

MODELING OF COVID-19 PANDEMIC INDICES AND THEIR RELATIONSHIPS WITH SOCIO-ECONOMIC INDICATORS

Mikhail Pomazanov¹

DOI: <https://doi.org/10.31410/ERAZ.S.P.2020.11>

Abstract: *This paper presents non-classical models for estimating and forecasting COVID-19 pandemic indices. These models have been successfully tested on country data where the pandemic is nearing completion. In particular, an effective algorithm for mortality index evaluation is also presented. This index is usually replaced by more simple estimates such as, for instance, „the number of deaths divided by the number of infected”; however, while the virus is at the stage of its rapid distribution, such superficial approaches are incorrect. Model indicators of the infection itself allow us to predict not only the apogee of the epidemic and the end of the quarantine period, but also the maximum number of infected people in some country (continent) during the height of the epidemic.*

The second part of the paper is devoted to an attempt to build regression models to explain (with using 100+ country socio-economic indicators taken from the World Bank data) the behavior of the epidemic spread indices. It is shown that the maximum number of infected people in the country is well predicted (R-square is close to 90%); and, moreover, migration indicators and the number of international air take-offs are effective regressors. Other indicators, for example, the mortality index, are difficultly modeled; nevertheless, it has a significant relationship with socio-economic factors.

The presented paper might be valuable for making effective decisions to forestall some future pandemics or even the „second wave” of COVID-19.

Keywords: *COVID-19, Pandemic, Modeling, Socio-economic indicators.*

1. INTRODUCTION

Among classical epidemiological models, there are three main types of deterministic (parametric) models for infectious diseases that are spread by direct human-to-human contact in the population. They are presented in Herbert and Hethcote, 1989. The abbreviation „S. I. R” means the proportion of healthy S, the proportion of infected I and the proportion of recovered/deceased R who have immunity, $S + I + R = 100\%$. These three types of models differ in that the first does not take into account the immunity of the SIS-type („healthy”-”infected”-”healthy”) $S + I = 100\%$, the model SIR („healthy”-”infected”-”recovered”) is used for diseases where infection gives permanent immunity; the third type of models takes into account natural birth/death in the population, and this type is focused on long-term modeling of epidemics. When a disease modeled by SIR passes through a population in a relatively short time (less than one year), this outbreak is called an epidemic. Since the epidemic occurs relatively quickly, the model does not include birth and death (life dynamics). Epidemics are common in diseases such as influenza, measles, rubella, and chicken pox. An overview, including non-parametric models, is given in Choisy et al., 2007.

Classical parametric models have serious drawbacks in order to model the COVID-19 global pandemic. One of the main issues is the lack of consideration in the model of active disciplinary actions to prevent the development of the epidemic. In the models, the parameters of mutual infection are

¹ Higher School of Economics National Research University, Main campus: 20 Myasnitskaya Ulitsa, Moscow, Russia

assumed to be constant, i.e. «the diseases are left to themselves». Due to obvious empirical observations boundary conditions on the simulation results are imposed; namely, at the beginning of the epidemic the number of infected as the number of infected (sick at a given time) increases exponentially, starting from a certain conditional „zero” level, that can be chosen empirically, for example - 100 infected people. The number of infected (patients) at the end of the outbreak can be considered tending to zero or allow a small background level of the growth rate of the total number of infected. The overall level of the total number of infected people covers a small percentage of the population, $S > I + R$, so it does not make sense to incorporate a percentage with immunity in the model (i.e. SIR is not suitable). For the SIS model, the solution of the differential equation in finite functions might exist (the Bernoulli equation in the SIS model), but it has the property of exponential growth (at the beginning of development) only at such a parameter value that the level of infected people does not tend to zero after the peak of the disease. The next paragraph will present an alternative parametric infection model that is being successfully tested in the Ebola virus epidemic (Chapter 3).

During an outbreak of a new or emerging infectious agent such as COVID-19, one of the most important epidemiological quantities to be determined is the mortality rate (indicator), which is the percentage of cases that eventually die from this disease. This ratio is often estimated using the combined number of cases and deaths at one time, such as those compiled daily by the world health organization during the COVID-19 epidemic (WHO. Coronavirus disease, 2019). However, simple estimates of the fatality ratio obtained from these reports can be misleading if the result is unknown for the infected but not recovered proportion of patients at the time of analysis. Estimates obtained during the SARS epidemic by dividing the number of deaths by the total number of reported cases were much lower (3-5 percent during the first few weeks of the global outbreak) than estimates obtained using appropriate statistical methods, and varied significantly across countries. Moreover, as the epidemic progressed, these naive statistical estimates falsely indicated an increase in the death rate, which fueled an already high level of public anxiety among the affected population. Therefore, this assessment requires modeling.

Ghani et al. (2005) propose a method for this, based on the Kaplan-Meyeri survival model, and evaluate its indicators using data from the 2003 severe acute respiratory syndrome epidemic in Hong Kong. In Chapter 4, we will propose an effective and simpler method for estimating the mortality index, which uses the result of approximating the time series of the number of infected, recovered, and deceased patients obtained from the model discussed in Chapter 2. The method defines two indices – the mortality index and the indicative period of recovery (conditional hospitalization).

