Predicting the Dynamics of Covid-19 Propagation in Azerbaijan based on Time Series Models

SAKINA BABASHOVA
Turkish World of Economics Faculty, Department of Economics
Azerbaijan State University of Economics
Baku, Istiqlaliyyat str.6, AZ1001
AZERBAIJAN

Abstract: - The study is dedicated to developing an econometric model that can be used to make medium-term forecasts about the dynamics of the spread of the coronavirus in different countries, including Azerbaijan. We examine the number of COVID-19 cases and deaths worldwide to understand the data's intricacies better and make reliable predictions. Though it's essential to quickly obtain an acceptable (although not perfect) prediction that shows the critical trends based on incomplete and inaccurate data, it is practically impossible to use standard SIR models of the epidemic spread. At the same time the similarity of the dynamics in different countries, including those which were several weeks ahead of Azerbaijan in the epidemic situation, and the possibility of including the heterogeneity factors into the model allowed as early as March 2020 to develop the extrapolation working relatively well on the medium-term horizon. The SARS-CoV-2 virus, which causes COVID-19, has affected societies worldwide, but the experiences have been vastly different. Countries' health-care and economic systems differ significantly, making policy responses such as testing, intermittent lockdowns, quarantine, contact tracing, mask-wearing, and social distancing. The study presented in this paper is based on the Exponential Growth Model method, which is used in statistical analysis, forecasting, and decision-making in public health and epidemiology. This model was created to forecast coronavirus spread dynamics under uncertainty over the medium term. The model predicts future values of the percentage increase in new cases for 1–2 months. Data from previous periods in the United States, Italy, Spain, France, Germany, and Azerbaijan were used. The simulation results confirmed that the proposed approach could be used to create medium-term forecasts of coronavirus spread dynamics. The main finding of this study is that using the proposed approach for Azerbaijan, the deviation of the predicted total number of confirmed cases from the actual number was within 3-10 percent. Based on March statistics on the spread of the coronavirus in the US, 4 European countries: Italy, Spain, France, Germany (most susceptible to the epidemic), and Azerbaijan, it was shown how the trajectory would deviate exponentially from a shape; a trial was carried out to identify and assess the key factors that characterize countries. One of the unexpected results was the impact of quarantine restrictions on the number of people infected. We also used the medium-term forecast set by the local government to assess the adequacy of health systems.

Keywords: - Coronavirus, epidemic spread, regression models, time series analysis, exponential growth model, prediction, applied econometrics.

Received: September 22, 2021. Revised: June 25, 2022. Accepted: July 6, 2022. Published: July 28, 2022.

1 Introduction

The coronavirus pandemic, which emerged at the end of 2019 in the Chinese city of Wuhan and spread almost all over the world in the spring of 2020, has no close analogs in recent decades, so it is tough to predict its dynamics and consequences, including the impact on the economy. The situation with both forecasting and the fight against the new virus is aggravated by its properties such as high infectivity. A rather long incubation period and a high proportion of asymptomatic carriers make the current official statistics very inaccurate, and practically exclude the use of standard models of the spread of epidemics [1], incl. spatial SIR-models [3] describing the dynamics of groups of susceptible, infected, and recovered persons. At the same time, a significant proportion of carriers with severe symptoms, a high mortality rate (also not yet accurately determined,) and the exceptional significance of the impact of the pandemic on the global economy and the economy of individual countries [4] determine the particular relevance of at least medium-term forecasts of the dynamics of the number of infected, including, in the context of developing measures to limit human contacts, monitoring the effectiveness or ineffectiveness of their impact on the rate of spread of the epidemic, and predicting the expected burden on the health system.

A feature of the new COVID-19 virus epidemic is the lack of statistics from previous years. In this regard, there is a problem with good use of the available information on the parameters of the developing epidemic, including in other countries of the world. Many scientific groups in the USA, China, and Europe are working on developing methods to predict the spread of a new virus in the short term. On February 11, 2020, the World Health Organization announced the Global Research and Innovation Forum and identified the most critical research goals in distribution epidemics [5]. In early March 2020, Science Translation Medicine published an editorial [6] in which its authors formulated research directions, the innovative results of which should contribute to the prevention of the spread of the epidemic. They also noted that comprehensive mathematical models that include complex pathogenic and socially significant variables require significant time and effort to develop and verify (often from months to several years). However, mathematical models predicting the dynamics of registration of new cases of COVID-19 in real-time began to appear in journals and online resources [7-12]. The article [7] provides estimates of the extent of the epidemic in Wuhan and other cities, including outside mainland China, to which the virus may have been exported from Wuhan. The authors predict the values of the population's domestic and global health risks from epidemics based on the SEIR (Susceptible-Exposed-Infectious-Recovered) model, taking into account possible scenarios for preventive intervention. The article [9] studied the dynamics of the spread of coronavirus in India using a system of differential equations with constant coefficients. Moreover, the concept of the primary reproduction number, applying the Pontryagin maximum principle to solve the problem of optimizing preventive measures. The work [10] draws attention to the similarity of the dynamics of the total number of infected, recovered, and dead people in China and Italy. It also analyzes the solutions of the system of differential equations adopted in the SIRD (Susceptible-Infected-Recovered-Deaths) model. It notes that although the SIRD model is rather crude, its use gives a good chance to reflect at least the general features of the evolution of the epidemic and predict the dynamics in real-time. The suggested technique aims to predict the peak in Italy in terms of the increase in new infections and the number of deaths over the entire period of the epidemic. The author's articles [10] hypothesize that any country that becomes a theater epidemic outbreak can be seen, at least as a first approximation, as an environment in which different population groups interact according to some general rules, regardless of geographical variations. The authors of the article [11] use the quantitative picture of the spread of the COVID-19 disease in China as a test case and infection data from eight countries to assess the evolution of the epidemiological process in each of these countries. This approach is based on the Gaussian hypothesis of virus spread and the basic SIR (Susceptible-Infected-Recovered) model. Considering difficulties in applying to predict the dynamics of the spread of COVID-19 deterministic models such as SIR, SEIR, and SIRD, which are built on the mechanisms of virus spread from individual to individual and use estimates distribution parameters of known viruses, our efforts will be focused on looking for other methods. An important observation formulated in [10] about the similarity parameters of the epidemic in China and Italy, as well as an analysis of statistics on the spread of coronavirus in the USA, Italy, Spain, France, and Germany led us to a hypothesis about the possible dependence of the growth rate of the leading indicators recorded in Azerbaijan on similar indicators in the countries surveyed.

