There are files for every month in years of interest (2021, 2022). These two years were selected because the vaccination in the Czech Republic started in December 2020, however there is no invoice for a vaccine in December 2020 which means that the first invoice was probably paid at the beginning of the year 2021.

Lists of invoices were downloaded (12 for each year) and merged based on the year to separate files. The identification of vaccines' expenses was made by looking for the producers of vaccines. Five leading suppliers were detected and expenditures for these companies will be computed, namely Pfizer, spol s r.o.; Moderna Biotech Spain S.L.; Janssen Pharmaceutica NV; AstraZeneca AB; and NOVAVAX CZ a.s. Supplier Janssen Pharmaceutica NV is a company in the Johnson & Johnson group, therefore it was publicly marked as Johnson & Johnson vaccine.

A summary of expenses is presented in Table A.27. It is apparent that the majority of expenses are connected with Pfizer company (83.4 %), followed by

Moderna (13.3 %) and the rest is equal to 3.3 %. The total amount of money spent on vaccines equals 14 838.5 million CZK. This is the cost of the vaccines and does not include the costs related to the operation of vaccination centres or other personnel costs.

6.5 Results

The model was constructed for each age group separately, however the proceeding is identical. Each group was divided into 3 sections according to the deaths in a particular year because it is impossible to determine the exact date of death in the excess statistics. The first excess deaths are provided for 2020 and therefore these statistics are used starting in 2021. Deaths recorded in 2021 are used in the calculation since 2022. Similarly for data from 2022. Life expectancies are used to demonstrate how many years a person in a particular age group is expected to live. According to Table A.25, 3 years were used for group 85+, 7.5 years are expected for age group 75-84, and 13 years for the youngest investigated group. Since the excess deaths analysis examines the period till 2030, the youngest group must be restricted. According to the excess deaths in 2020, only the group 75-84 was used for this year's calculations.
The last variable needed for the calculation is the average pension that was forecasted in Section 6.3. The amount of money not paid was calculated as the number of excess deaths times the average pension in the appropriate year. Table A.28, Table A.29, and Table A.30 show the distribution of unpaid money in the years of investigation. Each table is related to one of the age groups and also contains the sum of every year's excess deaths. The next Table A.31 summarizes the amount of money for the three age groups and their aggregates. Last but not least, it provides the total amount of money not paid due to increased mortality during the Covid-19 pandemic till 2030 which is equal to 5 761.9 million CZK.
Finally, the value of pensions that are not paid can compensate for the expenses that were spent on vaccines' purchases which should reduce the development of the Covid-19 pandemic. It serves as a protection against the more severe condition of the virus which could be associated with death. The purchase expenses have been described and their total value equals 14 838.5 million CZK till the end of 2022. The value is presented in Table A.27. Therefore, the money not paid in pensions could cover 39 % of costs associated with purchases of vaccines.

Chapter 7

Conclusion

This thesis aimed to investigate the effect of vaccination on the number of key statistics related to the Covid-19 pandemic - the number of new cases, new deaths, and hospitalization. A large dataset containing countries around the world was utilized with observations transformed into monthly intervals. Several methods were used in this section because of the panel data structure of the dataset. The inclusion of lags of the dependent variable in the world new cases model was accompanied by the adoption of the Difference GMM estimator. The other two world models were estimated using first-difference OLS regression because the dependent variable lags were insignificant. New cases and new deaths models appear to have a reducing effect of the first dose of vaccination and the hospitalization model showed evidence that also the second dose of vaccine reduces the number of people in hospitals.
Furthermore, analyses examining the Czech Republic and Israel separately with data filtered from World Analysis dataset based on daily observations were created because each country could have different outputs. Time series data are investigated and therefore different methodology has to be selected. OLS is considered appropriate and is used in all models. According to the new cases models, both countries show the negative estimate of new vaccinations with greater coefficients in Israel. Moreover, the insignificance of the further dose is apparent in both countries. New deaths and hospitalization models appear to have a reducing effect of full vaccination in the Czech Republic. Any lag of vaccination variables is not considered significant in the model of the new deaths in Israel, however the first vaccination has a decreasing effect on the number of hospitalizations. Average temperature strongly influences the number of dependent variables in all models.

In response to the investigation of the Czech Republic, an excess deaths analysis was carried out. It focuses on the estimation of excess deaths that were primarily or secondarily caused by the Covid-19 pandemic. Moreover, it evaluates the amount of money that will not be paid in pensions until 2030 due to the excess deaths. Finally, the excess deaths analysis compares the calculated amount of money to the expenses associated with vaccine's purchases, concluding that almost 40 % of these expenditures could be financed by unpaid pensions.

The author is aware that analyses created in this thesis are not without limitations. World Analysis was analysed after the aggregation of data into monthly intervals which obviously shortened the time dimension of the data. However, a future investigation could consider the time dimension and work with time series approaches. Unfortunately, it is not possible to take any variants of the SARS-CoV-2 virus into account because information about variants that were present in the confirmation of Covid-19 is not provided. Similarly, there are many types of vaccines, however it cannot be assigned precisely which one was used for a particular person. The excess deaths analysis has only limited insight into excess deaths consequences of the Covid-19 pandemic. Limitations could be associated with the forecasting accuracy because it cannot be certainly estimated what the average future pension will be or what the number of deaths would be without the Covid-19 pandemic. Further analysis investigation additional costs related to the Covid-19 pandemic in the Czech Republic would be beneficial. Despite the limitations, the results of the analyses support the importance of vaccination that creates protection against the Covid-19 pandemic.

Bibliography

Arellano, M. & S. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* **58(2)**: pp. 277–297.

Arellano, M. & O. Bover (1995): "Another look at the instrumental variable estimation of error-components models." *Journal of Econometrics* **68(1)**: pp. 29–51.

Blundell, R. & S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics* **87(1)**: pp. 115–143.

Chen, X., H. Huang, J. Ju, R. Sun, & J. Zhang (2022): "Impact of vaccination on the COVID-19 pandemic in U.S. states." *Scientific Reports* **12(1554)**. Available at https://doi.org/10.1038/s41598-022-05498-z,.

Czech National Bank (2023): "Prognosis cnb – winter 2023." Available at https://www.cnb.cz/cs/menova-politika/prognoza/.

Czech Social Security Administration (2022): "Pensions in detail." Available at https://www.cssz.cz/starobni-duchod-podrobne.

Czech Statistical Office (2022): "The number of deaths – weekly and monthly reported time series." Available at https://www.czso.cz/csu/czso/obypz_cr.

CZSO (2022): "Life expectancy." Available at https://www.czso.cz/csu/czso/obyut_cr.

Jain, V., A. Serisier, & P. Lorgelly (2022): "The Real-World Impact of Vaccination on COVID-19 Cases During Europe's Fourth Wave." *Int J Public Health* **67(1604793)**. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8877421/.

Kayano, T., M. Sasanami, T. Kobayashi, Y. K.Ko, K. Otani, & M. Suzuki (2022): "Number of averted COVID-19 cases and deaths attributable to reduced risk in vaccinated individuals in Japan." *The Lancet*

28(100571). Available at https://doi.org/10.1016/j.lanwpc.2022.100571.

Li, A. Y., T. C. Hannah, J. R. Durbin, N. Dreher, F. M. McAuley, N. F. Marayati, Z. Spiera, M. Ali, A. Gometz, J. Kostman, & T. F. Choudhri (2020): "Multivariate analysis of factors affecting covid-19 case and death rate in u.s. counties: The significant effects of black race and temperature." *medRxiv* .

Ministry of Health of the Czech Republic (2022): "Covid-19: Overview of the current situation in the czech republic." Available at https://onemocneni-aktualne.mzcr.cz/covid-19.

Ministry of Health of the Czech Republic (2023): "Invoices." Available at https://www.mzcr.cz/category/uredni-deska/faktury/.