2. THE INFECTION MODEL AND PEAK PARAMETERS

To explain the dynamics of the spread of a short-term epidemic, it is proposed to use the corrected classical Lotka-Volterra model (L-V)², which describes the interaction of two species, one of which is a predator, and the other is a victim (for example, the ecological system „carp-pike” or „hares-lynx”).

In the case of an epidemic, the victim is the number(percentage) of the population available for infection $S(t)$, which at the time of t is equal to 1, $S(\check{T}) = 1$. It is assumed that the population takes measures to dissociate the „victim” by the simplest law

² The model arose historically (1931) in connection with an attempt to explain the fluctuations of fish catch in the Adriatic Sea (Volterra, 1976). The same system of differential equations was proposed by Lotka a little earlier (1924), but Volterra much more fully analyzed this system.

$$S'(t) = -\alpha \cdot S(t) \quad (1)$$

where α is the constant responsible for the effectiveness of the measures to be taken.

The "predator" in this model is the number of infected $H(t)$. The predator, similar to the L-V model, reproduces by law

$$N'(t) = \gamma \cdot \alpha \cdot S(t) \cdot N(t) \quad (2)$$

where γ is the constant responsible for the scale of infection.

It is obvious that the solution of equations (1) and (2) will be the function responsible for the absolute number of infected, of the form

$$N(t) = \hat{N} \cdot e^{\gamma(1-e^{-\alpha(t-\hat{t})})} \quad (3)$$

where \hat{N} – the number of infected at the time of the conditional beginning of the epidemic \hat{t} . It is clear that \hat{N} , \hat{t} are not independent near the conditional beginning of the epidemic, but can be changed without changing the type of function (3) by rationing taking into account α , γ . Indeed, for $\alpha(t - \hat{t}) \ll 1$, function (3) will have an exponential development of $N(t) \cong \hat{N} \cdot e^{-\gamma\alpha(t-\hat{t})}$ in which $N e^{\{\gamma\alpha\hat{t}\}}$ is an invariant. Then it is reasonable to give an obvious meaning, namely, to put empirically $\hat{N}=100$, and to denote \hat{t} as t_{100} . Therefore, for the normalized function of the number of infected (3) $N(t) = 100 \cdot \exp(\gamma(1 - e^{-\alpha(t-t_{100})}))$ three infection constants t_{100}, α, γ are subjects to determination. These constants responsible for the development of the epidemic should be determined by the dynamics of infection by approximating by minimizing the distance functional, which will be discussed below.

The next practical step is to determine the peak infection parameters. The first important parameter is the moment of the peak infection rate for the model (3). In other words, we are looking for the moment t^* , at which the maximum $N'(t)$ is reached. The problem is easily solved by solving the equation $N''(t) = 0$. The solution is the value

$$t^* = \hat{t} + \frac{\ln(\gamma)}{\alpha}, \quad (4)$$

which at the peak gives the maximum increase in the number of cases per day

$$\dot{N}^* = \hat{N} \cdot \alpha \cdot e^{\gamma-1}. \quad (5)$$

the \hat{N} parameter estimates the maximum level of load on the medical infrastructure.

Further, it is possible to predict the maximum number of infected people for the entire time of the epidemic, which will be

$$N_{max} = \hat{N} \cdot e^{\gamma} \quad (6)$$

The timing of recovery from the epidemic, which regulates socio-economic decisions to ease quarantine measures, self-isolation, and so on, can be set for simple practical reasons. Namely, by the percentile of growing infections up to N_{max} . I.e., the term T_p is calculated from the percentile definition

$$p = \frac{N_{max} - \hat{N} \cdot e^{\gamma(1 - e^{-\alpha(T_p - \hat{t})})}}{N_{max}} \quad (7)$$

From (7) follow for small p a simple formula of benchmarks for the timing of measures to get out of restrictions

$$T_p = t^* + \frac{1}{\alpha} \ln\left(\frac{1}{p}\right), \quad (8)$$

where t^* is the peak of the epidemic (4). As a reasonable option, you can take the first forecast of the beginning of easing the restrictions at the level of $p=10\%$ (Finish of high security "90%"), the second – the forecast of removing restrictions at the level of $p=1\%$ (Finish of self-isolation "99%") for (8).

Using the presented elementary formulas, it is possible to predict the most important characteristics of the epidemic during its development, daily updating the parameters based on newly received data.

To determine the best values of the epidemic parameters, an optimizing functional is selected that simulates the distance to the empirical infection curve ($C(t_i)$ $i = 1 \dots T$, T is the period of current observation) in the standard, taking into account the approximation of derivatives (i.e. infection rates) as well. The functionality is offered in the form of:

$$F(\vec{p}) = \sum_{i=1}^T \frac{(C(t_i) - N(t_i, \vec{p}))^2}{C(T)^2} + \lambda \cdot \sum_{i=2}^T \left(C(t_i) - C(t_{i-1}) - \dot{N}(t_i - \frac{1}{2}, \vec{p}) \right)^2 \cdot \frac{(i-1)^2}{C(t_i)^2} \quad (9)$$

where \vec{p} – vector of parameters $(\alpha, \gamma, \hat{t})$,

t_i – daily data publication dates,

t_1 = the first date when \hat{N} of infected ($C(t_i) \geq \hat{N}$, e.g., $\hat{N} = 100$),

$C(t_i)$ – number of infected,

T – current rate of infection outbreak.

