Seminar of I. Vekua Institute
of Applied Mathematics
REPORTS, Vol. 49, 2023

THE SPREAD OF CORONAVIRUS (COVID-19): CHALLENGES AND PRACTICAL EXPERIENCE

Gabelaia A.

Abstract. Authors' practical experience in predicting coronavirus (Kovid-19) (for both, the World and Georgia) based on modern mathematical models and the well-known computer program EViews-10 ("Econometric Views") is discussed. The possibility of modifying the developed models of coronavirus prediction, which aims to increase the prediction horizon, is considering.

Keywords and phrases: Coronavirus, prediction, prediction models, prediction horizon.

AMS subject classification (2010): 91B02.

The problem of predicting the spread of coronavirus (COVID-19) in the world, we have been discussing since February 13, 2020 [1-2]. However, given that the virus was mainly spread only in China during this period, we used a logistic function for forecasting [2]. This gave us a pretty good result, considering that the number of infected people in China as of September 15 was 85214! (I.e. within 7 months of forecasting, the forecast error with respect to the real value came out to be only 0.25% !).

This model had the form:

INFIC=(85000*580)/((85000-580)*EXP(-C(1)*@TREND)+580)

where INFIC indicated the total number of infection cases (in the world) for the current moment (total cases), and @TREND was the time variable. (It should be noted that all our basic calculations were performed on the basis of the well-known computer program EViews-10).

However, later, when the virus spread all over the world and the dynamics of its spread became very complicated, we had to use shorter-term forecasting models, in particular, ARIMA (integrated autoregression and moving average) models [4] (with the addition of trend components) and periodically adjust the forecast estimates.

In particular, in the second stage, we tried to predict the total number of infected people for the month of March, already based on the data of January - February (more precisely, 22/01-29/02 2020), based on the aforementioned ARIMA - type models.

The trend model of the infic variable, with an autoregressive term, took the form for this period:

INFIC = 2419.82056382*@TREND

+ [AR(1)=0.934518885427,UNCOND,ESTSMPL="1/23/2020 2/29/2020"],

where AR(1) denoted the first-order autoregressive variable.

It should be noted, that the coefficient of determination of this regression turned out to be very high (of the order of 0.99), the t-statistics of the parameters equal to

6.4, 10.8 and 5.8, respectively, only the Darbin-Watson statistic came out low, which indicated that the regression was not immune to systematic errors. Obviously, its error diagram gave a certain idea about the accuracy of the mentioned model.

As for the accuracy of our prognostic estimates for the first decade of March, the so-called average error of approximation (percentage of error modules with respect to real values) was 2 %! which, obviously, was a good enough indicator.

In addition, at this stage, we found prognostic estimates for the number of patients with this virus for the current moment, ie. for values of the ac variable as well.

In general, similarly, in the period of January - August 2020, we conducted prognostic assessments in 13 stages, for different periods, the duration of which did not exceed one month. It should be noted that overall the accuracy of these prognostic assessments was quite high.

As for the dynamics of the main characteristics of the spread of the virus (in terms of days) in the period of January - August 2020, it looked like this (see Fig. 1-4):

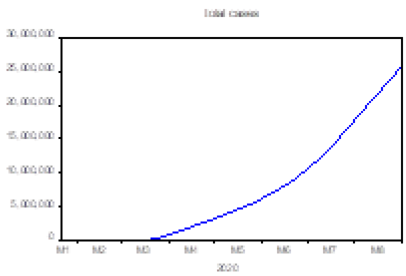

Figure 1: dynamics of the total number of infection cases in the period (infic) of January-August 2020 (in terms of days)

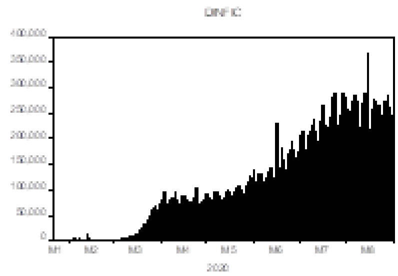

Figure 2: Dynamics of the “speed” d(infic) of the increase in the total number of infection cases in the period January-August 2020 (in terms of days)