Ministry of Labour and Social Affairs (2023): "Development of pensions in time." Available at https://www.mpsv.cz/documents/20142/4552532/TZ_valorizace_ÄDerven_15.2.2023+%281%29.pdf/0129a64f-7c91-bf5e-a27c-d0633ab3f319.

National Oceanic and Atmospheric Administration (2023): "Integrated surface dataset (global)." Available at https://www.ncei.noaa.gov/access/search/data-search/global-hourly.

Nickell, S. (1981): "Biases in dynamic models with fixed effects." *Econometrica* **49(6)**: pp. 1417–26.

Ritchie, H., E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian, & M. Roser (2020): "Coronavirus pandemic (covid-19)." *Our World in Data* Available at https://ourworldindata.org/coronavirus. [Accessed 2022-01-16].

Roodman, D. (2009a): "How to do xtabond2: An introduction to difference and system gmm in stata." *The Stata Journal* **9(1)**: pp. 86–136.

Roodman, D. (2009b): "A note on the theme of too many instruments*." *Oxford Bulletin of Economics and Statistics* **71(1)**: pp. 135–158.

Roser, M. (2021): "What is the covid-19 stringency index." Available at https://ourworldindata.org/metrics-explained-covid19-stringency-index.

Rustagi, V., M. Bajaj, Tanvi, P. Singh, R. Aggarval, M. F. AlAjmi, &

A. Hussain (2021): "Analyzing the Effect of Vaccination Over COVID Cases and Deaths in Asian Countries Using Machine Learning Models." *PubMed Central* **11(806265)**. Available at `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8877421/`.

Toharudin, T., R. S. Pontoh, R. E. Caraka, S. Zahroh, P. Kendogo, N. Sijabat, M. D. P. Sari, P. U. Gio, M. Basyuni, & B. Pardamean (2021): "National vaccination and local intervention impacts on covid-19 cases." *Sustainability* **13(15)**.

Velasco, J. M., W.-C. Tseng, & C.-L. Chang (2021): "Factors affecting the cases and deaths of COVID-19 victims." *Int J Environ Res Public Health* **18(2)**.

Watson, O. J., G. Barnsley, J. Toor, A. B. Hogan, P. Winskill, & A. C. Ghani (2022): "Global impact of the first year of COVID-19 vaccination: a mathematical modelling study." *The Lancet Infectious Diseases* **22(9)**: pp. 1293 – 1302. Available at `https://doi.org/10.1016/j.lanwpc.2022.100571`.

Weatherbase (2022): "Travel weather averages (weatherbase)." Available at `https://www.weatherbase.com/`. [Accessed 2021-01-16].

WHO Collaborating Centre for Vaccine Safety (2021): "How quickly does the vaccine work and how long does the protection last." Available at urlhttps://www.covid19infovaccines.com/en-posts/how-quickly-does-the-vaccine-work-and-how-long-does-the-protection-last.

Wooldridge, J. M. (2016): "Introductory econometrics: A modern approach."

Worldometer (2023): "Covid-19 coronavirus pandemic." Available at `https://www.worldometers.info/coronavirus/`.

World Health Organization (2022): "Statement for healthcare professionals: How covid-19 vaccines are regulated for safety and effectiveness (revised march 2022)." Available at urlhttps://www.who.int/news/item/17-05-2022-statement-for-healthcare-professionals-how-covid-19-vaccines-are-regulated-for-safety-and-effectiveness.

Şahin, M. (2020): "Impact of weather on COVID-19 pandemic in Turkey." *National Library of Medicine* Available at `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7169889/`.

Appendix A

Tables

Table A.1: List of countries in hospitalization dataset

Austria	Belgium	Bolivia	Bulgaria	Croatia
Cyprus	Czechia	Denmark	Estonia	Finland
France	Hungary	Iceland	Ireland	Israel
Italy	Latvia	Lithuania	Luxembourg	Malaysia
Malta	Netherlands	Norway	Portugal	Serbia
Slovakia	Slovenia	South Africa	Spain	Sweden
Switzerland				

Table A.2: List of countries in dataset

Albania	Angola	Argentina	Aruba	Austria
Azerbaijan	Bahamas	Bahrain	Bangladesh	Barbados
Belarus	Belgium	Belize	Bermuda	Bhutan
Bolivia	Bosnia and Herzegovina	Bulgaria	Burkina Faso	Burundi
Cambodia	Cape Verde	Chile	Colombia	Costa Rica
Cote d'Ivoire	Croatia	Cuba	Cyprus	Czechia
Denmark	Djibouti	Dominica	Dominican Republic	Ecuador
El Salvador	Estonia	Eswatini	Ethiopia	Faeroe Islands
Fiji	Finland	France	Gabon	Georgia
Ghana	Greece	Guatemala	Haiti	Hungary
Iceland	India	Indonesia	Iran	Iraq
Ireland	Israel	Italy	Jamaica	Japan
Jordan	Kazakhstan	Kenya	Kosovo	Kuwait
Laos	Latvia	Lebanon	Libya	Liechtenstein
Lithuania	Luxembourg	Malawi	Malaysia	Mali
Malta	Mauritania	Moldova	Mongolia	Morocco
Mozambique	Myanmar	Namibia	Nepal	Netherlands
New Zealand	Niger	Norway	Pakistan	Panama
Paraguay	Philippines	Portugal	Qatar	Russia
Rwanda	Saudi Arabia	Senegal	Serbia	Singapore
Slovakia	Slovenia	South Africa	South Korea	South Sudan
Spain	Sri Lanka	Suriname	Sweden	Switzerland
Taiwan	Thailand	Timor	Togo	Trinidad and Tobago
Tunisia	Turkey	Uganda	Ukraine	United Arab Emirates
Uruguay	Vietnam	Zambia	Zimbabwe	

Table A.3: Correlation matrix - world data

	Stri. Index	Average Temp.	N. Cases	N. tests	N. vaccinated	N. Deaths	N. vaccinated fully	Hospital.
Stringency Index	1							
Average Temperature	−0.053	1						
New Cases	0.32	−0.33	1					
New tests	0.088	−0.38	0.65*	1				
New vaccinated	0.48*	−0.013	0.13	0.2	1			
New Deaths	0.4*	−0.31	0.73*	0.4*	0.24	1		
New vaccinated fully	0.43*	−0.028	0.14	0.23	0.76*	0.22	1	
Hospitalization	0.21	−0.46*	0.57*	0.14	−0.029	0.71*	−0.12	1

Note: All variables apart from Stringency Index and Average temperature are in logs per thousand people
* significant at %

Table A.4: Specification of tests - new cases model

Wooldridge test (FE)*	0.00
Wooldridge test (FD)*	0.00
Breusch-Pagan test*	0.00
LM test*	0.00
F test*	0.00
Hausman test*	0.03
Arellano-Bond test (D-GMM)**	0.0

* All the tests were performed on the static model
** Tests the second order serial correlation in the dynamic model with only one lag of the dependent variable included

Table A.5: Specification of tests - new deaths model

Wooldridge test (FE)*	0.00
Wooldridge test (FD)*	0.00
Breusch-Pagan test*	0.00
LM test*	0.00
F test*	0.00
Hausman test*	0.003
Arellano-Bond test (D-GMM)**	0.81

* All the tests were performed on the static model
** Tests the second order serial correlation in the dynamic model with only one lag of the dependent variable included

Table A.6: Specification of tests - hospitalization model

Wooldridge test (FE)*	0.00
Wooldridge test (FD)*	0.37
Breusch-Pagan test*	0.01
LM test*	0.00
F test*	0.00
Hausman test*	0.00
Arellano-Bond test (D-GMM)**	0.005

* All the tests were performed on the static model
** Tests the second order serial correlation in the dynamic model with only one lag of the dependent variable included