2 Research Methodology

The baseline study was carried out based on March data on the number of infected in the United States, and the four European countries most affected by the epidemic - Italy, Spain, France, Germany, and Azerbaijan. Its purpose was to build a mid-term forecast of the dynamics of coronavirus infection in Azerbaijan, where at that time the epidemic had just begun, and the number of cases did not exceed the thousandth mark. It was important to estimate at least the order of the numbers for the number of infected and the timing of reaching the plateau since the values in different sources from the end of March to the beginning of April differed not even several times, but tens and hundreds of times. At the same time, the publication of this study to a wider audience one year after the calculations has different aims. First, it is important to demonstrate that it is possible to obtain a forecast of acceptable quality based on a simple extrapolation of data and analogies between countries before the available statistics allow the use of more complex and correct tools (while, of course, you need to understand the limitations of extrapolation and that it only works before the tendency changes). Secondly, econometric models make it possible to identify the significance of certain factors in the context of influencing the resulting indicator, and these results may be important, including for taking certain decisions by the authorities. Thirdly, having real statistics on the dynamics of the spread of the virus in different countries, we see its discrepancies with preliminary estimates, which gives grounds for adjusting measures aimed both at combating the pandemic and its economic consequences. The official statistics on the number of detected cases of infection, presented on the website https://www.worldometers.info/coronavirus [17], were used as the initial data. We realized that these statistics were incomplete and inaccurate. Probably, the actual number of infected in asymptomatic and mild forms exceeds the official figures by several times.

However, the statistics presented quite accurately reflected the tendencies occurring in reality, including the dynamics of the spread of the epidemic, which means that it was possible to focus on them.

Let's summarize the data for the countries most prone to the epidemic, as well as for Azerbaijan in Table 1 (see next page).

2.1 Exponential Growth Model

A growth curve is an empirical model of a quantity's evolution over time. Growth curves are widely used in biology to analyze quantities such as population size in population ecology, demography for population growth analysis, and individual body height in physiology for personal growth analysis. Growth is a fundamental property of many systems, including economic expansion, epidemic spread, crystal formation, adolescent growth, and stellar mass condensation. One of the simple models in which the population grows at a constant proportional rate over time is the exponential growth (unlimited population growth) model [22]. Depending on whether reproduction is assumed to be continuous or periodic, the relationship can be expressed in one of two ways [25]. Exponential growth produces a continuous curve of increase or decrease, the slope of which varies in direct proportion to population size.

$$P_t = y = P_0 e^{rt} \tag{1}$$

where r is the constant rate of growth, P_0 is the initial population size, and t and P_t represent time and population at time t, respectively (Method 1). Another type of exponential curve is shown below.

$$P_t = y = P_0 k^t \tag{2}$$

where $k = \left(\frac{P_n}{P_0}\right)^{1/n}$ and thus the growth rate in Eq. (2) is not a constant growth rate.

In the initial stage, the spread of the virus occurs by the laws of exponential growth. And an exponential model of the form of

$$y = \theta_0 e^{\theta_1 t} \tag{3}$$

which corresponds to the situation of a constant daily increase in the number of infected persons, and can be considered as a benchmark. The differences between the countries consisted only in the initial level and growth rates, which were easy to calculate from the March data, as well as to make a mid-term approximation.

With the current COVID-19 outbreak, we are hearing about exponential growth. In this study, an attempt was made to understand and analyze the data using an exponential growth curve. The reason for using an exponential growth curve to study the pattern of COVID-19 incidence is that epidemiologists have studied these events. It is well known that the first period of an epidemic follows exponential growth. The exponential growth function is not always a perfect representation of the epidemic. Because the exponential curve only fits the outbreak at the beginning, we attempted to fit it first and then studied the logarithmic growth curve. At some point, recovered people will no longer spread the virus, and when someone has been infected, the virus's growth will cease. Logarithmic growth is distinguished by increasing growth in the early period followed by decreasing growth after the point of inflection [23]. In the case of the coronavirus, for example, the maximum limit would be the total number of exposed people in Azerbaijan, because once everyone is infected, the virus's growth will be halted. Following that, the increasing rate of the curve begins to decline and eventually reaches a minimum.

Table 1. The total number of persons infected with COVID-19 by countries, March 2020 (Source: Compiled by the author based on [17])

Date	USA	Italy	Spain	France	Germany	Azerbaijan
1-Mar-20	75	1701	84	130	130	3
2-Mar-20	100	2036	120	191	165	3
3-Mar-20	124	2502	165	212	203	3
4-Mar-20	158	3089	228	285	262	3
5-Mar-20	221	3858	282	423	545	6
6-Mar-20	319	4636	401	653	670	9
7-Mar-20	435	5883	525	949	800	9
8-Mar-20	541	7375	674	1209	1040	9
9-Mar-20	704	9172	1231	1412	1224	9
10-Mar-20	994	10149	1695	1784	1565	11
11-Mar-20	1301	12462	2128	2281	1966	11
12-Mar-20	1630	15113	2950	2876	2745	15
13-Mar-20	2183	17660	4209	3361	3675	15
14-Mar-20	2771	21157	5753	4499	4599	19
15-Mar-20	3617	24747	7753	5423	5813	23
16-Mar-20	4604	27980	9191	6633	7272	25
17-Mar-20	6357	31506	11178	7730	9367	34
18-Mar-20	9317	35713	13716	9134	12327	34
19-Mar-20	13898	41035	17147	10995	15320	44
20-Mar-20	19551	47021	21571	12612	19848	44
21-Mar-20	24418	53578	25496	14459	22364	53
22-Mar-20	33840	59138	29909	16689	24873	65
23-Mar-20	44189	63927	35480	19856	29056	72
24-Mar-20	55398	69176	42058	22302	32991	87
25-Mar-20	68905	74386	50105	25233	37323	93
26-Mar-20	86379	80589	57786	29155	43938	122
27-Mar-20	105217	86498	65719	32964	50871	165
28-Mar-20	124788	92472	73232	37575	57695	182
29-Mar-20	144980	97689	80110	40174	62435	209
30-Mar-20	168177	101739	87956	44550	66885	277
31-Mar-20	193353	105792	95923	52128	71808	298

3 Results & Discussion

3.1 Exponential Growth Model Results

Let us summarize in Table 2 the results of calculations using the exponential growth model, estimated on March 2020 data for each of the countries.