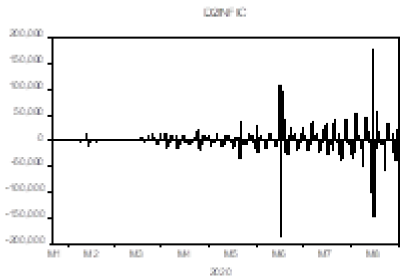

Figure 3: Dynamics of “acceleration” d(infic,2) of the total number of infection cases in the period January-August 2020 (in terms of days)

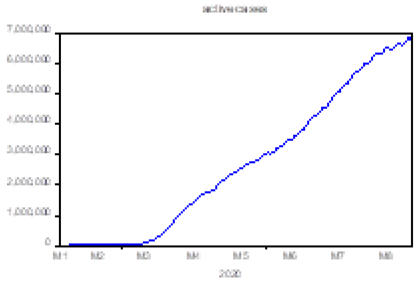

Figure 4: Dynamics of the number of people infected with the virus (ac) in the period January - August 2020 (in terms of days)

However, as practice has shown, the models we used showed high enough accuracy in the perspective of a month at most (then their accuracy dropped). On the other hand, considering that the virus was not “going to stop” in the near term, the problem of increasing the forecasting horizon was on the agenda.

Therefore, it might make sense to consider such a new indicator as, for example, “the average daily increase in the number of infected people during the month”. This would allow us to predict this indicator for a horizon containing several months, especially since according to the central limit theorem of probability theory, the distribution of this indicator should be close to normal, which should somewhat simplify the task of making reliable prognostic assessments for it.

In addition, it was possible to increase the prediction horizon based on an indicator such as the total number of infected people at the end of a period (in this case, a month).

However, this of course meant that the accuracy of such forecasts would have to increase as relevant (months) of information accumulated.

In particular, in the interests of increasing the forecasting horizon, we initially considered the problem of forecasting such an indicator as “average daily increase in the number of infected people during the month” (SDTC).

We tried to predict this indicator for the first quarter of 2021 (although in this case the statistics were very small, and therefore the accuracy of the prediction was not expected!) which, of course, was practically confirmed.

In addition, we tried to forecast this indicator for the second quarter of 2021 (obviously, based on the data of 2020 and the first quarter of 2021). In particular, the trend model of this indicator, with the addition of a moving average, looked like this:

SDTC = 42753.8683336*@TREND

+ [MA(3)=-1,UNCOND,ESTSMPL=”2020M01 2021M03”],

where AR(3) denoted the third-order autoregressive variable.

However, the coefficient of determination of this regression was of the order of 0.9, the t-statistic of the trend parameter was quite high, only the Darbin-Watson statistic was low (of the order of 1.5), which indicated the possibility of systematic error in the model.

As for the second estimate of this indicator (which was initially optimistic, but over time became pessimistic!), his equation had the form:

$$\text{LOG(SDTC)} = \text{-0.0227969509625*}(@TREND)^2 + \text{0.564327466074*@TREND}$$
$$+ \text{ 9.7369576181 } + \text{ [AR(2)=-0.757862290481,MA(2)}$$
$$= \text{-0.662599126347, UNCOND,ESTSMPL="2020M01 2021M03"]},$$

where AR(2) is the second-order autoregression variable, and MA(2) is the second-order moving average variable.

As for the parameters of this regression, its coefficient of determination came out to be of the order of 0.93, the t-statistics of the parameter were quite high, and the Darbin-Watson statistic was of the order of 1.76 (which indicated the possibility of some systematic errors in the model).

As for the quantification of the accuracy of these forecast estimates, the average error of approximation of our average forecast estimates for the second quarter of 2021 was 17 %, which was twice as good as compared to the same figure in the previous quarter!

At the same time, taking into account that the dynamics of the average daily increase in the number of infected people turned out to be quite difficult, we switched to predicting such an indicator as the total number of infected people at the end of the period (in this case, month), which we denoted by tcp. This, in our opinion, is justified at least due to the fact that the basic dynamics of this indicator is much simpler (due to its monotonicity (non-decreasing!)) than the dynamics of the average daily increase in the total number of infected people (in terms of months).

