INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

WORKFORCE COMPOSITION EFFECTS ON COVID-19 SPREAD: A SIR MODEL AND AN APPLICATION TO THE CZECH REGIONS

Oliver Škultéty

IES Working Paper 15/2025

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Škultéty O. (2025): " Workforce Composition Effects on Covid-19 Spread: a SIR model and an Application to the Czech Regions " IES Working Papers 15/2025. IES FSV. Charles University.

Workforce Composition Effects on Covid-19 Spread: a SIR model and an Application to the Czech Regions

Oliver Škultéty
Institute of Economic Studies, Charles University, Prague
E-mail: skultety.oliver@gmail.com

August 2025

Abstract:

One of the main types of government intervention aimed at fighting the COVID-19 pandemic were the social distancing measures. This article presents a novel approach to examining the relationship between social distancing and workforce composition. The model used is a SIRD model with dynamically optimizing individuals. The model is estimated for each Czech region and calibrated for the ’social distancing measure’ so that the model predictions best fit the data. This measure is then compared in a regression with the share of skilled labour in each region. This article finds that a significant correlation between social distancing preferences and workforce composition.

JEL: I12
Keywords: SIR, social distancing, Covid-19, workforce composition

1. Introduction

The COVID-19 pandemic has been a major crises between the years 2020 and 2022. The pandemic took millions of lives and caused various fiscal and financial crises. The pandemic was heterogeneous in it's effect on countries and even regions within countries. For instance, in the early waves of the pandemic, USA reported around 18 times more deaths per capita than Australia (Bilinski & Emanuel 2020). One of the main aspects causing this heterogeneity is the difference between government intervention. There are many types of such interventions. One type tries to affect transmissibility. These would be measures such as mandatory wearing of face masks. Other tried to enforce social distancing, either by quarantine, lock-downs or traveling restrictions.

A main factor influencing effectiveness of social distancing measures is social distancing compliance. Individuals try to optimize their private benefits during the pandemic. If the utility from social activity is high enough, a rational individual would violate government mandated social distancing and they would engage with others. There are numerous factors influencing social distancing compliance, such as news media (Simonov *et al.* 2020), political beliefs (Painter & Qiu 2020), social media (Gualda *et al.* 2021), mental capacities (Xie *et al.* 2020) and moral beliefs (Murphy *et al.* 2020) (Barrios *et al.* 2021).

Some professions are better able to be carried out remotely than other. It has been shown that COVID-19 disproportionately affected employment of workers in those professions that are largely unable to be done from home and that involve close physical contact (Mongey *et al.* 2021). The goal of this article is to find how is social distancing affected by workforce composition.

There are papers examining the link between social distancing and income (Weill *et al.* 2020), socioeconomic status (Huang *et al.* 2020) or proportion of essential workers (Garnier *et al.* 2021). All of these papers utilized mobile device location data to measure social distancing. The method used in this article doesn't require such data. Instead, we rely on the model to correctly predict social activity.

The used model is a SIRD model for a well-mixed population modified with social distancing which was proposed by Farboodi *et al.* (2021). In the model, individuals dynamically optimize, choosing the optimal level of social activity. Their overall utility is determined by their utility from engaging with others minus the disutility from potentially getting COVID-19. We estimate a measure of social distancing by estimating the model between the 30.8.2021 and 12.5.2022 for each Czech region separately, choosing such a k/U that the resulting predictions of new daily infections best fit the data obtained from MZČR (Ministry of Health of the Czech Republic). Using this method, we acquire an estimate of measure of social distancing for each region. This means that we can compare regions in terms of the preferences of their population over social activity compared to health risks.

Finally, we can regress Share of Skilled Labour on a measure of social distancing . We find a positive correlation between the social distancing measure and the Share of Skilled Labour variable, even after controlling for vaccination and population density, which is consistent with results of the literature (Garnier *et al.* (2021), Huang *et al.* (2020),Weill *et al.* (2020)) concerning the relationship between workforce composition and social distancing. There are two main contributions of this article. First of all, we utilize a novel approach, estimating social activity from the model and we find results consistent with the literature. The other contribution is that the study is concerned with the Czech Republic, while other studies primarily concentrated on the USA. This article presents a method of investigating the topic in countries where data on social distancing of individuals is difficult to come by.

The article starts with chapter 2 which introduces the epidemiological model used. The chapter 3 contains a discussion of imports of various parameters necessary for the model estimation and the estimation approach. In chapter 4, the results of modelling are discussed as well as various limitations of the approach. Chapter 5 is the conclusion with discussion of possible future work expanding on the approach in this article.

2. Epidemiological Model

To model social distancing compliance, individuals will optimize with respect to a couple of variables. Each individual derives some utility from engaging with other people, for instance by meeting in restaurants, in the theatre or by traveling. On the other hand, they also risk getting the disease, which poses disutility by spending a limited time sick or dying. The possibility of catching the disease depends on the number of people engaging in social activity, the number of infected people in the population and the transmission rate. These variables coupled with the standard SIR model enable us to modify SIR with social distancing compliance.

We begin constructing the modified SIR model by a couple of simplifying assumptions for modelling individuals decisions:

- The population is well mixed and recovered never lose immunity.
- We assume that individual subjects don't know whether they are infected or not, but they know whether they are recovered.
- People also have possibility to die from the disease, denoted by π. We define $D(t)$ to describe the proportion of individuals who died due to the disease
- Individual social activity is described by a function $a(t)$. Utility in our model will only depend on $a(t)$ and we will denote it $u(a(t))$. This is a major simplification, since utility from social activity should depend on social activity of others.

The basic SIR model contains 3 compartments: Susceptible, Infected and Recovered. It describes a course of a pandemic using a system of differential equations. Susceptible are people who aren't infected, but don't have any immunity - they can become infected. Infected are the people who have the disease and can infect others. Recovered are those who were infected but are no longer or who are no longer infectious.

For our model, we will modify the SIRD model. The only difference between a SIR model described and the SIRD model is the compartment $D(t)$, which denotes deceased individuals. The SIRD model can be described using following dynamics:

$$S'(t) = -\beta S(t)I(t)$$

$$I'(t) = \beta S(t)I(t) - \gamma I(t)$$

$$R'(t) = (1 - \pi)\gamma I(t)$$

$$D'(t) = \pi\gamma I(t)$$

Our model uses the SIRD model coupled with the 'changing the transmission rate' approach for introducing social distancing into the SIR model, where compared to the SIRD model, the transmission rate β depends on time. The core dynamics of the model are:

$$S'(t) = -A^2(t)\beta S(t)I(t)$$

$$I'(t) = A^2(t)\beta S(t)I(t) - \gamma I(t)$$

$$R'(t) = (1 - \pi)\gamma I(t)$$

$$D'(t) = \pi\gamma I(t)$$

$A^2(t)$ serves a similar purpose as β in the models described previously. When compartments are interacting, the amount of interactions is multiplied by each compartments social activity. Because people don't know whether they are susceptible or infected, both compartments' social activity will be the same. A person's decision on their social activity is also influenced by the disutility of contracting the disease. This disutility can be caused by various aspects of the disease, such as the discomfort of experiencing symptoms of the disease, long lasting effects on the disease (in the case of COVID-19 the so-called "Post-COVID syndrome") and finally the potential death. In the model in (Farboodi *et al.* 2021), the only negative effect considered is the possibility of death. In our model, the disutility from contracting COVID-19 is estimated, which means that we should use the extended definition, as we don't need to compute the expected value of contracting COVID-19. We also aren't able to compute it exogenously, as we would need to multiply probability of dying from COVID-19 π with the statistical value of life, which is not estimated for each Czech region.

The beliefs regarding individual's state are given by variables n_i, n_s, n_r. These beliefs are rational, meaning that they accurately represent the probability of an individual of being in each state. The contribution of an individual towards the overall pandemic is negligible in a person's decision making in the model. All individuals know in advance how will the pandemic progress and they optimize accordingly. Each individual should go through at most two states. In the first, they are either susceptible or infectious. In this state, they face the risk of getting infected with COVID-19 and they will choose their social activity accordingly. The other state is when they know they are recovered. In this state, the individual simply chooses $a(t)$ that maximizes $u(a(t))$.

Last component we need is time discounting. The decision-making process will be modelled as if the individuals optimize until infinity, but there is a heavy time-discounting. The time-discounting is caused by the possibility of developing better cures in the future as well as by the uncertainty in the state of the pandemic in longer time frame. The variables $S(t), I(t), R(t), D(t)$ are all divided by the population N for ease of interpretation of results of numerical computation.

The function that individuals optimize is modelled in a following way.

- We take expected values of being in each of the two states, meaning that we take probabilities of being in each state and multiply them by their respective utility. For the recovered state, the expected value is $n_r u(a(t))$. The individual doesn't face any disutility, so they optimize their social activity freely.

- For the other state, the individual's expected utility is $(n_s(t)+n_i(t))u(a(t))-\gamma n_i(t)k$. The term $(n_s(t)+n_i(t))u(a(t))$ is the probability of being susceptible of infected times utility from social activity. The term $\gamma n_i(t)k$ is the probability of being infected times the probability of leaving the infected compartment, denoted by γ times expected value of leaving the infected compartment, which is probability of dying times statistical value of life.

- All of the terms are multiplied by $e^{-\rho t}$, which is the time discounting.

We can now set-up the integral that all individuals try to maximize:

$$\int_0^\infty [(n_s(t) + n_i(t))u(a(t)) + n_r(t)u(a(t)) - \gamma n_i(t)k]e^{-\rho t}\,\mathrm{d}t \tag{2.1}$$

$$n_s'(t) = -\beta a(t) n_s(t) A(t) I(t) \tag{2.2}$$

$$n_i'(t) = \beta a(t) n_s(t) A(t) I(t) - \gamma n_i(t) \tag{2.3}$$

$$n_r'(t) = (1-\pi)\gamma n_i(t) \tag{2.4}$$

Equations (2.2), (2.3) and (2.4) are similar to equations describing core dynamics of the model and their logic is similar as well. An individual's probability of getting infected is given by their amount of social activity $a(t)$, the transmission rate and their probability of being susceptible. In one time unit t, interactions are given by the multiple of my social activity $a(t)$ times the amount of infected $I(t)$ times their social activity $A(t)$. In order to get infected, an individual needs to be susceptible, so we multiply the expression by $n_s(t)$. We also multiply the expression by the probability of transmitting the disease β. Using the same logic, we get the equations (2.3) and (2.4).

Using these expressions, we can derive the system of differential equations that we will try to estimate. The Hamiltonian for these equations is:

$$\begin{aligned} H(n_s(t), n_i(t), a(t), \lambda_s(t), \lambda_i(t)) = (n_s(t) + n_i(t))u(a(t)) - \gamma n_i(t) k \\ - \lambda_s(t)\beta a(t) n_s(t) A(t) I(t) + \lambda_i(t)(\beta a(t) n_s(t) A(t) I(t) - \gamma n_i(t)) \end{aligned} \tag{2.5}$$

We can use the fact that: "There are three necessary first order conditions for optimal control. First, the derivative of the Hamiltonian with respect to the control variable is zero... Additionally, the derivatives with respect to the state variables and are equal to minus the time derivative of the costate, with a correction for discounting" Farboodi *et al.* (2021). This gives us:

$$(n_s(t) + n_i(t))u'(a(t)) = (\lambda_s(t) - \lambda_i(t))\beta n_s(t) A(t) I(t) \tag{2.6}$$

$$\rho(\lambda_s(t) - \lambda_s'(t)) = u(a(t)) + (\lambda_i(t) - \lambda_s(t))\beta a(t) A(t) I(t) \tag{2.7}$$

$$\rho(\lambda_i(t) - \lambda_i'(t)) = u(a(t)) - \gamma(k + \lambda_i(t)) \tag{2.8}$$

And the transversality conditions:

$$\lim_{t\to-\infty} e^{t\rho}\lambda_s(t) n_s(t) = \lim_{t\to-\infty} e^{t\rho}\lambda_i(t) n_i(t) = 0 \tag{2.9}$$

These equations determine an individual's optimal level of social activity. Since we assume all individuals to be homogeneous, we can now substitute $A(t)$ for $a(t)$. In a society of homogeneous individuals, everybody has the same probability of being susceptible or infected. Everybody knows how many people are in each compartment, so the personal probability can be substituted for the amount of people in the compartment divided by the population: $n_s(t) = S(t)$ and $n_i(t) = I(t)$ Using this substitution, we get the system of differential equations that we estimate:

$$S'(t) = -\beta S(t)A^2(t)I(t) \tag{2.10}$$

$$I'(t) = \beta S(t)A^2(t)I(t) - \gamma I(t) \tag{2.11}$$

$$(S(t) + I(t))u'(A(t)) = (\lambda_s(t) - \lambda_i(t))\beta S(t)A(t)I(t) \tag{2.12}$$

$$\rho(\lambda_s(t) - \lambda_s'(t)) = u(A(t)) + (\lambda_i(t) - \lambda_s(t))\beta A^2(t)I(t) \tag{2.13}$$

$$\rho(\lambda_i(t) - \lambda_i'(t)) = u(A(t)) - \gamma(k + \lambda_i(t)) \tag{2.14}$$

$$\lim_{t\to-\infty} e^{t\rho}\lambda_s(t)S(t) = \lim_{t\to-\infty} e^{t\rho}\lambda_i(t)I(t) = 0 \tag{2.15}$$

3. Estimation Methodology

The data is taken from the MZČR website, which is the official website of the Ministry of Health of the Czech Republic. The data consists of the daily amounts of positively tested individuals, the amount of recovered and the amount of dead due to COVID-19, all three of which for each region. These statistics do not fully describe the state of the pandemic, as not all positive cases are tested. Some cases are without symptoms and symptomatic cases are often unreported. This means that for the model to be calibrated accurately, we need some estimate of the true number of cases.

A good way to gauge the true number of cases is by conducting seroprevalence studies. In this type of studies, a random sample from a population is taken and a presence of antibodies is measured. By using this technique repeatedly, the researchers are then able to estimate the number of real cases. What is extremely useful for SIR-based models of the pandemic is that the share of people without antibodies can be used as a good estimate of the share of susceptible population if we account for loss of immunity over time. In this paper, two results from seroprevalence studies are utilized. First, a study done by the Piler *et al.* (2022), focusing on the presence of COVID-19 antibodies in the population of the Czech Republic. The authors state that 51 % of responders from a sample of 19548 tested positive for COVID-19 antibodies in the months of February and March of 2021. Another essential figure used in the modeling is the ratio of 1.7 measured by a meta-analysis of multiple seroprevalence studies done by Bergeri *et al.* (2021). This figure describes the estimated ratio between people with antibodies and reported cases from mass testing in High Income European countries in Q4 of 2020 and Q1 of 2021. The 51 % is added with 1.7 times the number of cases between 30.3.2021 and 30.8.2021 to estimate the number of susceptible at the beginning of the wave we are interested in. The whole expression is then multiplied by 0.95 to estimate loss of immunity Dan *et al.* (2021). Over the whole estimation of the model, all real-world data is multiplied by the factor of 1.7.