Table 2. Average growth rates of the number of infected by countries and approximation
(Source: Compiled by the author)

Country	Growth rate, %	Growth in 30 days, times	Approximation for 30.04, thousand people
USA	27.6	1483.3	286800
Italy	14.1	51.6	5463
Spain	24.2	657.8	63097
France	19.5	210.6	10976
Germany	21.7	359.7	25832
Azerbaijan	23.1	513.8	1201

This cannot even be called a forecast - the growth rates in the first weeks of the spread of the virus can be very high until a certain critical level is reached, then they gradually decrease. At the same time, all European countries, with insignificant specific features, in contrast to China and other countries of the Far East, where extremely strict quarantine measures were introduced, and rapid identification and localization of foci were carried out, move along approximately the same trajectory, adjusted for the time, which is also possible take into account in the model [24].

Let us make the following clarification: we will calculate the average initial rate of exponential growth not according to data for March 1-31, but for a fixed number of days (for example, 15 or 30 days) from the moment the country exceeded the threshold of 1000 detected cases (before that, random daily fluctuations were too large, and data is too sensitive to single large-scale infections). In different countries, this happened at different times, as shown in Table 3, where, along with the average daily rate of increase in the number of detected cases, the moment of crossing the threshold is indicated. Also, the last column of Table 3 shows how much the growth rates decreased when switching from a two-week to a monthly modeling horizon.

The reported initial growth rates may serve as a rough indicator of the rate at which the virus spreads. In particular, in Azerbaijan, they were slightly lower than in key European countries. Moreover, another advantage could be called a temporary difference of 4 weeks - there was a little more time in Azerbaijan to assess the risks and take measures to localize the foci of the spread of the virus.

Table 3. The average rate of increase in the number of infected after crossing the threshold
(Source: Compiled by the author)

Country	Date of crossing the threshold	Rate of increase for 15 days, %	Rate of increase for 30 days, %	Growth rate decrease , %
USA	11-Mar-20	29.4	20.6	8.8
Italy	29- Feb-20	20	14.8	5.2
Spain	9-Mar-20	23.5	15.6	7.9
France	8-Mar-20	18.9	14.6	4.3
Germany	8-Mar-20	23.5	15.8	7.7
Azerbaijan	11-Apr-20	17.1	14.6	2.5

Unfortunately, this was not sufficient to prevent the epidemic (this, in particular, can be seen in the smallest decrease in the growth rate among all countries when moving from a 15-day to a 30-day horizon). The size of the country is also an objective reason. When the epidemic ends in some regions, an outbreak may occur in others and the process continues. At the same time, exponential growth cannot last forever, and even for a medium-term forecast, more complex models should be considered.

In particular, the growth rate is gradually decreasing from the initial high level to lower values. In the simplest version of the model, this decrease can be linear. The data showed that the base rate of growth at the time the threshold reached 1000 infected was 25.7% per day, slightly differing across countries and decreasing daily by an average of 0.99 percentage points. However, a decreasing linear function always sooner or later goes into the negative region - assuming the same decrease in 40 days in the USA, 37 in Spain, 33 in Italy and Germany, 31 in France, and 26 in Azerbaijan. Moreover, this cannot be the case for the indicator of the dynamics of cumulative dependence (we consider as a resulting indicator the total number of infected people detected since the beginning of the epidemic and not the number of patients at the moment), so it is desirable to change the model specification. As a modified version, we consider the exponential decrease in the relative increase in y from time t. Let us also take into account the decrease in growth depending on the proportion x of infected people in the country. The mechanisms of the effect of this indicator can be very different, but in general, this corresponds to the negative relationship of the limit indicators with the current level of the cumulative value, which is characteristic of many processes, taking into account a large number of undetected asymptomatic cases, as well as the high proportion of cases in the capital and large metropolitan areas with much lower morbidity in the regions and, especially, in the outback. We will also take into account the change in the system for measuring the number of infected in the United States during March 17-23, which led to a surge in the number of registered cases, and the timing of the introduction of the main quarantine measures (March 21 - in the United States, February 23 - in Italy, March 14 - in Spain, March 17 - in France, March 16 - in Germany, March 23 - in Azerbaijan) with a time lag of 5 days (a certain period passes from the moment of infection to detection). The resulting equation (4) looks like this:

$$\widetilde{ln\, y_t} = \underset{(0.099)}{-1.405} \underset{(0.0064)}{-0.025t} - \underset{(0.0001)}{0.0005x_t} + \underset{(0.105)}{0.571m_t}$$

$$\underset{(0.063)}{-0.049q_{t-5}} + \underset{(0.115)}{0.365z_t^{(1)}} + \underset{(0.115)}{0.227z_t^{(2)}} + \underset{(0.113)}{0.519z_t^{(3)}}$$

$$\underset{(0.112)}{+0.096z_t^{(4)}} + \underset{(0.112)}{0.224z_t^{(5)}} + \underset{(0.110)}{0.0015z_t^{(6)}} + \varepsilon_t \quad (4)$$