We suggest to use the following initial values to find the minimum distance:

$$\alpha = 0, \gamma = \ln\left(\frac{C(T)}{\hat{N}}\right), \hat{t} = t_1.$$

The parameter λ is chosen empirically and characterizes weight smoothing rate of infection by function (3). Authors' experience on the basis of empirical calculations suggests $\lambda \in (1; 4)$.

3. TESTING A PARAMETRIC MODEL OF INFECTION ON THE SPIKES OF THE EBOLA EPIDEMIC

The model (3) is tested on data on the dynamics of Ebola virus infection (WHO Situation Reports. Number of Cases and Deaths in Guinea, Liberia, and Sierra Leone during the 2014-2016 West Africa Ebola Outbreak). Testing is performed on each date T during the development of the epidemic, without using the values (known to date) after the dates. The result is shown in Figure 1. Charts are arranged in a cascade for easy identification of changes in forecast parameters along the abscissa's axis, which is identical for all charts. On the first (top), the current value of the forecast for the peak date of infection for the model is calculated for the dynamic data. The dotted line

shows the final result (fact). The straight line marked with dots is the current time when the forecast was made. On the second-the forecast value of the maximum number of infected C_{max} in the aggregate of foci of infection (African countries). Below the graphs (thin lines and dots) are the infection dynamics: the model (as of the end date) and the fact. On the third (lower) – the number of infections per day: the model and the fact. This is calculated for the end date of the epidemic.

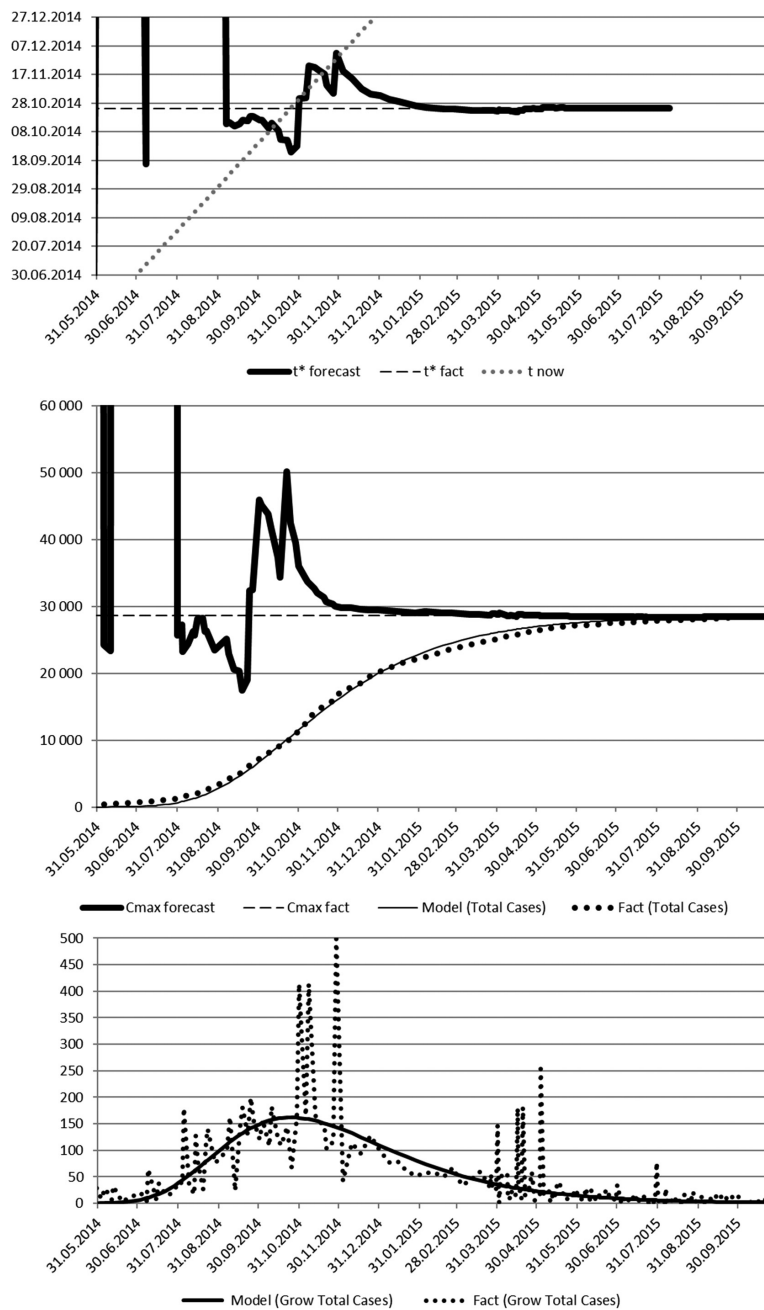


Figure 1. Modeling of peak parameters for the epidemic of the Ebola virus.

Data source: WHO Situation Reports

It can be seen that the infection dynamics is approximated fairly accurately by the model (3). Forecasts of the peak values of the maximum infection rate begin to approach the exact values when the fluctuations in forecasts subside. The approach to the exact values of ± 10 thousand people (up to 30%) begins with the period of approach of the epidemic period to the peak of infection. Before reaching the peak, forecast values are highly volatile and not significant.