In this article, the time frame which we try to estimate spans from the 30.8.2021 to 12.5.2022. Both dates are chosen as to describe the last two waves of the pandemic. Both these waves are taken together, as the time between the waves is relatively short and the amount of new daily cases does not decrease as significantly as in the previous waves. The start and end dates of the estimation are chosen arbitrarily as for the model predictions to best fit the data.

The model is described using parameters β, γ, k/U and ρ and variables $S(t)$, $I(t)$, $\lambda_s(t)$, $\lambda_i(t)$ and $A(t)$. The parameters β and γ are taken from previous literature. γ is taken as 1/7. This result comes from a study done by Lauer *et al.* (2020), who estimated an incubation period of 5.1 days. The parameter can be computed by taking one over the number of days between becoming infected and recovering. The model setting presumes that individuals don't know whether they are infectious or still susceptible. This assumption is unrealistic, especially when home tests are widely available. Not all people engage in self-quarantine when they start developing symptoms. The slightly lower γ aims to capture average behavior, as people engage in self-quarantine, but they don't get tested instantly after developing symptoms or they don't fully stop with all interaction with susceptible individuals.

β can be estimated through the R_0 and γ. The basic reproduction number represents the average number of new infections generated by one person at the beginning of a pandemic. Many studies try to estimate the measure with widely different results. Kong *et al.* (2021) find a R_0 for the Czech Republic of 1.96. Locatelli *et al.* (2021) estimate R_0 to be 2.2 (95 % CI: 1.9-2.6) for Western Europe. Meta-analysis of studies concerning the R_0 done by He *et al.* (2020) finds R_0 to be 3.15 and another meta-analysis written by Alimohamadi *et al.* (2020) finds mean R_0 from articles of 3.38. The choice of $R_0 = 2.5$ done in this article is based on the upper estimate of the two mentioned articles concerning Western Europe and the Czech Republic. The choice of higher than mean estimate is to take into account that new mutations of COVID -19 often have higher transition rates. An article by (van Oosterhout *et al.* 2021, pg. 1) states: "In recent weeks, several new strains of SARS-CoV-2, the causative agent of COVID-19, have emerged. These variants have evolved an increased transmission rate compared to the original strains". The formula for it in our model is $R_0 = \mathrm{A}(0)\mathrm{S}(0)\beta/\gamma$. At the beginning of a pandemic, A(0) and S(0) are close to 1, which means that we can compute β as $R_0 * \gamma$. β is therefore computed as $2.5/7 = 0.35714285714$.

The parameter k/U is unknown and it is estimated by minimizing the sum of least squares of predictions of the model about new daily infections and modified data. In the model from Farboodi *et al.* (2021) only k is present, so the utility function used in this article is a generalization of one of two of their proposed utility functions. The utility function used here is:

$$u(a(t)) = -U/2(a(t) - 1)^2 \tag{3.1}$$

The utility function is chosen in this way because it has a couple of useful properties. First of all, it is concave. We would expect that if one does not have any social contact, increasing their social activity by 0.1 would bring higher utility than if a person with a lots of social contact. Second, the function has a single peak. This enables us to easily find the maximum, which is at $a(t) = 1$. Thirdly, both low and excessive amounts of social activity negatively affect utility. This property also well describes reality, as people are both social creature but with a need for privacy. Last useful property is that all individuals will choose maximally social activity of 1. Choosing any more is dominated by choosing $a(t) = 1$, because the utility from the function $u(a(t))$ is higher and there is less chance of becoming infected.

The parameter k denotes the value of statistical life multiplied by the probability of dying from COVID -19. In the model, an individual compares the expecting value of contracting COVID -19 with utility from social activity. It is useful to use a more general version of a utility function proposed by Farboodi *et al.* (2021) in the approach in this paper. Without utilizing any data on social activity of individuals during the COVID-19 crisis, we can't choose any specific utility function. This function is chosen not only due to its appropriate properties, such as its concave shape, but also due to ease of use in computation. The U parameter is not estimated itself, but it is useful to keep it in the model to highlight the fact that the estimate of k/U will not only be affected by the statistical value of life and the probability of dying due to COVID -19. k/U is also affected by attitude towards social activity of people in the specific region, which can widely differ between regions.

The model is composed of a system of four differential equations (2.10, 2.11, 2.13, 2.14) and one equation for social activity (2.12). To estimate a unique solution to a system of differential equations, we need a starting condition, which is a point in time where we know values of all of the variables. Unfortunately, we don't have such a point in time in this model. This is caused by the co-state variables, for which we do not have real world data. The optimal solutions to the model follow equations (2.10-2.14) and the transversality conditions (2.15). The condition is satisfied if we pick stationary terminal values (derivatives of co-state variables equal zero). This choice of values of costate variables is not necessary for the transversality condition to hold (if the co-state variables would grow for instance linearly, the condition would still hold), but this choice of terminal values enables us to estimate the model.

By solving equations (2.13) and (2.14) using this condition, we obtain terminal values of λ_s and λ_i. Terminal time is denoted by T. It would be possible to use real-world data to estimate terminal $S(T)$ and $I(T)$, but this approach wouldn't often yield any useful results. The modelled pandemic is very sensitive to these two values, which means that the model might better fit the data if we calibrate for one of them. The calibration of parameters in this model is done by comparing the modelled new infections with modified data of real infections. Since we want vectors of values of these two variables to be as close as possible, we will choose the terminal value of $I(T)$ to be a number of infected individuals at the end date of the chosen wave and we will try to estimate the terminal $S(T)$ so that the model predictions best fit a modified real-world data.

The system of differential equations that describes the model does not have an analytical solution. This means that we have to utilize a numerical approach to approximating a solution to the model. The method used in this paper is the Euler method. The Euler method works by splitting the time into discrete units and it is described by the equation:

$$f(t+1) = f(t) + \Delta f'(t)$$

Given that we know $f'(t)$. Δ is the time between two discrete time units. If we want to increase accuracy, we split the time into smaller time intervals.

This method is used in a backwards-shooting algorithm. Since we don't have the initial values of the co-state variables, but we know the beginning state of S and I, we will have to start the estimation from the end of the pandemic and estimate backwards to the beginning. The algorithm takes the determined values of co-state variables and $I(T)$, and a guess of $S(T)$ and describes backwards in time a system of functions. The length of the pandemic is unknown, which means that we don't have a time where we know we have to stop the algorithm. The algorithm stops after we find time before terminal T where both S and I are sufficiently close to their real-world equivalents, namely when real world $S(t)$ and $I(t)$ rounded to three decimal places equals the modelled $S(t)$ and $I(t)$ to three decimal places. The two parameters that are estimated are k/U and $S(T)$.

To evaluate the model predictions for different values of k/U and $S(T)$, we compare predictions of new infections with modified real-world data using the Sum of Squares and we find the combination of k/U and $S(T)$ that attains the minimum value of the sum. We can't use a linear regression, because the modelled new daily infections' dependence on k/U and $S(T)$ can't be expressed linearly, due to the fact that the differential equations describing our model don't have an analytic solution. The optimization was done using the *optimize* function in R. The *optimize* function takes a function of one variable and finds a minimum of the function using a combination of the golden section search and the successive parabolic interpolation. It takes values of the function at different values of the variable and finds the value of its minimum using the aforementioned methods. The function used is the Sum of Squares of real-world data and the model predictions new daily infections. We can express it as:

$$f(N(T), k/U) = \sum_{t=1}^{255} (newI_d(t) - newI_m(t, N(T), k/U))^2 \tag{3.2}$$

In (3.1), the variable $newI_d(t)$ denotes the new daily infections taken from MZČR data multiplied by 1.7 at day t of our chosen time interval, which spans from 30.8.2021 to 12.5.2022. The variable $newI_m(t, N(T), k/U)$ captures the new daily infections predicted by the model at time t, given $N(T)$ and k/U. If we choose specific values of $N(T)$ and k/U, we get model prediction of daily new infections. The sum ends after 255 days, which is the length of the time interval. Since *optimize* finds a minimum of a single variable function, we first specify a new function, which uses *optimize* on the equation (3.1) for a given k/U. This new function takes k/U as an argument and it finds the $N(T)$ for which equation (3.1) attains a minimum. We use *optimize* on this new function to find k/U.

The model was estimated for each region. In figure 3.1, we can see that even with optimal parameters k/U and $S(T)$, the model is unable to predict two waves. This result is consistent among all 14 regions. While the model can be modified to capture multiple waves, it would be much more complicated to choose values of key parameters such as β, because of the presence of new variants of COVID-19.

The setting of individual choice in this model suggests that higher k/U encourages more social distancing, because the trade off between social activity

Figure 3.1: Daily new cases in the capital city Prague between 30.8.2021 and 12.5.2022. Black points denote model new cases, red points denote real-world data.

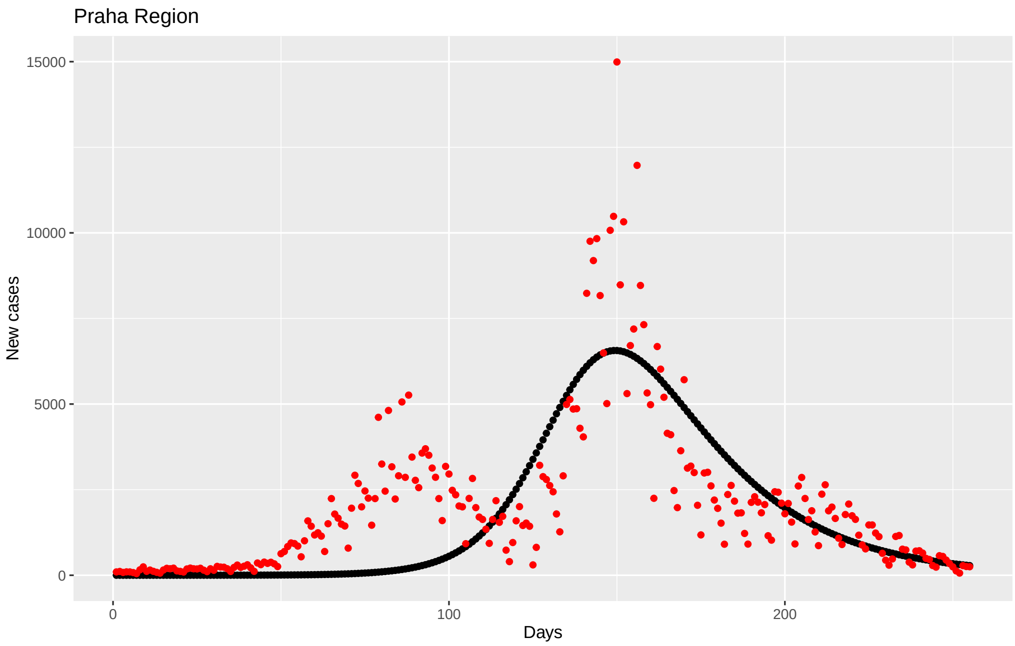
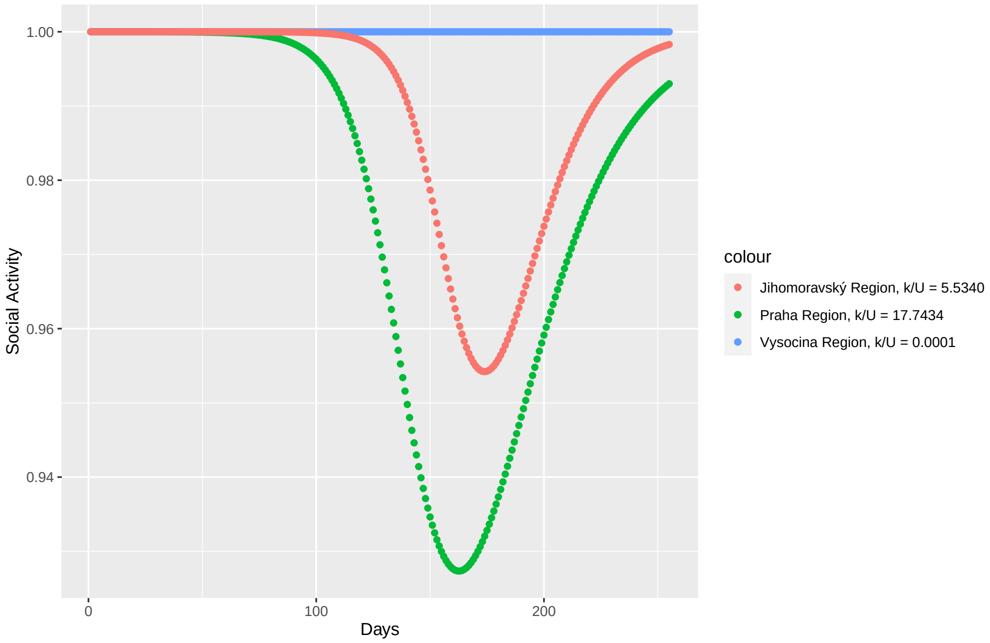


Figure 3.2: Model-predicted daily social activity in selected regions between 30.8.2021 and 12.5.2022



and catching COVID-19 is worse. The model results support this hypothesis, as can be seen in figure 3.2..

Any mathematical model used to describe a pandemic is inherently limited, because there are numerous factors influencing a spread of a disease and one model can't include all of them. One of the limitations of this model compared to other models is how low the number of compartments is in this model. There are models that include exposed population (Li *et al.* 2021), that differentiate between asymptomatic and symptomatic cases (Gaeta 2020) or even both (Hromádková 2020). Including these categories in this model could provide better predictions, although at the cost of making the model much more computationally complex and difficult to estimate. The model as is does not enable multiple waves of the disease. There are two main biases if we estimate the last two waves.

The model does not take into account vaccination. Vaccination decreases the amount of susceptible individuals continuously. The way the model is estimated, the parameters β and γ are already determined from other studies, we don't calibrate them. Social distancing decreases the amount of interactions, which decreases the amount of new infections. If we increase k/U, the social distancing increases, which means that the social activity $A(t)$ decreases. This in turn means that the amount of infected individuals $I(t)$ is lower throughout the modelled pandemic. With k being close or equal to zero, individuals are not concerned with the prospect of dying from COVID-19 and they set their activity $A(t)$ to one for all t. In this situation, we get the same results as in a SIRD model.

Not including vaccination into the model provides positive bias to the estimate of k/U. Vaccination decreases the size of the susceptible class $S(t)$, which in turn decreases the amount of new infections. Without any social distancing or vaccines, there would be many more daily infections. The model overestimates k/U, because the difference between potential infections and real infections is explained only through social distancing, while in reality, part of them are decreased due to vaccines.

The model also does not take into account different variants. Some of them differ from one another even in model parameters, such as mortality, the transition rate β or recovery rate γ (van Oosterhout *et al.* 2021). Newer variants, such as the omicron variant, has higher transition rate and lower mortality then variants present at the beginning of the pandemic. This means that the $k/U$is underestimated. With the same sizes of all other parameters and social distancing $A(t)$, an estimation of the model with higher transition rate has a higher number of daily infections. If we used true β of the newer variants, we would estimate higher k/U, as it would need to balance the increase in daily infections as to match the real-world data. This implies that $k/U$is underestimated.