here, $\tilde{y}_t = y_t / y_{t-1} - 1$ is the relative increase in the number of infected, t is the day since the threshold of 1000 infected was exceeded, x_t - is the number of infected per million people, and m_t is a dummy variable for the period of change in the measurement system in the United States (taking a single value in the period from 17 to 23 March), q_{t-5} is a dummy variable equal to one for the period when quarantine measures are in effect, with an offset of 5 days (from March 26 in the US, etc.), $z_t^{(1)},\ldots,z_t^{(6)}$ is a dummy variable for the USA, Italy, Spain, France, Germany, and Azerbaijan, respectively. Under the estimates of the coefficients, their standard errors are indicated in parentheses. The equation (4) presented above suggests that the base (at the time of passing the thousandth threshold) average daily growth rate of the number of infected is 35.4% in the USA ($e^{-1.405+0.365}$ = 0.354), 30.8% - in Italy, 41.2 % - in Spain, 27.0% - in France, 30.7% - in Germany and 24.5% - in Azerbaijan. At the same time, every day the growth is reduced by 2.56% (note, it is a percentage, not a percentage point!). The proportion of those infected has a significant negative effect. The control for changes in the measurement system in the United States increased the accuracy of the model. At the same time, contrary to what was expected, the data provided did not reveal a significant impact of quarantine measures. The t-statistic, calculated as the ratio of the coefficient estimate to its standard error, equal to 0.049/0.063 = 0.784, means that one cannot trust the negative sign of the coefficient. Any restrictive measures (prohibition of mass events, closure of shopping centers, restaurants, cinemas, sports complexes, and other public places, the transition of several industries, including the education system, to online mode, restrictions on movement, etc.) slow down the spread of the virus, reduce the maximum number of active cases and allow to prevent the collapse of the medical system. On the other hand, they increase the duration of the epidemic and the economic costs associated with reduced economic activity. Therefore, a very important question is how effective the stringency of the constraints is. It is hypothesized that the insignificance of the factor of restrictions may be associated with an inaccurate specification of the model, for example, an erroneous lag between their introduction and the slowdown in the spread of the virus. However, if the lag is increased or decreased, the significance of the introduction of restrictions not only increases but typically decreases or even becomes of the opposite sign (for example, with a lag of fewer than 2 days). Data on the t-statistics of the coefficient in restrictive measures depending on the lag are presented in Table 4.

Table 4. t-statistics of the coefficient in restrictive measures depending on the lag (days)
(Source: Compiled by the author)

0	1	2	3	4	5	6	7	8
0.168	0.692	-0.058	-0.210	-0.399	-0.777	-0.568	-0.506	-0.936

The second hypothesis is related to possible inaccuracies in using a dummy variable to take into account the introduced quarantine measures, which takes only values of zero or one, as well as in indicating the timing of the introduction of these measures since the indicated dates were selected solely based on media reports without serious additional analysis. The site __[18] provides an isolation index reflecting the severity of restrictions and taking values from zero (no restrictions at all) to 100 (use of all measures simultaneously in the strongest edition). Its values normalized to the interval [0; 1] are given in Table 5.

Table 5. Government Response Stringency Index for March 1-31, 2020 by Country
(Source: Compiled by the author based on [18])

Date	USA	Italy	Spain	France	Germany	Azerbaijan
1-Mar-20	0.083	0.699	0.111	0.194	0.250	0.194
2-Mar-20	0.111	0.699	0.111	0.287	0.250	0.194
3-Mar-20	0.111	0.699	0.111	0.287	0.250	0.306
4-Mar-20	0.111	0.745	0.111	0.287	0.250	0.306
5-Mar-20	0.204	0.745	0.111	0.287	0.250	0.306
6-Mar-20	0.204	0.745	0.111	0.287	0.287	0.306
7-Mar-20	0.204	0.745	0.111	0.287	0.329	0.306
8-Mar-20	0.204	0.745	0.111	0.287	0.329	0.306
9-Mar-20	0.204	0.745	0.250	0.287	0.329	0.306
10-Mar-20	0.204	0.824	0.458	0.287	0.329	0.361
11-Mar-20	0.218	0.852	0.458	0.287	0.329	0.361
12-Mar-20	0.301	0.852	0.458	0.287	0.329	0.361
13-Mar-20	0.301	0.852	0.458	0.426	0.329	0.361
14-Mar-20	0.357	0.852	0.671	0.482	0.329	0.528
15-Mar-20	0.412	0.852	0.671	0.482	0.329	0.528
16-Mar-20	0.523	0.852	0.690	0.556	0.421	0.528
17-Mar-20	0.551	0.852	0.718	0.907	0.421	0.528
18-Mar-20	0.551	0.852	0.718	0.907	0.523	0.528
19-Mar-20	0.671	0.852	0.718	0.907	0.551	0.611
20-Mar-20	0.671	0.917	0.718	0.907	0.579	0.611
21-Mar-20	0.727	0.917	0.718	0.907	0.681	0.611
22-Mar-20	0.727	0.917	0.718	0.907	0.732	0.611
23-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
24-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
25-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
26-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
27-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
28-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
29-Mar-20	0.727	0.917	0.718	0.907	0.732	0.685
30-Mar-20	0.727	0.917	0.852	0.907	0.732	0.685
31-Mar-20	0.727	0.917	0.852	0.907	0.732	0.685

However, the use of the isolation index did not significantly change the significance of the quarantine measures. Moreover, with a lag of more than 5 days, the refined model even gave positive values of the corresponding coefficient. The data on t-statistics for this indicator depending on the lag (the period from zero to 8 days between the introduction of restrictions and the impact on the number of infected was considered) are summarized in Table 6.