Table A.7: Breusch-Pagan tests - Czechia

Breusch-Pagan test - New cases*	0.00
Breusch-Pagan test - New deaths*	0.00
Breusch-Pagan test - Hospitalization*	0.00

Table A.8: Breusch-Pagan tests - Israel

Breusch-Pagan test - New cases*	0.00
Breusch-Pagan test - New deaths*	0.00
Breusch-Pagan test - Hospitalization*	0.00

Table A.9: Variance inflation factors - new cases model - Czechia

$\log(New\_tests_t)$	2.035
$\log(New\_vaccinated_{(t-14)})$	5.557
$\log(New\_vaccinated_{(t-30)})$	8.193
$\log(New\_fully\_vaccinated_{(t-14)})$	3.597
$\log(New\_fully\_vaccinated_{(t-30)})$	6.733
$Average\_temperature_t$	1.716
$Stringency\_index_{(t-14)}$	11.963
$Stringency\_index_{(t-30)}$	11.519

Table A.10: Variance inflation factors - deaths model - Czechia

$\log(New\_cases_t)$	2.737
$\log(New\_cases_{(t-7)})$	3.008
$\log(New\_cases_{(t-14)})$	2.672
$\log(New\_vaccinated_{(t-14)})$	5.427
$\log(New\_vaccinated_{(t-30)})$	8.375
$\log(New\_fully\_vaccinated_{(t-14)})$	3.372
$\log(New\_fully\_vaccinated_{(t-30)})$	6.716
$Average\_temperature_t$	2.988
$Stringency\_index_{(t-14)}$	12.296
$Stringency\_index_{(t-30)}$	11.892

Table A.11: Variance inflation factors - hospitalization model - Czechia

$\log(New\_cases_t)$	2.777
$\log(New\_cases_{(t-7)})$	3.041
$\log(New\_cases_{(t-14)})$	2.701
$\log(New\_vaccinated_{(t-14)})$	5.414
$\log(New\_vaccinated_{(t-30)})$	8.370
$\log(New\_fully\_vaccinated_{(t-14)})$	3.374
$\log(New\_fully\_vaccinated_{(t-30)})$	6.721
$Average\_temperature_t$	3.029
$Stringency\_index_{(t-14)}$	12.203
$Stringency\_index_{(t-30)}$	11.706

Table A.12: Variance inflation factors - new cases model - Israel

New_tests_t	1.396
$\log(New\_vaccinated_{(t-14)})$	5.595
$\log(New\_vaccinated_{(t-30)})$	6.457
$\log(New\_fully\_vaccinated_{(t-14)})$	1.950
$\log(New\_fully\_vaccinated_{(t-30)})$	4.672
$Average\_temperature_t$	1.291
$Stringency\_index_{(t-14)}$	8.662
$Stringency\_index_{(t-30)}$	10.180

Table A.13: Variance inflation factors - new deaths model - Israel

$\log(New\_cases_t)$	2.303
$\log(New\_cases_{(t-7)})$	2.468
$\log(New\_cases_{(t-14)})$	2.172
$\log(New\_vaccinated_{(t-14)})$	5.776
$\log(New\_vaccinated_{(t-30)})$	7.178
$\log(New\_fully\_vaccinated_{(t-14)})$	1.893
$\log(New\_fully\_vaccinated_{(t-30)})$	4.780
$Average\_temperature_t$	1.399
$Stringency\_index_{(t-14)}$	8.741
$Stringency\_index_{(t-30)}$	10.839

Table A.14: Variance inflation factors - hospitalization model - Israel

$\log(New\_cases_t)$	2.326
$\log(New\_cases_{(t-7)})$	2.479
$\log(New\_cases_{(t-14)})$	2.174
$\log(New\_vaccinated_{(t-14)})$	5.735
$\log(New\_vaccinated_{(t-30)})$	7.189
$\log(New\_fully\_vaccinated_{(t-14)})$	1.887
$\log(New\_fully\_vaccinated_{(t-30)})$	4.776
$Average\_temperature_t$	1.399
$Stringency\_index_{(t-14)}$	8.802
$Stringency\_index_{(t-30)}$	10.885

Table A.15: Estimation results of new cases model - Czechia

	Dependent variable:	
	$\log(New\_cases_t)$	
	OLS	*Robust SE*
	(1)	(2)
$\log(New\_tests_t)$	0.473^{***}	0.473^{***}
	(0.079)	(0.084)
$\log(New\_vaccinated_{(t-14)})$	-0.112^{**}	-0.112
	(0.056)	(0.071)
$\log(New\_vaccinated_{(t-30)})$	-0.163^{**}	-0.163^{**}
	(0.064)	(0.077)
$\log(New\_fully\_vaccinated_{(t-14)})$	-0.044	-0.044
	(0.040)	(0.035)
$\log(New\_fully\_vaccinated_{(t-30)})$	-0.061	-0.061
	(0.062)	(0.079)
$Average\_temperature_t$	-0.134^{***}	-0.134^{***}
	(0.009)	(0.019)
$Stringency\_index_{(t-14)}$	0.024^{**}	0.024^{*}
	(0.009)	(0.013)
$Stringency\_index_{(t-30)}$	0.002	0.002
	(0.009)	(0.011)
Constant	-2.971^{***}	-2.971^{***}
	(0.381)	(0.704)
Observations	520	
R^2	0.572	
Adjusted R^2	0.566	
Residual Std. Error	1.200 (df = 511)	
F Statistic	85.440^{***} (df = 8; 511)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.16: Estimation results of new deaths model - Czechia

	Dependent variable:	
	$\log(New\_deaths_t)$	
	OLS	*Robust SE*
	(1)	(2)
$\log(New\_cases_t)$	0.713***	0.713***
	(0.027)	(0.037)
$\log(New\_cases_{(t-7)})$	−0.081***	−0.081***
	(0.028)	(0.031)
$\log(New\_cases_{(t-14)})$	−0.097***	−0.097***
	(0.026)	(0.035)
$\log(New\_vaccinated_{(t-14)})$	0.129***	0.129***
	(0.031)	(0.046)
$\log(New\_vaccinated_{(t-30)})$	0.101***	0.101
	(0.037)	(0.076)
$\log(New\_fully\_vaccinated_{(t-14)})$	−0.072***	−0.072**
	(0.022)	(0.036)
$\log(New\_fully\_vaccinated_{(t-30)})$	−0.089**	−0.089
	(0.035)	(0.064)
$Average\_temperature_t$	−0.048***	−0.048***
	(0.007)	(0.011)
$Stringency\_index_{(t-14)}$	0.032***	0.032***
	(0.005)	(0.007)
$Stringency\_index_{(t-30)}$	−0.015***	−0.015**
	(0.005)	(0.007)
Constant	−5.621***	−5.621***
	(0.196)	(0.369)
Observations	511	
R^2	0.847	
Adjusted R^2	0.844	
Residual Std. Error	0.677 (df = 500)	
F Statistic	276.833*** (df = 10; 500)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.17: Estimation results of hospitalization model - Czechia

	Dependent variable:	
	$\log(Hospitalizations_t)$	
	OLS	*Robust SE*
	(1)	(2)
$\log(New\_cases_t)$	0.250^{***}	0.250^{***}
	(0.027)	(0.061)
$\log(New\_cases_{(t-7)})$	0.078^{***}	0.078^{**}
	(0.028)	(0.037)
$\log(New\_cases_{(t-14)})$	-0.005	-0.005
	(0.027)	(0.042)
$\log(New\_vaccinated_{(t-14)})$	0.122^{***}	0.122^{**}
	(0.031)	(0.061)
$\log(New\_vaccinated_{(t-30)})$	0.134^{***}	0.134
	(0.037)	(0.109)
$\log(New\_fully\_vaccinated_{(t-14)})$	-0.078^{***}	-0.078
	(0.022)	(0.057)
$\log(New\_fully\_vaccinated_{(t-30)})$	-0.172^{***}	-0.172^{*}
	(0.036)	(0.097)
$Average\_temperature_t$	-0.106^{***}	-0.106^{***}
	(0.007)	(0.016)
$Stringency\_index_{(t-14)}$	0.034^{***}	0.034^{***}
	(0.005)	(0.012)
$Stringency\_index_{(t-30)}$	-0.028^{***}	-0.028^{**}
	(0.005)	(0.013)
Constant	-0.908^{***}	-0.908^{*}
	(0.198)	(0.507)
Observations	516	
R^2	0.820	
Adjusted R^2	0.817	
Residual Std. Error	0.684 (df = 505)	
F Statistic	230.800^{***} (df = 10; 505)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.18: Estimation results of new cases model - Israel