The first graph shows that the maximum begins to form when the forecast of the peak date intersects with the observation date. On an earlier observation date, the peak forecast is possible, but less reliable.

On empirical observations of the passage of the COVID-19 infection peak, after the formation of the peak date that intersects with the observation date, the final date of the peak recedes back (about a week) and a steady trend begins to decrease the intensity of infection, i.e. the regression of the epidemic.

4. MODEL INDEX MORTALITY

There are three parameters for monitoring that are updated daily, at time t . This is the number of cases in which the virus is confirmed (Confirmed, $C(t)$), the number of recovered (Recovered, $R(t)$), and the number of deaths (Deaths, $D(t)$). Up-to-date data and graphs of these parameters can be seen on the Johns Hopkins CSSE University website (Johns Hopkins CSSE, 2019). These parameters are sufficient to determine the death rate as a percentage of patients with $RIP(t)$ and determine the average signal duration of the disease, i.e. hospitalization, $T(t)$. Due to the spread of the disease and the fact that new cases have not yet died, the value of $RIP(t)$ will be greater than the minimum value of low $RIP(t) = D(t) / C(t)$ and less than the limit of mortality among those who have already been ill or died, equal to up $RIP(t) = D(t) / (D(t) + R(t))$. LowRIP and upRIP parameters are called naive estimates in the literature. If the spread of the epidemic stops and it takes a long time, these parameters will be equal to $upRIP = lowRIP = RIP$. However, this is not the case during the development of the epidemic, $lowRIP(t) < RIP(t) < upRIP(t)$. Therefore, a model is required to evaluate $RIP(t)$ and $T(t)$, which will also be significant indicators of the effectiveness of the disease treatment process, a measure of the threat to the life of any person who may become a victim of the epidemic, and a macroeconomic factor that affects the economy and GDP as a whole.

To model the desired parameters, certain simplifying assumptions must be proposed. They should be as natural as possible and preferably simple. So:

- First – the sick patient is hospitalized for a certain time T , the same for all, after this period, he either recovers (Recovered) or is dead (Deaths).
- Second – during the period of illness $s = 0 \dots T$, the proportion of dying patients is distributed evenly and is equal to $d(s) = RIP \cdot s / T$, i.e. at the end of the period, this proportion is just equal to $d(T) = RIP$.

Then the number of recovered patients at the current (or past) time t can be formed only from those who were infected with the term T in the countdown back (on the date $t-T$)

$$R(t) = C(t-T) \cdot (1-RIP) \quad (9)$$

This will be the first equation in which T and RIP are unknown.

On the other hand, the number of deaths will be formed from those who became ill time T or more ago and from those who became ill recently but die gradually. Then,

$$D(t) = \left(C(t-T) + \frac{1}{T} \sum_{s=t-T+1}^t (C(s) - C(s-1)) \cdot (t-s) \right) \cdot RIP \quad (10)$$

where $C(s)-C(s-1)$ is the number of new cases for each day of the last t period. This will be the second equation of the model (9-10), which will make it possible to determine the desired two parameters RIP and T at each moment t (i.e. $RIP(t)$ and $T(t)$).

As the calculations showed, this problem is uniquely solved and gives the result shown in the graphs figures 2 and 3.

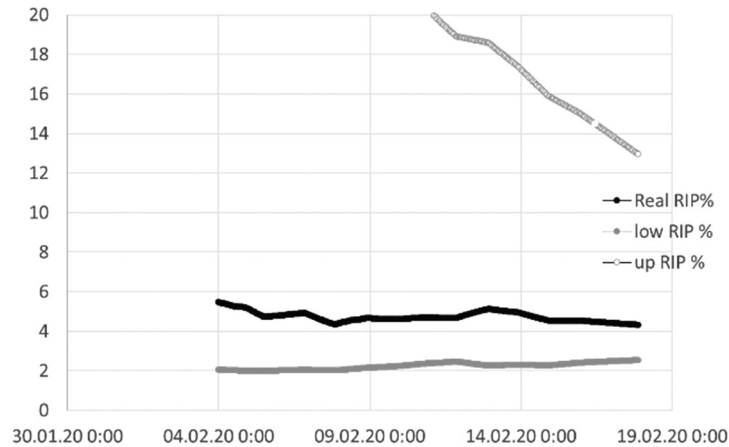


Figure 2. Dynamics of the true mortality rate from coronavirus (as a percentage of cases) for a certain period (in the center) in comparison with the lower and upper bounds obtained from a primitive arithmetic calculation.

Source: author's calculations as of 18.02.2020, given by Johns Hopkins CSSECOVID-19 China

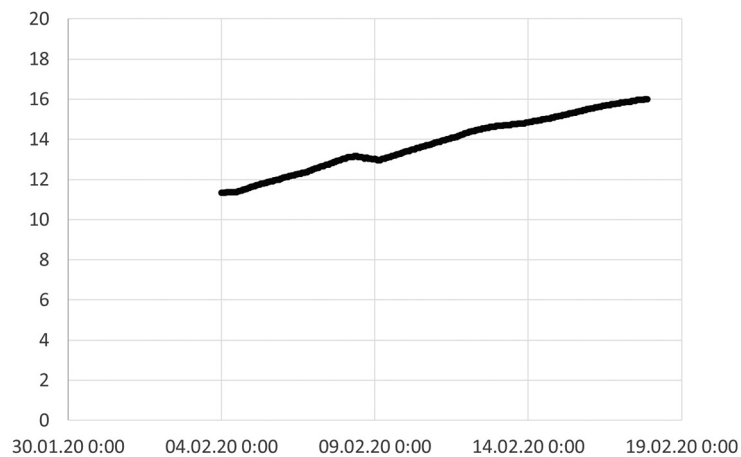


Figure 3. Average calculated dynamics of the duration of hospitalization of a recovering patient (in days).