Table 6. t-statistics of the coefficient under constraints depending on the lag (days) for the model, taking into account the isolation index (Source: Compiled by the author)

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
-0.850	-0.159	-0.442	-0.100	-0.007	0.125	0.059	0.193	0.365

Thus, the available data did not reveal a significant relationship between the severity of quarantine measures and the scale of the epidemic. This is indirectly revealed by the fact that the level of spread of the virus (the number of detected cases per 1 million people) is approximately the same as in countries with relatively strict restrictions - Italy (the maximum index level is 0.935), France (0.907), Russia (0.870), the average is the United Kingdom (0.759), USA (0.745), Germany (0.732) and the lowest - Sweden (0.407) and Belarus (0.194). Taking into account the fact that even in China not all restrictions were introduced (the maximum isolation index was 0.819, although the introduced restrictions were strictly observed), and in other Asian countries the values were even lower (Hong Kong - 0.667, Japan - 0.472), probably a more important factor is precisely the basic measures - restrictions on holding mass events, wearing a mask in public places, transfer to online services, etc. - and their unconditional implementation everywhere. At the same time, many of the restrictions introduced, including in Baku, i.e., bans on single walks in parks, access control, etc. - do not lead to a decrease in the scale of the epidemic.

3.2 Forecasting

Let us move on to forecasting. We will demonstrate a medium-term forecast for each of the countries based on the base equation (2) with a dummy variable for quarantine measures and a lag of 5 days. This forecast was presented on April 1, 2021. Some of its results (forecasts for April 15, May 1, May 15, and June 1, 2021) are presented in Table 7.

Table 7. Forecast of the number of infected people for the specified date in the equation (2), persons (Source: Compiled by the author)

Country/Date	**15.04.21**	**1.05.21**	**15.05.21**	**1.06.21**
USA	695863	1096024	1299450	1445955
Italy	174274	217809	240786	258072
Spain	179861	224934	246820	262629
France	126103	184489	216504	240651
Germany	181223	262127	304735	336261
Azerbaijan	1222	1824	2808	6845

Since a year has passed since the forecast, it is possible to assess its accuracy. Table 8 demonstrates the percentage deviation of actual values from the forecast for the specified dates.

Table 8. Deviation of the actual number of infected persons from the forecast in the equation (2), % (Source: Compiled by the author)

Country/Date	**15.04.21**	**1.05.21**	**15.05.21**	**1.06.21**
USA	-5.9	3.5	14.8	29.6
Italy	-5.2	-4.8	-7	-9.6
Spain	0.4	8	11.2	9.2
France	-15.8	-29.4	-34.4	-36.8
Germany	-25.6	-37.4	-42.3	-45.4
Azerbaijan	2.5	1.6	5.8	-20.9

Taking into account the replacement of the dummy variable q_t of the quarantine measures by the isolation index $\tilde{q}_t$, the equation (5) will be as follows:

$$\widetilde{ln\,y_t} = \underset{(0.141)}{-1.411} \underset{(0.0078)}{-0.0285t} - \underset{(0.0001)}{0.0005x_t}$$

$$+\underset{(0.105)}{0.585m_t} + \underset{(0.170)}{0.021\tilde{q}_{t-5}} + \underset{(0.136)}{0.375z_t^{(1)}}$$

$$+\underset{(0.116)}{0.207z_t^{(2)}} + \underset{(0.126)}{0.512z_t^{(3)}} + \underset{(0.128)}{0.102z_t^{(4)}}$$

$$+\underset{(0.140)}{0.231z_t^{(5)}} + \underset{(0.114)}{0.0027z_t^{(6)}} + \varepsilon_t \qquad (5)$$

At the same time, there is no significant difference between equations (4) and (5). In particular, the increase in the number of infected people is decreasing daily not by 2.56%, but by 2.85%. There are other small quantitative differences as well. As Table 9 shows, for some countries the forecast is getting slightly better, for others, on the contrary, it is slightly worsening.

Table 9. Deviation of the actual number of infected persons from the forecast in the equation (5), %
(Source: Compiled by the author)

Country/Date	15.04.21	1.05.21	15.05.21	1.06.21
USA	-5.9	4.5	16.7	32.5
Italy	-3.7	-2.3	-4	-6.1
Spain	0.4	8.5	12.1	10.6
France	-15.3	-28.3	-32.9	-34.8
Germany	-25.2	-36.5	-41.1	-43.8
Azerbaijan	2.7	2.4	3.7	-15.8

At the same time, if the forecast for April can be considered acceptable, then in the forecast for May, and even more so in the longer-term forecast, significant systematic biases are found. In Germany and partly in France, the epidemic began to come to an end faster than it was seen in March. At the same time, in the United States and especially in Azerbaijan, the departure from the trajectory of exponential growth is slower than expected. Moreover, although the growth has slowed down (and has practically become linear), it continues, and to date, the lag behind most European countries in terms of tendency is not 2-3 weeks, but more than 1.5 months.

The 1.5 million forecasted figures for the USA for June, which seemed to be significantly overestimated in March, were exceeded. Brazil, which at the end of May took second place in the world, and several other countries follow the same trajectory.

What is the basis for clustering countries with a faster and slower recovery from a pandemic? The size of the country can be suggested as a hypothesis. There is a meaningful explanation for this. Large countries are very heterogeneous, so while in some parts (for example, in the capital or several major metropolitan areas) reaching the plateau has already occurred, in other parts the outbreak is just beginning. On the contrary, at the initial stage, the number of cases is reduced, since the epidemic has not yet affected a significant part of the country. Open borders between regions with different levels of morbidity aggravate the situation.

In the end, it turns out that size matters, and in large countries, the decrease in the growth of the number of infected persons is slower. For example, this can be modeled by dividing the coefficient at t by the area of the country to some small degree α. If we set the parameter α equal to 0.1 (this means that a 10 times larger country will be characterized by a 20% slower decrease in the growth rate of the number of infected: $0{,}1^{0.1} \approx 0{,}7943$), then the modified model specification will look like this equation (6):

$$\widetilde{ln}\, y_t = \underset{(0.137)}{-1.411} \underset{(0.0070)}{-0.0285}t - \underset{(0.0001)}{0.0004}x_t + \underset{(0.104)}{0.596}m_t - \underset{(0.156)}{0.011}\tilde{q}_{t-5} + \underset{(0.124)}{0.315}z_t^{(1)} + \underset{(0.119)}{0.263}z_t^{(2)} + \underset{(0.124)}{0.520}z_t^{(3)} + \underset{(0.127)}{0.130}z_t^{(4)} + \underset{(0.140)}{0.270}z_t^{(5)} + \underset{(0.135)}{0.0025}z_t^{(6)} + \varepsilon_t \quad (6)$$

Here S_i is the area of the *i*-country. By varying, the value of the parameter α, it is possible to some extent to enhance or weaken the influence of the size of the country. With α = 0, we get the original equation (2). If α = 0.2, the model will predict large numbers of people infected in the United States and accelerate the end of the epidemic in Germany and Azerbaijan. The forecast for α = 0.1 and the deviation from it are presented in Tables 10 and 11.