Since we started finding forecast estimates of this quantity in the first days of May 2021, we made its forecast for the second quarter of 2021 (i.e. based on January 2020-March 2021 data), that is, for April (based on the corresponding model), we made an ex post forecast estimate, in order to have a means of comparison with the received prognostic assessment.

In particular, the trend equation of this indicator (TCP), with the addition of the corresponding member of the moving average, had the form:

$$\text{TCP} = \text{-1703180.96828*@TREND} + \text{807507.387172*}(@TREND)^2$$
$$+ \text{ [MA(1)=0.919288114933,UNCOND,ESTSMPL="2020M01 2021M03"]}.$$

However, the coefficient of determination of this regression turned out to be very high (of the order of 0.998). The t-statistics of the parameters were also quite high, only the Darbin-Watson statistic was low (of the order of 1.6).

As for the optimistic prognostic estimates of this indicator, the corresponding autoregressive equation with a trend component had the form:

$$\text{TCP} = \text{1.36442167005*TCP(-1) - 0.480876472847*TCP(-2)}$$
$$+ \text{ 1830732.9788*@}TREND\text{- 4339333.23072},$$

where TCP(-1) and TCP(-2) denote the first and second order autoregressive variables, respectively.

The coefficient of determination of this regression turned out to be very high (of the order of 0.997), the t-statistics of the main parameters were also satisfactory, and the Darbin-Watson statistic was of the order of 1.96, which indicates the high accuracy of the model.

As a quantitative feature of the accuracy of these forecast estimates, the average error of approximation of our average forecast estimates for the second quarter of 2021 was 2.1 %, which is clearly a much better figure than the one obtained above for the average daily increase in the number of infected people.

Note. As our practice showed, it was relatively easy to predict the total number of infected people by the end of the period (month), even compared to predicting the average daily increase in the number of infected people. However, on the other hand, this indicator is a less visible indicator from the point of view of the analysis of the trend of the spread of the pandemic. So, for this purpose, we could use again the magnitude of the average daily increase in the number of infected people, or (easier to calculate!) the increase in the total number of infected people at the end of the period (the "rate of growth") d(tcp), which can also be easily calculated for the obtained prognostic estimates (e.g. for average prognostic assessments). In addition, for a more complete analysis of the dynamic characteristics of the process (just as we did before!) we could consider the rate of "acceleration" of the increase in the total number of infected people during the period d(tcp,2), the calculation of which (if using a computer software package such as Eviews) is very easy, including for prognostic evaluations. Further, we have found TCP forecast estimates for the second half of 2021.

As for the accuracy of the received prognostic assessments of this indicator, the approximation error of the average prognostic assessment for this period amounted to 0.75%, which should be considered a good result.

After that, we were able to find now, TCP forecast estimates for the second half of 2021.

It should be noted that in this case (made 6 months earlier!), the so-called average error of approximation of optimistic prognostic estimates for the period July-December 2021 was only 1.23%, which is a very good indicator in itself!

In addition, the so-called Kovid The Omicron strain, which began to spread in late 2001 and was characterized by dramatically high prevalence rates, made it virtually impossible to obtain more or less accurate prognostic estimates (based on statistical methods) for the first quarter of 2022 (in months).

Indeed, the prognostic estimates for the first half of 2022 based on the data of the end of 2021 turned out to be catastrophically inaccurate (too optimistic!).

In particular, in January 2022, the average daily rate of the total number of infection cases increased to a record 2,913,165, and in February-March, respectively, it decreased to only 2,082,771 and 1,649,391.

Taking into account the above, we tried to refine our forecast estimates for the second quarter of 2022 (taking into account the results of the first quarter). This time our prognostic assessments turned out to be relatively "moderately" inaccurate (however, this time they turned out to be pessimistic!).

As for our prognostic assessments for the III and IV quarters of 2002, this time they turned out to be quite accurate (especially in the III quarter).

In addition, we found prognostic estimates of the spread of Covid for the I and II quarters of the current year 2023. Unfortunately, our prognostic assessments were quite inaccurate this time as well (it turned out to be pessimistic then!).