	Dependent variable:	
	$\log(New\_cases_t)$	
	OLS	*Robust SE*
	(1)	(2)
New_tests_t	0.063^{***}	0.063
	(0.012)	(0.050)
$\log(New\_vaccinated_{(t-14)})$	-0.334^{***}	-0.334^{**}
	(0.077)	(0.148)
$\log(New\_vaccinated_{(t-30)})$	-0.577^{***}	-0.577^{***}
	(0.083)	(0.175)
$\log(New\_fully\_vaccinated_{(t-14)})$	0.035	0.035
	(0.044)	(0.052)
$\log(New\_fully\_vaccinated_{(t-30)})$	0.194^{**}	0.194
	(0.076)	(0.130)
$Average\_temperature_t$	-0.053^{***}	-0.053^{*}
	(0.013)	(0.030)
$Stringency\_index_{(t-14)}$	0.019	0.019
	(0.012)	(0.034)
$Stringency\_index_{(t-30)}$	0.064^{***}	0.064^{*}
	(0.014)	(0.036)
Constant	-6.350^{***}	-6.350^{***}
	(0.602)	(1.585)
Observations	532	
R^2	0.371	
Adjusted R^2	0.361	
Residual Std. Error	1.805 (df = 523)	
F Statistic	38.505^{***} (df = 8; 523)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.19: Estimation results of new deaths model - Israel

	Dependent variable:	
	$\log(New\_deaths_t)$	
	OLS	*Robust SE*
	(1)	(2)
$\log(New\_cases_t)$	0.367^{***}	0.367^{***}
	(0.029)	(0.038)
$\log(New\_cases_{(t-7)})$	0.035	0.035
	(0.030)	(0.034)
$\log(New\_cases_{(t-14)})$	-0.045	-0.045
	(0.028)	(0.038)
$\log(New\_vaccinated_{(t-14)})$	0.068	0.068
	(0.044)	(0.047)
$\log(New\_vaccinated_{(t-30)})$	0.100^{**}	0.100
	(0.048)	(0.070)
$\log(New\_fully\_vaccinated_{(t-14)})$	0.026	0.026
	(0.024)	(0.023)
$\log(New\_fully\_vaccinated_{(t-30)})$	-0.068	-0.068
	(0.043)	(0.051)
$Average\_temperature_t$	-0.003	-0.003
	(0.008)	(0.011)
$Stringency\_index_{(t-14)}$	0.048^{***}	0.048^{***}
	(0.007)	(0.010)
$Stringency\_index_{(t-30)}$	-0.037^{***}	-0.037^{***}
	(0.008)	(0.010)
Constant	-6.833^{***}	-6.833^{***}
	(0.374)	(0.604)
Observations	522	
R^2	0.494	
Adjusted R^2	0.484	
Residual Std. Error	1.001 (df = 511)	
F Statistic	49.929^{***} (df = 10; 511)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.20: Estimation results of hospitalization model - Israel

	Dependent variable:	
	$\log(Hospitalization_t)$	
	OLS	*Robust SE*
	(1)	(2)
$\log(New\_cases_t)$	0.147***	0.147***
	(0.013)	(0.041)
$\log(New\_cases_{(t-7)})$	0.079***	0.079***
	(0.013)	(0.016)
$\log(New\_cases_{(t-14)})$	0.032***	0.032
	(0.012)	(0.029)
$\log(New\_vaccinated_{(t-14)})$	−0.114***	−0.114***
	(0.019)	(0.043)
$\log(New\_vaccinated_{(t-30)})$	−0.048**	−0.048
	(0.021)	(0.046)
$\log(New\_fully\_vaccinated_{(t-14)})$	0.033***	0.033
	(0.010)	(0.022)
$\log(New\_fully\_vaccinated_{(t-30)})$	−0.006	−0.006
	(0.019)	(0.034)
$Average\_temperature_t$	−0.020***	−0.020*
	(0.003)	(0.010)
$Stringency\_index_{(t-14)}$	0.034***	0.034***
	(0.003)	(0.009)
$Stringency\_index_{(t-30)}$	−0.010***	−0.010
	(0.003)	(0.007)
Constant	−3.081***	−3.081***
	(0.161)	(0.535)
Observations	528	
R^2	0.763	
Adjusted R^2	0.759	
Residual Std. Error	0.435 (df = 517)	
F Statistic	166.888*** (df = 10; 517)	
Note:	*p<0.1; **p<0.05; ***p<0.01	
	Standard errors in parentheses	

Table A.21: Number of deaths in Czechia

Age groups	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0-14	379	340	354	365	399	382	376	387	339	323	396
15-44	2 914	2 868	2 672	2 725	2 665	2 593	2 711	2 644	2 567	2 751	3 139
45-64	16 168	15 678	14 444	14 418	13 460	13 540	13 278	12 944	12 914	15 434	15 369
65-74	18 763	20 017	19 609	21 236	20 797	21 756	22 506	21 908	22 758	27 899	26 975
75-84	27 193	27 218	25 210	26 859	24 727	25 776	26 214	26 526	28 504	34 453	38 288
85+	23 474	24 672	23 945	27 297	25 422	28 014	28 320	28 120	29 325	30 838	35 721
Sum	88 891	90 793	86 234	92 900	87 470	92 061	93 405	92 529	96 407	111 698	119 888

Table A.22: Estimated deaths in Czechia

Age groups	2020	2021	2022
0-14: Forecasted	392	396	400
0-14: Real	339	323	396
15-44: Forecasted	2565	2529	2494
15-44: Real	2 567	2 751	3 139
45-64: Forecasted	12 193	11 738	11 282
45-64: Real	12 914	15 434	15 369
65-74: Forecasted	22 992	23 473	23 955
65-74: Real	22 758	27 899	26 975
75-84: Forecasted	25 673	25 553	25 432
75-84: Real	28 504	34 453	38 288
85+: Forecasted	29 431	30 158	30 886
85+: Real	29 325	30 838	35 721
Sum: Forecasted	93 245	93 847	94 449
Sum: Real	96 407	111 698	119 888

Table A.23: Excess of deaths in Czechia

Age groups	2020	2021	2022
65-74	-234	4 426	3 020
75-84	2 831	8 900	12 856
85+	-106	680	4 835

Table A.24: Life Expectancy in Czechia

Sex	Year	65	70	75	80	85	90	95	100	105
Men	2019	16.3	13.0	10.0	7.4	5.2	3.6	2.5	1.8	1.5
	2020	15.2	12.0	9.1	6.6	4.6	3.2	2.3	1.7	1.4
	2021	14.5	11.4	8.7	6.5	4.6	3.3	2.3	1.8	1.5
Women	2019	19.9	15.9	12.2	8.8	6.1	4.0	2.6	1.8	1.5
	2020	19.2	15.1	11.5	8.2	5.6	3.7	2.5	1.7	1.4
	2021	18.6	14.8	11.3	8.2	5.6	3.8	2.5	1.8	1.4