Table 10. Forecast of the number of infected persons for the specified date in the equation (4) with α = 0.1, persons
(Source: Compiled by the author)

Country/Date	**15.04.21**	**1.05.21**	**15.0521**	**1.06.21**
USA	787348	1290333	1550012	1740681
Italy	170778	210693	230700	244800
Spain	184247	231092	252931	267925
France	125692	183196	213884	236126
Germany	178494	256129	295553	323208
Azerbaijan	1091	1485	2586	6653

Table 11. Deviation of the actual number of infected people from the forecast in the equation (4), % (Source: Compiled by the author)

Country/Date	15.04.21	1.05.21	15.0521	1.06.21
USA	-16.8	-12.1	-3.7	7.7
Italy	-3.3	-1.5	-3	-4.7
Spain	-1.9	5.1	8.5	7
France	-15.5	-28.9	-33.6	-35.6
Germany	-24.5	-35.9	-40.6	-43.1
Azerbaijan	-10.7	-18.6	-7.9	-2.8

In addition, for greater clarity, we will present these data in the graphs (see Fig.1-6).

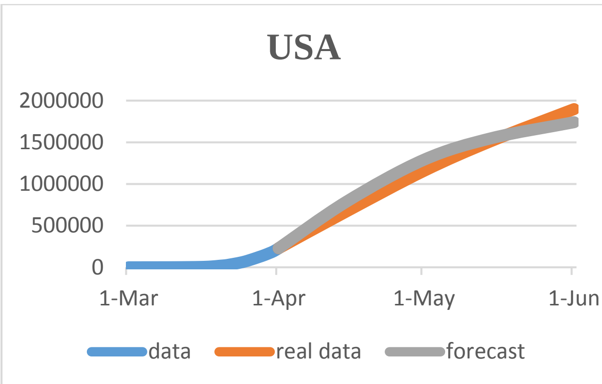

Fig. 1: Comparison of predicted and actual data of the USA (Source: Compiled by the author)

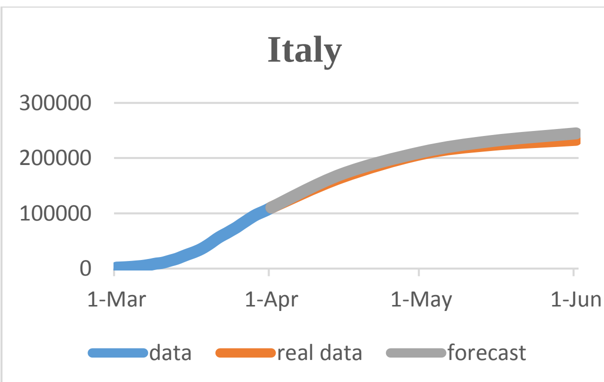

Fig. 2: Comparison of predicted and actual data of Italy (Source: Compiled by the author)

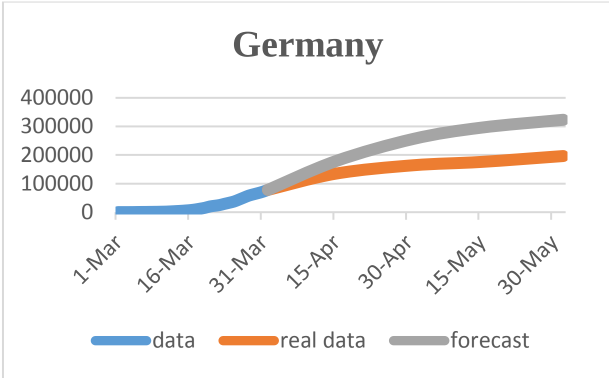

Fig. 3: Comparison of predicted and actual data of Germany (Source: Compiled by the author)

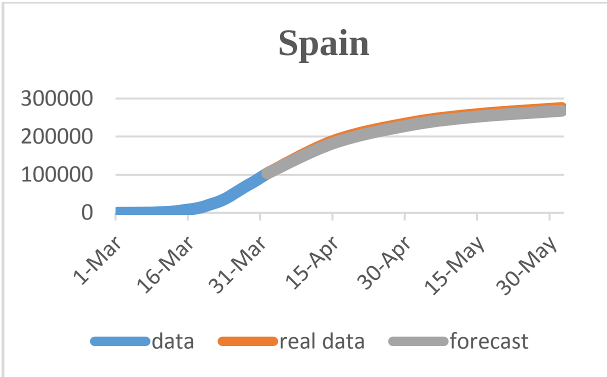

Fig. 4: Comparison of predicted and actual data of Spain (Source: Compiled by the author)

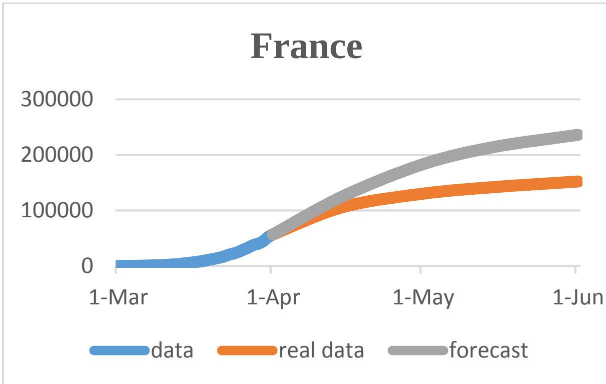

Fig. 5: Comparison of predicted and actual data of France (Source: Compiled by the author)