In this article, the measure used for workforce composition is the share of high skilled labor in the employed population in the region. A variable used for high skilled labor are the groups with competence level 3 or higher according to the International Standard Classification of Occupations (CZ-ISCO) - level 1 - Major group. The included groups are also assumed to be the most able to engage in social distancing by conducting their work remotely, compared with the other groups.

From observing data in Table 3.1, we can already determine one outlier, which is Praha Region in terms of Vaccines per Capita. The SoHSL denotes Share of Skilled labour in table 3.1. Leaving Prague out of the sample would be costly to the variance of explanatory variables. It is not an outlier in both k/U and Share of Skilled Labour, so it is better to leave it included.

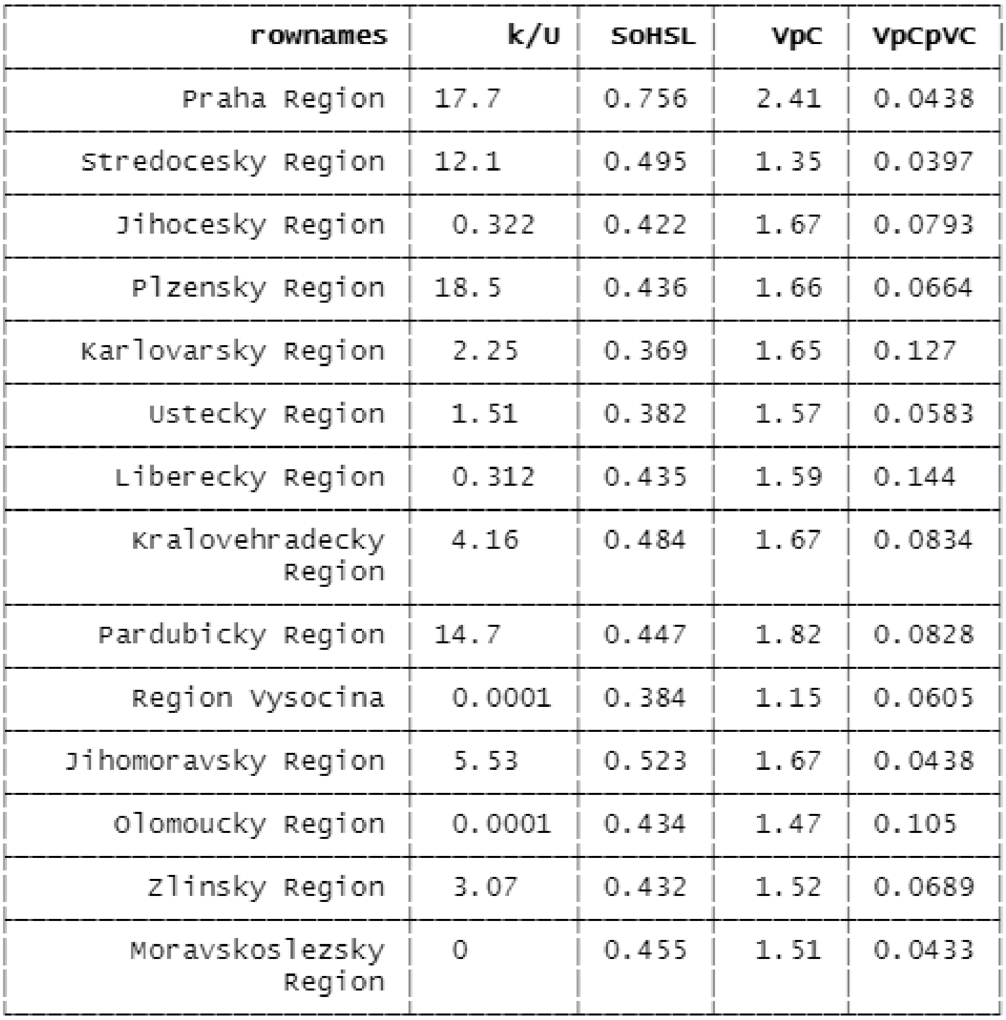
As noted before, the model does not include vaccination, which causes bias in the estimated k/U. There is notable heterogeneity between regions in terms of vaccination, which means that the bias would affect k/U in each region by the different degrees. Attitudes towards vaccination can also be linked with the share of high skilled labor, as people with higher education are more likely to get vaccinated (Humer *et al.* 2021). If we choose a linear regression model in the form of:

$$k/U = \beta_0 + \beta_1 * shareofhighskilledlabor + \beta_2 * ratioofvaccinated + u$$

the analysis will not suffer from the bias, as we include vaccination.

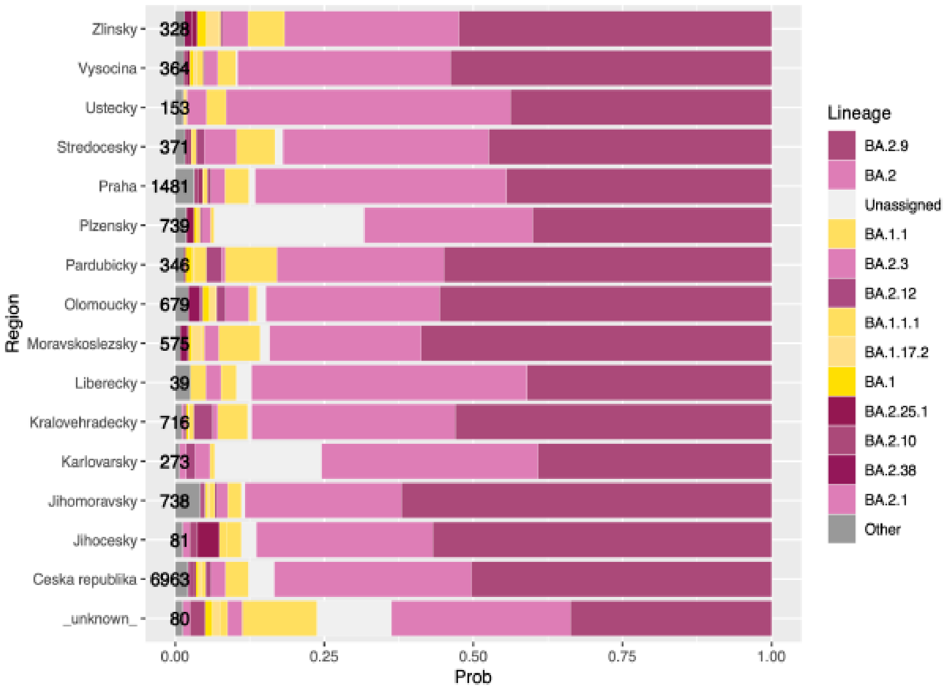
A second main source of bias was identified in the model where the mutations and variants of COVID-19 are not taken into account. The model estimates one pandemic with only one set of parameters of β and γ. The distribution of different variants is not homogeneous between regions so it could be a potential source of a bias, but the identification of properties of different variants of COVID-19 and their spread across the Czech regions is beyond the scope of this article. For illustration, a figure 3.4 shows complexity of spread of different COVID-19 variants across Czech regions.

Table 3.1: Variables for each region. Variable explanations: k/U - the measure of willingness of individuals to engage in social distancing, SoHSL - Share of High Skilled Labor, VpC - Vaccines per Capita, VpCpVC - Vaccines per Capita per Vaccination Center



rownames	k/U	SoHSL	VpC	VpCpVC
Praha Region	17.7	0.756	2.41	0.0438
Stredocesky Region	12.1	0.495	1.35	0.0397
Jihocesky Region	0.322	0.422	1.67	0.0793
Plzensky Region	18.5	0.436	1.66	0.0664
Karlovarsky Region	2.25	0.369	1.65	0.127
Ustecky Region	1.51	0.382	1.57	0.0583
Liberecky Region	0.312	0.435	1.59	0.144
Kralovehradecky Region	4.16	0.484	1.67	0.0834
Pardubicky Region	14.7	0.447	1.82	0.0828
Region Vysocina	0.0001	0.384	1.15	0.0605
Jihomoravsky Region	5.53	0.523	1.67	0.0438
Olomoucky Region	0.0001	0.434	1.47	0.105
Zlinsky Region	3.07	0.432	1.52	0.0689
Moravskoslezsky Region	0	0.455	1.51	0.0433

Figure 3.3: Representation of different COVID-19 variants in each region in the 22nd week of 2022



Source: Pačes (2022).

Table 3.2: Regressions

	k/U		
	Only SoSL	With Vaccination	With Density
	(1)	(2)	(3)
Share of Skilled Labor	42.283^{**}	24.758	48.098
	(17.141)	(27.485)	(41.378)
Vaccines per Capita		7.719	11.023
		(9.379)	(10.487)
Population density			-0.005
			(0.006)
Constant	-13.758	-18.193^{*}	-32.826
	(8.056)	(9.784)	(21.540)
Observations	14	14	14
R^2	0.336	0.375	0.410
Adjusted R^2	0.281	0.261	0.233
Residual Std. Error	5.884 (df = 12)	5.965 (df = 11)	6.080 (df = 10)
F Statistic	6.085^{**} (df = 1; 12)	3.299^{*} (df = 2; 11)	2.313 (df = 3; 10)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4. Results

Figure 3.1 shows the comparison between the estimated pandemic and the real-world data. The model is successful in capturing the new daily cases in the second wave depicted, but unsuccessful in showing the first wave. The fact that there are multiple waves of COVID-19 might be affected primarily by the government measures and emergence of new variants. The aspect in which the model fails is the social activity. Even without additional data and measures, we can state that the social activity decreased to some degree in all regions, while the model suggests that in regions such as the region Vysocina, the Liberecky region, the Jihocesky region and the Olomoucky region have k/U close to zero, which corresponds to almost no decrease in social activity, as can be seen in Figure 3.2.

This result could be explained in several ways:

- First of all, the coefficient β might be understated. The β coefficient was estimated using the basic reproduction number R_0. Different studies were not unanimous in its estimation. We used an upper limit for early estimate of R_0 for the Czech Republic and rich European countries.

- We already established that this might still be understated, as since the publication of these studies, newer variants emerged with potentially lower R_0 and also β. Lower β would imply a lower k/U, as the number of daily infections is increasing in β and decreasing in k/U. The k/U would be estimated to be lower to counteract less infections from lower β.

- Another reason for the low k/U might be that the number of real cases might be lower than used to estimate the model. We multiplied the number of daily cases by 1.7 in order to get closer to calibrating for all cases, not only the reported ones. While this number might be more appropriate to earlier waves, people could be more used to getting tested in those

> two last waves. The bans of traveling without negative tests and prevalence of self-tests might have encouraged the more thorough testing and the amount of unreported cases might have been lower. The population might have also been more used to getting tested. If we calibrated for fewer cases, the k/U would be higher, as the model would overestimate the number of cases otherwise.

To amend this issue, one could acquire proxy data for the social activity and then calibrate the model for that data. This approach is much closer to the one in Farboodi *et al.* (2021). The problem with this approach is that the estimated daily infections might be largely overestimated. For instance, in the Farboodi *et al.* (2021) paper, the authors demonstrated how the model is successful in modelling the social activity, even without any calibration if we set $U = 1$. The problem with their estimation is that their model predicted roughly 2/3 of the US population to get infected in the first wave. This wasn't the case, as even for 30.6.2022, there were roughly 89 million confirmed cumulative cases, which isn't even 1/3 of USA population.

After estimating the model for each region, we acquire Figure 3.2. The Vaccines per Capita variable represents the cumulative number of vaccines administered in the region until 12.5.2022 divided by the population of the region. This measure serves as a proxy to the amount of vaccinated in each region. The difference between the proxy and the other variable can be seen in the case of the Stredocesky region. Intuitively, there is no reason for the Vaccines per Capita measure to be so low in the Stredocesky region - there is a large degree of migration between Prague and Stredocesky region. We also wouldn't expect more sceptical attitudes towards vaccination in this region, since many people there work in the capital city. The reason why the Vaccines per Capita measure is so low can be explained by the inter-region migration. People from the Stredocesky region could have traveled to the Prague region to get their vaccines. The administrative center of the region is Prague, which makes it unique among other regions, because their administrative centers are in cities only belonging to the region.

It is thus also possible to use the VpCpVC measure, which is an acronym denoting Vaccines per Capita per Vaccination Center. This measure is constructed using the following formula:

$$VpCpVC = \frac{NumberofVaccines}{Population * NumberofVaccinationCenters}$$

This formula should ensure better estimation of the number of vaccinated people in the region, because the regions with relatively higher amounts of vaccination centers should attract more visitors from other regions. This measure didn't produce substantially different results to the three regressions included.

One other control variable considered was the population density. This variable should be correlated with the Share of Skilled Labour variable, because those regions with higher population density will have larger cities, which attract skilled labour. It is also reasonable to assume that population density affects social distancing. People in regions with higher population density will be more likely to live in big cities. The mixing of the population in these cities could be more thorough than in regions where the population is more spread out over villages and smaller cities. The problem with using this variable is that due to it's high correlation with Share of Skilled Labour, we risk the colinearity bias. With already a low amount of observations, adding this variable would risk overfiting, so the variable is left out. After adding to the regression, the adjusted r-squared doesn't increase, further discouraging it's inclusion in the regression.

The conclusions drawn from these regressions suffer some limitations. First of all, there are only 14 observations, which means that we can't use a large number of variables. If we use too few variables, we risk the results being affected by biases. For example, not including vaccination measure causes an omitted variable bias, as described before. If we include too many variables, we run the risk of over-fitting. All three regressions maintain the same sign for the effect of Share of Skilled Labor. Regressions (2) and (3) don't contain any significant predictors and the overall F-test is also non-significant. Since the amount of observations for the regressions is very small, we are limited in what we can infer. All the regressions suggest that there is a positive correlation between the Share of Skilled Labor and k/U, which depicts unwillingness to engage in social distancing. If there is a higher share of skilled labor in a region, we would expect the inhabitants of that region to comply more in social distancing. The sign of the correlation didn't change after including the vaccination variable into the regression, suggesting that the result holds even when accounting for differences regarding vaccination in each region.

5. Conclusion and Discussion

Individual countries and even regions within countries had heterogeneous experiences in the COVID-19 pandemic. The way a disease spreads is not only determined by it's biological aspects, but also by the characteristics of the population it spreads in. One of the main determinants of the spread of the pandemic is social activity. Most people in the world experienced a lock-down, a quarantine and other government mandated social distancing measures with various degrees of effectiveness in stopping the spread of COVID-19. This article presents a novel approach to investigating the relationship between workforce composition and social distancing during the COVID-19.