Table A.25: Life expectancy by age groups in Czechia

	Men			Women			Average		
Year	65-74	75-84	85+	65-74	75-84	85+	65-74	75-84	85+
2019	13.1	7.5	2.9	16.0	9.0	3.2	14.5	8.3	3.1
2020	12.1	6.8	2.7	15.3	8.4	3.0	13.7	7.6	2.8
2021	11.6	6.6	2.7	14.9	8.4	3.0	13.2	7.5	2.9
2022	11.6	6.6	2.7	14.9	8.4	3.0	13.2	7.5	2.9

Table A.26: Average pensions in Czechia

Year	Average pension
2010	10 123
2011	10 552
2012	10 778
2013	10 970
2014	11 075
2015	11 348
2016	11 460
2017	11 850
2018	12 418
2019	13 468
2020	14 479
2021	15 425
2022	18 061
2023	20 188
Predictions	
2024	20 996
2025	21 625
2026	22 274
2027	22 740
2028	23 174
2029	23 638
2030	24 110

Table A.27: Vaccines' expenditures in Czechia

in mil. CZK	Pfizer	Moderna	Johnson & Johnson	AstraZeneca	Novavax
2021	5 227	1 441.4	161.2	52.7	0
2022	7 147	532.8	68.3	82	126.1
Sum	12 374	1 974.2	229.5	134.7	126.1
Total sum			14 838.5		

Table A.28: Pension distribution excess deaths for group 65-74 - Czechia

in mil. CZK	2022	2023	2024	2025	2026	2027	2028	2029	2030
Deaths from 2021	79.933	89.346	92.920	95.708	98.579	100.55	102.561	104.613	106.705
Sum					870.914				
Deaths from 2022	-	60.968	63.407	65.309	67.268	68.614	69.986	71.386	72.813
Sum					539.75				

Table A.29: Pension distribution excess deaths for group 75-84 - Czechia

in mil. CZK	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
Deaths from 2020	43.667	51.129	57.151	59.437	61.22	63.057	64.318	32.802	-	-
Sum					432.78					
Deaths from 2021	-	160.751	179.682	186.869	192.476	198.25	202.215	206.259	105.192	-
Sum					1 431.693					
Deaths from 2022	-	-	259.536	269.917	278.015	286.355	292.083	297.924	303.883	154.98
Sum					2 142.693					

Table A.30: Pension distribution excess deaths for group 85+ - Czechia

in mil. CZK	2022	2023	2024	2025	2026
Deaths from 2021	12.278	13.724	14.273	-	-
Sum			40.274		
Deaths from 2022	-	97.619	101.524	104.569	-
Sum			303.712		

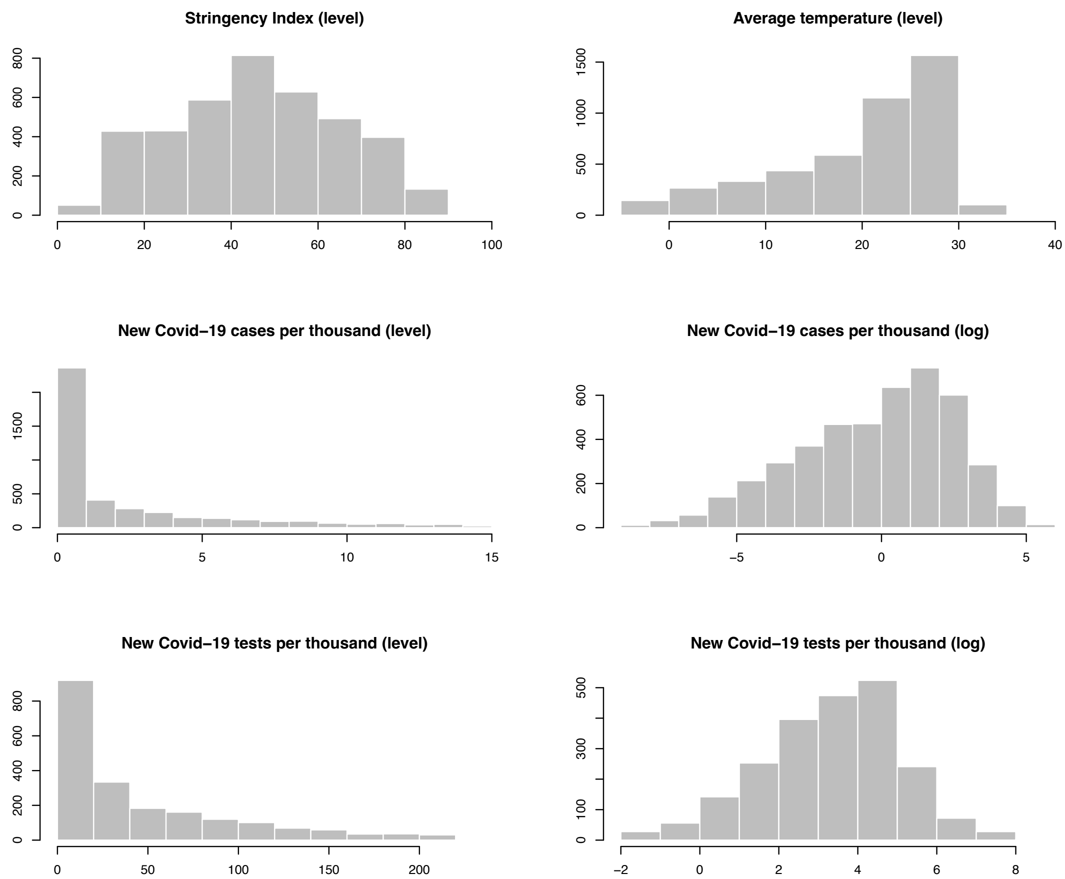
Table A.31: Pension distribution - Czechia

in mil. CZK	65-74	75-84	85+
Deaths from 2020	-	432.780	-
Deaths from 2021	870.914	1 431.693	40.274
Deaths from 2022	539.75	2 142.693	303.712
Sum	1 410.7	4 007.2	344
Total sum		5 761.9	