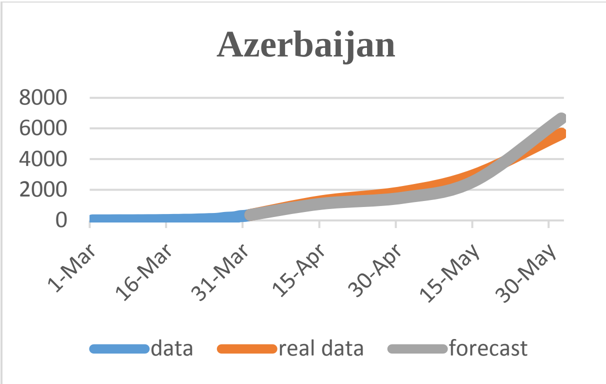

Fig. 6: Comparison of predicted and actual data of Azerbaijan (Source: Compiled by the author)

In general, the graphs demonstrate a sufficiently high accuracy for forecasting based on March data. Suffice it to say that as of March 31, the number of detected cases in the United States was 10 times fewer and in Azerbaijan 18 times fewer than at the beginning of June. Forecasts for Germany and France turned out to be slightly less accurate - two countries that at the end of March were supposed to be the next leaders after Italy and Spain in terms of the number of infected, but according to official statistics, at the beginning of June they are respectively in 9th and 12th places. The process of reaching a plateau in these countries occurred much faster and at a lower level than it was predicted. At the same time, the current statistics are also not final, since it contains such artifacts as the impossible in reality reduction in the cumulative number of detected cases on April 29 and June 2, as well as unlikely sharp fluctuations in the levels of the series. In addition, if the current method of tracking the infected is recognized as more correct, most likely the March data should also be retrospectively corrected, which will change the forecast. Unlike four European countries, where the model very accurately revealed the shape of the curve (and for Italy and Spain also the quantity), a different scenario was realized in Azerbaijan. The exponential tendency changed to a linear one (which means reaching a plateau - the number of new cases coincides with the number of recovered patients). Among other things, this means a significant delay in overcoming the epidemic compared to European countries and the need for significant measures to support the economy. Indeed, the depth of the emerging problems and the speed of economic recovery depends on the duration of the epidemic and the actions of the state. If restrictive measures continue for 1-2 months or the state, through fiscal and monetary policy, does not allow a downward spiral to unfold, the crisis can have a V-shape with fairly rapid recovery after the restrictions are lifted. Otherwise, especially if the crisis causes significant problems in the banking sector, it may take an L-shape and turn into a prolonged depression. The situation is complicated by the costs of restrictions imposed by the state during the pandemic which are unevenly distributed, and among the most affected companies, there is a very high proportion of small and medium-sized businesses, which are usually not included in the lists of systemically important industries and at the same time do not have a financial cushion, which means a high probability of their bankruptcy during a prolonged (even 3-4 months) suspension of activities.

3.3 Discussion

The author of this study provided a brief forecast of the possible cumulative number of COVID-19 confirmed cases of this epidemic worldwide. Because this epidemic is widespread, the author published two months ahead of the forecast in the time series model. The author forecasted a total number of confirmed cases from April 1, 2020, to June 1, 2020, based on the data model until April 1, 2020. This prediction had a confidence level of around 95%, which was adequate for the prediction. The outcome demonstrated that prediction accuracies and, as a result, multiple-step forecasting were high. Our research found that the longer the training time, the better the forecasting. The model revealed that the width of the prediction intervals decreased on average as more data was included for forecasts. However, if the data were reliable and there was no second transmission, the time series model predicted that the COVID-19 outbreak would have the same number of confirmed cases worldwide. The findings of the study exceeded our expectations. The hypothesis about the effect of countries' percentage growth dynamics on the future dynamics of the total number of confirmed cases in country-followers was confirmed. This problem allowed us to make 4–8-week forecasts for the spread of the epidemic in Azerbaijan, with three predicting intervals. Similar modeling in terms of percentage growth dynamics could be done for other countries with a long enough lead time.

4 Conclusion

Active steps taken by the state are especially important, including the allocation of unconditional transfers (allowing the most affected segments of the population to survive and at the same time creating consumer demand, preventing the crisis from spreading to industries not affected by restrictions), the abolition (or at least a reduction) of taxes payable by small and medium-sized businesses, subject to several conditions, first of all, maintaining employment and paying salaries (which increases the number of those who work during the crisis), the implementation of measures that minimize the costs of companies forced to stop their economic activity during the crisis, for the fastest and full launch of production after the end of restrictive measures. Perhaps, the basic set of measures presented is not ideal in the presence of complete information and sufficient time to make decisions. At the same time, it is incomplete. In particular, it does not include measures already implemented, incl. in the field of medicine, to expand the capacity of the health care system, or to support specific industries such as aviation or tourism [20], [21]. At the same time, in a real situation of a lasting pandemic (as shown by the study), with severe time pressure and existing imperfect institutions, the proposed measures, despite their costly characteristic, will allow accelerating the recovery of the economy and by the end of 2021 to approach pre-crisis monthly production levels, avoiding bankruptcy in a significant share of the business, which threatens much higher costs from the state.

A better understanding of the progress of the epidemic in the country can be obtained by analyzing the progress of the epidemic at the regional level. In conclusion, if the current mathematical model results can be validated within the range provided here, then the social distancing and other prevention and treatment policies that the central and various state governments and people are currently implementing should be continued until no new cases are seen. The migration of urban to rural and rich to poor populations should be closely monitored and controlled. There are many assumptions about population homogeneity in terms of urban/rural or rich/poor that do not capture variations in population density in mathematical models. If several protective measures are not implemented effectively, this rate may be altered. However, the government of Azerbaijan has already taken various protective measures, including the establishment of a quarantine facility, to slow the spread of COVID-19, and we can hope that the country will be successful in slowing the spread of this pandemic.