The article uses a model that takes the SIRD model with a well-mixed population. This model is enhanced by including a social activity measure, which modifies the transmission rate. Rational individuals with perfect information dynamically optimize their social activity over the course of the pandemic. Core model parameters are calibrated to best fit the new daily infections from 30.8.2021 to 12.5.2022. The calibrated parameter of interest is k/U, which denotes the statistical value of life divided by the subjective value of social activity. We first take the modified real-world data and we find the value of the social distancing measure k/U such that the model best fits the data. We use this method to estimate k/U for each region. By estimating k/U, we get a measure of preferences regarding social distancing that can be compared with workforce composition. We then regress k/U on share of skilled labour, vaccination per capita and population density.

We find a positive correlation between share of skilled labour and social distancing. This result is consistent with Garnier *et al.* (2021), who finds that poorer counties in the USA with more poverty and essential workers engage less in social distancing. Similarly, the result is also consistent with Huang *et al.* (2020) and Weill *et al.* (2020). This article provides a new method to identify social distancing without use of mobile device data. It also contributes by examining the relationship between social distancing and workforce composition in the Czech Republic, while most of the literature focuses solely on the USA.

The article still leaves room for exploration with the used approach. We only estimated one wave of the COVID-19 pandemic in the Czech regions. It would be theoretically possible to estimate k/U for multiple waves of the pandemic, which would gives us many more observations in the form of panel data. Another possibility is to include regions from other countries, or even estimate k/U for each country as a whole. Utilizing these approaches would potentially enable usage of more control variables, which would make the regression more rigorous.

Bibliography

ALIMOHAMADI, Y., M. TAGHDIR, & M. SEPANDI (2020): "Estimate of the basic reproduction number for covid-19: a systematic review and meta-analysis." *Journal of Preventive Medicine and Public Health* **53(3)**: p. 151.

BARRIOS, J. M., E. BENMELECH, Y. V. HOCHBERG, P. SAPIENZA, & L. ZINGALES (2021): "Civic capital and social distancing during the covid-19 pandemicâ†." *Journal of public economics* **193**: p. 104310.

BERGERI, I., M. WHELAN, H. WARE, L. SUBISSI, A. NARDONE, H. C. LEWIS, Z. LI, X. MA, M. VALENCIANO, B. CHENG *et al.* (2021): "Global epidemiology of sars-cov-2 infection: a systematic review and meta-analysis of standardized population-based seroprevalence studies, jan 2020-oct 2021." *medRxiv* .

BILINSKI, A. & E. J. EMANUEL (2020): "Covid-19 and excess all-cause mortality in the us and 18 comparison countries." *Jama* **324(20)**: pp. 2100–2102.

DAN, J. M., J. MATEUS, Y. KATO, K. M. HASTIE, E. D. YU, C. E. FALITI, A. GRIFONI, S. I. RAMIREZ, S. HAUPT, A. FRAZIER *et al.* (2021): "Immunological memory to sars-cov-2 assessed for up to 8 months after infection." *Science* **371(6529)**: p. eabf4063.

FARBOODI, M., G. JAROSCH, & R. SHIMER (2021): "Internal and external effects of social distancing in a pandemic." *Journal of Economic Theory* **196**: p. 105293.

GAETA, G. (2020): "A simple sir model with a large set of asymptomatic infectives." *arXiv preprint arXiv:2003.08720* .

GARNIER, R., J. R. BENETKA, J. KRAEMER, S. BANSAL *et al.* (2021): "Socioeconomic disparities in social distancing during the covid-19 pandemic in the united states: observational study." *Journal of medical Internet research* **23(1)**: p. e24591.

Gualda, E., A. Krouwel, M. Palacios-Gálvez, E. Morales-Marente, I. Rodríguez-Pascual, & E. B. García-Navarro (2021): "Social distancing and covid-19: Factors associated with compliance with social distancing norms in spain." *Frontiers in psychology* **12**.

He, W., G. Y. Yi, & Y. Zhu (2020): "Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for covid-19: Meta-analysis and sensitivity analysis." *Journal of medical virology* **92(11)**: pp. 2543–2550.

Hromádková, v. (2020): "Initiative model anticovid-19 for Č r, model a overall epidemiological view, technical description."

Huang, V., S. Sutermaster, Y. Caplan, H. Kemp, D. Schmutz, & S. K. Sgaier (2020): "Social distancing across vulnerability, race, politics, and employment: How different americans changed behaviors before and after major covid-19 policy announcements." *MedRxiv* .

Humer, E., A. Jesser, P. L. Plener, T. Probst, & C. Pieh (2021): "Education level and covid-19 vaccination willingness in adolescents." *European Child & Adolescent Psychiatry* pp. 1–3.

Kong, J. D., E. W. Tekwa, & S. A. Gignoux-Wolfsohn (2021): "Social, economic, and environmental factors influencing the basic reproduction number of covid-19 across countries." *PloS one* **16(6)**: p. e0252373.

Lauer, S. A., K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, & J. Lessler (2020): "The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application." *Annals of internal medicine* **172(9)**: pp. 577–582.

Li, W., J. Gong, J. Zhou, L. Zhang, D. Wang, J. Li, C. Shi, & H. Fan (2021): "An evaluation of covid-19 transmission control in wenzhou using a modified seir model." *Epidemiology & Infection* **149**.

Locatelli, I., B. Trächsel, & V. Rousson (2021): "Estimating the basic reproduction number for covid-19 in western europe." *Plos one* **16(3)**: p. e0248731.

Mongey, S., L. Pilossoph, & A. Weinberg (2021): "Which workers bear the burden of social distancing?" *The Journal of Economic Inequality* **19(3)**: pp. 509–526.

Murphy, K., H. Williamson, E. Sargeant, & M. McCarthy (2020): "Why people comply with covid-19 social distancing restrictions: Self-interest or duty?" *Australian & New Zealand Journal of Criminology* **53(4)**: pp. 477–496.

van Oosterhout, C., N. Hall, H. Ly, & K. M. Tyler (2021): "Covid-19 evolution during the pandemic–implications of new sars-cov-2 variants on disease control and public health policies."

Painter, M. & T. Qiu (2020): "Political beliefs affect compliance with covid-19 social distancing orders." *Covid Economics* **4**: pp. 103–123.

Pačes, K. (2022): "Varianty sars-cov-2 v České republice." .

Piler, P., V. Thon, L. Andrýsková, K. Doležel, D. Kostka, T. Pavlík, L. Dušek, H. Pikhart, M. Bobák, S. Matic *et al.* (2022): "Nationwide increases in anti-sars-cov-2 igg antibodies between october 2020 and march 2021 in the unvaccinated czech population." *Communications Medicine* **2(1)**: pp. 1–7.

Simonov, A., S. K. Sacher, J.-P. H. Dubé, & S. Biswas (2020): "The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic." *Technical report*, National Bureau of Economic Research.

Weill, J. A., M. Stigler, O. Deschenes, & M. R. Springborn (2020): "Social distancing responses to covid-19 emergency declarations strongly differentiated by income." *Proceedings of the National Academy of Sciences* **117(33)**: pp. 19658–19660.

Xie, W., S. Campbell, & W. Zhang (2020): "Working memory capacity predicts individual differences in social-distancing compliance during the covid-19 pandemic in the united states." *Proceedings of the National Academy of Sciences* **117(30)**: pp. 17667–17674.

Appendix A

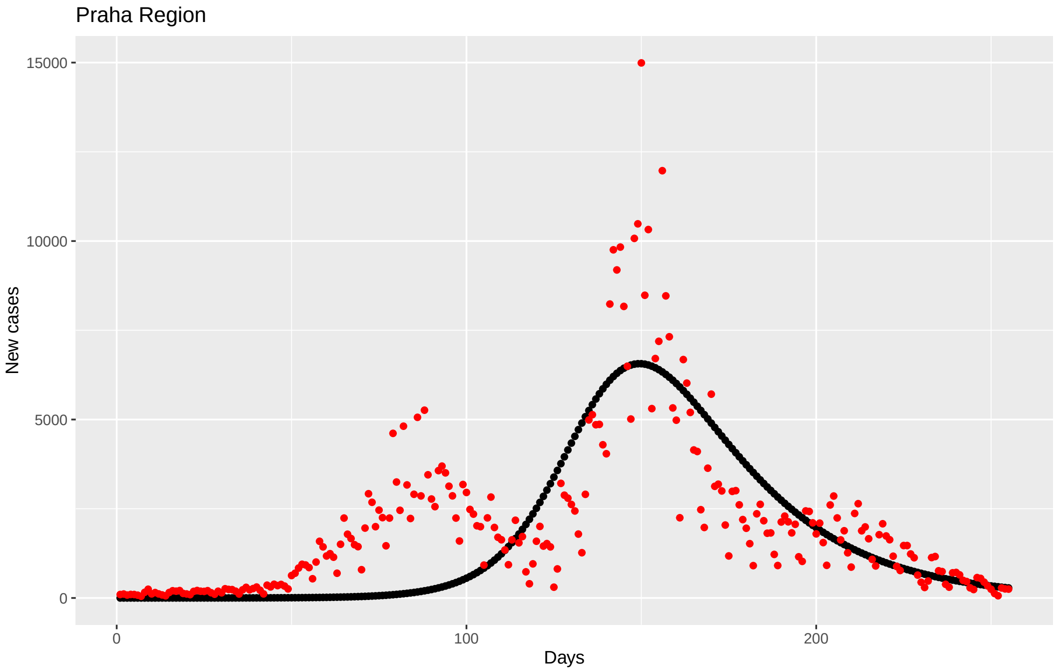
The coding for this article was compiled in R. The model is estimated as described in chapter 3.2. Because the equations for costate variables are unstable, we develop two models simultaneously arbitrarily close to one another. If these two models separate by too large of a degree, the development of the model is stopped and new values of S(t) are chosen between those two models S(t). From this point, other variables are made equal between the two models and the development continues.

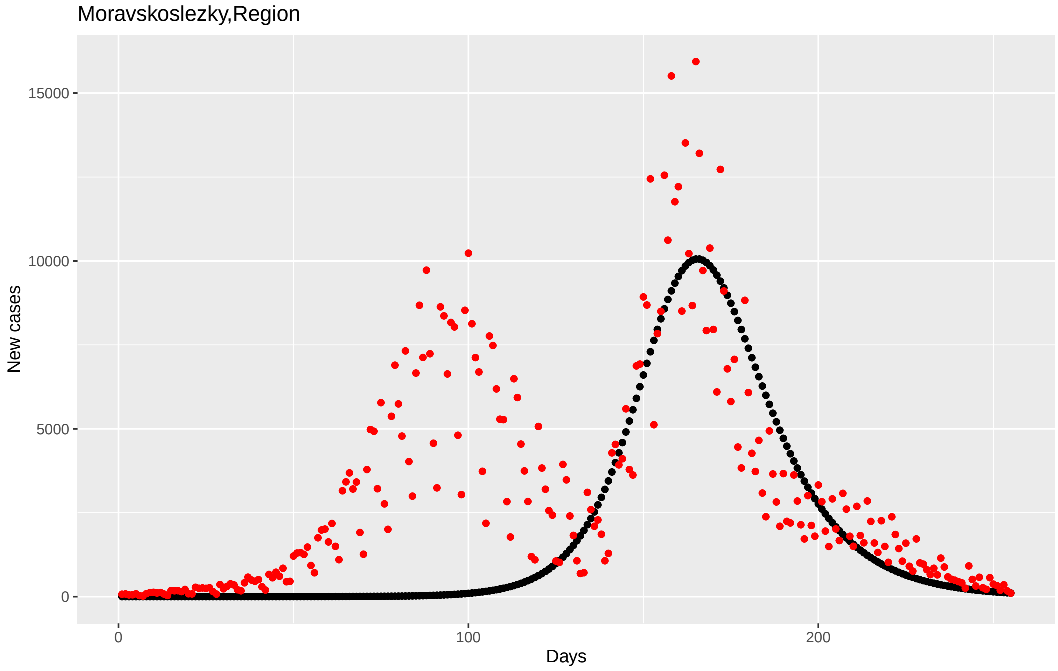
The estimation in section 2 is repeated 14 times, once for each region. The detailed comments for one such estimation can be send upon request.

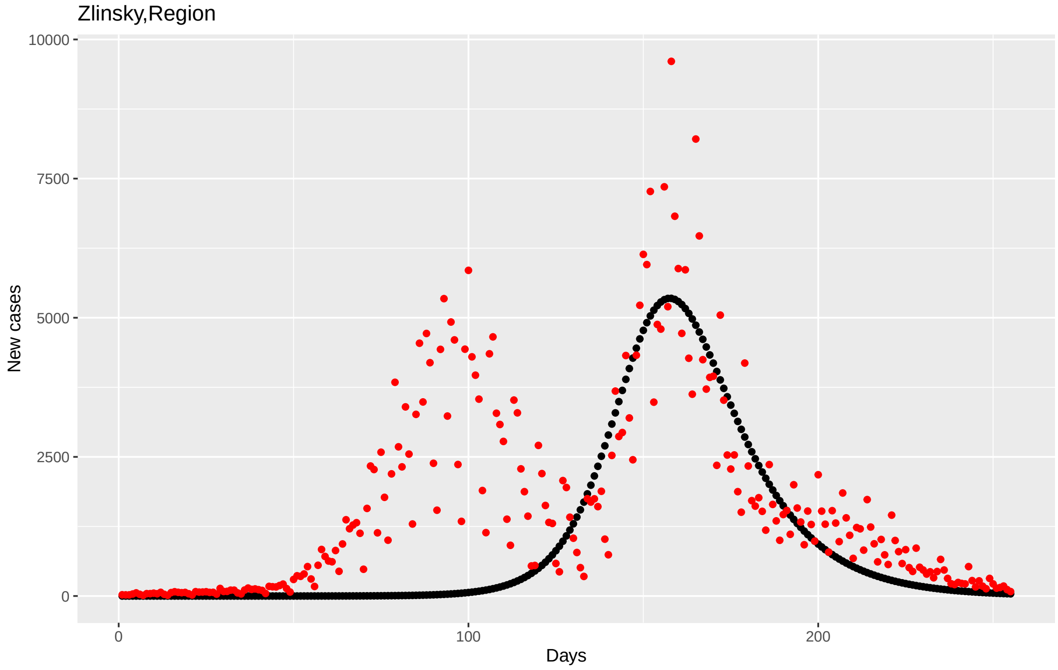
The red dotes denoted real new infections multiplied by 1.7, while the black dots denote model predictions in the models below.