Source: author's Calculations on 18.02.2020, data from Johns Hopkins CSSE COVID-19 China

From figure 2 it can be seen that there is a tendency to decrease the true mortality rate from 5.5% to 4.5%, but it is too early to say that it is stable in the presented period of the report. Figure 3 shows an increase in the duration of hospitalization, apparently related to medical policies.

For practical implementation of calculations of mortality indices that are resistant to non-smooth dynamics of recorded cases of infection, recovery and death³, it is recommended to switch to the continuous functions $C(t)$, $R(t)$ and $D(t)$.

³ Data is updated no more than once a day.

One way is to approximate these functions by smoothing splines (Hastie and Tibshirani, 1990). An alternative and easier-to-implement interpolation possibility in this problem is provided by using a parametric approximation function of type (3) for all three time series $C(t)$, $R(t)$ and $D(t)$, since $R + D$ is, in fact, a lagging function of $C(t)$, D is a fraction of $C(t)$. In any case, the RIP will be of interest primarily from the point of view of dynamics in the development of the epidemic, and the dynamics will be correctly reflected with a simpler than spline smoothing interpolation. The accuracy requirements for calculating the index are not as high.

When switching to continuous functions, equations (9-10) will be converted to integral relations, the first of which will determine the index $T(t)$ of the period of recovery/hospitalization/death. $T(t)$ will be the solution of a nonlinear equation with unknown T obtained after transformation (9-10):

$$R(t) = C(t - T) \cdot \left(1 - \frac{D(t) \cdot T}{\int_{t-T}^t C(s) ds} \right) \quad (11)$$

After solving equation (11), which is the only one, it is possible to uniquely determine the lethality index using the formula

$$RIP(t) = \frac{D(t) \cdot T(t)}{\int_{t-T(t)}^t C(s) ds} \quad (12)$$

The proposed (11-12) method of calculating the mortality index from the point of view of practice has an obvious advantage in that it is mainly determined by the last period of the disease and infection/recovery/death of all patients, determined by the index period of hospitalization. This means that the proposed RIP index should respond promptly to changes in the conditions of the disease course, associated with the appearance of effective treatment, mutations of the infection carrier, reaching the limit of medical resources and reducing the quality of treatment, changing the number of patients, etc.

Here are some examples of the dynamics of the RIP, lowRIP and upRIP index calculated for countries with early COVID-19 infection as of the reporting date 20.04.2020, as presented in figure 4.

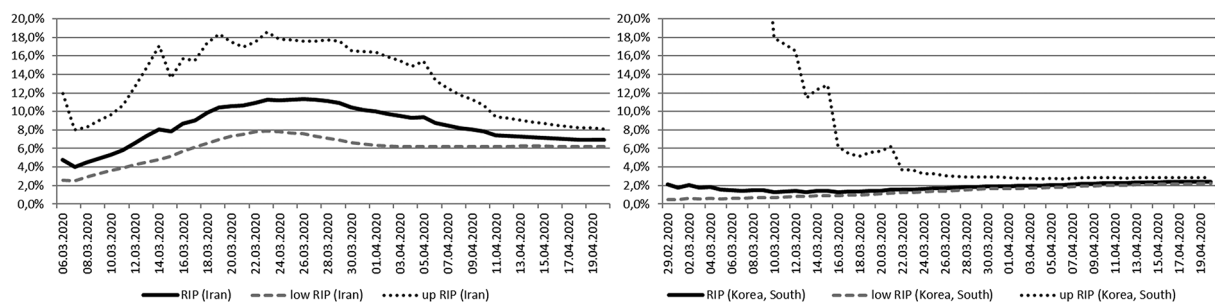


Figure 4. COVID-19 lethality indices for the countries of Iran (left) and South Korea (right).

Source: author's calculations on 29.04.2020,
 data from Johns Hopkins CSSE COVID-19 Iran, South Korea.

From the presented examples, a significant difference in the behavior of RIP indices in both dynamics and absolute value is visible. So, in Iran, the index first rose, then fell, and in South Korea - the opposite. However, the indices differ by 3-4 times in absolute values.

5. EXAMPLES IN TIME-RESULTS OF MODELING OF COVID-19 INDICES IN EUROPEAN COUNTRIES (DATA FROM 15.06.2020)

In order not to clutter this study with reporting material, the author uses two European countries Italy and the United Kingdom as demonstration examples of COVID-19 propagation modeling, as presented in figure 5.