The study looked into the COVID-19 growth rate in detail and forecasted the number of confirmed cases, intending to inform the public about the situation. The author discovered that the COVID-19 confirmed cases curve would continue to rise, urging everyone to be more aware of the virus. Finally, our most recent data-driven estimates have remained reasonably constant. The time series model predicted the global stage of the outbreak. It was most likely due to the epidemic's wide-ranging influence. The projection was predicated on the assumption that current mitigating efforts would be maintained. Many studies have been conducted for short-term forecasting periods such as 5, 10, and 15 days. In this study, the author used data from the previous month to forecast the next two months. If the data set is large, it can accurately predict long periods. Both the short- and medium-term forecasts capture well the epidemic trajectory across different waves of COVID-19 infections with small relative errors over the forecast horizon. The medium-term forecasts of COVID-19 mortality can be used in conjunction with the short-term forecasts as a useful planning tool as countries continue to relax stringent public health measures implemented to contain the pandemic. Furthermore, the exponential growth model demonstrated excellent accuracy in time series analysis prediction, which previous models could not achieve.

As a result, this model should be used to forecast future analysis of any dataset. The limitations observed during the prediction were a relatively small dataset, and the prediction was based on a pandemic with a high variation in the data set. The output would be more accurate if the dataset were more extensive and less variable. Future researchers can use COVID-19 to investigate prediction models such as artificial neural networks (ANN), Bayesian networks, and Support Vector Machines (SVM). This model can also predict future pandemics and patients with any disease.

References:

[1] Brauer F., Compartmental models in epidemiology. *Mathematical epidemiology*, Springer, Berlin, Heidelberg, 2008, pp. 19−79.

[2] Bailey, Norman T.J., *The mathematical theory of epidemics.* Griffin, 1957, pp. 80-170.

[3] Kermack W., McKendrick A., A contribution to the mathematical theory of epidemics, *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character*, Vol. 115, No. 772, 1927, pp. 700−721.

[4] Anderson R.M, May R.M., *Infectious diseases of humans: dynamics and control*. Oxford University Press; 1992.

[5] Novel Coronavirus Global Research and Innovation Forum: Towards a Research Roadmap. WHO. Available at: www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novelcoronavirus-2019-ncov (accessed: June 15, 2020).

[6] Layne S. P., Hyman J. M., Morens D. M., Taubenberger J. K., New coronavirus outbreak: Framing questions for pandemic prevention. *Sci. Transl. Med*., 2020, vol. 12, Iss. 534, no. eabb1469. https://doi.org/10.1126/scitranslmed.abb1469

[7] Wu J.T., Leung K., Leung G. M., Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. *Lancet*, 2020, vol. 395, Iss. 10225, pp. 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9

[8] Models of Infectious Disease Agent Study. MIDAS Coordination Center. Available at: https://midasnetwork.us/ (accessed: June 15, 2020).

[9] Mandal M., Jana S., Nandi S. K., Khatua A., Adak S., Kar T. K., A model-based study on the dynamics of COVID-19: Prediction and Control. Chaos, Solitons and Fractals, 2020, vol. 136, no. 109889. https://doi.org/10.1016/j.chaos.2020.109889

[10] Fanelli D., Piazza F., Analysis and forecast of COVID-19 spreading in China, Italy, and France. Chaos, Solitons and Fractals, 2020, vol. 134, no. 109761. https://doi.org/10.1016/j.chaos.2020.109761

[11] Bekirosab S., Kouloumpou D., SBDiEM: A new mathematical model of infectious disease dynamics. Chaos, Solitons and Fractals, 2020, vol. 136, no. 109828. https://doi.org/10.1016/j.chaos.2020.109828

[12] Barmparis G., Tsironis G., Estimating the infection horizon of COVID-19 in eight countries with a data-driven approach. Chaos, Solitons and Fractals, 2020, vol. 135, no. 109842. https://doi.org/10.1016/j.chaos.2020.109842

[13] Kondratyev M., Forecasting methods and models of disease spread. *Computer Research and Modeling*, vol. 5, Iss. 5, 2013, pp. 863–882.

[14] Schmidt R., Waligora T., Influenza forecast: Case-Based Reasoning or statistics? *Proceedings of the 11th International Conference on knowledge-based intelligent information and engineering systems. Pt I. Series Lecture Notes in Computer Science*, vol. 4692, 2007, pp. 287–294.

[15] Viboud C., Boelle P., Carrat F., Valleron A., Flahault A., Prediction of the spread of influenza epidemics by the method of analogs. *American Journal of Epidemiology*, vol. 158, Iss. 10, 2003, pp. 996– 1006.

[16] Johns Hopkins Coronavirus Resource Center. Available at: https://coronavirus.jhu.edu/data (accessed: May 29, 2020).

[17] Available at: https://www.worldometers.info/coronavirus

[18] Available at: https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

[19] Akbulaev N., Mammadov I., Aliyev V., Economic impact of COVID-19. *Sylwan*,164, 2020, p. 5.

[20] Akbulaev N., Aliyev V., The Effect of Coronavirus SARS-CoV-2 in the Tourism Industry in Africa. *Journal of Environmental Management and Tourism*, 11.8, 2021, 1938-1947.

[21] Akbulaev N., Mammadov I., Aliyev V., Impact of COVID-19 on the tourism industry (Spain, Italy, and Turkey). *Scientific works of the Free Economic Society of Russia*, 223.3, 2020, 380-390.

[22] Fong, S.J., et al., 2020. Finding an accurate early forecasting model from a small dataset: a case of 2019-nCoV novel coronavirus outbreak. Int. J. Interact. Multimed. Artif. Intell. 6 (1), 132. https://doi.org/10.9781/ijimai.2020.02.002.

[23] Petropoulos, F., Makridakis, S., Forecasting the novel coronavirus COVID-19. *PLoS One* 15 (3), 2020, pp. 1–8. https://doi.org/10.1371/journal.pone.0231236.

[24] Zhan, C., et al., Prediction of COVID-19 spreading profiles in South Korea, Italy, and Iran by data-driven coding. *PLoS One* 15 (7 July), 2020, pp. 1–17. https://doi.org/10.1371/journal.pone.023

[25] Shryock H., Siegel J., Methods and Materials of Demography. *U.S. Dept. of Commerce*, Bureau of the Census Washington, 1973.