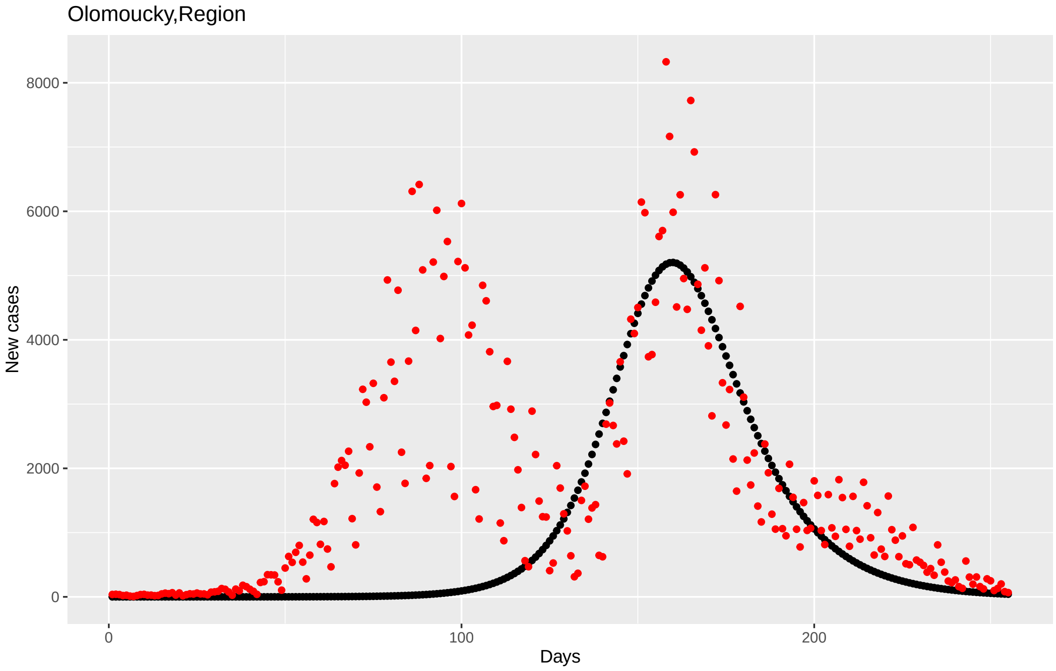
Appendix B

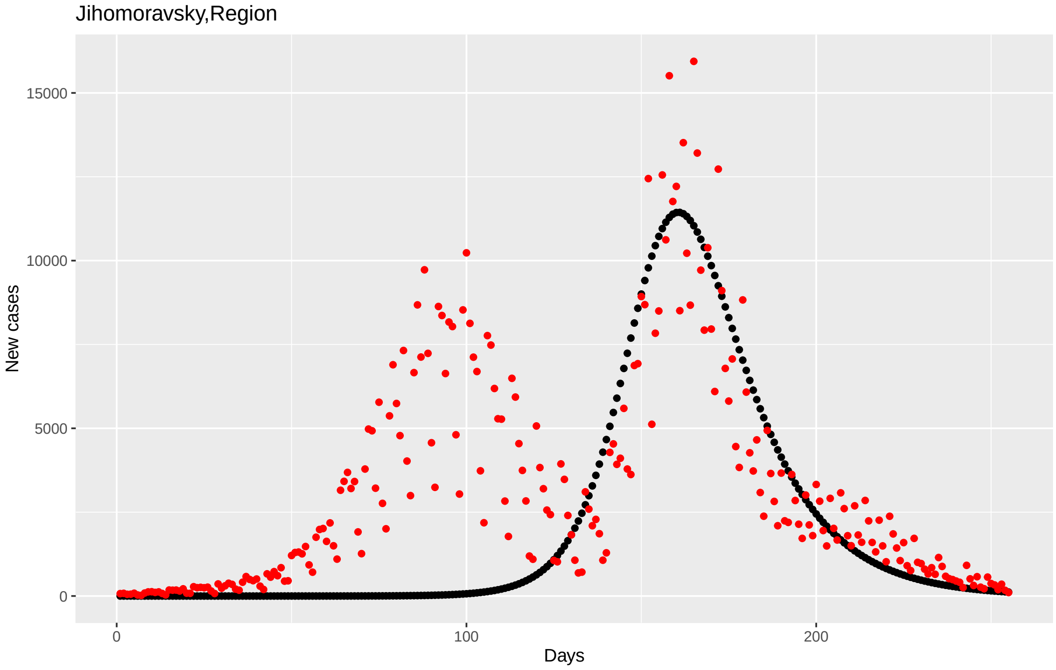
The red dotes denoted real new infections multiplied by 1.7, while the black dots denote model predictions.