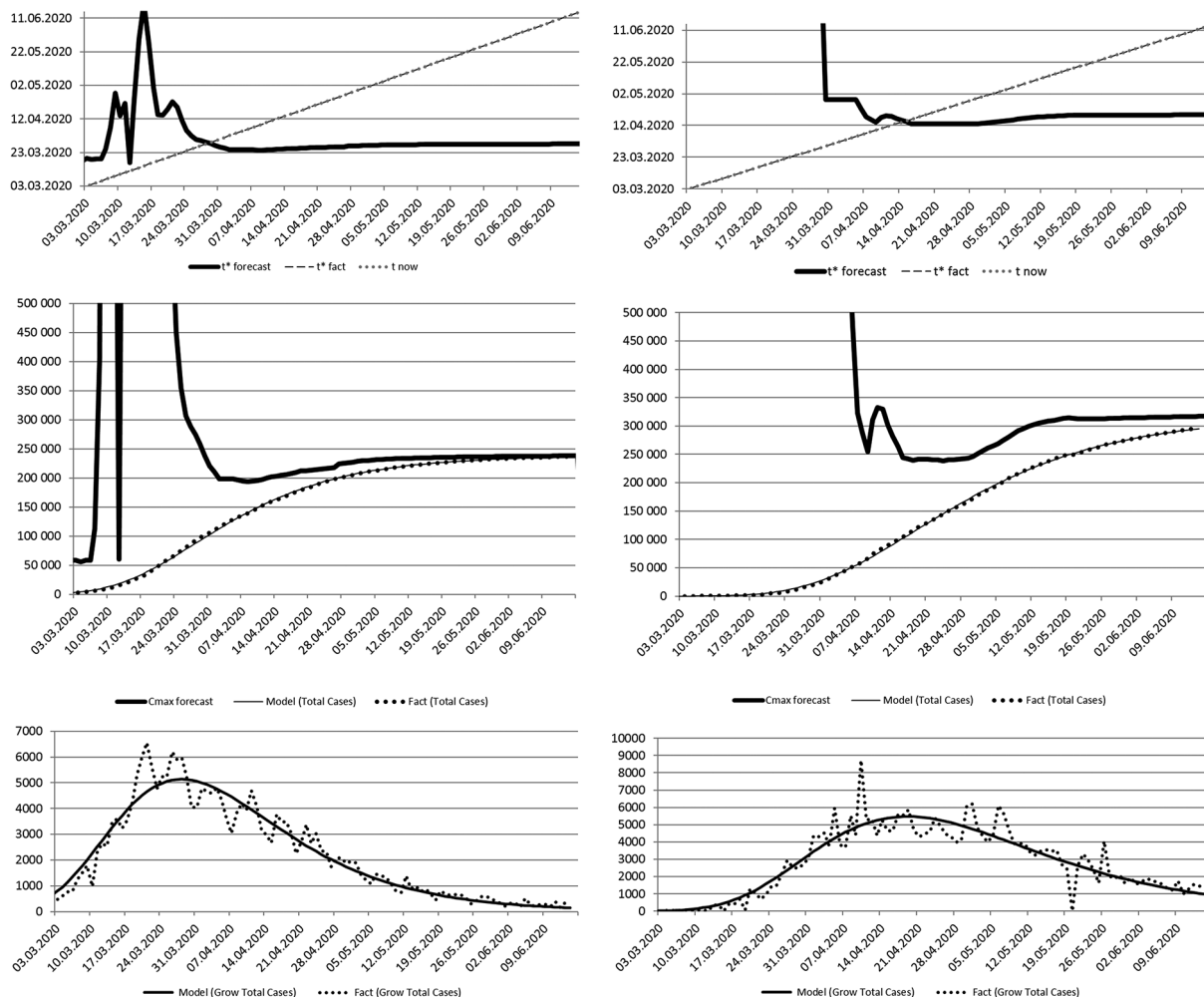


Figure 5. Modeling of peak COVID-19 parameters for Italy (left) and UK (right).

Source: Author's calculations as of 15.06.2020, data from Johns Hopkins CSSE COVID-19

From the first graphs above, it can be seen that the model forecast of peak infection rates becomes relatively stable after the forecast date reaches the date of the peak itself. After this period, the forecast for the peak date is adjusted slightly. Also, the forecast of the maximum number of infected people becomes relatively stable after passing the peak of the epidemic.

However, the presented static model (1), (2) is based on the assumption of a monotonous impact of the socio-economic regime of self-isolation on the containment of the spread of the pandemic. This is not always true, so repeated outbreaks of infection are possible, which are unpredictable and may be the result of mistakes by governments to mitigate the regime, mass riots, natural disasters, etc. The analysis of these processes and their consequences on the dynamics of the spread of the epidemic is the subject of separate scientific research.

Table 1 shows several important parameters of the COVID-19 pandemic for the largest European countries.

Table 1. Some parameters of COvid-19 in the largest European countries.

Country	C now	RIP	Peak data forecast	Start: t_100	Peak - t_100, days	Finish of a high security "90%" (forecast)	Finish self-isolation "99%" (forecast)	max C forecast in % of the population	Deaths forecast
United Kingdom	297 342	26,5%	18.04.2020	05.03.2020	45	06.06.2020	25.07.2020	0,48%	84 071
Italy	236 989	15,1%	28.03.2020	22.02.2020	35	06.05.2020	14.06.2020	0,39%	36 028
France	194 153	17,6%	03.04.2020	09.03.2020	25	02.05.2020	30.05.2020	0,28%	34 084
Germany	187 518	4,8%	30.03.2020	03.03.2020	27	29.04.2020	30.05.2020	0,23%	9 074

Source: author's calculations as of 15.06.2020, data from Johns Hopkins CASE COVER-19

From Table 1 it can be seen that the level of true RIP mortality, calculated according to the model of item 4., for the UK is significantly higher than the average level, the minimum is found in Germany. The period of „rampant” pandemic «Peak - t_100» before the peak period is also significantly higher in the UK than in other countries, as well as the percentage of infected people from the population.