According to the final report, the dynamics of the total number of infection cases by the end of the period for the period January 2020-June 2023 (in months) is shown in fig. 5, the real dynamics of the average daily increase in the total number of infection cases, for the period February 2020-June 2023 - fig. at 6.

According to the final report, the dynamics of the total number of infection cases by the end of the period for the period January 2020-June 2023 (in months) is shown in fig. 5, the real dynamics of the average daily increase in the total number of infection cases, for the period February 2020-June 2023 - fig. at 6.

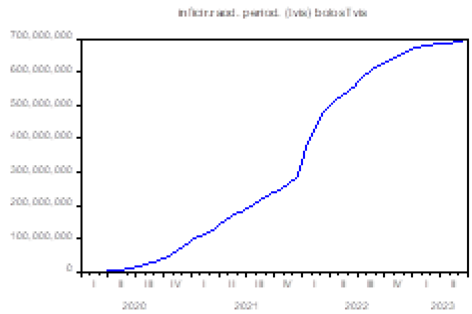

Figure 5: Dynamics of the total number of infection cases by the end of the period for the period January 2020 - June 2023 (in months)

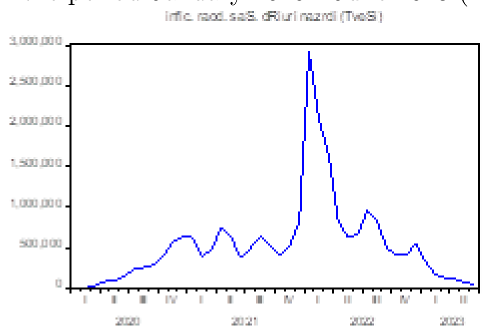

Figure 6: Real dynamics of the average daily increase in the total number of infection cases (in terms of months) for the period February 2020-June 2023

As we can see, the dynamics of the total number of infection cases at the end of the period was stable enough for the end of the given period, which is clearly visible from the real dynamics of the average daily increase in the total number of infection cases in the corresponding period. In particular, starting from January 2022 and ending in June 2023, there was a decreasing trend of this indicator, and according to June data, it amounted to 45,671.

Apparently, therefore, on May 6, 2023, the World Health Organization declared the end of the global health emergency of COVID-19.

In this regard, we also stopped the above-discussed attempts to predict this process. However, according to official statistical data, the pandemic has not ended, and as of September 1, 2023, the total number of inficirebis cases reached 694,630,525, and the number of active cases reached 21,282,956. In addition (by the end of the period) critical cases amounted to 0.2%, and the so-called Mortality of completed cases was 1%. The total number of dead people reached 6, 911, 550.

Note. It should be noted that the d(infic,2) indicator of the "acceleration" of the spread of the virus is not directly found in the global database "COVID-19 CORONAVIRUS PANDEMIC", which we mainly used in this paper, so this indicator is calculated by us (by EViews-10 using). Therefore, it can be considered that the corresponding graphic image represents our small contribution to the description of the process of spreading the virus! The same can be said for the SDTC and TCP indicators that we use above (and which are not directly obtained from the database referenced above)!

Finally, we should note that the working files of EViews developed within the framework of this paper (and its analogue created for Georgia!) can be used for the so-called "ex post" analysis of accuracy of other (alternative methods!) of forecasting.

REFERENCES

1. Gabelaia A. Problem of predicting the spread of Coronavirus (COVID-19). *Semin. Inst. Appl. Math. Rep.*, **46** (2020), 17-26.

2. Gabelaia A. The problem of modifying models for prediction the spread of Coronavirus (COVID-19). *Semin. Inst. Appl. Math. Rep.*, **47** (2021), 26-30.

3. Verhulst P. F. Recherches Mathématiques sur La Loi D'Accroissement de la Population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, **18**, 1 (1845), 1-45.

4. Gabelaia A., Gabelaia L. Econometrical analyzes and Forecasting package EViews fundamentals (Georgian). Tbilisi, 2017.

Received 05.09.2023; revised 27.09.2023; accepted 17.10.2023

Author's address:

A. Gabelaia
Georgian Technical University
77, M. Kostava St., Tbilisi 0175
Georgia
E-mail: agabelaia@gtu.ge