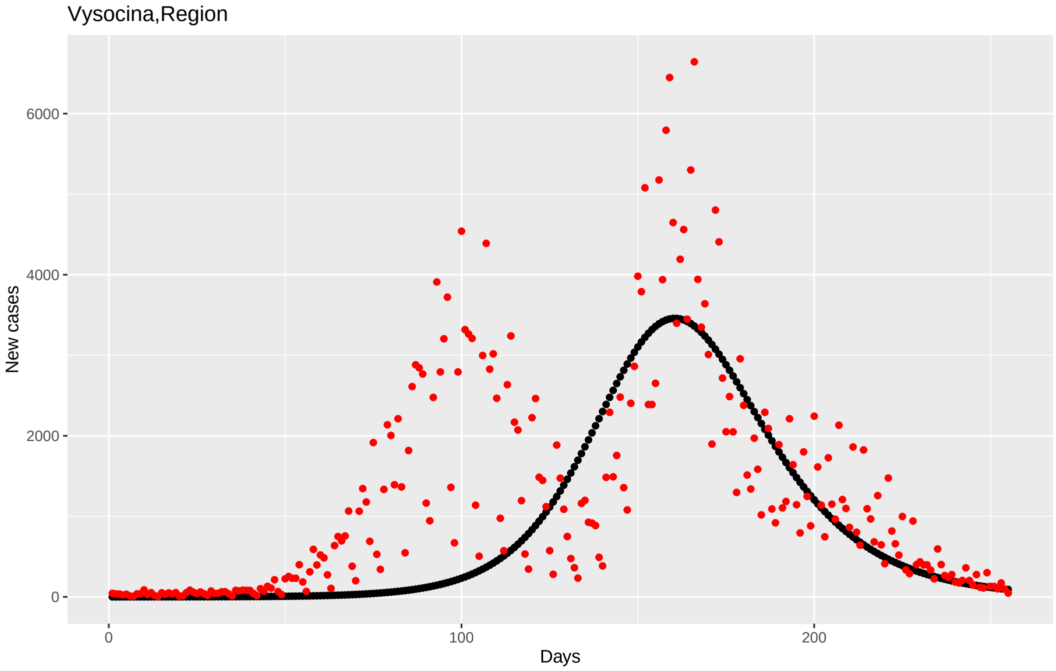


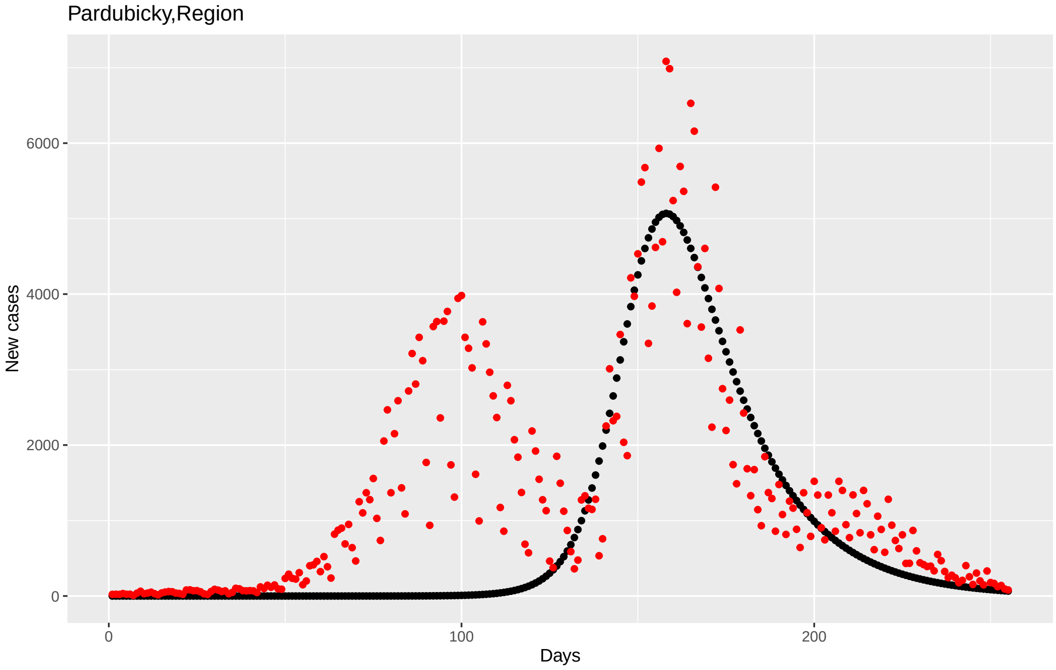

Moravskoslezky,Region
15000
10000
5000
0
New cases
0
100
200
Days

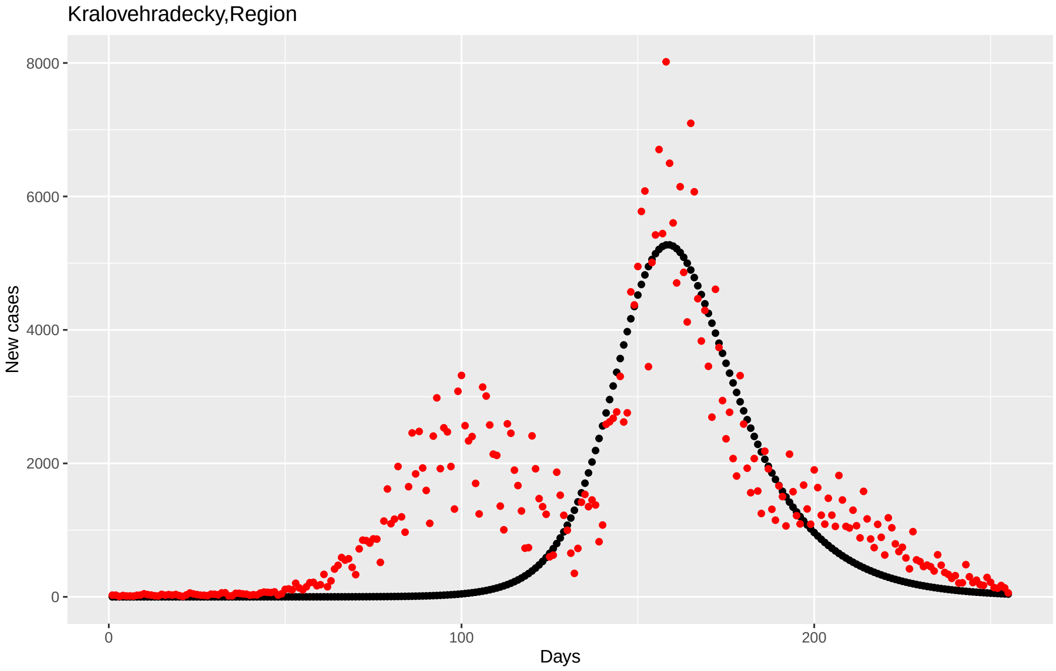

Zlinsky,Region
10000
7500
5000
2500
0
New cases
0
100
200
Days

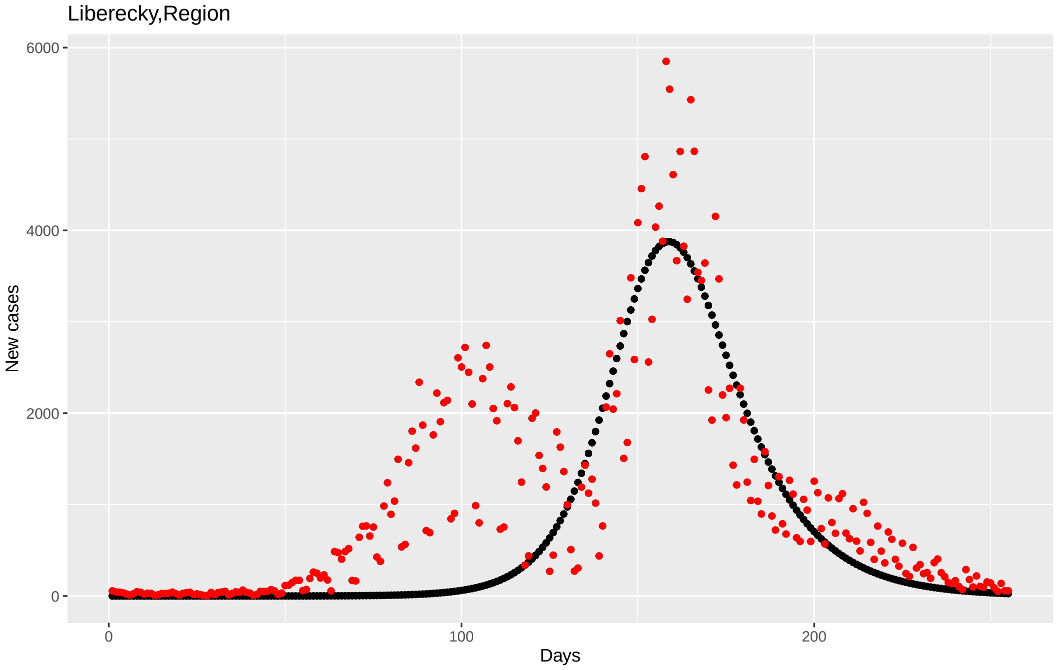

Olomoucky,Region
New cases
8000
6000
4000
2000
0
0
100
200
Days

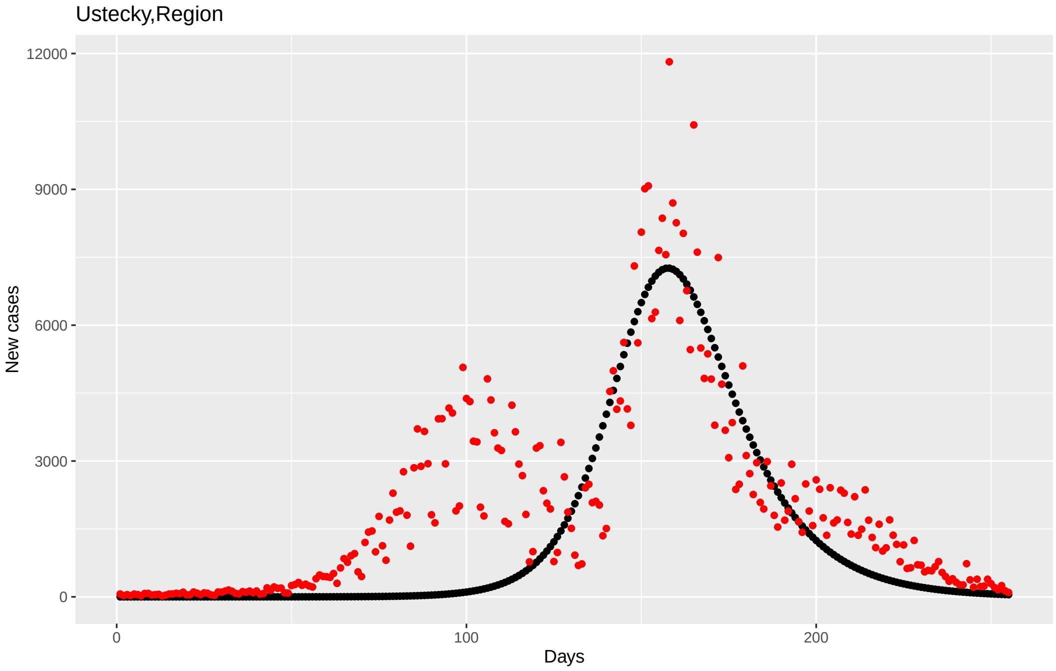

Jihomoravsky,Region
New cases
15000
10000
5000
0
0
100
200
Days

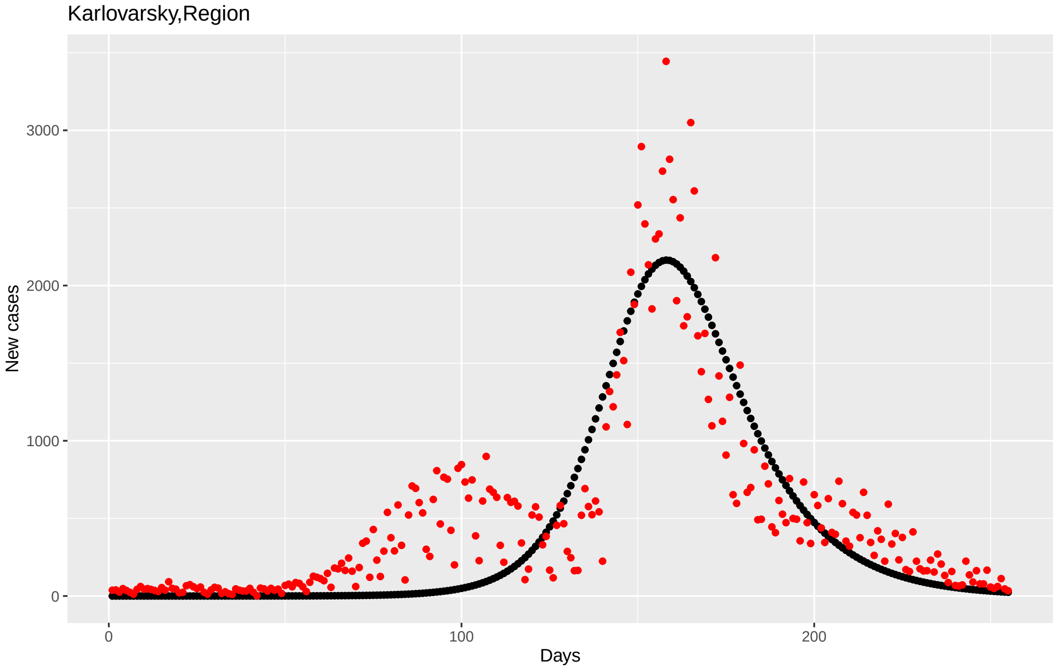

Vysocina,Region
New cases
6000
4000
2000
0
0
100
200
Days

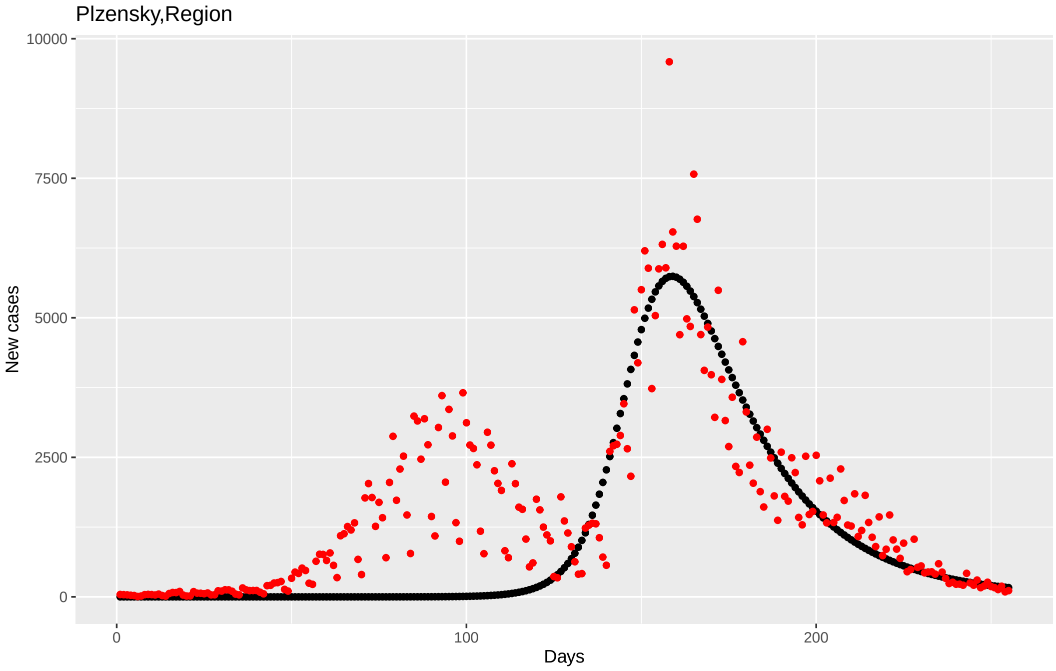

Pardubicky,Region
New cases
6000
4000
2000
0
0
100
200
Days

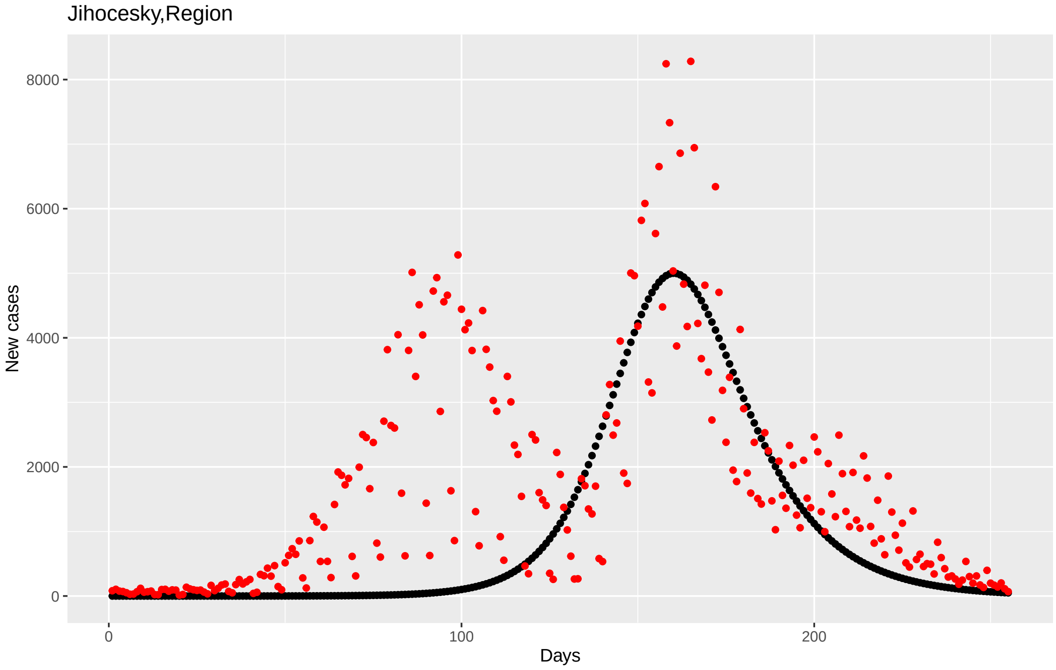

Kralovehradecky,Region
New cases
8000
6000
4000
2000
0
0
100
200
Days


Liberecky,Region
New cases
6000
4000
2000
0
0
100
200
Days


Ustecky,Region
New cases
12000
9000
6000
3000
0
0
100
200
Days


Karlovarsky,Region
New cases
3000
2000
1000
0
0
100
200
Days


Plzensky,Region
New cases
10000
7500
5000
2500
0
0
100
200
Days