The next chapter will present a result on the relationship of pandemic parameters with socio-economic factors in the statistics of countries that have passed the peak of the pandemic on the current date of 15.06.2020.

6. STUDY OF THE RELATIONSHIP OF SOCIO-ECONOMIC INDICATORS WITH THE PARAMETERS OF THE COVID-19 PANDEMIC

Before you start presenting research results, you need to make a disclaimer related to the preliminary nature of these results. At the time of writing, the COVID-19 pandemic is far from its final phase in the most countries of the World. Therefore, the study of the relationship with socio-economic indicators is based on the forecast (model) values of COVID-19 parameters. By the time this work is published, certain refinements may be relevant. Nothing stands still, but the value of the research lies, among other things, in the method of building leading conclusions made in «in-time» mode. The value of such conclusions is particularly significant if they can prevent future mistakes in socio-economic containment of the spread of the disease.

Of the 84 countries that were affected by the pandemic, with the number of officially infected more than 1000 people on 15.06.2020, only 60 countries that participate in the statistical study passed the peak of infection. Exogenous variables are the parameters of the pandemic. These parameters are expected to change slightly over time for those countries where the peak of the pandemic has already passed by the time of the study.

Exogenous variables (author's calculations on 17.05.2020)

- **RIP** – Fatality (lethality) Index in a given country (model forecast)
- **α** - The constant is responsible for the effectiveness of the measures taken in a given country (model forecast)
- **$\alpha \cdot \gamma$** - Pandemic Rate Index (or detection rate of number of infected) in a given country (model forecast)
- **Cmax** - Maximum number of infected people in a given country (model forecast)

- The endogenous parameters for the regression model will be socio-economic indicators presented by the World Bank (The World Bank. Indicators). There are more than 150 of these indicators, and the study involved indicators calculated for the closest date to the present, which is available in the World Bank database.

Endogenous variables (regressors, World Bank):

- *Agriculture & Rural Development*
Agricultural land (% of land area)
Agricultural methane emissions
etc.
- *Climate Change*
CO2 emissions
Disaster risk reduction progress score
Ease of doing business index
Electric power consumption
etc.
- *Social Development*
Public Sector
Trade
Health
etc.

A nonlinear multiple regression model is constructed, and the minimum number of indicators that are significant for exogenous variables is selected. The quality level of the model is estimated by the R-square coefficient of determination. Additional tests are made, the main of which is the Breusch-Pagan test (Breusch and Pagan, 1979), to check for the presence of heteroscedasticity of random errors in the regression model. Economic consistency is analyzed.

Specifications of regression models

1. Non-linear multiple regression
- $$Y = A + \sum a_i x_i + \sum \sum b_{i,j} x_i x_j + \varepsilon$$
2. All regressors are statistically significant (p-value)
 3. **Regression Quality Level** (High = R-square > 90%, Intermediate = 50% < R-square < 90%, Low = R-square < 50%)
 4. **Heteroscedasticity** (Yes = Breusch - Pagan test, significance > 10%, No is less 10%)
 5. **Economic consistency** of model coefficient signs

The aim of the study is to identify factors and their combinations that explain the COVID-19 pandemic parameters, as well as pandemic parameters that cannot be meaningfully explained using World Bank indicators. More than 1 million models of various combinations of regressors are studied.

The results of the study are presented in Table 2.

For the „True mortality index” parameter, Rip was not able to find any meaningful model. The quality of the regression was low. The economic consistency of signs in variables was ambiguous. To measure the effectiveness of measures taken (to control the pandemic) it was also not possible to find a regression model of any significant quality.

For the parameter $\alpha \cdot \gamma$, which is responsible for Pandemic Rate (or detection rate of number of infected), we found a satisfactory model with an R-square level of 70%. Even more success was achieved for the parameter of the maximum number of infected citizens of those countries that were included in the study. We managed to find a whole series of very effective models that have a high determination (R-square of 90%), a low level of heteroscedasticity and economic consistency of regressors.

Table 2. General result of regression modeling of covid-19 exogenous parameters

Exogenous variables	Interpretation	Regression Quality Level	Heteroscedasticity	Economic consistency
RIP	Fatality (lethality) Index	Low	Yes	Ambiguously
α	The effectiveness of the measures taken	Low	Yes	Yes
$\alpha \cdot \gamma$	Pandemic Rate Index (or detection rate of number of infected)	Intermediate	No	Yes
Cmax	Maximum number of infected people	High	No	Yes

Table 3. The most powerful regressors explaining the parameters of the COVID-19 pandemic identified during the active period for 60 countries (15.06.2020)