Appendix B

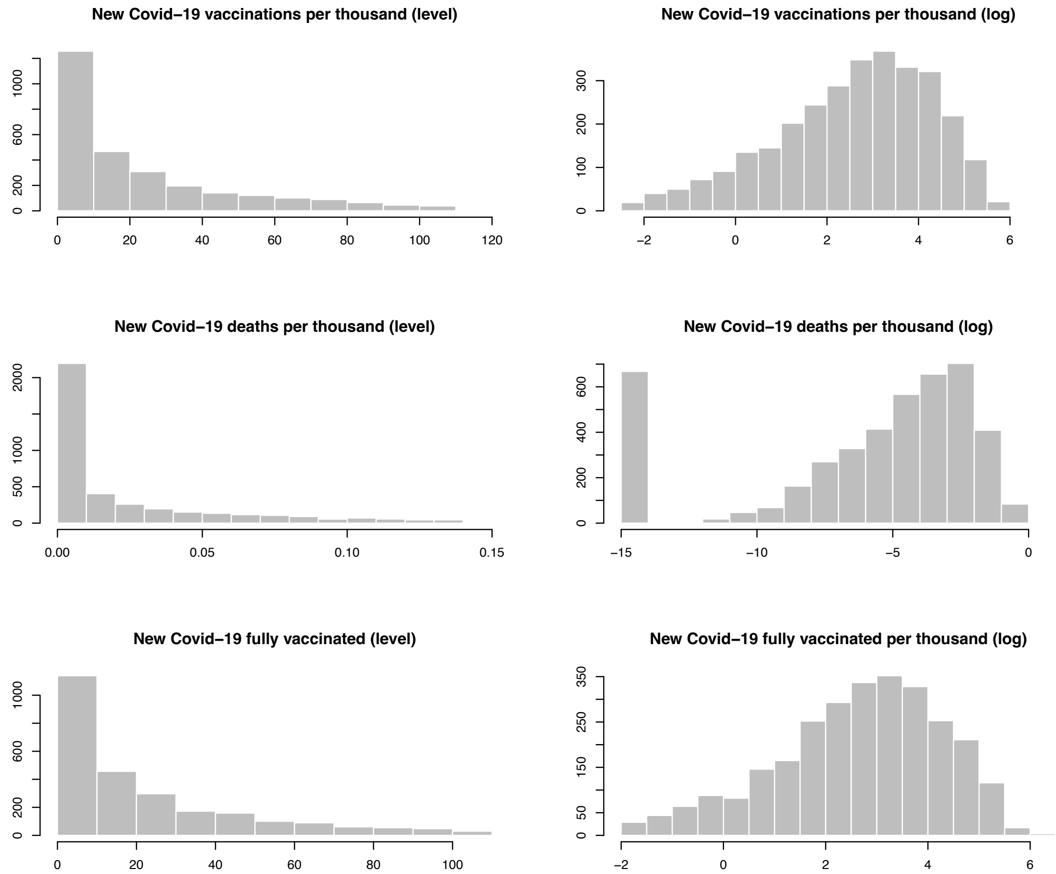
Figures

Figure B.1: Histograms world data - part 1



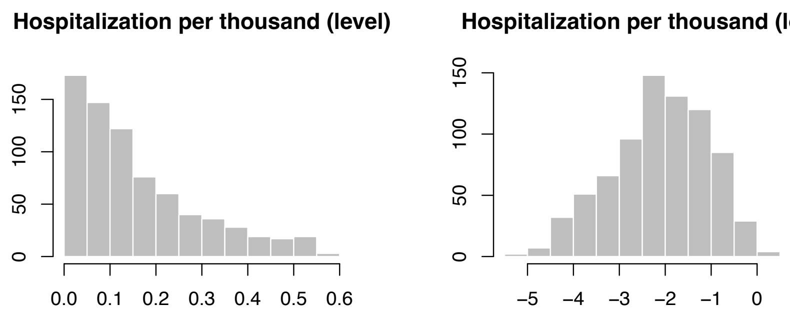
Source: Authors' computations based on the compiled data set

Figure B.2: Histograms world data - part 2



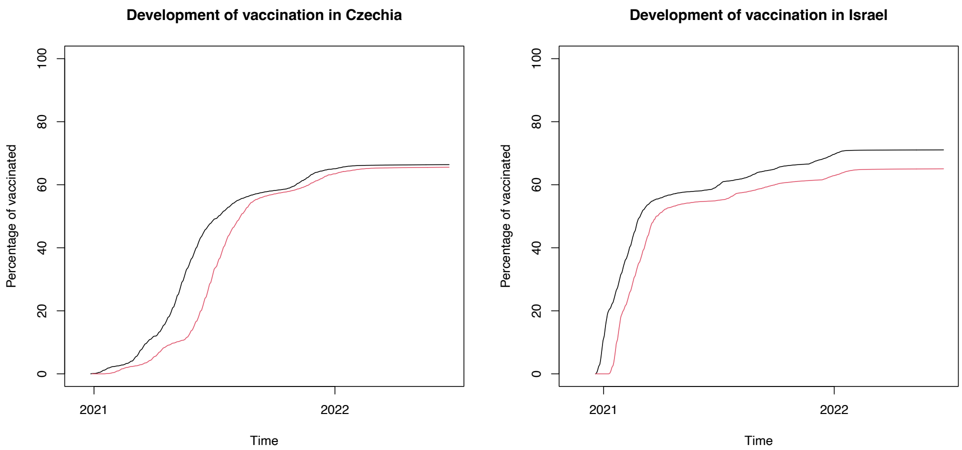
Source: Authors' computations based on the compiled data set

Figure B.3: Histograms world data - part 3



Source: Authors' computations based on the compiled data set

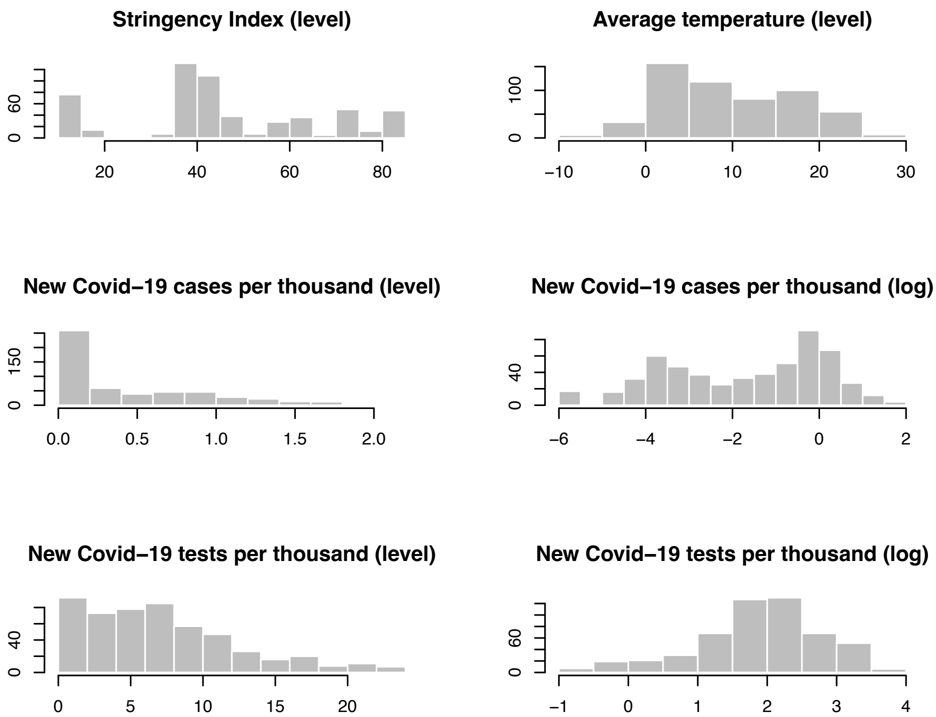
Figure B.4: Vaccination development



Source: Authors' computations based on the compiled data set

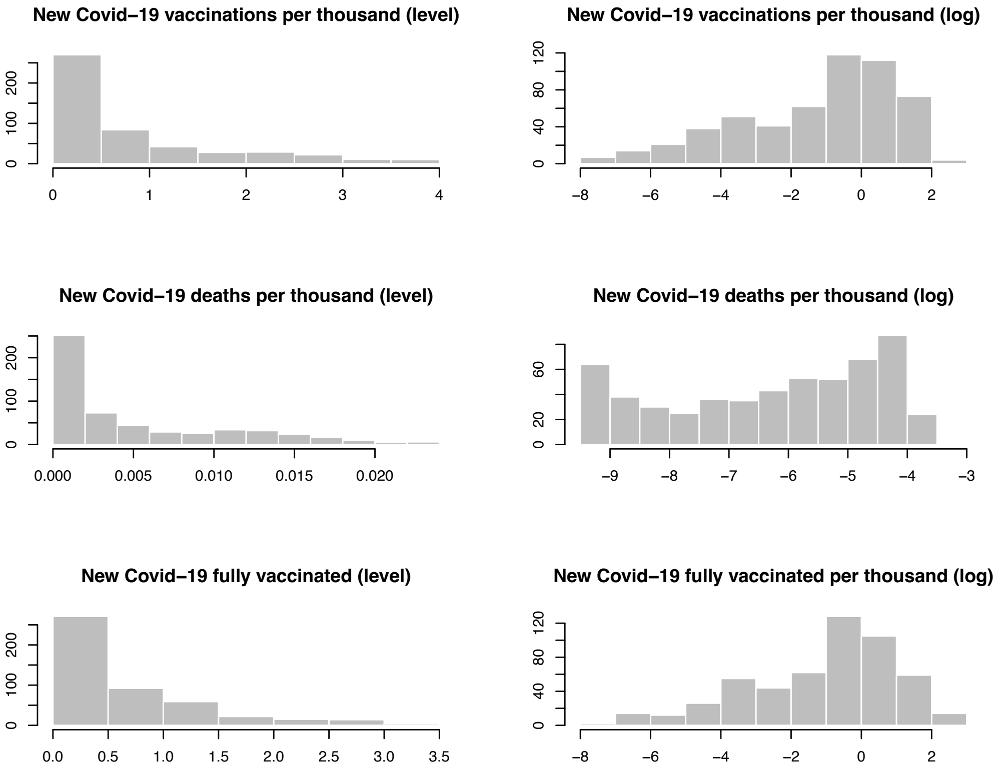
Black line - first vaccination x Red line - fully vaccinated people

Figure B.5: Histograms Czechia data - part 1



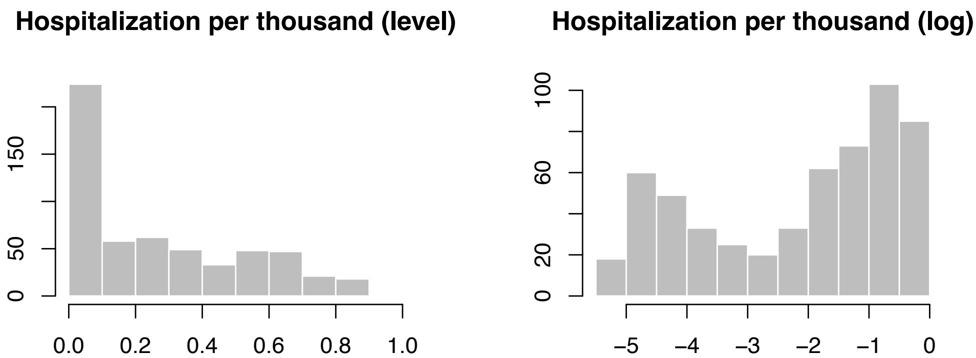
Source: Authors' computations based on the compiled data set

Figure B.6: Histograms Czechia data - part 2



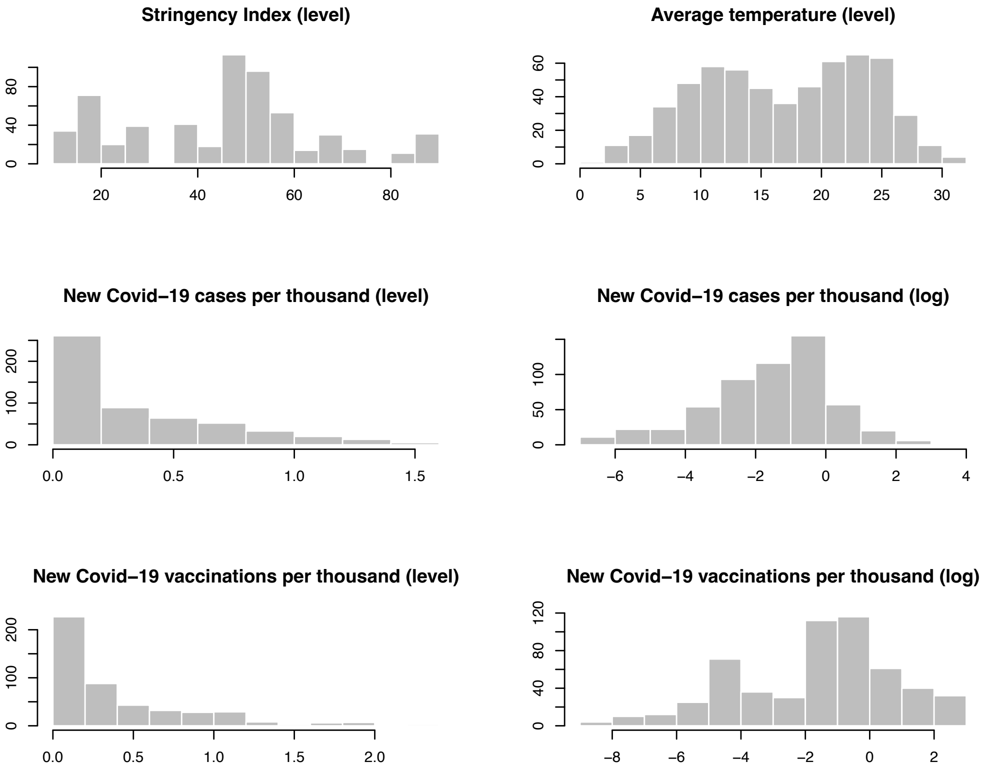
Source: Authors' computations based on the compiled data set

Figure B.7: Histograms Czechia data - part 3



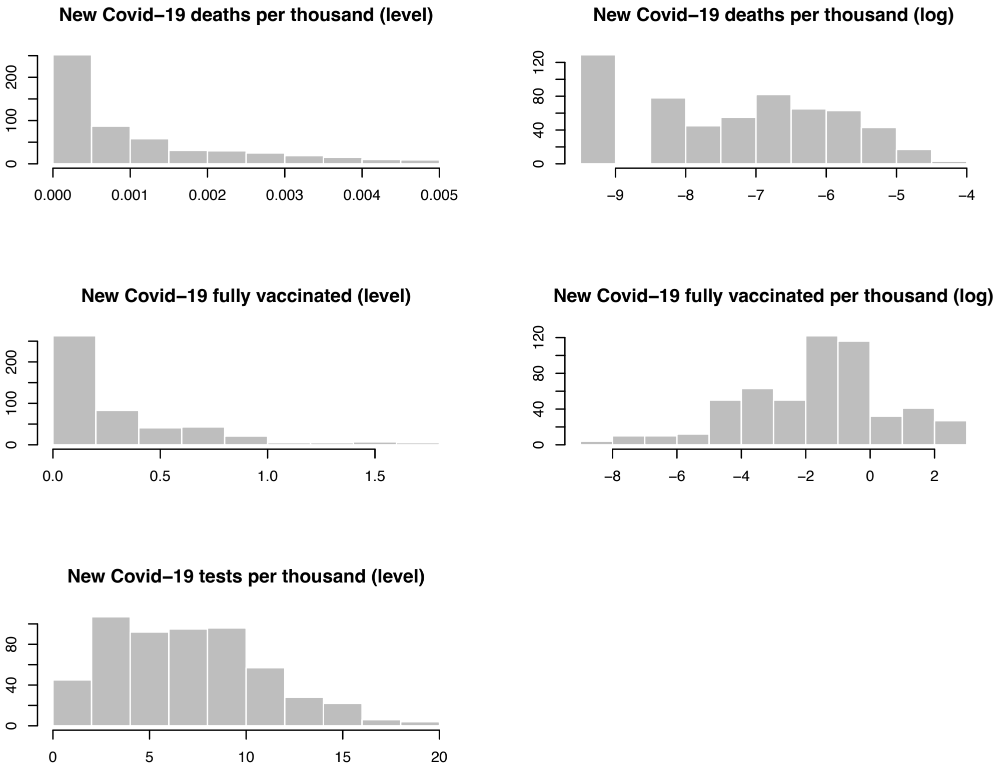
Source: Authors' computations based on the compiled data set

Figure B.8: Histograms Israel data - part 1



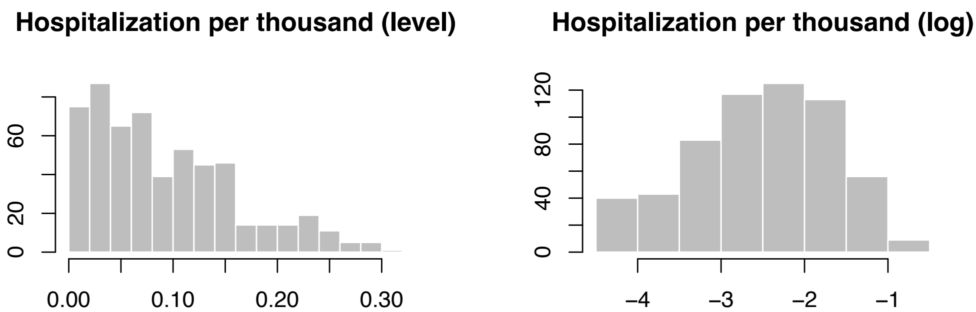
Source: Authors' computations based on the compiled data set