Jihocesky,Region
New cases
8000
6000
4000
2000
0
0
100
200
Days

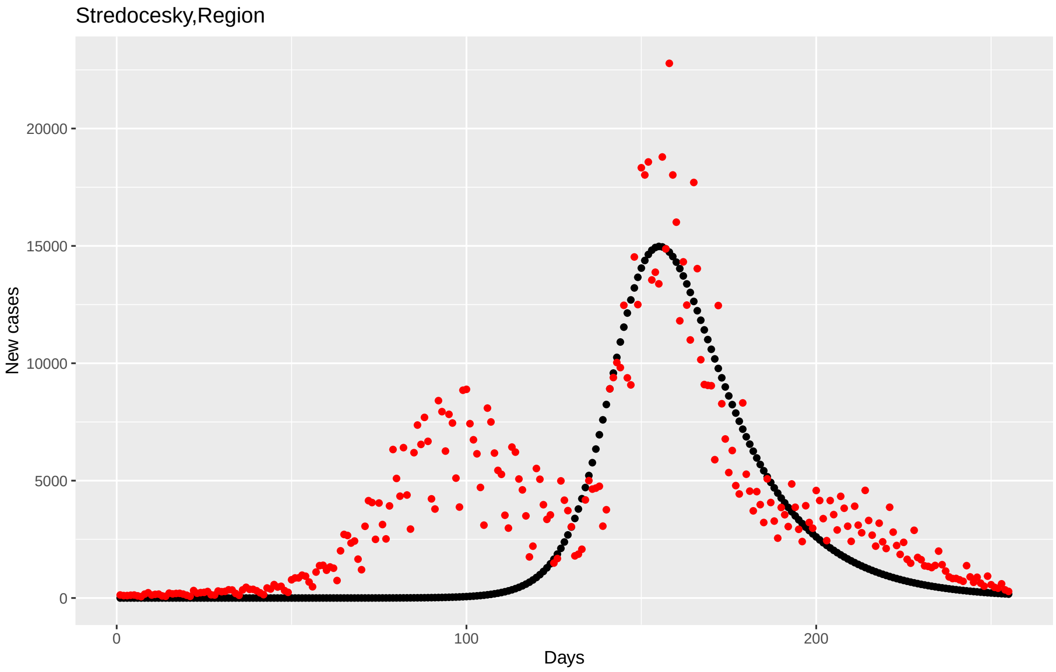


Figure 5.1: Social activity between 30.8.2021 and 12.5.2022: Praha Region

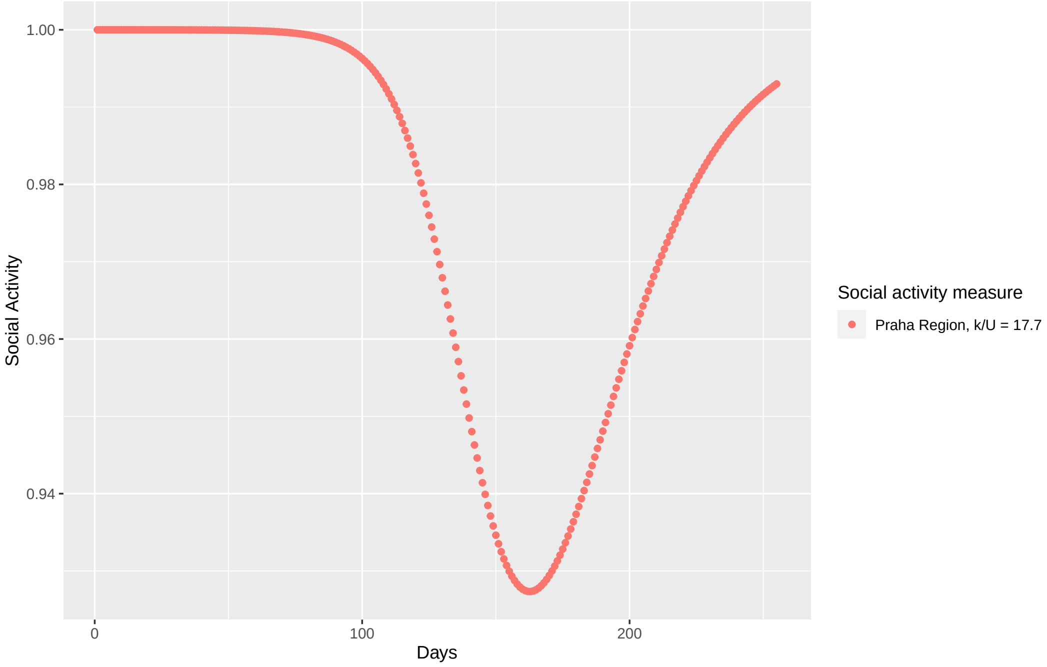


Figure 5.2: Social activity between 30.8.2021 and 12.5.2022: Stredocesky Region

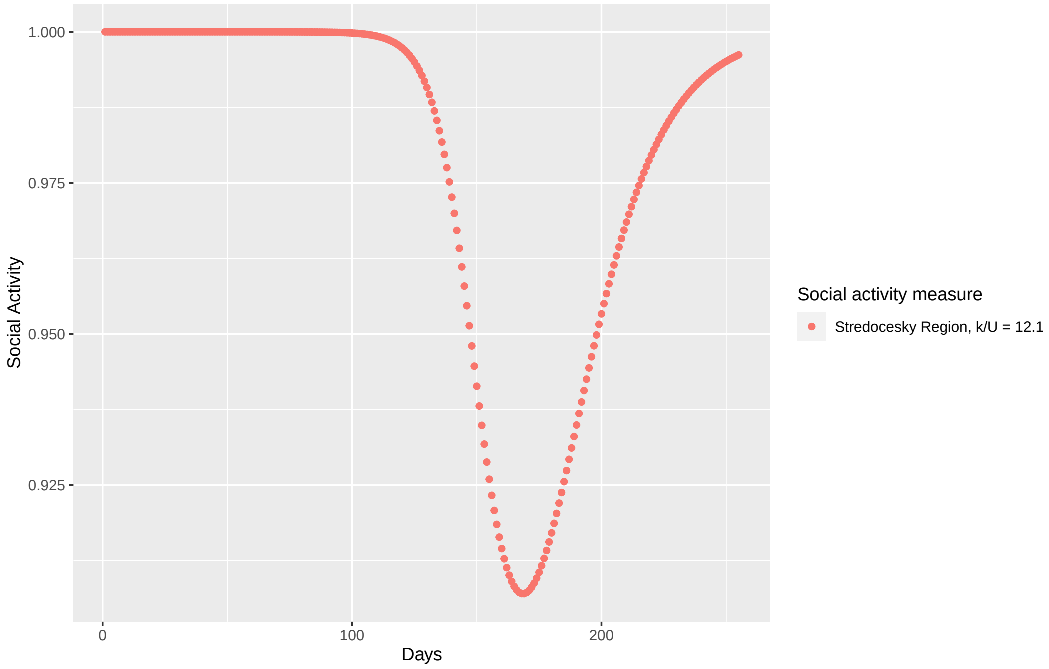


Figure 5.3: Social activity between 30.8.2021 and 12.5.2022: Jihocesky Region

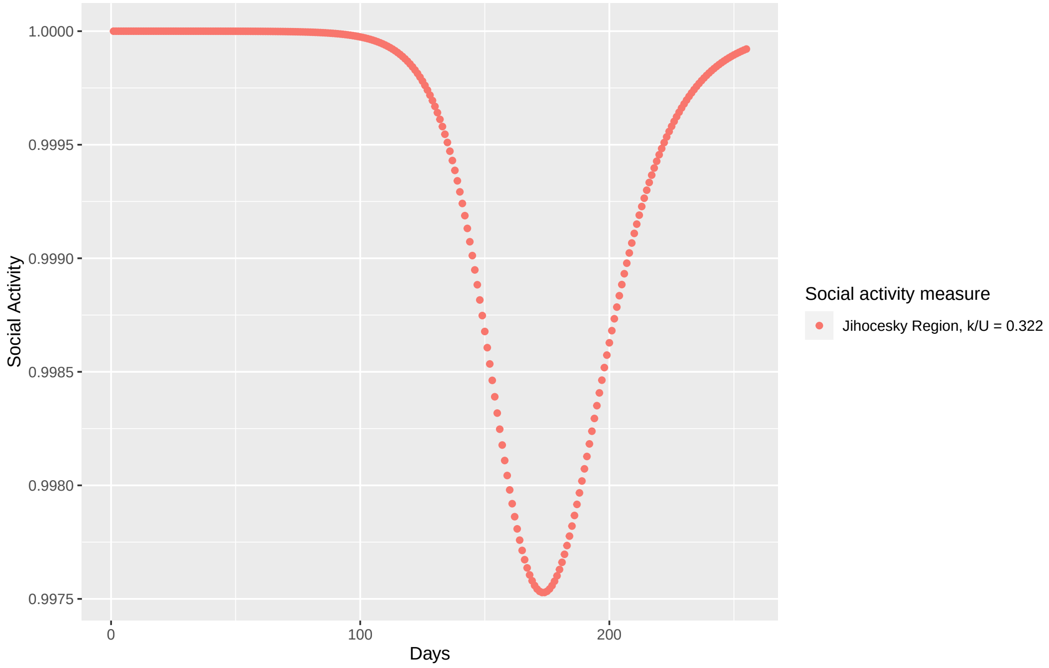


Figure 5.4: Social activity between 30.8.2021 and 12.5.2022: Plzensky Region

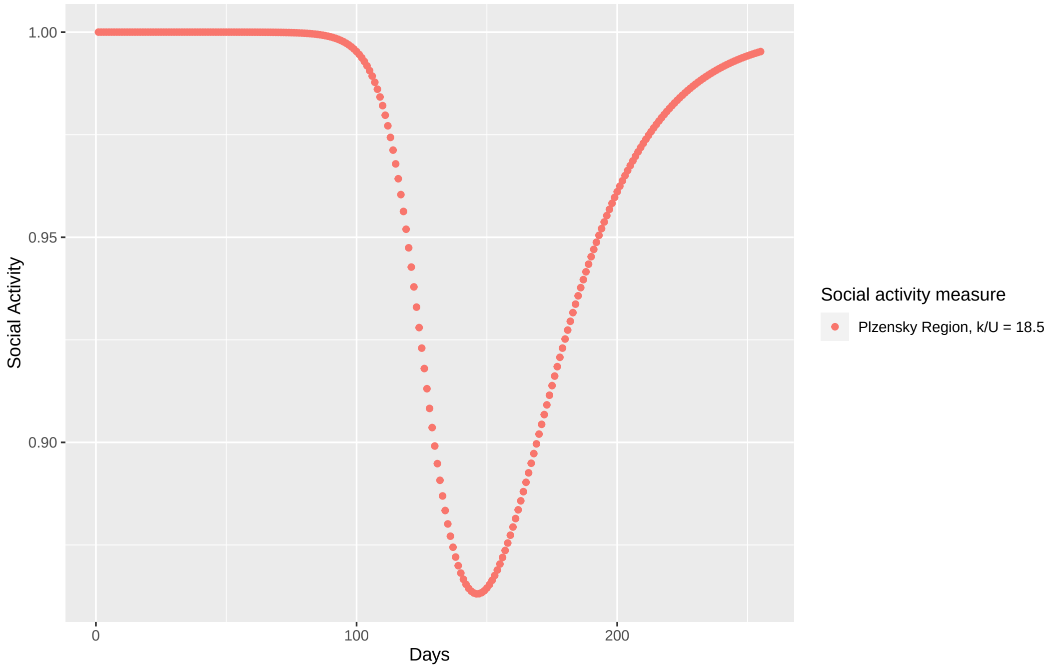


Figure 5.5: Social activity between 30.8.2021 and 12.5.2022: Karlovarsky Region

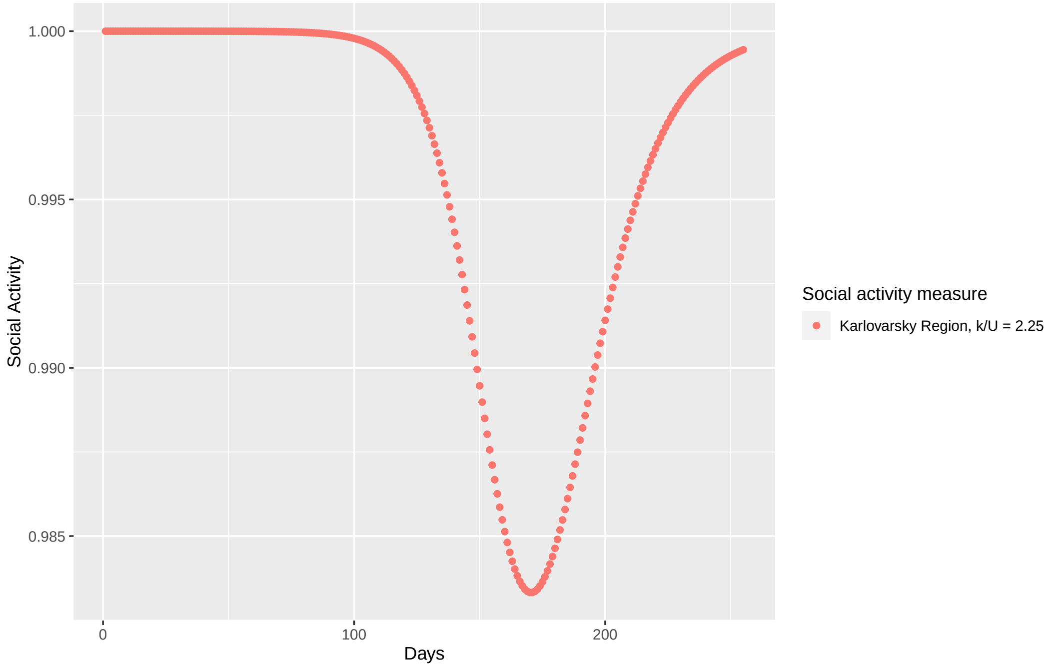


Figure 5.6: Social activity between 30.8.2021 and 12.5.2022: Ustecky Region

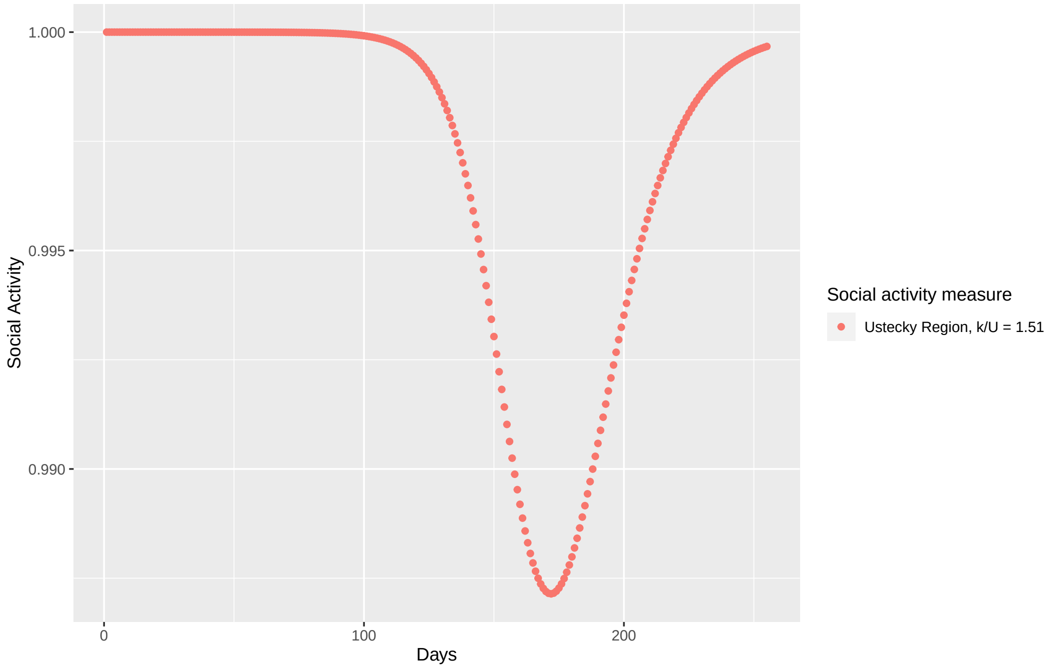


Figure 5.7: Social activity between 30.8.2021 and 12.5.2022: Liberecky Region

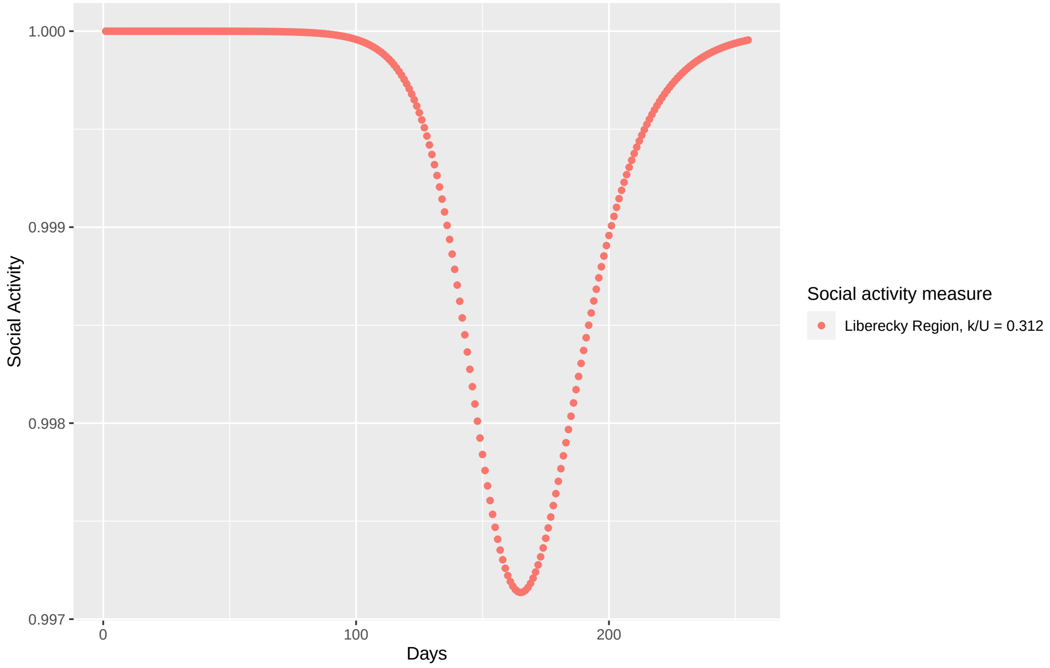


Figure 5.8: Social activity between 30.8.2021 and 12.5.2022: Kralovehradecky Region