Target variable		Regressor№1	Regressor№2	Regressor№3
Pandemic Rate Index (or detection rate of number of infected)	69%	Railways, passengers carried (million passenger-km) x Specialist surgical workforce (per 100,000 population)	GDP per capita, PPP (current international \$) x ln Hospital beds (per 1,000 people)	Life expectancy at birth, total (years) x International tourism, number of arrivals
		GDP per capita, PPP (current international \$) x ln Hospital beds (per 1,000 people)	International tourism, number of arrivals x ln Agricultural land (% of land area)	Railways, passengers carried (million passenger-km) x GDP 2018
		Air transport, passengers carried x Agriculture, forestry, and fishing, value added (% of GDP)	Life expectancy at birth, total (years) x Life expectancy at birth, total (years)	International tourism, number of arrivals x ln Hospital beds (per 1,000 people)
		Railways, passengers carried (million passenger-km) x Specialist surgical workforce (per 100,000 population)	GDP per capita, PPP (current international \$) x ln Hospital beds (per 1,000 people)	International tourism, number of arrivals x ln Life expectancy at birth, total (years)
		Railways, passengers carried (million passenger-km) x Specialist surgical workforce (per 100,000 population)	GDP per capita, PPP (current international \$) x ln Hospital beds (per 1,000 people)	International tourism, number of arrivals

max C forecast	89%	Netmigration x Population	CO2 emissions (metric tons per capita) x Population	
		Netmigration x Population	Air transport, registered carrier departures worldwide x Population	International tourism, number of arrivals x International tourism, expenditures (% of total imports)
		Air transport, passengers carried x International tourism, expenditures (% of total imports)	Netmigration x Population	

The purpose of this part of the study is not to present «Plug and Play» forecast models, but only to show what socio-economic factors initially determine the scale and speed of the pandemic. And they turned out to be logical, and hypothetically they should have been proposed, without any statistical research. But the value lies in the fact that the determining factors are confirmed by statistical research.

So, we make sure on the statistics COVID-19, which is the number of infected in the country is influenced by uncontrollable factors – population industrial capacity of the country and its production (as expressed in GDP, CO2, etc.)

And managed ones, namely:

- • power communication (migration, the number of human exchanges through the travel and transportation including railway transport), tourism, etc.

Including visible factors responsible for the speed of social reaction. These are the country’s medical parameters:

- • In Hospital beds (per 1,000 people)
- • Life expectancy at birth, total (years)

These are the key points that you should pay attention to first of all to prevent consequences. If tourism is harmful (in terms of the risk of epidemics), then this is true. But it is not possible to stop it, it is necessary to strengthen infection control. This is also obvious for the level of development of medicine.

7. CONCLUSION

The theoretical part of the presented work offers a very relevant, in our opinion, model that allows us to make adequate predictions about the spread of the epidemic at the moment of its active phase. This is especially relevant for the operational management of the socio-economic risks posed by the epidemic (pandemic). Unlike classic epidemic models, the presented model takes into account socio-economic counteraction to infection, focused on creating restrictions in the form of self-isolation, quarantine measures, etc. The model is static, in the sense that the factor of the reaction of society is assumed to be constant. In those countries where this hypothesis is acceptable, infection dynamics are observed in full accordance with the model. We

understand that in practice this is not quite true, therefore, second and other subsequent waves that are observed are possible. How and from what they come is a topic of future research. The paper proposes a certain basis for further, more complex modeling.

The practical part of study showed that countries with developed infrastructure for international and domestic movements, as well as high migration, have the maximum number of infected. It clearly demonstrates the need for greater monitoring of these activities to prevent future pandemics.

REFERENCES

- Breusch T. S., Pagan A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation". *Econometrica*. 47 (5): 1287–1294. doi:10.2307/1911963. JSTOR 1911963. MR 0545960.
- Choisy M., Guégan J.-F., and Rohani P. (2007) Mathematical Modeling of Infectious Diseases Dynamics. Chapter 22, *Encyclopedia of Infectious Diseases: Modern Methodologies*, by M. Tibayrenc. Copyright © 2007 John Wiley & Sons, Inc.
- Ghani C. et al. (2005) Methods for Estimating the Case Fatality Ratio for a Novel, Emerging Infectious Disease. *American Journal of Epidemiology*, Volume 162, Issue 5, 1 September 2005, Pages 479–486, <https://doi.org/10.1093/aje/kwi230>
- Hastie T. J., Tibshirani R. J. (1990) *Generalized Additive Models*. — Chapman and Hall, ISBN 0-412-34390-8.
- Herbert W., Hethcote (1989) Three Basic Epidemiological Models. *Applied Mathematical Ecology*. pp 119-144. Editors Simon A. Levin Thomas G. Hallam Louis J. Gross. Springer-Verlag Berlin Heidelberg DOI <https://doi.org/10.1007/978-3-642-61317-3>
- Johns Hopkins CSSE (2019) Novel Coronavirus COVID-19 (2019-nCoV) Data Repository, <https://github.com/CSSEGISandData/2019-nCoV>
- The World Bank. Indicators. <https://data.World Bank.org/indicator?tab=all>
- Volterra, V. (1976) *The Mathematical Theory of the Struggle for Existence*. Science, Moscow, 288.
- WHO Situation Reports. Number of Cases and Deaths in Guinea, Liberia, and Sierra Leone during the 2014-2016 West Africa Ebola Outbreak, <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html>
- WHO. Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>