Figure B.9: Histograms Israel data - part 2



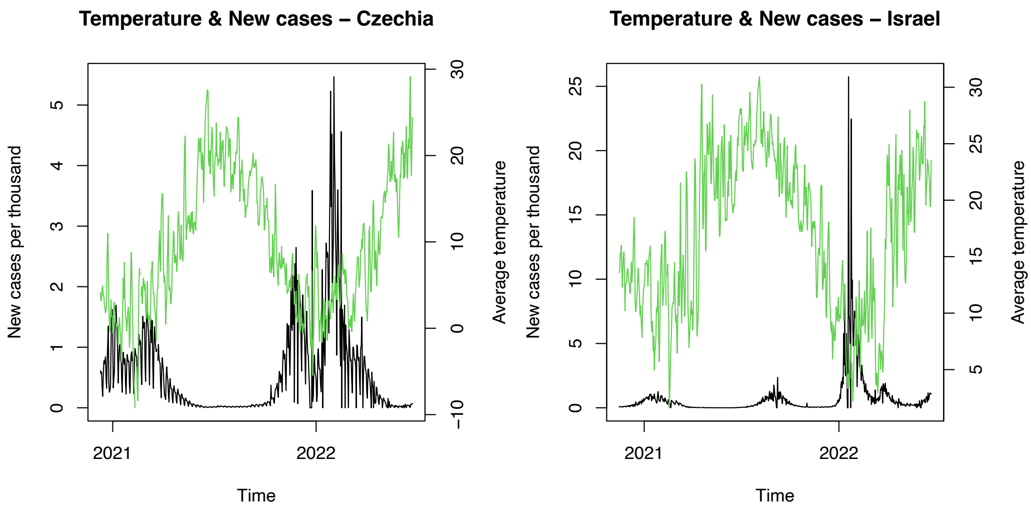
Source: Authors' computations based on the compiled data set

Figure B.10: Histograms Israel data - part 3



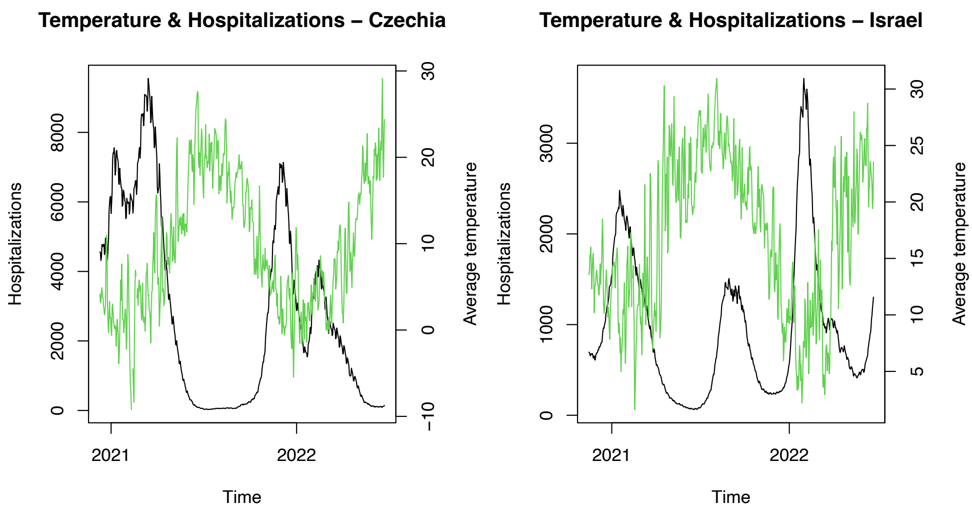
Source: Authors' computations based on the compiled data set

Figure B.11: Comparison of new cases and average temperature in time



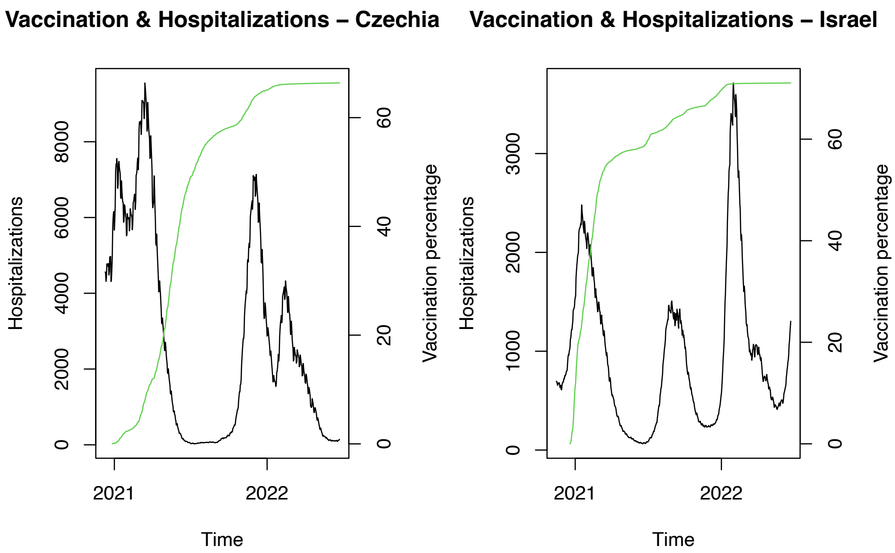
Source: Authors' computations based on the compiled data set

Figure B.12: Comparison of hospitalization and average temperature in time



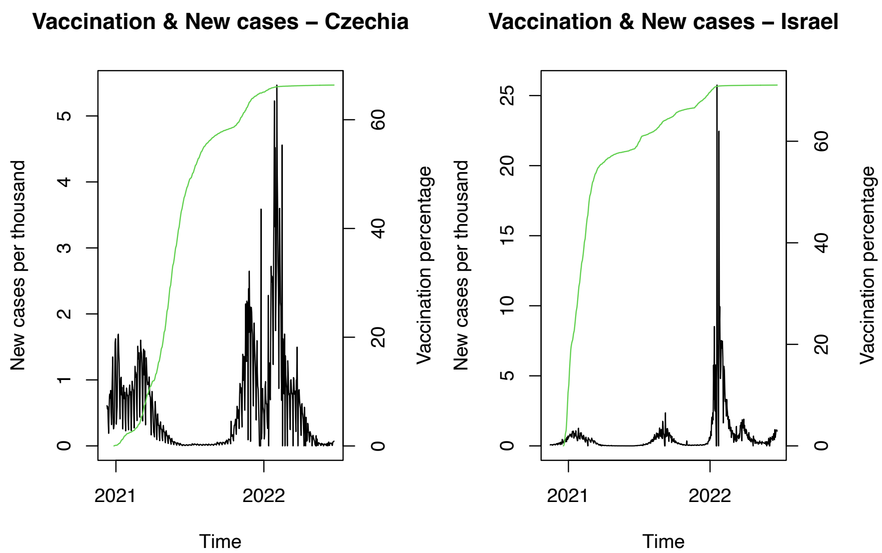
Source: Authors' computations based on the compiled data set

Figure B.13: Comparison of hospitalization and vaccination percentage in time



Source: Authors' computations based on the compiled data set

Figure B.14: Comparison of new cases and vaccination percentage in time



Source: Authors' computations based on the compiled data set

Figure B.15: Development of deaths - Czechia