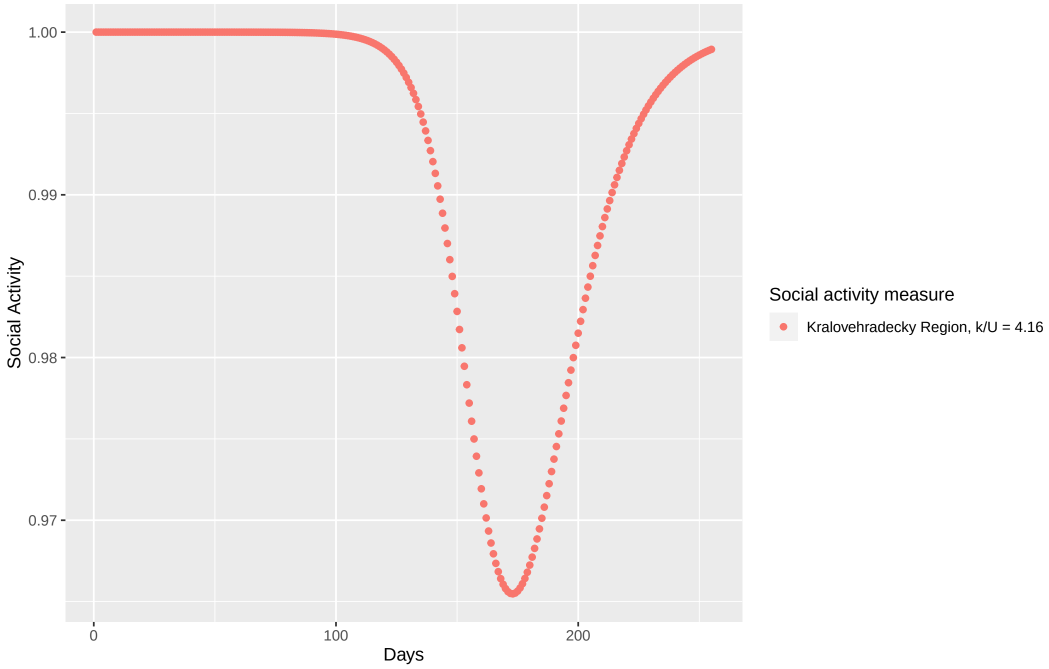


Figure 5.9: Social activity between 30.8.2021 and 12.5.2022: Pardubicky Region

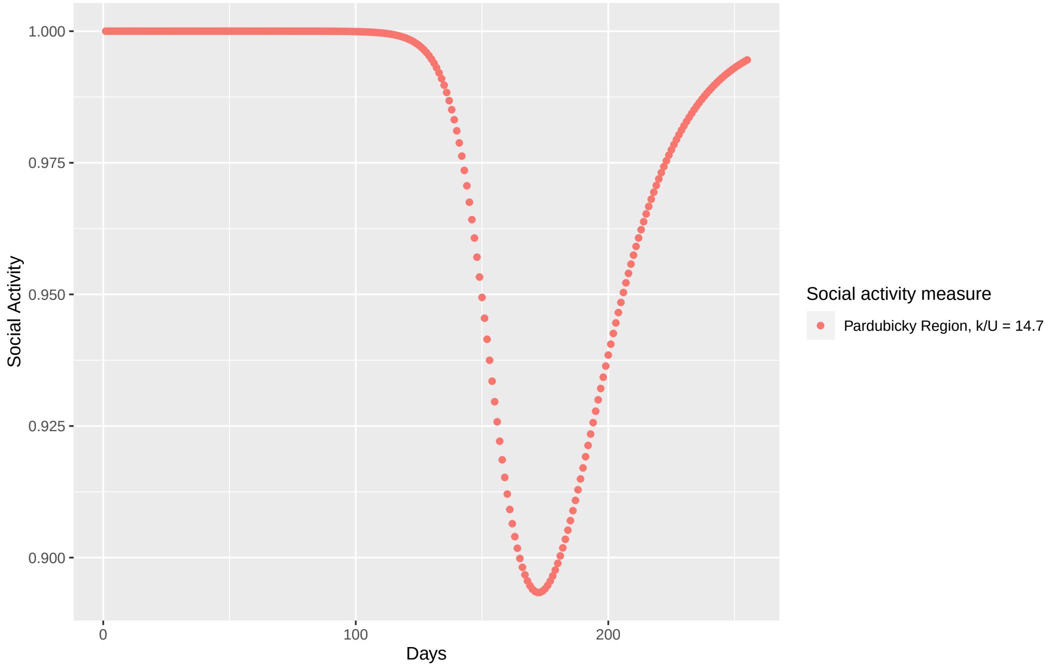


Figure 5.10: Social activity between 30.8.2021 and 12.5.2022: Region Vysocina

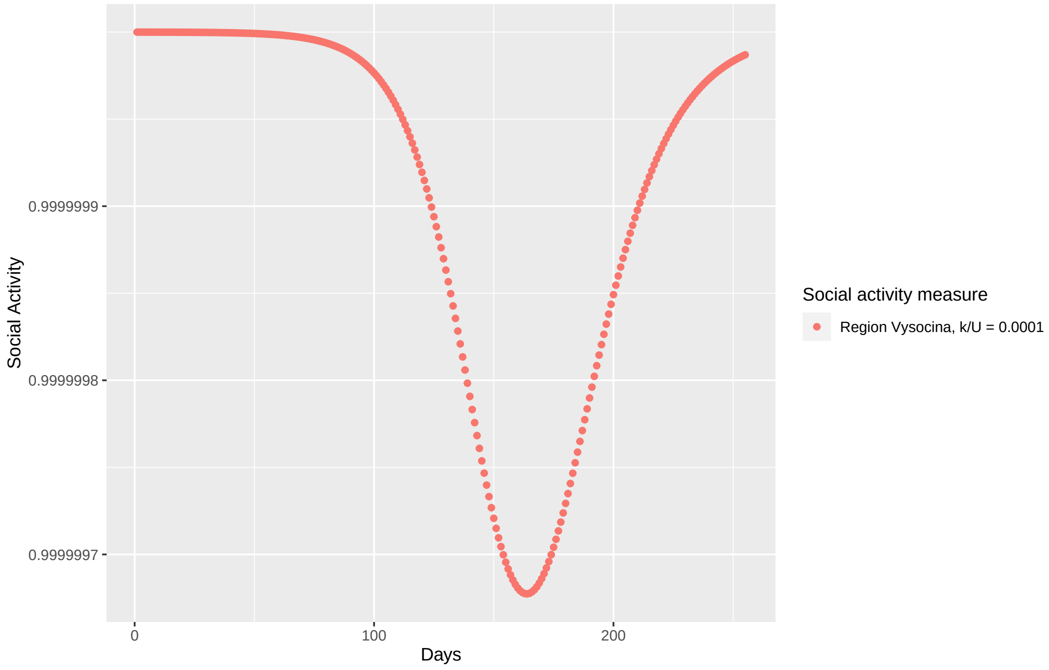


Figure 5.11: Social activity between 30.8.2021 and 12.5.2022: Jihomoravsky Region

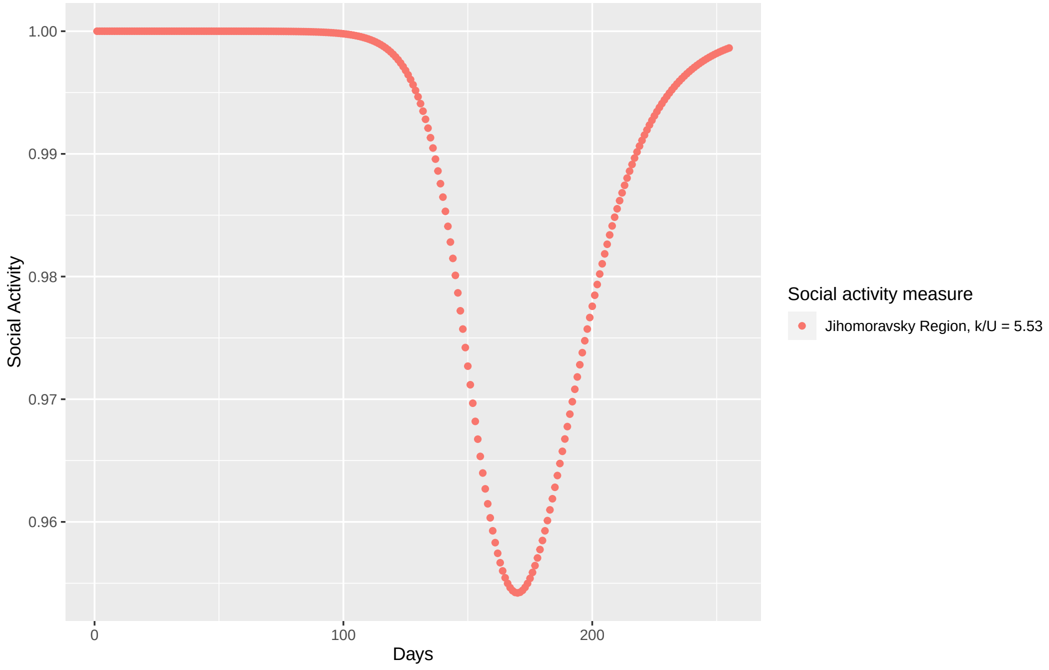


Figure 5.12: Social activity between 30.8.2021 and 12.5.2022: Olomoucky Region

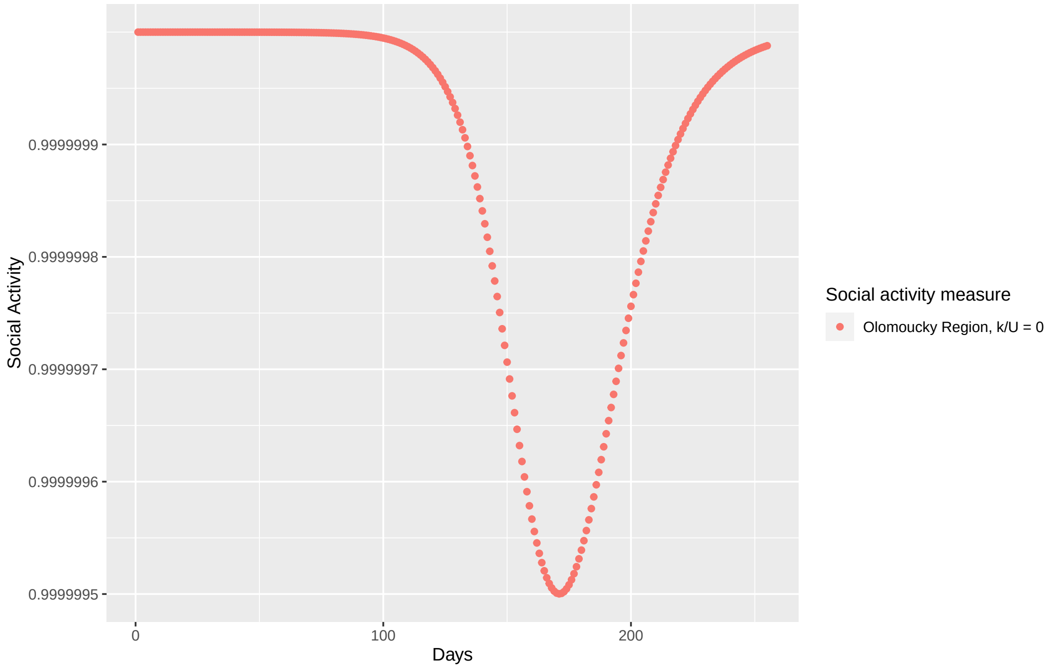


Figure 5.13: Social activity between 30.8.2021 and 12.5.2022: Zlinsky Region

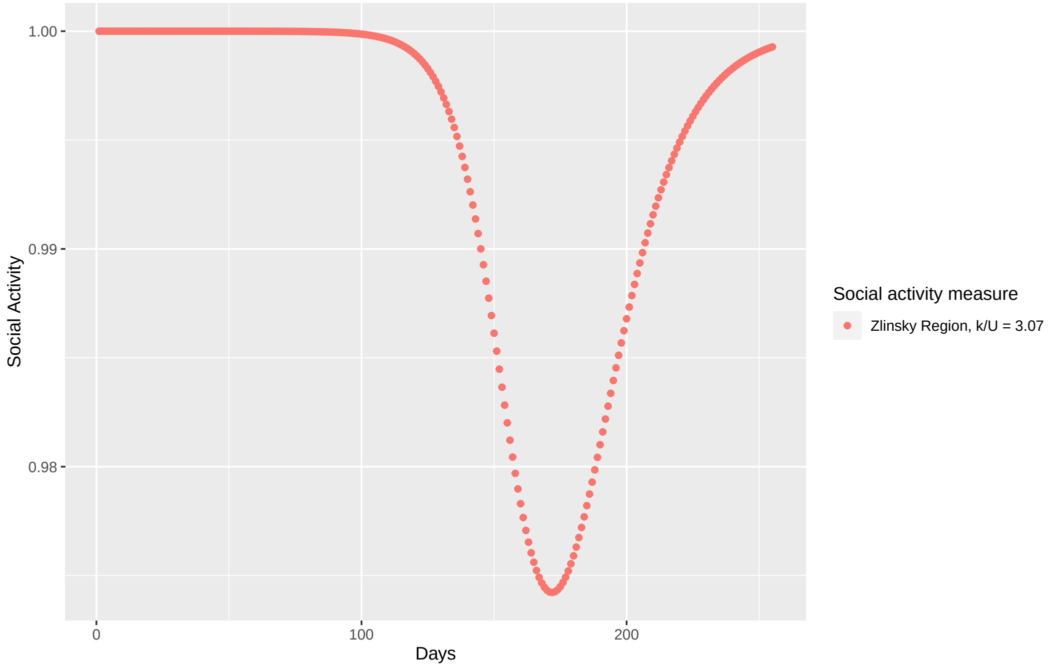
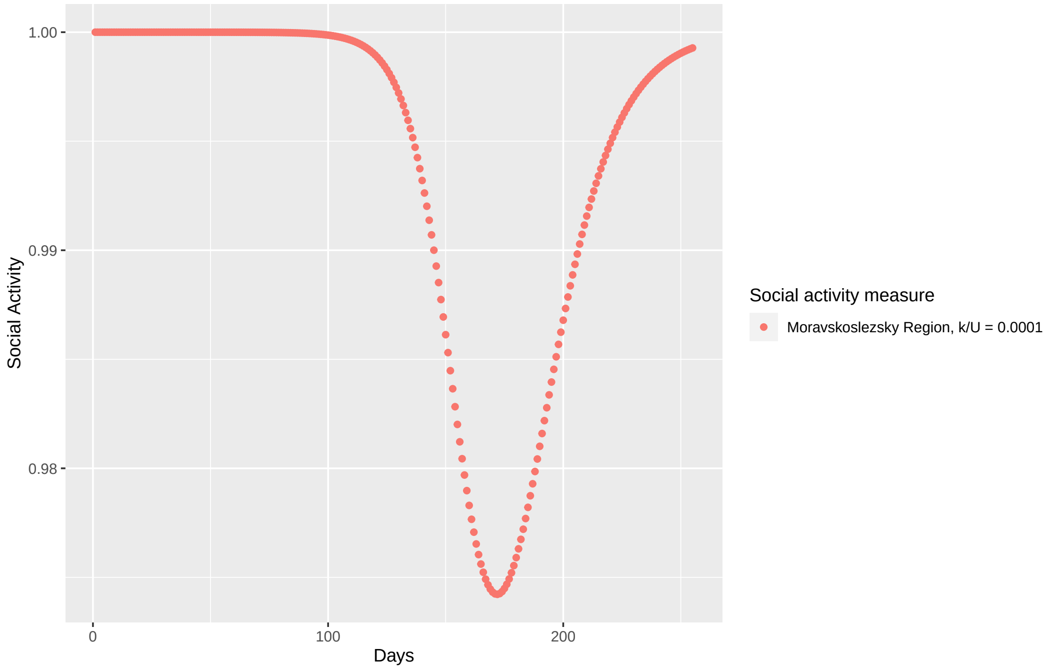


Figure 5.14: Social activity between 30.8.2021 and 12.5.2022: Moravskoslezsky Region



IES Working Paper Series

2025

1. Kseniya Bortnikova, Josef Bajzik, Evzen Kocenda: *How Do Event Studies Capture Impact of Macroeconomic News in Forex Market? A meta-Analysis*
2. Zuzana Meteláková, Adam Geršl: *Does Bank Regulation and Supervision Impact Income Inequality? Cross-Country Evidence*
3. Tersoo David Iorngurum: *Inflation Differentials in the African Economic Community*
4. Lorena Skufi, Adam Gersl: *Does FX Hedge Mitigate the Impact of Exchange Rate Changes on Credit Risk? Evidence from a Small Open Economy*
5. Meri Papavangjeli: *From Skies to Markets Implications of Extreme Weather Events for Macroeconomic and Financial Imbalances in CESEE Countries*
6. Matej Opatrny, Milan Scasny: *Bridging the Gap: A Novel M2/LIHC Hybrid Indicator Unveils Energy Poverty Dynamics - Case Study of the Czech Republic*
7. Vojtěch Mišák: *Temperature and Productivity in Soccer*
8. Klara Kantova, Tomas Havranek, Zuzana Irsova: *The Elasticity of Substitution between Native and Immigrant Labor: A Meta-Analysis*
9. Suren Karapetyan, Matej Bajgar: *Unicorn Exits and Subsequent Venture Capital Investments*
10. Jonáš Čekal, Adam Geršl: *The Effects of Crisis Management Measures on the Economy: Evidence from Past Crises*
11. *Samuel Fiifi Eshun, Evzen Kocenda, Princewill Okwoche, Milan Ščasný: Price and Income Elasticities of Industrial Energy Demand in New EU Member States*
12. Lukáš Janásek: *Gradient-Based Reinforcement Learning for Dynamic Quantile*
13. Anna Drahozalová: *Balance of Payments Pressures on the Nominal Exchange Rate: A Fresh Look at Old Ideas*
14. Samuel Fiifi Eshun, Evžen Kočenda: *What Really Drives Financial Inclusion? Evidence from a Meta-Analysis of 3,817 Estimates*
15. Oliver Škultéty: *Workforce Composition Effects on Covid-19 Spread: a SIR model and an Application to the Czech Regions*

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz